Novel Statistical Methods for Causal Inference Based on Truncated and Censored Time-to-Event Data

Eric Morenz

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Marco Carone, Chair
Lianne Sheppard
Ting Ye

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Novel Statistical Methods for Causal Inference
Based on Truncated and Censored Time-to-Event Data

Eric Morenz

Chair of the Supervisory Committee:

Marco Carone
Department of Biostatistics

In this dissertation, we discuss two settings in which association parameters based on traditional models can be challenging to interpret causally, even when exposure level is randomized. The first setting pertains to human autopsy studies, in which a time-varying outcome is only measured at death. When the exposure affects survival, the distribution of outcome measurement times differs across exposure levels, and can produce misleading comparisons. We propose to study the causal effect of the exposure on the outcome process using a contrast inspired by the mediation literature and that we refer to as the natural-time direct effect. As a first step, we derive nonparametric debiased machine learning approaches for inference on survival integrals using left-truncated right-censored data. We then use these methods to assess the effect of the APOE-4 gene on autopsy measures of Braak staging using data from the Adult Changes in Thought study. In the second setting, a point-exposure of varying doses occurs at a random time, and we wish to characterize the (possibly exposure time-specific) effect of dose on survival. In this context, standard survival models are often not amenable to interpretably addressing the scientific question at hand. As an alternative, we propose to use an accelerated residual failure time model. We develop methods for inference on model parameters indexing a parametric accelerated residual failure time model using left-truncated right-censored data, and provide a causal interpretation under natural causal conditions. Using data from the Life Span Study, a long-term prospective cohort study of survivors of the atomic bombings of Hiroshima and Nagasaki, we use these methods to study how a point-exposure to radiation of differing doses affects survival.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to everyone who supported and encouraged me throughout the completion of this dissertation. In particular, to the chair, my advisor Dr. Marco Carone, I would not have taken this journey with anyone else. I am deeply grateful for the support and encouragement I received along the way.

A heartfelt thanks to the members of my dissertation committee, Drs. Lianne Sheppard, Ting Ye, Benjamin French, and Steve Mooney, for their insightful feedback and constructive criticism. Your kindness and engagement greatly enriched the quality of this work.

This work would not have been possible without the help of numerous individuals from the University of Washington and Radiation Effects Research Foundation that contributed directly or indirectly to this project. In particular, thank you to Drs. Benjamin French, Amanda Phipps, John Cologne, Eric Grant, Alina Brenner, Ritsu Sakata, and Hiromi Sugiyama. A special thank you to the colleagues, contributors, and mentors at the University of Washington that made the past few years some of the most enjoyable to date—Brian Williamson, Yiqun Chen, Steven Wang, Spencer, Hansen, Sijia Li, and Xiudi Li. The collaborative and supportive spirit within our community has been a constant source of motivation.

To my family and friends, thank you for your unwavering encouragement, understanding, and patience. Your persistent support sustained me through the ups and downs; your joy reminded me to savour the little moments together. To my parents, your encouragement gave me the confidence to take on this challenge. To my siblings, your thoughtfulness reminded me to take time off and enjoy the moment. Each of you contributed in your own way; without you I would not have made it where I am today.

Finally, to Maria Gracia Gauto, your love and understanding has been everything. Your belief in me fueled my determination to see this through to completion. This dissertation is as much yours as it is mine, and I am profoundly grateful for each contribution, large and small. Thank you for seeing through this significant chapter in my life.

DEDICATION

To my mentors, my friends, and my family

Chapter 1

DEBIASED MACHINE LEARNING FOR SURVIVAL FUNCTIONALS BASED ON LEFT-TRUNCATED RIGHT-CENSORED DATA

Learning causal effects on time-to-event endpoints can be challenging because survival times may be partially observed due to censoring and systematically biased due to truncation. In this work, we present nonparametric estimators of a counterfactual survival time distribution for use when the data are subject to covariate-dependent left truncation and right censoring. Our inferential procedures explicitly allow the integration of flexible machine learning tools for nuisance estimation, and enjoy certain robustness properties. The approach we propose can be directly used to make pointwise or uniform inference on smooth summaries of the counterfactual survival time distribution, and can be valuable even in the absence of interventions, when summaries of a marginal survival distribution are of interest. We showcase how our procedures can be used to learn a variety of inferential targets, and illustrate their performance in simulation studies.

1.1 Introduction

In biomedical studies, the outcome of interest is often the time elapsed between an initiating event and a terminating event. For example, investigators may wish to study the time from some exposure or treatment (e.g., administration of vaccine) until a particular clinical event (e.g., onset of symptomatic disease). In particular, they may be interested in determining the effect of a treatment on the event time. Even in the context of a randomized trial, in which the design ensures that the relationship between treatment and event time is unconfounded, the analysis of time-to-event data remains challenging because event times are typically only partially observed in some study participants. Indeed, some participants may exit the study during the course of follow-up, or may not yet have experienced the event of interest by the end of the study, in which case their event times are right-censored. Right-censoring complicates the identification of the time-to-event distribution — notably, required assumptions about the censoring mechanism may fail to hold even in randomized trials — and ensuing procedures for assumption-lean statistical inference are also much more involved. The problems that arise due to incomplete observation of terminating events are compounded when the study does not include randomization, in which case appropriate deconfounding, whenever possible, must also be incorporated into statistical procedures.

In many observational studies, in addition to right censoring, the available data are subject to left truncation, wherein only participants for whom the event time is larger than a corresponding truncation time can be recruited into the study. This may occur, for example, due to delayed entry into a prospective study or to the use of a cross-sectional sampling scheme. Unlike censoring, which results in partially observed data but has no bearing on who may be sampled, truncation implies a restriction on the sampling mechanism, and usually renders the sampling population biased relative to the target population. Indeed, truncation induces systematic selection bias into the study design, with an over-representation of participants with a longer event time. Failure to account for left truncation can result in severely biased inferences and misleading scientific conclusions.

While the field of survival analysis is mature, with many decades of rigorous methodological developments pertaining to the analysis of left-truncated right-censored data, most existing works have relied heavily either on semiparametric and parametric modeling assumptions, or on strong uninformativeness assumptions about the censoring and truncation mechanisms. Furthermore, while there has been a growing literature at the intersection of survival analysis and causal inference, the focus has been almost exclusively on data subject to right censoring without truncation — see, e.g., Westling et al. (2023) for a sampling of such existing methods. In this work, we contribute to addressing this gap by developing novel nonparametric statistical methods for estimating causal effect summaries with left-truncated right-censored data.

In the developments below, we propose debiased machine learning techniques for nonparametric inference on smooth summaries of a counterfactual time-to-event distribution using left-truncated right-censored data. The class of summaries we consider is broad and includes, in particular, commonly reported estimands, such as survival probabilities, restricted means, and percentiles, as well as more complex functionals. Notably, the methods we develop allow informative censoring and truncation insofar as can be explained by recorded covariates — in other words, the censoring and truncation mechanisms may be covariate-dependent. They also allow the use of flexible learning algorithms for estimating involved nuisance functions without compromising the calibration of resulting statistical inferences. This is desirable since the use of such algorithms can mitigate the risk of systematic bias possibly resulting from inconsistent estimation of such nuisance functions.

We note that Wang et al. (2022) recently made important advances in the development of debiased machine learning methods for use with left-truncated data. However, their work focuses on inference for a marginal (rather than counterfactual) survival function, and their procedures neither facilitate flexible estimation of the censoring mechanism nor generally achieve the efficiency bound in the presence of right censoring. As such, our work extends theirs by both including consideration of summaries of a counterfactual time-to-event distribution and restoring efficiency even in the presence of both left truncation and

right censoring. We also note that our work can be seen as a natural generalization of the recent work of Westling et al. (2023), which develops flexible techniques for nonparametric efficient inference on a counterfactual survival function using right-censored data without truncation. While traditional risk set-based methods for the analysis of right-censored data can often be effortlessly extended to the analysis of left-truncated right-censored data, this is not necessarily the case for other methods, including those based on influence functions or Neyman orthogonalization, as in this work.

This chapter is organized as follows. In Section 1.2, we define the estimand of interest and discuss its identification in contexts in which the time-to-event random variable is observed subject to possibly covariate-dependent left truncation and right censoring. In Section 1.3, we derive a linearization of the survival integral parameter viewed as a functional of the observed data distribution. In Section 1.4, we use this linearization to construct two distinct cross-fitted inferential procedures that explicitly allows the incorporation of machine learning methods. In Section 1.5, we establish certain large-sample properties of the proposed procedures, including both pointwise and uniform distributional results, and extend these results to a larger class of smooth survival functionals. In Section 1.6, we discuss several analytic examples of our general results, whereas in Section 1.7, we present results from numerical experiments to illustrate the operating characteristics of our procedures. We conclude with final remarks in Section 1.8.

1.2 Statistical setup and identification

1.2.1 Notation and examples

The ideal data unit is $X := (T, C, W, A, Z) \sim P_{X,0}$, where $Z \in \mathcal{Z} \subseteq \mathbb{R}^p$ denotes a vector of baseline covariates, $A \in \{0, 1\}$ is a binary exposure level indicator, $W \in [0, \infty)$ and $C \in (0, \infty)$ are the truncation and censoring times, respectively, and the event time (or survival time) is $T \in (0, \infty)$. Here, $P_{X,0}$ denotes the true (unknown) distribution of X in the *target* population. For a fixed $a_0 \in \{0, 1\}$ and known kernel function $\varphi : \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$, we

begin by studying inference on the survival integral

$$\theta_0 := \iint \varphi(t,z)\, F_{X,0}(dt \,|\, a_0, z)\, H_{X,0}(dz) \;, \tag{1.1}$$

where we define $F_{X,0}(t \,|\, a, z) := P_{X,0}\,(T \leq t \,|\, A = a, Z = z)$ and $H_{X,0}(z) := P_{X,0}\,(Z \leq z)$ pointwise. We note here that the ideal data unit may have been taken to simply be (T, A, Z) since the estimand of interest depends only on the conditional distribution of T given (A, Z) and on the marginal distribution of Z, and if neither truncation nor censoring act on the data unit, the value of (C, W) is irrelevant. Nevertheless, for notational convenience in developments below, we define X to also include (C, W).

Survival integrals encompass several estimands of interest at the intersection of survival analysis and causal inference. Under typical causal conditions, including that, within each stratum of Z, the counterfactual event time $T(a_0)$ corresponding to the intervention that sets $A = a_0$ is independent of A, and that $A = a_0$ occurs with positive probability, θ_0 identifies the counterfactual mean value $\mathbb{E}_0\left[\varphi(T(a_0), Z)\right]$ computed under the joint distribution $\mathbb{P}_0$ of $(T(a_0), Z)$. Various choices of φ yield different causal estimands of practical interest. As a special case, by considering the exposure A to be degenerate at a_0, the survival integral trivially corresponds instead to moments of the joint distribution of (T, Z), estimands that arise in traditional survival analyses. Specific examples of estimands that motivate our work and are later discussed include:

(1) the marginal survival probability $P_{X,0}\,(T > t)$;

(2) the Brier score $E_{X,0}\left\{I(T \geq \tau) - b(Z)\right\}^2$ of a given function $b : \mathcal{Z} \to [0, 1]$ for predicting survival at time τ (Brier et al., 1950; Gerds and Schumacher, 2006);

(3) the counterfactual survival probability $\mathbb{P}_0\left\{T(a_0) > t\right\}$ (Westling et al., 2023).

Later, we build upon our results on survival integrals to develop inferential methods for nonlinear survival functionals. This extension allows us to tackle many more estimands of interest. Examples of such estimands that we study in greater detail include:

(4) the median counterfactual event time (Díaz, 2017);

(5) a model-agnostic measure of dependence of $T(a_0)$ on a_0 (Vansteelandt et al., 2022).

Example 1 is the primary target of inference in classical survival analysis, although here we wish to allow possibly covariate-dependent censoring and truncation. Example 2 arises in the evaluation of prediction models in survival analysis, and emphasizes the value of allowing the kernel value $\varphi(t, z)$ to depend on both t and z. Examples 3 and 4 are commonly reported summaries of the counterfactual survival distribution. Example 5 is a novel parsimonious measure of the causal effect of A on T inspired by recent work on assumption-lean Cox regression (Vansteelandt et al., 2022).

We refer to X as an ideal data unit because statistical inference would be straightforward if it were directly observed. However, in practice, this is rarely the case. In many prospective cohort studies, the sampling distribution of X is a systematically biased version of its target distribution due to left-truncation resulting, for example, from delayed entry or cross-sectional sampling. Additionally, only a coarsened version O of X is observed due to right-censoring. Thus, in order to derive procedures for statistical inference, we must first represent θ_0, which is explicitly defined as a summary of the *ideal data* distribution $P_{X,0}$, as a summary of the *observed data* distribution P_0, thereby establishing identification under suitable conditions. We refer to the distribution P_0 as defining the *observable* population.

1.2.2 Identification

The observed data structure is $O := (Y, \Delta, W, A, Z) \sim P_0$ with $Y := \min(T, C)$ and $\Delta := I(T \leq C)$, and results from left-truncation and right-censoring of the ideal data unit X. Specifically, only individuals with $Y \geq W$ can be sampled — those who are neither censored nor experience the terminating event before possible recruitment into the study — and the event time T is subject to right-censoring by C. We assume that $C \geq W$ with $P_{X,0}-$probability one, since we are interested primarily in settings in which censoring is a study-induced nuisance and only operates on individuals who can possibly be recruited into

the study. As such, the sampling conditions $Y \geq W$ and $T \geq W$ are equivalent. The observed data distribution P_0 is obtained from the target population distribution $P_{X,0}$ through the relationship

$$P_0\,(dy, d\delta, dw, da, dz)$$
$$= \frac{I(y \geq w)}{P_{X,0}\,(T \geq W)} \left[\delta \int_{c \geq y} P_{X,0}\,(dy, dc, dw, da, dz) + (1-\delta) \int_{t \geq y} P_{X,0}\,(dt, dy, dw, da, dz) \right] \nu(d\delta)$$

for $\delta \in \{0,1\}$, where ν is the counting measure on $\{0,1\}$. As indicated above, even in the absence of censoring, the sampling distribution P_0 of the observed data unit does not coincide with the target distribution $P_{X,0}$ because individuals with $T < W$ are systematically excluded. Individuals with larger values of T are therefore over-represented in the observable population relative to the target population. Throughout this chapter, the observed data consist of n independent draws $O_1, O_2, \ldots, O_n$ from P_0.

We now consider the problem of recovering $P_{X,0}$ from P_0 on relevant portions of its support, some of which may not be fully recoverable. For example, right-censoring often precludes the identification of the right tail of the time-to-event distribution. Nevertheless, θ_0 may still be identified. To formalize these issues, we first define for a generic random variable B the lower and upper support bounds

$$\begin{aligned} \underline{\tau}_B(a,z) &:= \sup\{u : P_{X,0}\,(B \geq u \,|\, A = a, Z = z) = 1\}\,; \\ \overline{\tau}_B(a,z) &:= \sup\{u : P_{X,0}\,(B \geq u \,|\, A = a, Z = z) > 0\}\,. \end{aligned}$$

We also denote by $\overline{\tau}_C(a,z) := \sup\{u : P_{X,0}\,(C \geq u \,|\, A = a, Z = z, T \geq W) > 0\}$ the upper bound for the support of the censoring distribution, and by $\pi_{X,0}(z) := P_{X,0}(A = a_0 \,|\, Z = z)$ the propensity score for each $z \in \mathcal{Z}$. We make the following support recovery conditions for identifiability:

(A1) for $a \in \{0,1\}$ and $P_{X,0}-$almost every value $z \in \mathcal{Z}$, it holds that:

(i) $\underline{\tau}_W(a,z) \leq \underline{\tau}_T(a,z)\,;$

(ii) $\overline{\tau}_W(a,z) + \alpha \leq \min\{\overline{\tau}_T(a,z), \overline{\tau}_C(a,z)\}$ for some $\alpha > 0$;

(iii) $t \mapsto \varphi(t,z)$ is constant for $t \geq \overline{\tau}_C(a,z)$;

(A2) for $P_{X,0}-$almost every value $z \in \mathcal{Z}$, it holds that $\pi_{X,0}(z) > 0$.

These conditions can be interpreted heuristically as follows. First, if $\underline{\tau}_W(a,z) > \underline{\tau}_T(a,z)$, then individuals with exposure $A = a$, covariate vector $Z = z$ and event time $T = t$ such that $\underline{\tau}_T(a,z) \leq t < \underline{\tau}_W(a,z)$ are systematically excluded in the observable population. As such, the target conditional time-to-event distribution function $F_{X,0}(t \,|\, a,z)$ cannot be identified for any $t > 0$, and neither can θ_0 if the set of such values has positive $P_{X,0}-$probability. Second, if $\min\{\overline{\tau}_T(a,z), \overline{\tau}_C(a,z)\} < \overline{\tau}_W(a,z)$, then individuals with exposure level $A = a$, covariate vector $Z = z$ and truncation time $W = w$ such that $\min\{\overline{\tau}_T(a,z), \overline{\tau}_C(a,z)\} \leq w < \overline{\tau}_W(a,z)$ are systematically excluded in the observable population. As such, the right tail of the target conditional truncation distribution function cannot be identified. This is problematic because, as we will see below, identification of the marginal covariate distribution $H_{X,0}$ — and thus of θ_0 — hinges on that of the conditional truncation distribution. Third, for any $z \in \mathcal{Z}$, $F_{X,0}(t \,|\, a,z)$ can only be identified up to $\overline{\tau}_C(a,z)$ since values of T above $\overline{\tau}_C(a,z)$ can never be observed, and so, unless $t \mapsto \varphi(t,z)$ is constant for $t > \overline{\tau}_C(a,z)$ and $P_{X,0}-$almost every $z \in \mathcal{Z}$, θ_0 also cannot typically be identified. These facts motivate the need for condition (A1). Finally, it is necessary that $\pi_{X,0}(z) > 0$ in order to be able to learn $F_{X,0}(t \,|\, a,z)$ without relying on extrapolating assumptions, since otherwise no inference could ever be made from the subpopulation defined by $(A,Z) = (a,z)$; this motivates condition (A2).

Beyond support recovery conditions, identification hinges fundamentally on the vector Z of baseline covariates being sufficiently rich to account for any dependence between T and (C,W). Specifically, we introduce the following additional conditions on the censoring and truncation mechanisms:

(B1) T and W are independent given (A,Z) $P_{X,0}-$almost surely;

(B2) T and C are independent given (W,A,Z) and $T \geq W$ $P_{X,0}-$almost surely.

We note that distributional constraints on C are only imposed in the observable population, that is, the subpopulation of individuals for whom $T \geq W$. In fact, C need not even be defined for individuals with $T < W$ since censoring only ever affects those with $T \geq W$. Under conditions (A1)–(A2) and (B1)–(B2), $P_{X,0}$ may be expressed in terms of P_0. For $0 < t < \min\{\overline{\tau}_T(z), \overline{\tau}_C(z)\}$, the target conditional distribution function $F_{X,0}(t \,|\, a_0, z)$ is identified via conditional product-integration (Gill and Johansen, 1990) by

$$\tilde{F}_0(t \,|\, a_0, z) := 1 - \prod_{u \leq t} \left[1 - \frac{F_{1,0}(du \,|\, a_0, z)}{R_0(u \,|\, a_0, z)}\right],$$

where $F_{1,0}(u \,|\, a, z) := P_0\left(Y \leq y, \Delta = 1 \,|\, A = a, Z = z\right)$ is an observable conditional sub-distribution function and $R_0(u \,|\, a, z) := P_0\left(Y \geq u \geq W \,|\, A = a, Z = z\right)$ is an observable conditional at-risk probability. As indicated above, the target conditional truncation distribution function, defined pointwise as $G_{X,0}(w \,|\, a, z) := P_{X,0}\left(W \leq w \,|\, A = a, Z = z\right)$, is needed to recover the target covariate distribution. For any $w \geq 0$, it can be expressed as $G_{X,0}(w \,|\, a, z) \propto_w \int_{u \leq w} S_{X,0}(u \,|\, a, z)^{-1} G_0(du \,|\, a, z)$, where we denote the observable conditional truncation distribution function by $G_0(w \,|\, a, z) := P_0\left(W \leq w \,|\, A = a, Z = z\right)$, the survival function corresponding to $F_{X,0}$ by $S_{X,0} := 1 - F_{X,0}$, and $\propto_w$ refers to proportionality in w for fixed a and z. The identification of $F_{X,0}$ over $[0, \min\{\overline{\tau}_T(a, z), \overline{\tau}_C(a, z)\})$ — and thus, by condition (A1), over its subset $[0, \overline{\tau}_W(a, z))$ — then implies that of $G_{X,0}$. The target exposure-covariate distribution function $J_{X,0}(a, z) := P_{X,0}\left(A \leq a, Z \leq z\right)$ can be expressed as a reweighted version of its observable counterpart; indeed, we have that

$$\begin{aligned} J_{X,0}(da, dz) &\propto_{a,z} \frac{J_0(da, dz)}{\int S_{X,0}(u \,|\, a, z)\, G_{X,0}(du \,|\, a, z)} \\ &\propto_{a,z} J_0(da, dz) \int \frac{G_0(du \,|\, a, z)}{1 - \tilde{F}_0(u \,|\, a, z)} \end{aligned} \tag{1.2}$$

with $J_0(a, z) := P_0\left(A \leq a, Z \leq z\right)$ denoting the observable exposure-covariate distribution function; here, $\propto_{a,z}$ refers to proportionality in (a, z). In particular, (1.2) implies an identification $\tilde{H}_0$ of the target covariate distribution $H_{X,0}$ using that $H_{X,0}(dz) = \int_a J_{X,0}(da, dz)$ for

each z. These expressions suffice to identify θ_0 as the summary $\iint \varphi(t,z)\,\tilde{F}_0(dt\,|\,a_0,z)\,\tilde{H}_0(dz)$ of the observed data distribution P_0. Rather than focusing on the ideal data distribution of the censoring random variable, which as discussed earlier need not even be defined, we note that the observable conditional censoring survival function Q_0, defined pointwise as $Q_0(c\,|\,w,a,z) := P_0\,(C > c\,|\,W = w, A = A, Z = z)$ can be identified using conditional product-integration as used in $\tilde{F}_0$ but without truncation and reversing the value of Δ to $1-\Delta$. Additional details on these identification results are provided in Part A of the Appendix A.

1.3 Study of the target parameter

The above identification formulas motivate us to study the observed data parameter

$$\Psi : P \mapsto \iint \varphi(t,z)\,\tilde{F}_P(dt\,|\,a_0,z)\,\tilde{H}_P(dz)\ . \tag{1.3}$$

Here, $\tilde{F}_P$ is defined pointwise as

$$\tilde{F}_P(t\,|\,a_0,z) := 1 - \prod_{u \le t}\left[1 - \frac{F_{1,P}(du\,|\,a_0,z)}{R_P(u\,|\,a_0,z)}\right],$$

where $F_{1,P}$ and R_P are defined pointwise as $F_{1,P}(u\,|\,a,z) := P\,(Y \le u, \Delta = 1\,|\,A = a, Z = z)$ and $R_P(u\,|\,a,z) := P\,(Y \ge u \ge W\,|\,A = a, Z = z)$, respectively. Further, $\tilde{H}_P$ is defined pointwise as $\tilde{H}_P(dz) := \int_a \bar{\gamma}_P(a,z)\,J_P(da,dz)$ with $J_P(a,z) := P\,(A \le a, Z \le z)$, $\gamma_P(a,z) := \int \tilde{S}_P(w\,|\,a,z)^{-1} G_P(dw\,|\,a,z)$, $\gamma_P := \iint \gamma_P(a,z)\,J_P(da,dz)$ and $\bar{\gamma}_P(a,z) := \gamma_P(a,z)/\gamma_P$, where $\tilde{S}(t\,|\,a,w) := 1 - \tilde{F}(t\,|\,a,w)$ is the survival function corresponding to $\tilde{F}_P$ and $G_P(w\,|\,a,z) := P\,(W \le w\,|\,A = a, Z = z)$ is the observable conditional truncation distribution function. Under conditions (A1)–(A2) and (B1)–(B2), the survival integral θ_0 is identified by $\psi_0 := \Psi(P_0)$. Thus, in the remainder of this chapter, we focus on developing inferential methods for ψ_0 and related estimands.

We wish to employ flexible learning strategies to avoid unnecessarily strong modeling

assumptions on the data-generating mechanism P_0. As such, in order to carry out valid nonparametric efficient inference, we develop debiased machine learning methods for this problem. As a first step, we derive a linearization of the parameter mapping Ψ around P_0 based on the nonparametric efficient influence function of Ψ at $P = P_0$ (Pfanzagl, 1982). This linearization is critical for guiding the construction of our estimation procedure and elucidating the conditions under which this procedure has desirable statistical properties.

Before tackling the problem in its generality, it is instructive to first examine the simpler setting in which the support $\mathcal{Z}$ of the covariate vector is finite. In such case, for any fixed $z_0 \in \mathcal{Z}$, the inner integral $\int \varphi(t, z)\, F_0(dt \,|\, a_0, z_0)$ can be estimated nonparametrically under the conditions we have introduced so far using the stratum-specific Kaplan-Meier integral $\int \varphi(t, z)\, F_n(dt \,|\, a_0, z_0)$, where $F_n(t \,|\, a_0, z)$ denotes the Kaplan-Meier estimator of $F_0(t \,|\, a_0, z_0)$ computed using only data from stratum $(A, Z) = (a_0, z_0)$. Under certain regularity conditions, this stratum-specific Kaplan-Meier integral can be shown to be regular and asymptotically linear (Reid, 1981b; Stute, 1994) with influence function given by

$$(z, a, w, \delta, y) \mapsto \frac{I(a = a_0, z = z_0)}{P\,(A = a_0, Z = z_0)} \phi_{\mathrm{KM},P}(L_{P,\varphi})(z, a, w, \delta, y) \ ,$$

where for any P we denote by $L_{P,\varphi} : (y, a, z) \mapsto \int_y^\infty \tilde{S}_P(u \,|\, a, z) \varphi(du, z)$ and we define pointwise, for any function $m : \mathcal{Y} \times \{0, 1\} \times \mathcal{Z} \to \mathbb{R}$,

$$\phi_{\mathrm{KM},P}(m)(z, a, w, \delta, y) := -\frac{\delta m(y, a, z)}{R_P(y \,|\, a, z)} + \int I_{[w,y]}\,(u) \frac{m(u, a, z)}{R_P(u \,|\, a, z)}\, \tilde{\Lambda}_P(du \,|\, a, z)$$

with $\tilde{\Lambda}_P(t \,|\, a, z) := \int_{u \le t} \tilde{S}_P(u \,|\, a, z)^{-1} \tilde{F}_P(du \,|\, a, z)$ denoting the cumulative hazard function corresponding to $\tilde{F}_P$. In our constructions and theoretical results below, the function $\phi_{\mathrm{KM},P}$ appears prominently, as we will now see.

Our linearization results involve additional notation that we now introduce. We define the observable survival regression $\mu_P(z) := \int \varphi(t, z)\, \tilde{F}_P(dt \,|\, a_0, z)$, the observable propensity

score $\pi_P(z) := P(A = a_0 \,|\, Z = z)$, and the partial truncation weight function

$$\gamma_{P,\natural}(y, a, z) := \int I_{[y,\infty)}(u)\, \tilde{S}_P(u \,|\, a, z)^{-1} G_P(du \,|\, a, z) \ .$$

As we establish in the following theorem, the nonparametric linearization of Ψ hinges critically on the nonparametric efficient influence function of Ψ at P, which can be written as $\phi_P := \phi_{1,P} + \phi_{2,P}$ with

$$\begin{aligned}
\phi_{1,P} &: (z, a, w, y, \delta) \mapsto \frac{I(a = a_0)}{\pi_P(z)} \bar{\gamma}_P(z) \phi_{\mathrm{KM},P}(L_{P,\varphi})(z, a, w, y, \delta) \\
\phi_{2,P} &: (z, a, w, y, \delta) \mapsto \frac{\mu_P(z) - \Psi(P)}{\gamma_P} \left[\frac{1}{\tilde{S}_P(w \,|\, a, z)} - \phi_{\mathrm{KM},P}(\gamma_{P,\natural})(z, a, w, y, \delta) \right] ,
\end{aligned}$$

and where $\bar{\gamma}_P(z) := \sum_a \bar{\gamma}_P(a, z) \pi_P(a \,|\, z)$. The nonparametric linearization of $\Psi(P)$ around $P = P_0$ involves a second-order remainder term that can be written as $R(P, P_0) := R_1(P, P_0) + R_2(P, P_0) + R_3(P, P_0) + R_4(P, P_0)$ with

$$\begin{aligned}
R_1(P, P_0) &:= \int L_{P,\varphi}(y, a, z) \left\{ \frac{\pi_0(z) \bar{\gamma}_P(z) \nu_P(y, a, z)}{\pi_P(z) \bar{\gamma}_0(z) \nu_0(y, a, z)} - 1 \right\} \left(\frac{\tilde{S}_0}{\tilde{S}_P} - 1 \right) (dy \,|\, a_0, z)\, \tilde{H}_0(dz) \,; \\
R_2(P, P_0) &:= \int \xi_P(z) \gamma_{P,\natural}(y, a, z) \left\{ 1 - \frac{\nu_P(y, a, z)}{\nu_0(y, a, z)} \right\} \left(\frac{\tilde{S}_0}{\tilde{S}_P} - 1 \right) (dy \,|\, a, z)\, \tilde{H}_0(dz) \,; \\
R_3(P, P_0) &:= \int \xi_P(z) \left[\frac{1}{\tilde{S}_0(w \,|\, a, z)} - \frac{1}{\tilde{S}_P(w \,|\, a, z)} \right] (G_P - G_0)(dw \,|\, a, z)\, J_0(da, dz) \\
&\quad - \int \xi_P(z) \left\{ \tilde{S}_P(y \,|\, a, z) - \tilde{S}_0(y \,|\, a, z) \right\}^2 \frac{\tilde{G}_P(dw \,|\, a, z)}{\tilde{S}_0(w \,|\, a, z) \tilde{S}_P^2(w \,|\, a, z)}\, J_0(da, dz) \,; \\
R_4(P, P_0) &:= \left(\frac{\gamma_0 - \gamma_P}{\gamma_0} \right) \{ R_{4,1}(P, P_0) + R_{4,2}(P, P_0) + R_{4,3}(P, P_0) \} \ ,
\end{aligned}$$

where we define

$$\begin{aligned}
R_{4,1}(P, P_0) &:= \iint \xi_P(z) \{ \gamma_P(a, z) - \gamma_0(a, z) \}\, J_0(da, dz) \,; \\
R_{4,2}(P, P_0) &:= \int \sum_{a \in \{0,1\}} \xi_P(z) \gamma_P(a, z) \{ \pi_P(a \,|\, z) - \pi_0(a \,|\, z) \}\, H_0(dz) \,;
\end{aligned}$$

$$R_{4,3}(P, P_0) := \int \sum_{a \in \{0,1\}} \xi_P(z) \pi_P(a \mid z) \gamma_P(a, z) \left(H_P - H_0\right)(dz) ;$$

and also write $\xi_P(z) := \{\mu_P(z) - \Psi(P)\}/\gamma_P$ and $\nu_P(y, a, z) := \tilde{S}_P(y \mid a, z)/R_P(y \mid a, z)$.

Theorem 1. *Suppose that conditions (A1)–(A2) and (B1)–(B2) hold. Then, the survival integral parameter $P \mapsto \Psi(P)$ is pathwise differentiable in a nonparametric model with efficient influence function ϕ_P, and for each P, the linearization*

$$\Psi(P) - \Psi(P_0) = \int \phi_P(o) \left(P - P_0\right)(do) + R(P, P_0) \ .$$

We note that the efficient influence function ϕ_P we have provided above agrees with existing results for special cases. In the absence of left-truncation, it coincides with results provided in Gerds et al. (2017) for a general survival integral — this is seen by taking the truncation distribution to be degenerate at zero, in which case $\phi_{2,P}$ simplifies to $(z, a, w, y, \delta) \mapsto \mu_P(z) - \Psi(P)$ — and in Westling et al. (2023) for a counterfactual survival probability, obtained by taking $\varphi(t, z) = I(t \leq t_0)$ for a fixed value t_0. In the absence of right-censoring and any treatment intervention, our result agrees with results presented in Wang et al. (2022) for a marginal survival probability.

1.4 Proposed estimation procedure

Equation (1.3) expresses ψ_0 in terms of components of the observed data distribution P_0, which are themselves functions of components of the ideal data distribution $P_{X,0}$. Thus, an estimator of ψ_0 can thus be obtained by substituting an estimator of relevant components of P_0 into (1.3), or an estimator of relevant components of $P_{X,0}$, namely $F_{X,0}$ and $H_{X,0}$, into (1.1). Additional parametrizations of P_0 — for example, using a combination of components of P_0 and of $P_{X,0}$ — can also be considered, each leading to a strategy for estimating ψ_0 with relative advantages and disadvantages. Here, we consider a particular parametrization that we believe provides a balance between implementability and desirable statistical properties,

as we elaborate below. This parametrization is characterized by the following combination of components of P_0 and $P_{X,0}$:

1. the observable conditional covariate distribution function H_0;
2. the observable conditional exposure probability π_0;
3. the observable conditional truncation distribution function G_0;
4. the observable conditional censoring survival function Q_0;
5. the target conditional time-to-event distribution function $F_{X,0}$ (equivalently, survival function $S_{X,0}$), considered only on the observable region $[0, \min\{\overline{\tau}_T(a_0, z), \overline{\tau}_C(a_0, z)\})$ for each covariate value $z \in \mathcal{Z}$.

For notational convenience, we denote the vector $(H_0, \pi_0, G_0, Q_0, F_{X,0})$ of nuisance functions as η_0. We note first that this is indeed a valid parametrization in the sense that two observed data distributions are the same if and only if they agree in this parametrization. We also note that components of this parametrization are variationally-independent in the sense that fixing the value of a subset of components of η_0 does not constrain the values that the remaining components of η_0 can take. The survival integral value ψ_0 can be expressed in terms of η_0 as

$$\frac{\iiint \varphi(t, z)\, F_{X,0}(dt \,|\, a_0, z) \int S_{X,0}(u \,|\, a, z)^{-1} G_0(du \,|\, a, z)\, J_0(da, dz)}{\iiint S_{X,0}(u \,|\, a, z)^{-1} G_0(du \,|\, a, z)\, J_0(da, dz)} \tag{1.4}$$

with $J_0(da, dz) = [(1-a)\{1-\pi_0(z)\} + a\pi_0(z)]\, H_0(dz)\, \nu(da)$ itself a function of η_0. Similarly, the efficient influence function $\phi_0 := \phi_{P_0}$ of $P \mapsto \Psi(P)$ under sampling from P_0 can be expressed as a function of η_0 using the fact that we can write

$$R_0(u \,|\, a, z) := S_{X,0}(u \,|\, a, z) \int_0^u \frac{Q_0(u \,|\, w, a, z)}{S_{X,0}(w \,|\, a, z)}\, G_0(dw \,|\, a, z)\ .$$

We write $\psi_0 = \psi_{\eta_0}$ and $\phi_0 = \phi_{\eta_0}$ to emphasize that ψ_0 and ϕ_0 can be computed based on η_0. This parametrization is convenient for the purpose of estimation since, on one hand, H_0, π_0 and G_0 can be estimated using off-the-shelf regression algorithms based on the observed

data, and on the other hand, Q_0 and $F_{X,0}$ can be estimated using regression methods for survival data subject to right censoring or right censoring and left truncation, respectively. Furthermore, it facilitates the construct of estimators of ψ_0 that enjoy certain robustness properties, as discussed in Section 1.5.

We wish to incorporate flexible learning strategies in our estimation procedure to minimize the risk of systematic bias stemming from the use of misspecified parametric or semiparametric nuisance models. Once an estimator η_n of η_0 is obtained, the naive plug-in estimator ψ_{η_n}, obtained by replacing η_0 by η_n in the form of ψ_{η_0}, could be considered. We refer to such an estimator as naive since in general η_n need not be tailored to the end goal of estimating ψ_0. Furthermore, if η_n is estimated flexibly, it is often the case that ψ_{η_n} is overly biased and fails to even be $n^{\frac{1}{2}}$–consistent. Debiasing tools are typically used to address this challenge. Here, we employ the one-step debiasing approach based on the efficient influence function (Ibragimov and Has' Minskii, 1981; Pfanzagl, 1982) as well as the optimal estimating equations framework (Van der Laan and Robins, 2003).

A standard one-step debiased estimator of ψ_0 is given by $\psi_{\eta_n} + \frac{1}{n}\sum_{i=1}^{n}\phi_{\eta_n}(O_i)$. While the simplicity of this estimator is appealing, its asymptotic linearity is only guaranteed to hold under a stringent cap on the flexibility of the procedures used to yield η_n. To circumvent this constraint, cross-fitting can be incorporated into the construction of the one-step debiased estimator (Zheng and Laan, 2011; Chernozhukov et al., 2018). In its simplest form, this is achieved by partitioning the sample into two subsamples, using one subsample to obtain η_n and the other to build the one-step debiased estimator, repeating this construction with the roles of the subsamples reversed, and finally averaging the two estimators obtained. This procedure can be naturally extended to involve partitioning the sample into $K \geq 2$ subsamples of approximately equal sizes. Specifically, to compute the K-fold cross-fitted one-step debiased estimator, we first randomly partition the index set $\{1, 2, \ldots, n\}$ into K subsets, say $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_K$, of roughly equal sizes $n_1, n_2, \ldots, n_K$. Then, for each $k = 1, 2, \ldots, K$, an estimate $\eta_{k,n} = (H_{k,n}, \pi_{k,n}, G_{k,n}, Q_{k,n}, S_{k,n})$ of η_0 is obtained using only observations with indices not in $\mathcal{V}_k$, and the estimate $\psi^*_{k,n} := \psi_{\eta_{k,n}} + \frac{1}{n_k}\sum_{i\in\mathcal{V}_k}\phi_{\eta_{k,n}}(O_i)$

of ψ_0 is calculated. Finally, the average $\psi_n^* := \frac{1}{K}\sum_{k=1}^K \psi_{k,n}^*$ of fold-specific estimates is taken to be the final estimate of ψ_0. For any fixed $K \geq 2$, the cross-fitted one-step debiased estimator is guaranteed to be asymptotically linear without the need to limit the range of algorithms used to estimate η_0 — details are provided in Section 1.5.

Alternatively, we consider a second estimator ψ_n^{**} based on solving the efficient influence function estimating equation. To be precise, denoting by $\phi_{\eta,\psi}$ the efficient influence function ϕ_P where all nuisances are replaced by corresponding components of η but the parameter value $\Psi(P)$ is instead replaced by ψ, the fold-specific estimator $\psi_{k,n}^{**}$ is the solution in ψ of the equation

$$\sum_{i \in \mathcal{V}_k} \phi_{\eta_{k,n},\psi}(O_i) = 0 \ ,$$

and the estimator ψ_n^{**} is taken to be the average $\frac{1}{K}\sum_{k=1}^K \psi_{k,n}^{**}$ of the fold-specific estimators. Because for each fixed realization o of the data unit and each fixed nuisance η the mapping $\psi \mapsto \phi_{\eta,\psi}(o)$ is linear in ψ, ψ_n^{**} admits a closed-form expression: specifically, we find that $\psi_{k,n}^{**}$ is given explicitly by

$$\frac{\sum_{i \in \mathcal{V}_k} \left[\frac{I(A_i = a_0)}{\pi_{k,n}(Z_i)} \gamma_{k,n}(Z_i) \phi_{\mathrm{KM},\eta_{k,n}}(L_{\eta_{k,n},\varphi})(O_i) + \mu_n(Z_i) \left\{ \frac{1}{S_{k,n}(W_i \mid A_i, Z_i)} - \phi_{\mathrm{KM},\eta_{k,n}}(\gamma_{k,n,\natural})(O_i) \right\} \right]}{\sum_{i \in \mathcal{V}_k} \left\{ \frac{1}{S_{k,n}(W_i \mid A_i, Z_i)} - \phi_{\mathrm{KM},\eta_{k,n}}(\gamma_{k,n,\natural})(O_i) \right\}} .$$

Because $\phi_{\eta,\psi}$ does not have the form $\tilde{\phi}_\eta - \psi$ for any function $\tilde{\phi}_\eta$ indexed by η but not ψ, the estimators ψ_n^* and ψ_n^{**} are distinct. As we will see below, these estimators not only differ in their value on given samples but also in at least one key statistical property.

1.5 Large-sample inferential theory

1.5.1 Pointwise statistical inference

We study conditions under which the proposed estimators ψ_n^* and ψ_n^{**} are asymptotically linear and nonparametric efficient estimators of the survival integral ψ_0. We denote by $\eta_\infty := (H_0, \pi_\infty, G_\infty, Q_\infty, S_{X,\infty})$ the common limit in-probability of the split-specific nuisance

estimators $\eta_{1,n}, \eta_{2,n} \ldots, \eta_{K,n}$, and by $\overline{\tau}(a,z) := \min\{\overline{\tau}_T(a,z), \overline{\tau}_C(a,z)\}$ the maximum possible follow-up time in the subpopulation of individuals with $(A,Z) = (a,z)$. We will refer to the following conditions on the nuisance estimators:

(C1) the following consistency conditions hold:

$$\text{(a)} \quad \max_k E_0 \left| \frac{\bar{\gamma}_{k,n}(a_0, Z)}{\pi_{k,n}(Z)} - \frac{\bar{\gamma}_{\infty}(a_0, Z)}{\pi_{\infty}(Z)} \right|^2 \xrightarrow{P} 0\,;$$

$$\text{(b)} \quad \max_k E_0 \left[\sup_{y \in [0, \overline{\tau}(a_0, Z)]} \left| \frac{L_{k,n,\varphi}(y, a_0, Z)}{\tilde{S}_{k,n}(y \mid a_0, Z)} - \frac{L_{\infty,\varphi}(y, a_0, Z)}{\tilde{S}_{\infty}(y \mid a_0, Z)} \right| \right]^2 \xrightarrow{P} 0\,;$$

$$\text{(c)} \quad \max_k E_0 \left[\sup_{y \in [0, \overline{\tau}(A, Z)]} \left| \frac{\tilde{S}_{k,n}(y \mid A, Z)}{R_{k,n}(y \mid A, Z)} - \frac{\tilde{S}_{\infty}(y \mid A, Z)}{R_{\infty}(y \mid A, Z)} \right| \right]^2 \xrightarrow{P} 0\,;$$

$$\text{(d)} \quad \max_k E_0 \left[\sup_{y \in [\underline{\tau}_T(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{\bar{\gamma}_{k,n,\natural}(y, A, Z)}{\tilde{S}_{k,n}(y \mid A, Z)} - \frac{\bar{\gamma}_{\infty,\natural}(y, A, Z)}{\tilde{S}_{\infty}(y \mid A, Z)} \right| \right]^2 \xrightarrow{P} 0\,;$$

$$\text{(e)} \quad \max_k E_0 \left[\sup_{y \in [\underline{\tau}_W(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{1}{\tilde{S}_{k,n}(y \mid A, Z)} - \frac{1}{\tilde{S}_{\infty}(y \mid A, Z)} \right| \right]^2 \xrightarrow{P} 0\,;$$

$$\text{(f)} \quad \max_k E_0 \left[\sup_{u \in [\underline{\tau}_W(A, Z), \overline{\tau}(A, Z)]} |G_{k,n}(u \mid A, Z) - G_{\infty}(u \mid A, Z)| \right]^2 \xrightarrow{P} 0\,.$$

(C2) there exists some constant $\kappa \in (0, \infty)$ for which the following inequalities hold with P_0–probability tending to one:

$$\text{(a)} \quad \frac{\bar{\gamma}_{k,n}(a_0, Z)}{\pi_{k,n}(Z)}, \frac{\bar{\gamma}_{\infty}(a_0, Z)}{\pi_{\infty}(Z)} \leq \kappa$$

$$\text{(b)} \quad \sup_{y \in [0, \overline{\tau}(a_0, Z)]} \left| \frac{L_{k,n,\varphi}(y, a_0, Z)}{\tilde{S}_{k,n}(y \mid a_0, Z)} \right|, \sup_{y \in [0, \overline{\tau}(a_0, Z)]} \left| \frac{L_{\infty,\varphi}(y, a_0, Z)}{\tilde{S}_{\infty}(y \mid a_0, Z)} \right| \leq \kappa$$

$$\text{(c)} \quad \sup_{y \in [0, \overline{\tau}(A, Z)]} \left| \frac{\tilde{S}_{k,n}(y \mid A, Z)}{R_{k,n}(y \mid A, Z)} \right|, \sup_{y \in [0, \overline{\tau}(A, Z)]} \left| \frac{\tilde{S}_{\infty}(y \mid A, Z)}{R_{\infty}(y \mid A, Z)} \right| \leq \kappa$$

$$\text{(d)} \quad \sup_{y \in [\underline{\tau}_T(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{\bar{\gamma}_{k,n,\natural}(y, A, Z)}{\tilde{S}_{k,n}(y \mid A, Z)} \right|, \sup_{y \in [\underline{\tau}_T(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{\bar{\gamma}_{\infty,\natural}(y, A, Z)}{\tilde{S}_{\infty}(y \mid A, Z)} \right| \leq \kappa$$

(e) $\dfrac{1}{\tilde{S}_{k,n}(\overline{\tau}_W(A,Z)\,|\,A,Z)}, \dfrac{1}{\tilde{S}_{\infty}(\overline{\tau}_W(A,Z)\,|\,A,Z)} \leq \kappa$

(f) $|\varphi(Y,Z)|, \int |\varphi(dy,Z)| \leq \kappa$

(C3) the limits of the nuisance estimators agree with the true nuisances as follows:

(a) $S_{X,\infty}(Y\,|\,A,Z) = S_{X,0}(Y\,|\,A,Z)$ and $G_{\infty}(Y\,|\,A,Z) = G_0(Y\,|\,A,Z)$ P_0–almost surely;

(b) $\pi_{\infty}(Z) = \pi_0(Z)$ P_0–almost surely;

(c) $Q_{\infty}(Y\,|\,A,Z) = Q_0(Y\,|\,A,Z)$ P_0–almost surely;

(C4) $\max_k R(\eta_{k,n}, \eta_0) = o_P(n^{-\frac{1}{2}})$.

The following theorem describes the large-sample (pointwise) inferential properties of estimators ψ_n^* and ψ_n^{**} under appropriate conditions.

Theorem 2. *Suppose that conditions (A1)–(A2) and (C1)–(C2) hold.*

(i) If conditions (C3a)–(C3b) hold, then ψ_n^* is a consistent estimator of ψ_0.

(ii) If condition (C3a) holds, then ψ_n^{**} is a consistent estimator of ψ_0.

(iii) If conditions (C3)–(C4) hold, then ψ_n^* and ψ_n^{**} are asymptotically linear estimators of ψ_0 with common influence function ϕ_0, that is,

$$\psi_n^* \,=\, \psi_n^{**} + o_P(n^{-\frac{1}{2}}) \,=\, \psi_0 + \frac{1}{n}\sum_{i=1}^{n} \phi_0(O_i) + o_P(n^{-\frac{1}{2}}) \;.$$

In particular, this implies that $n^{\frac{1}{2}}\,(\psi_n^* - \psi_0\,)$ and $n^{\frac{1}{2}}\,(\psi_n^{**} - \psi_0)$ converge in distribution to a normal random variable with mean zero and variance $\sigma_0^2 := var_0\{\phi_0(O)\} < \infty$.

A simple estimator of σ_0^2 can be constructed as $\sigma_n^2 := \frac{1}{n}\sum_{i=1}^{n}\{\phi_{\eta_n}(O_i) - \bar{\phi}_n\}^2$ with $\bar{\phi}_n := \frac{1}{n}\sum_{i=1}^{n}\phi_n(O_i)$. Alternatively, a cross-fitted counterpart of σ_n^2 with possibly improved finite-

sample performance can be obtained as

$$\sigma^2_{n,*} := \frac{1}{K}\sum_{k=1}^{K}\frac{1}{n_k}\sum_{i\in\mathcal{V}_k}\{\phi_{\eta_{k,n}}(O_i) - \bar{\phi}_{k,n}\}^2$$

with $\bar{\phi}_{k,n} := \frac{1}{n_k}\sum_{i\in\mathcal{V}_k}\phi_{\eta_{k,n}}(O_i)$ for $k = 1, 2, \ldots, K$. Wald confidence intervals with asymptotic coverage $1-\alpha$ can then be constructed as $(\psi_n - q_\alpha\sigma_n n^{-\frac{1}{2}}, \psi_n + q_\alpha\sigma_n n^{-\frac{1}{2}})$, where q_α denotes the $(1-\frac{\alpha}{2})$–quantile of the standard normal distribution. Here, $\sigma_{n,*}$ can also be used instead of σ_n.

Beyond providing a template for making inference about ψ_0, the result above highlights that ψ_n^* and ψ_n^{**} both enjoy some degree of robustness to inconsistent nuisance estimation. Interestingly though, despite the fact that these two estimators are asymptotically equivalent when all nuisance estimators are consistent for their intended target, they have differing behavior when this is not the case. For example, the one-step estimator ψ_n^* retains its consistency for ψ_0 even when the censoring distribution is inconsistently estimated provided the time-to-event distribution, truncation distribution, and treatment propensity score are estimated consistently. In contrast, the estimating equations-based estimator ψ_n^{**} is consistent for ψ_0 provided the time-to-event and truncation distributions are estimated consistently, irrespective of how poorly the propensity score and censoring distributions may be estimated. As such, ψ_n^{**} exhibits strictly greater robustness than ψ_n^* in terms of consistency.

The conditions imposed in Theorem 2 can be scrutinized in the context of each application at hand. Condition (C1) requires the weak consistency of certain transformations of the nuisance estimators to their respective (possibly off-target) limits, often in some uniform sense that depends partly on the kernel function φ defining the estimand of interest. Condition (C2) requires that these transformations of nuisance estimators as well as their limits be bounded above, at least in large samples, so that all terms involved in the linearization of the survival integral estimator are controlled. Required support recovery assumptions ensure that these conditions hold for the true nuisance values, whereas condition (C2) requires that the same also be true of the nuisance limits. Condition (C3) is useful to describe various

patterns of consistent or inconsistent estimation of certain nuisance components under which consistency of the survival integral estimator may be preserved, as discussed above. Finally, condition (C4) is a generic condition on the rate of convergence of nuisance estimators — whether or not it holds in practice depends on the degree of smoothness or structure that the nuisance functions satisfy and whether the nuisance estimators used are able to leverage that structure effectively to achieve fast enough convergence.

1.5.2 Uniform statistical inference

We now study conditions under which we can make inference for a class of survival integrals simultaneously. Suppose that $\{\varphi_s : s \in \mathcal{S}\}$ is a collection of kernel functions from $[0,\infty) \times \mathcal{Z}$ to $\mathbb{R}$ indexed by a set $\mathcal{S}$, and that we are interested in learning about a collection of survival integral values $\{\psi_0(s) : s \in \mathcal{S}\}$, where $\psi_0(s)$ is the value of $\Psi(P_0)$ corresponding to kernel function $\varphi = \varphi_s$. In most applications, $\mathcal{S}$ is finite-dimensional but that is not a requirement for the developments below. For example, the set of kernels giving rise to the joint distribution function $\mathbb{F}_0$ of $(T(a_0), Z)$, namely $(t,z) \mapsto \mathbb{P}_0\{T(a_0) \leq t, Z \leq z\}$, over $[0,\tau) \times \mathcal{Z}$ is specified by $\varphi_s : (u,v) \mapsto I(u \leq t, v \leq z)$ for $s := (t,z)$ ranging in $\mathcal{S} := [0,\tau) \times \mathcal{Z}$. The conditions outlined so far pertain to inference for a fixed index s. To ensure valid inference uniformly over a range of values, we require the following additional conditions, where for any given kernel function $\varphi = \varphi_s$ we explicitly define $L_{s,P}$ pointwise as $L_{s,P}(y \,|\, a, z) := \int_y^\infty \tilde{S}_P(t \,|\, a, z)\, \varphi_s(dt, z)$, and denote by $\phi_{s,P}$ the nonparametric efficient influence function of $P \mapsto \Psi(P)$ with kernel $\varphi = \varphi_s$ under sampling from P, and by $R_s(P, P_0)$ the corresponding linearization remainder. We will make use of the conditions below:

(D1) the following consistency condition hold:

$$\text{(a)} \quad \max_k E_0 \left[\sup_{y \in [0, \overline{\tau}(a_0, Z)]} \sup_{s \in \mathcal{S}} \left| \frac{L_{s,k,n,\varphi}(y, a_0, Z)}{\tilde{S}_{k,n}(y \,|\, a_0, Z)} - \frac{L_{s,\infty,\varphi}(y, a_0, Z)}{\tilde{S}_\infty(y \,|\, a_0, Z)} \right| \right]^2 \xrightarrow{P} 0\,;$$

(D2) there exists some constant $\kappa \in (0,\infty)$ for which the following inequalities hold

with P_0–probability tending to one:

$$\text{(a)} \quad \sup_{s\in\mathcal{S}} \sup_{y\in[0,\overline{\tau}(a_0,Z)]} \left| \frac{L_{s,k,n}(y \mid a_0, Z)}{S_{k,n}(y \mid a_0, Z)} \right|, \sup_{s\in\mathcal{S}} \sup_{y\in[0,\overline{\tau}(a_0,Z)]} \left| \frac{L_{s,0}(y, a_0, Z)}{S_0(y \mid a_0, Z)} \right| \leq \kappa$$

(D3) $\max_k \sup_{s\in\mathcal{S}} R_s(\eta_{k,n}, \eta_0) = o_P(n^{-\frac{1}{2}})$;

(D4) The set of functions $\{\phi_{s,0} : s \in \mathcal{S}\}$ forms a P_0–Donsker class.

Denoting the estimator of $\psi_0(s)$ by $\psi_n(s)$, either corresponding to ψ_n^* or ψ_n^{**} based upon kernel $\varphi = \varphi_s$, we define the standardized process $\mathbb{B}_n := \{\mathbb{B}_n(s) : s \in \mathcal{S}\}$ pointwise as $\mathbb{B}_n(s) := n^{\frac{1}{2}}\{\psi_n(s) - \psi_0(s)\}$. The result below provides conditions under which $\mathbb{B}_n$ converges weakly to the same Gaussian process as the empirical process $\mathbb{B}_n^+ := \{\mathbb{B}_n^+(s) : s \in \mathcal{S}\}$ defined pointwise as $\mathbb{B}_n^+(s) := n^{-\frac{1}{2}} \sum_{i=1}^n \phi_{s,0}(O_i)$. Below, $\ell^\infty(\mathcal{S})$ refers to the space of uniformly bounded functions from $\mathcal{S}$ to $\mathbb{R}$.

Theorem 3. *Suppose that condition (A1) holds for each $\varphi = \varphi_s$ with $s \in \mathcal{S}$, and that conditions (C1a)–(C1c), (C2a), (C3) and (D1)–(D4) also hold. Then, the process $\mathbb{B}_n$ converges weakly to a tight mean-zero Gaussian process with covariance function $\sigma_0(s,t) := E_0[\phi_{s,0}(O)\phi_{t,0}(O)]$ in the space $\ell^\infty(\mathcal{S})$ relative to the supremum norm over $\mathcal{S}$.*

This result can be used to numerically construct confidence sets for $\{\psi_0(s) : s \in \mathcal{S}\}$ similarly as described in Westling et al. (2023). Asymptotically valid fixed-width bands can be readily obtained using a Wald construction and an estimate of relevant quantiles of $\sup_{s\in\mathcal{S}} |\mathbb{B}_n(s)|$. Alternatively, variable-width bands could be obtained by instead considering re-scaled process $\bar{\mathbb{B}}_n := \{\bar{\mathbb{B}}_n(s) : s \in \mathcal{S}\}$ with $\bar{\mathbb{B}}_n(s) := \sigma(s,s)^{-\frac{1}{2}}\bar{\mathbb{B}}_n(s)$, as in Westling et al. (2023). We note that conditions (D1), (D2) and (D3) are uniform counterparts to conditions (C1b), (C2b) and (C4). Condition (D4) instead puts a constraint on the complexity of the collection consisting of the nonparametric efficient influence function under sampling from P_0 for each survival integral parameter considered. When $\mathcal{S}$ is finite-dimensional, this is achieved, for example, if $s \mapsto \phi_{s,0}$ satisfies a certain Lipshitz condition (see Example 19.7

of Van der Vaart, 2000). More generally, this condition holds if $o \mapsto \phi_{s,0}(o)$ has uniform sectional variation norm uniformly bounded over $s \in \mathcal{S}$.

1.5.3 Extension to smooth functionals

The nonparametric inferential procedures we have described pertain to survival integral estimands, which correspond to linear functionals of the joint distribution distribution $\mathbb{F}_0$ of $(T(a_0), Z)$. However, in some applications, the relevant estimand may be a nonlinear functional of this same distribution. Fortunately, through the delta method, our results on linear functionals can be directly used to tackle a large class of nonlinear functionals. We demonstrate how the established results readily permit the study of parameters that can be expressed as sufficiently smooth functionals of $\mathbb{F}_0$.

We denote by $\mathcal{F}$ the collection of distribution functions on $\mathbb{R} \times \mathcal{Z}$ restricted to the identification subset $\{(t, z) : 0 \leq t \leq \overline{\tau}_C(a_0, z), z \in \mathcal{Z}\} \subseteq \mathbb{R} \times \mathcal{Z}$. Suppose that $\Theta : \mathcal{F} \mapsto \mathbb{R}$ is a given functional, and that we are interested in nonparametric inference on $\theta_0 := \Theta(\mathbb{F}_0)$. The results obtained so far describe how to construct, whenever possible, a uniformly asymptotically linear and regular estimator $\mathbb{F}_n$ of $\mathbb{F}_0$. Provided $\Theta : \mathcal{F} \to \mathbb{R}$ is sufficiently smooth, it is reasonable to expect that $\theta_n := \Theta(\mathbb{F}_n)$ is an asymptotically linear and regular estimator of θ_0. Such a result is formalized in the following theorem.

Theorem 4. *Suppose that the conditions of Theorem 3 hold, and that Θ is Hadamard differentiable at $\mathbb{F}_0$ relative to the supremum norm. Then, it holds that*

$$\theta_n - \theta_0 = \frac{1}{n} \sum_{i=1}^{n} \partial\Theta \left(\mathbb{F}_0; s \mapsto \phi_{s,0}(O_i) \right) + o_P(n^{-\frac{1}{2}}) \ , \tag{1.5}$$

where $\partial\Theta(\mathbb{F}_0; h)$ denotes the Gâteaux derivative of Θ at $\mathbb{F}_0$ in a given direction $h : \mathcal{S} \to \mathbb{R}$. In particular, this implies that $n^{\frac{1}{2}}(\theta_n - \theta_0)$ tends to a mean-zero normal random variable with variance given by $\sigma^2_\Theta := var_0 \left[\partial\Theta \left(\mathbb{F}_0; s \mapsto \phi_{s,0}(O) \right) \right]$.

In Section 1.6, we explicitly discuss the implications of this result in the context of

Examples 4 and 5, two motivating examples provided in Section 1.2 that feature nonlinear functionals arising in applications. Before proceeding, we use the result above to describe how to study problems in which the estimand can be expressed as the solution of an estimating equation.

1.5.4 Application to estimating equations

Suppose that the estimand of interest m_0 can be expressed as the solution in m of the population estimating equation

$$\iint U_m(t,z)\,\mathbb{F}_0(dt,dz) = 0\ ,$$

where for each $m \in \mathbb{R}$ the function U_m maps from $\mathbb{R} \times \mathcal{Z}$ to $\mathbb{R}$. Quantiles of the marginal distribution of $T(a_0)$, for example, can be expressed in this manner, as explicitly discussed later. Similarly as before, to ensure identification, it must be the case that, for $P_{X,0}$–almost every $z \in \mathcal{Z}$, the mapping $t \mapsto U_m(t,z)$ is constant for $t > \overline{\tau}_C(a_0,z)$. Under certain regularity conditions, the solution m_n of the empirical version of this estimating equation in m, $\iint U_m(t,z)\,\mathbb{F}_n(dt,dz) = 0$, is an asymptotically linear and regular estimator of m_0 with influence function given by

$$\phi_{m,0} : (z,a,w,y,\delta) \mapsto -\left[\frac{\partial}{\partial m}\iint U_m(t,z)\,\mathbb{F}_0(dt,dz)\bigg|_{m=m_0}\right]^{-1} \tilde{\phi}_{m,0}(z,a,w,y,\delta)\ ,$$

where the unscaled influence function $\tilde{\phi}_{m,0}$ is defined as

$$\begin{aligned}\tilde{\phi}_{m,0} : (z,a,w,y,\delta) \mapsto{}& \frac{I(a=a_0)}{\pi_0(z)}\bar{\gamma}_0(z)\phi_{\mathrm{KM},0}(L_{U,0})(z,a,w,y,\delta)\\ &+ \frac{\mu_{a_0,0}(z)}{\gamma_0}\left[\frac{1}{\tilde{S}_0(w\,|\,a,z)} - \phi_{\mathrm{KM},0}(\gamma_{0,\natural})(z,a,w,y,\delta)\right]\end{aligned}$$

and we write $L_{U,0}(y,a,z) := \int_y^\infty \mathbb{S}_0(u,z)U_{m_0}(du,z)$, and $\mu_{a_0,0}(z) := \int U_{m_0}(t,z)\,\mathbb{F}_0(dt,z)$.

1.6 Revisiting motivating examples

Example 1: marginal survival probability

To begin, we consider perhaps the simplest survival functional, the marginal survival probability $P_{X,0}(T > t)$ for some $t > 0$. Considered as a function of t, this estimand consists of the marginal survival function, which describes the entire time-to-event distribution and is commonly reported, for example, in studies of the natural history of disease in a population. A marginal survival probability is typically estimated nonparametrically using the Kaplan-Meier estimator, which handles left truncation in addition to right censoring through a straightforward risk-set adjustment (Kaplan and Meier, 1958; Tsai et al., 1987). However, this estimator builds upon marginal independence between T and (W, C), and is inconsistent when this condition fails to hold. Here, we allow dependence between T and (W, C) so long as independence holds within strata defined by a baseline covariate vector Z.

In the last decades, several authors have developed methods for estimating a marginal survival probability allowing covariate-dependent right censoring (e.g., Robins et al., 1993; Murray and Tsiatis, 1996; Zeng, 2004; Moore and van der Laan, 2009) or covariate-dependent left truncation (e.g., Chaieb et al., 2006; Mackenzie, 2012; Vakulenko-Lagun et al., 2022). However, to the best of our knowledge, there is currently no method in the literature that accounts for covariate-dependent right censoring and left truncation, facilitates the conduct of valid inference even when flexible learning strategies are used to estimate involved nuisance functions, and also enjoys robustness to inconsistent estimation of certain nuisance functions. The recent work of Wang et al. (2022) proposes an approach that comes closest to achieving these desiderata, though their procedure does not appear to generally allow flexible estimation of the conditional censoring distribution and robustness to its consistent estimation.

Under conditions (B1)–(B2), the marginal survival probability $S_{X,0}(\tau)$ at fixed time τ can be expressed as the survival integral $\iint I(t \geq \tau)\, F_{X,0}(dt \mid z)\, H_{X,0}(dz)$ corresponding to kernel function $\varphi_\tau(t, z) := I(t \geq \tau)$. Using Theorem 1, and taking A to be degenerate in order to

recover the simpler setting in which there is no exposure on which we wish to intervene, the nonparametric efficient influence function of the corresponding parameter under sampling from P_0 is given by

$$o \mapsto -S_{X,0}(\tau \,|\, z)\bar{\gamma}_0(z) \left[\frac{\delta I(y \leq \tau)}{R_0(y \,|\, z)} - \int_w^{y \wedge \tau} \frac{\Lambda_{X,0}(du \,|\, z)}{R_0(u \,|\, z)} \right] + \frac{S_{X,0}(\tau \,|\, z) - S_{X,0}(\tau)}{\gamma_0} \left[\frac{1}{S_{X,0}(w \,|\, z)} - \phi_{\mathrm{KM},0}(\gamma_{0,\natural})(z, w, y, \delta) \right].$$

We can use our uniform results to construct an estimator and make inference on the marginal survival function $t \mapsto S_{X,0}(t)$ in an interval within which identification is possible. To enforce monotonicity of the resulting estimator, the resulting function-valued estimator can be projected into the space of monotone functions using isotonic regression (Westling et al., 2020).

Example 2: Brier score

The Brier score is defined as the survival integral $\iint [I(t \geq \tau) - b(z)]^2 \, F_{X,0}(dt \,|\, z) \, H_{X,0}(dz)$ corresponding to kernel function $\varphi_\tau(t, z) := [I(t \geq \tau) - b(z)]^2$, and has been used to quantify the predictive performance of a given algorithm $z \mapsto b(z)$ for predicting whether the event will occur by some fixed time τ (Brier et al., 1950). Here again, we consider A to be degenerate since there is no exposure on which we wish to intervene in this example. The Brier score can be used to compare the performance of several prediction models. In the absence of left truncation, Gerds and Schumacher (2006) derived an inferential procedure allowing conditionally-dependent right-censoring. In the more general case of left-truncated right-censored data, Theorem 1 implies that the nonparametric influence function for the corresponding parameter under sampling from P_0 is given by

$$o \mapsto -S_{X,0}(\tau \,|\, z)\bar{\gamma}_0(z) \left\{ \frac{\delta I(y \leq \tau)[1 - 2b(z)]}{R_0(y \,|\, z)} - \int_w^{y \wedge \tau} \frac{[1 - 2b(z)]\Lambda_{X,0}(du \,|\, z)}{R_0(u \,|\, z)} \right\} + \frac{[S_{X,0}(\tau \,|\, z) - b(z)]^2 - \kappa_\tau}{\gamma_0} \left[\frac{1}{S_{X,0}(w \,|\, z)} - \phi_{\mathrm{KM},0}(\gamma_{0,\natural})(z, w, y, \delta) \right],$$

where $\kappa_\tau := E_{X,0}[\varphi_\tau(T, Z)]$. Since it is often used heuristically to compare the performance of competing prediction algorithms, the Brier score is often not itself the target of inference in a given problem. Nevertheless, our results make it straightforward to perform formal tests to rigorously compare candidate prediction algorithms.

Example 3: counterfactual survival probability

The counterfactual (or treatment-specific) survival probability $\mathbb{S}_0(\tau) := \mathbb{P}_0\{T(a_0) \geq \tau\}$, defined as the probability of survival beyond a given time τ under an intervention that sets treatment (or exposure) to a specified level, is a useful summary to study treatment effects in the context of time-to-event endpoints. Differences or ratios in counterfactual survival probabilities across treatment levels are commonly used in clinical trials to quantify treatment effects.

In addition to adjustment needed for possibly covariate-dependent right-censoring and left-truncation, in order to identify the counterfactual survival probability, adjustment for possible confounding between treatment level and survival is also required. Under typical causal conditions, including positivity and conditional randomization, the counterfactual survival distribution is identified by the survival integral $\iint I(t \geq \tau)\, F_{X,0}(dt \,|\, a_0, z)\, H_{X,0}(dz)$ corresponding to kernel function $\varphi_\tau(t, z) := I(t \geq \tau)$. This is the same expression as in Example 1 but without degeneracy of A. Westling et al. (2023) derived a nonparametric efficient estimation procedure for a counterfactual survival probability based on right-censored data. In view of Theorem 1, the nonparametric efficient influence function of the corresponding parameter under sampling from P_0 is given by

$$o \mapsto -S_{X,0}(\tau \,|\, a_0, z)\bar{\gamma}_0(z)\frac{I(a = a_0)}{\pi_0(z)}\left[\frac{\delta I(y \leq \tau)}{R_0(y \,|\, a_0, z)} - \int_w^{y \wedge \tau} \frac{\Lambda_{X,0}(du \,|\, a_0, z)}{R_0(u \,|\, a_0, z)}\right] + \frac{S_{X,0}(\tau \,|\, a_0, z) - \mathbb{S}_0(\tau)}{\gamma_0}\left[\frac{1}{S_{X,0}(w \,|\, a_0, z)} - \phi_{\mathrm{KM},0}(\gamma_{0,\natural})(z, a, w, y, \delta)\right].$$

When there is no truncation, this influence function simplifies and agrees exactly with that

provided in Theorem 2 of Westling et al. (2023). When A is degenerate, we recover the expression obtained in Example 1. As in Example 1, our results readily described how to perform uniform inference for the counterfactual survival function over an interval under which identification is possible.

Example 4: median counterfactual event time

Due to the presence of right censoring, summaries of the (marginal or counterfactual) time-to-event distribution that depend on its right tail — for example, the mean event time — are typically not identified. Because it generally circumvents this right-tail issue, and also because it affords an interpretation that is relevant in many scientific problems, the median event time is often used instead of the mean survival time in survival analysis. Nonparametric inference on the median marginal event time with right-censored data has a long history in survival analysis, dating at least as far back as Reid (1981a). Corresponding results for the median counterfactual event time are far more recent, with Díaz (2017) and Shepherd and Moreno-Betancur (2022) studying the problem in the absence of censoring and truncation.

Here, we are specifically interested in nonparametric inference on the median of the counterfactual time-to-event distribution, namely the median m_0 corresponding to the distribution function $\mathbb{F}_0$ of $T(a_0)$, based on left-truncated right-censored data. To the best of our knowledge, this problem has not been studied before. Of course, as discussed in Examples 1 and 3, results obtained for the median counterfactual event time imply results for the median marginal event time. While the median counterfactual event time cannot be expressed as a survival integral, it can be framed as the solution of an estimating equation based on a survival integral. Specifically, since $\int U_m(t)\,\mathbb{F}_0(dt) = 0$ if and only if $m = m_0$ with $U_m(t) := I(t \leq m) - 0.5$ under mild conditions, results from the last section can be directly used to characterize the median counterfactual event time parameter. We first note

that

$$\begin{aligned}\frac{\partial}{\partial m}\int U_m(t)\,\mathbb{F}_0(dt)\Big|_{m=m_0} &= \frac{\partial}{\partial m}\iint U_m(t)\,F_{X,0}(dt\,|\,a_0,z)\,H_{X,0}(dz)\Big|_{m=m_0}\\ &= \int f_{X,0}(m_0\,|\,a_0,z)\,H_{X,0}(dz) \;=\; \int f_{X,0}(m_0\,|\,a_0,z)\,\tilde{H}_0(dz)\;,\end{aligned}$$

which is simply the identification of the density function of $T(a_0)$ evaluated at m_0. Results from Section 1.5.4 then yield that, under sampling from P_0, the nonparametric influence function of the counterfactual median event time parameter is given by

$$\begin{aligned}o \mapsto -\frac{1}{\int f_{X,0}(m_0\,|\,a_0,z)\tilde{H}_0(dz)}\Bigg\{&\frac{I(a=a_0)\bar{\gamma}_0(z)}{\pi_0(a_0\,|\,z)}\phi_{KM}(L_{U,0})(z,a,w,y,\delta)\\ &+\frac{0.5-S_{X,0}(m_0\,|\,a_0,z)}{\gamma_0}\left[\frac{1}{S_{X,0}(w\,|\,a,z)}-\phi_{KM}(\gamma_{0,\natural})(z,a,w,y,\delta)\right]\Bigg\}\;,\end{aligned}$$

where we have defined $L_{U,0}(z,a,y) := -S_{X,0}(m_0\,|\,a_0,z)I(y\leq m_0)$. In the absence of a treatment intervention and truncation, and under unconditionally independent right censoring, it can be shown that this expression agrees with the influence function of the median marginal event time provided in Reid (1981a). In the absence of censoring and truncation, this expression reduces to results for the standard G-computed median, as discussed extensively in Díaz (2017) and Shepherd and Moreno-Betancur (2022).

Example 5: model-agnostic measure of dependence of $T(a_0)$ on a_0

Our final example serves to illustrate that the class of parameters covered by our theoretical results is large enough to include methodologically challenging estimands of scientific interest. An assumption-lean approach to Cox regression recently proposed in Vansteelandt et al. (2022) is based on a nonparametric projection estimand that, in the case of a binary exposure A, summarizes $(t,z)\mapsto \log\Lambda_{X,0}(t\,|\,1,z)-\log\Lambda_{X,0}(t\,|\,0,z)$, the difference in log conditional cumulative hazard functions. In this spirit, we consider here a novel version of this estimand,

defined as the contrast $\mathbb{L}(1) - \mathbb{L}(0)$ with

$$\mathbb{L}(a_0) := \int \omega(t) \log[-\log \mathbb{P}_0\{T(a_0) > t\}]\, dt$$

for $a_0 \in \{0, 1\}$ and some fixed weight function ω such that $\int \omega(t)\, dt = 1$. We have used here that $-\log \mathbb{P}_0\{T(a_0) > t\}$ equals the counterfactual cumulative hazard function at t under continuity of the distribution function. The function ω is taken to emphasize scientifically relevant values of t and to also restrict attention to values of t at which the distribution functions of $T(0)$ and $T(1)$ are identified.

Notably, this estimand cannot be expressed as a survival integral; nevertheless, it is generally a Hadamard differentiable functional of the counterfactual distribution function. Specifically, writing $\Theta(\mathbb{F}) := \int \omega(t) \log[-\log\{1 - \mathbb{F}(t)\}]\, dt$, the corresponding parameter is the difference between the evaluation of Θ on the counterfactual distribution function of $T(1)$ and on that of $T(0)$. In view of Theorem 4, it is not difficult to show that, under sampling from P_0, this parameter has nonparametric efficient influence function given by

$$o \mapsto \int \omega(t) \left[\frac{\phi_{1,t}(o)}{\{1 - \mathbb{F}_1(t)\} \log\{1 - \mathbb{F}_1(t)\}} - \frac{\phi_{0,t}(o)}{\{1 - \mathbb{F}_0(t)\} \log\{1 - \mathbb{F}_0(t)\}} \right] dt \ ,$$

where $\phi_{a_0,t}$ is the nonparametric efficient influence function of the parameter identifying the cumulative distribution $\mathbb{F}_{a_0}$ of $T(a_0)$ at t, presented explicitly in Example 3. While this result applies when ω is fixed, it is possible to derive extended results for a weight function indexed by P_0 itself although those calculations would typically need to be done on a case-by-case basis unless the dependence of the weight function on P_0 is very smooth.

1.7 Numerical illustrations

We now present results from a simulation study based on Example 1 described above. Specifically, we make inference about a marginal survival function in the absence of a treatment variable but including covariate-dependent censoring and truncation. We considered settings

with high, low and no truncation levels, and high and low censoring levels.

Covariates Z_1, Z_2, and Z_3 are independent random variables distributed uniformly on the set $\{-1, +1\}$. Given covariate vector $Z = z$, the study entry time variable W is distributed as $10U$, where U is a Beta random variable with parameters $\alpha = a(z_1)$ and $\beta = b(z_1)$. In the low truncation setting (25% truncation), we set $a(z_1) = 1$ and $b(z_1) = 1 + 2I(z_1 < 0)$; in the high truncation setting (50% truncation), we set $a(z_1) = 1 + I(z_1 > 0)$ and $b(z_1) = 1$. Given $W = w$ and $Z = z$, the censoring time is taken $C = W + D$ where D is an independent random variable generated from a Gamma distribution with shape k and scale $\lambda_C = \exp\{-\frac{1}{10}(z_1 + z_2 + z_3)\}$, where in each simulation setting k was chosen to yield a low (25%) or high (50%) censoring rate. Given covariate vector $Z = z$, we independently simulated the event time T from a Gamma distribution with shape 6 and scale $\lambda_T = \exp\{\frac{1}{10}(z_1 + z_2 + z_3)\}$.

The marginal survival function was estimated at times corresponding to the first four quintiles of the population event-times. Performance was assessed using the following metrics:

(A) empirical bias, scaled by $n^{\frac{1}{2}}$;

(B) empirical variance, scaled by n;

(C) pointwise confidence interval coverage;

(D) average confidence interval width.

We note here that, in the absence of a treatment variable A, estimators ψ_n^* and ψ_n^{**} are actually equal to each other, and so, all results pertain to this common estimation procedure. Each nuisance function featuring in the construction of the debiased estimator were estimated using flexible approaches. The conditional time-to-event and censoring survival functions were estimated using global survival stacking with a learner library consisting of the empirical mean, GLM, GAM, splines, random forests, and gradient-boosted trees. For comparison, a correctly specified parametric regression model was also included. The conditional entry time

distribution was estimated using the stratified empirical distribution function. The five-fold cross-fitted one-step estimator was compared to the Kaplan-Meier estimator, which assumes independent censoring and truncation.

Across all six scenarios, we observe similar patterns for simulation results. The debiased global stacking estimator, as implemented in Section 1.5.1, has negligible bias, appropriate coverage, and has similar variability as the debiased estimator based on correctly specified parametric nuisance estimators. This is not surprising, since nuisance estimators only play a second-order role in the behavior of the debiased estimator — in fact, as long as nuisance functions are estimated sufficiently well, the debiased estimator with estimated nuisances is asymptotically equivalent to the (oracle) debiased estimator based on the true nuisances. In this setting, cross-fitting does not appear to greatly change the results but this is not expected to always be the case. As expected, the Kaplan-Meier is biased in all settings but resulting confidence intervals can undercover or overcover depending on which of the bias and variance dominate.

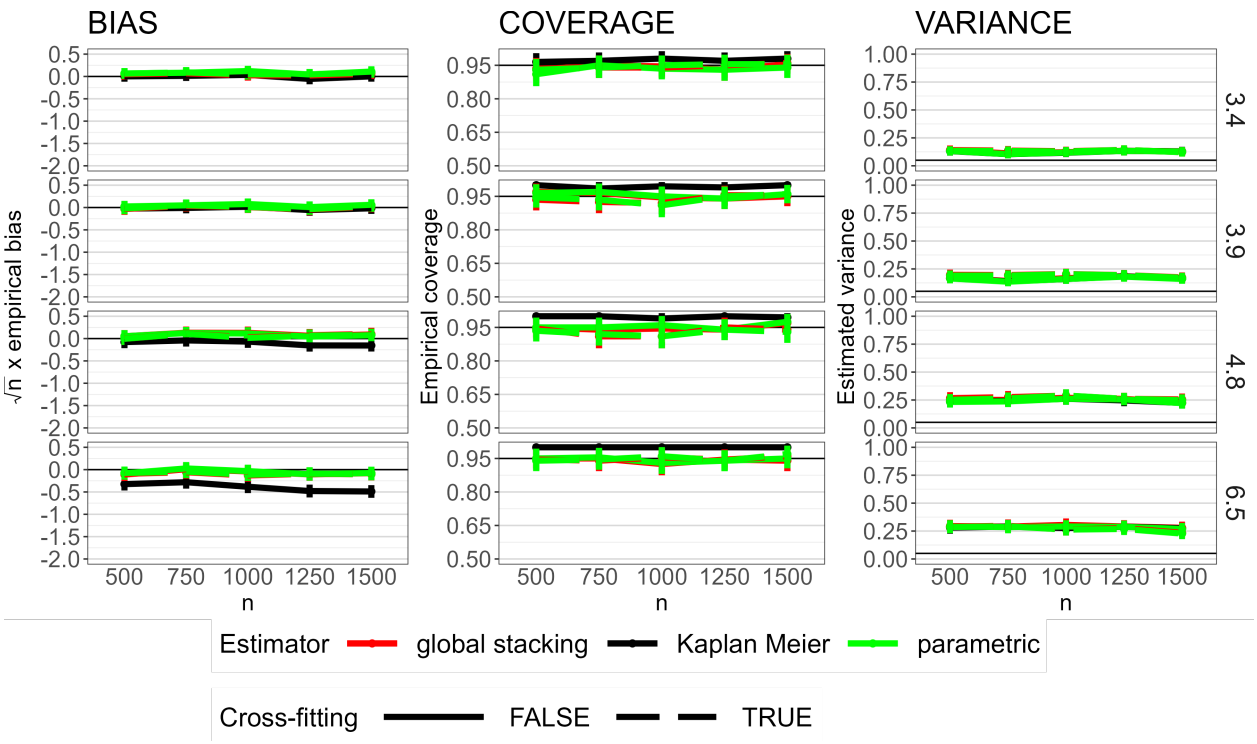

Figure 1.1: Summary of inferential performance metrics for scenario with 25% censoring and no truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

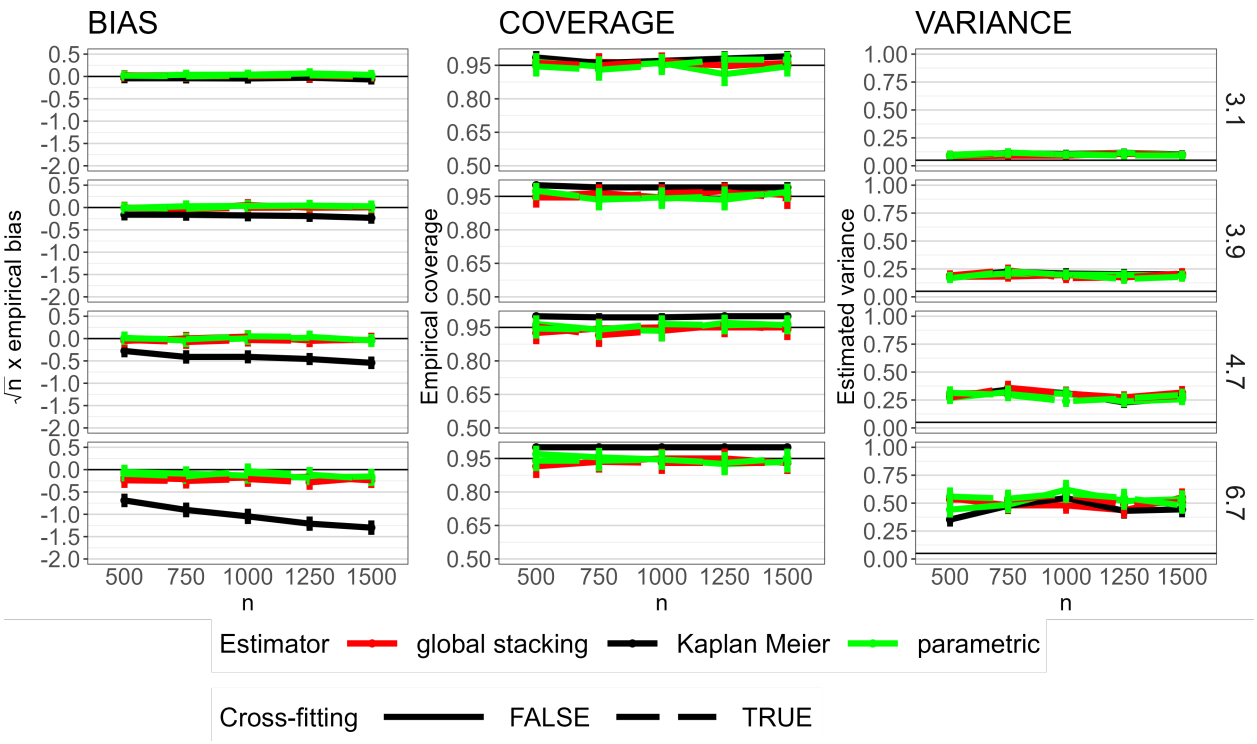

Figure 1.2: Summary of inferential performance metrics for scenario with 50% censoring and no truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

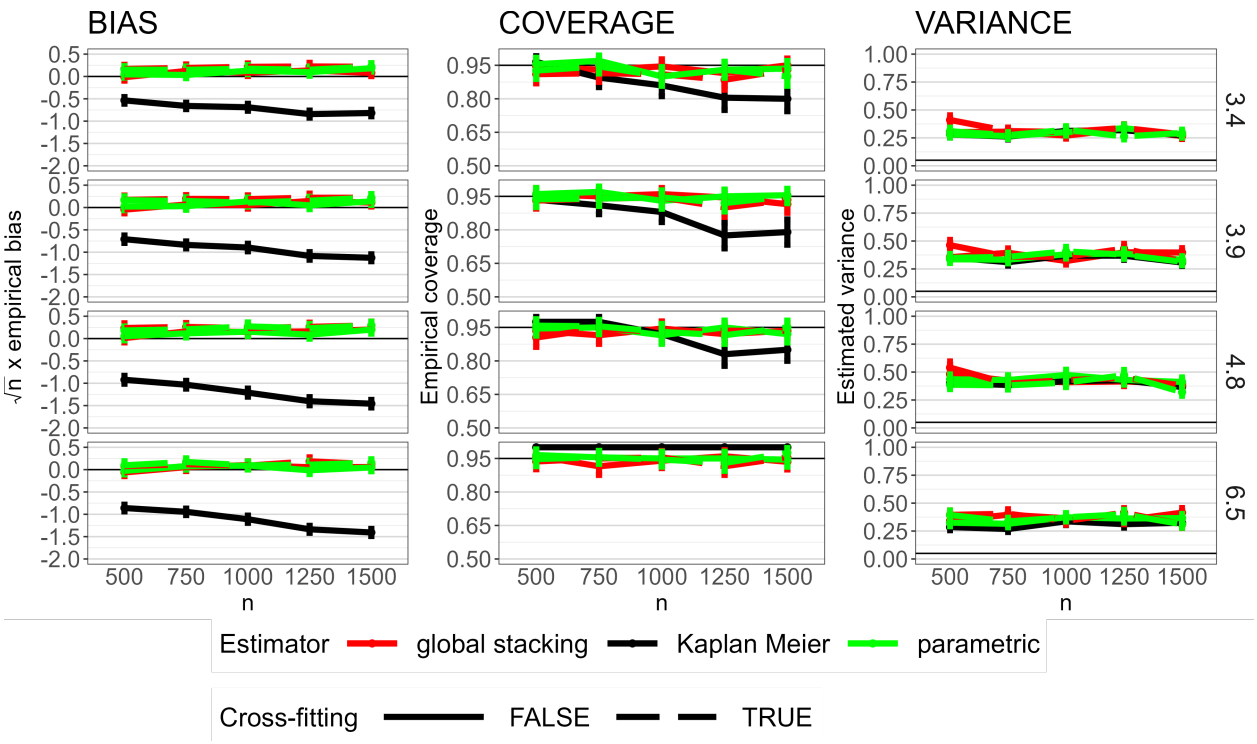

Figure 1.3: Summary of inferential performance metrics for scenario with 25% censoring and 25% truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

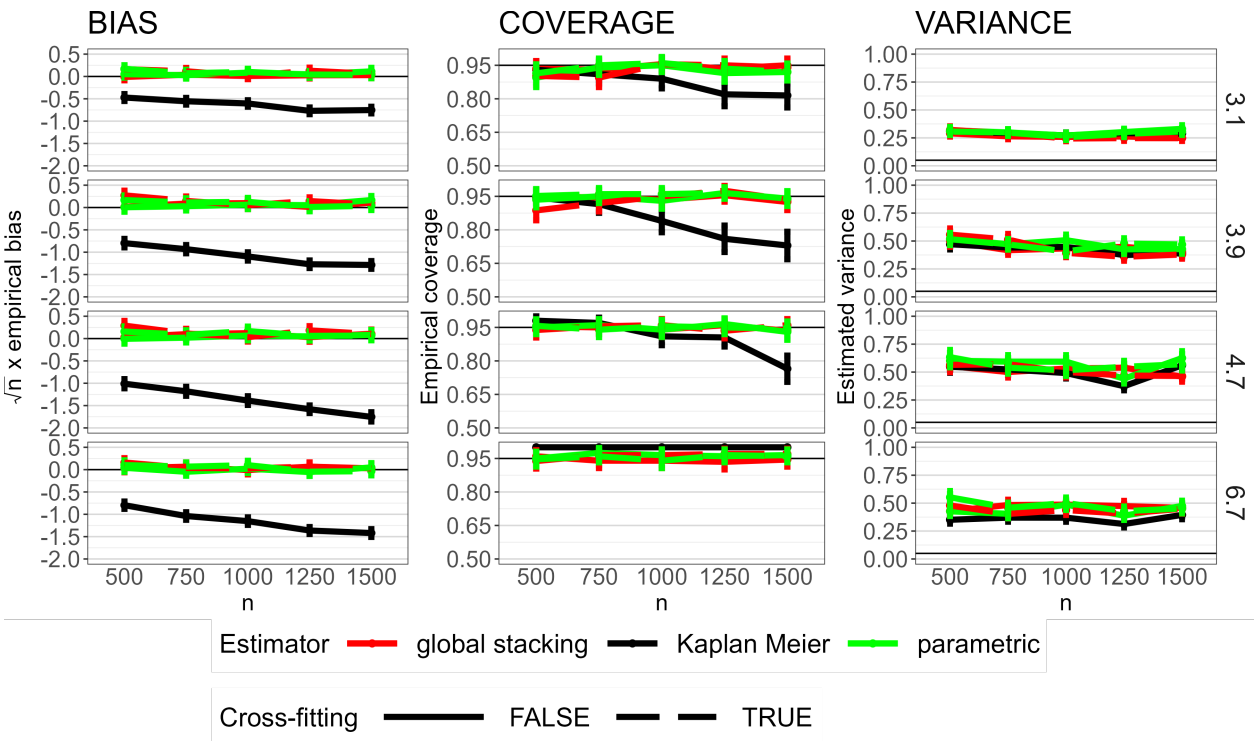

Figure 1.4: Summary of inferential performance metrics for scenario with 50% censoring and 25% truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

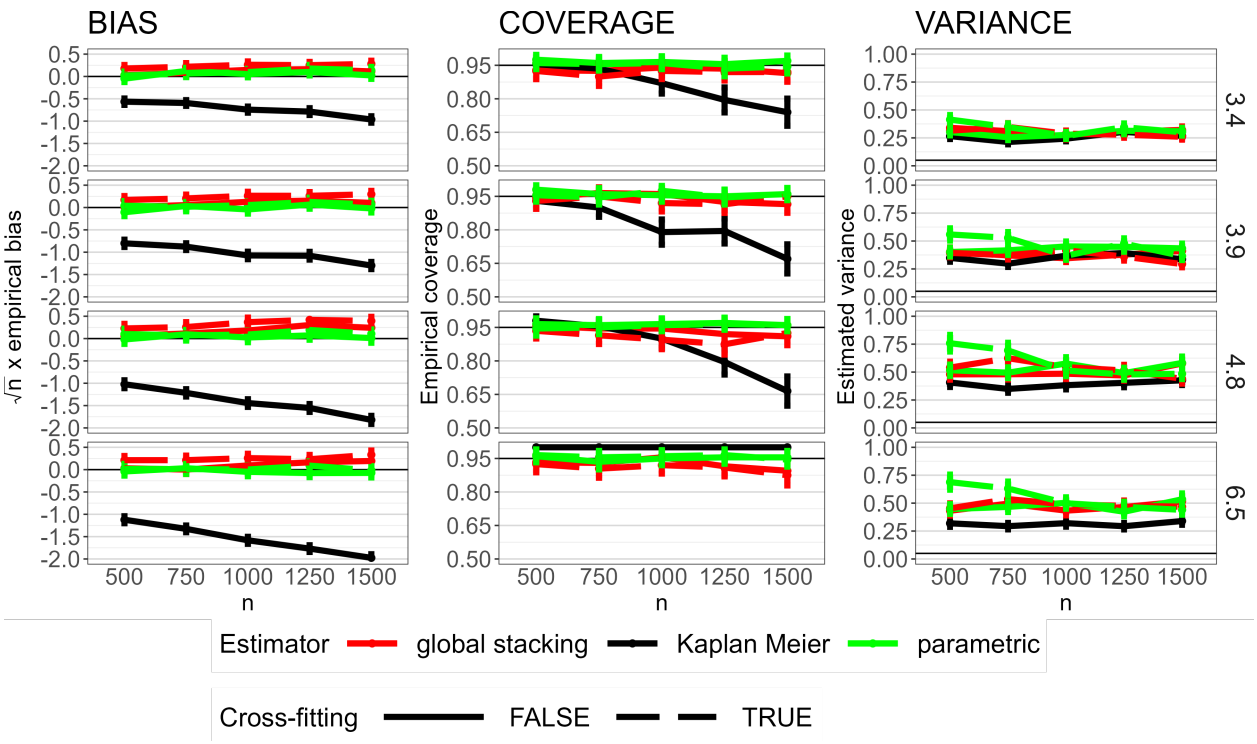

Figure 1.5: Summary of inferential performance metrics for scenario with 25% censoring and 50% truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

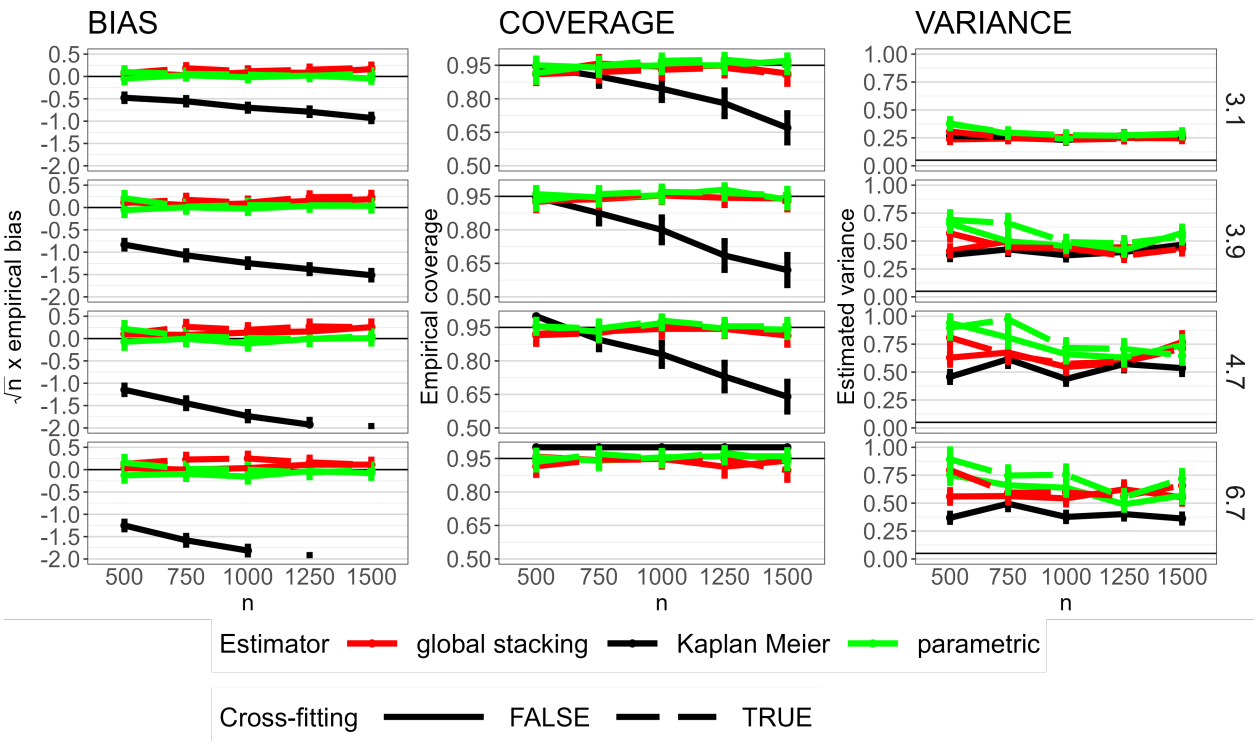

Figure 1.6: Summary of inferential performance metrics for scenario with 50% censoring and 50% truncation. Including empirical bias, coverage, and variance of the different estimation procedures.

1.8 Concluding remarks

We have developed and studied debiased machine learning methods for making statistical inference on summaries of a counterfactual time-to-event distribution using left-truncated right-censored data. These methods require, as an intermediate step, estimation of various nuisance functions. However, in view of the Neyman-orthogonal construction used, the use of machine learning techniques is allowed to estimate these nuisances, thereby reducing the risk of inconsistent estimation. In particular, through the use of adaptive ensemble learning, this fact renders more realistic the rate conditions imposed on the nuisance estimators.

Identification of a summary of the time-to-event distribution requires that the set of available covariates be sufficiently rich to explain any dependence between the treatment allocation mechanism and counterfactual outcomes, between the truncation and event times, and between the censoring and event times. While in this work we have focused on baseline covariates exclusively, our methods could be extended to allow covariates recorded at (post-baseline) study entry to possibly inform the relationship between the censoring and event times. Identification also requires that the counterfactual time-to-event distribution itself be identified over a sufficient large portion of its support to allow computation of the summary of interest. Right censoring often precludes identification of the right tail of a time-to-event distribution, rendering unidentifiable summaries that depend on the right tail, such as moments of the time-to-event distribution. Interestingly, in some cases, left truncation can help restore the identifiability of this right tail; this occurs, for example, when left truncation arises due to cross-sectional sampling, and the censoring mechanism only acts on the portion of the event time under follow-up (i.e., from study entry and on) and is independent of the study entry time. Thus, in any given application, it is important to determine the extent to which the time-to-event support may be recovered in a given application, and to consider its implications on which summary can be identified.

The two estimation procedures we derived exhibited some level of robustness to the estimation of involved nuisance functions. Specifically, they allowed a certain degree of inconsis-

tent nuisance estimation under which the summary of interest is still estimated consistently. Of the estimators proposed, we noted that estimating equations-based estimator is qualitatively more robust than the one-step estimator — this is an interesting example of a setting in which two constructive approaches for nonparametric inference, while equivalent when all nuisances are estimated sufficiently well, differ in behavior when that is not the case. Both procedures considered required consistent estimation of the target conditional time-to-event distribution and the observable conditional truncation distribution; in other words, neither exhibits robustness to inconsistent estimation of these nuisances. In future work, it is important to consider how the use of different parametrizations may lead to different — and possibly more permissive — robustness profiles. Additionally, while the robustness discussed here pertains to preservation of consistency, it may be fruitful to also consider how to achieve preservation of asymptotic linearity so that robust confidence intervals and p-values may also be constructed, along the lines of Benkeser et al. (2017), for example.

While the class of summaries we considered in this chapter is broad, it does not include all functionals for which parametric-rate inference is possible. For example, some summaries that depend inextricably on the counterfactual time-to-event density function fall outside the class considered, and appear more difficult to tackle in generality. Similarly, it is challenging to characterize inference for survival integrals for which the kernel φ depends on the underlying distribution $P_{X,0}$. However, such survival integrals do arise in contemporary applications, and the developments provided here serve as important building blocks for the study of such integral estimands.

Chapter 2

DEBIASED MACHINE LEARNING FOR A CAUSAL CONTRAST BASED ON AUTOPSY DATA

It is often of interest to characterize how an exposure relates to a particular disease-related biomarker using observational data. While many biomarkers can be recorded *in vivo*, others can only be measured by analyzing tissues collected during autopsy. This is the case, for example, for neurological biomarkers that require sampling brain tissues. When the scientific goal is to quantify the effect of an exposure (e.g., genetic mutation, smoking, air pollution) on such biomarker, complications arise because biomarker data can only be collected when a participant dies. Comparing observed biomarker values across exposure groups can be highly misleading when the exposure under consideration affects survival, possibly even resulting in a reversal in the direction of association. In this work, we propose a causal inference framework for studying the effect of a point-exposure on a time-varying biomarker process that can only be sampled at death. We derive nonparametric efficient inferential procedures based on either prospective or retrospective studies, and establish the large-sample properties of the proposed procedures. We illustrate the operating characteristics of these procedures through simulation studies, and use our proposed methods to analyze the effect of the APOE-4 gene hyperphosphorylated tau protein accumulation in the brain.

2.1 Introduction

Whenever it is of interest to study the impact of various exposures on the risk of a particular disease, investigators at times seek to instead characterize how those exposures impact biomarkers established to be intimately tied to the disease process. This is particular useful in contexts in which the clinical endpoint occurs late in life, is relatively rare, or is difficult to diagnose clinically. While many biomarkers can be recorded *in vivo*, others can only be measured by analyzing tissue collected during autopsy. This is the case, for example, for neurological biomarkers that can only be measured on brain tissue samples. When the scientific goal is to quantify the effect of an exposure (e.g., genetic markers, smoking, air pollution) on such biomarker, outcome data collection can only occur when a participant dies. Regardless of whether the exposure is randomized, comparing observed biomarker values across exposure groups can be highly misleading when the exposure under consideration affects survival. This occurs because such naive comparison ignores that observation times tend to be different across exposure groups and that biomarker values typically vary temporally.

A notable example of the contradictions that can arise in estimating the effect of a toxic exposure on biomarkers measured at autopsy arises in the Alzheimer's disease literature. A review of the literature quantifying the relationship between smoking and biomarkers of Alzheimer's disease and related dementia using post-mortem data noted inconsistencies in the direction of the estimated effect of smoking (Chang et al., 2014). Previous investigations found smoking to have both protective (Ulrich et al., 1997) and harmful (Tyas et al., 2003) effects on Alzheimer's disease progression. In the case of incident Alzheimer's disease, various theories have been posited to explain such counter-intuitive results (Hernán et al., 2008; Chang et al., 2012), yet complexities inherent to the autopsy setting have been largely unexplored. Reported differences in the autopsy setting could stem, for example, from the fact that a variety of statistical estimands have been used across studies, rendering results not directly comparable and more or less susceptible to autopsy-induced biases.

Here, we describe a framework for the analysis of studies in which the time-varying out-

come of interest is only recorded once at a random time, and there may be a systematic difference in observation times across exposure levels. For the autopsy setting, wherein the time-varying outcome is measured as death, we define a causal estimand for describing the effect of exposure on the outcome process and discuss its interpretation. We detail conditions under which this estimand is identified in the context of a prospective observational study with neither censoring not truncation, and find that the identification formula agrees with that of the natural direct effect, a causal parameter featured prominently in causal mediation analysis. More importantly, we derive identification formulas for use in two more complex designs — a prospective design including possible delayed entry and incomplete follow-up, and a retrospective design. Both settings are challenging to analyze because the outcomes observed at autopsy represent a biased sample of the population due to truncation, the systematic exclusion of individuals based on their survival time. In this sense, while developed with the analysis of autopsy data in mind, the methods developed here can also be used in other problems in which the natural direct effect is of interest but the mediator is observed subject to a combination of censoring and truncation. We utilize a debiased machine learning approach to statistical inference to allow the use of flexible learning strategies in order to minimize reliance on unnecessary statistical modeling assumptions in the process of learning causal effects from autopsy studies.

We note that the inconsistencies that can arise in the study of autopsy data have been investigated using various strategies. Researchers have highlighted certain sampling biases that may be occur in autopsy studies due, for example, to the autopsy consent that the patient or their next-of-kin must provide (Haneuse et al., 2009) or to the truncation mechanism that characterizes many retrospective autopsy samples (Rennert and Xie, 2019). Nevertheless, beyond the epistemology of inferring causes-of-death from autopsy data (Rothman and Greenland, 2005), there has been minimal explicit discussion of what may be suitable causal estimands for consideration in autopsy studies. In this chapter, we aim to bridge this gap in the literature by providing a statistical framework for studying the causal effect of a point-exposure on a time-varying biomarker process observed only at death.

The chapter is organized as follows. In Section 2.2, we explicitly describe the challenges in interpreting simple causal summaries of the effect of a binary exposure on outcomes measured at autopsy, introduce the causal inference framework we build upon, and present identification results for the simplest autopsy study setting. In Sections 2.3 and 2.4, we describe two more realistic study designs for data collection, and derive the accompanying results for identification as well as nonparametric estimation and inference for each of these designs. Sections 2.3 and 2.4 both build upon results in Section 2.2 but can be read independently of each other. Simulations are presented in Section 2.5 to validate the properties of our inferential procedures. Finally, in Section 2.6 we analyze data from the Adult Changes in Thought (ACT) prospective community-based cohort to estimate the effect of a positive APOE-4 genetic expression on AD neurological biomarkers collected at autopsy.

2.2 Motivation and problem setup

2.2.1 Setting and notation

We begin by introducing the variables used in our causal framework for autopsy studies. We denote by $L \in \mathcal{L} \subseteq \mathbb{R}^p$ a vector of baseline covariates, and write the point-exposure of interest indicator as A. As a convention, we consider $t = 0$ to be the calendar time at exposure, and assume that L is collected immediately before exposure. Here, we focus on settings with a binary exposure — for example, $A \in \{0, 1\}$ may distinguish between exposed ($A = 1$) versus unexposed ($A = 0$) individuals. Adopting the Neyman-Rubin (Neyman, 1923; Rubin, 1974) potential outcomes framework, we denote by $T(a)$ the counterfactual time elapsed between exposure and death (hereafter simply referred to as survival time) corresponding to exposure level $A = a$. Similarly, we denote by $B_t(a)$ the counterfactual biomarker value at time t corresponding to exposure level $A = a$. We refer to $\{B_t(a) : t \in (0, T(a)]\}$ as the counterfactual biomarker process, noting that $B_t(a)$ is typically undefined beyond time of death $T(a)$. This motivates the definition of the complete counterfactual data unit as $\mathbb{X} := (L, A, T(0), T(1), \{B_t(0) : 0 < t \leq T(0)\}, \{B_t(1) : 0 < t \leq T(1)\})$; we denote by $P_{\mathbb{X},0}$

the distribution of $\mathbb{X}$ characterizing the target population of interest. Of note, this data unit includes the entire counterfactual biomarker process corresponding to each possible exposure level.

In autopsy studies, by design, certain measurements are only observed at time of death. In our setup, the point value of the biomarker process is only recorded when the patient dies, which motivates the definition of $B(a) := B_{T(a)}(a)$, the counterfactual biomarker value measured at death. The counterfactual autopsy data unit is thus given by $X := (L, A, T(0), T(1), B(0), B(1))$, and we denote by $P_{X,0}$ the distribution of X induced by $P_{\mathbb{X},0}$. Under ideal sampling conditions, where there is no loss to follow-up and the study population is representative of the target population, the observed data unit is $Z := (L, A, T, B)$, where $T := T(A)$ and $B := B(A) = B_T(A)$ are the observed survival time and biomarker value at death, respectively. We denote by $P_{Z,0}$ the distribution of the observable data unit Z induced by $P_{X,0}$. The ideal observable data consist of n independent draws $Z_1, Z_2, \ldots, Z_n$ from $P_{Z,0}$. In this article, we focus on inference in the context of two sampling designs, each of which generate a coarsened or biased version of these 'ideal' observable data.

2.2.2 Causal parameter of interest

In settings in which exposure is randomized, a naive analysis comparing biomarker values collected at autopsy across exposure levels may focus on the contrast

$$E_{X,0}\{B(1) - B(0)\} = E_{Z,0}(B \mid A = 1) - E_{Z,0}(B \mid A = 0) \ ,$$

which corresponds to the average treatment effect of the exposure on the biomarker value at autopsy. To illustrate why the interpretation of this estimand may be problematic, in Figure 2.1, a hypothetical scenario is depicted in which:

(i) irrespective of exposure level, biomarker values increase with time;

(ii) at any time, biomarker values are higher under exposure;

(iii) the exposed tend to have shorter survival time than the unexposed.

Despite these stipulations, the mean counterfactual biomarker value at death is the same for the exposed and unexposed. In fact, if in this illustration the difference between the two survival time distributions were increased, or the difference between the two latent biomarker curves were decreased, the mean counterfactual biomarker value at death would be greater in the unexposed. The example is constructed to highlight the critical role of the outcome measurement timing when analyzing autopsy data and the unfair comparison that ensues when contrasting the exposed and unexposed at their natural survival times. Indeed, the average treatment effect on the autopsy-measured biomarker value can be rewritten as

$$E_{X,0}\left\{B(1) - B(0)\right\} = E_{\mathbb{X},0}\left\{B_{T(1)}(1) - B_{T(0)}(0)\right\} ,$$

underscoring the difference in measurement times of the biomarker process across exposure levels.

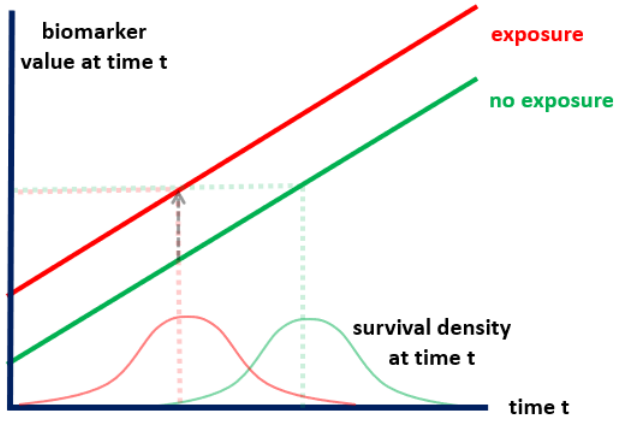

Figure 2.1: Hypothetical scenario to illustrate how a naive comparison across exposure levels may yield problematic findings. Darker solid lines each represent a curve of $t \mapsto B_t(a)$ for some value $a \in \{0, 1\}$, whereas lighter solid curves each represent a density function of $T(a)$ for some value $a \in \{0, 1\}$. Horizontal dashed lines denote the mean of $B(a)$, and vertical dashed lines denote the mean of $T(a)$.

The interpretation of a parameter that quantifies an effect partly due to systematic differences in survival time — the time at which the outcome is observed — may not be useful to describe the causal effect of the exposure on the biomarker process. In the example (Figure 2.1), the true mean difference of the biomarker values collected at autopsy is zero, even though by construction biomarker levels are always higher in the exposed versus unexposed. We expect a meaningful summary to reveal a positive effect of exposure on biomarker levels in such setting. This observation is distinct from existing methodological contributions to the analysis of autopsy data, which typically seek to unravel sampling biases when inferring on more traditional regression parameters (Haneuse et al., 2009). Rather, we focus on formulating a causal estimand more specifically suited to addressing the underlying scientific question and on establishing how to learn such an estimand in the context of various sampling complexities.

As an alternative to the average treatment effect, we consider a causal parameter that incorporates survival time standardization across exposure groups. Specifically, we focus on the contrast

$$\theta_0 := E_{\mathbb{X},0}\left\{B_{T(1)}(1) - B_{T(1)}(0)\right\} ,$$

which we refer to as the *natural-time direct effect* (NTDE). The nomenclature emphasizes the parallel to the natural direct effect (NDE) (Greenland et al., 1999; Rubin, 2004; VanderWeele, 2015a), which is commonly used to describe the effect of an exposure on an outcome through mechanisms not involving a particular mediating variable. In this spirit, the NTDE compares the average counterfactual biomarker values corresponding to exposure and non-exposure *at the natural survival time of an individual under assignment to exposure*. Here, in the parallel to mediation analysis, survival time may be considered to play the role of the mediator. The standardized contrast is represented by the vertical arrow in Figure 2.1, denoting the difference between counterfactual outcomes at time $T(1)$. The choice to standardize the autopsy outcomes at the survival time under exposure is motivated by the fact that we are considering settings in which the exposure is potentially harmful (and never beneficial) to

survival. In such cases, it is reasonable to expect an individual to survive at least until time $T(1)$ even under non-exposure. If instead the exposure is potentially beneficial (and never harmful) to survival, standardization at $T(0)$ would be more suitable.

Framing the causal analysis of autopsy data using ideas from the mediation literature, it is natural to wonder whether the time-dependent comparison

$$\beta_0(t) := E_{\mathbb{X},0}\left\{B_t(1) - B_t(0)\right\} ,$$

which mirrors the definition of the controlled direct effect, may provide an alternative contrast of interest. In mediation analysis, the controlled direct effect estimand is sometimes preferred because its interpretation involves a more fine-tuned control of the mediator. However, the interpretation of $\beta_0(t)$ as a time-dependent controlled direct effect can be problematic in the autopsy setting. Indeed, at any time t, the populations of individuals surviving to time t under assignment to exposure versus non-exposure may differ, so that in general $\beta_0(t)$ represents a contrast across different populations. As a remedy, it may be tempting to define $\beta_0(t)$ conditionally upon survival to time t under both assignment to exposure and to non-exposure, or symbolically, on $\min\{T(1), T(0)\} \geq t$. While this may render $\beta_0(t)$ a fairer contrast at any given t, the curve of $\beta_0(t)$ versus t still suffers from the same problem: the populations for which $\min\{T(1), T(0)\} \geq t$ differ across t. Thus, the curve may fail to provide a meaningful interpretation and may even be misleading in practice. Investigators could instead specify a fixed time τ at which $\beta_0(\tau)$ should be interpreted. While this approach resolves the interpretability issue, the choice of τ is arbitrary and subject to a trade-off. Small values of τ describe a large eligible population but correspond to fewer observed autopsy outcomes. In contrast, large values of τ are expected to yield a greater number of observable autopsy outcomes but the population surviving to τ is more limited and therefore less relevant to answering the scientific question at hand. Defining the estimand in terms of each individual's (random) counterfactual survival time addresses both these limitations, as the entire population is included in the definition of the estimand, and the time of comparison

is practically meaningful. For these reasons, we consider the NTDE to be a more natural estimand to address causal questions in autopsy studies.

2.2.3 Identification in ideal observational setting

Before tackling more complex sampling schemes, we consider the problem of recovering the causal estimand of interest from the distribution $P_{Z,0}$ of the ideal observable data unit Z, since this will serve as a springboard onto the more challenging settings. The causal estimand θ_0 represents a contrast of mean counterfactual biomarker values across exposure levels. Identification of the parameter in terms of the distribution $P_{Z,0}$ from observable data can be difficult because, even in studies in which investigators can randomize the exposure level A, the survival-outcome relationship may still be confounded. This challenge can be overcome if a sufficiently rich covariate vector L is available such that deconfounding is possible, a strategy often employed in mediation analysis (Tchetgen and Shpitser, 2012). This requirement can be formalized in terms of the following causal identification conditions:

(A0) the stable unit treatment value assumption holds;

(A1) the inequality $T(1) \leq T(0)$ holds $P_{X,0}$–almost surely;

(A2) the following conditions hold for each $a \in \{0, 1\}$ and $P_{X,0}$–almost every $\ell \in \mathcal{L}$:

- **i.** $P_{X,0}(A = a \,|\, L = \ell) > 0$;
- **ii.** $T(a)$ and A are independent given $L = \ell$ under $P_{X,0}$.

Condition (A0) represents a standard requirement for relating counterfactual and observable random variables as well as for defining meaningful average causal contrasts at the individual level. More precisely, this conditions requires consistency, which specifies that $B = B(1)A + B(0)(1 - A)$ and $T := T(1)A + T(0)(1 - A)$, and absence of any interference, which states that the exposure level of an individual cannot have an effect on the outcome of any other individual. Condition (A1) is a monotonicity assumption that imposes that the exposure never be beneficial to survival in the target population; a similar condition is leveraged in

the use of instrumental variables for causal inference (Angrist et al., 1996). If investigators are uncertain *a priori* of the direction of the effect of exposure on survival, an alternative approach consists of defining the causal estimand conditionally on the subset of individuals for whom condition (A1) is satisfied. Condition (A2) requires positivity of the probability of exposure and conditional exchangeability of the exposure-survival relationship within each stratum of the population defined by a sufficiently rich vector L of potential confounders. Of course, this condition is automatically satisfied if the exposure status is randomized at baseline.

Additional conditions are needed on the biomarker-survival relationship for causal identification. We denote by $\mathcal{S}(a, \ell)$ the support of the distribution of T given $A = a$ and $L = \ell$ implied by $P_{Z,0}$. We introduce the following conditions:

(A3) the inclusion $\mathcal{S}(1, \ell) \subseteq \mathcal{S}(0, \ell)$ holds for $P_{Z,0}$–almost every $\ell \in \mathcal{L}$;

(A4) for $P_{Z,0}$–almost every $\ell \in \mathcal{L}$, the following conditions hold for each $a \in \{0, 1\}$ and $P_{Z,0}$–almost every $t \in \mathcal{S}(0, \ell)$:

- **i.** $B_t(a)$ and A are independent given $T(a) = t$ and $L = \ell$;
- **ii.** $B_t(0)$ and $T(a)$ are independent given $T(0) \geq t$ and $L = \ell$.

Condition (A3) ensures that the counterfactual biomarker process $\{B_t(0)\}$ is well-defined at $T(1)$ in the target population — this can be viewed as requiring the absence of immortal time windows, within which it is possible to experience death under exposure level $A = 1$ but not $A = 0$. Condition (A4)i. requires conditional exchangability of the biomarker process and exposure level within each stratum defined by covariate and survival time values. This condition ensures that the counterfactual biomarker value observed at autopsy has a comparable distribution across exposure levels within strata defined by death time and covariate vector values. Condition (A4)ii. expresses two different independence requirements. For $a = 1$, this implies a modified *cross-world* independence condition, commonly required in mediation analysis (Richardson and Robins, 2013; Pearl, 2014; VanderWeele, 2015b). For $a = 0$, this instead requires that, within strata defined by covariate values, and among

unexposed survivors at any given time t, there be no direct effect of the latent biomarker value $B_t(0)$ on residual survival.

2.2.4 Observed data parameter

The stated conditions (A0) – (A4) are sufficient for identification of the natural-time direct effect in the ideal observable data setting, and suggests that we study the observed-data parameter

$$\psi_0 := E_{Z,0}\Big\{ E_{Z,0}\Big[E_{Z,0}\left(B \,|\, T, A=1, L\right) - E_{Z,0}\left(B \,|\, T, A=0, L\right) \,\Big|\, A=1, L \Big] \Big\} \,. \tag{2.1}$$

The identification formula given above is identical to that arising in the study of the NDE in mediation analysis (VanderWeele, 2015a). Nevertheless, several of the causal identifiability conditions stated above are novel — this is not surprising since the setting we are studying is not in fact one of mediation analysis. Indeed, the survival time is not a true mediator; rather, it defines the time interval over which the latent biomarker process is defined, and thus indexes the observation time for the autopsy outcome. Yet, the survival time plays the role of the mediating variable in the observable data identification. As a consequence, all existing results for statistical inference on the NDE, including those of Petersen et al. (2006); Tchetgen and Shpitser (2012); Zheng and van der Laan (2012), can be directly used to study the natural-time direct effect in this ideal observable data setting and under the stated causal identification conditions. As such, for brevity of discussion, we do not reproduce these results explicitly.

We note here that, when certain simple regression models hold, ψ_0 can be expressed in terms of regression coefficients. For example, if the biomarker regression model

$$E_{Z,0}(B \,|\, T=t, A=a, L=\ell) = \alpha_0 + \alpha_T t + \alpha_A a + \alpha_L \ell \,,$$

then ψ_0 is simply equal to the exposure regression coefficient α_A, which agrees with common

practice in parametric mediation analysis (Baron and Kenny, 1986; Imai et al., 2010). However, even in the ideal data setting, regression coefficients can provide misleading conclusions when the regression model is incorrectly specified (Zhao et al., 2010). Our work circumvents the pitfalls of interpreting ψ_0 on the basis of restrictive (and unnecessary) modeling assumptions by using a nonparametric formulation of the problem and inferential strategy. This approach also differs from other strategies of assessing potential bias from autopsy studies using model-based approaches (Haneuse et al., 2009; Rennert and Xie, 2019).

Researchers have previously investigated the properties of the NDE under various scenarios, including in the context of missing outcomes (Emsley et al., 2010), instrumental variables (Small, 2011; Frölich and Huber, 2017), and unmeasured confounding (VanderWeele, 2013). However, there has only been limited work on the problem of inference for the NDE identification formula in settings in which the mediator is subject to a combination of censoring and truncation. Existing work have been restricted to parametric modeling approaches (Wang and Zhang, 2011; Schultzberg, 2016; Wang and Shete, 2018) and inverse weighting approaches (Wang et al., 2019; Chernofsky et al., 2021). To the best of our knowledge, nonparametric efficient inference for this problem has not been previously studied. Such study is necessary for inference on the NTDE in autopsy studies, because the observed data typically do not arise from the ideal observable data distribution $P_{Z,0}$. Usually, only a coarsened version O of X may observed due to right censoring. Additionally, in many applications, the sampling distribution of X from which O is obtained is systematically biased relative to the target population due to either left or right truncation. The following sections provide conditions and results for statistical inference on the NTDE under prospective and retrospective sampling designs.

2.3 Prospective study design

2.3.1 Description of sampling design

The first design we consider involves prospective follow-up of individuals enrolled during a study period. Individuals in the target population may be recruited into the study at their own time zero or later, and are followed from enrollment until death, end of the study or loss to follow-up. We assume that a participant's baseline covariates can be measured even if the participant entered the study after time zero, and that the outcome is either measured at autopsy or missing. The observed data structure is $O := (L, A, W, Y, \Delta, U)$ with $\Delta := I(T \leq C)$, $Y := \min(T, C)$ and $U := \Delta B$ denoting the event indicator, observed follow-up time and observed outcome, respectively, and C and W are censoring and study entry (truncation) times, respectively. In view of delayed entry, only individuals that satisfy the left truncation condition $W \leq T$ can possibly be recruited, resulting in a systematically biased sample of the target population (Tsai et al., 1987). If an individual is lost to follow-up or has not died by the end of the study — that is, $\Delta = 0$ for this individual — their survival time T is only partially known, and their biomarker value B at death is missing (and set to zero by convention). Below, to simplify the discussion, we focus on the setting in which $W = 0$ for every participant, in which all participants are followed from time zero and on. However, all results in this chapter can be extended to the general case of delayed entry using the results derived in Chapter 1. Details are provided in Appendix B.2.

2.3.2 Notation and identification

Additional conditions are required to identify the natural-time direct effect when the data available are subject to right censoring of the survival time. To streamline the presentation of the additional conditions needed, we redefine the full observable data unit as $Z := (L, A, T, C, B)$ — this is simply an augmented version of the data unit provided in the previous section, namely also including the censoring time C. The identification presented in Section 2.2 remains valid using this augmented definition of the ideal observable

data unit Z, as the expression for ψ_0 given in Equation 2.1 is simply a summary of the distribution $P_{Z,0}$ that depends only on the distribution of (L, A, T, B) implied by $P_{Z,0}$. We suppose that the observed data consist of independent draws $O_1, O_2, \ldots, O_n$ from the distribution P_0 of O implied by $P_{Z,0}$.

We consider first the problem of recovering $P_{Z,0}$ from the observed data distribution P_0. To do so, we define the upper bound of the support of the conditional censoring time distribution

$$\tau(a, \ell) := \sup\{c : P_{Z,0}\left(C \geq c \,|\, A = a, L = \ell\right) > 0\},$$

which restricts the survival time values that can even be observed. For each $\ell \in \mathcal{L}$, we define pointwise the propensity score $\pi_{Z,0}(a \,|\, \ell) := P_{Z,0}(A = a \,|\, L = \ell)$ and the biomarker regression function $\varphi_{Z,0}(t, a, \ell) := E_{Z,0}(B \,|\, T = t, A = a, L = \ell)$. We make use of the following additional conditions, which we suppose to hold for $P_{Z,0}$–almost every $\ell \in \mathcal{L}$:

(B1) $(a, t) \mapsto \varphi_{Z,0}(t, a, \ell)$ is constant for $t > \tau(a, \ell)$;

(B2) $\pi_{Z,0}(a \,|\, \ell) > 0$ for each $a \in \{0, 1\}$;

(B3) (T, B) and C are independent given $(A, L) = (a, \ell)$ for each $a \in \{0, 1\}$.

Condition (B1) requires that, within strata defined by exposure and covariate vector values, the mean biomarker value at different survival times be the same for all times beyond the maximum observable time. This allows identification of the estimand of interest even when the censoring mechanism precludes identification of the right tail of the survival time distribution. If this assumption is implausible, then the inferential procedures described below can be viewed as targeting instead a version of the natural-time direct effect wherein time is restricted to be no greater than some fixed upper bound at which identification is possible, akin to the restricted mean survival time (Andersen et al., 2004; Díaz et al., 2019). While it resolves the lack of identification possibly resulting from right censoring, this modified estimand may be more challenging to interpret. Condition (B2) — typically referred to as a positivity condition — requires that, at the population level, it be possible to observe all

exposure levels within each subpopulation defined by values of the covariate vector. If individuals in certain subgroups of the population can never experience a particular exposure level, this condition would be violated; to remedy such violation, the target population could be redefined to exclude such subgroups. Finally, condition (B3) requires that, within strata defined by exposure and covariate vector values, the censoring random variable be independent of the survival and biomarker random variables. This condition allows the censoring mechanism to be informative but only insofar as this informativeness is entirely explained by the exposure and covariate vector values. The censoring variable induces a dual coarsening mechanism — right-censoring of the survival times and missingness of the biomarker values — and this condition is critical to be able to use the partial information provided by the observed data unit O to recover relevant portions of the underlying ideal observable data distribution $P_{Z,0}$.

Suppose that conditions (B1)–(B3) hold. For $0 < t < \tau(a, \ell)$, the target conditional time-to-event distribution function $F_{Z,0}(t \,|\, a, \ell) := P_{Z,0}(T \leq t \,|\, A = a, L = \ell)$ is identified via conditional product-integration (Gill and Johansen, 1990) by

$$\tilde{F}_0(t \,|\, a, \ell) := 1 - \prod_{u \leq t} \left[1 - \frac{F_{1,0}(du \,|\, a, \ell)}{R_0(u \,|\, a, \ell)} \right],$$

where $F_{1,0}(u \,|\, a, \ell) := P_0\left(Y \leq y, \Delta = 1 \,|\, A = a, L = \ell\right)$ is an observable conditional subdistribution function and $R_0(u \,|\, a, \ell) := P_0\left(Y \geq u \,|\, A = a, L = \ell\right)$ is an observable conditional at-risk probability. Because the data collection mechanism neither biases nor coarsens the target covariate distribution $H_{Z,0}(\ell) := P_{Z,0}(L \leq \ell)$, the latter is simply identified by $H_0(\ell) := P_0(L \leq \ell)$. The outcome regression is identified by the observable regression $\varphi_0(y, a, \ell) := E_0(U \,|\, \Delta = 1, Y = y, A = a, L = \ell)$.

Theorem 5. *If conditions (A0)–(A4) and (B1)–(B3) hold, then it holds that*

$$\theta_0 = \iint \left[\varphi_0(y, 1, \ell) - \varphi_0(y, 0, \ell)\right] \tilde{F}_0(dy \,|\, 1, \ell)\, H_0(d\ell) \ .$$

Beyond the natural-time direct effect, this result provides a nonparametric identification for the natural direct effect when the mediator is subject to right censoring. The corresponding result for settings in which the mediator is subject to both right censoring and left truncation is provided in the appendix.

2.3.3 Study of the parameter

The identification result above suggests that we study the statistical parameter

$$\Psi_{a_0} : P \mapsto \iint \varphi_P\,(y, a_0, \ell)\,\tilde{F}_P(dy \,|\, 1, \ell)\,H_P(d\ell)\ , \tag{2.2}$$

where, for any given distribution P, the functions φ_P, $\tilde{F}_P$ and H_P are obtained by substituting P in place of P_0 in the definition of φ_0, $\tilde{F}_0$ and H_0. Theorem 5 indicates that we can write $\theta_0 = \psi_{1,0} - \psi_{0,0}$ with $\psi_{a_0,0} := \Psi_{a_0}(P_0)$ for $a_0 \in \{0, 1\}$. In view of this fact, in the remainder of this section, we focus on developing inferential methods for $\psi_{a_0,0}$.

We note that even though $\psi_{a_0,0}$ is explicitly a summary of the observed-data distribution P_0, since P_0 is implied by the full observable-data distribution $P_{Z,0}$, $\psi_{a_0,0}$ can be represented in terms of components of P_0, of $P_{Z,0}$, or a combination thereof. This gives rise to multiple possible parametrizations for $\Psi_{a_0}(P)$. In this work, we focus on parametrization in terms of components of the full observed-data distribution since it provides greater interpretability for conditions arising in our study. For a given full observed-data distribution P_Z, we define pointwise the propensity score $\pi_Z(a \,|\, \ell) := P_Z(A = a \,|\, L = \ell)$, the conditional censoring survival function $Q_Z(y \,|\, a, \ell) := P_Z(C > y \,|\, A = a, L = \ell)$, and the conditional time-to-event distribution and survival functions $F_Z(y \,|\, a, \ell)$ and $S_Z(y \,|\, a, \ell) := 1 - F_Z(y \,|\, a, \ell)$ with corresponding density function $f_Z(y \,|\, a, \ell)$ and cumulative hazard function $\Lambda_Z(u \,|\, a, \ell)$. We also define the survival regression function $\mu_Z(a, \ell) := \int \varphi_Z(y, a, \ell)\,F_Z(dy \,|\, 1, \ell)$ and partial reverse regression function $D_{Z,\varphi}(y, a, \ell) := \int_y^\infty S_Z(u \,|\, 1, \ell)\,\varphi_Z(du, a, \ell)$.

As we establish below, a linearization of $P \mapsto \Psi_{a_0}(P)$ is possible based on the influence

function $\phi_{a_0,P} := \phi_{a_0,1,P} + \phi_{a_0,2,P} + \phi_{a_0,3,P}$, where we define

$$\phi_{a_0,1,P} : o \mapsto \frac{\delta}{Q_{Z,P}(y \mid a_0, \ell)} \frac{I(a = a_0)}{\pi_{Z,P}(a_0 \mid \ell)} \frac{f_{Z,P}(y \mid 1, \ell)}{f_{Z,P}(y \mid a_0, \ell)} [u - \varphi_{Z,P}(y, a_0, \ell)]$$
$$\phi_{a_0,2,P} : o \mapsto -\frac{I(a = 1)}{\pi_{Z,P}(1 \mid \ell)} \left[\frac{\delta D_{Z,\varphi,P}(y, a_0, \ell)}{S_{Z,P}(y \mid 1, \ell) Q_{Z,P}(y \mid 1, \ell)} - \int_0^y \frac{D_{Z,\varphi,P}(u \mid a_0, \ell) \Lambda_{Z,P}(du \mid 1, \ell)}{S_{Z,P}(u \mid 1, \ell) Q_{Z,P}(u \mid 1, \ell)} \right]$$
$$\phi_{a_0,3,P} : o \mapsto \mu_{Z,P}(a_0, \ell) - \Psi_{a_0}(P) .$$

Equipped with these definitions, we can provide a description of the behavior of the parameter $\Psi_{a_0}(P)$ in a neighborhood of $P = P_0$ within a nonparametric model.

Theorem 6. *Suppose that Conditions (A0)–(A4) and (B1)–(B3) hold. Then, the functional* $P \mapsto \Psi_{a_0}(P)$ *is a pathwise differentiable parameter with nonparametric efficient influence function* ϕ_P*. Furthermore, for each* P*, the linearization*

$$\Psi_{a_0}(P) - \Psi_{a_0}(P_0) = \int \phi_{a_0,P}(o)(P - P_0)(do) + R(P, P_0)$$

holds with second-order remainder $R(P, P_0)$ *defined in the Appendix.*

Results from this theorem enable us to derive a debiased machine learning inferential strategy for learning the natural-time direct effect from a prospective cohort study design.

2.3.4 Estimation and inference

Equation 2.2 expresses the identified natural-time direct effect value in terms of components of the observed data distribution P_0. A straightforward estimation strategy would consist of substituting each nuisance appearing in (2.2) by a corresponding nuisance estimator. However, despite its simplicity, an estimator obtained in this manner suffers from important drawbacks. In order to ensure nondegenerate root-n distribution results, only the simplest of learning approaches (e.g., parametric models) can be used when building nuisance estimators for use with such estimator. Any flexible learning approaches (e.g., machine learning) can typically not be used, and rendering systematic bias due to model misspecification more

likely and thereby possibly invalidating inference. Additionally, the simple plug-in strategy results in an estimator that does not benefit from any robustness to inconsistent estimation of nuisance functions. Both these drawbacks are related to the fact that, without additional work, the behavior of the simple plug-in estimator is sensitive to the behavior of each of its constituent nuisance function estimators. In contrast, the use of debiased machine learning results in an estimator that is relatively insensitive to its nuisance estimators, thereby allowing valid inference with machine learning tool and ascribing a degree of robustness to the resulting estimator.

In contrast to the estimation strategy we outlined in Chapter 1, here we consider a representation of ψ_0 and of the relevant efficient influence function provided in Theorem 6 in terms of nuisance functions indexed by $P_{Z,0}$. The debiased machine learning procedure we describe relies on an estimator $\eta_n := (\varphi_n, F_n, Q_n, \pi_n)$ of the nuisance vector $\eta_0 := (\varphi_{Z,0}, F_{Z,0}, Q_0, \pi_{Z,0})$. To discuss robustness properties of our proposed procedure, we suppose that η_n converge in probability to some nuisance vector value $\eta_\infty := (\varphi_\infty, F_\infty, Q_\infty, \pi_\infty)$ that may differ from η_0. Among several possible strategies, the influence function $\phi_{a_0,P}$ could be used in a one-step debiasing procedure or instead as an estimating function for $\psi_{a_0,0}$. In the setting including only right censoring of the survival time along with missingness of the outcome for individual with censored times, these two distinct strategies result in exactly the same estimator. Specifically, the resulting debiased machine learning estimator of the evaluation of the NTDE identification formula $\psi_0 := \psi_{1,0} - \psi_{0,0}$ is given by

$$\psi_n^* := \frac{1}{n} \sum_{i=1}^{n} \left[\mu_n(1, L_i) - \mu_n(0, L_i) \right] + \frac{1}{n} \sum_{i=1}^{n} \left[\phi_{1,n}(O_i) - \phi_{0,n}(O_i) \right],$$

where we have defined $\mu_n : (a, \ell) \mapsto \int \varphi_n(u, \ell) \, F_n(du \,|\, a, \ell)$ and $\phi_{a_0,n} := \phi_{a_0,1,\eta_n} - \phi_{a_0,2,\eta_n}$ with ϕ_{a_0,j,η_n} referring to $\psi_{a_0,j,P}$ but with relevant components of P replaced by corresponding components of η_n. We describe the implementation of this estimator in greater detail in Algorithm 1. We note here that the first summand in the definition of ψ_n^* is simply the plug-in estimator based on the identification formula whereas the second summand is the

debiasing term that endows the estimator with improved statistical properties.

Algorithm 1 Outline of estimation procedure

1: Obtain estimate φ_n of the outcome regression, π_n of the propensity score, and Q_n of the censoring survival function;
2: Obtain estimate F_n of the survival distribution function, Λ_n of the cumulative hazard function, and m_n of the density ratio function;
3: Obtain estimate μ_n of the survival regression integral and $D_{n,\varphi}$ of the partial reverse regression function based on numerical approximations of integrals

$$\mu_n := \int_u \varphi_n(u,\ell) F_n(du \,|\, 1,\ell) \text{ and } D_{n,\varphi} := S_n(t \,|\, 1,\ell)\varphi(t,\ell) - \int_{u \geq t} \varphi_n(u,\ell) F_n(du \,|\, 1,\ell)\,;$$

4: Construct plug-in estimate $\psi_n := \frac{1}{n}\sum_{i=1}^n [\mu_n(1,L_i) - \mu_n(0,L_i)]$;
5: Calculate estimated influence function $\phi_{a_0,n} := \phi_{a_0,1,n} + \phi_{a_0,2,n} + \phi_{a_0,3,n}$ with

$$\phi_{a_0,1,n}(\ell,a,y,\delta,u) = \frac{\delta}{Q_n(y \,|\, a_0,\ell)} \frac{I(a=a_0)}{\pi_n(a_0 \,|\, \ell)} m_n(y,a_0,\ell)\,[u - \varphi_n(y,a_0,\ell)]$$

$$\phi_{a_0,2,n}(\ell,a,y,\delta,u) = -\frac{I(a=1)}{\pi_n(1 \,|\, \ell)} \left[\frac{\delta D_{n,\varphi}(y,a_0,\ell)}{S_n(y \,|\, 1,\ell) Q_n(y \,|\, 1,\ell)} - \int_0^y \frac{D_{n,\varphi}(u \,|\, a_0,\ell)\Lambda_n(du \,|\, 1,\ell)}{S_n(u \,|\, 1,\ell) Q_n(u \,|\, 1,\ell)} \right]$$

and $\phi_{a_0,3,n}(\ell,a,y,\delta,u) = \mu_n(a_0,\ell) - \psi_n$;
6: Calculate the debiased estimate

$$\psi_n^* := \psi_n + \frac{1}{n}\sum_{i=1}^n \{\phi_{1,n}(O_i) - \phi_{0,n}(O_i)\}\ .$$

The following theorem states that ψ_n admits large-sample distributional approximations that make it suitable as a basis for constructing confidence intervals and conducting hypothesis tests under certain conditions described in Appendix B.2.

Theorem 7. *Under conditions (B5)–(B8) enumerated in Appendix B.2, ψ_n^* is an asymptotically linear estimator of ψ_0 with influence function $\phi_{1,P_0} - \phi_{0,P_0}$.*

This theorem immediately implies, by the weak law of large numbers, that ψ_n^* is a consistent estimator of ψ_0, and by the central limit theorem, that $n^{\frac{1}{2}}(\psi_n - \psi_0)$ tends to a mean-zero normal random variable with variance $\sigma_0^2 := var_0\,[\phi_{1,P_0}(O) - \phi_{0,P_0}(O)]$.

To establish Theorem 7, a remainder term emanating from the linearization provide in Theorem 6 must be controlled along with an empirical process term that arises in the study of $\psi_n - \psi_0$. For the former, we require that nuisance function estimators converge to the true nuisance functions at a sufficiently fast rate in an appropriate sense. The use of ensemble learning strategies can help achieve this condition by adaptively leveraging a variety of different structures (e.g., smoothness, sparsity) to improve estimation rates. For the latter, the learning strategies used to estimate nuisance functions must be constrained in complexity so that the Donsker conditions imposed can be satisfied. Such conditions can preclude the use of particularly flexible learning strategies. In order to circumvent this condition and allow a greater range of learning approaches to be used, the estimator ψ_n can be constructed using cross-fitting (see, e.g., Zheng and Laan, 2011), as described explicitly in Chapter 1. In the context of the simulation study presented below, $F_{Z,0}$ and Q_0 are estimated based on Cox proportional hazards models, whereas both $\varphi(t, a, \ell)$ and $\pi(a \mid \ell)$ are estimated using a generalized addative model and an empirical estimator, respectively. This ensure that these regularity conditions are satisfied irrespective of the dimension of L. Of course, in practice, a practitioner may not know that the Cox model holds in a given application, which would motivate the use of more complex (e.g., machine learning) strategies for nuisance estimation.

While Theorem 7 describes the behavior of ψ_n when all relevant nuisance functions are estimated consistently, it is interesting to note that ψ_n retains consistency even when certain nuisance functions are inconsistently estimated. Specifically, consistency of ψ_n to the target limit ψ_0 only requires only two out of the three statements below to be true:

(i) $\varphi_\infty(Y, A, L) = \varphi_{Z,0}(Y, A, L)$ with P_0–probability one;

(ii) $(Q_\infty(Y \mid A, L), \pi_\infty(A \mid L)) = (Q_{Z,0}(Y \mid A, L), \pi_{Z,0}(A \mid L))$ with P_0–probability one;

(iii) $F_\infty(Y \mid A, L) = F_{Z,0}(Y \mid A, L)$ with P_0–probability one.

This robustness property is similar to the multiple robustness that arises in nonparametric inference for the natural direct effect in the absence of censoring — see Tchetgen and Shpitser (2012) for details. It does differ, however, in that to benefit from consistent estimation of

the propensity score, the censoring survival function must also be consistently estimated. In the context of randomized clinical trials, it may be possible to ensure that Condition (ii) is satisfied by design if censoring can only occur administratively (e.g., as imposed by study investigators or end of the study period).

We note that while the two debiased machine learning strategies alluded to above result in the same estimator in the context of right censoring alone, this is not the case when there is also left truncation in the sampling scheme due to delayed entry into the prospective study. Specifically, in this more complex scenario, the two distinct strategies yield estimators that are not only different but also have different robustness profiles — this is discussed explicitly in Chapter 1.

2.4 Retrospective study design

2.4.1 Description of sampling design

Prospective study designs, as described in the previous section, can be challenging to implement because of the need for lengthy follow-up in order to observe a sufficiently large number of deaths and thereby record autopsy outcomes. Even under ideal conditions, the cost of follow-up may be expensive and logistically complicated, leading to minimal information about individuals with longer survival times. As an alternative, practitioners may instead take a cross-sectional sample of an existing database of autopsy records. In such case, an individual may only possibly recruited in the study if their survival time T is no greater than the time W elapsed between their time zero and the time at which the database is queried; symbolically, sampling is conditional on the criterion $T \leq W$, which is an example of selection bias due to right truncation. Such retrospective design can be conducted quickly and does not require prospective follow-up, often making it resource-efficient. However, when one of the exposure levels is rare, in order to recruit enough individuals having experienced that exposure level so that reasonably precise inferences can be drawn, a large sample may be required. In contrast, in a prospective follow-up study, investigators may determine by de-

sign the relative representation of individuals with different exposure levels. Finally, we note that while the analysis of outcomes collected from retrospective autopsy studies has been discussed before (Rennert and Xie, 2019), the analytic approaches adopted have relied on strong model structure in order to produce valid inference. Here, we adopt a nonparametric approach and use debiased machine learning tools for estimation and inference.

2.4.2 Notation and identification

In the presence of right truncation, the observed data are drawn from the subpopulation of individuals for whom $T \leq W$. For all sampled individuals, we suppose that there is access to baseline information (L, A) and that the outcome is measured at time T. We define an augmented version of the full data unit, $Z := (L, A, W, T, B) \sim P_{Z,0}$; as in the previous section, the expression for the target parameter in terms of the full data distribution is unchanged by this updated definition of the data unit. The observed data consist of independent draws $O_1, O_2, \ldots, O_n$ from the conditional distribution P_0 of Z given $T \leq W$ induced by $P_{Z,0}$.

We consider the problem of recovering relevant portions of $P_{Z,0}$ from the observed data distribution P_0. To do so, we must be able to recover the support $\mathcal{S}(a, \ell)$ of the conditional target survival time distribution. We define the upper support bounds

$$\tau_T(a, \ell) := \sup\{t : P_{Z,0}(T \geq t \,|\, A = a, L = \ell) > 0\},$$
$$\tau_W(a, \ell) := \sup\{w : P_{Z,0}(W \geq w \,|\, A = a, L = \ell) > 0\},$$

which we use to describe the support from which observations may be observed. We denote by $\pi_{Z,0}(a \,|\, \ell) := P_{Z,0}(A = a \,|\, L = \ell)$ the propensity score and $\varphi_{Z,0}(t, a, \ell) := E_{Z,0}(B \,|\, T = t, A = a, L = \ell)$ the biomarker regression. We suppose that, for $P_{Z,0}$–almost every $\ell \in \mathcal{L}$ and for each $a \in \{0, 1\}$, the following conditions hold:

(C1) for some $\alpha > 0$, it holds that $\tau_W(a, \ell) > \tau_T(a, \ell) + \alpha$;

(C2) it holds that $\pi_{Z,0}(a \mid \ell) > 0$;

(C3) (T, B) and W are independent given $(A, L) = (a, \ell)$.

Condition (C1) requires that, within strata defined by the exposure and covariate vector values, the sampling scheme permits the entire support of the target survival time distribution to be recoverable. Similarly as in the prospective design, if the sampling design systematically excludes a portion of the population, then it is not possible to recover information from that subpopulation without extrapolation. In such cases, the inferential procedures described below can be viewed as targeting a version of the natural-time direct effect wherein time is restricted to be no greater than some fixed upper bound at which identification is possible. Condition (C2) requires that, at a population level, it is possible to observe all exposure levels within each subpopulation defined by the values of the covariate vector. As discussed before, if individuals in certain subgroups of the population can never experience a particular exposure level, this condition would be violated; to remedy such violation, the target population could be redefined to exclude such subgroups. Finally, condition (C3) requires that, within strata defined by the exposure and covariate vector values, both the truncation-survival relationship and the truncation-outcome relationship be independent. When time zero represents birth or any fixed age (e.g., 65 years, as in our motivating example), under cross-sectional retrospective sampling, the truncation time conveys information about an individual's birth cohort. In such case, we interpret this condition as requiring sufficient information in the collection of covariates to describe the variability in outcomes across different birth cohorts.

Under conditions (C1)–(C3), the target distribution $P_{Z,0}$ may be expressed in terms of P_0. For $t \in (0, \infty)$, the target conditional survival distribution $F_{Z,0}(t \mid a, \ell)$ is identified via conditional product-integration integral (Woodroofe, 1985) by

$$\tilde{F}_0(t \mid a, \ell) := \prod_{u \geq t} \left\{ 1 - \frac{F_0(du \mid a, \ell)}{R_0(u \mid a, \ell)} \right\} ,$$

where $F_0(u \,|\, a, \ell) := P_0(T \leq u \,|\, A = a, L = \ell)$ and $R_0(u \,|\, a, \ell) := P_0(T \leq u \leq W \,|\, A = a, L = \ell)$ both describe summaries of the observed data distribution and are implicitly defined by the sampling inclusion $T \leq W$. An added complexity, due to the systematic bias induced by the truncation of event times, is that the observed covariate distribution is no longer representative of the target covariate distribution. We must express the target covariate distribution in terms of the observed covariate distribution. This can be done using inverse-probability-of-inclusion weighting — for example, as in Chan and Wang (2012) and Cheng and Wang (2012) — leading to $\tilde{H}_0(d\ell) := \gamma(\ell) H_0(d\ell)$ for weight function $\ell \mapsto \gamma(\ell) := \sum_{a \in \{0,1\}} \gamma(a, \ell) \pi_0(a \,|\, \ell)$, where $\gamma(a, \ell) := -\int \tilde{F}_0(w \,|\, a, \ell)^{-1} G_0(dw \,|\, a, \ell)$ with $G_0(w \,|\, a, \ell) := P_0(W > w \,|\, A = a, L = \ell)$ the observed data truncation survival function, and $\gamma := \int \gamma(\ell) H_0(d\ell)$ is a normalizing constant. We note that $\gamma(a, \ell)$ is in fact equal to $P_{Z,0}(T \leq W \,|\, A = a, L = \ell)^{-1}$.

Theorem 8. *If conditions (A0)–(A4) and (C1)–(C3) hold for, then it holds that*

$$\theta_0 = \iint [\varphi_0(t, 1, \ell) - \varphi_0(t, 0, \ell)] \tilde{F}_0(dt \,|\, 1, \ell) \tilde{H}_0(d\ell) \ .$$

2.4.3 Study of the parameter

The identification result above suggests that we study the statistical parameter

$$\Psi_{a_0} : P \mapsto \iint \varphi_P(t, a_0, \ell) \, \tilde{F}_P(dt \,|\, 1, \ell) \, \tilde{H}_P(d\ell) \,, \tag{2.3}$$

where, for any given distribution P, the functions φ_P, $\tilde{F}_P$ and $\tilde{H}_P$ are obtained by substituting P in place of P_0 in the definition of φ_0, $\tilde{F}_0$ and $\tilde{H}_0$. Theorem 8 indicates that we can write $\theta_0 = \psi_{1,0} - \psi_{0,0}$ with $\psi_{a_0,0} := \Psi_{a_0}(P_0)$ for $a_0 \in \{0, 1\}$. In view of this fact, in the remainder of this section, we focus on developing inferential methods for $\psi_{a_0,0}$.

In this work, as in the previous section, we focus on parametrization in terms of components of the full observed-data distribution since it provides greater interpretability for conditions arising in our study. However, in practice it may prove at times prove more convenient to use a mixed parametrization for estimation. For a given full observed-data

distribution P_Z, we define pointwise the propensity score $\pi_Z(a \,|\, \ell) := P_Z(A = a \,|\, L = \ell)$, the truncation survival function $G_Z(w \,|\, a, \ell) := P_Z(W > w \,|\, A = a, L = \ell)$, and the survival distribution function $F_Z(t \,|\, a, \ell) := P_Z(T \leq t \,|\, A = a, L = \ell)$ and its corresponding density function $f_Z(t \,|\, a, \ell)$ and reverse-time cumulative hazard function $\Gamma_Z(t \,|\, a, \ell) := \log F_Z(t \,|\, a, \ell)$. We also define pointwise the survival regression function $\mu_Z(a, \ell) := \int \varphi_Z(t, a, \ell)\, F_Z(dt \,|\, 1, \ell)$, the marginal inclusion probability $\gamma_Z := P_Z(T \leq W)$, and the partial regression function $D_{Z,\varphi} : (t, a, \ell) \mapsto \int_0^t F_Z(u \,|\, 1, \ell)\, \varphi_Z(du, a, \ell)$.

As we establish below, the nonparametric linearization of $P \mapsto \Psi_{a_0}(P)$ hinges on the influence function $\phi_{a_0,P} := \frac{1}{\gamma_Z}(\phi_{a_0,1,P} + \phi_{a_0,2,P} + \phi_{a_0,3,P})$, where we define

$$
\begin{aligned}
\phi_{a_0,1,P}(o) &: o \mapsto \frac{I(a = a_0)}{\pi_Z(a_0 \,|\, \ell)} \frac{f_Z(t \,|\, 1, \ell)}{f_Z(t \,|\, a_0, \ell)} \frac{1}{G_Z(t \,|\, a_0, \ell)} \left[b - \varphi_Z(t, a_0, \ell) \right]; \\
\phi_{a_0,2,P}(o) &: o \mapsto -\frac{I(a = 1)}{\pi_Z(1 \mid \ell)} \left[\frac{D_{Z,\varphi}(t, a_0, \ell)}{F_Z(t \,|\, 1, \ell) G_Z(t \,|\, 1, \ell)} - \int_t^w \frac{D_{Z,\varphi}(u, a_0, \ell)}{G_0(u \,|\, a, l) F_Z(u \mid 1, \ell)} \Gamma_Z(du \,|\, 1, \ell) \right]; \\
\phi_{a_0,3,P}(o) &: o \mapsto \left[\mu_Z(a_0, \ell) - \Psi_{a_0}(P) \right] \left[\frac{1}{F_Z(w \,|\, a, \ell)} + \phi_{F,Z}(t, w, a, \ell) \right]; \\
\phi_{F,Z,P}(o) &: o \mapsto \frac{1 - G_Z(t \,|\, a, \ell)}{F_Z(t \,|\, a, \ell) G_Z(t \mid a, \ell)} - \int_t^w \frac{[1 - G_Z(u \,|\, a, \ell)]}{F_Z(u \,|\, a, \ell) G_Z(u \,|\, a, \ell)} \Gamma_Z(du \,|\, a, \ell)\,.
\end{aligned}
$$

Equipped with these definitions, we provide a description of the behavior of the parameter $\Psi_{a_0}(P)$ in a neighborhood of $P = P_0$ within a nonparametric model.

Theorem 9. *Suppose that conditions (A0)–(A4) and (C1)–(C3) hold. Then, the functional $P \mapsto \Psi_{a_0}(P)$ is a pathwise differentiable parameter with nonparametric efficient influence function $\phi_{a_0,P}$. Furthermore, for each P, the linearization*

$$\Psi_{a_0}(P) - \Psi_{a_0}(P_0) = \int \phi_{a_0,0}(o)(P - P_0)(do) + R(P, P_0)$$

holds with second-order remainder $R(P, P_0)$ defined in the Appendix.

Results from this theorem enable us to derive a debiased machine learning inferential strategy for learning the NTDE from a retrospective cohort study design.

2.4.4 Estimation and inference

Similarly as before, we outline a strategy for estimating the value θ_0 of the NTDE identification formula parameter when data are collected through a cross-sectional retrospective cohort sample. In contrast to the prospective sampling design, the one-step debiased estimation procedure and Neyman-orthogonal estimating equations approach do not necessarily yield equivalent estimators. In this work, we use an estimating equations approach as it provides an intuitive multiple robustness property. Still, as in the prospective setting, estimation of the target parameter requires that we first estimate the relevant components of the data-generating distribution $P_{Z,0}$. The debiased machine learning procedure we describe relies on an estimator $\eta_n := (\varphi_n, F_n, G_n, \pi_n)$ of the nuisance vector $\eta_0 := (\varphi_{Z,0}, F_{Z,0}, G_0, \pi_{Z,0})$. To discuss robustness properties of our proposed procedure, we suppose that η_n converges in probability to some nuisance vector value $\eta_\infty := (\varphi_\infty, F_\infty, G_\infty, \pi_\infty)$ that may differ from η_0. We consider the estimating function given by the mapping $U(\psi; o) : \psi \mapsto \phi_{1,0,\psi}(o) - \phi_{0,0,\psi}(o)$, where $\phi_{a_0,0,\psi} := \phi_{a_0,1,P_0} + \phi_{a_0,2,P_0} + \phi_{a_0,3,P,\psi}$ with $\phi_{a_0,3,P,\psi}$ defined exactly like $\phi_{a_0,3,P_0}$ but with $\Psi_{a_0}(P_0)$ replaced by ψ. It can be shown that the target parameter value ψ_0 uniquely solves the estimating equation $E_0\{U(\psi; O)\} = 0$. The resulting empirical estimating equation has a unique, closed-form solution, and the resulting estimator is asymptotically linear under certain regularity conditions — details are provided in Appendix B.2. We provide an overview of the estimation strategy for retrospective sampling designs in Algorithm 2.

Theorem 10. *Under conditions (C4)–(C7) enumerated in Appendix B.2, ψ_n^* is an asymptotically linear estimator of ψ_0 with influence function $\phi_{1,P_0} - \phi_{0,P_0}$.*

As before, this theorem immediately implies, by the weak law of large numbers, that ψ_n^* is a consistent estimator of ψ_0, and by the central limit theorem, that $n^{\frac{1}{2}}(\psi_n^* - \psi_0)$ tends to a mean-zero normal random variable with variance $\sigma_0^2 := var_0\left[\phi_{1,P_0}(O) - \phi_{0,P_0}(O)\right]$. The chosen construction admits a multiple robustness property. Specifically, consistency of ψ_n^* to the target limit ψ_0 requires only two out of the three statements below to be true:

(i) $\varphi_\infty(Y, A, L) = \varphi_{Z,0}(Y, A, L)$ with P_0–probability one;

(ii) $(G_\infty(Y \mid A, L), \pi_\infty(A \mid L)) = (G_{Z,0}(Y \mid A, L), \pi_{Z,0}(A \mid L))$ with P_0–probability one;

(iii) $F_\infty(Y \mid A, L) = F_{Z,0}(Y \mid A, L)$ with P_0–probability one.

Algorithm 2 Outline of estimation procedure

1: Obtain estimate φ_n of the outcome regression, π_n of the propensity score, and G_n of truncation time distribution;

2: Obtain estimate S_n of the survival distribution function, Γ_n of the backwards hazard function and m_n of the density ratio function;

3: Obtain estimate μ_n of the survival regression integral, $D_{n,\varphi}$ of the partial survival regression function, and $\phi_{F,n}$ of the weight function based on numerical approximations of integrals

$$\begin{aligned}
\mu_n &:= \int_u \varphi_n(u, \ell) F_n(du \mid 1, \ell) \\
D_{n,\varphi} &:= F_n(t \mid 1, \ell)\varphi(t, \ell) - \int_{u \le t} \varphi_n(u, \ell) F_n(du \mid 1, \ell) \\
\phi_{F,n} &:= F_n(w \mid a, \ell)^{-1} + \frac{1 - G_n(t \mid a, \ell)}{F_n(t \mid a, \ell) G_n(t \mid a, \ell)} - \int_{t \le u \le w} \frac{[1 - G_n(u \mid a, \ell)]\Gamma_n(du \mid a, z)}{F_n(u \mid a, \ell) G_n(u \mid a, \ell)}
\end{aligned}$$

4: We define the estimating function $U_n(\psi; O_i) : \psi \mapsto \phi_{1,n,\psi} - \phi_{0,n,\psi}$ such that $\phi_{a_0,n,\psi} := \phi_{a_0,1,n,\psi} + \phi_{a_0,2,n,\psi}$, and

$$\begin{aligned}
\phi_{a_0,1,n}(o) &= \frac{1}{G_n(t \mid a_0, \ell)} \frac{I(a = a_0)}{\pi_n(a_0 \mid \ell)} m_n(t, a_0, \ell) \left[b - \varphi_n(t, a_0, \ell)\right] \\
&\quad - \frac{I(a = 1)}{\pi_n(1 \mid \ell)} \left[\frac{D_{n,\varphi}(t, a_0, \ell)}{F_n(t \mid 1, \ell) G_n(t \mid 1, \ell)} - \int_{t \le u \le w} \frac{D_{n,\varphi}(u \mid a_0, \ell)\Gamma_n(du \mid 1, \ell)}{F_n(u \mid 1, \ell) G_n(u \mid 1, \ell)} \right] \\
\phi_{a_0,2,n,\psi}(o) &= \left(\mu_n(a_0, \ell) - \psi\right) \phi_{F,n}(\ell, a, w, t)
\end{aligned}$$

5: The estimator ψ_n^*, defined as

$$\psi_n^* := \frac{\sum_{i=1}^n [\phi_{1,1,n}(O_i) - \phi_{0,1,n}(O_i)] + \phi_{F,n}(O_i) \left[\mu_n(1, L_i) - \mu_n(0, L_i)\right]}{\sum_{i=1}^n \phi_{F,n}(O_i)},$$

is the solution of the equation $\frac{1}{n} \sum_{i=1}^n U_n(\psi; O_i) = 0$

However, in order to benefit fully from this robustness property, additional work may be needed to devise nuisance function estimators that are themselves multiply-robust. Indeed,

in our current implementation, the nuisance estimator π_n is constructed using an estimator of the observed-data propensity as well as estimators of the survival and truncation distributions — the latter are used to relate the observed-data propensity score to the target propensity score. Therefore, for the proposed estimator to be consistent, condition (iii) must be satisfied along with at least one of conditions (i)–(ii). Determining whether the target population propensity score could be estimated directly remains an important open question.

2.5 Simulation study

2.5.1 Description of data-generating mechanism

We conducted a series of numerical studies using simulated data to assess the finite sample performance of the proposed methods. To construct a random data unit, we first generated a covariate vector $L := (L_1, L_2, L_3)$ with L_1 and L_2 independent random variables distributed uniformly on the set $\{-1, +1\}$ and L_3 a Bernoulli random variable with conditional success probability $\text{expit}(L_1)$ given (L_1, L_2). Given $(L_1, L_2, L_3) = (\ell_1, \ell_2, \ell_2)$, we generated A as a Bernoulli random variable with conditional success probability given by $\text{expit}[-1 + \log\{1 + \exp(\ell_1) + \exp(-\ell_2)\}]$. Given $(A, L_1, L_2, L_3) = (a, \ell_1, \ell_2, \ell_3)$, we generate T according to a Gamma random variable with shape parameter 6 and scale parameter

$$\lambda_T(a, \ell_1, \ell_2, \ell_3) = \exp\left(\frac{-1 + 3a + 2\ell_2 - \ell_3}{5}\right).$$

given $(T, A, L_1, L_2, L_3) = (t, a, \ell_1, \ell_2, \ell_3)$, we generated Z from a normal distribution with mean $\varphi_{Z,0}(t, a, \ell_1, \ell_2, \ell_3) = 1 + t + \frac{3}{2}\sin(\frac{3}{2}t) - 2a - \ell_2 + \ell_3$ and unit variance. In this particular data-generating mechanism, there is no interaction in the biomarker regression model between A and T; as such, the NTDE value is simply given by the value -2 of coefficient of A in the biomarker regression model. A visual depiction of the variables at play is provided in Figure 2.2. Conditions (A2)–(A4) are satisfied by this data-generating mechanism.

We further introduce systematic bias in the sampling mechanism by including left truncation of survival times in addition to right censoring. Given $(A, L_1, L_2, L_3) = (a, \ell_1, \ell_2, \ell_3)$, we

generated the truncation variable W as $10U$, where U is a Beta random variable with parameters $\alpha(a, \ell_1, \ell_2, \ell_3)$ and $\beta(a, \ell_1, \ell_2, \ell_3)$ selected to achieve a certain truncation level. For the low truncation scenario (25% truncation), we set $\alpha(a, \ell_1, \ell_2, \ell_3) = 1$ and $\beta(a, \ell_1, \ell_2, \ell_3) = 1 + 2I(\ell_1 < 0)$, whereas for the high truncation scenario (50% scenario), we set $\alpha(a, \ell_1, \ell_2, \ell_3) = 1 + I(\ell_1 > 0)$ and $\beta(a, \ell_1, \ell_2, \ell_3) = 1$. Given $(T, W, A, L_1, L_2, L_3) = (t, w, a, \ell_1, \ell_2, \ell_3)$, the censoring time C was generated as $W + V$, where V follows a Gamma distribution with scale parameter $\lambda_C = \exp(1 + 0.1L_2 - 0.1L_3)$ and shape parameter chosen to yield a censoring rate of 25% (low censoring) or 50% (high censoring).

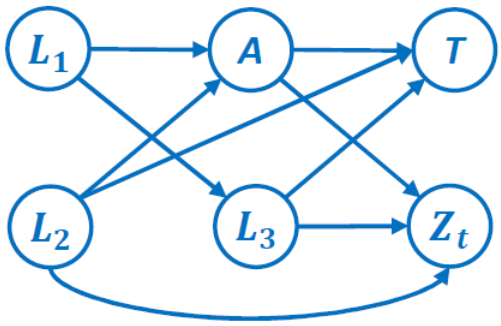

Figure 2.2: Directed acyclic graph representing the data-generatin mechanism underlying the simulation study, excluding truncation and censoring variable.

2.5.2 Results of simulation studies

We simulated 100 datasets with sample size n taking values in $\{250, 500, 750, 1000\}$. We computed the proposed estimator ψ_n^* of the the identified NTDE value $\psi_0 = -2$. For each nuisance parameter, we detail the estimation approach used for simulation. We used a generalized additive model (GAM) with identify and logistic links, respectively, to estimate the outcome regression $\varphi_{Z,0}(t, a_0, \ell)$ and observed data propensity score $\pi_0(\ell)$. In practice, as

suggested before, we recommend using ensemble learning strategies. As a caveat, whenever L is high-dimensional, propensity score estimates may require trimming to avoid values close to the boundary of $(0, 1)$. The survival, truncation and censoring distributions were estimated using penalized spline functions for survival data using the `survPen` package in `R`.

To quantify the performance of the proposed estimator in this simulation study, we examined the finite sample bias and confidence interval coverage. We compared the performance of the proposed estimator using GAM and penalized spline nuisance estimators to one based on correctly-specified parametric models for all nuisance functions. To emphasize the need for debiasing, we also compared the proposed debiased estimator to the plug-in estimator without debiasing. In all settings, the proposed estimator was found to be approximately unbiased and constructed 95% confidence intervals had empirical coverage near the nominal level, similarly as the procedure based on parametric nuisance function estimators. In the high truncation setting, some under-coverage was observed in confidence intervals at smaller sample sizes due to numerical instability in the nuisance function estimators; in the larger sample sizes, the coverage reached the nominal level. As expected, the undebiased plug-in estimator was excessively biased, which resulted in poor interval coverage, thus emphasizing the need for debiasing when using flexible algorithms for learning nuisance functions.

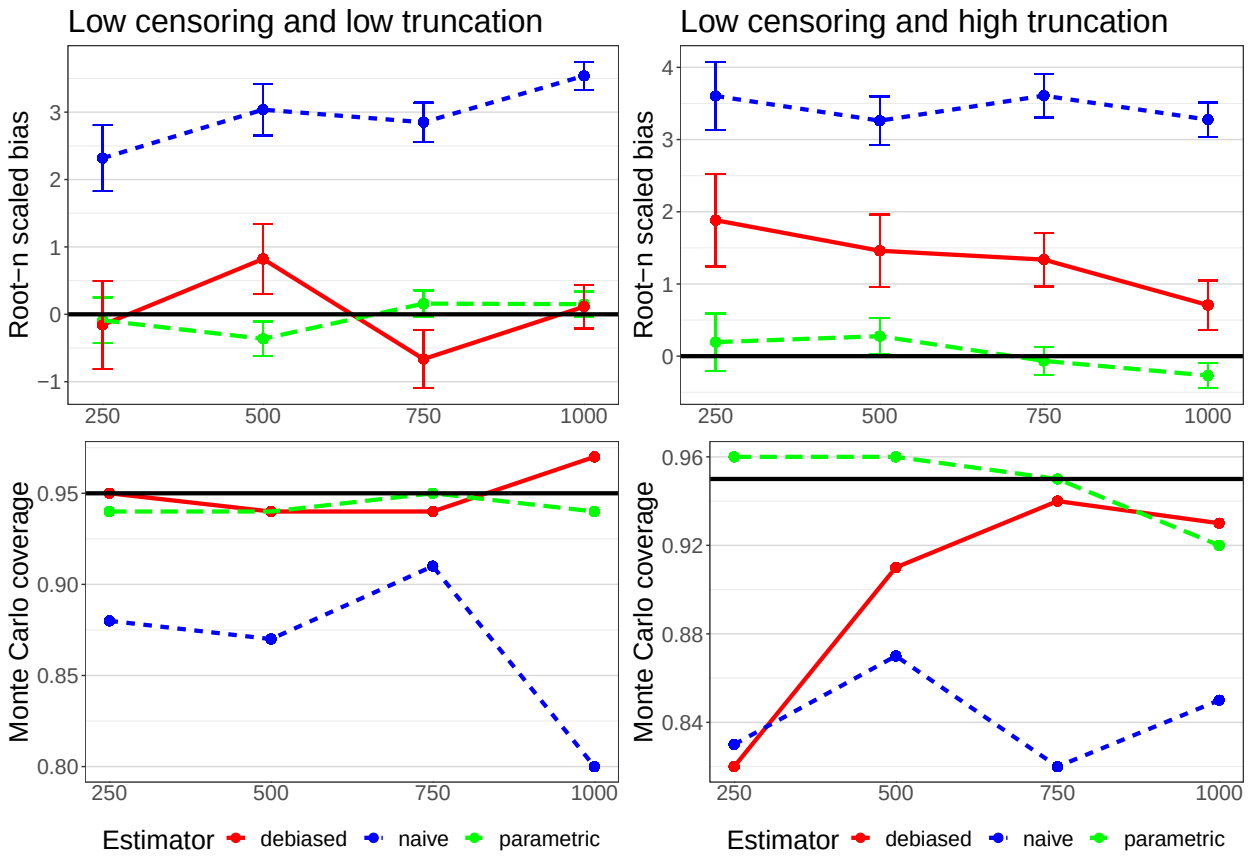

Figure 2.3: Empirical bias and coverage from simulation study for setting in which the survival times are observed subject to 25% censoring, and 25% truncation (left), 50% truncation (right). Values on the x-axis denote sample sizes.

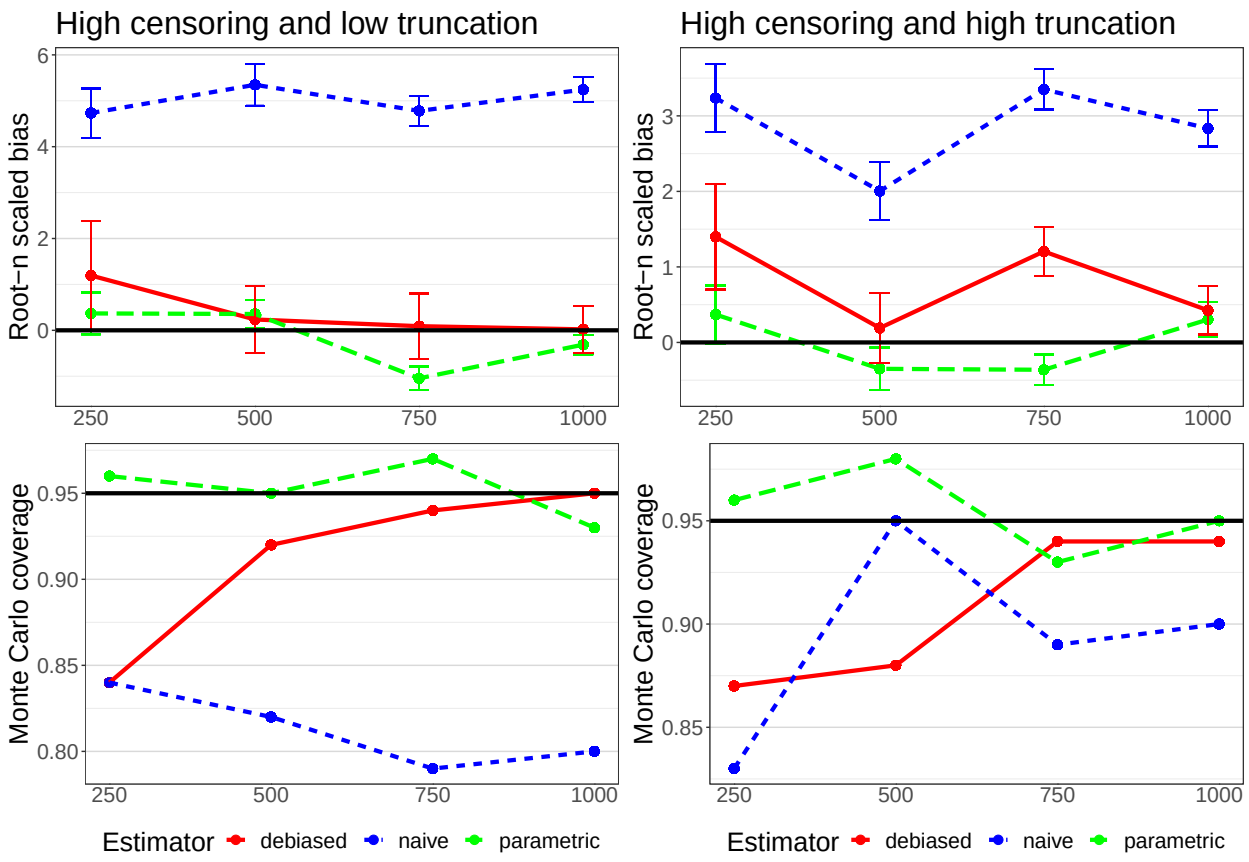

Figure 2.4: Empirical bias and coverage from simulation study for setting in which the survival times are observed subject to 50% censoring, and 25% truncation (left), 50% truncation (right). Values on the x-axis denote sample sizes.

2.6 Results from the ACT study

2.6.1 Study description

Alzheimer's disease and related dementias (ADRD), a group of neurological diseases involving progressive cognitive decline (involving, for example, memory loss, behavioral changes and speech difficulties), is the fifth leading cause of death worldwide and one of the greatest sources of morbidity and dependency among elderly individuals (Gaugler et al., 2022). It represents a major burden for patient caregivers and health systems. Alzheimer's disease is characterized by the presence of extracellular amyloid β_{1-42} (Aβ_{1-42}) plaques and interneuronal tau aggregations (neurofibrillary tangles) in the brain. These alterations, among others, disrupt cell-to-cell communication and transport, and trigger pathological inflammatory processes (Selkoe, 1991; Serrano-Pozo et al., 2011). The presence of the ϵ4–allele of the gene encoding apolipoprotein E (APOE-4) has been associated with increased mortality and a decline in cognitive function (Tilvis et al., 2004). Neuritic plaque density in the cerebral cortex is commonly quantified using the Consortium to Establish a Registry for Alzheimer's disease (CERAD) score (taking values between 1 and 3) (Mirra et al., 1991), whereas the neurofibrillary tangle distribution is assessed using the Braak score (taking values between 1 and 6) (Braak and Braak, 1991). Here, using the method developed in this paper and autopsy data from the Adult Changes in Thought (ACT) study, we sought to understand the effect of a positive APOE-4 gene expression on these two established correlates of cognitive decline and Alzheimer's disease progression.

The ACT study is a longitudinal community-based prospective cohort study conducted to study brain aging and the epidemiology of dementia (Kukull et al., 2002). The ACT study cohort is comprised of an urban and suburban elderly population ($>$ 65 years at enrollment) randomly recruited from a well-established health maintenance organization (Kaiser Permanente of Washington) in the Puget Sound area. Enrollment of cognitively intact (defined as Cognitive Abilities Screening Instrument score of $>$85 or consensus diagnosis of "not demented") individuals began in 1994-1996 and continued until 2020. Participants could

consent to an autopsy upon death during the study period; this was discussed at study enrollment and follow-up visits. Consent was also confirmed by the next-of-kin at the time of participant death, as required by Washington State law. For this analysis, we use data from all participants in the cohort but only had autopsy outcomes for individuals having both consented to autopsy and died at the time of last data freeze (October 2021). For the 921 participants with evaluable autopsy brain tissues, preservation and evaluation procedures have been described elsewhere (Sonnen et al., 2007).

2.6.2 Methods

In our analysis, we made certain assumptions to facilitate study of the natural-time direct effect (NTDE) of APOE-4 on AD neuropathologic markers. We defined time $t = 0$ to be age 65 since all participants had to be at leat 65 years old at recruitment, and assumed that the available covariates were either measured at baseline or represented relatively unchanging features over time. Given our current identification strategy, we assumed that the exposure A is not a cause of any component of the covariate vector L, which possibly confounds the outcome-mortality relationship. The covariates used in the full analysis include:

- $L_1 = \{$ BMI status, diabetes diagnosis, history of cardiovascular disease, hypertension, education, smoking history, alcohol history, birth cohort, sex, race $\}$;
- $L_2 = \{$ race, birth cohort $\}$.

A DAG representation of the relationship between the variables of interest is illustrated in Figure 2.5; the structure described ensures that conditions (A2)–(A4) hold.

To implement the estimation procedure proposed in this paper, the outcome regression function was estimated using splines with all other variables (i.e., covariates, exposure) possibly interacting with APOE-4; the propensity score and truncation distribution were estimated empirically; and the survival and censoring distributions were estimated using survival splines implemented on the hazard scale. These estimators performed well in simulation and were considered sufficiently numerically stable.

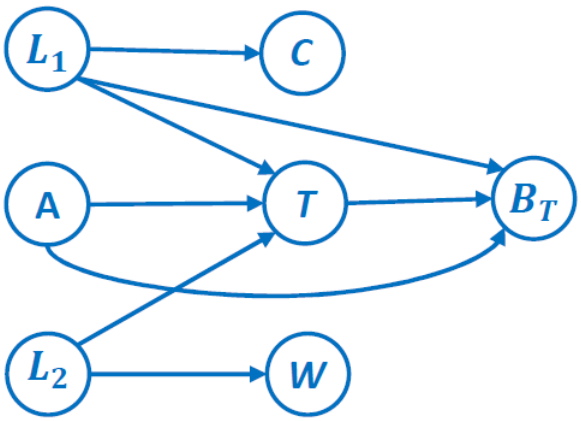

Figure 2.5: Directed acyclic graph of the ACT study autopsy data analysis. Variables include $A :=$ APOE-4 gene presence, $T :=$ survival time, $B_T :=$ neuropathological marker at autopsy, $(L_1, L_2) :=$ potential confounders, $W :=$ enrollment time, $C :=$ censoring time.

Selection bias is a challenge in the analysis of autopsy outcomes collected from patients enrolled in the ACT study cohort (Tsuang et al., 2005; Haneuse et al., 2009). As indicated before, participant or next-of-kin consent was required to record an individual's AD neuropathologic markers at autopsy. To address the missingness in autopsy outcomes due to lack of consent, we used an inverse-probability-of-consent weighting approach using variables in L_2. For stability, estimated consent probabilities were truncated within the interval $(0.1, 0.9)$ to reduce the risk of few individuals having undue influence in the analysis; this approach has been used in previous ACT autopsy data analyses (Shaffer et al., 2021b).

2.6.3 Results from ACT study

Characteristics of individuals in the ACT study cohort as a whole (Shaffer et al., 2021a) and within the subgroup of participants having provided consent to autopsy (Shaffer et al., 2021b) have been previously described—the summary characteristics presented in the referenced work used data collected before September 2018. The key variables used in our NTDE

analysis are summarized in Table B.3. On average, for each baseline covariate, there was less than 7% missingness across participants. These missing values were imputed by the mode for categorical variables and the mean for quantitative variables. The cohort consisted of older individuals at enrollment, with a majority being 75 years of age at enrollment. Participants who died during follow-up were older, with a median age of 88.2 years (range from 66 to 108 years old). Participants spanned birth cohorts between 1890-1950, with a large portion (40%) born between 1930-1950. The cohort included 89.3% white participants; 40.1% participants with a positive APOE ϵ4–allele; 33.1% participants with a recorded autopsy consent; and 16.1% participants with recorded autopsy measures. The distribution of observed Braak scores and CERAD scores are summarized in Table B.3. As the APOE-4 gene is defined at birth, the effect of APOE-4 on AD neuropathologic markers observed at autopsy is unconfounded. When excluding participants who did not consent to autopsy, the average treatment effect of a positive APOE-4 gene expression on Braak score at death was estimated to be 0.44 (95% CI: 0.18 to 0.70); when accounting for consent, the estimate was instead 0.47 (95% CI: 0.20 to 0.75). In contrast, the NTDE of a positive APOE-4 gene expression on Braak score was estimated to be 0.58 (95% CI: 0.31 to 0.85) when excluding participants who did not provide consent, and instead 0.45 (95% CI: 0.09 to 0.81) when adjusting for consent. Results for the CERAD score were similar and described in Appendix B.1.

To interpret the autopsy average treatment effect and NTDE as an effect of APOE-4 on AD neuropathologic markers, we return to the causal question described in Section 2.2. In Figure 2.6, we highlight the similarity of the outcome and survival time distributions to the hypothetical scenario we constructed and summarized in Figure 2.1. The estimated survival distribution for ACT study participants with a positive APOE-4 gene expression was slightly more skewed to the left compared to those that have negative gene expression. In this setting, the difference in Braak scores by APOE-4 gene expression could have led to a reversal in direction of association if the difference in mortality had been greater. In this setting, the difference in survival was relatively small across exposure groups, which resulted in similar estimates of the autopsy ATE and NTDE. However, in general, the interpretation

of the autopsy ATE can still lead to confusion, while the NTDE is defined to account for potential differences in mortality.

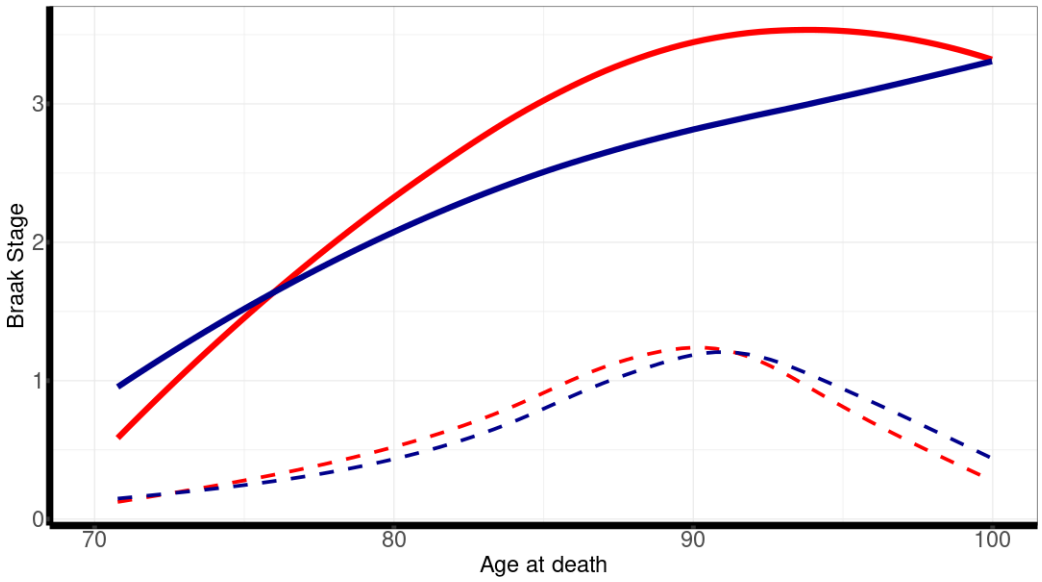

Figure 2.6: Red indicates the APOE-4 positive group, while Blue denotes the APOE-4 negative group. Solid lines indicate the marginal mean Braak stage observed at time-of-death, while dashed lines are overlayed survival time densities.

2.7 Concluding remarks

We have developed and studied a framework for summarizing the causal effect of a point exposure on a time-varying biomarker process observed only at death. The causal estimand we proposed quantifies a marginal contrast in counterfactual outcomes after eliminating exposure-induced differences in survival, and provides a natural scientific interpretation. In the absence of survival complications, the identified expression for the natural-time direct effect is exactly that for the natural direct effect even though the definition presented does not explicitly describe a mediation problem. Regardless, the results for nonparametric identification, estimation and inference we derived can be viewed as an extension of existing result from the mediation literature to settings in which the mediating variable is subject to

censoring, truncation or both.

The natural-time direct effect is preferable to a naive contrast across exposure groups and should be considered as possible target of interest in prospective and retrospective autopsy designs. The causal framework we build upon draws attention to the necessity of accounting for covariates that may confound the survival and biomarker outcome relationship when outcomes are observed at the time of death. Unadjusted comparisons can result in biased estimates due to systematic differences in observed outcomes, even in simple randomized trials. We believe that the identification formula for the natural-time direct effect is a reasonable statistical parameter even if the causal conditions stated do not strictly hold. Indeed, this estimand still describes a standardized summary of differences in autopsy outcomes, which is typically more interpretable than a marginal comparison across exposure groups in this problem.

In Chapter 1 we developed and studied debiased machine learning methods for making statistical inference on summaries of a counterfactual time-to-event distribution using left-truncated right-censored data. These results were critical to developing methods for debiased machine learning of the natural-time direct effect under the prospective and retrospective study designs considered. To the best of our knowledge, this work is the first to present a derivation of the nonparametric efficient influence function for the natural direct effect in the context of a right-censored mediating variable, let alone of left- or right-truncation on the mediating variable.

Our application focused on the point-exposure genetic biomarker APOE-4 and its potential effect on biomarkers of neurodegeneration. In this work, we do not provide results that accommodate time-varying exposures or time-varying covariates. In future work, we aim to extend our results to a time-varying setting for both exposure and covariates, as we believe this may lead to causal conditions that are more likely to hold in studies of toxic environmental exposures (Carone et al., 2020). These results could be used to establish the effect of important environmental exposures, such as air pollution or smoking, on cognitive decline. It may also be fruitful to consider alternative strategies for identifying the natural-

time direct effect estimand under different conditions that may be more natural in certain applications, including the use of instrumental variables. Current work in the mediation literature on stochastic effects may provide possible avenues for relaxing the most restrictive causal assumption (i.e., cross-world independence) used in this Chapter.

Chapter 3

QUANTIFYING RESIDUAL LIFETIME ACCELERATION FROM A POINT-EXPOSURE

Analyses based on the proportional hazard model are most common for studying time-to-event outcomes. Critics of this model have noted the difficulty in ascribing a causal interpretation to the hazard ratio even under ideal conditions. Given the increased interest of practitioners in formally studying causal relationships, even in observational contexts, models for which the statistical estimand of interest lends itself more readily to a causal interpretation, such as the accelerated failure time (AFT) model, may be preferred. However, in its standard implementation, the AFT model does not naturally handle situations in which the exposure occurs at a random point in time. To address this gap, we propose a novel *residual accelerated failure time model* that quantifies the association between a point-exposure and acceleration of residual lifetime, possibly as a function of circumstances of the exposure event itself, including the age at which exposure occurred. We show that parameters of this model admit a natural causal interpretation under typical conditions for causal identification. Focusing on a parametric implementation of the model, we show that maximum likelihood estimation can be used to conduct valid inference for model parameters in the presence of left truncation and right censoring. We illustrate the use of our proposed model and methods by quantifying the age-dependent association between radiation dose and acceleration of residual lifetime among Japanese survivors of the 1945 atomic bombings of Hiroshima and Nagasaki.

3.1 Introduction

A major objective of environmental epidemiology is to quantify the human health effects of physical, biologic, and chemical exposures from the external environment. Results from environmental epidemiology studies have established ambient air pollution, lead dust, and ionizing radiation, among others, as risk factors for adverse health outcomes and early mortality. Researchers conducting environmental epidemiology studies consider a number of important factors that are out of their control due to the observational nature of the design. For example, the age at which an individual is exposed and the extent of exposure (e.g., dose) can have a significant bearing on the effect of exposure on health outcomes. Therefore, it may be important to better understand the manner in which age-at-exposure and dose impact the effect of an environmental exposure in order to quantify its population-level impact. For example, the Life Span Study is a prospective cohort study of Japanese survivors of the 1945 atomic bombings of Hiroshima and Nagasaki that has quantified the long-term human health effects of acute radiation exposure (Ozasa et al., 2019). Analyses of the Life Span Study data with hazard-based statistical models have consistently demonstrated that increasing radiation dose leads to decrements in survival, and that age-at-exposure (i.e., age at the time of the atomic bombings) modifies the association between radiation dose and mortality (Ozasa et al., 2012).

When dealing with time-to-event outcomes, analysts often resort to the use of the Cox proportional hazards model (Cox, 1972) and the accelerated failure time (AFT) model (Wei, 1992). To account for differences in exposure effect due to variations in age-at-exposure, the implementation of such models would need to allow for such time dependence. In the Cox model, this can be easily done since the conditional definition of the hazard function naturally allows for incorporation of time-varying covariates. Even though the standard AFT model does not allow the use of time-varying covariates, modifications that do have been put forth. Most notably, the time-dependent accelerated failure time (TD-AFT) model was proposed to quantify associations between a time-dependent continuous exposure and time-to-event

outcomes (Cox and Oakes, 1984; Robins, 1992). However, neither the Cox model nor the TD-AFT model provide an entirely satisfying interpretation when making inference on how age-at-exposure modifies the exposure-outcome association. On one hand, when accommodating dependence of the hazard ratio on age through specification of a time-varying hazard ratios in a Cox model, the resulting time-varying parameter estimates are challenging to interpret causally and can be subject to bias from differential selection (Hernán, 2010). On the other hand, the formulation and implementation of the original TD-AFT specify a general time-dependent exposure model, which does not immediately accommodate age-specific effects of an acute exposure.

We overcome these important limitations by proposing a novel extension of the AFT model that explicitly allows a point-exposure to occur post-baseline. In particular, this allows us to study how an acute exposure alters the residual survival of exposed populations based on exposure at different points in life. Framing the problem with this structure is useful to quantify how the effect of a continuous exposure on residual lifetime is modified as a function of exposure timing. For this reason, we refer to the model as a *residual accelerated failure time (R-AFT)* model and provide conditions under which the model parameters admit a causal interpretation equivalent to those in structural failure time models (Robins, 1992; Lok et al., 2004).

The article is organized as follows. In Section 3.2, we formulate the R-AFT model for quantifying the age-specific association between a continuous exposure (e.g., radiation dose) and residual lifetime. We describe conditions under which the model parameters admit a causal interpretation and contrast the proposed model to existing methods. Parametric procedures for estimation and inference are described in Section 3.3. In Section 3.4, we examine the finite sample performance of the proposed estimation procedure in a simulation study. In section 3.5, we conduct an illustrative analysis of the Life Span Study data. Concluding remarks are provided in Section 3.6.

3.2 Proposed model

3.2.1 Notation

We are interested in quantifying the effect of an acute exposure on survival among individuals in a target population. For each individual in this population, we denote by Z a vector of baseline covariates, and by W the time elapsed between birth and possible enrollment into the study. If recruitment happens at a fixed point τ in calendar time, then W is simply given by $\tau - B$ with B denoting birth date. We denote by $E \in \{0, 1\}$ an indicator of exposure, and for an individual with $E = 1$, we denote by A and X the age at exposure and exposure dose, respectively. For any individual with $E = 0$, A and X are undefined. We denote by T the total lifetime and by C the age at which the individual is possibly lost to follow-up or otherwise right-censored. We define the data unit $Q := (Z, E, X, A, W, T, C)$ and denote its (unknown) distribution by P_0. The observed data unit is $O := (Z, E, X, A, W, Y, \Delta)$, where $Y := \min\{T, C\}$ and $\Delta := I(T \leq C)$. We denote by P_0^* the distribution of O conditional on the criterion $W \leq T$ and induced by P_0. The observed data consist of a sample of n independent draws $O_1, O_2, \ldots, O_n$ from P_0^*.

We suppose that the full-data generating mechanism satisfies the following conditions P_0–almost surely:

(A1) T and W are independent given (A, X, E, Z);

(A2) T and C are independent given (W, A, X, E, Z) and $T \geq W$;

(A3) $\min\{P_0(T \geq W \mid E), P_0(T \leq C \mid E, T \geq W)\} > 0$.

Condition (A1) requires conditional independence between total lifetime and truncation time, whereas condition (A2) requires conditional independence between the the total lifetime and censoring time within the observable (i.e., truncated) population. In particular, this allows censoring to only pertain for observable individuals; this is important since in many cases censoring is undefined outside the context of the study. A discussion of various possible assumptions for identification under left truncation and right censoring can be found in Qian

and Betensky (2014). Condition (A3) eliminates scenarios under which no individual could be sampled from the population of interest or no total lifetime could be fully observed. Under the conditions (A1)–(A3), the conditional density of (Y, Δ) given $(W, A, X, E, Z) = (w, a, x, e, z)$ can be expressed as

$$p_0^*(y, \delta \,|\, w, a, x, e, z) = \lambda_0(y \,|\, a, x, e, z)^{\delta} \frac{S_0(y \,|\, a, x, e, z)}{S_0(w \,|\, a, x, e, z)} \ ,$$

where λ_0 and S_0 denote the conditional hazard and survival functions corresponding to the distribution of T given (A, X, E, Z).

3.2.2 Model formulation

To motivate the R-AFT model, we first describe the null relationship from which we wish to measure deviations using model parameters. In the Life Span Study, survivors who were within three kilometers of the atomic hypocenter were potentially exposed to radiation; survivors who were more than three kilometers away were enrolled as an internal reference group and matched to proximal survivors on city, sex and age-at-exposure (Ozasa et al., 2019). If exposure is entirely inconsequential and does not alter an individual's life course, we may expect that

$$P_0\left(T > t \,|\, A = a, X = x, E = 1, Z = z\right) = P_0\left(T > t \,|\, T > a, E = 0, Z = z\right) \tag{3.1}$$

for each (z, x, a) and $t > a$. This states that, under the null setting in which the exposure has no effect, the survival profile of an individual exposed to dose x at age a is the same as that of an unexposed individual who survived until at least age a. In other words, the only information derived from an individual's exposure is the age they must have attained.

We define the dilated residual lifetime $u_\phi(z, x, a, t) := (t - a)\phi(z, x, a) + a$ for individuals in the exposed population. The acceleration factor $\phi(z, x, a) > 0$ quantifies the extent to which residual lifetime is accelerated (or decelerated) upon exposure. We propose the R-AFT model to relate the exposed and unexposed populations as

$$P_0\left(T > t \,|\, A = a, X = x, E = 1, Z = z\right) = P_0\left(T > u_\phi(z, x, a, t) \,|\, T > a, E = 0, Z = z\right)$$

for each (z, x, a) and $t > a$. Because residual lifetime survival probabilities are defined conditionally on survival to age a, this can be alternatively restated in terms of hazards as

$$\lambda(t \,|\, a, x, 1, z) = \phi(z, x, a)\lambda_0(u_\phi(z, x, a, t) \,|\, z)\,,$$

where $\lambda_0(\cdot \,|\, z)$ is the lifetime hazard function corresponding to the unexposed population with covariate profile z. The R-AFT model expresses the survival probability among the exposed population according to the survival probability among the unexposed based on a dilated or contracted residual time-scale. If $\phi(z, x, a) = 1$, the null setting is recovered. If $\phi(z, x, a) > 1$, then the exposure is harmful and results in a contraction of time upon exposure, whereas if instead $\phi(z, x, a) < 1$, the exposure is protective and a dilation of residual time occurs.

The acceleration function ϕ is typically the component of primary scientific interest in the R-AFT model. Analogously to the traditional AFT model, the acceleration factor $\phi(z, x, a)$ can be represented as the ratio of residual moments, namely as

$$\phi(z, x, a) = \frac{E_0\left[\, m(T - a) \,|\, T > a, E = 0, Z = z \,\right]}{E_0\left[\, m(T - a) \,|\, A = a, X = x, E = 1, Z = z \,\right]}\,, \tag{3.2}$$

for any monotone function $m : \mathbb{R} \to \mathbb{R}$; it can also be expressed as a ratio of corresponding quantiles. In particular, $\phi(z, x, a)^{-1}$ can be interpreted, among individuals with covariate profile z, as the relative difference in mean (or median) residual lifetime comparing individuals exposed to dose x at age a to unexposed individuals who survived until at least age a. For example, if $\phi(z, x, a) = 2$, then among individuals with covariate value z, the mean (or median) residual lifetime is halved among those exposed to dose x at age a relative to the unexposed who survived to age a.

For the sake of providing a simple, parsimonious summary of the contraction function ϕ, we formulate a parametric model that relates the acceleration factor to a parsimonious

summary of the covariates. A natural parametrization for the acceleration function is given by the log-linear form

$$\log \phi_\beta(z, x, a) = \beta + \beta_A a + \beta_X x + \beta_Z z \,,$$

which provides a simple representation of the interplay between residual lifetime and the exposure dose, age-at-exposure and baseline covariate vector. Under correct specification of the R-AFT model, since

$$\frac{\phi_\beta(z, x+1, a)}{\phi_\beta(z, x, a)} = \exp(\beta_X) \;,$$

the exponentiated coefficient $\exp(-\beta_X)$ represents the relative median residual survival comparing two exposed populations with same age-at-exposure, same baseline covariate vector, and exposure doses that differ by one unit. Of course, the log-linear model can be modified to include nonlinear or interaction terms similarly as with other regression models.

The straightforward interpretation of β_X described above does not readily extend to coefficients β_A and β_A because the denominator in (3.2) is not the same for subpopulations of individuals who differ by one unit in either their covariate vector or age-at-exposure values. The exponentiated term $\exp(-\beta - \beta_A a - \beta_Z z)$ represents the relative median residual survival comparing, among individuals with covariate vector value z, subpopulations of individuals exposed at age a to a null dose and unexposed populations surviving until at least age a. This difference in interpretations of the model coefficients differs from those of the Cox and AFT models, for which each coefficient can typically be interpreted as the ratio between *any* two populations that differ by one unit in the corresponding covariate. Due to this important distinction, the coefficients in the R-AFT model are most naturally viewed as measures of exposure interactions — and as we outline below, under certain conditions, as effect modifiers of exposure.

3.2.3 Causal interpretation of model parameters

An appealing property of the R-AFT model is that, under certain causal conditions, the acceleration factor $\phi(z, x, a)$ has a natural causal interpretation. Adopting the Neyman-Rubin potential outcomes framework (Neyman, 1923; Rubin, 1974), we denote by $T(0)$ the counterfactual survival time under no exposure, and conditional on $T(0) > a$, the counterfactual survival time $T(a, x, e)$ corresponding to an exposure at x and age a. We introduce the following conditions, under which a nonparametric causal interpretation of the acceleration factor $\phi(z, x, a)$ will be possible:

(B0) Interventions on one individual do not have repercussions on outcomes in other individuals.

(B1) The relationship $T = T(0) + I(T(0) > A, E = 1)[T(A, X, 1) - T(0)]$ between unintervened survival time T and counterfactual survival times holds P_0–almost surely.

(B2) $P_0\left(E = 1 \,|\, Z\right) > 0$ P_0–almost surely;

(B3) for P_0–almost every $A = a$ in the support of $T(0)$ and every $(X, E) = (x, e)$, $T(a, x, e)$ and (A, X, E) are conditionally independent given $T(0) > a$ and Z holds P_0–almost surely.

Condition (B0) describes the 'no interference' assumption, which essentially requires that an individual's counterfactual lifetime be independent of interventions carried out in any other individual. This most commonly fails in the context of infectious diseases. Condition (B1) is a statement of consistency of the observed survival outcomes, which indicates that the survival time observed corresponds to the counterfactual survival time corresponding to the exposure actually experienced. The statement of consistency is more complicated in our context because exposure only possibly happens at random time A, which must be accounted for carefully. Together, conditions (B0) and (B1) describe a version of the Stable Unit Treatment Value Assumption for this context. Condition (B2) is a positivity assumption, which requires that exposure or non-exposure both be possible in all strata defined by values of the baseline

covariate vector. Condition (B3) is a requirement that there be no unmeasured confounding between the exposure vector (A, X, E) and lifetimes among individuals who survive to age a under no exposure.

Theorem 11. *Under conditions (B0)–(B2), the acceleration factor admits the following representation as a causal contrast:*

$$\phi(z, x, a) = \frac{E_0\left[m(T(0) - a) \,|\, T(0) > a, Z = z\right]}{E_0\left[m(T(x, a, 1) - a) \,|\, T(0) > a, Z = z\right]} \ ,$$

where $m : \mathbb{R} \to \mathbb{R}$ is any monotone function.

This result states that, within the subpopulation of individuals who survive until at least age a in the absence of the exposure and who have baseline covariate value z, the reciprocal of the acceleration factor $\phi(z, x, a)$ can be interpreted as the relative change in mean (or median) residual lifetime that results from enforcing exposure at dose x at age a rather than assigning no exposure. Notably, the interpretation of $\phi(z, x, a)$ does not lend itself to a temporal interpretation in age a, as the reference population varies with a. This is a common problem in the interpretation of causal parameters in survival settings (Young et al., 2020). The causal interpretation of $\phi(z, x, a)$ agrees with the expression provided in structural nested failure time models, a class of structural distribution models used for causal inference (Lok et al., 2004; Vansteelandt and Joffe, 2014; Vock et al., 2013). In settings in which it is not possible to adjust for a suitably rich vector Z of covariates, the acceleration factor retains an interpretation as a conditional association between residual lifetime and exposure.

3.2.4 Comparison to alternative models

The R-AFT model is formulated to emphasize an interpretable age-dependent association between a point exposure occurring at a random time and residual survival. In contrast, the Cox and AFT model — the most popular statistical models for studying time-to-event endpoints — do not provide such interpretable summaries. The Cox model readily handles

time-dependent exposures but does not produce interpretable causal summaries for the effects of dose (Hernán, 2010). Further, hazards are limited by their interpretation being subject to differential selection bias (Robins, 1998).

While it is natural to wonder whether the R-AFT model is equivalent to an AFT model for the residual lifetime $T-A$, in reality, the two models are distinct. To illustrate this point, consider the acceleration factor $\phi(a) = \exp(\beta_A a)$ in the R-AFT model. Under the null setting in which exposure does not affect survival, by design, it will be the case that $\beta_A = 0$. However, when implementing an AFT model on the residual survival time scale with acceleration factor $\phi_{AFT}(a, e) = \exp(\beta_A ae)$ — the interaction term here is the relevant one since we are contrasting the exposed and unexposed population at a fixed age— the parameter β_A may not be zero. The difference in interpretation of these corresponding parameters results from the fact that age-at-exposure is a component of total lifetime. Therefore, age-at-exposure A and residual survival time $T - A$ are expected to be negatively correlated in most scenarios — and certainly in the null setting. Because in the absence of dose and baseline covariates the acceleration factor in the AFT model on the residual time is given by

$$\phi_{AFT}(a) = \frac{E_0\left[m(T-A) \mid A = 0, E = 0\right]}{E_0\left[m(T-A) \mid A = a, E = 1\right]} ,$$

the acceleration factor $\phi_{AFT}(a)$ contrasts individuals exposed at age a to those unexposed at age 0 rather than to those who have survived at least to age a, as in the R-AFT model. In the standard AFT model, the age-at-exposure parameters are not interpretable due to the inherent negative association between age and residual survival. In contrast, our formulation of the R-AFT model *explicitly* describes the association between a point-exposure occurring at a random time and residual survival.

We note that the R-AFT model can be viewed as an extension of the time-dependent AFT model — a modification of the standard AFT that allows the inclusion of time-varying exposures — wherein the age-at-exposure is of scientific interest as an effect modifier of the exposure (Cox and Oakes, 1984). The resulting causal interpretation is similar to the one

derived from the structural nested failure time model proposed by Robins (1992), itself an extension of the time-dependent AFT model. The primary difference between the R-AFT model and the time-dependent AFT model framing of the problem is that the latter directly equates the counterfactual survival times across exposure levels via a mapping $h_\gamma(Z, X, A, T)$ for some acceleration function indexed by γ. The structural nested failure time model imposes a time scaling on individual counterfactual survival times, which might be an overly restrictive structure. However, imposing an acceleration model for the unintervened survival time distribution might be more natural to practitioners as it relates two observable populations; specifically, the R-AFT model relates the observable residual survival time distributions. The R-AFT model specification is specifically devised to assist practitioners in quantifying how differences in residual survival between populations is modulated by age-at-exposure. The model formulation allows for flexible specification of the acceleration factor $\phi(z, x, a)$ to describe effect-modifiers of a point-exposure. While the acceleration factor $\phi(z, x, a)$ may not have a causal interpretation when there are unmeasured confounders, it can nevertheless be used to describe meaningful adjusted relationships in settings in which age-at-exposure is a possible effect modifier of survival.

3.3 Estimation and inference

3.3.1 Identification and estimation

In Section 3.2.2, we highlighted that the R-AFT model is indexed by the acceleration factor ϕ and the lifetime survival function S among the unexposed. We suppose that the true acceleration factor ϕ_0 lies in a d–dimensional subspace $\{\phi_\beta : \beta \in B \subseteq \mathbb{R}^d\}$ so that $\phi_0 = \phi_{\beta_0}$ for some $\beta_0 \in B$. We further suppose that the unexposed survival function S_0 lies in a p–dimensional subspace $\{S_\eta : \eta \in H \subseteq \mathbb{R}^p\}$ so that $S_0 = S_{\eta_0}$ for some $\eta_0 \in H$. We collate these index parameters as $\theta_0 := (\beta_0, \eta_0)$ and refer to a generic candidate value for θ_0 as θ. We denote the remaining portions of the data-generating mechanism that do not contribute to the likelihood function as η_0, and note that η_0 is orthogonal to θ_0. Letting η_0 be unrestricted,

we denote by $\mathcal{P}$ the resulting model for P_0.

We suppose the data consist of n independent draws $O_1, O_2, \ldots, O_n$ from P_0. For any distribution that lies within the R-AFT model $P \in \mathcal{P}$, when conditions (A1) – (A3) hold, the contribution to the relevant portion of the log-likelihood function for an individual with observation $o := (y, \delta, w, a, x, e, z)$ is

$$\ell(\theta; o) := \delta \log \lambda_\eta(u(o;\beta) \,|\, z)\} + \log \left\{ \frac{S_\eta(u(o;\beta) \,|\, z)}{S_\eta(v(o;\beta) \,|\, z)} \right\},$$

where $u(o;\beta) = (y-a)\phi_\beta(a,x,z)^e + a$ and $v(o;\beta) = (w-a)\phi_\beta(a,x,z)^e + a$ denote the contracted lifetime and truncation time, respectively. The remaining contribution to the likelihood pertain to the censoring, truncation and covariate distribution, and without additional knowledge (e.g., about the truncation or covariate distribution), does not contain information about θ_0.

We denote by $\dot{\ell}_\theta(o) := \frac{\partial}{\partial\theta}\ell(\theta; o)$ the score function for the parameter θ, and introduce the following regularity conditions:

(A4) for some σ–finite measure μ, P_θ has finite μ–density p_θ satisfying the M-estimator regularity conditions (Van der Vaart, 2000);

(A5) the score function class $\{\dot{\ell}_\theta : \theta \in \Theta\}$ is P_0–Donsker;

(A6) for each $\epsilon > 0$, $\inf\{|E_0[\dot{\ell}_\theta(O)]| : \theta \in \Theta \text{ with } |\theta - \theta_0| \geq \epsilon\} > 0$.

The conditions ensure that model is sufficiently well-behaved such that true parameter θ_0 is identified and estimable from the observed data.

Lemma 1. *Under conditions (A1)–(A6), the maximum likelihood estimator θ_n, which solves the score equation $\frac{1}{n}\sum_{i=1}^n \dot{\ell}_{\theta_n}(O_i) = 0$, is a consistent estimator of θ_0.*

The estimator θ_n can be characterized as an estimating equations-based estimator (or Z-estimator) (Qin and Lawless, 1994; Vaart and Wellner, 1996) and converges to the true parameter θ_0 in probability. However, to perform valid inference, distributional results are also needed. These are provided in the next theorem.

3.3.2 Statistical inference

We suppose that the information matrix $I(\theta_0) := E_0\left[\dot{\ell}_{\theta_0}(O)^\top \dot{\ell}_{\theta_0}(O)\right]$ is invertible.

Theorem 12. *Under conditions (A1)–(A6), if $P_0 \in \mathcal{P}$ and $\theta \mapsto I(\theta)^{-1}$ is continuous at $\theta = \theta_0$, then θ_n is a regular and asymptotically linear estimator that satisfies*

$$\sqrt{n}\,(\theta_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I(\theta_0)^{-1} \dot{\ell}_{\theta_0}(O_i) + o_p(1) \ .$$

In particular, Theorem 12 implies that $\sqrt{n}\,(\theta_n - \theta_0)$ converges in distribution to a mean-zero normal random variable with variance-covariance matrix

$$\Sigma_0 := I(\theta_0)^{-1} E_0\left[\dot{\ell}_{\theta_0}(O)^\top \dot{\ell}_{\theta_0}(O)\right] I(\theta_0)^{-1} \ .$$

The sufficient conditions for asymptotic linearity of the estimator θ_n detailed in Theorem 12 cover a number of parametric models for the acceleration factor and unexposed survival distribution. As an alternative, we propose a simple checklist for practitioners to verify if a chosen parametric model can be used when building a R-AFT model. The following conditions are sufficient for conditions (A4)–(A6) to hold:

1. denoting by $\{f_\eta(\cdot \mid z)\}$ the class of density functions defining the parametric model for T given $E = 0$ and $Z = z$, and defining $I(\eta) := -E_0\left[\frac{\partial^2}{\partial \eta^2} \log f_\eta(T \mid Z)\right]$, it holds that $I(\eta_0)$ is finite and $I(\eta)$ is invertible in a neighborhood of η_0;
2. in a compact set $B_0 \subseteq B$, for P_0–almost surely $(Z, X, A) = (z, x, a)$, $\beta \mapsto \phi_\beta(z, x, a)$ is Lipschitz differentiable over B_0.

Verifying the two conditions is straightforward in most applications. The Weibull and other common parametric families satisfy Condition 1, with detailed results for the Weibull distribution found in Example 5.4.3 in Van der Vaart (2000). Condition 2 is satisfied for most functions used, including log linear models with interactions.

For ease in implementation, we consider a model-based estimator of the variance-covariance matrix Σ_0 given by

$$\Sigma_{n,model} := \left[\frac{1}{n}\sum_{i=1}^{n} \dot{\ell}_{\theta_n}(O_i)^\top \dot{\ell}_{\theta_n}(O_i)\right]^{-1}.$$

This estimator can be obtained using existing software. However, the result of Theorem 12 also suggest the model-robust 'sandwich' variance estimator

$$\Sigma^2_{n,robust} = I_n(\theta_n)^{-1}\Sigma^{-1}_{n,model} I_n(\theta_n)^{-1}$$

with $I_n(\theta_n) := -\frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{\theta_n}(O_i)$. When the model $\mathcal{P}$ is correctly specified, both variance estimators are consistent although the model-based estimator is expected to be more efficient. When conditions (A1)–(A3) hold but the R-AFT model does not hold, then θ_n converges to θ_0^*, the closest $\theta \in \Theta$ to θ_0 in a Kullback-Leibler sense (White, 1982) and then only the model-robust variance estimator is generally consistent.

3.4 Numerical studies

In this section, we present empirical results on the performance of the maximum likelihood estimator of R-AFT model parameters in finite sample settings. We make use of the `Flexsurv` (version 2.2) package in `R` (version 3.5.1), which evaluates the likelihood and score functions for a collection of parametric distributions. To evaluate the maximum likelihood, we used the Nelder-Mead algorithm within the `optim` function. We construct model-based 95% Wald-type confidence intervals using the model-based standard errors $\sigma^2_{n,model}$ as described in Section 3.3.2. We consider a data-generating mechanism according to the following sequential pecification:

1. baseline covariates: generate $Z_1 = z_1 \sim$ Bernoulli(0.5) and $Z_2 = z_2 \sim$ Uniform(0,1), independent of each other;
2. exposure indicator: generate $E = e \sim$ Bernoulli(0.5);

3. age-at-exposure: if $e = 1$, generate $A = a \sim \min[\text{Weibull}(2.5,\ 45), 100]$;
4. dose: if $e = 1$, generate $X = x$ as $I(0.001 < \tilde{X} < 4)\tilde{X} + 4I(\tilde{X} \geq 4)$ with $\tilde{X}$ a lognormal(-4, 2.3) random variable;
5. unexposed survival time: if $e = 0$, generate $T = t \sim$ Weibull(5, 70);
6. exposed survival time: if $e = 1$, generate $T = t \sim a+U$ with U following the conditional distribution of $\tilde{U} - a$ given $\tilde{U} \geq a$ for $\tilde{U} \sim$ Weibull(5, $70/\phi(z, x, a)$);
7. truncation time: generate $W = w \sim a + 30\times$Beta(3, 10);
8. censoring time: generate $C = c \sim$ w+Beta(20, 10)$\times\beta_C$.

The collection of baseline covariates represent potential confounders or effect modifiers in the Life Span Study data. The covariate Z_1 is analogous to sex, which modifies the acceleration factor. The variable Z_2 is distributed according to a Uniform(0,1) and represents a proxy variable for socioeconomic status. Age-at-exposure A is chosen to have a distribution centered near 25 so that it is comparable to the observed data on age-at-exposure. Dose X is generated according to a log-normal with mean and standard deviation values matching those observed in the Life Span Study dataset.

We generated 1000 datasets of size $n \in \{250, 1000, 10000\}$ and examined performance for two choices of acceleration factors $\log \phi_\beta(z, x, a)$, namely

1. Model 1: $\log \phi_{\beta,1}(z, x, a) = \beta_0 + \beta_{age}A + \beta_{dose,1}X + \beta_{Z1}Z_1 + \beta_{Z2}Z_2$
2. Model 2: $\log \phi_{\beta,2}(z, x, a) = \beta_0 + \beta_{age}A + \beta_{dose,2}X + \beta_{Z1}Z_1 + \beta_{Z2}Z_2 + \beta_{dose:Z1}XZ_1$,

which differ in whether or not they include an interaction term. In addition, we vary the censoring distribution coefficient $\beta_C \in \{70, 80, 90\}$, which results in different censoring levels. For brevity, we present only the performance of the MLE β_n of the dose main effect. The results presented in Table 3.1 and Figure 3.1 demonstrate that the method results are unbiased and have appropriate coverage of the acceleration factor coefficients in the simulation scenarios considered.

Size (n)	$\Delta = 0$	Mean $\beta_{dose,1}$	Coverage $\beta_{dose,1}$	Mean $\beta_{dose,2}$	Coverage $\beta_{dose,2}$
250	27%	0.520	94.3 (%)	0.539	93.7
	17%	0.517	94.1 (%)	0.544	93.6
	10%	0.516	94.4 (%)	0.544	95.2
1000	27%	0.502	94.4 (%)	0.501	97.6
	17%	0.503	94.4 (%)	0.509	94.3
	10%	0.502	94.5 (%)	0.509	94.2
10000	27%	0.499	95.7 (%)	0.501	95.1
	17%	0.499	96.9 (%)	0.500	95.9
	10%	0.499	95.9 (%)	0.501	95.2

Table 3.1: Performance of parametric estimation strategy in simulated data setting, true $\beta_{dose} = 0.5$ in both settings. Performance is evaluated over 1000 replications. Column one indicates the percentage of missing outcomes due to censoring, columns two and four show mean of the estimator, columns three and five show empirical coverage.

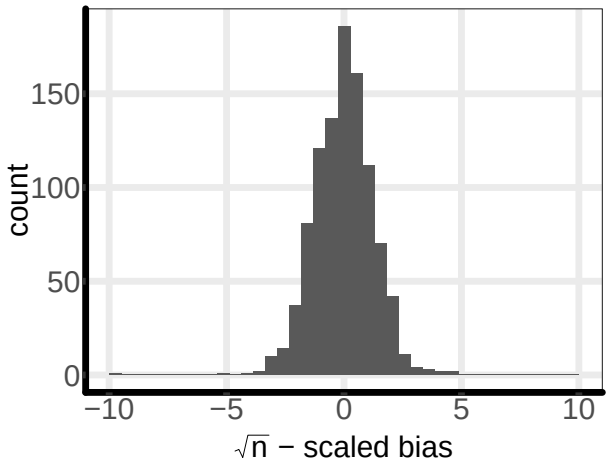

Figure 3.1: Empirical distribution of β_{dose} coefficient estimates over 1000 replications

3.5 Results from the Life Span Study

The Life Span Study is a prospective cohort study of Japanese survivors of the 1945 atomic bombings of Hiroshima and Nagasaki (Ozasa et al., 2019). The study cohort includes 93,741 survivors who were located within 10 kilometers of the hypocenter in either city at the time of the bombings. Participants were recruited through a supplementary survey to the 1950 National Census of Japan; their follow-up began on 1 October 1950. For each participant, follow-up continued until the date of death obtained from the Japanese National Family Registry system, which is virtually complete. The primary outcome for our analysis was death due to any cause. Participants were considered to be censored when they emigrated from Japan, when they attained 110 years of age, or at the end of the study period on 31 December 2009. DS02R1 radiation doses were estimated from the DS02 dosimetry system based on a participant's age-at-exposure; the city in which they were exposed; the ground distance between the bomb's hypocenter and their reported location; the amount and type of shielding between the participant and the blast; and the orientation of the participant relative to the direction of the blast (Cullings et al., 2006, 2017). Weighted absorbed dose to the colon (in gray) measured the absorbed energy per unit mass of tissue calculated using the DS02R1 radiation doses. Participants with an unknown dose (n=7,021) — primarily due to complex shielding situations not accommodated by the dosimetry system — were excluded, resulting in a sample size of 86,559. We defined the 'unexposed' (or reference) group as survivors at distal locations (i.e., more than three kilometers from the hypocenter) (Cologne and Preston, 2001); estimated radiation doses are set to zero for these participants. Relevant covariates included city (Hiroshima or Nagasaki), government-reported sex (male or female), and age-at-exposure (in years).

In line with the data structure presented in Section 3.2.2, the data were prospectively collected from the target population based on a national census. Due to the length of time of collection, we account for left truncation from the time at which the census was taken. Cohort follow-up begins 5 years after the time-of-exposure, which restricts the study to measuring

the long-term survival outcomes due to radiation exposure excluding acute effects. As such, we modify the definition of survival to be conditional on the subset of individuals who were alive prior to 1945 and after 1950.

Coefficient	$\widehat{\beta}^m$ (s.e.)	$\widehat{\beta}^f$ (s.e.)
β_0	-0.11 (0.006)	-0.13 (0.009)
β_1	0.18 (0.04)	-0.07 (0.06)
β_2	0.56 (0.12)	1.23 (0.18)
β_3	-0.05 (0.14)	-0.21 (0.19)
β_4	-0.01 (0.018)	0.08 (0.018)
β_5	0.056 (0.02)	-0.006 (0.02)
β_6	-0.012 (0.006)	-0.0002 (0.007)

Table 3.2: Estimates of model coefficients for RERF data stratified by sex m :=male, f :=female. The age variables $(\beta_1, \beta_2, \beta_3)$ are on the scale of (years/100), while dose variables $(\beta_4, \beta_5, \beta_6)$ are in gray. s.e. denotes the standard error estimates.

We note that in the Life Span Study cohort government-reported sex $s \in \{m, f\}$ is an established effect modifier of radiation dose and modulates survival in the reference population (Ozasa et al., 2012). A Weibull distribution was assumed for the survival times among the unexposed population. The association of radiation dose with mortality is known to depend on age-at-exposure and might not be linear (Ozasa et al., 2012). Therefore, we fit a simple yet flexible log-linear model for the acceleration factor, which included a cubic specification for age-at-exposure a and for radiation dose x:

$$\log \phi^s_\beta(a, x) = \beta^s_0 + \beta^s_1 a + \beta^s_2 a^2 + \beta^s_3 a^3 + \beta^s_4 x + \beta^s_5 x^2 + \beta^s_6 x^3$$

We provide parameter estimates and standard error estimates in Table 3.2 and summarize the findings by quantifying residual survival within specific populations. We interpret the results using both the relative and absolute differences in estimated median residual survival

as a function of radiation dose for individuals exposed at the same fixed age.

At 0 gray, the median relative residual survival decreases as age-at-exposure increases (Figure 3.2). In particular, male survivors exposed to a dose of 0 gray (i.e., in proximal locations) at age 30 years had a 1% reduction in median residual survival compared to unexposed survivors (i.e., in distal locations) who were alive at age 30 years. Among all age-at-exposure groups, relative median residual survival decreased as dose increased. For example, men exposed to 2 gray at age 30 years had a 11% reduction in median residual survival compared to men exposed to 0 gray at age 30 years. Compared to unexposed survivors who were alive at age 30 years, the reduction in median residual survival was also roughly 11% (i.e. 1 - 0.89 $\times$ 0.99).

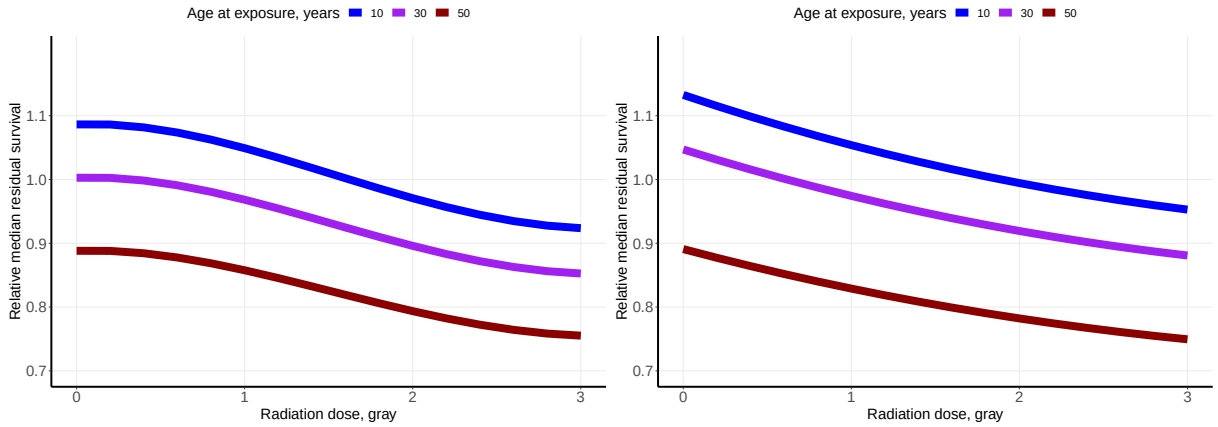

Figure 3.2: Estimated relative median residual survival by varying radiation exposure doses among men (left) and women (right) when compared to an unexposed individual of the same sex.

Alternatively, we make use of the parametric model structure to summarize the median number of residual years lost due to radiation exposure (Figure 3.3). Among all age-at-exposure groups, the median difference in residual years worsened as radiation dose increased. For example, men exposed to 2 Gy at age 30 years lost 4.8 median residual years compared to men exposed to 0 Gy at age 30 years. The results shown in Figure 3.3 are consistent with earlier analyses of the Life Span Study data (Ozasa et al., 2012), which reported that higher

radiation dose is associated with greater decrements in survival, survivors exposed at younger ages are particularly sensitive to radiation, and women have a greater radiation-associated loss of life compared to men.

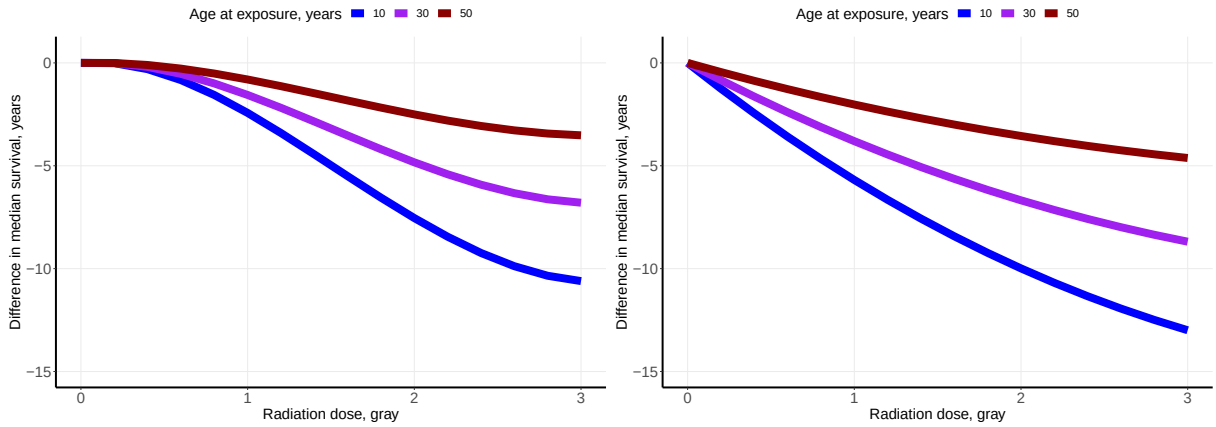

Figure 3.3: Estimated difference in median residual survival by varying radiation exposure doses among men (left) and women (right) when compared to a 0 dose exposed individual of the same sex.

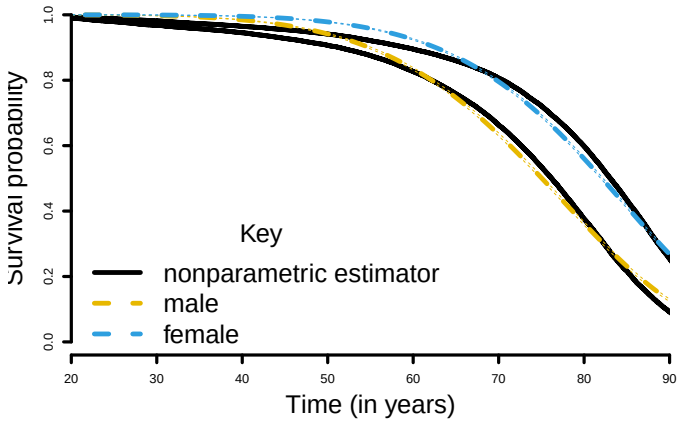

Figure 3.4: Estimated survival curves of unexposed population of men and women under Weibull distribution compared to Kaplan-Meier.

A comparison between the parametric estimated survival distribution and the nonparametric Kaplan-Meier among the nonexposed group is presented in Figure 3.4. This figure suggests that the Weibull model provides a decent fit to the observed data, though an improved fit may be possible with a more complex parametric model for the unexposed survival time.

3.6 Concluding remarks

We have introduced the residual accelerated failure time model as a statistical tool to quantify the association between a point-exposure and residual life expectancy. It is especially useful in settings in which the exposure occurs at a random time and has an effect possibly modulated by dose and age-at-exposure. In such cases, the acceleration factor quantifies the relative mean (or median) residual survival as a function of dose for a given age-at-exposure and covariate level compared to an unexposed group with the same covariate profile that survived to the same age. The advantage of the R-AFT model is that the coefficients for age-at-exposure are interpretable as effect modifiers of exposure — such interpretation does not readily follow from use of either a traditional AFT model or Cox model.

We propose a parametric modeling strategy in this chapter and establish the asymptotic linearity of the maximum likelihood estimator of model parameters. This naturally gives rise to both model-based and model-robust approaches to inference. We use the developed model to study the cohort of participants in the Life Span Study, namely examining the relationship between radiation dose exposure and residual survival. This application illustrates well the need for the proposed model since it involves a point-exposure occurring at a random time. It is of interest to develop inferential methods for use in semiparametric formulations of the proposed model, which would make fewer modeling assumptions without jeopardizing the interpretability of model coefficients. However, such extension requires significant analytic developments and is left as future work.

BIBLIOGRAPHY

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004). Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime data analysis*, 10:335–350. 53

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455. 49

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173. 51

Benkeser, D., Carone, M., Laan, M. V. D., and Gilbert, P. (2017). Doubly robust nonparametric inference on the average treatment effect. *Biometrika*, 104(4):863–880. 39

Beran, R. (1981). Nonparametric regression with randomly censored survival data. 148

Braak, H. and Braak, E. (1991). Demonstration of amyloid deposits and neurofibrillary changes in whole brain sections. *Brain pathology*, 1(3):213–216. 72

Brier, G. W. et al. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3. 5, 25

Carone, M., Dominici, F., and Sheppard, L. (2020). In pursuit of evidence in air pollution epidemiology: the role of causally driven data science. *Epidemiology (Cambridge, Mass.)*, 31(1):1. 77

Chaieb, L. L., Rivest, L.-P., and Abdous, B. (2006). Estimating survival under a dependent truncation. *Biometrika*, 93(3):655–669. 24

Chan, K. C. G. and Wang, M.-C. (2012). Estimating incident population distribution from prevalent data. *Biometrics*, 68(2):521–531. 63

Chang, C.-C. H., Zhao, Y., Lee, C.-W., and Ganguli, M. (2012). Smoking, death, and alzheimer's disease: a case of competing risks. *Alzheimer disease and associated disorders*, 26(4):300. 41

Chang, R. C.-C., Ho, Y.-S., Wong, S., Gentleman, S. M., and Ng, H.-K. (2014). Neuropathology of cigarette smoking. *Acta neuropathologica*, 127(1):53–69. 41

Cheng, Y.-J. and Wang, M.-C. (2012). Estimating propensity scores and causal survival functions using prevalent survival data. *Biometrics*, 68(3):707–716. 63

Chernofsky, A., Bosch, R. J., and Lok, J. J. (2021). Causal mediation analysis with mediator values below an assay limit. *arXiv preprint arXiv:2107.14782*. 51

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. 15

Cologne, J. B. and Preston, D. L. (2001). Impact of comparison group on cohort dose response regression: an example using risk estimation in atomic-bomb survivors. *Health physics*, 80(5):491–496. 95

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202. 80

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall/CRC. 81, 88

Cullings, H. M., Fujita, S., Funamoto, S., Grant, E. J., Kerr, G. D., and Preston, D. L. (2006). Dose estimation for atomic bomb survivor studies: its evolution and present status. *Radiation research*, 166(1):219–254. 95

Cullings, H. M., Grant, E. J., Egbert, S. D., Watanabe, T., Oda, T., Nakamura, F., Yamashita, T., Fuchi, H., Funamoto, S., Marumo, K., et al. (2017). Ds02r1: improvements to atomic bomb survivors' input data and implementation of dosimetry system 2002 (ds02) and resulting changes in estimated doses. *Health physics*, 112(1):56–97. 95

Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scandinavian Journal of Statistics*, pages 181–197. 148

Díaz, I. (2017). Efficient estimation of quantiles in missing data models. *Journal of Statistical Planning and Inference*, 190:39–51. 6, 27, 28

Díaz, I., Colantuoni, E., Hanley, D. F., and Rosenblum, M. (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime data analysis*, 25(3):439–468. 53

Emsley, R., Dunn, G., and White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical methods in medical research*, 19(3):237–270. 51

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1645–1666. 51

Gaugler, J., James, B., Johnson, T., Reimer, J., Solis, M., Weuve, J., Buckley, R. F., and Hohman, T. J. (2022). 2022 alzheimer's disease facts and figures. *Alzheimers & Dementia*, 18(4):700–789. 72

Gerds, T. A., Beyersmann, J., Starkopf, L., Frank, S., Laan, M. J., and Schumacher, M. (2017). The kaplan-meier integral in the presence of covariates: A review. *From Statistics to Mathematical Finance*, pages 25–41. 13

Gerds, T. A. and Schumacher, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040. 5, 25

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *The annals of statistics*, pages 1501–1555. 9, 54, 111, 115, 120, 125, 149, 156, 160, 161, 163

Greenland, S., Pearl, J., and Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, pages 37–48. 46

Haneuse, S., Schildcrout, J., Crane, P., Sonnen, J., Breitner, J., and Larson, E. (2009). Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology*, 32(3):229–239. 42, 46, 51, 74

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*, 21(1):13. 81, 88

Hernán, M. A., Alonso, A., and Logroscino, G. (2008). Commentary: cigarette smoking and dementia: potential selection bias in the elderly. *Epidemiology*, pages 448–450. 41

Ibragimov, I. A. and Has' Minskii, R. Z. (1981). *Statistical estimation: asymptotic theory*, volume 2. Springer Science & Business Media. 15

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309. 51

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481. 24

Kukull, W. A., Higdon, R., Bowen, J. D., McCormick, W. C., Teri, L., Schellenberg, G. D., Van Belle, G., Jolley, L., and Larson, E. B. (2002). Dementia and alzheimer disease incidence: a prospective cohort study. *Archives of neurology*, 59(11):1737–1746. 72

Lok, J., Gill, R., Van Der Vaart, A., and Robins, J. (2004). Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica*, 58(3):271–295. 81, 87

Mackenzie, T. (2012). Survival curve estimation with dependent left truncated data using cox's model. *The international journal of biostatistics*, 8(1). 24

Mirra, S. S., Heyman, A., McKeel, D., Sumi, S., Crain, B. J., Brownlee, L., Vogel, F., Hughes, J., Van Belle, G., Berg, L., et al. (1991). The consortium to establish a registry for alzheimer's disease (cerad): Part ii. standardization of the neuropathologic assessment of alzheimer's disease. *Neurology*, 41(4):479–479. 72

Moore, K. L. and van der Laan, M. J. (2009). Increasing power in randomized trials with right censored outcomes through covariate adjustment. *Journal of biopharmaceutical statistics*, 19(6):1099–1131. 24

Murray, S. and Tsiatis, A. A. (1996). Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics*, pages 137–151. 24

Neyman, J. S. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9.(translated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465-480). *Annals of Agricultural Sciences*, 10:1–51. 43, 86

Ozasa, K., Cullings, H. M., Ohishi, W., Hida, A., and Grant, E. J. (2019). Epidemiological studies of atomic bomb radiation at the radiation effects research foundation. *International journal of radiation biology*, 95(7):879–891. 80, 83, 95

Ozasa, K., Shimizu, Y., Suyama, A., Kasagi, F., Soda, M., Grant, E. J., Sakata, R., Sugiyama, H., and Kodama, K. (2012). Studies of the mortality of atomic bomb survivors, report 14, 1950–2003: an overview of cancer and noncancer diseases. *Radiation research*, 177(3):229–243. 80, 96, 97

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological methods*, 19(4):459. 49

Petersen, M. L., Sinisi, S. E., and van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology*, pages 276–284. 50

Pfanzagl, J. (1982). Lecture notes in statistics. *Contributions to a general asymptotic statistical theory*, 13. 11, 15

Qian, J. and Betensky, R. A. (2014). Assumptions regarding right censoring in the presence of left truncation. *Statistics & probability letters*, 87:12–17. 82

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *the Annals of Statistics*, 22(1):300–325. 90

Reid, N. (1981a). Estimating the median survival time. *Biometrika*, 68(3):601–608. 27, 28

Reid, N. (1981b). Influence function for censored data. *The Annals of Statistics*, 9(1):78–92. 11, 148

Rennert, L. and Xie, S. X. (2019). Bias induced by ignoring double truncation inherent in autopsy-confirmed survival studies of neurodegenerative diseases. *Statistics in medicine*, 38(19):3599–3613. 42, 51, 61

Richardson, T. S. and Robins, J. M. (2013). Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013. 49

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512. 148

Robins, J. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 79(2):321–334. 81, 89

Robins, J. M. (1998). Structural nested failure time models. *Encyclopedia of Biostatistics*, 6:4372–4389. 88

Robins, J. M. et al. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In *Proceedings of the biopharmaceutical section, American statistical association*, volume 24, page 3. San Francisco CA. 24

Rothman, K. J. and Greenland, S. (2005). Causation and causal inference in epidemiology. *American journal of public health*, 95(S1):S144–S150. 42

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688. 43, 86

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, 31(2):161–170. 46

Schultzberg, M. (2016). Causal effects in mediation analysiswith limited-dependent variables. 51

Selkoe, D. J. (1991). The molecular pathology of alzheimer's disease. *Neuron*, 6(4):487–498. 72

Serrano-Pozo, A., Frosch, M. P., Masliah, E., and Hyman, B. T. (2011). Neuropathological alterations in alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 1(1):a006189. 72

Shaffer, R. M., Blanco, M. N., Li, G., Adar, S. D., Carone, M., Szpiro, A. A., Kaufman, J. D., Larson, T. V., Larson, E. B., Crane, P. K., et al. (2021a). Fine particulate matter

and dementia incidence in the adult changes in thought study. *Environmental health perspectives*, 129(8):087001. 74

Shaffer, R. M., Li, G., Adar, S. D., Dirk Keene, C., Latimer, C. S., Crane, P. K., Larson, E. B., Kaufman, J. D., Carone, M., and Sheppard, L. (2021b). Fine particulate matter and markers of alzheimer's disease neuropathology at autopsy in a community-based cohort. *Journal of Alzheimer's Disease*, 79(4):1761–1773. 74

Shepherd, D. A. and Moreno-Betancur, M. (2022). Confounding-adjustment methods for the difference in medians. *arXiv preprint arXiv:2207.05940*. 27, 28

Small, D. S. (2011). Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables. *arXiv preprint arXiv:1109.1070*. 51

Sonnen, J. A., Larson, E. B., Crane, P. K., Haneuse, S., Li, G., Schellenberg, G. D., Craft, S., Leverenz, J. B., and Montine, T. J. (2007). Pathological correlates of dementia in a longitudinal, population-based sample of aging. *Annals of neurology*, 62(4):406–413. 73

Stute, W. (1994). The bias of kaplan-meier integrals. *Scandinavian Journal of Statistics*, pages 475–484. 11

Tchetgen, E. J. T. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*, 40(3):1816. 48, 50, 59

Tilvis, R. S., Kähönen-Väre, M. H., Jolkkonen, J., Valvanne, J., Pitkala, K. H., and Strandberg, T. E. (2004). Predictors of cognitive decline and mortality of aged people over a 10-year period. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(3):M268–M274. 72

Tsai, W.-Y., Jewell, N. P., and Wang, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4):883–886. 24, 52

Tsuang, D., Simpson, K. L., Li, G., Barnhart, R. L., Edland, S. D., Bowen, J., McCormick, W., Teri, L., Nochlin, D., Larson, E. B., et al. (2005). Evaluation of selection bias in an incident-based dementia autopsy case series. *Alzheimer disease and associated disorders*, 19(2):67. 74

Tyas, S. L., White, L. R., Petrovitch, H., Ross, G. W., Foley, D. J., Heimovitz, H. K., and Launer, L. J. (2003). Mid-life smoking and late-life dementia: the honolulu-asia aging study. *Neurobiology of aging*, 24(4):589–596. 41

Ulrich, J., Johannson-Locher, G., Seiler, W. O., and Stähelin, H. B. (1997). Does smoking protect from alzheimer's disease? alzheimer-type changes in 301 unselected brains from patients with known smoking history. *Acta neuropathologica*, 94(5):450–454. 41

Vaart, A. W. and Wellner, J. A. (1996). Weak convergence. In *Weak convergence and empirical processes*, pages 16–28. Springer. 90

Vakulenko-Lagun, B., Qian, J., Chiou, S. H., Wang, N., and Betensky, R. A. (2022). Nonparametric estimation of the survival distribution under covariate-induced dependent truncation. *Biometrics*, 78(4):1390–1401. 24

Van der Laan, M. J. and Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*, volume 5. Springer. 15

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press. 22, 90, 91, 137, 138, 159, 171, 174

VanderWeele, T. (2015a). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press. 46, 50, 148

VanderWeele, T. (2015b). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press. 49

VanderWeele, T. J. (2013). Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *European journal of epidemiology*, 28(2):113–117. 51

Vansteelandt, S., Dukes, O., Van Lancker, K., and Martinussen, T. (2022). Assumption-lean cox regression. *Journal of the American Statistical Association*, pages 1–10. 6, 28, 142

Vansteelandt, S. and Joffe, M. (2014). Structural nested models and g-estimation: the partially realized promise. *Statistical Science*, 29(4):707–731. 87

Vock, D. M., Tsiatis, A. A., Davidian, M., Laber, E. B., Tsuang, W. M., Finlen Copeland, C. A., and Palmer, S. M. (2013). Assessing the causal effect of organ transplantation on the distribution of residual lifetime. *Biometrics*, 69(4):820–829. 87

Wang, J., Ning, J., and Shete, S. (2019). Mediation analysis in a case-control study when the mediator is a censored variable. *Statistics in medicine*, 38(7):1213–1229. 51

Wang, J. and Shete, S. (2018). Estimation of indirect effect when the mediator is a censored variable. *Statistical methods in medical research*, 27(10):3010–3025. 51

Wang, L. and Zhang, Z. (2011). Estimating and testing mediation effects with censored data. *Structural Equation Modeling*, 18(1):18–34. 51

Wang, Y., Ying, A., and Xu, R. (2022). Doubly robust estimation under covariate-induced dependent left truncation. *arXiv preprint arXiv:2208.06836*. 3, 13, 24

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. 80

Westling, T., Luedtke, A., Gilbert, P. B., and Carone, M. (2023). Inference for treatment-specific survival curves using machine learning. *Journal of the American Statistical Association*, (just-accepted):1–26. 3, 4, 5, 13, 21, 26, 27

Westling, T., van der Laan, M. J., and Carone, M. (2020). Correcting an estimator of a multivariate monotone function with isotonic regression. *Electronic journal of statistics*, 14(2):3032. 25

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the econometric society*, pages 1–25. 92

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, 13(1):163–177. 62

Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., and Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in medicine*, 39(8):1199–1236. 87

Zeng, D. (2004). Estimating marginal survival function by adjusting for dependent censoring using many covariates. *Annals of Statistics*, 32(4):1533–1555. 24

Zhao, X., Lynch Jr, J. G., and Chen, Q. (2010). Reconsidering baron and kenny: Myths and truths about mediation analysis. *Journal of consumer research*, 37(2):197–206. 51

Zheng, W. and Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted Learning*, pages 459–474. Springer. 15, 59

Zheng, W. and van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *The international journal of biostatistics*, 8(1):1–40. 50

Appendix A

CHAPTER ONE APPENDIX

Appendix

A.1 Part A: identification

Before establishing the identification of θ_0, we show that components of the ideal data distribution $P_{X,0}$ can be identified in terms of the observed data distribution P_0.

Under conditions (A1)–(A2), in view of Theorem 11 of Gill and Johansen (1990), for any $t \in (0, \tau_2(z))$, we can write

$$1 - F_{X,0}(t \,|\, a, z) \;=\; S_{X,0}(t \,|\, a, z) \;=\; \prod_{u \in [0,t)} \{1 - \Lambda_{X,0}(du \,|\, a, z)\} \tag{A.1}$$

for $\Lambda_{X,0}(t \,|\, a, z) := \int_0^t \frac{F_{X,0}(du \,|\, a,z)}{S_{X,0}(u \,|\, a,z)}$. We first express the observed follow-up time subdistribution function in terms of $P_{X,0}$. Defining $F_0(t \,|\, w, a, z) := P_{X,0}\,(T \le t \,|\, W = w, A = a, Z = z, T \ge W)$ and $G_0(w \,|\, a, z) := P_{X,0}\,(W \le w \,|\, A = a, Z = z, T \ge W)$, we note that

$$F_0(dt \,|\, w, a, z) G_0(dw \,|\, a, z) = \frac{I_{[0,t]}(w) F_{X,0}(dt \,|\, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(u \,|\, a, z) F_{X,0}(du \,|\, a, z)}$$

under conditions (B1)–(B2). So, we have that

$$\begin{aligned} F_{1,0}(u \,|\, a, z) \;&=\; P_0\,(Y \le u, \Delta = 1 \,|\, A = a, Z = z) \\ &=\; P_{X,0}\,(T \le u, T \le C \,|\, A = a, Z = z, T \ge W) \\ &=\; \iint I_{[0,u]}(t) P_{X,0}\,(C \ge t \,|\, T = t, W = w, A = a, Z = z, T \ge W)\, F_0(dt \,|\, w, a, z) G_0(dw \,|\, a, z) \end{aligned}$$

$$
\begin{aligned}
&= \iint I_{[0,u]}(t) Q_0(t \,|\, w, a, z) F_0(dt \,|\, w, a, z) G_0(dw \,|\, a, z) \\
&= \frac{\iint I_{[w,u]}(t) Q_0(t \,|\, w, a, z) F_{X,0}(dt \,|\, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} \\
&= \frac{\int I_{[0,u]}(t) \left\{ \int I_{[0,t]}(w) Q_0(t \,|\, w, a, z) G_{X,0}(dw \,|\, a, z) \right\} F_{X,0}(dt \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} \,,
\end{aligned}
$$

which implies that

$$
F_{1,0}(du \,|\, a, z) = \frac{\int I_{[0,u]}(w) Q_0(u \,|\, w, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} F_{X,0}(du \,|\, a, z) \,.
$$

Next, we can write that

$$
\begin{aligned}
R_0(u \,|\, a, z) &= P_0(W \leq u \leq Y \,|\, A = a, Z = z) \\
&= \int I_{[0,u]}(w) P_{X,0}\left(T \geq u, C \geq u \,|\, W = w, A = a, Z = z, T \geq W\right) G_0(dw \,|\, a, z) \\
&= \iint I_{[w,t]}(u) Q_0(u \,|\, w, a, z) F_0(dt \,|\, w, a, z) G_0(dw \,|\, a, z) \\
&= \frac{\iint I_{[w,t]}(u) Q_0(u \,|\, w, a, z) I_{[0,t]}(w) F_{X,0}(dt \,|\, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(u \,|\, a, z) F_{X,0}(du \,|\, a, z)} \\
&= \frac{\int I_{[0,u]}(w) Q_0(u \,|\, w, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} S_{X,0}(u \,|\, a, z) \,.
\end{aligned}
$$

Thus, under conditions (B1)–(B2), we find that

$$
\tilde{\Lambda}_0(t \,|\, a, z) = \int_0^t \frac{F_{0,1}(du \,|\, a, z)}{R_0(u \,|\, a, z)} = \int_0^t \frac{F_{X,0}(du \,|\, a, z)}{S_{X,0}(u \,|\, a, z)} = \Lambda_{X,0}(t \,|\, a, z) \,,
$$

and so, in view of (A.1), $F_{X,0}(t \,|\, a, z)$ is identified by $\tilde{F}_0(t \,|\, a, z) := 1 - \prod_{u \in (0,t)} \{1 - \tilde{\Lambda}(t \,|\, a, z)\}$. The fact that $F_{X,0}$ is identified directly implies that the target conditional truncation distribution $G_{X,0}$ is itself identified in view of the fact that $G_{X,0}(du \,|\, a, z) \propto_{a,z} S_{X,0}(u \,|\, a, z)^{-1} G_0(du \,|\, a, z)$, which allows us to write that

$$
G_{X,0}(w \,|\, a, z) = \frac{1}{\gamma_0(a, z)} \int_0^w S_{X,0}(u \,|\, a, z)^{-1} G_0(du \,|\, a, z)
$$

with $\gamma_0(a,z) := \int S_{X,0}(u \,|\, a,z)^{-1} G_0(du \,|\, a,z)$ the appropriate normalizing constant. Similarly, identification of $F_{X,0}$ implies identification of the target joint exposure-covariate distribution function $J_{X,0}$ in view of the fact that

$$J_{X,0}(da,dz) \propto_{a,z} \frac{J_0(da,dz)}{\int S_{X,0}(u \,|\, a,z) G_{X,0}(du,a,z)} = \gamma_0(a,z) J_0(da,dz)$$

so that $J_{X,0}(da,dz) = \bar{\gamma}_0(a,z) J_0(da,dz)$ with $\bar{\gamma}_0(a,z) := \gamma_0(a,z) / \iint \gamma_0(a,z) J_0(da,dz)$. Of course, the target marginal covariate distribution $H_{X,0}$ is then itself identified from the identification for $J_{X,0}$ through marginalization. Since θ_0 is a functional of $F_{X,0}$ and $H_{X,0}$, the identification of the latter distributions directly implies that of θ_0.

The observable conditional censoring survival function Q_0 can also be identified using product-integration as in the well-known context of right-censoring without truncation. Defining with some abuse of notation the subdistribution function $F_{0,0}(u \,|\, w,a,z) := P_0(T \leq u, \Delta = 0 \,|\, A = a, Z = z)$ and at-risk probability function $R_0(u \,|\, w,a,z) := P_0(W \leq u \leq Y \,|\, W = w, A = a, Z = z)$, we have that

$$\begin{aligned} F_{0,0}(u \,|\, w,a,z) &= -\int I(c \leq u) P_0(T \geq c \,|\, W = w, A = a, Z = z, C = c)\, Q_0(dc \,|\, w,a,z) \\ &= -\int I(c \leq u) P_0(T \geq c \,|\, W = w, A = a, Z = z)\, Q_0(dc \,|\, w,a,z)\ , \end{aligned}$$

which implies that $F_{0,0}(du \,|\, w,a,z) = -P_0(T \geq u \,|\, W = w, A = a, Z = z)\, Q_0(du \,|\, w,a,z)$, and furthermore, for $u \geq w$, we have that

$$\begin{aligned} R_0(u \,|\, w,a,z) &= P_0(Y \geq u \,|\, W = w, A = a, Z = z) \\ &= P_0(C \geq u \,|\, W = w, A = a, Z = z)\, P_0(T \geq u \,|\, W = w, A = a, Z = z)\,. \end{aligned}$$

Thus, we find that $F_{0,0}(du \,|\, w,a,z) / R_0(u \,|\, w,a,z) = -Q_0(du \,|\, w,a,z) / Q_0(u \,|\, w,a,z)$, which

then implies that $Q_0(c \mid w, a, z)$ can be identified by the product-integral

$$\prod_{c \in [0,u)} \left\{ 1 - \frac{F_{0,0}(du \mid w, a, z)}{R_0(u \mid w, a, z)} \right\}.$$

A.2 Part B: proofs of theorems

To begin, we present certain key identities on the linearization of conditional survival integrals that will be used below.

Lemma 2. *For each $(a, z) \in \{0, 1\} \times \mathcal{Z}$ and any $0 \leq \alpha < \beta < \infty$, the following identities hold:*

$$
\begin{aligned}
&(a)\ \int \varphi(y,z)(\tilde{F}_P - \tilde{F}_0)(dy \,|\, a, z) = -\int L_{P,\varphi}(y,a,z)\frac{\tilde{S}_0(y \,|\, a, z)}{\tilde{S}_P(y \,|\, a, z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, a, z)\,; \\
&(b)\ \int \frac{(\tilde{S}_P - \tilde{S}_0)(w \,|\, a, z)}{\tilde{S}_P(w \,|\, a, z)^2} G_P(dw \,|\, a, z) = -\int \gamma_{P,\natural}(y,a,z)\frac{\tilde{S}_0(y \,|\, a, z)}{\tilde{S}_P(y \,|\, a, z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, a, z) \\
&(c)\ \left|\int_\alpha^\beta \omega(y,z)\left[\frac{L_{P,\varphi}(y,a,z)}{\tilde{S}_P(y \,|\, a, z)}\tilde{\Lambda}_P(dy \,|\, a, z) - \frac{L_{0,\varphi}(y,a,z)}{\tilde{S}_0(y \,|\, a, z)}\tilde{\Lambda}_0(dy \,|\, a, z)\right]\right| \\
&\quad \leq\ 3v_z(\omega,\alpha,\beta)\sup_{y\in[\alpha,\beta]}\left|\frac{L_{P,\varphi}(y,a,z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{L_{0,\varphi}(y,a,z)}{\tilde{S}_0(y \,|\, a, z)}\right| \\
&(d)\ \left|\int_\alpha^\beta \omega(y,z)\left[\frac{\gamma_{P,\natural}(y,a,z)}{\tilde{S}_P(y \,|\, a, z)}\tilde{\Lambda}_P(dy \,|\, a, z) - \frac{\gamma_{0,\natural}(y,a,z)}{\tilde{S}_0(y \,|\, a, z)}\tilde{\Lambda}_0(dy \,|\, a, z)\right]\right| \\
&\quad \leq\ 3v_z(\omega,\alpha,\beta)\sup_{y\in[\alpha,\beta]}\left|\frac{\gamma_{P,\natural}(y,a,z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{\gamma_{0,\natural}(y,a,z)}{\tilde{S}_0(y \,|\, a, z)}\right| \\
&\qquad + \frac{3v_z(\omega,\alpha,\beta)}{S_0(\beta \,|\, a, z)^3}\sup_{y\in[\alpha,\beta]}|G_P(y \,|\, a, z) - G_0(y \,|\, a, z)| \\
&\qquad + \frac{2v_z(\omega,\alpha,\beta)}{S_P(\beta \,|\, a, z)S_0(\beta \,|\, a, z)}\sup_{y\in[\alpha,\beta]}\left|\frac{1}{S_P(y \,|\, a, z)} - \frac{1}{S_0(y \,|\, a, z)}\right|,
\end{aligned}
$$

where $v_z(\omega, \alpha, \beta)$ denotes the maximum of the total variation and supremum norm of the function $y \mapsto \omega(y, z)$ over the interval $[\alpha, \beta]$.

Proof. The Duhamel equation (Theorem 6 of Gill and Johansen, 1990) indicates that

$$
(S - S_0)(y \,|\, a, z) = -S(y \,|\, a, z)\int_0^y \frac{S_0(u \,|\, a, z)}{S(u \,|\, a, z)}\,(\Lambda - \Lambda_0)\,(du \,|\, a, z)
$$

for any two continuous survival functions S and S_0 and their corresponding cumulative

hazard function Λ and Λ_0. The differential form of this equation is

$$(F-F_0)(dy\,|\,a,z) = S_0(y\,|\,a,z)(\Lambda-\Lambda_0)(dy\,|\,a,z) - \int_0^y \frac{S_0(u\,|\,a,z)}{S(u\,|\,a,z)}\,(\Lambda-\Lambda_0)\,(du\,|\,a,z)F(dy\,|\,a,z).$$

The latter equation allows us to write

$$\begin{aligned}
&\int \varphi(y,z)(\tilde{F}_P - \tilde{F}_0)(dy\,|\,a,z) \\
&= \int \varphi(y,z)\tilde{S}_0(y\,|\,a,z)(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy\,|\,a,z) \\
&\quad - \int \varphi(y,z) \int_0^y \frac{\tilde{S}_0(u\,|\,a,z)}{\tilde{S}_P(u\,|\,a,z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(du\,|\,a,z)\tilde{F}_P(dy\,|\,a,z) \\
&= \int \varphi(y,z)\tilde{S}_0(y\,|\,a,z)(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy\,|\,a,z) \\
&\quad - \iint I(u \le y)\varphi(y,z)\frac{\tilde{S}_0(u\,|\,a,z)}{\tilde{S}_P(u\,|\,a,z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(du\,|\,a,z)\tilde{F}_P(dy\,|\,a,z) \\
&= \int \tilde{S}_0(y\,|\,a,z)\left\{\varphi(y,z) - \frac{1}{\tilde{S}_P(y\,|\,a,z)}\int_y^\infty \varphi(u,z)\tilde{F}_P(du\,|\,a,z)\right\}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy\,|\,a,z) \\
&= -\int L_{\varphi,P}(y\,|\,a,z)\frac{\tilde{S}_0(y\,|\,a,z)}{\tilde{S}_P(y\,|\,a,z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy\,|\,a,z)\ ,
\end{aligned}$$

where we have used that

$$\begin{aligned}
\int_y^\infty \varphi(u,z)\tilde{F}_P(du\,|\,a,z) &= -\int_y^\infty \varphi(u,z)\tilde{S}_P(du\,|\,a,z) \\
&= -\varphi(u,z)\tilde{S}_P(u\,|\,a,z)\Big|_y^\infty + \int_y^\infty \tilde{S}_P(u\,|\,a,z)\varphi(du,z) \\
&= \varphi(y,z)\tilde{S}_P(y\,|\,a,z) + L_{\varphi,P}(y,a,z)\ .
\end{aligned}$$

and this establishes part (a). We again make use of the Duhamel equation and write that

$$\begin{aligned}
&\int \frac{(\tilde{S}_P - \tilde{S}_0)(w\,|\,a,z)}{\tilde{S}_P(w\,|\,a,z)^2}G_P(dw\,|\,a,z) \\
&= -\int \frac{1}{\tilde{S}_P(w\,|\,a,z)^2}\tilde{S}_P(w\,|\,a,z)\int_0^w \frac{\tilde{S}_0(u\,|\,a,z)}{\tilde{S}_P(u\,|\,a,z)}(\tilde{\Lambda}_P - \tilde{\Lambda}_0)(du\,|\,a,z)G_P(dw\,|\,a,z)
\end{aligned}$$

$$
\begin{aligned}
&= -\int \left\{ \int_u^\infty \frac{G_P(dw \,|\, a, z)}{\tilde{S}_P(w \,|\, a, z)} \right\} \frac{\tilde{S}_0(u \,|\, a, z)}{\tilde{S}_P(u \,|\, a, z)} (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(du \,|\, a, z) \\
&= -\int \gamma_{P,\natural}(u, a, z) \frac{\tilde{S}_0(u \,|\, a, z)}{\tilde{S}_P(u \,|\, a, z)} (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, a, z) \ ,
\end{aligned}
$$

which establishes part (b). Next, using the fact that

$$
\frac{L_\varphi(u)}{S(u)} \Lambda(du) = \frac{L_\varphi}{S}(du) + \varphi(du)
$$

for any survival function S, corresponding cumulative hazard function Λ and φ–integrated form L, we have that

$$
\begin{aligned}
&\int_\alpha^\beta \omega(y, z) \left[\frac{L_{P,\varphi}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} \tilde{\Lambda}_P(dy \,|\, a, z) - \frac{L_{0,\varphi}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \tilde{\Lambda}_0(dy \,|\, a, z) \right] \\
&= \int_\alpha^\beta \omega(y, z) \left[\frac{L_{P,\varphi}}{\tilde{S}_P}(dy \,|\, a, z) - \frac{L_{0,\varphi}}{\tilde{S}_0}(dy \,|\, a, z) \right] \\
&= \omega(y, z) \left[\frac{L_{P,\varphi}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{L_{0,\varphi}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right] \Bigg|_{y=\alpha}^\beta + \int_\alpha^\beta \left[\frac{L_{P,\varphi}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{L_{0,\varphi}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right] \omega(dy, z) \ .
\end{aligned}
$$

This then implies that

$$
\begin{aligned}
&\left| \int_\alpha^\beta \omega(y, z) \left[\frac{L_{P,\varphi}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} \tilde{\Lambda}_P(dy \,|\, a, z) - \frac{L_{0,\varphi}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \tilde{\Lambda}_0(dy \,|\, a, z) \right] \right| \\
&\qquad \leq \left[2 \sup_{y \in [\alpha, \beta]} |\omega(y, z)| + \int_\alpha^\beta |\omega(dy, z)| \right] \sup_{y \in [\alpha, \beta]} \left| \frac{L_{P,\varphi}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{L_{0,\varphi}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right| .
\end{aligned}
$$

which implies the claimed inequality in (c). Finally, using integration by parts, we write

$$
\begin{aligned}
&\int_\alpha^\beta \omega(y, z) \left[\frac{\gamma_{P,\natural}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} \tilde{\Lambda}_P(dy \,|\, a, z) - \frac{\gamma_{0,\natural}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \tilde{\Lambda}_0(dy \,|\, a, z) \right] \\
&= \omega(y, z) \left[\frac{\gamma_{P,\natural}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{\gamma_{0,\natural}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right] \Bigg|_{y=\alpha}^\beta \\
&\quad + \int_\alpha^\beta \left[\frac{\gamma_{P,\natural}(y, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{\gamma_{0,\natural}(y, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right] \omega(dy, z) + \int_\alpha^\beta \omega(y, z) \left[\frac{\gamma_{P,\natural}(dy, a, z)}{\tilde{S}_P(y \,|\, a, z)} - \frac{\gamma_{0,\natural}(dy, a, z)}{\tilde{S}_0(y \,|\, a, z)} \right]
\end{aligned}
$$

and furthermore expand

$$\int_\alpha^\beta \omega(y,z)\left[\frac{\gamma_{P,\natural}(dy,a,z)}{\tilde{S}_P(y\,|\,a,z)}-\frac{\gamma_{0,\natural}(dy,a,z)}{\tilde{S}_0(y\,|\,a,z)}\right] = \int_\alpha^\beta \omega(y,z)\left[\frac{G_P(dy\,|\,a,z)}{\tilde{S}_P(y\,|\,a,z)^2}-\frac{G_0(dy\,|\,a,z)}{\tilde{S}_0(y\,|\,a,z)^2}\right]$$
$$= \int_\alpha^\beta \omega(y,z)\left[\frac{1}{\tilde{S}_P(y\,|\,a,z)^2}-\frac{1}{\tilde{S}_0(y\,|\,a,z)^2}\right]G_P(dy\,|\,a,z)+\int_\alpha^\beta \frac{\omega(y,z)}{S_0(y\,|\,a,z)^2}(G_P-G_0)(dy\,|\,a,z)$$
$$= \int_\alpha^\beta \omega(y,z)\left[\frac{1}{\tilde{S}_P(y\,|\,a,z)}-\frac{1}{\tilde{S}_0(y\,|\,a,z)}\right]\left[\frac{1}{\tilde{S}_P(y\,|\,a,z)}+\frac{1}{\tilde{S}_0(y\,|\,a,z)}\right]G_P(dy\,|\,a,z)$$
$$+\int_\alpha^\beta (G_P-G_0)(y\,|\,a,z)\left[\frac{1}{S_0(y\,|\,a,z)^2}\omega(dy,z)+\frac{2\omega(y,z)}{S_0(y\,|\,a,z)^3}F_0(dy\,|\,a,z)\right].$$

This allows us to write

$$\left|\int_\alpha^\beta \omega(y,z)\left[\frac{\gamma_{P,\natural}(y,a,z)}{\tilde{S}_P(y\,|\,a,z)}\tilde{\Lambda}_P(dy\,|\,a,z)-\frac{\gamma_{0,\natural}(y,a,z)}{\tilde{S}_0(y\,|\,a,z)}\tilde{\Lambda}_0(dy\,|\,a,z)\right]\right|$$
$$\leq \left[2\sup_{y\in[\alpha,\beta]}|\omega(y,z)|+\int_\alpha^\beta |\omega(dy,z)|\right]\sup_{y\in[\alpha,\beta]}\left|\frac{\gamma_{P,\natural}(y,a,z)}{\tilde{S}_P(y\,|\,a,z)}-\frac{\gamma_{0,\natural}(y,a,z)}{\tilde{S}_0(y\,|\,a,z)}\right|$$
$$+\sup_{y\in[\alpha,\beta]}|\omega(y,z)|\left[\frac{1}{\tilde{S}_P(\beta\,|\,a,z)}+\frac{1}{\tilde{S}_0(\beta\,|\,a,z)}\right]\sup_{y\in[\alpha,\beta]}\left|\frac{1}{\tilde{S}_P(y\,|\,a,z)}-\frac{1}{\tilde{S}_0(y\,|\,a,z)}\right|$$
$$+\left[\frac{\int_\alpha^\beta |\omega(dy,z)|}{S_0(\beta\,|\,a,z)^2}+\frac{2\sup_{y\in[\alpha,\beta]}|\omega(y,z)|}{S_0(\beta\,|\,a,z)^3}\right]\sup_{y\in[\alpha,\beta]}|G_P(y\,|\,a,z)-G_0(y\,|\,a,z)|\,,$$

thus implying the claimed inequality. □

Proof of Theorem 1

Let $P\in\mathcal{M}$ be given, and take $\{P_\epsilon : |\epsilon|\leq\delta\}$ to be a suitably smooth and bounded (i.e., Hellinger-differentiable) path with $P_{\epsilon=0}=P$ and score for ϵ at $\epsilon=0$ given by $h\in L_2^0(P)$, and let $h=h_1+h_2$ denote the $L_2^0(P)$–unique decomposition of h for which $o\mapsto h_1(y,\delta,w,a,z)$ and $o\mapsto h_2(a,z)$ are such that, P–almost surely, $E_P[h_1(O)\,|\,A,Z]=0$, $E_P[h_2(O)]=0$, $var_P[h_1(O)\,|\,A,Z]<\infty$, and $var_P[h_2(O)]<\infty$. We wish to compute the pathwise derivative

$$\frac{\partial}{\partial\epsilon}\Psi(P_\epsilon)\Big|_{\epsilon=0} = \frac{\partial}{\partial\epsilon}\iint \varphi(t,z)\tilde{F}_\epsilon(dt\,|\,a_0,z)\tilde{H}_\epsilon(dz)\Big|_{\epsilon=0}$$

$$= \frac{\partial}{\partial \epsilon} \iiint \varphi(t,z) \tilde{F}_\epsilon(dt \,|\, a_0, z) \bar{\gamma}_\epsilon(a,z) J_\epsilon(da, dz) \Big|_{\epsilon=0}$$

where here and below we use the shorthand notation A_ϵ to refer to A_{P_ϵ} for any relevant quantity A_P indexed by P. Furthermore, under mild regularity conditions allowing interchange of integral and derivative operations, this pathwise derivative can be decomposed as (1) + (2) + (3) with

$$\begin{aligned}
(1) &= \iint \frac{\partial}{\partial \epsilon} \int \varphi(t,z) \tilde{F}_\epsilon(dt \,|\, a_0, z) \Big|_{\epsilon=0} \bar{\gamma}_P(a,z) J_P(da, dz) \\
(2) &= \iiint \varphi(t,z) \tilde{F}_P(dt \,|\, a_0, z) \, \frac{\partial}{\partial \epsilon} \bar{\gamma}_\epsilon(a,z) \Big|_{\epsilon=0} J_P(da, dz) \\
(3) &= \iiint \varphi(t,z) \tilde{F}_P(dt \,|\, a_0, z) \bar{\gamma}_P(a,z) \, \frac{\partial}{\partial \epsilon} J_\epsilon(da, dz) \Big|_{\epsilon=0} .
\end{aligned}$$

Below, we study each of these summands separately.

By integration by parts, we first note that $\int \varphi(t,z) \tilde{F}_\epsilon(dt \,|\, a_0, z) = \varphi(0,z) + \int \tilde{S}_\epsilon(t \,|\, a_0, z) \varphi(dt, z)$, and so, we can equivalently write

$$(1) = \iiint \frac{\partial}{\partial \epsilon} \tilde{S}_\epsilon(t \,|\, a_0, z) \Big|_{\epsilon=0} \varphi(dt, z) \bar{\gamma}_P(a,z) J_P(da, dz) \; .$$

To compute the pathwise derivatives of $\epsilon \mapsto \tilde{S}_\epsilon(t \,|\, a_0, z)$, we first consider the pathwise derivative of $\epsilon \mapsto \tilde{\Lambda}_\epsilon(t \,|\, a_0, z)$, where $\tilde{\Lambda}_\epsilon$ is the cumulative hazard function corresponding to $\tilde{S}_\epsilon$, defined as $\tilde{\Lambda}_\epsilon(t \,|\, a_0, z) := \int_0^t R_\epsilon(u \,|\, a_0, z)^{-1} F_{1,\epsilon}(du \,|\, a_0, z)$. We can show that

$$\begin{aligned}
\frac{\partial}{\partial \epsilon} \tilde{\Lambda}_\epsilon(t \,|\, a_0, z) \Big|_{\epsilon=0} &= \int_0^t \frac{\frac{\partial}{\partial \epsilon} F_{1,\epsilon}(du \,|\, a_0, z) \big|_{\epsilon=0}}{R_P(u \,|\, a_0, z)} - \int_0^t \frac{\frac{\partial}{\partial \epsilon} R_\epsilon(u \,|\, a_0, z) \big|_{\epsilon=0}}{R_P(u \,|\, a_0, z)^2} F_{1,P}(du \,|\, a_0, z) \\
&= E_P \left[\left\{ \frac{\Delta I(Y \le t)}{R_P(Y \,|\, a_0, z)} - \int_0^t \frac{I(W \le u \le Y)}{R_P(u \,|\, a_0, z)} \tilde{\Lambda}_P(du \,|\, a_0, z) \right\} h_1(O) \,\middle|\, A = a_0, Z = z \right] \\
&= E_P \left[-\phi_{\mathrm{KM},P}((u,a,z) \mapsto I(u \le t))(O) h_1(O) \,|\, A = a_0, Z = z \right]
\end{aligned}$$

by first showing that

$$\frac{\partial}{\partial\epsilon}F_{1,\epsilon}(u\,|\,a_0,z)\bigg|_{\epsilon=0} \;=\; \iiint \frac{\delta I(y\le u)}{R_P(y\,|\,a_0,z)}h_1(y,\delta,w,a_0,z)P(dy,d\delta,dw\,|\,a_0,z)$$

and

$$\frac{\partial}{\partial\epsilon}R_\epsilon(u\,|\,a_0,z)\bigg|_{\epsilon=0} \;=\; \iiint\int_0^t \frac{I(w\le u\le y)}{R_P(u\,|\,a_0,z)}\tilde{\Lambda}_P(du\,|\,a_0,z)h_1(y,\delta,w,a_0,z)P(dy,d\delta,dw\,|\,a_0,z).$$

This then implies, using Theorem 8 of Gill and Johansen (1990), we that

$$\begin{aligned}\frac{\partial}{\partial\epsilon}\tilde{S}_\epsilon(t\,|\,a_0,z)\bigg|_{\epsilon=0} &= -\tilde{S}_P(t\,|\,a_0,z)\,\frac{\partial}{\partial\epsilon}\tilde{\Lambda}_\epsilon(t\,|\,a_0,z)\bigg|_{\epsilon=0}\\ &= E_P\left[-\tilde{S}_P(t\,|\,a_0,z)\phi_{\mathrm{KM},P}((u,a,z)\mapsto I(u\le t))(O)h_1(O)\,\middle|\,A=a_0,Z=z\right].\end{aligned}$$

In particular, this allows us to compute

$$\begin{aligned}\int \frac{\partial}{\partial\epsilon}\tilde{S}_\epsilon(t\,|\,a_0,z)\bigg|_{\epsilon=0}\varphi(dt,z) &= E_P\left[-\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h_1(O)\,|\,A=a_0,Z=z\right]\\ &= E_P\left[-\frac{I(A=a_0)}{P(A=a_0\,|\,Z=z)}\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h_1(O)\,\middle|\,Z=z\right]\end{aligned}$$

and therefore, we have that

$$\begin{aligned}(1) &= \iint E_P\left[-\frac{I(A=a_0)}{P(A=a_0\,|\,Z=z)}\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h_1(O)\,\middle|\,Z=z\right]\bar{\gamma}_P(a,z)J_P(da,dz)\\ &= E_P\left[E_P\left[-\frac{I(A=a_0)}{P(A=a_0\,|\,Z=z)}\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h_1(O)\,\middle|\,Z\right]\bar{\gamma}_P(A,Z)\right]\\ &= E_P\left[-\frac{I(A=a_0)}{P(A=a_0\,|\,Z=z)}\bar{\gamma}_P(Z)\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h_1(O)\right],\end{aligned}$$

where we have defined $\bar{\gamma}_P : z \mapsto E_P[\bar{\gamma}_P(A,Z)\,|\,Z = z]$. From the result

$E_P[\phi_{\mathrm{KM},P}(L_{P,\varphi})(O) \mid A = a_0, Z] = 0$ P–almost surely, we can write

$$(1) = E_P\left[-\frac{I(A = a_0)}{P(A = a_0 \mid Z = z)}\bar{\gamma}_P(Z)\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)h(O)\right]$$

in view of the fact that $h_2(O)$ is only a function of (A, Z).

We now turn to computing the pathwise derivative of $\epsilon \mapsto \bar{\gamma}_\epsilon(a, z)$ at $\epsilon = 0$, which is critical for computing (2). We have that

$$\begin{aligned}
&\frac{\partial}{\partial\epsilon}\bar{\gamma}_\epsilon(a,z)\Big|_{\epsilon=0} = \frac{1}{\gamma_P}\left[\frac{\partial}{\partial\epsilon}\gamma_\epsilon(a,w)\Big|_{\epsilon=0} - \bar{\gamma}_P(a,z)\,\frac{\partial}{\partial\epsilon}\gamma_\epsilon\Big|_{\epsilon=0}\right] \\
&= \frac{\partial}{\partial\epsilon}\frac{\gamma_\epsilon(a,w)}{\gamma_P}\Big|_{\epsilon=0} - \frac{\bar{\gamma}_P(a,z)}{\gamma_P}\left\{\iint \frac{\partial}{\partial\epsilon}\gamma_\epsilon(\bar{a},\bar{z})\Big|_{\epsilon=0} J_P(d\bar{a},d\bar{z}) + \iint \gamma_P(\bar{a},\bar{z})\,\frac{\partial}{\partial\epsilon}J_\epsilon(d\bar{a},d\bar{z})\Big|_{\epsilon=0}\right\}
\end{aligned}$$

using that $\gamma_\epsilon = \iint \gamma_\epsilon(\bar{a},\bar{z})J_\epsilon(d\bar{a},d\bar{z})$, so that we can decompose $(2) = (2\mathrm{a}) + (2\mathrm{b}) + (2\mathrm{c})$ with

$$\begin{aligned}
(2\mathrm{a}) &= \frac{1}{\gamma_P}\iint \mu_P(z)\,\frac{\partial}{\partial\epsilon}\gamma_\epsilon(a,z)\Big|_{\epsilon=0} J_P(da,dz) \\
(2\mathrm{b}) &= -\Psi(P)\frac{1}{\gamma_P}\iint \frac{\partial}{\partial\epsilon}\gamma_\epsilon(a,z)\Big|_{\epsilon=0} J_P(da,dz) \\
(2\mathrm{c}) &= -\Psi(P)\frac{1}{\gamma_P}\iint \gamma_P(a,z)h_2(a,z)J_P(da,dz)\ .
\end{aligned}$$

In particular we note that

$$(2\mathrm{a}) + (2\mathrm{b}) = \iint \xi_P(z)\,\frac{\partial}{\partial\epsilon}\gamma_\epsilon(a,z)\Big|_{\epsilon=0} J_P(da,dz)$$

with $\xi_P(z) := \gamma_P^{-1}\{\mu_P(z) - \Psi(P)\}$. This expression involves the pathwise derivative of $\epsilon \mapsto \gamma_\epsilon(a, w)$ at $\epsilon = 0$, which we can be computed as

$$\begin{aligned}
&\frac{\partial}{\partial\epsilon}\gamma_\epsilon(a,z)\Big|_{\epsilon=0} = \frac{\partial}{\partial\epsilon}\int \frac{G_\epsilon(dw \mid a,z)}{\tilde{S}_\epsilon(w \mid a,z)}\Big|_{\epsilon=0} \\
&= \int \frac{1}{\tilde{S}_P(w \mid a,z)}\,\frac{\partial}{\partial\epsilon}G_\epsilon(dw \mid a,z)\Big|_{\epsilon=0} - \int \frac{\partial}{\partial\epsilon}\tilde{S}_\epsilon(w \mid a,z)\Big|_{\epsilon=0}\frac{G(dw \mid a,z)}{\tilde{S}_P(w \mid a,z)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \iiint \frac{1}{\tilde{S}_P(w \mid a, z)} \frac{\partial}{\partial \epsilon} P_\epsilon(dy, d\delta, dw \mid a, z)\Big|_{\epsilon=0} + \int \frac{\partial}{\partial \epsilon} \tilde{\Lambda}_\epsilon(w \mid a_0, z)\Big|_{\epsilon=0} \frac{G(dw \mid a, z)}{\tilde{S}_P(w \mid a, z)} \\
&= \iiint \frac{1}{\tilde{S}_P(w \mid a, z)} h_1(o) P(dy, d\delta, dw \mid a, z) \\
&\qquad + \int E_P\left[-\phi_{\mathrm{KM},P}((u, a, z) \mapsto I(u \le w))(O) h_1(O) \mid A = a, Z = z\right] \frac{G(dw \mid a, z)}{\tilde{S}_P(w \mid a, z)} \\
&= E_P\left[\left\{\frac{1}{\tilde{S}_P(W \mid a, z)} - \gamma_P(a, z) - \phi_{\mathrm{KM},P}(\gamma_{P,\natural})(O)\right\} h(O) \,\middle|\, A = a, Z = z\right].
\end{aligned}
$$

where we re-centered the first summand by $\gamma_P(a, z) = \int \tilde{S}_P(w \mid a, z)^{-1}) G_P(dw \mid a, z)$ in the penultimate step, which then allowed h_1 to be replaced by h since $h_2(O)$ only depends on (A, Z). We can therefore write that

$$
(2\mathrm{a}) + (2\mathrm{b}) = E_P\left[\xi_P(Z)\left\{\frac{1}{\tilde{S}_P(W \mid A, Z)} - \gamma_P(A, Z) - \phi_{\mathrm{KM},P}(\gamma_{P,\natural})(O)\right\} h(O)\right].
$$

Using the fact that $h_1(O)$ has mean zero conditional on (A, Z) and that $E_P[\bar{\gamma}_P(A, Z)] = 1$, we can write that

$$
(2\mathrm{c}) = -\Psi(P) \iint \bar{\gamma}_P(a, z) h_2(a, z) J_P(da, dz) = -E_P\left[\Psi(P)\left\{\bar{\gamma}_P(A, Z) - 1\right\} h(O)\right],
$$

from which we conclude that

$$
\begin{aligned}
(2) &= E_P\left[\left[\xi_P(Z)\left\{\frac{1}{\tilde{S}_P(W \mid A, Z)} - \gamma_P(A, Z) - \phi_{\mathrm{KM},P}(\gamma_{P,\natural})(O)\right\} - \Psi(P)\left\{\bar{\gamma}_P(A, Z) - 1\right\}\right] h(O)\right] \\
&= E_P\left[\left[\xi_P(Z)\left\{\frac{1}{\tilde{S}_P(W \mid A, Z)} - \gamma_P(A, Z) - \phi_{\mathrm{KM},P}(\gamma_{P,\natural})(O)\right\} - \Psi(P)\left\{\bar{\gamma}_P(A, Z) - 1\right\}\right] h(O)\right]
\end{aligned}
$$

The final term to compute, (3), has a simple form. Indeed, using a similar argument as above, it can be written as

$$
\begin{aligned}
(3) &= \iint \mu_P(z) \bar{\gamma}_P(a, z) h_2(a, z) J_P(da, dz) \\
&= E_P\left[\mu_P(Z) \bar{\gamma}_P(A, Z) h_2(A, Z)\right] = E_P\left[\left\{\mu_P(Z) \bar{\gamma}_P(A, Z) - \Psi(P)\right\} h(O)\right].
\end{aligned}
$$

Adding the expressions derived for each of (1), (2) and (3), as claimed, we find that the pathwise derivative of $\epsilon \mapsto \Psi(P_\epsilon)$ at $\epsilon = 0$ is given by

$$E_P\left[\left[-\frac{I(A=a_0)\bar{\gamma}_P(Z)}{P(A=a_0\,|\,Z=z)}\phi_{\mathrm{KM},P}(L_{P,\varphi})(O)+\xi_P(Z)\left\{\frac{1}{\tilde{S}_P(W\,|\,A,Z)}-\phi_{\mathrm{KM},P}(\gamma_{P,\natural})(O)\right\}\right]h(O)\right],$$

which is simply $E_P[\phi_P(O)h(O)]$ with ϕ_P as defined in the main text. This establishes the pathwise differentiability of $P \mapsto \Psi(P)$ relative to a nonparametric model as well as the fact that ϕ_P is the nonparametric efficient influence function of Ψ at P.

We now study the linearization of Ψ around P_0 based on ϕ_P. Specifically, we derive the form of the remainder term $R(P,P_0) := \Psi(P) - \Psi(P_0) - (P-P_0)\phi_P$ from this linearization. We begin by decomposing the difference $\Psi(P)-\Psi(P_0)$ as the sum (D1)+(D2)+(D3)+(D4) with

$$\begin{aligned}
(\mathrm{D1}) &= \iint \varphi(y,z)(\tilde{F}_P - F_0)(dy\,|\,a_0,z)\tilde{H}_0(dz)\\
(\mathrm{D2}) &= \iint \xi_P(z)\left\{\gamma_P(a,z)-\gamma_0(a,z)\right\}J_0(da,dz)\\
(\mathrm{D3}) &= \frac{\gamma_0-\gamma_P}{\gamma_0}\iint \xi_P(z)\gamma_0(a,z)J_0(da,dz)\\
(\mathrm{D4}) &= \iint \xi_P(z)\gamma_P(a,z)(J_P-J_0)(da,dz)\ .
\end{aligned}$$

First, using part (a) of Lemma 2, we can write

$$(\mathrm{D1}) = -\iint L_{P,\varphi}(y,a_0,z)\frac{\tilde{S}_0(y\,|\,a_0,z)}{\tilde{S}_P(y\,|\,a_0,z)}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a_0,z)\tilde{H}_0(dz)\ .$$

Next, we decompose (D2) as the sum (D2a) + (D2b) + (D2c) + (D2d) with

$$\begin{aligned}
(\mathrm{D2a}) &:= -\iiint \xi_P(z)\left\{\tilde{S}_P(w\,|\,a,z)-\tilde{S}_0(w\,|\,a,z)\right\}\frac{G_P(dw\,|\,a,z)}{\tilde{S}_P(w\,|\,a,z)^2}J_0(da,dz)\\
(\mathrm{D2b}) &:= \iiint \frac{\xi_P(z)}{\tilde{S}_P(w\,|\,a,z)}\left(G_P-G_0\right)(dw\,|\,a,z)J_0(da,dz)
\end{aligned}$$

$$\text{(D2c)} := \iiint \xi_P(z) \left\{ \frac{1}{\tilde{S}_0(w \,|\, a, z)} - \frac{1}{\tilde{S}_P(w \,|\, a, z)} \right\} (G_P - G_0)\,(dw \,|\, a, z) J_0(da, dz)$$

$$\text{(D2d)} := \iiint \left\{ \frac{\xi_P(z)}{\tilde{S}_0(w \,|\, a, z)} - \frac{\xi_P(z)}{\tilde{S}_P(w \,|\, a, z)} \right\} \left\{ \tilde{S}_0(w \,|\, a, z) - \tilde{S}_P(w \,|\, a, z) \right\} \frac{G_P(dw \,|\, a, z)}{\tilde{S}_P(w \,|\, a, z)} J_0(da, dz)$$

Using part (b) of Lemma 2, we can rewrite

$$\text{(D2a)} = \iint \xi_P(z) \left\{ \int \gamma_{P,\natural}(y, a, z) \frac{\tilde{S}_0(y \,|\, a, z)}{\tilde{S}(y \,|\, a, z)} (\tilde{\Lambda} - \tilde{\Lambda}_0)(dy \,|\, a, z) \right\} J_0(da, dz)\ .$$

We then note that we can decompose (D3) as the sum (D3a) + (D3b) with

$$\text{(D3a)} := \frac{\gamma_P - \gamma_0}{\gamma_0} \iint \xi_P(z) \left[\gamma_P(a, z) - \gamma_0(a, z) \right] J_0(da, dz)$$

$$\text{(D3b)} := \frac{\gamma_P - \gamma_0}{\gamma_0} \iint \xi_P(z) \gamma_P(a, z) \left(J_P - J_0 \right) (da, dz)$$

using the fact that $\iint \xi_P(z) \gamma_P(a, z) J_P(da, dz) = 0$.

We now compute the linear term $(P - P_0)\phi_P = -P_0\phi_P$, which can be decomposed as the sum (L1) + (L2) + (L3) + (L4), where we define

$$\text{(L1)} := - \iint \frac{\pi_0(z)}{\pi(z)} \frac{\bar{\gamma}_P(a_0, z)}{\bar{\gamma}_0(a_0, z)} \frac{\tilde{R}_0(y, a_0, z)}{\tilde{R}_P(y, a_0, z)} L_{P,\varphi}(y, a_0, z) (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, a_0, z) \tilde{H}_0(dz)$$

$$\text{(L2)} := \iiint \xi_P(z) \frac{\tilde{R}_0(y, a, z)}{\tilde{R}_P(y, a, z)} \gamma_{P,\natural}(y, a, z) (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, a, z) J_0(da, dz)$$

$$\text{(L3)} := \iiint \frac{\xi_P(z)}{\tilde{S}_P(w \,|\, a, z)} (G_P - G_0)(dw \,|\, a, z) J_0(da, dz)$$

$$\text{(L4)} := \iint \xi_P(z) \gamma_P(a, z) (J_P - J_0)(da, dz)\ .$$

We now scrutinize the terms appearing in $R(P, P_0) = \text{(D1)} + \text{(D2)} + \text{(D3)} + \text{(D4)} - \text{(L1)} - \text{(L2)} - \text{(L3)} - \text{(L4)}$. First, we observe that $\text{(D4)} - \text{(L4)} = 0$ and $\text{(D2b)} - \text{(L3)} = 0$. Next, we use $\nu_P(y, a, z) := \frac{\tilde{S}_P(y \,|\, a, z)}{\tilde{R}_P(y, a, z)}$ for simplicity and we note that

(D1) − (L1)

$$
\begin{aligned}
&= \iint L_{P,\varphi}(y,a_0,z)\left\{\frac{\pi_0(z)}{\pi(z)}\frac{\bar{\gamma}_P(a_0,z)}{\bar{\gamma}_0(a_0,z)}\frac{\tilde{R}_0(y,a_0,z)}{\tilde{R}_P(y,a_0,z)}-\frac{\tilde{S}_0(y\,|\,a_0,z)}{\tilde{S}_P(y\,|\,a_0,z)}\right\}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a_0,z)\tilde{H}_0(dz)\\
&= \iint L_{P,\varphi}(y,a_0,z)\left\{\frac{\pi_0(z)}{\pi(z)}\frac{\bar{\gamma}_P(a_0,z)}{\bar{\gamma}_0(a_0,z)}\frac{\nu_P(y,a_0,z)}{\nu_0(y,a_0,z)}-1\right\}\frac{\tilde{S}_0(y\,|\,a_0,z)}{\tilde{S}_P(y\,|\,a_0,z)}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a_0,z)\tilde{H}_0(dz)\\
&= \iint L_{P,\varphi}(y,a_0,z)\left\{\frac{\pi_0(z)}{\pi(z)}\frac{\bar{\gamma}_P(a_0,z)}{\bar{\gamma}_0(a_0,z)}\frac{\nu_P(y,a_0,z)}{\nu_0(y,a_0,z)}-1\right\}\left(\frac{\tilde{S}_P}{\tilde{S}_0}-1\right)(dy\,|\,a_0,z)\tilde{H}_0(dz)\\
&= R_1(P,P_0)\,,
\end{aligned}
$$

where we used the fact that

$$
\frac{\tilde{S}_0(y\,|\,a,z)}{\tilde{S}_P(y\,|\,a,z)}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a,z)=\left(\frac{\tilde{S}_0}{\tilde{S}_P}-1\right)(dy\,|\,a,z)\;,
$$

which is a consequence of the Duhamel equation in Theorem 6 of Gill and Johansen (1990). Using the same argument, we note that

$$
\begin{aligned}
&\text{(D2a)} - \text{(L2)}\\
&= \iiint \xi_P(z)\gamma_{P,\natural}(y,a,z)\left\{\frac{\tilde{S}_0(y\,|\,a,z)}{\tilde{S}_P(y\,|\,a,z)}-\frac{\tilde{R}_0(y\,|\,a,z)}{\tilde{R}_P(y\,|\,a,z)}\right\}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a,z)J_0(da,dz)\\
&= \iiint \xi_P(z)\gamma_{P,\natural}(y,a,z)\left\{1-\frac{\nu_P(y\,|\,a,z)}{\nu_0(y\,|\,a,z)}\right\}\frac{\tilde{S}_0(y\,|\,a,z)}{\tilde{S}_P(y\,|\,a,z)}(\tilde{\Lambda}_P-\tilde{\Lambda}_0)(dy\,|\,a,z)J_0(da,dz)\\
&= \iiint \xi_P(z)\gamma_{P,\natural}(y,a,z)\left\{1-\frac{\nu_P(y\,|\,a,z)}{\nu_0(y\,|\,a,z)}\right\}\left(\frac{\tilde{S}_P}{\tilde{S}_0}-1\right)(dy\,|\,a,z)J_0(da,dz) \;=\; R_2(P,P_0)\,.
\end{aligned}
$$

We also note that (D3) = (D3a) + (D3b) = $R_4(P,P_0)$ and that (D2c) + (D2d) = $R_3(P,P_0)$. As such, we find that the remainder from the linear approximation of $\Psi(P)-\Psi(P_0)$ by $(P-P_0)\phi_P$ is given by $\{(\text{D1})+(\text{D1})+(\text{D1})+(\text{D1})\}-\{(\text{L1})+(\text{L2})+(\text{L3})+(\text{L4})\}$, and this quantity coincides precisely with the form of $R(P,P_0)$ given in the theorem.

Proof of Theorem 2

We begin by studying the estimation error of ψ_n^*. First, by Theorem 1, we note that $\psi_{\eta_{n,k}} - \psi_0 = -P_0\phi_{\eta_{n,k}} + R(\eta_{n,k}, \eta_0)$ for each $k = 1, 2, \ldots, K$, and so, denoting $\bar{\phi}_\infty := \phi_\infty - P_0\phi_\infty$, we can write

$$\begin{aligned}
\psi_n^* - \psi_0 &= \frac{1}{K}\sum_{k=1}^{K}\left(\psi_{\eta_{n,k}} + \mathbb{P}_{n,k}\phi_{\eta_{n,k}} - \psi_0\right) \\
&= \frac{1}{K}\sum_{k=1}^{K}\left\{(\mathbb{P}_{n,k} - P_0)\phi_\infty + (\mathbb{P}_{n,k} - P_0)(\phi_{\eta_{n,k}} - \phi_\infty) + R(\eta_{n,k}, \eta_0)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\bar{\phi}_\infty(O_i) + r_{an} + r_{bn} + r_{cn}\ ,
\end{aligned}$$

where we have defined $r_{an} := \frac{1}{K}\sum_{k=1}^{K}(\mathbb{P}_{n,k} - \mathbb{P}_n)\phi_\infty$, $r_{bn} := \frac{1}{K}\sum_{k=1}^{K}(\mathbb{P}_{n,k} - P_0)(\phi_{\eta_{n,k}} - \phi_\infty)$ and $r_{cn} := \frac{1}{K}\sum_{k=1}^{K} R(\eta_{k,n}, \eta_0)$. We first show that $r_{an} = o_P(n^{-1/2})$. To see this, we note that since we can always find $n_1, n_2, \ldots, n_K$ such that $|n_k K - n| \leq K$ for each $k = 1, 2, \ldots, n$, we have that

$$\begin{aligned}
|r_{an}| &= \left|\sum_{k=1}^{K}\sum_{i\in V_k}\left(\frac{1}{Kn_k} - \frac{1}{n}\right)\phi_\infty(O_i)\right| \leq \max_k\left|\frac{n}{Kn_k} - 1\right|\frac{1}{n}\sum_{i=1}^{n}|\phi_\infty(O_i)| \\
&\leq \left(\frac{K}{n-K}\right)\frac{1}{n}\sum_{i=1}^{n}|\phi_\infty(O_i)| = O_P(n^{-1}) \\
&= o_P(n^{-1/2})
\end{aligned}$$

in view of the fact that $var_0\{\phi_\infty(O)\} < \infty$. Next, we show that $r_{bn} = o_P(n^{-1/2})$ under Conditions (C1)–(C4). Denoting by $\mathcal{D}_k := \cup_{j\notin k}\mathcal{V}_j$ the portion of the dataset used to construct $\eta_{n,k}$ and writing $A_{n,k} := n^{1/2}(\mathbb{P}_{n,k} - P_0)(\phi_{\eta_{n,k}} - \phi_\infty)$, by Chebyshev's inequality, for any $\varepsilon > 0$, we have that

$$P_0\left(|A_{n,k}| > \varepsilon \,|\, \mathcal{D}_k\right) \leq \frac{var_0\left(A_{n,k} \,|\, \mathcal{D}_{n,k}\right)}{\varepsilon^2} \leq \frac{n}{n_k}\cdot\frac{P_0\left(\phi_{n,k} - \phi_\infty\right)^2}{\varepsilon^2} \longrightarrow 0$$

provided $P_0\left(\phi_{n,k} - \phi_\infty\right)^2$ tends to zero in probability and using that $n/n_k \to K < \infty$. By the Bounded Convergence Theorem, it follows that $A_{n,k}$ tends to zero in probability since we can write

$$P_0\left(|A_{n,k}| > \varepsilon\right) = E_0\left[P_0\left(|A_{n,k}| > \varepsilon \,|\, \mathcal{D}_k\right)\right] \longrightarrow 0$$

for each $\varepsilon > 0$. This implies that $r_{bn} = n^{-1/2}\frac{1}{K}\sum_{k=1}^{K} A_{n,k} = o_P(n^{-1/2})$, as claimed, provided we can show that $P_0\left(\phi_{n,k} - \phi_\infty\right)^2 = o_P(1)$. To do so, we first observe that we can express $\phi_{n,k} - \phi_\infty$ as the sum $U_{1,n,k} + U_{2,n,k} + \ldots + U_{11,n,k}$, where we define

$$U_{1,n,k} : o \mapsto I(a = a_0)\left[\frac{\bar{\gamma}_{n,k}(a,z)}{\pi_{n,k}(z)} - \frac{\bar{\gamma}_\infty(a,z)}{\pi_\infty(z)}\right]\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(z,a,w,y,\delta)$$

$$U_{2,n,k} : o \mapsto \frac{I(a = a_0, \delta = 1)\bar{\gamma}_{n,k}(a,z)\tilde{S}_\infty(y \,|\, a,z)}{\pi_{n,k}(z)R_\infty(y \,|\, a,z)}\left[\frac{L_{n,k,\varphi}(y,a,z)}{\tilde{S}_{n,k}(y \,|\, a,z)} - \frac{L_{\infty,\varphi}(y,a,z)}{\tilde{S}_\infty(y \,|\, a,z)}\right]$$

$$U_{3,n,k} : o \mapsto \frac{I(a = a_0, \delta = 1)\bar{\gamma}_{n,k}(a,z)L_{n,k,\varphi}(y,a,z)}{\pi_{n,k}(z)S_{n,k}(y \,|\, a,z)}\left[\frac{S_{n,k}(y \,|\, a,z)}{\tilde{R}_{n,k}(y \,|\, a,z)} - \frac{S_\infty(y \,|\, a,z)}{\tilde{R}_\infty(y \,|\, a,z)}\right]$$

$$U_{4,n,k} : o \mapsto -\frac{I(a = a_0)\bar{\gamma}_{n,k}(a,z)}{\pi_{n,k}(z)}\int_w^y \left[\frac{\tilde{S}_{n,k}(u \,|\, a,z)}{R_{n,k}(u \,|\, a,z)} - \frac{\tilde{S}_\infty(u \,|\, a,z)}{R_\infty(u \,|\, a,z)}\right]\frac{L_{\infty,\varphi}(u,a,z)}{\tilde{S}_\infty(u \,|\, a,z)}\tilde{\Lambda}_\infty(du \,|\, a,z)$$

$$U_{5,n,k} : o \mapsto -\frac{I(a = a_0)\bar{\gamma}_{n,k}(a,z)}{\pi_{n,k}(z)} \int_w^y \frac{\tilde{S}_{n,k}(u \,|\, a,z)}{R_{n,k}(u \,|\, a,z)}\left[\frac{L_{n,k,\varphi}(u,a,z)}{\tilde{S}_{n,k}(u \,|\, a,z)}\tilde{\Lambda}_{n,k}(du \,|\, a,z) - \frac{L_{\infty,\varphi}(u,a,z)}{\tilde{S}_\infty(u \,|\, a,z)}\tilde{\Lambda}_\infty(du \,|\, a,z)\right]$$

$$U_{6,n,k} : o \mapsto \frac{\mu_{n,k}(z) - \mu_\infty(z)}{\gamma_\infty}\left[\frac{1}{S_\infty(w \,|\, a,z)} - \phi_{\mathrm{KM},\infty}(\gamma_{\infty,\natural})(z,a,w,y,\delta)\right]$$

$$U_{7,n,k} : o \mapsto \mu_{n,k}(z)\left[\frac{1}{\tilde{S}_{n,k}(w \,|\, a,z)\gamma_{n,k}} - \frac{1}{\tilde{S}_\infty(w \,|\, a,z)\gamma_\infty}\right]$$

$$U_{8,n,k} : o \mapsto \frac{\delta\mu_{n,k}(z)\tilde{S}_\infty(y \,|\, a,z)}{R_\infty(y \,|\, a,z)}\left[\frac{\gamma_{n,k,\natural}(y,a,z)}{\gamma_{n,k}\tilde{S}_{n,k}(y \,|\, a,z)} - \frac{\gamma_{\infty,\natural}(y,a,z)}{\gamma_\infty\tilde{S}_\infty(y \,|\, a,z)}\right]$$

$$U_{9,n,k} : o \mapsto -\frac{\delta\mu_{n,k}(z)\gamma_{n,k,\natural}(y,a,z)}{\gamma_{n,k}\tilde{S}_{n,k}(y \,|\, a,z)}\left[\frac{\tilde{S}_{n,k}(y \,|\, a,z)}{R_{n,k}(y \,|\, a,z)} - \frac{\tilde{S}_\infty(y \,|\, a,z)}{R_\infty(y \,|\, a,z)}\right]$$

$$U_{10,n,k} : o \mapsto \mu_{n,k}(z)\int_w^y \left[\frac{\tilde{S}_{n,k}(u \,|\, a,z)}{R_{n,k}(u \,|\, a,z)} - \frac{\tilde{S}_\infty(u \,|\, a,z)}{R_\infty(u \,|\, a,z)}\right]\frac{\gamma_{\infty,\natural}(u \,|\, a,z)}{\gamma_\infty\tilde{S}_\infty(u \,|\, a,z)}\tilde{\Lambda}_\infty(du \,|\, a,z)$$

$$U_{11,n,k} : o \mapsto \mu_{n,k}(z) \int_w^y \frac{\tilde{S}_{n,k}(u \mid a, z)}{R_{n,k}(u \mid a, z)} \left[\frac{\gamma_{n,k,\natural}(u, a, z)}{\gamma_{n,k}\tilde{S}_{n,k}(u \mid a, z)} \tilde{\Lambda}_{n,k}(du \mid a, z) - \frac{\gamma_{\infty,\natural}(u, a, z)}{\gamma_{\infty}\tilde{S}_{\infty}(u \mid a, z)} \tilde{\Lambda}_{\infty}(du \mid a, z) \right]$$

with $\mu_{n,k}(z) := \int \varphi(u, z)(F_{n,k} - F_0)(du \mid a_0, z)$. By the triangle inequality, we then have that

$$P_0 \left(\phi_{n,k} - \phi_{\infty}\right)^2 \leq \{(P_0 U_{1,n,k}^2)^{\frac{1}{2}} + (P_0 U_{2,n,k}^2)^{\frac{1}{2}} + \ldots + (P_0 U_{11,n,k}^2)^{\frac{1}{2}}\}^2,$$

and so, we can focus on bounding each $P_0 U_{j,n,k}^2$ separately. We define $\overline{\tau}(z) := \min\{\overline{\tau}_C(z), \overline{\tau}_T(z)\}$ and the (random) bounding terms $M_{1,n,k}, M_{2,n,k}, \ldots, M_{6,n,k}$ given by

$$\begin{aligned}
M_{1,n,k}^2 &:= E_0 \left| \frac{\bar{\gamma}_{n,k}(a_0, Z)}{\pi_{n,k}(Z)} - \frac{\bar{\gamma}_{\infty}(a_0, Z)}{\pi_{\infty}(Z)} \right|^2 \\
M_{2,n,k}^2 &:= E_0 \left[\sup_{y \in [0, \overline{\tau}(a_0, Z)]} \left| \frac{L_{n,k,\varphi}(y, a_0, Z)}{\tilde{S}_{n,k}(y \mid a_0, Z)} - \frac{L_{\infty,\varphi}(y, a_0, Z)}{\tilde{S}_{\infty}(y \mid a_0, Z)} \right| \right]^2 \\
M_{3,n,k}^2 &:= E_0 \left[\sup_{y \in [0, \overline{\tau}(A, Z)]} \left| \frac{\tilde{S}_{n,k}(y \mid A, Z)}{R_{n,k}(y \mid A, Z)} - \frac{\tilde{S}_{\infty}(y \mid A, Z)}{R_{\infty}(y \mid A, Z)} \right| \right]^2 \\
M_{4,n,k}^2 &:= E_0 \left[\sup_{y \in [\underline{\tau}_T(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{\bar{\gamma}_{n,k,\natural}(y, A, Z)}{\tilde{S}_{n,k}(y \mid A, Z)} - \frac{\bar{\gamma}_{\infty,\natural}(y, A, Z)}{\tilde{S}_{\infty}(y \mid A, Z)} \right| \right]^2 \\
M_{5,n,k}^2 &:= E_0 \left[\sup_{y \in [\underline{\tau}_W(A, Z), \overline{\tau}_W(A, Z)]} \left| \frac{1}{\tilde{S}_{n,k}(y \mid A, Z)} - \frac{1}{\tilde{S}_{\infty}(y \mid A, Z)} \right| \right]^2 \\
M_{6,n,k}^2 &:= E_0 \left[\sup_{u \in [\underline{\tau}_W(A, Z), \overline{\tau}(A, Z)]} |G_{n,k}(u \mid A, Z) - G_{\infty}(u \mid A, Z)| \right]^2,
\end{aligned}$$

where E_0 is a P_0–expectation over the random data unit (W, A, Z) drawn independently of $\eta_{n,k}$. Below, we restrict our attention to the portion of the sample space on which $\tilde{F}_0(t \mid a_0, z)$ is identified — this is the relevant event to focus on since it has P_0–probability tending to one by Condition (A1). Before proceeding, we note that $|\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(z, a_0, w, y, \delta)| \leq V_{1,\infty}(y, z) + V_{2,\infty}(y, w, z)$, where we write

$$V_{1,\infty}(y, z) := \left| \frac{L_{\infty,\varphi}(y, a_0, z)}{R_{\infty}(y \mid a_0, z)} \right| = \left| \frac{L_{\infty,\varphi}(y, a_0, z)}{S_{\infty}(y \mid a_0, z)} \right| \left| \frac{S_{\infty}(y \mid a_0, z)}{R_{\infty}(y \mid a_0, z)} \right|$$

$$
\begin{aligned}
V_{2,\infty}(y,w,z) &:= \left| \int_w^y \frac{L_{\infty,\varphi}(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} \Lambda_\infty(du \,|\, a_0, z) \right| \\
&= \left| \int_w^y \frac{L_{\infty,\varphi}(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} S_\infty(u \,|\, a_0, z) \frac{\Lambda_\infty(du \,|\, a_0, z)}{S_\infty(u \,|\, a_0, z)} \right| \\
&= \left| \int_w^y L_{\infty,\varphi}(u \,|\, a_0, z) \frac{S_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} \frac{1}{S_\infty}(du \,|\, a_0, z) \right| \\
&\leq \sup_{u \in [w,y]} \left| \frac{S_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} \right| \int_w^y |L_{\infty,\varphi}(u \,|\, a_0, z)| \frac{1}{S_\infty}(du \,|\, a_0, z) \ .
\end{aligned}
$$

Now, for each z over $[0, \overline{\tau}(z)]$, $u \mapsto \varphi(u,z)$ is assumed to have finite variation, and so, we can write $\varphi(\cdot, z) = \varphi_1(\cdot, z) - \varphi_2(\cdot, z)$ for non-decreasing functions $\varphi_1(\cdot, z)$ and $\varphi_2(\cdot, z)$ with finite variation. This implies that we can write $L_{\infty,\varphi} = L_{\infty,\varphi_1} - L_{\infty,\varphi_2}$ with $L_{\infty,\varphi_1}, L_{\infty,\varphi_2} \geq 0$, which implies that $|L_{\infty,\varphi}(u \,|\, a_0, z)| \leq L_{\infty,\varphi_1}(u \,|\, a_0, z) + L_{\infty,\varphi_2}(u \,|\, a_0, z)$ and

$$
\begin{aligned}
&\int_w^y |L_{\infty,\varphi}(u \,|\, a_0, z)| \frac{1}{S_\infty}(du \,|\, a_0, z) \\
&\quad \leq \int_w^y L_{\infty,\varphi_1}(u \,|\, a_0, z) \frac{1}{S_\infty}(du \,|\, a_0, z) + \int_w^y L_{\infty,\varphi_2}(u \,|\, a_0, z) \frac{1}{S_\infty}(du \,|\, a_0, z) \ .
\end{aligned}
$$

By integration by parts, we have that

$$
\begin{aligned}
&\int_w^y L_{\infty,\varphi_j}(u \,|\, a_0, z) \frac{1}{S_\infty}(du \,|\, a_0, z) \\
&\quad = \frac{L_{\infty,\varphi_j}(y \,|\, a_0, z)}{S_\infty(y \,|\, a_0, z)} - \frac{L_{\infty,\varphi_j}(w \,|\, a_0, z)}{S_\infty(w \,|\, a_0, z)} - \int_w^y \frac{1}{S_\infty(u \,|\, a_0, z)} L_{\infty,\varphi_j}(du \,|\, a_0, z) \\
&\quad \leq 2 \sup_{u \in [w,y]} \left| \frac{L_{\infty,\varphi_j}(u \,|\, a_0, z)}{S_\infty(u \,|\, a_0, z)} \right| + \int_w^y \varphi_j(du, z) \ \leq \ 2 \sup_{u \in [w,y]} \left| \frac{L_{\infty,\varphi}(u \,|\, a_0, z)}{S_\infty(u \,|\, a_0, z)} \right| + \|\varphi_j(\cdot, z)\|_{v,[0,\overline{\tau}(z)]} \ ,
\end{aligned}
$$

which then implies that

$$
\int_w^y |L_{\infty,\varphi}(u \,|\, a_0, z)| \frac{1}{S_\infty}(du \,|\, a_0, z) \ \leq \ 4 \sup_{u \in [w,y]} \left| \frac{L_{\infty,\varphi}(u \,|\, a_0, z)}{S_\infty(u \,|\, a_0, z)} \right| + \|\varphi(\cdot, z)\|_{v,[0,\overline{\tau}(z)]}
$$

and therefore that $|\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(z,a_0,w,y,\delta)|$ is bounded above by

$$\sup_{u\in[w,y]}\left|\frac{S_\infty(u\,|\,a_0,z)}{R_\infty(u\,|\,a_0,z)}\right|\left[5\sup_{u\in[w,y]}\left|\frac{L_{\infty,\varphi}(u\,|\,a_0,z)}{S_\infty(u\,|\,a_0,z)}\right|+\|\varphi(\cdot,z)\|_{v,[0,\overline{\tau}(z)]}\right].$$

In view of the above, we have that

$$P_0U_{1,n,k}^2\;\le\;E_0\left[\left|\frac{\bar\gamma_{n,k}(a_0,Z)}{\pi_{n,k}(Z)}-\frac{\bar\gamma_\infty(a_0,Z)}{\pi_\infty(Z)}\right|^2\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(Z,a_0,W,Y,\Delta)^2\right]\;\le\;36\,\kappa^4M_{1,n,k}^2\;.$$

We can write that

$$P_0U_{2,n,k}^2\;\le\;E_0\left[\left|\frac{L_{n,k,\varphi}(Y,a_0,Z)}{\tilde S_{n,k}(Y\,|\,a_0,Z)}-\frac{L_{\infty,\varphi}(Y,a_0,Z)}{\tilde S_\infty(Y\,|\,a_0,Z)}\right|\frac{\bar\gamma_{n,k}(a_0,Z)}{\pi_{n,k}(Z)}\frac{\tilde S_\infty(Y\,|\,a_0,Z)}{R_\infty(Y\,|\,a_0,Z)}\right]^2\;\le\;\kappa^4M_{2,n,k}^2\;.$$

We can also write that

$$P_0U_{3,n,k}^2\;\le\;E_0\left[\left|\frac{\tilde S_{n,k}(Y\,|\,a_0,Z)}{R_{n,k}(Y\,|\,a_0,Z)}-\frac{\tilde S_\infty(Y\,|\,a_0,Z)}{R_\infty(Y\,|\,a_0,Z)}\right|\frac{\bar\gamma_{n,k}(a_0,Z)}{\pi_{n,k}(Z)}\frac{L_{n,k,\varphi}(Y,a_0,Z)}{\tilde S_{n,k}(Y\,|\,a_0,Z)}\right]^2\;\le\;\kappa^4M_{3,n,k}^2\;.$$

We have that

$$\begin{aligned}P_0U_{4,n,k}^2\;&=\;E_0\left[\frac{\bar\gamma_{n,k}(a_0,Z)}{\pi_{n,k}(Z)}\int_W^Y\left[\frac{\tilde S_{n,k}(u\,|\,a_0,Z)}{R_{n,k}(u\,|\,a_0,Z)}-\frac{\tilde S_\infty(u\,|\,a_0,Z)}{R_\infty(u\,|\,a_0,Z)}\right]L_{\infty,\varphi}(u,a_0,Z)\frac{1}{S_\infty}(du\,|\,a_0,Z)\right]^2\\
&\le\;E_0\left[\frac{\bar\gamma_{n,k}(a_0,Z)}{\pi_{n,k}(Z)}\sup_{u\in[W,Y]}\left|\frac{\tilde S_{n,k}(u\,|\,a_0,Z)}{R_{n,k}(u\,|\,a_0,Z)}-\frac{\tilde S_\infty(u\,|\,a_0,Z)}{R_\infty(u\,|\,a_0,Z)}\right|\int_W^Y|L_{\infty,\varphi}(u\,|\,a_0,Z)|\frac{1}{S_\infty}(du\,|\,a_0,Z)\right]^2\\
&\le\;\frac{25\kappa^4}{\gamma_{n,k}^2}M_{3,n,k}^2\;.\end{aligned}$$

Next, using part (c) of Lemma 1, and writing for convenience

$$\Theta_{n,k}(du):=\frac{L_{n,k,\varphi}(u,a_0,Z)}{\tilde S_{n,k}(u\,|\,a_0,Z)}\tilde\Lambda_{n,k}(du\,|\,a_0,Z)\;\;\text{and}\;\;\Theta_\infty(du):=\frac{L_{\infty,\varphi}(u,a_0,Z)}{\tilde S_\infty(u\,|\,a_0,Z)}\tilde\Lambda_\infty(du\,|\,a_0,Z)\;,$$

we have that

$$P_0 U_{5,n,k}^2 \leq E_0 \left[\frac{\bar{\gamma}_{n,k}(a_0, Z)}{\pi_{n,k}(Z)} \int_W^Y \frac{\tilde{S}_{n,k}(u \,|\, a_0, Z)}{R_{n,k}(u \,|\, a_0, Z)} \left(\Theta_{n,k} - \Theta_\infty\right)(du) \right]^2.$$

To make further progress on this bound, we note that

$$\begin{aligned}
&\left| \int_w^y \frac{\tilde{S}_{n,k}(u \,|\, a_0, z)}{R_{n,k}(u \,|\, a_0, z)} (\Theta_{n,k} - \Theta_\infty)(du) \right| \\
&\leq \left| \int_w^y \left[\frac{\tilde{S}_{n,k}(u \,|\, a_0, z)}{R_{n,k}(u \,|\, a_0, z)} - \frac{\tilde{S}_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} \right] (\Theta_{n,k} - \Theta_\infty)(du) \right| + \left| \int_w^y \frac{\tilde{S}_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} (\Theta_{n,k} - \Theta_\infty)(du) \right| \\
&\leq \sup_{u \in [w,y]} \left| \frac{\tilde{S}_{n,k}(u \,|\, a_0, z)}{R_{n,k}(u \,|\, a_0, z)} - \frac{\tilde{S}_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} \right| \int_w^y |(\Theta_{n,k} - \Theta_\infty)(du)| \\
&\qquad + \left| \int_w^y \frac{\tilde{S}_\infty(u \,|\, a_0, z)}{R_\infty(u \,|\, a_0, z)} (\Theta_{n,k} - \Theta_\infty)(du) \right|.
\end{aligned}$$

Using arguments used above, we observe that

$$\begin{aligned}
\int_w^y |(\Theta_{n,k} - \Theta_\infty)(du)| &\leq \int_w^y |\Theta_{n,k}(du)| + \int_w^y |\Theta_\infty(du)| \\
&= \int_w^y \frac{|L_{n,k,\varphi}(u \,|\, a_0, z)|}{\tilde{S}_{n,k}(u \,|\, a_0, z)} \tilde{\Lambda}_{n,k}(du \,|\, a_0, z) + \int_w^y \frac{|L_{\infty,\varphi}(u \,|\, a_0, z)|}{\tilde{S}_\infty(u \,|\, a_0, z)} \tilde{\Lambda}_\infty(du \,|\, a_0, z) \\
&\leq 4 \sup_{u \in [w,y]} \left| \frac{L_{\infty,\varphi}(u \,|\, a_0, z)}{S_\infty(u \,|\, a_0, z)} \right| + 4 \sup_{u \in [w,y]} \left| \frac{L_{n,k,\varphi}(u \,|\, a_0, z)}{S_{n,k}(u \,|\, a_0, z)} \right| + 2 \|\varphi(\cdot, z)\|_{v,[0,\bar{\tau}(z)]} \leq 10\kappa
\end{aligned}$$

and also, in view of part (c) of Lemma 1, that

$$\left| \int_W^Y \frac{\tilde{S}_\infty(u \,|\, a_0, Z)}{R_\infty(u \,|\, a_0, Z)} (\Theta_{n,k} - \Theta_\infty)(du) \right| \leq 3\kappa \sup_{u \in [W,Y]} \left| \frac{L_{n,k,\varphi}(u, a_0, Z)}{\tilde{S}_{n,k}(u \,|\, a_0, Z)} - \frac{L_{\infty,\varphi}(u, a_0, Z)}{\tilde{S}_\infty(u \,|\, a_0, Z)} \right|$$

holds P_0–almost surely. As a consequence, we find that

$$P_0 U_{5,n,k}^2 \leq E_0 \left[\frac{\bar{\gamma}_{n,k}(a_0, Z)}{\pi_{n,k}(Z)} \int_W^Y \frac{\tilde{S}_{n,k}(u \,|\, a_0, Z)}{R_{n,k}(u \,|\, a_0, Z)} \left(\Theta_{n,k} - \Theta_\infty\right)(du) \right]^2$$

$$\leq \kappa^4(10M_{3,n,k} + 3M_{2,n,k})^2$$

Before studying the remaining terms, we note that

$$\begin{aligned}
\mu_{n,k}(z) - \mu_\infty(z) \;&=\; \int \varphi(y,z)(F_{n,k} - F_0)(dy \,|\, a_0, z) - \int \varphi(y,z)(F_\infty - F_0)(dy \,|\, a_0, z) \\
&= -\left[\int \frac{L_{n,k,\varphi}(y, a_0, z)}{S_{n,k}(y \,|\, a_0, z)} S_0(y \,|\, a_0, z)(\Lambda_{n,k} - \Lambda_0)(dy \,|\, a_0, z)\right. \\
&\qquad \left. - \int \frac{L_{\infty,\varphi}(y, a_0, z)}{S_\infty(y \,|\, a_0, z)} S_0(y \,|\, a_0, z)(\Lambda_\infty - \Lambda_0)(dy \,|\, a_0, z)\right] \\
&= \int \left[\frac{L_{n,k,\varphi}(y, a_0, z)}{S_{n,k}(y \,|\, a_0, z)} - \frac{L_{\infty,\varphi}(y, a_0, z)}{S_\infty(y \,|\, a_0, z)}\right] S_0(y \,|\, a_0, z)\Lambda_0(dy \,|\, a_0, z) \\
&\qquad - \int S_0(y \,|\, a_0, z)\left[\frac{L_{n,k,\varphi}(y, a_0, z)}{S_{n,k}(y \,|\, a_0, z)}\Lambda_{n,k}(dy \,|\, a_0, z) - \frac{L_{\infty,\varphi}(y, a_0, z)}{S_\infty(y \,|\, a_0, z)}\Lambda_\infty(dy \,|\, a_0, z)\right],
\end{aligned}$$

and so, in view of part (c) of Lemma 2, we find that

$$|\mu_{n,k}(z) - \mu_\infty(z)| \;\leq\; 4 \sup_{y \in [0, \bar{\tau}(z)]} \left| \frac{L_{n,k,\varphi}(y, a_0, z)}{S_{n,k}(y \,|\, a_0, z)} - \frac{L_{\infty,\varphi}(y, a_0, z)}{S_\infty(y \,|\, a_0, z)} \right|.$$

Similarly as above, before proceeding, we note that $|\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(z, a, w, y, \delta)|$ is bounded above by $V_{1,\infty}(y, a, z) + V_{2,\infty}(y, w, a, z)$, where we write

$$\begin{aligned}
V_{1,\infty}(y, a, z) \;&:=\; \left|\frac{\gamma_{\infty,\natural}(y, a, z)}{R_\infty(y \,|\, a, z)}\right| \;=\; \left|\frac{\gamma_{\infty,\natural}(y, a, z)}{S_\infty(y \,|\, a, z)}\right| \left|\frac{S_\infty(y \,|\, a, z)}{R_\infty(y \,|\, a, z)}\right| \\
V_{2,\infty}(y, w, a, z) \;&:=\; \left|\int_w^y \frac{\gamma_{\infty,\natural}(u \,|\, a, z)}{R_\infty(u \,|\, a, z)} \Lambda_\infty(du \,|\, a, z)\right| \\
&=\; \left|\int_w^y \frac{\gamma_{\infty,\natural}(u \,|\, a, z)}{R_\infty(u \,|\, a, z)} S_\infty(u \,|\, a, z) \frac{\Lambda_\infty(du \,|\, a, z)}{S_\infty(u \,|\, a, z)}\right| \\
&=\; \left|\int_w^y \gamma_{\infty,\natural}(u \,|\, a, z) \frac{S_\infty(u \,|\, a, z)}{R_\infty(u \,|\, a, z)} \frac{1}{S_\infty}(du \,|\, a, z)\right| \\
&\leq\; \sup_{u \in [w,y]} \left|\frac{S_\infty(u \,|\, a, z)}{R_\infty(u \,|\, a, z)}\right| \int_w^y \gamma_{\infty,\natural}(u \,|\, a, z) \frac{1}{S_\infty}(du \,|\, a, z) \;,
\end{aligned}$$

and since we have that

$$\begin{aligned} 0 \leq \int_w^y \gamma_{\infty,\natural}(u \mid a,z)\frac{1}{S_\infty}(du \mid a,z) &= \frac{\gamma_{\infty,\natural}(u \mid a,z)}{S_\infty(u \mid a,z)}\bigg|_{u=w}^{y} - \int_w^y \frac{1}{S_\infty(u \mid a,z)}\gamma_{\infty,\natural}(du \mid a,z) \\ &= \frac{\gamma_{\infty,\natural}(u \mid a,z)}{S_\infty(u \mid a,z)}\bigg|_{u=w}^{y} + \int_w^y \left[\frac{1}{S_\infty(u \mid a,z)}\right]^2 G_\infty(du \mid a,z) \\ &\leq 2 \sup_{u\in[w,y]} \left|\frac{\gamma_{\infty,\natural}(u \mid a,z)}{S_\infty(u \mid a,z)}\right| + \left[\frac{1}{S_\infty(\overline{\tau}_W(z) \mid a,z)}\right]^2, \end{aligned}$$

it follows that $|\phi_{\mathrm{KM},\infty}(L_{\infty,\varphi})(Z,A,W,Y,\Delta)| \leq \kappa(2\kappa+\kappa^2) < \infty$ P_0–almost surely. Using this fact, we then can write that

$$\begin{aligned} P_0 U_{6,n,k}^2 &= \gamma_0^{-2} E_0 \left[[\mu_{n,k}(Z) - \mu_\infty(Z)] \left[\frac{1}{S_\infty(W \mid A,Z)} - \phi_{\mathrm{KM},\infty}(\gamma_{\infty,\natural})(Z,A,W,Y,\Delta)\right]\right]^2 \\ &\leq \left[\frac{\kappa+\kappa(2\kappa+\kappa^2)}{\gamma_0}\right]^2 E_0\left[\mu_{n,k}(Z) - \mu_\infty(Z)\right]^2 \leq 16\left[\frac{\kappa+\kappa(2\kappa+\kappa^2)}{\gamma_0}\right]^2 M_{2,n,k}^2 . \end{aligned}$$

Next, we have that

$$\begin{aligned} P_0 U_{7,n,k}^2 &= E_0\left[\mu_{n,k}(Z)\left|\frac{1}{\tilde{S}_{n,k}(W \mid A,Z)\gamma_{n,k}} - \frac{1}{\tilde{S}_\infty(W \mid A,Z)\gamma_\infty}\right|\right]^2 \\ &\leq E_0\left[\mu_{n,k}(Z)\left|\frac{1}{\tilde{S}_{n,k}(W \mid A,Z)}\left(\frac{1}{\gamma_{n,k}} - \frac{1}{\gamma_\infty}\right) + \frac{1}{\gamma_\infty}\left[\frac{1}{\tilde{S}_\infty(W \mid A,Z)} - \frac{1}{\tilde{S}_\infty(W \mid A,Z)}\right]\right|\right]^2 \\ &\leq \left[\left|\frac{1}{\gamma_{n,k}} - \frac{1}{\gamma_\infty}\right|\left[E_0\left|\frac{\mu_{n,k}(Z)}{\tilde{S}_{n,k}(W \mid A,Z)}\right|^2\right]^{\frac{1}{2}} + \left[E_0\left|\frac{\mu_{n,k}(Z)}{\gamma_\infty}\left[\frac{1}{\tilde{S}_\infty(W \mid A,Z)} - \frac{1}{\tilde{S}_\infty(W \mid A,Z)}\right]\right|^2\right]^{\frac{1}{2}}\right]^2 \\ &\leq \left[2\kappa^2\left|\frac{1}{\gamma_{n,k}} - \frac{1}{\gamma_\infty}\right| + 2\kappa\frac{1}{\gamma_\infty}M_{5,n,k}\right]^2 . \end{aligned}$$

Using a similar expansion, it can be shown that

$$P_0 U_{8,n,k}^2 \leq 2\kappa^2\left[2\kappa^2\left|\frac{1}{\gamma_{n,k}} - \frac{1}{\gamma_\infty}\right| + 2\kappa\frac{1}{\gamma_\infty}M_{4,n,k}\right]^2$$

We have that

$$P_0 U_{9,n,k}^2 \;=\; E_0\left[|\mu_{n,k}(Z)|\left|\frac{\bar{\gamma}_{n,k,\natural}(Y,A,Z)}{\tilde{S}_{n,k}(Y\,|\,A,Z)}\right|\left|\frac{\tilde{S}_{n,k}(Y\,|\,A,Z)}{R_{n,k}(Y\,|\,A,Z)}-\frac{\tilde{S}_{\infty}(Y\,|\,A,Z)}{R_{\infty}(Y\,|\,A,Z)}\right|\right]^2 \;\leq\; 4\kappa^4 M_{3,n,k}^2\;.$$

Next, we note that

$$\begin{aligned}
&\left|\int_w^y\left[\frac{\tilde{S}_{n,k}(u\,|\,a,z)}{R_{n,k}(u\,|\,a,z)}-\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}\right]\frac{\gamma_{\infty,\natural}(u\,|\,a,z)}{\gamma_{\infty}\tilde{S}_{\infty}(u\,|\,a,z)}\tilde{\Lambda}_{\infty}(du\,|\,a,z)\right| \\
&\leq\; \gamma_{\infty}^{-1}\sup_{u\in[w,y]}\left|\frac{\tilde{S}_{n,k}(u\,|\,a,z)}{R_{n,k}(u\,|\,a,z)}-\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}\right|\int_w^y\gamma_{\infty,\natural}(u\,|\,a,z)\frac{1}{S_{\infty}}(du\,|\,a,z) \\
&\leq\; \gamma_{\infty}^{-1}\sup_{u\in[w,y]}\left|\frac{\tilde{S}_{n,k}(u\,|\,a,z)}{R_{n,k}(u\,|\,a,z)}-\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}\right|\left[2\sup_{u\in[w,y]}\frac{\gamma_{\infty,\natural}(u\,|\,a,z)}{S_{\infty}(u\,|\,a,z)}+\left[\frac{1}{S_{\infty}(\overline{\tau}_W(z)\,|\,a,z)}\right]^2\right]
\end{aligned}$$

using the fact that

$$\begin{aligned}
0\;\leq\;\int_w^y\gamma_{\infty,\natural}(u\,|\,a,z)\frac{1}{S_{\infty}}(du\,|\,a,z) \;&=\; \frac{\gamma_{\infty,\natural}(u\,|\,a,z)}{S_{\infty}(u\,|\,a,z)}\bigg|_{u=w}^{y}-\int_w^y\left[\frac{1}{S_{\infty}(u\,|\,a,z)}\right]^2 G_{\infty}(du\,|\,a,z) \\
&\leq\; 2\sup_{u\in[w,y]}\frac{\gamma_{\infty,\natural}(u\,|\,a,z)}{S_{\infty}(u\,|\,a,z)}+\left[\frac{1}{S_{\infty}(\overline{\tau}_W(z)\,|\,a,z)}\right]^2.
\end{aligned}$$

Thus, using that $|\mu_{\infty}(z)|\leq\|u\mapsto\varphi(u,z)\|_{\infty}$, we find that

$$P_0 U_{10,n,k}^2\;\leq\;\left[\frac{\|(u,z)\mapsto\varphi(u,z)\|_{\infty}(2\kappa+\kappa^2)}{\gamma_{\infty}}\right]^2 M_{3,n,k}^2\;.$$

To study the final term, similarly as in the study of $P_0 U_{5,n,k}^2$, we first note that

$$\begin{aligned}
&\left|\int_w^y\frac{\tilde{S}_{n,k}(u\,|\,a,z)}{R_{n,k}(u\,|\,a,z)}(\Theta_{n,k}-\Theta_{\infty})(du)\right| \\
&\leq\;\left|\int_w^y\left[\frac{\tilde{S}_{n,k}(u\,|\,a_0,z)}{R_{n,k}(u\,|\,a,z)}-\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}\right](\Theta_{n,k}-\Theta_{\infty})(du)\right|+\left|\int_w^y\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}(\Theta_{n,k}-\Theta_{\infty})(du)\right| \\
&\leq\;\sup_{u\in[w,y]}\left|\frac{\tilde{S}_{n,k}(u\,|\,a,z)}{R_{n,k}(u\,|\,a,z)}-\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}\right|\int_w^y|(\Theta_{n,k}-\Theta_{\infty})(du)|+\left|\int_w^y\frac{\tilde{S}_{\infty}(u\,|\,a,z)}{R_{\infty}(u\,|\,a,z)}(\Theta_{n,k}-\Theta_{\infty})(du)\right|,
\end{aligned}$$

where we have now redefined $\Theta_{n,k}$ and Θ_∞ via the differentials

$$\Theta_{n,k}(du) := \frac{\gamma_{n,k,\natural}(u,a,z)}{\tilde{S}_{n,k}(u \,|\, a,z)} \tilde{\Lambda}_{n,k}(du \,|\, a,z) \;\text{ and }\; \Theta_\infty(du) := \frac{\gamma_{\infty,\natural}(u,a,z)}{\tilde{S}_\infty(u \,|\, a,z)} \tilde{\Lambda}_\infty(du \,|\, a,Z) \;.$$

Using similar arguments as used above, we observe that

$$\begin{aligned}
&\int_w^y |(\Theta_{n,k} - \Theta_\infty)(du)| \;\leq\; \int_w^y |\Theta_{n,k}(du)| + \int_w^y |\Theta_\infty(du)| \\
&= \int_w^y \frac{\gamma_{n,k,\natural}(u,a,z)}{\tilde{S}_{n,k}(u \,|\, a,z)} \tilde{\Lambda}_{n,k}(du \,|\, a,z) + \int_w^y \frac{\gamma_{\infty,\natural}(u,a,z)}{\tilde{S}_\infty(u \,|\, a,z)} \tilde{\Lambda}_\infty(du \,|\, a,z) \\
&\leq\; 2 \sup_{u \in [w,y]} \left| \frac{\gamma_{n,k,\natural}(u \,|\, a,z)}{S_{n,k}(u \,|\, a,z)} \right| + 2 \sup_{u \in [w,y]} \left| \frac{\gamma_{\infty,\natural}(u \,|\, a,z)}{S_\infty(u \,|\, a,z)} \right| + \left[\frac{1}{S_\infty(\overline{\tau}_W(z) \,|\, a,z)} \right]^2 + \left[\frac{1}{S_{n,k}(\overline{\tau}_W(z) \,|\, a,z)} \right]^2 \\
&\leq\; 4\kappa + 2\kappa^2
\end{aligned}$$

and also, in view of part (d) of Lemma 2, that $\left| \int_W^Y \frac{\tilde{S}_\infty(u \,|\, A,Z)}{R_\infty(u \,|\, A,Z)} (\Theta_{n,k} - \Theta_\infty)(du \,|\, A,Z) \right|$ is P_0–almost surely bounded above by

$$\begin{aligned}
3\kappa \sup_{u \in [W,Y]} \left| \frac{\gamma_{n,k,\natural}(u \,|\, A,Z)}{S_{n,k}(u \,|\, A,Z)} - \frac{\gamma_{\infty,\natural}(u \,|\, A,Z)}{S_\infty(u \,|\, A,Z)} \right| + 3\kappa^4 \sup_{u \in [W,Y]} |G_{n,k}(u \,|\, A,Z) - G_\infty(u \,|\, A,Z)| \\
+ 2\kappa^4 \sup_{u \in [W,Y]} \left| \frac{1}{S_{n,k}(u \,|\, A,Z)} - \frac{1}{S_\infty(u \,|\, A,Z)} \right| .
\end{aligned}$$

As a consequence, we find that

$$\begin{aligned}
P_0 U_{11,n,k}^2 \;&\leq\; E_0 \left[\mu_{n,k}(Z) \int_W^Y \frac{\tilde{S}_{n,k}(u \,|\, A,Z)}{R_{n,k}(u \,|\, A,Z)} \left[\Theta_{n,k}(du \,|\, A,Z) - \Theta_\infty(du \,|\, A,Z) \right] \right]^2 \\
&\leq\; 4\kappa^2 \left[(4\kappa + 2\kappa^2) M_{3,n,k} + 3\kappa M_{4,n,k} + 3\kappa^4 M_{6,n,k} + 2\kappa^4 M_{5,n,k} \right]^2 .
\end{aligned}$$

In view of all the inequalities derived, we see that if $M_{1,n,k}, M_{2,n,k}, \ldots, M_{6,n,k}$ tend to zero in probability, then so does each $P_0 U_{j,n,k}^2$ for $j = 1, 2, \ldots, 11$ and thus $P_0(\phi_{n,k} - \phi_\infty)^2$ itself tends to zero in probability, as required to complete the proof.

We prove criteria (i), we do so by using an showing that $P_0 \phi_{\eta_\infty, \psi_0} = 0$ under condition

(C3a). We can write each of the equations

$$\begin{aligned} F_{1,\infty}(u \,|\, a, z) &:= P_\infty(Y \leq u, \Delta = 1 \,|\, A = a, Z = z) \\ &= \frac{\int I_{[0,u]}(t) \left\{ \int I_{[0,t]}(w) Q_\infty(t \,|\, w, a, z) G_{X,0}(dw \,|\, a, z) \right\} F_{X,0}(dt \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} \,, \end{aligned}$$

and

$$\begin{aligned} R_\infty(u \,|\, a, z) &:= P_\infty(W \leq u \leq Y \,|\, A = a, Z = z) \\ &= \frac{\int I_{[0,u]}(w) Q_\infty(u \,|\, w, a, z) G_{X,0}(dw \,|\, a, z)}{\int G_{X,0}(t \,|\, a, z) F_{X,0}(dt \,|\, a, z)} S_{X,0}(u \,|\, a, z) \,, \end{aligned}$$

and therefore observe that $F_{1,\infty}(du \,|\, a, z)/R_\infty(u \,|\, a, z) = F_{X,0}(du \,|\, a, z)/S_{X,0}(u \,|\, a, z)$. We observe further that $\gamma_{\infty,\natural} = \gamma_{0,\natural}$ under condition (C3a), and then observe that

$$\begin{aligned} \mathbb{E}_0 \left[\phi_{KM,\eta_\infty}(L_{\eta_\infty,\varphi})(O_i) \,|\, A = a, Z = z \right] &= \int \frac{L_{\eta_\infty,\varphi}(u, a, z) R_0(u \,|\, a, z)}{R_\infty(u \,|\, a, z)} \left\{ \frac{F_{1,\infty}(du \,|\, a, z)}{R_\infty(u \,|\, a, z)} - \frac{F_{1,0}(du \,|\, a, z)}{R_0(u \,|\, a, z)} \right\} \\ &= 0 \,, \end{aligned}$$

$$\begin{aligned} \mathbb{E}_0 \left[\phi_{KM,\eta_\infty}(\gamma_{\eta_\infty,\natural})(O_i) \,|\, A = a, Z = z \right] &= \int \frac{\gamma_{\eta_\infty,\natural}(u, a, z) R_0(u \,|\, a, z)}{R_\infty(u \,|\, a, z)} \left\{ \frac{F_{1,\infty}(du \,|\, a, z)}{R_\infty(u \,|\, a, z)} - \frac{F_{1,0}(du \,|\, a, z)}{R_0(u \,|\, a, z)} \right\} \\ &= 0 \,. \end{aligned}$$

We next note that under condition (C3a) that $\mu_\infty = \mu_0$ and therefore

$$\begin{aligned} P_0 \phi_{\eta_\infty,\psi_0} &= \gamma_\infty^{-1} \mathbb{E}_0 \left[(\mu_\infty(Z) - \psi_0) \gamma_0(A, Z) \right] \\ &= \frac{\gamma_0}{\gamma_\infty} \mathbb{E}_0 \left[(\mu_\infty(Z) - \psi_0) \frac{\gamma_0(A, Z)}{\gamma_0} \right] = 0 \,, \end{aligned}$$

and therefore completing the proof of (i).

Criteria (ii) follows from evaluating the remainder term $R(P_\infty, P_0)$ directly given condi-

tions (C3a) and (C3b). It is evident that each of $R_1(P_\infty, P_0), R_2(P_\infty, P_0)$, and $R_3(P_\infty, P_0)$ are 0 when $(C3a)$ holds. Because $\gamma_\infty = \gamma_0$ when both conditions hold then $R_4(P_\infty, P_0) = 0$. Thus we have proved criteria (ii).

The equivalence of the approaches is established by observing that when the plug-in estimator $\psi_{\eta_{k,n}}$ is consistent we can write

$$\begin{aligned}\psi^*_{k,n} :=& \psi_{\eta_{k,n}} + n_k^{-1} \sum_{i \in \mathcal{V}_k} \left[\phi_{1,\eta_{k,n}}(O_i) + \gamma_{k,n}^{-1}(\mu_{k,n}(Z_i) - \psi_{\eta_{k,n}})(S_{k,n}(W_i \mid A_i, Z_i)^{-1} + \phi_{KM,\eta_{k,n}}(\gamma_{k,n,\natural}(O_i))\right] \\ =& \psi_{\eta_{k,n}} + \gamma_{k,n}^{-1} n_k^{-1} \sum_{i \in \mathcal{V}_k} \left[(\psi^{**}_{k,n} - \psi_{\eta_{k,n}})(S_{k,n}(W_i \mid A_i, Z_i)^{-1} + \phi_{KM,\eta_{k,n}}(\gamma_{k,n,\natural})(O_i)\right] ,\end{aligned}$$

which means that

$$\psi^*_n - \psi^{**}_n := \frac{1}{K} \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i \in \mathcal{V}_k} \left[\psi^{**}_{k,n} - \psi_{\eta_{k,n}}\right] \left[\gamma_{k,n}^{-1} S_{k,n}(W_i \mid A_i, Z_i)^{-1} + \gamma_{k,n}^{-1} \phi_{KM,\eta_{k,n}}(\gamma_{k,n,\natural})(O_i) - 1\right] ,$$

and Conditions (C1)–(C4) are sufficient to establish that $\frac{1}{n_k} \sum_{i \in \mathcal{V}_k} \left\{\gamma_{k,n}^{-1} S_{k,n}(W_i \mid A_i, Z_i)^{-1} - 1\right\} = O_P(n_k^{-1/2})$ and $\frac{1}{n_k} \sum_{i \in \mathcal{V}_k} \phi_{KM,\eta_{k,n}}(\gamma_{k,n,\natural})(O_i) - = O_P(n_k^{-1/2})$, which then establishes that $\psi^*_n - \psi^{**}_n = o_P(n^{-1/2})$ and the asymptotic equivalence of both estimators.

Proof of Theorem 3

We first show that the stochastic processes $\mathbb{B}_n := \{\sqrt{n}\,[\psi_n(s) - \psi_0(s)] : s \in (0, \tau)\}$ and $\bar{\mathbb{B}}_n := \{n^{-1/2} \sum_i \phi_{s,0}(O_i) : s \in (0, \tau)\}$ are asymptotically equivalent. Under Condition (D2), $\bar{\mathbb{B}}_n$ converges weakly to $\mathbb{B}_0$ relative to the supremum norm. The limiting distribution $\mathbb{B}_0$ is a mean-zero Gaussian process with covariance function $\sigma_0^2(s,t) := \mathbb{E}_0\{\phi_{s,0}(O)\phi_{t,0}(O)\}$, in view of Theorem 19.3 (Van der Vaart, 2000). Let g be any function bounded by M_0, and Lipschitz with constant L_0. We know that $\mathbb{E}_0\left\{g(\bar{\mathbb{B}}_n)\right\}$ converges to $\mathbb{E}_0\{g(\mathbb{B}_0)\}|$ from the convergence

of $\bar{\mathbb{B}}_n$ to $\mathbb{B}_0$, therefore the inequality

$$|\mathbb{E}_0\left\{g(\mathbb{B}_n)\right\} - \mathbb{E}_0\left\{g(\mathbb{B}_0)\right\}| \leq |\mathbb{E}_0\left\{g(\mathbb{B}_n) - g(\bar{\mathbb{B}}_n)\right\}| + |\mathbb{E}_0\left\{g(\bar{\mathbb{B}}_n)\right\} - \mathbb{E}_0\left\{g(\mathbb{B}_0)\right\}|$$

will establish uniform convergence if we can show that $|\mathbb{E}_0\left\{g(\mathbb{B}_n) - g(\bar{\mathbb{B}}_n)\right\}|$ tends to zero. We note that condition (D4)) and the definition of the function g ensures that

$$|\mathbb{E}_0\left\{g(\mathbb{B}_n) - g(\bar{\mathbb{B}}_n)\right\}| \leq \mathbb{E}_0\left\{\min(L_0 n^{1/2}\|r_n\|_\infty, 2M_0)\right\}$$

and it straightforward to establish that $\mathbb{E}_0\left\{\min(L_0 n^{1/2}\|r_n\|_\infty, 2M_0)\right\} \to 0$. Because $\mathbb{E}_0\left\{g(\mathbb{B}_n)\right\}$ converges in expectation to $\mathbb{E}_0\left\{g(\mathbb{B}_0)\right\}$ for any bounded and Lipschitz function g, we conclude that $\mathbb{B}_n$ converges weakly to $\mathbb{B}_0$ relative to the supremum norm.

Following this result, Condition (D1) is a modification of Condition (C1) that ensure the upper bound holds uniformly. We define $\mathbb{B}_n^* ::= \{\sqrt{n}\frac{\psi_n(s)-\psi_0(s)}{\sigma_0(s)} : s \in (0,\tau)\}$ and seek to accomplish the task of establishing that $\mathbb{B}_n^*$ converges weakly to a mean-zero standard process ($\mathbb{B}_0^*$) relative to the supremum norm. We define $\sigma_L := \inf_{s\in(0,\tau)} \sigma_0(s)$, which is strictly positive. By the definition of a Lipschitz function, the following bound holds

$$|g(\mathbb{B}_n^*) - g(\mathbb{B}_0^*)| \leq \frac{L_0}{\sigma_L}\|\mathbb{B}_n - \mathbb{B}_0\|_\infty$$

which implies that $\mathbb{B}_n^*$ converges to a standardized Gaussian process defined by $\{\mathbb{B}_0(s)/\sigma_0(s) : s \in (0,\tau)\}$ when $\mathbb{B}_n$ converges to $\mathbb{B}_0$. The above result therefore implies that $\mathbb{B}_n^*$ converges weakly to a standard mean-zero Gaussian process relative to the supremum norm.

Proof of Theorem 4

We first observe that the distribution function is linear $I(t \leq s) \mapsto \int I(t \leq s)\mathbb{F}_0(dt)$ and therefore has Gateaux derivative $\dot{\phi}_{\mathbb{F},0}(o) := I(t \leq s) \mapsto \phi_{s,0}(o)$, where $\phi_{s,0}(o)$ defines the influence function for $\mathbb{F}_0(s)$ pointwise (presented in Example 1). We use Theorem 20.8 of Van der Vaart (2000), Theorem 2 of this article, and the fact that Gateaux derivatives

are linear in their second argument to show

$$\begin{aligned}\Phi(\mathbb{F}_n) - \Phi(\mathbb{F}_0) &= \dot{\Phi}(\mathbb{F}_0; \mathbb{F}_n - \mathbb{F}_0) + o_P\left(n^{-1/2}\right) \\ &= \dot{\Phi}\left(\mathbb{F}_0; \frac{1}{n}\sum_{i=1}^{n} \dot{\phi}_{\mathbb{F},0}(o) + R_s(\eta_n, \eta_0)\right) + o_P\left(n^{-1/2}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \dot{\Phi}\left(\mathbb{F}_0; \dot{\phi}_{\mathbb{F},0}(o)\right) + o_P\left(n^{-1/2}\right) ,\end{aligned}$$

where the final argument follows from the condition that $\sup_s R_s(\eta_n, \eta_0) \implies \dot{\Phi}(\mathbb{F}_0; R_s(\eta_n, \eta_0)) = o_P\left(n^{-1/2}\right)$, which completes the proof.

A.3 Part C: reparametrization

We establish the identification, variation independence and reparametrization of the observed data distribution P_0 in terms of the chosen nuisance parameters $\eta(P) := (F_{X,0}, Q_0, G_0, J_0)$. We first write

$$P_0(do) = P_0(dy, d\delta \,|\, w, a, z) G_0(dw \,|\, a, z) J_0(da, dz) ,$$

which leaves the conditional distribution $P_0(dy, d\delta \,|\, w, a, z)$ left to identify in terms of the distributions $T \,|\, A, Z$ and $C \,|\, W, A, Z, T \geq W$. This is done by noting the following equality

$$\begin{aligned}P_0(Y \leq y, \Delta = 1 \,|\, w, a, z) &= \int P_0(T \leq c, T \leq y \,|\, C = c, W = w, A = a, Z = z) P_0(dc \,|\, w, a, z) \\ (A1) &= \int P_0(T \leq c \wedge y \,|\, W = w, A = a, Z = z) P_0(dc \,|\, w, a, z) \\ (A2) &= \int \frac{P_{X,0}(w \leq T \leq c \wedge y \,|\, A = a, Z = z)}{S_{X,0}(w \,|\, a, z)} P_0(dc \,|\, w, a, z) ,\end{aligned}$$

which gives the desired result for $\delta = 1$. Similarly for $\delta = 0$, we have

$$P_0(Y \leq y, \Delta = 0 \,|\, w, a, z) = \int P_0(C \leq t, C \leq y \,|\, W = w, A = a, Z = z) F_0(dt \,|\, w, a, z)$$

$$= \int [1 - Q_0(c \wedge y \,|\, w, a, z)] \frac{F_{X,0}(dt \,|\, a, z)}{S_{X,0}(w \,|\, a, z)}$$

Using the results from identification we can write the observed at risk probability as

$$\tilde{R}_0(u \,|\, a, z) := S_{X,0}(u \,|\, a, z) \int_0^u Q_0(u \,|\, w, a, z) \frac{G_0(dw \,|\, a, z)}{S_{X,0}(w \,|\, a, z)} \,,$$

which together with the previous results gives a representation of the parameter and influence function as functionals of $\eta(P)$.

A.4 Part D: example of functionals

Corollary 1

The corollary 1 follows directly from Theorem 4, but we include the details as an example application of the result. We first set $P_\epsilon = P + \epsilon h$ and note that $\Phi(\Psi(P_\epsilon; P_\epsilon) = 0$ for each ϵ and that $\Phi(\Psi(P); P) = \Phi(\Psi(P); \mathbb{F})$, such that $\mathbb{F}$ is the distribution function that agrees with P. When the parameter $\Psi(P_\epsilon)$ is differentiable then,

$$\begin{aligned} 0 = & \left. \frac{d}{d\epsilon} \Phi(\Psi(P_\epsilon); P_\epsilon) \right|_{\epsilon=0} \\ = & \left. \frac{d}{d\epsilon} \Phi(\Psi(P_\epsilon); P) \right|_{\epsilon=0} + \left. \frac{d}{d\epsilon} \Phi(\Psi(P); P_\epsilon) \right|_{\epsilon=0} \\ = & \left. \frac{d}{d\epsilon} \Phi(\epsilon; P) \right|_{\epsilon=\Psi(P)} \cdot \left. \frac{d}{d\epsilon} \Psi(P_\epsilon) \right|_{\epsilon=0} + \dot{\Phi}(\mathbb{F}; h) \end{aligned}$$

and we can solve for the derivative of interest. The existence of a solution and regularity conditions under which differentiablity and inversion is possible are omitted but provided for specific examples. Rearranging the terms

$$\left. \frac{d}{d\epsilon} \Psi(P_\epsilon) \right|_{\epsilon=0} = \dot{\Psi}(P; h) = - \left[\left. \frac{d}{d\epsilon} \Phi(\epsilon; P) \right|_{\epsilon=\Psi(P)} \right]^{-1} \dot{\Phi}(\mathbb{F}; h) \,,$$

which completes the proof.

Example 1

The derivation follows from Theorem 1 and evaluating the function $L_{0,\mathbb{S}}(y, a, z)$ and $L_{0,\mathbb{F}}(y, a, z)$ for a fixed τ. as in the example. It is straight forward to show that

$$\begin{aligned} L_{0,\mathbb{S}}(y, a, z) &= I(y \leq \tau) S_{X,0}(\tau \,|\, a, z) \\ L_{0,\mathbb{F}}(y, a, z) &= -I(y \leq \tau) S_{X,0}(\tau \,|\, a, z)\,, \end{aligned}$$

which is sufficient for the result of example 1 and provides the details for the result of Theorem 4.

Example 2

Following the previous example, we evaluate $L_{0,\varphi}(y, a, z)$ where $\varphi(y, z) := [I(y \geq \tau) - b(z)]^2$. We first note that $\varphi(y, z) = I(y \geq \tau) - 2I(y \geq \tau)b(z) - b(z)^2$, therefore it is straight forward to use the result from example 1 that

$$L_{0,\varphi}(y, a, z) = I(y \leq \tau) S_{X,0}(\tau \,|\, a, z) \left\{1 - 2b(z)\right\}\,,$$

which we substitute into the result of Theorem 1 and this completes the result.

Example 3

The derivation follows from Theorem 1 and using Example 1.

Example 4

The linear term $\phi_{U,0}(o)$ is straightforward to evaluate, since following examples 1 and 2 the function $L_{U,0}(z, a, y) = S_{X,0}(y \,|\, a, z) I(y \leq m_0)$. Further, we observe that $\mu_0(z) := \int I(y \leq m_0) - 0.5 F_{X,0}(dy \,|\, a_0, z) = F_{X,0}(m_0 \,|\, a_0, z) - 0.5$, which we can rearrange to produce the result

in the article. Finally, the derivative

$$\begin{aligned}\frac{d}{d\epsilon}\Phi(\epsilon;P_0)\bigg|_{\epsilon=m_0} &= \frac{d}{d\epsilon}\int I(y\leq\epsilon)F_{X,0}(dy\,|\,a_0,z)\tilde{H}_0(dz)\bigg|_{\epsilon=m_0}\\ &= \int f_{X,0}(\epsilon\,|\,a_0,z)\tilde{H}_0(dz)\bigg|_{\epsilon=m_0}\end{aligned}$$

and this completes the result. If we set $w=0$ with probability 1, and keep the same notation and assumptions then the result simplifies to

$$\phi_{m,0}(o_i) := -\left[\int f_{X,0}(m_0\,|\,a_0,z)H_0(dz)\right]^{-1}\left\{\frac{I(a=a_0)}{\pi_0(a_0\,|\,z)}\phi^*_{KM}(L_{U,0})(o)+0.5-S_{X,0}(m_0\,|\,a_0,z)\right\},$$

such that $\phi^*_{KM}(L_{U,0})(o) := S_{X,0}(m_0\,|\,a_0,z)\left\{\frac{\delta I(y\leq m_0)}{S_{X,0}(y\,|\,a,z)Q_0(y\,|\,a,z)}-\int_0^{y\wedge m_0}\frac{\Lambda_{X,0}(du\,|\,a,z)}{S_{X,0}(u\,|\,a,z)Q_0(u\,|\,a,z)}\right\}$, a useful for application in clinical trials where truncation is less prominent.

Example 5

The derivation of the influence function of $\mathbb{L}_{a_0}:\mathbb{F}_{a_0}(t)\mapsto\frac{1}{\tau_2-\tau_1}\int_{\tau_1}^{\tau_2}\log\{-\log[1-\mathbb{F}_{a_0}(t)]\}$ follows from Theorem 4 and the Gateaux derivative. The Gateaux derivative in direction $\mathbb{F}_n-\mathbb{F}_0$

$$\begin{aligned}\dot{\mathbb{L}}_{a,0}(\mathbb{F}_0;\mathbb{F}_n-\mathbb{F}_0) &= -\int\frac{1}{\{1-\mathbb{F}_0(t)\}\log\{1-\mathbb{F}_0(t)\}}(\mathbb{F}_n-\mathbb{F}_0)(dt)\\ &= -\frac{1}{n}\sum_{i=1}^n\int\frac{\phi_{\mathbb{F},0}(o_i;t)}{\mathbb{S}_0(t)\log\mathbb{S}_0(t)}dt\,,\end{aligned}$$

where $\phi_{\mathbb{F},0}(o_i;t)$ is the pointwise influence function of the distribution $\mathbb{F}_0(t)$. The details of Example 1 complete the proof. The choice to define the parameter over a standardized interval (τ_1,τ_2) is for identification, the topic is also discussed in (Vansteelandt et al., 2022).

Appendix B

CHAPTER TWO SUPPLEMENTARY MATERIALS AND APPENDIX

B.1 Supplementary materials

We present additional information about the ACT study cohort in this section. Figure B.1 contains estimated marginal survival probability curves for each group. Tables B.2 and B.3 denote summaries of the relevant confounding variables and outcomes included in the analysis. Table B.1 contains the results of the analysis comparing the APOE-e4 positive group to the APOE-e4 negative group. In every analysis, both the NTDE and ATE estaimted an increase in CERAD scores among the group with a positive APOE-4 gene expression.

APOE-4	Adjusting for consent	Not adjusting
Mean, (95% CI)		
NTDE estimate on CERAD	0.51 (0.11-0.92)	0.41 (0.27-0.55)
ATE estimate on CERAD	0.41 (0.25-0.57)	0.40 (0.24-0.56)

Table B.1: Estimates of the natural time direct effect and the autopsy average treatment effect of APOE-4 on CERAD scores measured at death.

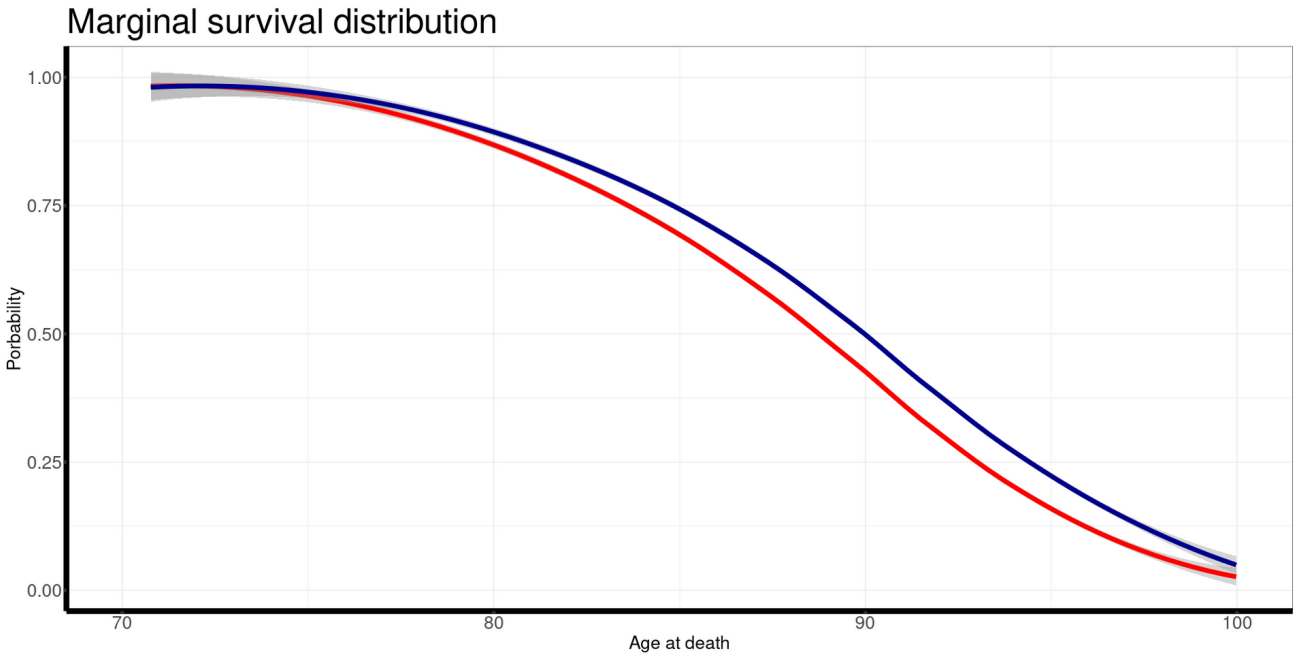

Figure B.1: Red indicates the APOE-4 positive group, while Blue denotes the APOE-4 negative group. Solid lines denote the estimated marginal survival probability within each group, estimated using using survival splines implemented on the hazard scale.

	- APOE e4 (N=3418)	+ APOE e4 (N=2291)	Overall (N=5709)
Braak score			
Mean (SD)	3.50 (1.62)	4.05 (1.69)	3.67 (1.66)
Median [Min, Max]	4.00 [0, 6.00]	4.00 [0, 6.00]	4.00 [0, 6.00]
Missing	2829 (82.8%)	2036 (88.9%)	4865 (85.2%)
CERAD score			
Mean (SD)	1.44 (1.13)	1.91 (1.05)	1.58 (1.13)
Median [Min, Max]	1.00 [0, 3.00]	2.00 [0, 3.00]	2.00 [0, 3.00]
Missing	2827 (82.7%)	2035 (88.8%)	4862 (85.2%)
Autopsy consent			
No	2779 (81.3%)	2009 (87.7%)	4788 (83.9%)
Yes	639 (18.7%)	282 (12.3%)	921 (16.1%)
Sex			
Female	1918 (56.1%)	1396 (60.9%)	3314 (58.0%)
Male	1500 (43.9%)	895 (39.1%)	2395 (42.0%)
Race			
Non-White	324 (9.5%)	281 (12.3%)	605 (10.6%)
White	3090 (90.4%)	2006 (87.6%)	5096 (89.3%)
Missing	4 (0.1%)	4 (0.2%)	8 (0.1%)
BMI			
Normal	1070 (31.3%)	765 (33.4%)	1835 (32.1%)
Obese	856 (25.0%)	587 (25.6%)	1443 (25.3%)
Overweight	1388 (40.6%)	838 (36.6%)	2226 (39.0%)
Underweight	28 (0.8%)	28 (1.2%)	56 (1.0%)
Missing	76 (2.2%)	73 (3.2%)	149 (2.6%)
Education			
< 13	1012 (29.6%)	545 (23.8%)	1557 (27.3%)
13-16	1402 (41.0%)	917 (40.0%)	2319 (40.6%)
17+	1003 (29.3%)	829 (36.2%)	1832 (32.1%)
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
History of diabetes			
No	2867 (83.9%)	2009 (87.7%)	4876 (85.4%)
Yes	551 (16.1%)	282 (12.3%)	833 (14.6%)
History of heart disease			
No	2490 (72.8%)	1760 (76.8%)	4250 (74.4%)
Yes	928 (27.2%)	531 (23.2%)	1459 (25.6%)

Table B.2: Summary characteristics of the ACT study cohort participant population. Groups are stratified based on the presence of a positive APOE-e4 gene expression.

	- APOE e4 (N=2212)	+ APOE e4 (N=1172)	Overall (N=3384)
Braak score			
Mean (SD)	3.50 (1.62)	4.05 (1.69)	3.67 (1.66)
Median [Min, Max]	4.00 [0, 6.00]	4.00 [0, 6.00]	4.00 [0, 6.00]
Missing	1623 (73.4%)	917 (78.2%)	2540 (75.1%)
CERAD score			
Mean (SD)	1.44 (1.13)	1.91 (1.05)	1.58 (1.13)
Median [Min, Max]	1.00 [0, 3.00]	2.00 [0, 3.00]	2.00 [0, 3.00]
Missing	1621 (73.3%)	916 (78.2%)	2537 (75.0%)
Autopsy consent			
No	1573 (71.1%)	890 (75.9%)	2463 (72.8%)
Yes	639 (28.9%)	282 (24.1%)	921 (27.2%)
Sex			
Female	1229 (55.6%)	706 (60.2%)	1935 (57.2%)
Male	983 (44.4%)	466 (39.8%)	1449 (42.8%)
Race			
Non-White	181 (8.2%)	138 (11.8%)	319 (9.4%)
White	2030 (91.8%)	1034 (88.2%)	3064 (90.5%)
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
BMI			
Normal	699 (31.6%)	387 (33.0%)	1086 (32.1%)
Obese	533 (24.1%)	303 (25.9%)	836 (24.7%)
Overweight	903 (40.8%)	423 (36.1%)	1326 (39.2%)
Underweight	25 (1.1%)	19 (1.6%)	44 (1.3%)
Missing	52 (2.4%)	40 (3.4%)	92 (2.7%)
Education			
< 13	835 (37.7%)	407 (34.7%)	1242 (36.7%)
13-16	937 (42.4%)	498 (42.5%)	1435 (42.4%)
17+	439 (19.8%)	267 (22.8%)	706 (20.9%)
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
History of diabetes			
No	1809 (81.8%)	976 (83.3%)	2785 (82.3%)
Yes	403 (18.2%)	196 (16.7%)	599 (17.7%)
History of heart disease			
No	1478 (66.8%)	782 (66.7%)	2260 (66.8%)
Yes	734 (33.2%)	390 (33.3%)	1124 (33.2%)

Table B.3: Summary characteristics of the individuals that have died among the ACT study cohort. Groups are stratified based on the presence of a positive APOE-e4 gene expression.

B.2 Appendix

Part A: Ideal data setting

Details of identification result

We denote by $F_{Z,0}(t \mid \ell)$ the distribution function of the counterfactual random variable $T(1) \mid L$, and by $F_{Z,0}(t \mid a, \ell))$ the observed conditional distribution function of $T \mid A = 1, L = \ell$. Further, we denote by $H_{Z,0}(\ell) : P_{Z,0}(L \leq \ell)$ the distribution of the covariates L.

Lemma 3. *When conditions (A0) – (A4) hold, the natural time direct effect θ_0 is identified by ψ_0 define in Equation 2.1.*

Proof. Condition (A0) specifies that the observed autopsy outcome is $B = AB_{T(1)}(1) + (1 - A)B_{T(0)}0$ and the observed survival time is $T = AT(1) + (1 - A)T(0)$, then

$$\begin{aligned} E_{\mathbb{X},0}[B_{T(1)}(1) - B_{T(1)}(0)] &= E_{\mathbb{X},0}\left\{E_{\mathbb{X},0}\left[B_{T(1)}(1) - B_{T(1)}(0) \mid T(1), L\right]\right\} \\ \text{using (A1)- (A3)} \;\; &= \iint E_{\mathbb{X},0}\left[B_t(1) - B_t(0) \mid T(1) = t, L = \ell\right] F_{Z,0}(dt \mid \ell) H_{Z,0}(d\ell) \end{aligned}$$

Under conditions (A0)- (A3), we can write $F_{Z,0}(t \mid \ell) = F_{Z,0}(t \mid 1, \ell)$, then by condition i.

$$E_{\mathbb{X},0}\left[B_t(1) \mid T(1) = t, L = \ell\right] = E_{Z,0}\left(B \mid T = t, A = 1, L = \ell\right),$$

and when conditions (A3)- (A4) hold

$$\begin{aligned} E_{\mathbb{X},0}\left[B_t(0) \mid T(1) = t, L = \ell\right] &= E_{\mathbb{X},0}\left[B_t(0) \mid T(1) = t, T(0) \geq t, L = \ell\right] \\ &= E_{\mathbb{X},0}\left[B_t(0) \mid T(0) = t, T(0) \geq t, L = \ell\right] \\ &= E_{Z,0}\left(B \mid T = t, A = 0, L = \ell\right), \end{aligned}$$

combining the terms completes the proof. □

Part B: Prospective data setting

Details of identification result

The strategy used for identification of Theorem 1 incorporates strategies from the causal inference mediation literature (Robins, 1986; VanderWeele, 2015a) along with survival analysis (Beran, 1981; Reid, 1981b; Dabrowska, 1987). Before we establish the identification of ψ_0, we show that the components of the ideal data distribution $P_{Z,0}$ can be identified in terms of the observed data distribution P_0. We expand the conditions presented in the paper to account for left truncation of the survival times. We use the same notation as in the retrospective data setting, where W denotes the left-truncation times and P_0 denotes the data distribution implied by $P_{Z,0}$ in which the observed data satisfy $T \geq W$.

We first define for a generic random variable U the lower and upper support bounds

$$\begin{aligned} \underline{\tau}_U(a,\ell) &:= \sup\{u : P_{Z,0}\left(U \geq u \,|\, A=a, L=\ell\right) = 1\}\,; \\ \overline{\tau}_U(a,\ell) &:= \sup\{u : P_{Z,0}\left(U \geq u \,|\, A=a, L=\ell\right) > 0\}\,. \end{aligned}$$

We also denote by $\overline{\tau}_C(a,\ell) := \sup\{u : P_{Z,0}\left(C \geq u \,|\, A=a, L=\ell, T \geq W\right) > 0\}$ the upper bound for the support of the censoring distribution, and by $\pi_{Z,0}(a_0 \,|\, \ell) := P_{Z,0}(A=a_0 \,|\, L=\ell)$ the propensity score for each $\ell \in \mathcal{L}$. We make the following support recovery conditions and independence assumptions for identifiability:

(B1) for $P_{Z,0}-$almost every value $\ell \in \mathcal{L}$ and $a \in \{0,1\}$, it holds that:

- (i) $\underline{\tau}_W(a,\ell) \leq \underline{\tau}_T(a,\ell)\,;$
- (ii) $\overline{\tau}_W(a,\ell) + \alpha \leq \min\{\overline{\tau}_T(a,\ell), \overline{\tau}_C(a,\ell)\}$ for some $\alpha > 0\,;$
- (iii) $\varphi(t,\ell)$ is constnant for each $t \geq \overline{\tau}_C(a,\ell)\,;$

(B2) for $P_{Z,0}-$almost every value $\ell \in \mathcal{L}$, it holds that $\pi_{Z,0}(a_0 \,|\, \ell) > 0$.

(B3) (T,B) and W are independent given (A,L). Further (T,B) and C are independent given (W,A,L) and $T \geq W$.

Under conditions (B1)–(B2), according to the product integral form of Theorem 11 of Gill and Johansen (1990), for any $t \in (0, \tau(a,\ell) := \min\{\overline{\tau}_T(a,\ell), \overline{\tau}_C(a,\ell)\})$, we can write

$$1 - F_{Z,0}(t \,|\, a,\ell) := S_{Z,0}(t \,|\, a,\ell) = \prod_{u \in [0,t)} \{1 - \Lambda_{Z,0}(du \,|\, a,\ell)\} \;, \tag{B.1}$$

for $\Lambda_{Z,0}(t \,|\, a,\ell) := \int_0^t \frac{F_{Z,0}(du \,|\, a,\ell)}{S_{Z,0}(u \,|\, a,\ell)}$. We first express the observed follow-up time subdistribution function in terms of $P_{Z,0}$. Defining $F_0(t \,|\, w,a,\ell) := P_{Z,0}\left(T \le t \,|\, W = w, A = a, L = \ell, T \ge W\right)$ and $G_0(w \,|\, a,\ell) := P_{Z,0}\left(W \le w \,|\, A = a, L = \ell, T \ge W\right)$, we note that

$$F_0(dt \,|\, w,a,\ell) G_0(dw \,|\, a,\ell) = \frac{I_{[0,t]}(w) F_{Z,0}(dt \,|\, a,\ell) G_{Z,0}(dw \,|\, a,\ell)}{\int G_{Z,0}(u \,|\, a,\ell) F_{Z,0}(du \,|\, a,\ell)}$$

under conditions (B1)– (B3), where $G_{Z,0}(w \,|\, a,\ell) := P_{Z,0}(W \le w \,|\, A = a, L = \ell)$ denotes the target population truncation distribution. In Chapter 1, the censoring distribution is not defined in the target population, but only in the observed data population. As such a target version of the censoring distribution is not explicitly defined. We have that

$$\begin{aligned}
F_{1,0}(u \,|\, a,\ell) &= P_0\left(Y \le u, \Delta = 1 \,|\, A = a, L = \ell\right) \\
&= P_{Z,0}\left(T \le u, T \le C \,|\, A = a, L = \ell, T \ge W\right) \\
&= \iint I_{[0,u]}(t) Q_0(t \,|\, w,a,\ell) F_0(dt \,|\, w,a,\ell) G_0(dw \,|\, a,\ell) \\
&= \frac{\int I_{[0,u]}(t) \left\{ \int I_{[0,t]}(w) Q_0(t \,|\, w,a,\ell) G_{Z,0}(dw \,|\, a,\ell) \right\} F_{Z,0}(dt \,|\, a,\ell)}{\int G_{Z,0}(t \,|\, a,\ell) F_{Z,0}(dt \,|\, a,\ell)} \;,
\end{aligned}$$

which implies that

$$F_{1,0}(du \,|\, a,\ell) = \frac{\int I_{[0,u]}(w) Q_0(u \,|\, w,a,\ell) G_{Z,0}(dw \,|\, a,\ell)}{\int G_{Z,0}(t \,|\, a,\ell) F_{Z,0}(dt \,|\, a,\ell)} F_{Z,0}(du \,|\, a,\ell) \;.$$

Next, we can write that

$$\begin{aligned} R_0(u \,|\, a, z) &= P_0(W \leq u \leq Y \,|\, A = a, L = \ell) \\ &= \iint I_{[w,t]}(u) Q_0(u \,|\, w, a, \ell) F_0(dt \,|\, w, a, \ell) G_0(dw \,|\, a, \ell) \\ &= \frac{\int I_{[0,u]}(w) Q_0(u \,|\, w, a, \ell) G_{Z,0}(dw \,|\, a, \ell)}{\int G_{Z,0}(t \,|\, a, \ell) F_{Z,0}(dt \,|\, a, \ell)} S_{Z,0}(u \,|\, a, \ell) \ . \end{aligned}$$

Thus, under conditions (B1)– (B2), we find that

$$\tilde{\Lambda}_0(t \,|\, a, \ell) = \int_0^t \frac{F_{0,1}(du \,|\, a, \ell)}{R_0(u \,|\, a, \ell)} = \int_0^t \frac{F_{Z,0}(du \,|\, a, \ell)}{S_{Z,0}(u \,|\, a, \ell)} = \Lambda_{Z,0}(t \,|\, a, \ell) \ ,$$

and so, in view of (B.1), $F_{Z,0}(t \,|\, a, \ell)$ is identified by $\tilde{F}_0(t \,|\, a, \ell) := 1 - \prod_{u \in (0,t)} \{1 - \tilde{\Lambda}(t \,|\, a, \ell)\}$. The fact that $F_{Z,0}$ is identified directly implies that the target conditional truncation distribution $G_{Z,0}$ is itself identified in view of the fact that $G_{Z,0}(du \,|\, a, \ell) \propto_{a,\ell} S_{Z,0}(u \,|\, a, \ell)^{-1} G_0(du \,|\, a, \ell)$, which allows us to write that

$$G_{Z,0}(w \,|\, a, \ell) = \frac{1}{\gamma_0(a, \ell)} \int_0^w S_{Z,0}(u \,|\, a, \ell)^{-1} G_0(du \,|\, a, \ell)$$

with $\gamma_0(a, \ell) := \int S_{Z,0}(u \,|\, a, \ell)^{-1} G_0(du \,|\, a, \ell)$ the appropriate normalizing constant. Similarly, identification of $F_{Z,0}$ implies identification of the target joint exposure-covariate distribution function $J_{Z,0}$ in view of the fact that

$$J_{Z,0}(da, dz) \propto_{a,\ell} \frac{J_0(da, dz)}{\int S_{Z,0}(u \,|\, a, \ell) G_{Z,0}(du, a, \ell)} = \gamma_0(a, \ell) J_0(da, dz)$$

so that $J_{Z,0}(da, dz) = \bar{\gamma}_0(a, \ell) J_0(da, dz)$ with $\bar{\gamma}_0(a, \ell) := \gamma_0(a, \ell) / \iint \gamma_0(a, \ell) J_0(da, dz)$. Of course, the target marginal covariate distribution $H_{Z,0}$ is then itself identified from the identification for $J_{Z,0}$ through marginalization. Since ψ_0 is a functional of $F_{Z,0}$ and $H_{Z,0}$, the identification of the latter distributions directly implies that of ψ_0.

The observable conditional censoring survival function Q_0 can also be identi-

fied using product-integration as in the well-known context of right-censoring without truncation. Defining with some abuse of notation the subdistribution function $F_{0,0}(u \,|\, w, a, \ell) := P_0\left(T \leq u, \Delta = 0 \,|\, A = a, \ell = z\right)$ and at-risk probability function $R_0(u \,|\, w, a, \ell) := P_0\left(W \leq u \leq Y \,|\, W = w, A = a, \ell = z\right)$, we have that

$$\begin{aligned} F_{0,0}(u \,|\, w, a, \ell) &= -\int I(c \leq u) P_0\left(T \geq c \,|\, W = w, A = a, \ell = z, C = c\right) Q_0(dc \,|\, w, a, \ell) \\ &= -\int I(c \leq u) P_0\left(T \geq c \,|\, W = w, A = a, \ell = z\right) Q_0(dc \,|\, w, a, \ell) \ , \end{aligned}$$

which implies that $F_{0,0}(du \,|\, w, a, \ell) = -P_0\left(T \geq u \,|\, W = w, A = a, \ell = z\right) Q_0(du \,|\, w, a, \ell)$, and furthermore, for $u \geq w$, we have that

$$\begin{aligned} R_0(u \,|\, w, a, \ell) &= P_0\left(Y \geq u \,|\, W = w, A = a, \ell = z\right) \\ &= P_0\left(C \geq u \,|\, W = w, A = a, \ell = z\right) P_0\left(T \geq u \,|\, W = w, A = a, \ell = z\right). \end{aligned}$$

Thus, we find that $F_{0,0}(du \,|\, w, a, \ell)/R_0(u \,|\, w, a, \ell) = -Q_0(du \,|\, w, a, \ell)/Q_0(u \,|\, w, a, \ell)$, which then implies that $Q_0(c \,|\, w, a, \ell)$ can be identified by the product-integral

$$\prod_{c \in [0,u)} \left\{ 1 - \frac{F_{0,0}(du \,|\, w, a, \ell)}{R_0(u \,|\, w, a, \ell)} \right\}.$$

Theorem 13. *Given the conditions (B1)- (B3) hold for $a \in \{0, 1\}$ then,*

$$\psi_0 = \iint \left[\varphi_0(y, 1, \ell) - \varphi_0(y, 0, \ell)\right] \tilde{F}_0(dy \,|\, 1, \ell) \tilde{H}_0(d\ell)$$

P_0-almost surely.

Proof. The representation of the conditional mean is dependent on conditions (B2)- (B3),

such that

$$\begin{aligned} E_{Z,0}(B \,|\, T=t, A=a_0, L=\ell, T \geq W) &= E_{Z,0}(B \,|\, A=a_0, T=t, C>t, L=\ell, T \geq W) \\ &= E_0(B \,|\, A=a_0, \Delta=1, Y=t, L=\ell, T \geq W) \\ &= E_0(B \,|\, A=a_0, \Delta=1, Y=t, L=\ell) \end{aligned}$$

and we can replace B with U when conditioning on $\Delta = 1$. Using this result and the previous identification results for $\tilde{F}$ and $\tilde{H}$, we complete the proof. The proof of Theorem 5 follows immediately from the previous argument when $W = 0$ with probability one. □

We have provided a representation of the target distribution $P_{Z,0}$ in terms of observable nuisance parameters. Using the results of Chapter 1, we will also go in the reverse direction and results equations in terms of the target parameters. Given condition (B3) we can write,

$$\pi_{Z,0}(a \,|\, \ell) H_{Z,0}(d\ell) = \frac{\gamma_0(a,\ell)}{\gamma_0} \pi_0(a \,|\, \ell) H_0(d\ell)$$

where $\gamma_0 := E_0[\gamma_0(A, L)]$. Summing over $a \in \{0, 1\}$ gives the result $\tilde{H}_0(d\ell) := \frac{\gamma_0(\ell)}{\gamma_0} H_0(d\ell) = H_{Z,0}(d\ell)$, and $\frac{\gamma_0(a,\ell)}{\gamma_0(\ell)} \pi_0(a \,|\, \ell) = \pi_{Z,0}(a \,|\, \ell)$, where $\gamma_0(\ell) := E_0[\gamma_0(A, L) \,|\, L = \ell]$. Therefore each of the relevant contributions of the target distribution has a representation in terms of the observed data parameters, which is sufficient for identification.

Linearization of target parameter

To simplify the presentation of the results we introduce an additional condition

(B4) C and W are independent given (A, L) and $T \geq W$.

We use this condition to factor the term

$$\int I_{[0,u]}(w) Q_0(u \,|\, w, a, \ell) G_{Z,0}(dw \,|\, a, \ell) = Q_0(u \,|\, a, \ell) G_{Z,0}(u \,|\, a, \ell)\,,$$

and clarify the contributions of each term. In this section we show that under conditions (B1) – (B4) the linearization term $\phi_{a_0,P} := \gamma_Z^{-1}(\phi_{a_0,1,P} + \phi_{a_0,2,P} + \phi_{a_0,3,P})$ is the nonparametric influence function of the parameter $\Psi_{a_0}(P)$, where we define $\gamma_Z := P_Z(T \geq W)$ and

$$\begin{aligned}
\phi_{a_0,1,P}(\ell, a, y, \delta, u) &:= \frac{\delta S(y \,|\, a_0, \ell)}{\tilde{R}_Z(y \,|\, a_0, \ell)} \frac{I(a = a_0)}{\pi_Z(a_0 \,|\, \ell)} \frac{f_Z(y \,|\, 1, \ell)}{f_Z(y \,|\, a_0, \ell)} \left[u - \varphi_Z(y, a_0, \ell)\right] \\
\phi_{a_0,2,P}(\ell, a, y, \delta) &:= -\frac{I(a = 1)}{\pi_Z(1 \,|\, \ell)} \left[\frac{\delta D_{Z,\varphi}(y, a_0, \ell)}{\tilde{R}_Z(y \,|\, 1, \ell)} - \int_w^y \frac{D_{Z,\varphi}(u \,|\, a_0, \ell) \Lambda_Z(du \,|\, 1, \ell)}{\tilde{R}_Z(u \,|\, 1, \ell)}\right] \\
\phi_{a_0,3,P}(\ell) &:= \left\{\mu_Z(a_0, \ell) - \Psi_{a_0}(P)\right\} \phi_{S,P}(1 - G_Z)(o_i)\,,
\end{aligned}$$

where $\phi_{S,P}(1 - G_Z) : o \mapsto \frac{1}{S_Z(w \,|\, a, \ell)} + \frac{\delta[1 - G_Z(y \,|\, a, \ell)]}{S_Z(y \,|\, a, \ell) Q_P(y \,|\, a, \ell) G_Z(y|a, \ell)} - \int_w^y \frac{[1 - G_Z(u \,|\, a, \ell)] \Lambda_Z(du \,|\, a, \ell)}{S_Z(u \,|\, a, \ell) Q_P(u \,|\, a, \ell) G_Z(u \,|\, a, \ell)}$ as in Chapter 1 and $\tilde{R}_Z(y \,|\, a, \ell) := S_Z(y \,|\, 1, \ell) Q_P(y \,|\, 1, \ell) G_Z(y \,|\, 1, \ell)$. When there is no truncation, when $W = 0$ with probability one, this result simplifies to the what is presented in Chapter 2.3.3 by observing that $G_Z(w \,|\, a, \ell) = 1$ and $\phi_{S,P}(1 - G_Z)(o) = 1$ almost-surely .

Lemma 4. *The component of the nonparametric gradient of the parameter $\Psi(P)$ that lies in the tangent space $U \,|\, \Delta, Y, A, L, T \geq W$ is*

$$\phi_{a_0,1}(o) = \frac{1}{\gamma_Z} \frac{\delta}{Q_P(y \,|\, a_0, \ell) G_Z(y \,|\, a_0, \ell)} \frac{I(a = a_0)}{\pi_Z(a_0 \,|\, \ell)} m_Z(y, a_0, \ell) \left[u - \varphi_z(y, a_0, \ell)\right]\,,$$

where $m_z(y, a_0, \ell) := \frac{f_Z(y \,|\, 1, \ell)}{f_Z(y \,|\, a_0, \ell)}$.

Proof. Let $P \in \mathcal{M}$ be given, and take $\{P_\epsilon : |\epsilon| > 0\}$ to be a suitably smooth and bounded (i.e., Hellinger-differentiable) path with $P_{\epsilon=0} = P$ and score for ϵ at $\epsilon = 0$ given by $h \in L_2^0(P)$, for which $o \mapsto h(y, \delta, w, a, \ell)$ such that P–almost surely, $E_P[h(O) \,|\, Y, \Delta, A, L] = 0$, $var_P[h(O) \,|\, Y, \Delta, A, L] < \infty$. We wish to compute the pathwise derivative

$$\frac{\partial}{\partial \epsilon} \Psi(P_\epsilon)\bigg|_{\epsilon=0} = \frac{\partial}{\partial \epsilon} \iint \varphi_\epsilon(t, \ell) \tilde{F}_\epsilon(dt \,|\, a_0, \ell) \tilde{H}_\epsilon(dz)\bigg|_{\epsilon=0}$$

where here and below we use the shorthand notation A_ϵ to refer to A_{P_ϵ} for any relevant

quantity A_P indexed by P. Furthermore, under mild regularity conditions allowing interchange of integral and derivative operations, this pathwise derivative can be decomposed as (P1) + (P2) with

$$\begin{aligned}(\text{P1}) &= \iint \frac{\partial}{\partial \epsilon} \int \varphi_\epsilon(t,\ell)\tilde{F}_0(dt \,|\, a_0,\ell)\Big|_{\epsilon=0} \tilde{H}_0(d\ell) \\ (\text{P2}) &= \iint \varphi_0(t,\ell)\, \frac{\partial}{\partial \epsilon}\left[\tilde{F}_\epsilon(dt \,|\, a_0,\ell)\tilde{H}_\epsilon(d\ell)\right]\Big|_{\epsilon=0}.\end{aligned}$$

We have already studied the summand (P2) in Appendix A, below we study (P1).

Replacing the identified expression of the survival distribution with its target equivalents we have that

$$(\text{P1}) = \int \frac{I(\delta=1)I(a=a_0)}{P_0(\Delta=1, A=a_0 \,|\, Y=y, L=\ell)} u h_0(o) P_0(du, d\delta, dw, da \,|\, y,\ell) F_{Z,0}(dy \,|\, 1,\ell)\tilde{H}_0(d\ell)$$

which nearly isolates the linearization term. Before centering, we simplify the expression of the density given the relationship

$$\begin{aligned}F_{Z,0}(dy \,|\, a_0, l) &= \frac{F_{1,0}(dy \,|\, a_0,\ell)}{R_0(y \,|\, a_0,\ell)} S_{Z,0}(y \,|\, a_0,\ell) \\ &= P_0(dy \,|\, \ell) \frac{P_0(\Delta=1, A=a_0 \,|\, Y=y, L=\ell)}{\gamma_{Z,0}(a_0,\ell) Q_0(y \,|\, a_0,\ell) G_{Z,0}(y \,|\, a_0,\ell)\pi_0(a_0 \,|\, \ell)} \\ &= P_0(dy \,|\, \ell) \frac{1}{\gamma_{Z,0}(\ell)\pi_{Z,0}(a_0 \,|\, \ell)} \frac{P_0(\Delta=1, A=a_0 \,|\, Y=y, L=\ell)}{Q_0(y \,|\, a_0,\ell) G_{Z,0}(y \,|\, a_0,\ell)},\end{aligned}$$

where $P_0(dy \,|\, \ell)$ denotes the density of $Y \,|\, L$. We define $m_{Z,0}(y,a,\ell) := \frac{f_{Z,0}(y \,|\, 1,\ell)}{f_{Z,0}(y \,|\, a,\ell)}$ and using the cancellation we can write

$$(\text{P1}) = E_P\left[\frac{I(\Delta=1) m_{Z,0}(Y,A,L)}{Q_0(Y \,|\, A=a_0, L=\ell) G_{Z,0}(Y \,|\, A=a_0, L=\ell)} \frac{I(A=a_0)}{P_{Z,0}(A=a_0 \,|\, L=\ell)} U h(O)\right].$$

Using the fact that $h(O)$ has mean zero conditional on (Y,Δ,A,L) and $T \geq W$, and that

the mean $E_P[U \mid Y = y, \Delta = 1, A = a_0, L = \ell] = \varphi_0(y, \ell)$, we can write

$$\text{(P1)} \;=\; E_P\left\{\frac{I(\Delta = 1)m_{Z,0}(Y, A, L)}{Q_0(Y \mid A = a_0, L = \ell)G_{Z,0}(Y \mid A = a_0, L = \ell)}\frac{I(A = a_0)\left[U - \varphi_0(y, \ell)\right]}{P_{Z,0}(A = a_0 \mid L = \ell)}h(O)\right\}.$$

Equation (P2) was studied in Appendix A.2 and when we write the observed data parameters in terms of target parameters we get the given expressions for $\phi_{a_0,2,P}$ and $\phi_{a_0,3,P}$. □

In the analysis of autopsy data we may have to account for potential differences in the observed autopsy outcomes among individuals that have consented compared to those that have not. To do this we introduce two variables to distinguish between outcomes that are missing due to lack of consent and those missing because they are still alive. Let $\Delta_1 = 1$ indicate observable outcomes among for those that consent, and $\Delta_2 = I(T \leq C)$ is the indicator of survival these are used to construct the observed autopsy outcomes $U := B\Delta_1\Delta_2$. We note that we can always write $E_0(U \mid A = a_0, T = y, L = \ell) := E_0(B \mid \Delta_1 = 1, \Delta_2 = 1, A = a_0, Y = y, L = \ell)P(\Delta_1 = 1 \mid \Delta_2 = 1, A = a_0, Y = y, L = \ell)P(\Delta_2 = 1 \mid A = a_0, Y = y, L = \ell)$. Following the above result we can rewrite the first contribution of the influence function as

$$\phi_{a_0,1}(o) = \frac{\delta_1}{\pi_{P,1}(1 \mid 1, a_0, y, \ell)}\frac{\delta_2}{Q_P(y \mid a_0, \ell)G_Z(y \mid a_0, \ell)}\frac{I(a = a_0)}{\pi_{Z,2}(a_0 \mid \ell)}m_Z(y, a_0, \ell)\left[u - \varphi_z(y, a_0, \ell)\right],$$

where $\pi_{P,1}(1 \mid 1, a_0, y, \ell) := P_Z(\Delta_1 = 1 \mid \Delta_2 = 1, A = a_0, Y = y, T = t, T \geq W)$ and $\pi_{Z,2}(a_0 \mid \ell) := P_Z(A = a_0 \mid L = \ell)$. Notably the consent probability $\pi_{P,1}$ is defined in the observable population.

We use the representation of the nonparametric efficient influence function $\phi_{a_0,P}$ of $\Psi(P)$ to show the additional contribution of $\varphi(t, \ell)$ no longer being a fixed function to the remainder.

Lemma 5. *The remainder* $R(P, P_0) := \Psi(P) - \Psi(P_0) + P_0\phi_{a_0,P}$ *can be written as* $R(P, P_0) =$

$R_1(P, P_0) + R_2(P, P_0) + R_3(P, P_0)$ *with*

$$R_1(P, P_0) := E_0 \left\{ \int [\varphi_Z(y, a_0, L) - \varphi_{Z,0}(y, a_0, L)] \left[1 - \frac{\nu_{Z,0}(y, a_0, L)}{\nu_Z(y, a_0, L)} \frac{m_Z(y, a_0, L)}{m_{Z,0}(y, a_0, L)} \right] F_{Z,0}(dy \,|\, 1, L) \right\}$$

$$R_2(P, P_0) := E_0 \left\{ \int \frac{D_{Z,\varphi}(y \,|\, a_0, L) S_{Z,0}(y \,|\, 1, L)}{S_Z(y \,|\, 1, L)} \left(\frac{\nu_{Z,0}(y, a_0, L)}{\nu_Z(y, a_0, L)} - 1 \right) (\Lambda_Z - \Lambda_{Z,0})(dy \,|\, 1, L) \right\}$$

where $\nu_Z(y, a_0, \ell) := \gamma_Z Q_P(y \,|\, a_0, \ell) G_Z(y \,|\, a_0, \ell) \pi_Z(a_0 \,|\, \ell)$ *and* $R_3(P, P_0)$ *includes the term* $R_2(P, P_0) + R_3(P, P_0) + R_4(P, P_0)$ *defined in Appendix A.*

Proof. We examine the first order difference of the parameter around P_0

$$\Psi(P) - \Psi(P_0) = \int [\mu_Z(a_0, \ell) - \mu_{Z,0}(a_0, \ell)] \tilde{H}_{Z,0}(d\ell) + \int \mu_Z(a_0, \ell)(\tilde{H}_Z - \tilde{H}_{Z,0})(d\ell)$$

and further expand the survival regression,

$$\begin{aligned} \mu_Z(a_0, \ell) - \mu_{Z,0}(a_0, \ell) = & \int [\varphi_Z(y, a_0, \ell) - \varphi_{Z,0}(y, a_0, \ell)] \tilde{F}_{Z,0}(dy \,|\, 1, \ell) \\ & + \int \varphi_{Z,0}(a_0, y, \ell)(\tilde{F}_Z - \tilde{F}_{Z,0})(dy \,|\, 1, \ell) \,. \end{aligned}$$

We have studied the remainder in detail in Appendix A of Chapter 1 and therefore only present the novel contribution. Using Lemma 2 of in the Appendix A and Theorem 6 (Gill and Johansen, 1990), we can write

$$\int \varphi_Z(a_0, y, \ell)(\tilde{F}_P - \tilde{F}_0)(dy \,|\, 1, \ell) = - \int \frac{D_{Z,\varphi}(y, a_0, \ell) \tilde{S}_0(y \,|\, 1, \ell)}{\tilde{S}_P(y \,|\, 1, \ell)} (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, 1, \ell)$$

Next, we write the expectation of the influence function as

$$\begin{aligned} \mathbb{P}_0 \phi_{a_0, P} = & \iint \frac{m_Z(y, a_0, \ell)}{m_{Z,0}(y \,|\, 1, l)} \frac{\nu_{Z,0}(y, a_0, \ell)}{\nu_Z(y, a_0, \ell)} [\varphi_{Z,0}(y, a_0, \ell) - \varphi_Z(y, a_0, \ell)] \tilde{F}_0(dy \,|\, 1, \ell) H_0(d\ell) \\ & + \iint \frac{D_{Z,\varphi}(y \,|\, a_0, \ell) \tilde{S}_0(y \,|\, 1, \ell)}{\tilde{S}_Z(y \,|\, 1, \ell)} \frac{\nu_{Z,0}(y, a_0, L)}{\nu_Z(y, a_0, L)} (\tilde{\Lambda}_P - \tilde{\Lambda}_0)(dy \,|\, 1, \ell) H_{Z,0}(d\ell) \\ & - \int \mu_Z(a_0, \ell)(H_Z - H_{Z,0})(d\ell) \,, \end{aligned}$$

which when combined with the first order difference gives the expression for the remainders $R_1(P, P_0)$ and $R_2(P, P_0)$ and completes the proof. □

Theorem 14. *If conditions (B1)–(B3) hold for $a_0 \in \{0, 1\}$ then, the functional $P \mapsto \Psi_{a_0}(P)$ is a pathwise differentiable parameter with influence function $\phi_{a_0,P}(o)$ and the linearization*

$$\Psi_{a_0}(P) - \Psi_{a_0}(P_0) = \int \phi_{a_0,0}(o)(P - P_0)(do) + R(P, P_0)$$

holds for $P \in \mathcal{M}$.

Proof. Using Lemma 3 from Section B.2 and Theorem 1 in Chapter 1, the representation of the influence function follows. □

Estimation and Inference

We propose the following conditions under which asymptotic inference is valid. We denote the data generating distribution implied by the estimated nuisance functions by $P_n := (\varphi_n, F_n, G_n, Q_n, \pi_n)$. Further, we denote by $P_\infty := (\varphi_\infty, F_\infty, G_\infty, Q_\infty, \pi_\infty)$ the limiting distribution of each estimated nuisance function. We suppose the following conditions hold P_0-almost surely.

(B5) The class of functions $\{o \mapsto (\phi_n - \phi_0)(o)\}$ is Donsker.

(B6) There exists a constant $\epsilon > 0$ such that for all $y \in (0, \tau(a, \ell))$ and almost every ℓ:

(i) $1/\pi_n(\ell) \leq \epsilon; \quad 1/\pi_\infty(\ell) \leq \epsilon;$

(ii) $1/Q_n(y \,|\, a, \ell) \leq \epsilon; \quad 1/Q_\infty(y \,|\, a, \ell) \leq \epsilon;$

(iii) $1/G_n(y \,|\, a, \ell) \leq \epsilon; \quad 1/G_\infty(y \,|\, a, \ell) \leq \epsilon;$

(iv) $D_{n,\varphi_n}(y, a, \ell)/S_n(y \,|\, a, \ell) \leq \epsilon; \quad D_{\infty,\varphi_\infty}(y, a, \ell)/S_\infty(y \,|\, a, \ell) \leq \epsilon.$

(B7) For almost-every (a, ℓ) and for all $y \in (0, \tau(a, \ell))$ the following hold

(i) $F_\infty(y \,|\, a, \ell) = F_{Z,0}(y \,|\, a, \ell),$

(ii) $Q_\infty(y \mid a, \ell) = Q_0(y \mid a, \ell)$, $G_\infty(y \mid a, \ell) = G_{Z,0}(y \mid a, \ell)$, and $\pi_\infty(\ell) = \pi_{Z,0}(\ell)$;

(iii) $\varphi_\infty(y \mid a, \ell) = \varphi_{Z,0}(y \mid a, \ell)$

(B8) For all $y \in (0, \tau(a, \ell))$ and each $a_0 \in \{0, 1\}$, it holds that

$$R(P_n, P_0) = o_P\left(n^{-1/2}\right),$$

where $R(P_n, P_0)$ is the remainder evaluated at the distribution implied by the estimation procedure.

We prove the robustness result from Section 2.3.3 in the absence of left truncation.

Lemma 6. *In the absence of left truncation, when $W = 0$ with probability one, if two of the three conditions of (B7) are satisfied, then $\psi_n^* \xrightarrow{\text{P}} \psi_0$. In this setting an empirical one-step estimator and the estimating equations estimator are identical.*

Proof. Showing the equivalence is accomplished by rearranging the term $\phi_{a_0,3,P}(\ell)$, and isolating the parameter. In the absence of left truncation the remainder term is exactly $R(P, P_0) = R_1(P, P_0) + R_2(P, P_0)$, therefore it is sufficient to show that $R(P_\infty, P_0) = 0$ when two out of three conditions are satisfied. If either $\varphi_\infty = \varphi_0$ or all of $(F_\infty = F_0, Q_\infty = Q_0, \pi_\infty = \pi_0)$ then $R_1(P_\infty, P_0) = 0$. If either $F_\infty = F_0$ or both $(Q_\infty = Q_0, \pi_\infty = \pi_0)$ then $R_2(P_\infty, P_0) = 0$. We need only two of the three conditions to ensure $R(P_\infty, P_0) = 0$.

□

Proof of Theorem 7

Specifically, we restate Theorem 7 as the following. When conditions (B5) - (B8) hold, the one-step estimator is asymptotically linear

$$n^{1/2}\left[\psi_n^* - \psi_0\right] \xrightarrow{\text{d}} N(0, \sigma_0^2)$$

where $\sigma_0^2 := E_0\left[\phi_{1,0}(O)^2 + \phi_{0,0}(O)^2 - 2\phi_{1,0}(O)\phi_{0,0}(O)\right]$.

Proof. For each treatment level $a_0 \in \{0, 1\}$, we can decompose the one-step estimator as

$$\psi_{a_0,n} - \psi_{a_0,0} = (\mathbb{P}_n - \mathbb{P}_0)\phi_{a_0,0} + (\mathbb{P}_n - \mathbb{P}_0)(\phi_{a_0,n} - \phi_{a_0,0}) + R(P_n, P_0)$$

Conditions (B5) – (B7) and Lemma 19.24 of Van der Vaart (2000) ensures that $(\mathbb{P}_n - \mathbb{P}_0)(\phi_{a_0,n} - \phi_{a_0,0}) = o_P\left(n^{-1/2}\right)$ and that

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P}_0)\phi_{a_0,0} \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2_{a_0,0}),$$

where $\sigma^2_{a_0,0} := E_0[\phi_{a_0,0}(O)^2]$, while condition (B8) and Lemma 5 ensure that $R(P_n, P_0) = o_P\left(n^{-1/2}\right)$. Using the delta method for the difference $\psi^*_n := \psi^*_{1,n} - \psi^*_{0,n}$ completes the proof. □

Lemma 7. *When conditions (B5) – (B8) hold and if* $\frac{1}{n}\sum_{i=1}^n \{\gamma_n^{-1}\phi_{F,n}(O_i) - 1\} = O_P\left(n^{-1/2}\right)$ *then, the estimating equations estimator* ψ^*_n *has the same limiting distribution as the one-step estimator* ψ_n*. Further the difference between the two estimators is*

$$\psi^*_n - \psi^*_n = \frac{1}{n}\sum_{i=1}^n \left\{\gamma_n^{-1}\phi_{F,n}(O_i) - 1\right\}\left\{\psi^*_n - \hat{\psi}\right\},$$

where $\hat{\psi}$ *is an arbitrary* consistent *estiamtor of* ψ_0*.*

Proof. We define $\phi_{1,n} := \phi_{1,1,n} - \phi_{0,1,n}$ and $\phi_{2,n} := \phi_{1,2,n} - \phi_{0,2,n}$ as the first two contributions of the influence function of $\Psi_1(P_n)$ and $\Psi_0(P_n)$. For a choice consistent plugin estiamtor $\hat{\psi}$, the one step estimator can be written as

$$\begin{aligned}
\psi^*_n &= \hat{\psi} + \frac{1}{n}\sum_{i=1}^n \gamma_n^{-1}\left[\phi_{1,n}(O_i) + \phi_{2,n}(O_i) + \left[\{\mu_n(1, L_i) - \mu_n(0, L_i)\} - \hat{\psi}\right]\phi_{F,n}(O_i)\right] \\
&= \hat{\psi} + \frac{1}{n}\sum_{i=1}^n \frac{\phi_{F,n}(O_i)}{\gamma_n}\left[\left[\{\mu_n(1, L_i) - \mu_n(0, L_i)\} - \hat{\psi}\right] - \left[\{\mu_n(1, L_i) - \mu_n(0, L_i)\} - \psi^*_n\right]\right],
\end{aligned}$$

where we've utilized the definition of the estimating equations estimator. Rearranging the

terms completes the proof

$$\psi_n^* - \psi_n^* = \frac{1}{n}\sum_{i=1}^{n}\left\{\gamma_n^{-1}\phi_{F,n}(O_i) - 1\right\}\left\{\psi_n^* - \hat{\psi}\right\},$$

and notably if $\gamma_n := \frac{1}{n}\sum_{i=1}^{n}\phi_{F,n}(O_i)$ then the estimating equations estimator is equivalent to a modified one-step. □

Part C: Retrospective data setting

Details of identification result

Many of the arguments presented in the retrospective data setting follow from the prospective setting. Two notable differences are the absence of censoring and the direction of truncation. As we will see this simplifies the results and only mildly changes the derivations. Before we establish the identification of ψ_0, we show that the components of the ideal data distribution $P_{Z,0}$ can be identified in terms of data distribution P_0. Under conditions (C1) – (C2), according to the product integral form of Theorem 11 of Gill and Johansen (1990), for any $t \in (0, \tau_T(a,\ell))$, we can write

$$\begin{aligned} F_{Z,0}(t \,|\, a,\ell) &:= P_{Z,0}(T \le t \,|\, a,\ell) = \prod_{u\in(t,\infty)}\{1 - \Gamma_{Z,0}(du \,|\, a,\ell)\} \\ G_{Z,0}(w \,|\, a,\ell) &:= P_{Z,0}(W > w \,|\, a,\ell) = \prod_{u\in(0,w)}\{1 - \Lambda_{Z,0}(du \,|\, a,\ell)\}, \end{aligned}$$

where $\Gamma_{Z,0}(t \,|\, a,\ell) := \int_t^\infty \frac{F_{Z,0}(du \,|\, a,\ell)}{F_{Z,0}(u \,|\, a,\ell)}$ is the backwards cumulative hazard of the event-time distribution and $\Lambda_{Z,0}(w \,|\, a,\ell) := -\int_0^w \frac{G_{Z,0}(du \,|\, a,\ell)}{G_{Z,0}(u \,|\, a,\ell)}$ is the cumulative hazard of the truncation times. Defining $F_0(t \,|\, a,\ell) := P_0(T \le t \,|\, A = a, L = \ell, T \le W)$, $G_0(w \,|\, a,\ell) := P_0(W > w \,|\, A = a, L = \ell)$, and $R_0(u \,|\, a,\ell) := P_0(T \le u \le W \,|\, A = a, L = \ell) = P_{Z,0}(T \le u \le W \,|\, A = a, L = \ell, T \le W)$, we will identify the target survival and truncation distributions using the

product-integral form via

$$\tilde{F}_0(t \,|\, a, \ell) := \prod_{u>t} \left\{ 1 - \frac{F_0(du \,|\, a, \ell)}{R_0(u \,|\, a, \ell)} \right\}$$
$$\tilde{G}_0(w \,|\, a, \ell) := \prod_{u<t} \left\{ 1 + \frac{G_0(du \,|\, a, \ell)}{R_0(u \,|\, a, \ell)} \right\} .$$

First, we can write the observed data distribution in terms of target distribution parameters, specifically as

$$P_0(d\ell, da, dw, dt, db) = -P_{Z,0}(db \,|\, w, t, \ell, a) I(t \leq w) \gamma_{Z,0} F_{Z,0}(dt \,|\, a, \ell) G_{Z,0}(dw \,|\, a, \ell) P_{Z,0}(d\ell, da) , \tag{B.2}$$

where $\gamma_{Z,0} := P(T \leq W)^{-1}$. We define the inclusion probability weights as $\gamma_{Z,0}(a, \ell) := P_{Z,0}(T \leq W \,|\, A = a, L = \ell)^{-1}$ such that when condition (C3) holds, the observed follow-up and truncation distributions can be written as

$$\begin{aligned} F_0(u \,|\, a, \ell) &= P_{Z,0}(T \leq u \,|\, A = a, L = \ell, T \leq W) \\ &= \gamma_{Z,0}(a, \ell) \int_0^u G_{Z,0}(t \,|\, a, \ell) F_{Z,0}(dt \,|\, a, \ell) \\ G_0(u \,|\, a, \ell) &= P_{Z,0}(W > u \,|\, A = a, L = \ell, T \leq W) \\ &= -\gamma_{Z,0}(a, \ell) \int_u^\infty F_{Z,0}(w \,|\, a, \ell) G_{Z,0}(dw \,|\, a, \ell) , \end{aligned}$$

and the observed risk probability

$$R_0(u \,|\, a, \ell) := \gamma_{Z,0}(a, \ell) G_{Z,0}(u \,|\, a, \ell) F_{Z,0}(u \,|\, a, \ell) .$$

Thus, under conditions (C1) – (C3), we find that

$$F_{Z,0}(t \,|\, a, \ell) = \prod_{u>t} \left\{ 1 - \frac{F_{Z,0}(du \,|\, a, \ell)}{F_{Z,0}(u \,|\, a, \ell)} \right\} = \prod_{u>t} \left\{ 1 - \frac{F_0(du \,|\, a, \ell)}{R_0(u \,|\, a, \ell)} \right\} = \tilde{F}_0(t \,|\, a, \ell)$$

and the result is similar for $\Lambda_{Z,0}(w \,|\, a, \ell)$, and following Theorem 11 of Gill and Jo-

hansen (1990) this gives a representation of the target survival distribution. We note that these identification results also give a representation for $\gamma_{Z,0}(a,\ell)$ via $\tilde{\gamma}_0(a,\ell) := \int \tilde{G}_0(t \,|\, a,\ell)\tilde{F}_0(dt \,|\, a,\ell)$.

Proof Theorem 8

Proof. The representation of the conditional mean is dependent on conditions (C2) – (C3), such that

$$\begin{aligned} E_{Z,0}(B \,|\, T=t, A=a_0, L=\ell) &= E_{Z,0}(B \,|\, T=t, W>t, A=a_0, L=\ell) \\ &= E_0(B \,|\, T=t, A=a_0, L=\ell)\,. \end{aligned}$$

To complete the identification result we present a representation of the target propensity and covariate variable distributions in terms of observed data summaries. Given conditions (C1) – (C3) we can write,

$$\pi_{Z,0}(a \,|\, \ell) H_{Z,0}(d\ell) = \frac{\gamma_0(a,\ell)}{\gamma_0} \pi_0(a \,|\, \ell) H_0(d\ell)$$

where $\gamma_0 := E_0[\gamma_0(A,L)]$. Summing over $a \in \{0,1\}$ gives the result $\tilde{H}_0(d\ell) = H_{Z,0}(d\ell)$, and $\frac{\gamma_0(a,\ell)}{\gamma_0(\ell)}\pi_0(a \,|\, \ell) = \pi_{Z,0}(a \,|\, \ell)$, where $\gamma_0(\ell) := E_0[\gamma_0(A,L) \,|\, L=\ell]$. Each term of the target distribution has a representation in terms of the observed data parameters, which completes the proof. □

Linearization of target parameter

In this section we show that under conditions (C1) – (C3) the linearization term $\phi_{a_0,P} := \gamma_Z^{-1}(\phi_{a_0,1,P} + \phi_{a_0,2,P} + \phi_{a_0,3,P})$ is the nonparametric influence function of the parameter $\Psi_{a_0}(P)$, where we define $\gamma_Z := P_Z(T \leq W)$, the partial integral kernel $D_{Z,\varphi} : (t, a_0, \ell) \mapsto$

$\int_0^t F_Z(u \mid 1, \ell)\varphi_Z(du, a_0, \ell)$. and

$$\begin{aligned}
\phi_{a_0,1}(\ell, a, t, w, b) &= \frac{I(a = a_0)}{\pi_Z(a_0 \mid \ell)} \frac{f_Z(t \mid 1, \ell)}{f_Z(t \mid a_0, \ell)} \frac{1}{G_Z(t \mid a, \ell)} \left[b - \varphi_Z(t, a_0, \ell)\right] \\
\phi_{a_0,2}(\ell, a, t, w) &= -\frac{I(a = 1)}{\pi_0(1 \mid \ell)} \left\{ \frac{D_{Z,\varphi}(t, a_0, \ell)}{F_Z(t \mid 1, \ell) G_Z(t \mid 1, \ell)} - \int_t^w \frac{D_{Z,\varphi}(u, a_0, \ell)}{G_Z(u \mid a, l) F_Z(u \mid 1, \ell)} \Gamma_Z(du \mid 1, \ell) \right\} \\
\phi_{a_0,3}(\ell, a, t, w) &= \left[\mu_Z(a_0, \ell) - \psi_{a_0,Z}\right] \phi_{F,P}(1 - G_Z)(o)\,,
\end{aligned}$$

where $\phi_{F,P}(1 - G_Z) : o \mapsto \frac{1}{F_Z(w \mid a,\ell)} + \frac{1 - G_Z(t \mid a,\ell)}{F_Z(t \mid a,\ell) G_Z(t \mid a,\ell)} - \int_t^w \frac{[1 - G_Z(u \mid a,\ell)]\Gamma_Z(du \mid a,\ell)}{G_Z(u \mid a,\ell) F_Z(u \mid a,\ell)}$ as in Part B. The result follows from the results of Chapter 1 in the absence of censoring. The right truncation case has the same linearization but uses the reverse hazard function instead of the usual hazard function. The argument for $\phi_{a_0,1}$ follow exactly from Appendix B.2. We present a modification of Lemma 2 (a) in Appendix A for the case of right truncation

Lemma 8. *For each $(a, \ell) \in \{0, 1\} \times \mathcal{L}$, the following identities hold*

$$\begin{aligned}
&(a)\ F_Z(t \mid a, \ell) - F_{Z,0}(t \mid a, \ell) = -F_Z(t \mid a, \ell) \int_t^\infty \frac{F_{Z,0}(u \mid a, \ell)}{F_Z(u \mid a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(du \mid a, \ell)\,; \\
&(b)\ \int \varphi(t, \ell)(F_Z - F_{Z,0})(dt \mid a, \ell) = -\int D_{F,\varphi}(t, a, \ell) \frac{F_{Z,0}(t \mid a, \ell)}{F_Z(t \mid a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(dt \mid a, \ell)\,.
\end{aligned}$$

Proof. Following the same argument as Theorem 6 of Gill and Johansen (1990) we can write for any two continuous distribution functions F_Z and $F_{Z,0}$ and their corresponding reverse cumulative hazard function Γ_Z and $\Gamma_{Z,0}$. Starting on the right hand side we use integration by parts to write

$$\begin{aligned}
\int_t^\infty \frac{F_{Z,0}(u \mid a, \ell)}{F_Z(u \mid a, \ell)} \Gamma_Z(du \mid a, \ell) &= -\left. \frac{F_{Z,0}}{F_Z}(u \mid a, \ell) \right|_t^{\tau_T(a,\ell)} + \int_t^\infty \frac{F_{Z,0}(du \mid a, \ell)}{F_Z(u \mid a, \ell)} \\
&= \frac{F_{Z,0}}{F_Z}(t \mid a, \ell) - 1 + \int_t^\infty \frac{F_{Z,0}(u \mid a, \ell)}{F_Z(u \mid a, \ell)} \Gamma_{Z,0}(du \mid a, \ell)\,.
\end{aligned}$$

Replacing this result into the right hand side of Equation (a) gives the result.

The differential form of this equation is

$$(F_Z - F_{Z,0})(dt \,|\, a, \ell) = F_{Z,0}(u \,|\, a, \ell)(\Gamma_Z - \Gamma_{Z,0})(du \,|\, a, \ell) - F_Z(dt \,|\, a, \ell) \int_t^\infty \frac{F_{Z,0}(u \,|\, a, \ell)}{F_Z(u \,|\, a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(du \,|\, a, \ell),$$

which allows us to write

$$\begin{aligned}
&\int \varphi(t, \ell)(F_Z - F_{Z,0})(dt \,|\, a, \ell) \\
&= \int \varphi(t, \ell) \left\{ F_{Z,0}(u \,|\, a, \ell)(\Gamma_Z - \Gamma_{Z,0})(du \,|\, a, \ell) - F_Z(dt \,|\, a, \ell) \int_t^\infty \frac{F_{Z,0}(u \,|\, a, \ell)}{F_Z(u \,|\, a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(du \,|\, a, \ell) \right\} \\
&= - \int \int_0^t \varphi(u, \ell) F_Z(du \,|\, a, \ell) \frac{F_{Z,0}(t \,|\, a, \ell)}{F_Z(t \,|\, a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(dt \,|\, a, \ell)
\end{aligned}$$

where we use integration by parts and swap the order of integration of the second term, similarly to Appendix A. □

We now study the linearization of Ψ_{a_0} around P_0 based $\phi_{a_0,P}$. Specifically, we derive the form of the remainder term $R(P, P_0) := \Psi_{a_0}(P) - \Psi_{a_0}(P_0) + P_0\phi_{a_0,P}$ from this linearization.

Lemma 9. *The remainder $R(P, P_0) := \Psi_{a_0}(P) - \Psi_{a_0}(P_0) + P_0\phi_{a_0,P}$ can be written and $R(P, P_0) = R_1(P, P_0) + R_2(P, P_0) + R_3(P, P_0) + R_4(P, P_0) + R_5(P, P_0)$ with*

$$\begin{aligned}
R_1(P, P_0) &:= E_{Z,0} \left\{ \int [\varphi_Z(y, a_0, L) - \varphi_{Z,0}(y, a_0, L)] \left[1 - \frac{\nu_{Z,0}(y, a_0, L)}{\nu_Z(y, a_0, L)} \frac{m_Z(y, a_0, L)}{m_{Z,0}(y, a_0, L)} \right] F_{Z,0}(dt \,|\, 1, L) \right\} \\
R_2(P, P_0) &:= E_{Z,0} \left\{ \int \frac{D_{Z,\varphi}(t, a_0, L) F_{Z,0}(t \,|\, 1, \ell)}{F_Z(t \,|\, 1, \ell)} \left\{ \frac{\nu_{Z,0}(t, 1, L)}{\nu_Z(t, 1L)} - 1 \right\} (\Gamma_Z - \Gamma_{Z,0})(dt \,|\, 1, L) \right\} \\
R_3(P, P_0) &:= \frac{\gamma_{Z,0}}{\gamma_Z} E_{Z,0} \left\{ \int \xi_Z(t, A, L) \left[\frac{\gamma_Z(A, L)}{\gamma_{Z,0}(A, L)} - \frac{F_{Z,0}(t \,|\, A, L) G_{Z,0}(t \,|\, A, L)}{F_Z(t \,|\, a, L) G_Z(t \,|\, A, L)} \right] \left(\frac{S_{Z,0}}{S_Z} - 1 \right) (dt \,|\, A, L) \right\} \\
R_4(P, P_0) &:= \frac{\gamma_{Z,0}}{\gamma_Z} E_{Z,0} \left\{ \int [\mu_Z(a_0, L) - \Psi_{a_0}(P_Z)] \left[\frac{F_Z(w \,|\, A, L) - F_{Z,0}(w \,|\, A, L)}{F_Z(w \,|\, A, L)} \right] (G_Z - G_{Z,0})(dw \,|\, A, L) \right\} \\
R_5(P, P_0) &:= \frac{\gamma_Z - \gamma_{Z,0}}{\gamma_Z} \int \mu_Z(a_0, \ell)(J_Z - J_{Z,0})(da, d\ell)
\end{aligned}$$

where we define $\xi_Z(t, a, \ell) := [\mu_Z(a_0, \ell) - \psi_{a_0,Z}(P_Z)][1 - G_Z(t \,|\, a, \ell)]$ and $\nu_Z(t, a_0, \ell) :=$

$\gamma_Z G_Z(t \mid a_0, \ell)\pi_Z(a_0 \mid \ell)$.

Proof. We begin by decomposing the difference $\Psi_{a_0}(P) - \Psi_{a_0}(P_0)$ as the sum (D1) + (D2) + (D3) + (D4) + (D5) with

$$\begin{aligned}
\text{(D1)} &= \int [\varphi_Z(t, a_0, \ell) - \varphi_{Z,0}(t, a_0, \ell)] F_{Z,0}(dt \mid 1, \ell) H_{Z,0}(d\ell) \\
\text{(D2)} &= \int \varphi_Z(t, a_0, \ell)[\tilde{F}_P - \tilde{F}_0](dt \mid 1, \ell) H_{Z,0}(d\ell) \\
\text{(D3)} &= \int \gamma_Z^{-1}[\mu_Z(a_0, \ell) - \psi_{a_0,Z}][\gamma_Z(a, \ell) - \gamma_{Z,0}(a, \ell)] J_0(da, d\ell) \\
\text{(D4)} &= \int [\mu_Z(a_0, \ell) - \psi_{a_0,Z}] \frac{\gamma_Z(a, \ell)}{\gamma_Z} (J_P - J_0) da, d\ell \\
\text{(D5)} &= \int [\mu_Z(a_0, \ell) - \psi_{a_0,Z}] \left(H_Z - H_{Z,0}\right)(d\ell) \frac{\gamma_{Z,0} - \gamma_Z}{\gamma_Z} .
\end{aligned}$$

We leave (D1) unchanged for now and move to (D2). Using Lemma 6, we can write

$$\text{(D2)} = - \iint D_{F,\varphi}(t, a, \ell) \frac{F_{Z,0}(t \mid a, \ell)}{F_Z(t \mid a, \ell)} (\Gamma_Z - \Gamma_{Z,0})(dt \mid a, \ell) H_{Z,0}(d\ell) .$$

Next we decompose (D3) as the sum (D3a) + (D3b) + (D3c) with

$$\begin{aligned}
\text{(D3a)} &:= - \int \gamma_Z^{-1}[\mu_Z(a_0, \ell) - \psi_{a_0,Z}][F_Z(w \mid a, \ell) - F_{Z,0}(w \mid a, \ell)] \frac{G_Z(dw \mid a, \ell)}{F_Z(w \mid a, \ell)} J_0(da, d\ell) \\
\text{(D3b)} &:= \int \frac{[\mu_Z(a_0, \ell) - \psi_{a_0,Z}]}{\gamma_Z F_Z(w \mid a, \ell)} \left(G_P - G_0\right)(dw \mid a, \ell) J_0(da, d\ell) \\
\text{(D3c)} &:= \frac{\gamma_{Z,0}}{\gamma_Z} \int [\mu_Z(a_0, \ell) - \psi_{a_0,Z}] \left[\frac{F_Z(w \mid a, \ell) - F_{Z,0}(w \mid a, \ell)}{F_Z(w \mid a, \ell)}\right] (G_Z - G_{Z,0})(dw \mid a, \ell) J_{Z,0}(da, d\ell) ,
\end{aligned}$$

and it is important to note that terms without a Z subscript denote the biased observed data version of those parameters. We leave the remaining difference terms unchanged.

We now compute the linear terms $-(P - P_0)\phi_{a_0,P} = P_0\phi_{a_0,P}$, which can be decomposed as the sum (L1) + (L2) + (L3), where we define

$$\text{(L1)} := \int \left\{ \frac{m_Z(t, a_0, \ell)}{m_0(t, a_0, \ell)} \frac{\nu_{Z,0}(t, a_0, \ell)}{\nu_Z(t, a_0, \ell)} \right\} [\varphi_{Z,0}(t, \ell) - \varphi_Z(t, \ell)] F_{Z,0}(dt \mid 1, \ell) H_{Z,0}(d\ell)$$

$$(\text{L2}) := \int D_{Z,\varphi}(t, a_0, \ell) \frac{F_{Z,0}(t \,|\, 1, z)}{F_Z(t \,|\, 1, z)} \frac{\nu_{Z,0}(t, 1, \ell)}{\nu_Z(t, 1, \ell)} \left(\Lambda_Z(dt \,|\, 1, \ell) - \Lambda_{Z,0}(dt \,|\, 1, \ell)\right) H_{Z,0}(d\ell)$$
$$(\text{L2}) := (\text{L2a}) + (\text{L2b}) + (\text{L2c}) \,,$$

such that

$$(\text{L3a}) := -\int \frac{[\mu_Z(a_0, \ell) - \psi_{a_0,Z}]}{\gamma_Z F_Z(w \,|\, a, \ell)} \left(G_P - G_0\right)(dw \,|\, a, \ell) J_0(da, d\ell)$$
$$(\text{L3b}) := -\int [\mu_Z(a_0, \ell) - \psi_{a_0,Z}] \frac{\gamma_Z(a, \ell)}{\gamma_Z} (J_P - J_0)(da, d\ell)$$
$$(\text{L3c}) := -\frac{\gamma_{Z,0}}{\gamma_Z} \int \xi_Z(t, a, \ell) \frac{F_{Z,0}(t \,|\, a, \ell) G_{Z,0}(t \,|\, a, \ell)}{F_Z(t \,|\, a, \ell) G_Z(t \,|\, a, \ell)} \left(\Gamma_Z - \Gamma_{Z,0}\right)(dt \,|\, a, \ell) J_{Z,0}(da, d\ell)$$

and we are using that $P_0(d\ell, da, dw, dt, db) = I(t \leq w)\gamma_{Z,0} P_{Z,0}(d\ell, da, dw, dt, db)$. We again note that terms without a Z subscript are observed data quantities.

We now scrutinize the terms appearing in $R(P, P_0) = (\text{D1}) + (\text{D2}) + (\text{D3}) + (\text{D4}) + (\text{D5}) + (\text{L1}) + (\text{L2}) + (\text{L3})$. First, we observe that $(D4) + (L3b) = 0$ and $(D3b) + (L3a) = 0$. Next we note that

$$R_1(P, P_0) = (D1) + (L1)$$
$$R_2(P, P_0) = (D2) + (L2)$$
$$R_3(P, P_0) = (D3a) + (L3c)$$
$$R_4(P, P_0) = (D3c)$$
$$R_5(P, P_0) = (D5)$$

□

Proof of Theorem 9

When conditions (C1) – (C3) hold for $a_0 \in \{0,1\}$ then, the functional $P \mapsto \Psi_{a_0}(P)$ is a pathwise differentiable parameter with influence function $\phi_{a_0,P}(o)$ and the linearization

$$\Psi_{a_0}(P) - \Psi_{a_0}(P_0) = \int \phi_{a_0,0}(o)(P - P_0)(do) + R(P, P_0)$$

holds for $P \in \mathcal{M}$.

Proof. Using Lemma 2 from Appendix B.2 and Theorem 1 in Chapter 1, the representation of the influence function follows. □

Estimation and inference

We propose the following conditions under which asymptotic inference is valid. We will denote the estimator described in Algorithm 2 as ψ_n^* to clarify that it is an estimating equation estimator, we use ψ_n as in Appendix B.2 to denote the one-step estimator. We denote the data generating distribution implied by the estimated nuisance functions by $P_n := (\varphi_n, F_n, G_n, \pi_n)$ and denote by $P_\infty := (\varphi_\infty, F_\infty, G_\infty, \pi_\infty)$ the limiting distribution of each estimated nuisance function. We suppose the following conditions hold P_0-almost surely.

(C4) The class of functions $\{o \mapsto (\phi_n - \phi_0)(o)\}$ is Donsker.

(C5) There exists a constant $\epsilon > 0$ such that for almost every $y \in (0, \tau(a,\ell))$ and (a,ℓ):

(i) $1/\pi_n(a \,|\, \ell) \leq \epsilon; \quad 1/\pi_\infty(a \,|\, \ell) \leq \epsilon;$

(ii) $1/G_n(y \,|\, a,\ell) \leq \epsilon; \quad 1/G_\infty(y \,|\, a,\ell) \leq \epsilon;$

(iii) $D_{n,\varphi_n}(y,a,\ell)/S_n(y \,|\, a,\ell) \leq \epsilon; \quad D_{\infty,\varphi_\infty}(y,a,\ell)/S_\infty(y \,|\, a,\ell) \leq \epsilon.$

(C6) For almost-every (a,ℓ) and for all $y \in (0, \tau(a,\ell))$ the following hold

(i) $F_\infty(y \,|\, a,\ell) = F_{Z,0}(y \,|\, a,\ell),$

(ii) $G_\infty(y \mid a, \ell) = G_{Z,0}(y \mid a, \ell)$, and $\pi_\infty(\ell) = \pi_{Z,0}(\ell)$;

(iii) $\varphi_\infty(y \mid a, \ell) = \varphi_{Z,0}(y \mid a, \ell)$

(C7) For all $y \in (0, \tau(a, \ell))$, it holds that

$$R(P_n, P_0) = o_P\left(n^{-1/2}\right) ,$$

where $R(P_n, P_0)$ is the remainder evaluated at the distribution implied by the estimation procedure.

Theorem 15. *If two of the three conditions of (C6) are satisfied, then* $\psi_n^* \xrightarrow{p} \psi_0$.

Proof. We first note that the linearization of the weight adjustment function $\phi_{F,P}$ has another equivalent expression $\phi_{G,P}$, where

$$\phi_{F,P}(o) = F_P(w \mid a, \ell)^{-1} + \left\{ \frac{1 - G_P(t \mid a, \ell)}{F_P(t \mid a, \ell) G_P(t \mid a, \ell)} - \int_t^w \frac{1 - G_P(u \mid a, \ell)}{F_P(u \mid a, \ell) G_P(u \mid a, \ell)} \Gamma_P(du \mid a, \ell) \right\}$$
$$\phi_{G,P}(o) = G_P(t \mid a, \ell)^{-1} + \left\{ \frac{1 - F_P(w \mid a, \ell)}{F_P(w \mid a, \ell) G_P(w \mid a, \ell)} - \int_t^w \frac{1 - F_P(u \mid a, \ell)}{F_P(u \mid a, \ell) G_P(u \mid a, \ell)} \Lambda_{G,P}(du \mid a, \ell) \right\}$$

and $\Gamma_P(dt \mid a, \ell) := F_P(dt \mid a, \ell)/F_P(t \mid a, \ell)$ and $\Lambda_{G,P}(dw \mid a, \ell) := -G_P(dw \mid a, \ell)/G_P(w \mid a, \ell)$.

$$\phi_{a_0,1}(\ell, a, t, w, b) = \frac{I(a = a_0)}{\pi_Z(a_0 \mid \ell)} \frac{f_Z(t \mid 1, \ell)}{f_Z(t \mid a_0, \ell)} \frac{1}{G_Z(t \mid a, \ell)} \left[b - \varphi_Z(t, a_0, \ell) \right]$$
$$\phi_{a_0,2}(\ell, a, t, w) = -\frac{I(a = 1)}{\pi_0(1 \mid \ell)} \left\{ \frac{D_{Z,\varphi}(t, a_0, \ell)}{F_Z(t \mid 1, \ell) G_Z(t \mid 1, \ell)} - \int_t^w \frac{D_{Z,\varphi}(u, a_0, \ell)}{G_Z(u \mid a, l) F_Z(u \mid 1, \ell)} \Gamma_Z(du \mid 1, \ell) \right\}$$
$$\phi_{a_0,3}(\ell, a, t, w) = \left[\mu_Z(a_0, \ell) - \psi_{a_0,Z} \right] \phi_{F,P}(1 - G_Z)(o) ,$$

We denote by $U(F_Z, G_Z, \pi_Z, \varphi_Z; o)$ the estimating function evaluated at the specified nuisance functions and at ψ_0. For each piece of the influence function we define a corresponding estimating function $(U_{a_0,1}, U_{a_0,2}, U_{a_0,3})$ such that $U(P; o) := U_1(P; o) - U_0(P; o)$

and where $U_{a_0}(P;o) := U_{a_0,1}(P;o) + U_{a_0,2}(P;o) + U_{a_0,3}(P;o)$. The goal is to show that $P_0U(F_\infty, G_\infty, \pi_\infty, \varphi_\infty)$ under the conditions of (C6). We observe that $E_0[B - \varphi_0(T,L) \,|\, A = a_0, L = \ell, T = t, T \geq W] = 0$ therefore using iterated expectations

$$\mathbb{P}_0 U_{a_0,1}(F_P, G_P, \pi_P, \varphi_0) = 0\,,$$

which shows one case. We next use Equation B.2 to show that

$$\begin{aligned}\mathbb{P}_0 U_{a_0,1}(F_0, G_0, \pi_0, \varphi_P) &= E_0 E_0 \left\{ \frac{f_{Z,0}(T \,|\, 1, L)}{f_{Z,0}(T \,|\, a_0, L)} \frac{I(A = a_0)[\varphi_{Z,0}(T, a_0, L) - \varphi_Z(T, a_0, L)]}{\gamma_{Z,0}\pi_{Z,0}(a_0 \,|\, L) G_{Z,0}(T \,|\, a_0, L)} \middle| A, L \right\} \\ &= \gamma_0 E_{Z,0} \left\{ \mu_{F_0,\varphi_Z}(L) - \mu_0(L) \right\}\,,\end{aligned}$$

and we will later show the cancellation. Next we use the definition of $R_0(u \,|\, a, \ell)$ and we note that

$$\begin{aligned}&E_0 \left\{ \int I(W \leq u \leq T) \frac{D_{0,\varphi_Z}(u, a_0, \ell)}{G_Z(u \,|\, 1, \ell) F_{Z,0}(u \mid 1, \ell)} \Gamma_{Z,0}(du \,|\, 1, \ell) \middle| A = 1, L = \ell \right\} \\ &= \gamma_{Z,0}(a, \ell) \int \frac{G_{Z,0}(u \,|\, a, \ell) D_{0,\varphi_Z}(u, a_0, \ell)}{G_Z(u \,|\, 1, \ell)} \Gamma_{Z,0}(du \,|\, 1, \ell)\,,\end{aligned}$$

and similarly $E_0\{\frac{D_{0,\varphi_Z}(t,a_0,\ell)}{F_0(t \,|\, 1,\ell) G_Z(t \,|\, 1,\ell)} \,|\, A = 1, \ell\} = \gamma_{Z,0}(a,\ell) \int \frac{G_{Z,0}(u \,|\, a,\ell) D_{0,\varphi_Z}(u,a_0,\ell)}{G_Z(u \,|\, 1,\ell)} \Gamma_{Z,0}(du \,|\, 1, \ell)\,,$ which using iterated expectations gives the result $\mathbb{P}_0 U_{a_0,2}(F_0, G_P, \pi_P, \varphi_P) = 0$. Next we study $U_{a_0,2}$ when F_∞ is mispecified, using the same conditional expectation result

$$\mathbb{P}_0 U_{a_0,2}(F_P, G_0, \pi_0, \varphi_0) = E_{Z,0} \left\{ \int \frac{D_{F_Z,\varphi_0}(t, a_0, L) F_{Z,0}(t \,|\, 1, L)}{F_Z(t \,|\, 1, L)} (\Gamma_{Z,0} - \Gamma_Z)(dt \,|\, 1, L) \right\}\,.$$

Using integration by parts we write

$$\begin{aligned}\int \frac{D_{F_Z,\varphi_0}(t, a_0, \ell) F_{Z,0}(t \,|\, 1, \ell)}{F_Z(t \,|\, 1, \ell)} \Gamma_Z(dt \,|\, 1, \ell) = &- \frac{D_{F_Z,\varphi_0}(t, a_0, \ell) F_{Z,0}(t \,|\, 1, \ell)}{F_Z(t \,|\, 1, \ell)} \Big|_0^{\tau_T(a,\ell)} \\ &+ \int F_{Z,0}(t \,|\, 1, \ell) \varphi_{Z,0}(dt, a_0, \ell) + \int \frac{D_{F_Z,\varphi_0}(t, a_0, \ell) F_{Z,0}(t \,|\, 1, \ell)}{F_Z(t \,|\, 1, \ell)} \Gamma_{Z,0}(dt \,|\, 1, \ell)\end{aligned}$$

$$= - D_{F_Z,\varphi_0}(\tau_T(a,\ell),a_0,\ell) + D_{F_0,\varphi_0}(\tau_T(a,\ell),a_0,\ell) = \mu_{F,\varphi_0}(a_0,\ell) - \mu_0(a_0,\ell)\,,$$

which gives the result

$$\mathbb{P}_0 U_{a_0,2}(F_P,G_0,\pi_0,\varphi_0) = \gamma_0 E_{Z,0}[\mu_{F_Z,\varphi_0}(L) - \mu_0(L)]\,,$$

We split the problem into each case separately for $U_{a_0,3}$. First using the same argument as $U_{a_0,2}$ we can show that

$$\begin{aligned}\mathbb{P}_0 U_{a_0,3}(F_0,G_P,\pi_P,\varphi_0) &= E_0\left\{[\mu_0(L) - \psi_{a_0,0}]E_0[\phi_{F,0}(W,T,A,L)\,|\,A,L]\right\}\\ &= -\gamma_0 E_{Z,0}\left\{\mu_0(L) - \psi_0\right\} = 0\,.\end{aligned}$$

We use the same argument but use the representation ϕ_G to show that

$$\begin{aligned}\mathbb{P}_0 U_{a_0,3}(F_P,G_0,\pi_0,\varphi_0) &= E_0\left\{[\mu_0(L) - \psi_{a_0,0}]E_0[\phi_{G,0}(W,T,A,L)\,|\,A,L]\right\}\\ &= -\gamma_0 E_{Z,0}\left\{\mu_{F,\varphi_0}(L) - \mu_0(L)\right\}\,.\end{aligned}$$

Finally, the last follows the same argument as the first case

$$\mathbb{P}_0 U_{a_0,3}(F_0,G_0,\pi_0,\varphi_P) = -\gamma_0 E_{Z,0}\left\{\mu_{F_0,\varphi_P}(L) - \mu_0(L)\right\}\,.$$

We now show that $P_0U(P_\infty) = 0$ by arguing that $P_0U_{a_0}(P_\infty) = 0$ when two of the three conditions hold.

(Case1) $(F_\infty,G_\infty,\pi_\infty) = (F_{Z,0},G_{Z,0},\pi_{Z,0})$ then

$$P_0 U_{a_0}(P_\infty) = \gamma_0 E_{Z,0}\left\{\mu_{F_0,\varphi_\infty}(L) - \mu_0(L)\right\} - \gamma_0 E_{Z,0}\left\{\mu_{F_0,\varphi_\infty}(L) - \mu_0(L)\right\} = 0$$

(Case2) $(G_\infty, \pi_\infty, \varphi_\infty) = (G_{Z,0}, \pi_{Z,0}, \varphi_{Z,0})$ then

$$P_0 U_{a_0}(P_\infty) = \gamma_0 E_{Z,0}\left\{\mu_{F_0,\varphi_\infty}(L) - \mu_0(L)\right\} - \gamma_0 E_{Z,0}\left\{\mu_{F_0,\varphi_\infty}(L) - \mu_0(L)\right\} = 0$$

(Case3) $(F_\infty, \varphi_\infty) = (F_{Z,0}, \varphi_{Z,0})$ then

$$P_0 U_{a_0}(P_\infty) = 0$$

and this completes the robustness result. □

Proof of Theorem 10

When conditions (C4) – (C7) hold, the one-step estimator is asymptotically linear

$$n^{1/2}\left[\psi_n^* - \psi_0\right] \xrightarrow{\mathrm{d}} N(0, \sigma_0^2)$$

where $\sigma_0^2 = E_0\left[\phi_{1,0}(O)^2 + \phi_{0,0}(O)^2 - 2\phi_{0,0}(O)\phi_{0,0}(O)^2\right]$, and the estimating equations estimator ψ_n^* is asymptotically equivalent to the one-step estimator ψ_n.

Proof. Using Theorem 8, we can decompose the one-step estimator as

$$\psi_n - \psi_0 = (\mathbb{P}_n - \mathbb{P}_0)\phi_{a_0,0} + (\mathbb{P}_n - \mathbb{P}_0)(\phi_{a_0,P_n} - \phi_{a_0,0}) + R(P_n, P_0)$$

Condition (C4) and Lemma 19.24 of Van der Vaart (2000) ensures that $(\mathbb{P}_n - \mathbb{P}_0)(\phi_{a_0,P_n} - \phi_{a_0,0}) = o_P\left(n^{-1/2}\right)$ and that

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P}_0)\phi_{a_0,0} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \sigma_0^2)\,.$$

Condition (C7) ensures that $R(P_n, P_0) = o_P\left(n^{-1/2}\right)$, which completes the proof. □

Lemma 10. *When conditions (C4) – (C7) hold and if* $\frac{1}{n}\sum_{i=1}^{n}\{\gamma_n^{-1}\phi_{F,n}(O_i)-1\} = O_P\left(n^{-1/2}\right)$ *then, the estimating equations estimator* ψ_n^* *defined in Algorithm 2 has the same limiting distribution as the one-step estimator* ψ_n*. Further the difference between the two estimators is*

$$\psi_n - \psi_n^* = \frac{1}{n}\sum_{i=1}^{n}\left\{\gamma_n^{-1}\phi_{F,n}(O_i)-1\right\}\left\{\psi_n^* - \hat{\psi}\right\},$$

where $\hat{\psi}$ *is a consistent plug-in for* ψ_0*.*

Proof. We define $\phi_{1,n} := \phi_{1,1,n} - \phi_{0,1,n}$ and $\phi_{2,n} := \phi_{1,2,n} - \phi_{0,2,n}$ as the first two contributions of the influence function of $\Psi_1(P_n)$ and $\Psi_0(P_n)$. For a choice plugin $\hat{\psi}$, the one step estimator can be written as

$$\begin{aligned}\psi_n &= \hat{\psi} + \frac{1}{n}\sum_{i=1}^{n}\gamma_n^{-1}\left[\phi_{1,n}(O_i) + \phi_{2,n}(O_i) + \left\{[\mu_n(1,L_i)-\mu_n(0,L_i)] - \hat{\psi}\right\}\phi_{F,n}(O_i)\right] \\ &= \hat{\psi} + \frac{1}{n}\sum_{i=1}^{n}\gamma_n^{-1}\left[-\left\{[\mu_n(1,L_i)-\mu_n(0,L_i)] - \psi_n^*\right\}\phi_{F,n}(O_i) + \left\{[\mu_n(1,L_i)-\mu_n(0,L_i)] - \hat{\psi}\right\}\phi_{F,n}(O_i)\right],\end{aligned}$$

where we've utilized the definition of the estimating equations estimator. Rearranging the terms completes the proof

$$\psi_n - \psi_n^* = \frac{1}{n}\sum_{i=1}^{n}\left\{\gamma_n^{-1}\phi_{F,n}(O_i)-1\right\}\left\{\psi_n^* - \hat{\psi}\right\},$$

and notably if $\gamma_n := \frac{1}{n}\sum_{i=1}^{n}\phi_{F,n}(O_i)$ then the estimating equations estimator is equivalent to a modified one-step. □

Appendix C

CHAPTER THREE SUPPLEMENTARY MATERIALS AND APPENDIX

Proof of identification

The Residual accelerated failure time model implies the relationship $\phi(z, x, a)$ can be written as in Equation 3.2. This relation is confirmed by the observation that for any increasing continuous function $m : [0, \infty] \mapsto [0, \infty]$

$$\begin{aligned} E[m(T-a) \mid A=a, & X=x, E=1, Z=z] = \\ & \int_0^\infty P(T-a > m^{-1}(u) \mid A=a, X=x, E=1, Z=z)du \\ & = \int_0^\infty P(T-a > \phi^{-1}(z,x,a)m^{-1}(u) \mid T>a, E=0, Z=z)du \\ & = E\left[m\Big((T-a)\phi(z,x,a)\Big) \mid T>a, E=0, Z=z\right]. \end{aligned}$$

This result, much like an Accelerated Failure Time (AFT) model relates $\phi(z, x, a)$ to the ratio of moments of the *residual* lifetime. However we can extend this result if we consider additional assumptions. If we consider the indicator function which is not continuous, namely $m(u) : u \mapsto \mathbb{1}(u > Q_\pi(z, x, a, e))$ where Q_π denotes the conditional π^{th} quantile of T, then

$$\begin{aligned} 1-\pi &= P\left(T-a > Q_\pi(z,x,a,1) \mid A=a, X=x, E=1, Z=z\right) \\ &= P\left(T-a > \phi^{-1}(z,x,a)Q_\pi(z,x,a,1) \mid T>a, E=0, Z=z\right) \\ &= P\left(T-a > Q_\pi(a,z,0) \mid T>a, E=0, Z=z\right), \end{aligned}$$

with $Q_\pi(a, z, 0)$ denoting the π^{th} residual quantile among the unexposed. Comparing the final two lines of the derivation completes the result.

Proof of Theorem 1

Under the conditions (A1)– (A3) and (B0)– (B2), the acceleration factor can be written as the causal contrast

$$\phi(z, x, a) = \frac{E\left[\, m(T(0) - a) \,|\, T(0) > a, Z = z \,\right]}{E\left[\, m(T(x, a, 1) - a) \,|\, T(0) > a, Z = z \,\right]} \,,$$

where $m : \mathbb{R} \to \mathbb{R}$ is any monotone function.

Proof. We begin with the numerator,

$$E\left[\, m(T(0) - a) \,|\, T(0) > a, Z = z \,\right] = E\left[\, m(T - a) \,|\, T > a, E = 0, Z = z \,\right],$$

when the conditions hold. The denominator is

$$\begin{aligned} E\left[\, m(T(x, a, 1) - a) \,|\, T(0) > a, Z = z \,\right] &= E\left[\, m(T - a) \,|\, T(0) > a, X = x, E = 1, Z = z \,\right] \\ &= E\left[\, m(T - a) \,|\, T > a, X = x, E = 1, Z = z \,\right], \end{aligned}$$

when the exposure occurs at age a. Using the previous identification result completes the proof. □

Proof of Theorem 2

Both Lemma 1 and Theorem 12 are direct consequences of parametric estimation theory. Theorem 5.23 in Van der Vaart (2000) describes the general results and conditions for asymptotic normality of M-estimators.