

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Causal Inference with Post-Treatment Complications: Statistical Strategies for Intercurrent Events

Permalink

<https://escholarship.org/uc/item/93w4r98n>

ISBN

9798189272219

Author

Lu, Sizhu

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Causal Inference with Post-Treatment Complications:
Statistical Strategies for Intercurrent Events

by

Sizhu Lu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Peng Ding, Chair

Professor Peter Bickel

Professor Avi Feller

Professor Sam Pimentel

Spring 2026

Causal Inference with Post-Treatment Complications:
Statistical Strategies for Intercurrent Events

Copyright 2026

by

Sizhu Lu

Abstract

Causal Inference with Post-Treatment Complications: Statistical Strategies for Intercurrent Events

by

Sizhu Lu

Doctor of Philosophy in Statistics

University of California, Berkeley

Professor Peng Ding, Chair

Post-treatment events are common in causal inference in both randomized trials and observational studies. They appear in many scientific problems, including noncompliance, truncation by death, mediation, and surrogate endpoint evaluation, and may reveal important information about treatment response, alter the interpretation of the outcome, or make the outcome unobservable. This dissertation studies three related problems involving post-treatment events, with the common goal of defining scientifically meaningful causal estimands and developing rigorous identification, semiparametric efficiency theory, and estimation and inference procedures.

The first chapter studies principal stratification with continuous post-treatment variables. Principal stratification is a strategy to address these challenges by adjusting for the potential values of the post-treatment variables, defined as the principal strata. It allows for characterizing treatment effect heterogeneity across principal strata and unveiling the mechanism of the treatment's impact on the outcome related to post-treatment variables. However, the existing literature has primarily focused on binary post-treatment variables, leaving the case with continuous post-treatment variables largely unexplored. This gap persists due to the complexity of infinitely many principal strata, which present challenges to both the identification and estimation of causal effects. We fill this gap by providing nonparametric identification and semiparametric estimation theory for principal stratification with continuous post-treatment variables. We propose to use working models to approximate the underlying causal effect surfaces and derive the efficient influence functions of the corresponding model parameters. Based on the theory, we construct doubly robust estimators and implement them in the R package `continuousPCE`.

The second chapter turns to intercurrent events in randomized trials. The analysis of randomized controlled trials is often complicated by intercurrent events (IEs) – events that occur after treatment initiation and affect either the interpretation or existence of outcome

measurements. Examples include treatment discontinuation or the use of additional medications. In two recent clinical trials for systemic lupus erythematosus with complications of IEs, we classify the IEs into two broad categories: effect-informative (e.g., treatment discontinuation due to adverse events or lack of efficacy) and effect-uninformative (e.g., treatment discontinuation due to external factors such as pandemics or relocation). To define a clinically meaningful estimand, we adopt tailored strategies for each category of IEs. For effect-informative IEs, which are often informative about a patient’s outcome, we use the *composite variable strategy* that assigns an outcome value indicative of treatment failure. For effect-uninformative IEs, we apply the *hypothetical strategy*, assuming their timing is conditionally independent of the outcome given treatment and baseline covariates, and hypothesizing a scenario in which such events do not occur. A central yet previously overlooked challenge is the presence of competing IEs, where the first IE censors all subsequent ones. Despite its ubiquity in practice, this issue has not been explicitly recognized or addressed in previous data analyses due to the lack of rigorous statistical methodology. In this paper, we propose a principled framework to formulate the estimand, establish its nonparametric identification and semiparametric estimation theory, and introduce weighting, outcome regression, and doubly robust estimators. We apply our methods to analyze the two systemic lupus erythematosus trials, demonstrating the robustness and practical utility of the proposed framework.

The third chapter studies a different type of combined-strategy estimand motivated by long-term chronic disease trials. In such trials, some IEs such as rescue medication, dose adjustment, or treatment modification are part of real-world treatment practice and are handled well using a treatment policy strategy. Other IEs, such as relocation or administrative withdrawal, are external to the scientific question and are more naturally handled using a hypothetical strategy. Unlike Chapter 2, where the occurrence of effect-informative IEs is treated as failure, here they are preserved as part of the treatment pathway. This leads to a different semi-competing data structure: if the effect-uninformative IE occurs first, it censors the effect-informative IE time and the outcome; if the effect-informative IE occurs first, subsequent information on the effect-uninformative IE and the final outcome may still be available. This structure requires different statistical tools from the competing-event setting. We formulate the corresponding estimand, provide nonparametric identification formulas under clear assumptions, construct several estimators, and derive the efficient influence function to guide robust and efficient estimation.

Together, the three chapters develop a unified perspective on post-treatment events: the role of each post-treatment event should determine the causal estimand, while the observed data structure determines the assumptions and statistical tools needed for identification and estimation.

To my parents and Mogeng.

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Overview	1
2 Principal stratification with continuous post-treatment variables: non-parametric identification and semiparametric estimation	4
2.1 Introduction	4
2.2 Setup and assumptions	6
2.3 Nonparametric identification and projection parameters	11
2.4 Semiparametric estimation based on EIF	17
2.5 Asymptotic theory and statistical inference	19
2.6 Simulation	22
2.7 Application to a randomized experiment	24
2.8 Extensions	26
3 Estimating treatment effects with competing intercurrent events in randomized controlled trials	28
3.1 Intercurrent events in randomized controlled trials	28
3.2 Setup and estimand	34
3.3 Nonparametric identification and basic estimators	37
3.4 A semiparametrically efficient and doubly robust estimator	41
3.5 Analyzing two phase-3 trials	46
3.6 Future directions	50
4 Combining treatment policy and hypothetical strategies to handle semi-competing intercurrent events	53
4.1 Introduction	53
4.2 Setup and nonparametric identification	57
4.3 Basic and augmented estimators	61

4.4	Semiparametric efficiency theory	62
4.5	Simulation	65
4.6	Discussion	67
Bibliography		68
A Appendix to Chapter 2		76
A.1	Additional results	76
A.2	Proofs	85
B Appendix to Chapter 3		116
B.1	Additional technical results	116
B.2	Additional data analysis results	119
B.3	Simulation	125
B.4	Proofs	129
C Appendix to Chapter 4		138
C.1	Proofs	138

List of Figures

2.1	Estimated PCE surface of the credit limit increase on future consumption	25
3.1	Bar plot showing the IE types and proportions	30
3.2	Illustration of three representative scenarios of intercurrent events in the immunology trial	31
4.1	Illustration of five representative scenarios of intercurrent events in the long-term chronic weight management trial	55
B.1	Kaplan–Meier-based cumulative incidence of IEs by treatment arm	124

List of Tables

2.1	Biases and RMSEs of the three estimators $\hat{\beta}_{1,\text{eif}}^1$, $\hat{\beta}_{1,\text{tp+pd}}^1$, and $\hat{\beta}_{1,\text{pd+om}}^1$ with different values of $\rho = 0, 0.2, 0.5$	27
3.1	Summary of the observed IE types and outcome	36
3.2	Treatment effect of Baricitinib on the primary outcome SRI4	47
3.3	Treatment effect of Baricitinib on the primary outcome SRI4 using non-responder rule and hypothetical strategy	49
4.1	Finite-sample performance of the four estimators across nuisance-model regimes. For each estimator, we report Monte Carlo bias and standard deviation (SD) based on 500 Monte Carlo samples with $n = 2000$	67
B.1	Classification of IEs in the two immunology trials	120
B.2	Baseline characteristics and IE summary by treatment arm and trial	123
B.3	Treatment effect of Baricitinib on the primary outcome SRI4 under the proposed hybrid estimand	125
B.4	IE classification robustness check for the primary outcome SRI4	126
B.5	Causal risk ratio and odds ratio for the primary outcome SRI4 under the proposed hybrid estimand	126
B.6	Wall-clock time for full-estimator evaluation	127
B.7	Model choices for the nuisance parameters	127
B.8	Finite sample performance of the four estimators	128

Acknowledgments

It's been years since I first arrived at Berkeley, but I still remember vividly how lucky I felt when I started. I want to express my deepest thanks to my advisor, professors, colleagues, friends, and family who offer their light along this journey. It would not have been possible without them.

First and foremost, I owe my deepest gratitude to my advisor Peng Ding. You introduced me to the world of causal inference and shaped not only my approach to research but also philosophy on life. I am constantly touched by your unwavering passion for research and science, and I have been trying to follow your example. Over the years, you gave me so much guidance, faith, support, and tools to become a better researcher. I will always treasure the memories of our long meetings in Evans 425, the stacks of papers we discussed, and the time we spent chatting, eating, and picking loquats and cherries together. Thank you for teaching me how to forage, and for sharing with me your wisdom, generosity, and joy over the years.

I am also deeply grateful to my committee members. Thank you so much to Sam Pimentel for always keeping your door open to me. Whenever I faced uncertainty or confusion, I knew I could come to your office for advice. You are a deeply wise researcher, and your kindness has meant a lot to me. I am also very grateful to Peter Bickel for our conversations and for his support over the years. I have always been touched by how deeply you care about statistics and about students, and I feel so lucky to have learned from you. Many thanks also to Avi Feller, for your thoughtful suggestions regarding my research and career. Your advice was especially important to me as I was making every big decision about my future. I am very grateful to Ting Ye for introducing me to the fascinating fields of clinical trials and biostatistics, and for welcoming me into her wonderful research group. I have learned so much from your guidance and gained so much from your support.

Several other faculty in the statistics department and at UC Berkeley have also placed huge impacts on my life. I would like to thank Bin Yu, whose advice and support throughout my PhD have meant so much to me. The experiences you shared and your encouragement for women in academia have been deeply inspiring. To me, you are a role model for how to live with both strength and kindness. I also want to thank Erin Hartman for your generous advice and encouragement while I was on the job market. I feel very lucky to have had your guidance during the time. I am also grateful to Vira Semenova for her support. I really enjoyed discussing research with you.

In addition, I would like to thank Zhichao Jiang for his help on our joint projects. I also thank Yanyao Yi, Yongming Qu, and Eli Lilly and Company for their generous support and the opportunity to collaborate. I am also deeply grateful to the administrative staff in our department, especially La Shana and Tanisha, for always being so helpful and patient with all the requests I had. Your support made many difficult moments much easier.

My PhD life would definitely not be nearly as joyful without my peer colleagues in the department. I feel so grateful to my dear friends Arisa Sadeghpour, Karissa Huang, Yaxuan Huang, Seunghoon Paik, Lei Shi, Addison Hu, Tiffany Tang, Melody Huang, Tianyu Guo, Licong Lin, Ruiqi Zhang, Huong Vu, Hyunsuk Kim, Andy Shen, Jimmy Butler, Dan Soriano, and Benji Lu. I have learned so much from you, and I will always be so grateful for your companion. Thank you for filling these years with encouragement, shared struggles, and countless moments of happiness. I'm also very grateful for help and support from Xiudi Li, Yiqun Chen, Yuchen Zhou, and Mingrui Zhang.

I also want to thank my friends Sijia Li, Yue Fang, Shihui Liu, and Chen Ling for their companion and faith in me. I am so lucky to have you always by my side, sharing laughter and tears. You made me know that I was never walking this journey alone.

Finally, I want to thank my parents for their unconditional and selfless love. They have always tried their best to give me everything they could, never doubted my choices, and remained the most constant and powerful source of support in my life. Thank you for loving me so fully and patiently. Thank you to my husband Mogeng for your endless companion and love. You have given me a place where all my joy and sadness can land, freely and fully. With you, I have been able to be vulnerable, ambitious, childish, tired, brave, and completely myself. Thank you for walking beside me through all of this. As always, my special thanks to my grandfather. I know you are always there, and I hope that I have made you proud.

Chapter 1

Overview

Post-treatment events are common in causal inference in both randomized trials and observational studies. They appear in many scientific problems, including noncompliance, truncation by death, mediation, and surrogate endpoint evaluation, and may reveal important information about treatment response, alter the interpretation of the outcome, or make the outcome unobservable. This dissertation studies three related problems involving post-treatment events, with the common goal of defining scientifically meaningful causal estimands and developing rigorous identification, semiparametric efficiency theory, and estimation and inference procedures. This dissertation studies three related problems involving post-treatment events. The common goal is to define causal estimands that match the scientific question and then develop rigorous statistical tools for identification, estimation, and inference. Across the three chapters, the central theme is that the role of a post-treatment event should determine how it enters the estimand, while the observed-data structure determines the assumptions and methods needed for identification and estimation.

The first chapter studies principal stratification with continuous post-treatment variables. Principal stratification is a strategy to address these challenges by adjusting for the potential values of the post-treatment variables, defined as the principal strata. It allows for characterizing treatment effect heterogeneity across principal strata and unveiling the mechanism of the treatment's impact on the outcome related to post-treatment variables. However, the existing literature has primarily focused on binary post-treatment variables, leaving the case with continuous post-treatment variables largely unexplored. This gap persists due to the complexity of infinitely many principal strata, which present challenges to both the identification and estimation of causal effects. We fill this gap by providing nonparametric identification and semiparametric estimation theory for principal stratification with continuous post-treatment variables. We propose to use working models to approximate the underlying causal effect surfaces and derive the efficient influence functions of the corresponding model parameters. Based on the theory, we construct doubly robust estimators and implement them in the R package `continuousPCE`. This chapter is adapted from my joint work with

Professors Zhichao Jiang and Peng Ding, published in the Journal of the Royal Statistical Society, Series B (Statistical Methodology) (Lu et al., 2026).

The second chapter turns to intercurrent events in randomized trials. The analysis of randomized controlled trials is often complicated by intercurrent events (IEs) – events that occur after treatment initiation and affect either the interpretation or existence of outcome measurements. Examples include treatment discontinuation or the use of additional medications. In two recent clinical trials for systemic lupus erythematosus with complications of IEs, we classify the IEs into two broad categories: effect-informative (e.g., treatment discontinuation due to adverse events or lack of efficacy) and effect-uninformative (e.g., treatment discontinuation due to external factors such as pandemics or relocation). To define a clinically meaningful estimand, we adopt tailored strategies for each category of IEs. For effect-informative IEs, which are often informative about a patient’s outcome, we use the *composite variable strategy* that assigns an outcome value indicative of treatment failure. For effect-uninformative IEs, we apply the *hypothetical strategy*, assuming their timing is conditionally independent of the outcome given treatment and baseline covariates, and hypothesizing a scenario in which such events do not occur. A central yet previously overlooked challenge is the presence of competing IEs, where the first IE censors all subsequent ones. Despite its ubiquity in practice, this issue has not been explicitly recognized or addressed in previous data analyses due to the lack of rigorous statistical methodology. In this paper, we propose a principled framework to formulate the estimand, establish its nonparametric identification and semiparametric estimation theory, and introduce weighting, outcome regression, and doubly robust estimators. We apply our methods to analyze the two systemic lupus erythematosus trials, demonstrating the robustness and practical utility of the proposed framework. This chapter is adapted from joint work with Professors Ting Ye and Peng Ding, and with Yanyao Yi, Yongming Qu, and Karen Liu, statisticians at Eli Lilly. The paper is currently under revision and has been resubmitted to the Journal of the American Statistical Association. It is also available on arXiv (Lu et al., 2025).

The third chapter studies a different type of combined-strategy estimand motivated by long-term chronic disease trials. In such trials, some IEs such as rescue medication, dose adjustment, or treatment modification are part of real-world treatment practice and are handled well using a treatment policy strategy. Other IEs, such as relocation or administrative withdrawal, are external to the scientific question and are more naturally handled using a hypothetical strategy. Unlike Chapter 2, where the occurrence of effect-informative IEs is treated as failure, here they are preserved as part of the treatment pathway. This leads to a different semi-competing data structure: if the effect-uninformative IE occurs first, it censors the effect-informative IE time and the outcome; if the effect-informative IE occurs first, subsequent information on the effect-uninformative IE and the final outcome may still be available. This structure requires different statistical tools from the competing-event setting. We formulate the corresponding estimand, provide nonparametric identification formulas under clear assumptions, construct several estimators, and derive the efficient influence function

to guide robust and efficient estimation. This chapter is adapted from an ongoing work with Professors Ting Ye and Peng Ding, and with Yanyao Yi and Yongming Qu at Eli Lilly.

Together, the three chapters develop a unified perspective on post-treatment events: the role of each post-treatment event should determine the causal estimand, while the observed data structure determines the assumptions and statistical tools needed for identification and estimation. Chapter 2 focuses on principal stratification, where the post-treatment variable defines latent causal subgroups. Chapter 3 studies IEs that require combining composite and hypothetical strategies in the presence of competing event times. Chapter 4 studies IEs that require combining treatment policy and hypothetical strategies under a semi-competing event structure. Although the three problems differ in their estimands and observed-data structures, they share a common methodological architecture: each begins by clarifying the causal question, then states assumptions under which the target estimand is identified, and finally develops estimators guided by semiparametric efficiency theory. This dissertation also emphasizes the distinction between estimands and assumptions. The estimand encodes the scientific question: which post-treatment events are part of the treatment effect, which should be removed through a hypothetical intervention, and which define relevant subgroups or composite outcomes. Identification assumptions, by contrast, describe the conditions under which the observed data can recover that estimand. Separating these two layers is essential. It prevents post-treatment events from being treated automatically as missing-data problems and clarifies when different statistical tools are needed.

Chapter 2

Principal stratification with continuous post-treatment variables: nonparametric identification and semiparametric estimation

2.1 Introduction

Principal stratification is an approach to adjusting for post-treatment variables in causal inference. Because the post-treatment variable is affected by the treatment, the adjustment cannot be simply achieved by conditioning on its observed value. Principal stratification, first proposed by Frangakis and Rubin (2002), addresses this issue by conditioning on the principal strata defined by the joint potential values of the post-treatment variable. The causal effects within principal strata, often referred to as the principal stratification average causal effects (PCEs), delineate causal mechanisms through the post-treatment variable while maintaining valid interpretations as subgroup causal effects. Over the past two decades, principal stratification has been applied in a wide range of problems, including noncompliance (Angrist et al., 1996; Frumento et al., 2012; Mealli and Pacini, 2013), truncation by death (Rubin, 2006; Ding et al., 2011; Wang et al., 2017; Yang and Small, 2016), missing data (Frangakis and Rubin, 1999; Mattei et al., 2014), mediation (Rubin, 2004; Gallop et al., 2009; Elliott et al., 2010; Mattei and Mealli, 2011; Daniels et al., 2012; Zigler et al., 2012), surrogate evaluation (Frangakis and Rubin, 2002; Gilbert and Hudgens, 2008; Huang and Gilbert, 2011; Li et al., 2010; Jiang et al., 2016), and vaccine efficacy analysis (Gilbert et al., 2024).

The identification of PCEs presents challenges due to the fundamental problem of causal inference: for each unit, only one potential value of the post-treatment variable is observed.

Theoretical investigations into the identification of PCEs have mostly focused on binary or discrete post-treatment variables (e.g., Cheng and Small, 2006; Imai, 2008; Ding et al., 2011; Mattei and Mealli, 2011; Mealli and Pacini, 2013; Ogburn et al., 2015; Mealli et al., 2016; Jiang et al., 2016; Yang and Small, 2016; Forastiere et al., 2016; Jiang et al., 2022). Jiang and Ding (2021) provide formal identification results for PCEs with a general post-treatment variable. However, their theory requires the existence of additional variables satisfying certain conditional independence assumptions, which can be difficult to meet in practice. In contrast to the lack of theory for continuous post-treatment variables, real-world applications frequently involve continuous post-treatment variables. For example, Efron and Feldman (1991), Jin and Rubin (2008), and Bartolucci and Grilli (2011) consider a noncompliance problem in which the post-treatment variable is the proportion of the drug or placebo actually received and takes continuous values ranging from zero to one. Schwartz et al. (2011) study the causal mechanism of the effect of physical activity on cardiovascular disease mediated by a continuous post-treatment body mass index.

Dichotomizing the post-treatment variable is commonly employed in practice (e.g., Sjölander et al., 2009; Jiang et al., 2022). The dichotomization approach, however, not only discards important information for understanding causal mechanisms but also poses challenges to valid inference due to the arbitrary choice of thresholds. Without dichotomization, the existing literature mostly employs modeling approaches without a formal discussion of identifiability. For instance, Conlon et al. (2014) and Jin and Rubin (2008) assume parametric models for the potential values of both the outcome and the post-treatment variable and specify informative priors for the parameters. Bartolucci and Grilli (2011) adopt similar outcome models but propose the use of a copula to relax restrictions on potential values of the post-treatment variable. Schwartz et al. (2011) further relax the restrictions on the post-treatment variable under a Dirichlet process mixture model within a semiparametric Bayesian framework. Subsequently, many more complex models have been proposed for the application of principal stratification in various scenarios with continuous post-treatment variables (e.g., Zigler et al., 2012; Comment et al., 2019; Kim et al., 2019), all of which rely heavily on modeling assumptions. However, estimates from these models may be unstable when their identifiability is not assured. Therefore, there is a need for methods capable of handling continuous post-treatment variables without relying on strong modeling assumptions.

We develop a framework for principal stratification with a continuous post-treatment variable. In Section 2.2, we introduce the notation and the identification assumptions. In Section 2.3, we establish the nonparametric identification of PCEs under the principal ignorability assumption, which is commonly used in practice (e.g., Egleston et al., 2009; Jo and Stuart, 2009; Jo et al., 2011; Stuart and Jo, 2015; Feller et al., 2017; Mattei et al., 2023; Cheng et al., 2024) and has been utilized in Jiang et al. (2022) to establish the identification of PCEs for a binary post-treatment variable. We also use a copula function to obtain the joint distribution of potential values of the post-treatment variable from their marginal dis-

tributions. It can be viewed as a generalization of the monotonicity assumption for a binary post-treatment variable (Jiang and Ding, 2021) and has also been popular in various settings (e.g., Roy et al., 2008; Bartolucci and Grilli, 2011; Daniels et al., 2012; Conlon et al., 2014; Yang and Ding, 2018).

The continuous post-treatment variable generates infinitely many principal strata, making the estimation of PCEs a difficult task. We propose to project the PCEs onto pre-specified working models to approximate the underlying complex causal effect surfaces. The projection parameter recovers the model-based parameter when the working model is correctly specified and remains a well-defined causal parameter even if the working model is misspecified. This approach has been adopted in other contexts of causal inference (Neugebauer and Laan, 2007; Kennedy et al., 2019; Ai et al., 2021; Ye et al., 2023; Boileau et al., 2025). Based on the identification formulas, we propose two estimators that depend on different components of the observed data distribution.

The existence of multiple estimators motivates the derivation of a more efficient estimator. In Section 2.4, we derive the efficient influence function (EIF) of the PCEs to motivate a new estimator. In Section 2.5, we study the asymptotic properties of the proposed estimator based on EIF. The estimator is doubly robust in that it is consistent if either the treatment probability or the outcome mean model is correctly specified, contingent on the correct specification of the principal density model, and is semiparametrically efficient if all models are correctly specified with requirements on their convergence rates. In Section 2.6, we evaluate the finite sample performance of various estimators through simulations. In Section 2.7, we apply the estimators to analyze a randomized experiment in the youth labor market in Uganda. In Section 2.8, we discuss some extensions.

We use the following notation. Let $\|r\|_2 = \{\int r(v)^2 dP(v)\}^{1/2}$ denote the $L_2(P)$ norm where $P(\cdot)$ denotes the distribution of the observed data. Let $p(\cdot)$ denote the probability density function of continuous random variables. Write $b_n = O_P(a_n)$ if b_n/a_n is bounded in probability and $b_n = o_P(a_n)$ if b_n/a_n converges to 0 in probability. Let $1(\cdot)$ be the indicator function. Let $\dot{f}(\eta) = \partial f(\eta)/\partial \eta$ denote the gradient of function f and $\ddot{f}(\eta) = \partial^2 f(\eta)/\partial \eta \partial \eta^\top$ denote the Hessian matrix of f with respect to η . Let $\dot{c}_u(u, v)$ and $\dot{c}_v(u, v)$ denote the partial derivative of a bivariate function $c(u, v)$ with respect to u and v , respectively.

2.2 Setup and assumptions

2.2.1 Setup

We generalize the principal stratification method in Frangakis and Rubin (2002) to the case when the post-treatment variable is continuous. For unit $i = 1, \dots, n$, let Z_i be the treatment, S_i the post-treatment variable, Y_i the outcome, and X_i a set of pre-treatment

covariates. We invoke the stable unit treatment value assumption and let Y_{zi} and S_{zi} denote the potential values of the outcome and the post-treatment variable if the treatment were set to $Z_i = z$, respectively. Then, $S_i = Z_i S_{1i} + (1 - Z_i) S_{0i}$ and $Y_i = Z_i Y_{1i} + (1 - Z_i) Y_{0i}$. We assume all variables are identically and independently sampled from a super population and drop the subscript i for simplicity.

We revisit four illustrative examples of real-world applications with a continuous post-treatment variable.

Example 2.1. *Efron and Feldman (1991) and Jin and Rubin (2008) study the treatment effect of cholestyramine on lowering cholesterol levels. Let Z denote the binary treatment where patients with $Z = 1$ if a patient is assigned active pills of the drug and $Z = 0$ if assigned placebo pills. Let Y denote the cholesterol reduction. Partial noncompliance arises in this trial because the patients may take only a proportion of the assigned dose. The post-treatment variable S represents the proportion of the assigned dose taken, ranging from zero to one. To measure the causal effect of the drug that is actually taken, it is crucial to adjust for S in the analysis.*

Example 2.2. *Gilbert and Hudgens (2008) and Gilbert et al. (2015) study the surrogate endpoint evaluation in randomized controlled trials, using an HIV vaccine trial and a herpes zoster vaccine trial, respectively. The binary treatment variable Z is the randomized assignment of the vaccine or placebo. The post-treatment variable S is the potential surrogate endpoint, which is some vaccine-induced immune responses, e.g., the 50% neutralization titers for HIV gp120 in the HIV trial and the varicella zoster virus antibody titer in the herpes zoster trial. The outcome Y is the infection of the disease. Gilbert and Hudgens (2008) and Gilbert et al. (2015) propose using principal stratification criteria to evaluate how well the post-treatment variable S serves as a surrogate for the outcome.*

Example 2.3. *Schwartz et al. (2011) study the effect of physical activity on cardiovascular disease and how the effect varies across levels of body mass index (BMI). In this study, the binary treatment variable Z denotes physical activity with $Z = 1$ for high-level exercisers and $Z = 0$ for low-level exercisers. The outcome Y is the indicator of cardiovascular disease. The scientific question of interest concerns the heterogeneity of the effect of physical activity on cardiovascular disease across different BMI level subgroups, denoted by a continuous post-treatment variable S . To address this, Schwartz et al. (2011) employ principal stratification to adjust for S using a Bayesian semiparametric model.*

Example 2.4. *Yin (2026) conduct a randomized experiment in China to learn the effect of credit supply on consumption behavior. They randomly assign credit card users to treatment and control groups, denoted by $Z = 1$ and $Z = 0$, respectively. Users in the treatment group are offered an increase in their available credit, while users in the control group maintain their current credit limits. The outcome of interest Y is the change in their average monthly consumption following the treatment. To assess how the effect of credit increase*

varies across users with different expectations about their future outcome, we adjust for the expected changes in monthly income six months after the experiment, defined as a continuous variable S . We will analyze the data in Section 2.7.

The principal stratum U is defined by the joint potential values of S , i.e., $U = (S_1, S_0)$. When S is continuous, the number of principal strata is infinity. For ease of notation, we write $U = (s_1, s_0)$ to denote the principal stratum ($S_1 = s_1, S_0 = s_0$). Our goal is to identify and estimate the PCE surface as a function of s_1 and s_0 :

$$\tau(s_1, s_0) = E\{Y_1 - Y_0 \mid U = (s_1, s_0)\} = m_1(s_1, s_0) - m_0(s_1, s_0),$$

where $m_z(s_1, s_0) = E\{Y_z \mid U = (s_1, s_0)\}$ are the two corresponding average potential outcome surfaces for $z = 0, 1$.

2.2.2 Assumptions

In this subsection, we discuss two crucial identification assumptions. We consider the scenario in which treatment ignorability holds for both the outcome and the post-treatment variable.

Assumption 2.1 (Treatment ignorability). $Z \perp\!\!\!\perp (S_0, S_1, Y_0, Y_1) \mid X$.

Assumption 2.1 is a standard identification assumption for the average causal effects $\tau_Y = E(Y_1 - Y_0)$ and $\tau_S = E(S_1 - S_0)$. It holds in randomized experiments by design. However, it is untestable in observational studies and requires the collection of all confounders between the treatment and the outcome, as well as those between the treatment and the post-treatment variable (Rosenbaum and Rubin, 1983). Under Assumption 2.1, we can identify the average causal effect of the treatment on the outcome, $\tau_Y = E\{E(Y \mid Z = 1, X) - E(Y \mid Z = 0, X)\}$, and that on the post-treatment variable, $\tau_S = E\{E(S \mid Z = 1, X) - E(S \mid Z = 0, X)\}$. However, the PCEs are generally not identifiable since we cannot observe S_1 and S_0 simultaneously.

When S is binary, the monotonicity assumption, i.e., $S_1 \geq S_0$, is sufficient for identifying the distribution of U (e.g., Angrist et al., 1996; Jiang et al., 2022), which serves as the basis for identifying the PCEs. With a continuous S , the distribution of U becomes a density function and is not identifiable even with monotonicity.

Under Assumption 2.1, we can identify only the conditional distribution of S_z given X by $p(S_z \mid X) = p(S \mid Z = z, X)$ for $z = 0, 1$. Therefore, we propose to use a copula function $\mathbb{C}_\rho(\cdot, \cdot)$ to establish a mapping between the joint distribution and the conditional distributions of the marginal potential values S_z :

$$P(S_1 \leq s_1, S_0 \leq s_0 \mid X) = \mathbb{C}_\rho\{F_1(s_1, X), F_0(s_0, X)\},$$

where the parameter ρ indexes the association between S_1 and S_0 given X (see Jiang and Ding, 2021; Roy et al., 2008; Bartolucci and Grilli, 2011; Yang and Ding, 2018; Sun et al., 2024 for similar copula approaches), and $F_z(s_z, X) = P(S_z \leq s_z \mid X)$ denotes the conditional cumulative distribution function of S_z given X for $z = 0, 1$.

The literature takes different perspectives on the parameter ρ , which governs the association in the copula function and thereby influences the identification of the distribution of U . For instance, Efron and Feldman (1991) assume that S_1 and S_0 for each unit are at the same percentile and impute the counterfactual from the marginal distributions of S_z , effectively specifying ρ as a known constant. In contrast, Bartolucci and Grilli (2011) treat ρ as a parameter and propose a parametric model to identify and estimate it. More recently, Sun et al. (2024) leverage repeated measurements of S for identification and estimation. However, these approaches rely on strong parametric models or additional information.

In this study, we assume that the copula function is known with a pre-specified sensitivity parameter ρ . This assumption, though strong, allows us to focus on scenarios where the joint density of (S_1, S_0) is identifiable. Without this assumption, it would be necessary to impose strong modeling assumptions on other parts of the observed data distribution, which we avoid in this work. Exploring alternative approaches without this assumption is left for future research. We denote the copula function as $\mathbb{C}(\cdot, \cdot)$ and omit the subscript ρ for simplicity.

In practice, the true value of ρ is typically unknown. Therefore, we recommend treating ρ as a sensitivity parameter and conducting sensitivity analyses to explore the impact of different ρ values (see Section 2.7 for an application). Given a pre-specified value of ρ , we can identify the density of U given X by

$$\begin{aligned} p\{U = (s_1, s_0) \mid X\} &= \frac{\partial^2 P(S_1 \leq s_1, S_0 \leq s_0 \mid X)}{\partial s_1 \partial s_0} = \frac{\partial^2 \mathbb{C}(F_1(s_1, X), F_0(s_0, X))}{\partial s_1 \partial s_0} \\ &= c(F_1(s_1, X), F_0(s_0, X)) p_1(s_1, X) p_0(s_0, X), \end{aligned} \quad (2.1)$$

where $c(u, v) = \partial^2 \mathbb{C}(u, v) / \partial u \partial v$ denotes the copula density function that equals the second order cross derivative of the copula function $\mathbb{C}(\cdot, \cdot)$, and $p_z(s_z, X) = p(S = s_z \mid Z = z, X)$ is the density function of S given $Z = z$ and X at $S = s_z$ for $z = 0, 1$. For a continuous S , $p\{U = (s_1, s_0) \mid X\}$ is the density function of U given X . We define $e(s_1, s_0, X) = p\{U = (s_1, s_0) \mid X\}$ and refer to it as the *principal density*. With a binary S , it reduces to the principal score in Follmann (2000), Hill et al. (2002), Joffe et al. (2007), Ding and Lu (2017), Feller et al. (2017), and Jiang et al. (2022).

With the known copula function and association parameter ρ , there is a one-to-one mapping between $e(s_1, s_0, X)$ and $\{p_1(s_1, X), p_0(s_0, X)\}$, and thus we will refer to both as principal density. By averaging the principal density over X , we can obtain the marginal density of U as $e(s_1, s_0) = p\{U = (s_1, s_0)\} = E\{e(s_1, s_0, X)\}$.

Next, we generalize the principal ignorability assumption with a binary post-treatment variable (Jo and Stuart, 2009; Jo et al., 2011; Stuart and Jo, 2015; Ding and Lu, 2017; Jiang et al., 2022) to the following assumption with a continuous post-treatment variable.

Assumption 2.2 (Principal ignorability). $E\{Y_1 \mid U = (s_1, s_0), X\} = E\{Y_1 \mid U = (s_1, s'_0), X\}$ for any s_0, s'_0 and $E\{Y_0 \mid U = (s_1, s_0), X\} = E\{Y_0 \mid U = (s'_1, s_0), X\}$ for any s_1, s'_1 .

Assumption 2.2 requires that, conditional on the covariates, the expectation of the potential outcome under treatment, Y_1 , does not vary across principal strata with the same value of S_1 . Similarly, the expectation of the potential outcome under control, Y_0 , does not vary across principal strata with the same value of S_0 . This assumption restricts only the expected values of potential outcomes, making it a weaker condition than the conditional independence assumption: $Y_z \perp\!\!\!\perp S_{1-z} \mid S_z, X$ for $z = 0, 1$.

Although we do not impose the stronger conditional independence assumption $Y_z \perp\!\!\!\perp S_{1-z} \mid S_z, X$ for $z = 0, 1$, considering it helps to interpret Assumption 2.2 and facilitates comparison with related assumptions. Under Assumption 2.1, the conditional independence $Y_z \perp\!\!\!\perp S_{1-z} \mid S_z, X$ is equivalent to $Y \perp\!\!\!\perp S_{1-z} \mid Z = z, S = s, X$. Therefore, principal ignorability requires that given the treatment, observed post-treatment variable, and covariates, the counterfactual post-treatment variable S_{1-z} is ignorable for the distribution of Y .

Principal ignorability resembles the cross-world independence assumption in causal mediation analysis, which requires the potential outcome to be independent of S_{1-z} given $Z = z$ and X . However, unlike principal ignorability, cross-world independence does not condition on the observed post-treatment variable S . While principal ignorability and cross-world independence do not generally imply each other, conditioning on additional variables is often believed to improve the plausibility of conditional independence assumptions (Rubin, 2007; Vander Weele and Shpitser, 2011). Moreover, Forastiere et al. (2018) show that when S is binary and satisfies monotonicity, cross-world independence implies a weaker version of principal ignorability. Therefore, the covariates used to justify cross-world independence in mediation analysis are often suitable for ensuring principal ignorability. As a general guideline, we recommend including all covariates that confound the relationship between the post-treatment variable and the outcome.

Principal stratification also differs from the mediation analysis conceptually. Mediation analysis requires the mediator to be a well-defined intervention, whereas principal stratification only requires the treatment to be a well-defined intervention. Furthermore, mediation analysis involves nested potential outcomes that may not correspond to any feasible experimental setting. In contrast, principal stratification considers only the potential values defined by the treatment.

Under Assumption 2.1, Assumption 2.2 is equivalent to

$$E\{Y_1 \mid U = (s_1, s_0), Z = 1, S = s_1, X\} = E\{Y_1 \mid U = (s_1, s'_0), Z = 1, S = s_1, X\}, \quad (2.2)$$

$$E\{Y_0 \mid U = (s_1, s_0), Z = 0, S = s_0, X\} = E\{Y_0 \mid U = (s'_1, s_0), Z = 0, S = s_0, X\}. \quad (2.3)$$

The observed stratum ($Z = 1, S = s_1$) is a mixture of the principal strata $U = (s_1, s_0)$ for all possible values of s_0 . Therefore, the conditional expectations in (2.2) simplify to the outcome mean $\mu_1(s_1, X) = E(Y \mid Z = 1, S = s_1, X)$. Similar argument applies to the observed stratum ($Z = 0, S = s_0$); the conditional expectations in (2.3) simplify to the outcome mean $\mu_0(s_0, X) = E(Y \mid Z = 0, S = s_0, X)$.

2.3 Nonparametric identification and projection parameters

2.3.1 Identification of PCEs

In this subsection, we discuss the nonparametric identification of the PCE surfaces. From (2.2) and (2.3), we can identify the surfaces of the average potential outcomes $m_z(s_1, s_0)$ by averaging the outcome means $\mu_z(s_z, X)$ over the distribution of X given U . This leads to the identification formulas in the following theorem.

Theorem 2.1. *Suppose that Assumptions 2.1 and 2.2 hold and $0 < \pi(x) < 1$ for all x in the support of X . Assume that $e(s_1, s_0, X)$ is identifiable. For any (s_1, s_0) such that $e(s_1, s_0) > 0$, we have*

$$E\{Y_z \mid U = (s_1, s_0)\} = E\left\{\frac{e(s_1, s_0, X)}{e(s_1, s_0)}\mu_z(s_z, X)\right\} \quad \text{for } z = 0, 1. \quad (2.4)$$

Theorem 2.1 expresses the distribution of X given U as a function of the principal density. Thus, the formulas rely on both the principal density and the outcome mean, extending those in Jiang et al. (2022) based on the principal score and the outcome mean models. In Theorem 2.1, we do not need $e(s_1, s_0) > 0$ for all (s_1, s_0) because we focus only on the strata with $e(s_1, s_0) > 0$.

The existence of alternative identification formulas in Jiang et al. (2022) motivates us to explore other identification formulas with a continuous S . The following theorem presents the identification formula based on the treatment probability and principal density.

Theorem 2.2. *Suppose that Assumptions 2.1 and 2.2 hold and $0 < \pi(x) < 1$ for all x in the support of X . Assume that the potential outcome surface $m_z(s_1, s_0)$, the principal density*

$e(s_1, s_0, X)$ and $p_z(s_z, X)$, and the marginal density $e(s_1, s_0)$ are continuous in s_1 and s_0 for $z = 0, 1$. For any (s_1, s_0) such that $e(s_1, s_0) > 0$, we have

$$\begin{aligned} E\{Y_1 \mid U = (s_1, s_0)\} &= \lim_{\epsilon \rightarrow 0} E\{Y_1 \mid S_1 \in B_\epsilon(s_1), S_0 \in B_\epsilon(s_0)\} \\ &= \lim_{\epsilon \rightarrow 0} E\left[\frac{P\{S_1 \in B_\epsilon(s_1), S_0 \in B_\epsilon(s_0) \mid X\}}{P\{S_1 \in B_\epsilon(s_1), S_0 \in B_\epsilon(s_0)\}} \frac{1\{S_1 \in B_\epsilon(s_1)\}}{P\{S_1 \in B_\epsilon(s_1) \mid X\}} \frac{ZY}{\pi(X)}\right] \\ &= \lim_{\epsilon \rightarrow 0} E\left\{\frac{e(s_1, s_0, X)}{e(s_1, s_0)} \frac{1(s_1 - \epsilon \leq S < s_1 + \epsilon)}{2\epsilon \cdot p_1(s_1, X)} \frac{ZY}{\pi(X)}\right\}, \end{aligned} \quad (2.5)$$

where $B_\epsilon(s_z) = \{s \in \mathbb{R} \mid s_z - \epsilon \leq s < s_z + \epsilon\}$ is the ϵ -ball centered at s_z for $z = 0, 1$.

We can obtain a similar identification formula for $E\{Y_0 \mid U = (s_1, s_0)\}$. Equation (2.5) expresses the average potential outcome under treatment in principal stratum $U = (s_1, s_0)$ as the limit of a weighted average of the outcome for the treated units. The weights have two main components: $1/\pi(X)$ corresponds to the treatment probability, while $e(s_1, s_0, X)/\{e(s_1, s_0)p_1(s_1, X)\}$ corresponds to the principal density. The indicator function $1(s_1 - \epsilon \leq S < s_1 + \epsilon)$ limits the identification formula to units with S in the neighborhood of s_1 , indicating that the conditional mean $E\{Y_1 \mid U = (s_1, s_0)\}$ is not a Fréchet differentiable parameter (for technical details, see Bickel et al., 1993). Similar argument applies to $E\{Y_0 \mid U = (s_1, s_0)\}$. Therefore, no regular asymptotically linear estimator exists for $\tau(s_1, s_0)$. Estimating $\tau(s_1, s_0)$ based on Theorem 2.2 requires nonparametric local methods, such as kernel methods, and the estimators cannot converge at the parametric rate.

Unlike Jiang et al. (2022), it is difficult to obtain an identification formula based only on the treatment probability and outcome model with a continuous S . Achieving this requires replacing the principal density in the identification formula of Theorem 2.1 with the treatment probability. However, the principal density $p\{U = (s_1, s_0) \mid X\} = c(F_1(s_1, X), F_0(s_0, X))p(S_1 = s_1 \mid X)p(S_0 = s_0 \mid X)$ is not a linear function of the conditional distributions of S_z as in the case of binary S . Instead, it depends on the copula function $c(\cdot, \cdot)$, which generally takes a complicated form. Therefore, the principal density is crucial for deriving the identification formulas of the PCE surface when the post-treatment variable S is continuous. We will further demonstrate this fact later in Section 2.4.

2.3.2 Projection parameters

Using the identification formulas in Theorems 2.1 and 2.2, we can construct nonparametric estimators for the average potential outcome surfaces $m_1(s_1, s_0)$ and $m_0(s_1, s_0)$. However, their efficient influence functions with finite variance do not exist (Bickel et al., 1993), leading to substantial finite-sample variability in their estimates. In addition, the interpretation of these average potential outcome surfaces can be challenging in practical applications due to their two-dimensional nature.

To improve estimation precision and achieve better visualization and interpretation, we project the mean surface $m_z(s_1, s_0)$ onto a pre-specified working model $f_z(s_1, s_0; \eta_z)$, where $\eta_z \in \mathbb{R}^q$ is a finite-dimensional parameter (Neugebauer and Laan, 2007; Ogburn et al., 2015; Kennedy et al., 2019; Ai et al., 2021; Ye et al., 2023; Boileau et al., 2025). In particular, we define our parameters of interest as

$$\eta_1 = \arg \min_{\eta \in \mathbb{R}^q} E \left[w_1(S_1, S_0) \{m_1(S_1, S_0) - f_1(S_1, S_0; \eta)\}^2 \right], \quad (2.6)$$

$$\eta_0 = \arg \min_{\eta \in \mathbb{R}^q} E \left[w_0(S_1, S_0) \{m_0(S_1, S_0) - f_0(S_1, S_0; \eta)\}^2 \right], \quad (2.7)$$

where $w_1(s_1, s_0)$ and $w_0(s_1, s_0)$ are some user-specified weight functions. The formulation in (2.6) and (2.7) does not require a correct specification of the working model, and the pre-specified weights and the working models could be different for the two mean surfaces. We can view the projection coefficient η_z as the solution to the nonlinear weighted least squares of the true population parameters $m_z(S_1, S_0)$ for $z = 0, 1$. Constant weight functions, where $w_1(s_1, s_0) = w_0(s_1, s_0) = 1$, lead to uniform weighting across the entire support of (S_1, S_0) . In contrast, if domain knowledge indicates that particular values of (s_1, s_0) are more important, we can employ non-constant weight functions to assign greater weights to them accordingly. From (2.6) and (2.7), η_z is the parameter that minimizes the discrepancy between the true mean surface m_z and the working model f_z , with the discrepancy defined by the choice of the weighting function w_z .

The identification of the projected parameters η_1 and η_0 follows directly from the identification of the average potential outcome surfaces in Theorems 2.1 and 2.2 if the optimization problems in (2.6) and (2.7) have unique solutions, which depends on the functional form of f_z and the support of $e(s_1, s_0)$.

Remark. We can also project the PCE surface $\tau(s_1, s_0)$ using a similar approach:

$$\eta_\tau = \arg \min_{\eta \in \mathbb{R}^q} E \left[w_\tau(S_1, S_0) \{\tau(S_1, S_0) - f_\tau(S_1, S_0; \eta)\}^2 \right]. \quad (2.8)$$

The quantities η_τ and $\eta_1 - \eta_0$ are closely related under some specific choices of f_1 , f_0 , and f_τ , but are different in general. An advantage of introducing the separately projected parameters in (2.6) and (2.7) is that we can also consider nonlinear causal estimands, for instance, $m_1(S_1, S_0)/m_0(S_1, S_0)$, and approximate it by the ratio between two projected mean surfaces $f_1(S_1, S_0; \eta_1)/f_0(S_1, S_0; \eta_0)$. Because of this, we focus on the separately projected parameters η_1 and η_0 defined in (2.6) and (2.7) in the main text. We relegate the result for η_τ and the discussion on the special case when these two ways of projection are equivalent to Section A.1.5 in the Supplementary Material.

2.3.3 Estimation of the projection parameters

Next, we introduce two estimators of the projected parameters corresponding to the two identification formulas in Theorems 2.1 and 2.2. These estimators rely on different components

of the observed data distribution. For simplicity, we use $\hat{\pi}$, $\hat{e}$, and $\hat{\mu}$ to denote the estimated treatment probability, principal density, and outcome mean, respectively, suppressing their dependence on the covariates X . As established earlier, we only need to estimate the conditional distribution of S_z for $z = 0, 1$ and apply the copula function to obtain the principal density estimator. Throughout this subsection, we assume the optimization problems in (2.6) and (2.7) have unique solutions.

We consider the estimation of η_z for $z = 0, 1$. From (2.6) and (2.7), we can write the projected quantity η_z as

$$\begin{aligned}\eta_z &= \arg \min_{\eta} E \left[w_z(S_1, S_0) \{m_z(S_1, S_0) - f_z(S_1, S_0; \eta)\}^2 \right] \\ &= \arg \min_{\eta} \iint w_z(s_1, s_0) \{m_z(s_1, s_0) - f_z(s_1, s_0; \eta)\}^2 e(s_1, s_0) ds_1 ds_0\end{aligned}\quad (2.9)$$

for $z = 0, 1$. Taking the first order derivative with respect to η on the objective function in (2.9), we have

$$\iint w_z(s_1, s_0) \dot{f}_z(s_1, s_0; \eta_z) e(s_1, s_0) \{m_z(s_1, s_0) - f_z(s_1, s_0; \eta_z)\} ds_1 ds_0 = 0. \quad (2.10)$$

Equation (2.10) motivates an estimating equation based on Theorem 2.1.

Lemma 2.1 (Estimating equation based on principal density and outcome mean models). *Suppose that Assumptions 2.1 and 2.2 hold and $0 < \pi(x) < 1$ for all x in the support of X . For $z = 0, 1$, define*

$$\begin{aligned}D_{z, \text{pd+om}}(Y, S, Z, X; \eta, e, \mu_z) \\ = \iint w_z(s_1, s_0) \dot{f}_z(s_1, s_0; \eta) e(s_1, s_0, X) \{\mu_z(s_z, X) - f_z(s_1, s_0; \eta)\} ds_1 ds_0.\end{aligned}\quad (2.11)$$

We have

$$E \{D_{z, \text{pd+om}}(Y, S, Z, X; \eta_z, e, \mu_z)\} = 0 \quad (2.12)$$

and thus $D_{z, \text{pd+om}}$ is a valid estimating equation for η_z for $z = 0, 1$.

Theorem 2.1 indicates that

$$m_z(s_1, s_0) - f_z(s_1, s_0; \eta) = E \left[\frac{e(s_1, s_0, X)}{e(s_1, s_0)} \{\mu_z(s_z, X) - f_z(s_1, s_0; \eta)\} \right]. \quad (2.13)$$

Plugging (2.13) into (2.10) then yields the validity of the estimating equation in Lemma 2.1. Equation (2.11) motivates the following two-step estimator.

Estimator 2.1 (Estimator based on principal density and outcome mean models). *We provide a two-step estimator $\hat{\eta}_{z,\text{pd}+\text{om}}$ based on the principal density and outcome mean model for $z = 0, 1$.*

- *Step 1: Estimate the principal density $\hat{e}_i$ and outcome mean $\hat{\mu}_{zi}$ for each unit i .*
- *Step 2: Obtain the estimator $\hat{\eta}_{z,\text{pd}+\text{om}}$ by plugging $\hat{e}_i$ and $\hat{\mu}_{zi}$ into (2.11) and solving for the empirical analog of the estimating equation (2.12)*

$$n^{-1} \sum_{i=1}^n D_{z,\text{pd}+\text{om}}(Y_i, S_i, Z_i, X_i; \hat{\eta}_{z,\text{pd}+\text{om}}, \hat{e}_i, \hat{\mu}_{zi}) = 0.$$

The consistency and root- n convergence rate of $\hat{\eta}_{z,\text{pd}+\text{om}}$ to η_z require the consistent estimation and root- n convergence rate of the principal density and outcome mean models.

Next, we construct estimating equations based on the identification formula in Theorem 2.2.

Lemma 2.2 (Estimating equation based on treatment probability and principal density). *Suppose that Assumptions 2.1 and 2.2 hold and $0 < \pi(x) < 1$ for all x in the support of X . Define*

$$\begin{aligned} & D_{1,\text{tp}+\text{pd}}(Y, S, Z, X; \eta, \pi, e) \\ &= \int w_1(S, s_0) \dot{f}_1(S, s_0; \eta) \frac{e(S, s_0, X)}{p_1(S, X)} \frac{Z}{\pi(X)} \{Y - f_1(S, s_0; \eta)\} ds_0 \end{aligned} \quad (2.14)$$

and

$$\begin{aligned} & D_{0,\text{tp}+\text{pd}}(Y, S, Z, X; \eta, \pi, e) \\ &= \int w_0(s_1, S) \dot{f}_0(s_1, S; \eta) \frac{e(s_1, S, X)}{p_0(S, X)} \frac{1-Z}{1-\pi(X)} \{Y - f_0(s_1, S; \eta)\} ds_1. \end{aligned} \quad (2.15)$$

We have

$$E \{D_{z,\text{tp}+\text{pd}}(Y, S, Z, X; \eta_z, \pi, e)\} = 0 \quad (2.16)$$

and thus $D_{z,\text{tp}+\text{pd}}$ is a valid estimating equation for η_z for $z = 0, 1$.

In Lemma 2.2, the weights based on principal density and treatment probability in (2.14) and (2.15) have a similar form as those in Theorem 2.2. These equations depend on the treatment probability and principal density while remaining agnostic of the outcome means, motivating the following two-step weighting estimator that does not require modeling the outcome.

Estimator 2.2 (Estimator based on treatment probability and principal density). *We provide a two-step estimator $\hat{\eta}_{z, \text{tp}+\text{pd}}$ based on the treatment probability and principal density for $z = 0, 1$.*

- *Step 1: Estimate the treatment probability $\hat{\pi}_i$ and principal density $\hat{e}_i$ for each unit i .*
- *Step 2: Obtain the estimator $\hat{\eta}_{z, \text{tp}+\text{pd}}$ by plugging $\hat{\pi}_i$ and $\hat{e}_i$ into (2.14) and (2.15) and solving for the empirical analog of the estimating equation (2.16)*

$$n^{-1} \sum_{i=1}^n D_{z, \text{tp}+\text{pd}}(Y_i, S_i, Z_i, X_i; \hat{\eta}_{z, \text{tp}+\text{pd}}, \hat{\pi}_i, \hat{e}_i) = 0.$$

The consistency and root- n convergence rate of $\hat{\eta}_{z, \text{tp}+\text{pd}}$ to η_z require the consistent estimation and root- n convergence rate of the treatment probability and principal density models.

When the corresponding nuisance parameters are consistently estimated at a root- n rate, we can construct variance estimators of $\hat{\eta}_{z, \text{pd}+\text{om}}$ and $\hat{\eta}_{z, \text{tp}+\text{pd}}$ using standard variance estimators derived from the estimating equations for Z-estimators. We can also obtain variance estimators using a nonparametric bootstrap.

Because estimating equations (2.14) and (2.15) involve the inverse of the treatment probability, $\hat{\eta}_{z, \text{tp}+\text{pd}}$ becomes unstable if the estimated treatment probability is close to zero or one. In this case, $\hat{\eta}_{z, \text{pd}+\text{om}}$ is comparatively more stable, though it relies on the extrapolation using the outcome mean model. Interestingly, the inverse of $p_z(S, X)$ in (2.14) and (2.15) does not lead to a similar problem because the ratios $e(S, s_0, X)/p_1(S, X)$ and $e(s_1, S, X)/p_0(S, X)$ are always bounded in $[0, 1]$.

When the estimated treatment probability is extreme, we can improve the empirical performance of the estimator by using stabilized weights. Specifically, instead of using the inverse of the estimated treatment probabilities in (2.14) and (2.15), we replace them with more stable weights that satisfy the estimating equations. This approach improves stability, as demonstrated in previous studies (Imai and Ratkovic, 2015; Zubizarreta, 2015; Chattopadhyay et al., 2020; Ben-Michael et al., 2021).

The computation of both $\hat{\eta}_{z, \text{pd}+\text{om}}$ and $\hat{\eta}_{z, \text{tp}+\text{pd}}$ requires numerical integration. Due to the additional layer of integration in (2.11) compared to (2.14) and (2.15), $\hat{\eta}_{z, \text{pd}+\text{om}}$ is generally more computationally intensive than $\hat{\eta}_{z, \text{tp}+\text{pd}}$.

2.4 Semiparametric estimation based on EIF

The availability of multiple estimators for η_z catalyzes devising the most efficient estimator. In this section, we first derive the EIF for η_z and then construct an estimator based on the EIF. For ease of notation, we only present the result for η_1 ; the results for η_0 and η_τ can be obtained similarly and are provided in the Supplementary Material. Define

$$\begin{aligned}\nu_1(s_1, s_0, X) &= w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\}, \\ r_u(s_1, s_0, S, X) &= 1 - \frac{\dot{c}_u(s_1, s_0, X)}{c(s_1, s_0, X)} \{1(S \leq s_1) - F_1(s_1, X)\}, \\ r_v(s_1, s_0, S, X) &= 1 - \frac{\dot{c}_v(s_1, s_0, X)}{c(s_1, s_0, X)} \{1(S \leq s_0) - F_0(s_0, X)\}.\end{aligned}$$

The term $\nu_1(s_1, s_0, X)$ is the integrand of the estimating equation in (2.11), and $r_u(s_1, s_0, S, X)$ and $r_v(s_1, s_0, S, X)$ are quantities related to S_1 and S_0 . The term $\dot{c}_u(s_1, s_0, X)/c(s_1, s_0, X)$ is equal to $\partial \log c(u, v)/\partial u$ evaluated at $F_1(s_1, X)$ and $F_0(s_0, X)$, which can be viewed as the score function of the principal density with respect to u . Similar interpretation applies to $\dot{c}_v(s_1, s_0, X)/c(s_1, s_0, X)$. We can show that $E\{r_u(s_1, s_0, S, X) \mid Z = 1, X\} = E\{r_v(s_1, s_0, S, X) \mid Z = 0, X\} = 1$. The following theorem presents the EIF for η_1 .

Theorem 2.3 (EIF for η_1). *Under the nonparametric model, the EIF for η_1 is*

$$\varphi_1(V) = H_1^{-1} \{\ell_1(X) + \ell_2(S, Z, X) + \ell_3(Y, S, Z, X)\}, \quad (2.17)$$

where

$$H_1 = E \left(w_1(S_1, S_0) \left[\dot{f}_1(S_1, S_0; \eta_1) \dot{f}_1(S_1, S_0; \eta_1)^T - \{m_1(S_1, S_0) - f_1(S_1, S_0; \eta_1)\} \ddot{f}_1(S_1, S_0; \eta_1) \right] \right)$$

and

$$\begin{aligned}\ell_1(X) &= \iint \nu_1(s_1, s_0, X) ds_1 ds_0, \\ \ell_2(S, Z, X) &= \frac{Z}{\pi(X)} \left\{ \frac{\int \nu_1(S, s_0, X) ds_0}{p_1(S, X)} - \iint \nu_1(s_1, s_0, X) r_u(s_1, s_0, S, X) ds_1 ds_0 \right\} \\ &\quad + \frac{1-Z}{1-\pi(X)} \left\{ \frac{\int \nu_1(s_1, S, X) ds_1}{p_0(S, X)} - \iint \nu_1(s_1, s_0, X) r_v(s_1, s_0, S, X) ds_1 ds_0 \right\}, \\ \ell_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \frac{\int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\}.\end{aligned}$$

We suppress the dependence of H_1 , ℓ_1 , ℓ_2 , and ℓ_3 on η_1 for simplicity. We show in the proof of Theorem 2.3 that $2H_1$ is equal to the Hessian matrix of the objective function in (2.6)

evaluated at η_1 . The three terms $\ell_1(X)$, $\ell_2(S, Z, X)$, and $\ell_3(Y, S, Z, X)$ correspond to the three levels of perturbation on different components of the observed conditional distributions. The function $\ell_1(X)$ is the same as the estimating equation in (2.11). The function $\ell_3(Y, S, Z, X)$ has a similar form to the estimating equation in (2.14) by replacing the residual $Y - f_1(S, s_0; \eta)$ inside the integral with $Y - \mu_1(S, X)$. The function $\ell_2(S, Z, X)$ takes a more complex form, partly due to the association between S_1 and S_0 specified in the copula function, which hinders a clear explanation. In the special case when S_1 and S_0 are independent conditional on X , $c(u, v) = 1$, $\dot{c}_u(u, v) = \dot{c}_v(u, v) = 0$, and thus $r_u(s_1, s_0, S, X) = r_v(s_1, s_0, S, X) = 1$. The function $\ell_2(S, Z, X)$ then simplifies to

$$\frac{Z}{\pi(X)} [u_1(S, X) - E\{u_1(S, X) \mid X, Z = 1\}] + \frac{1 - Z}{1 - \pi(X)} [u_1(S, X) - E\{u_1(S, X) \mid X, Z = 0\}],$$

where $u_1(s_z, X) = \int \nu_1(s_1, s_0, X) ds_{1-z} / p_z(s_z, X)$ for $z = 0, 1$. Thus, $\ell_2(S, Z, X)$ consists of the inverse probability weighting of the residuals of the ratio $u_1(s_z, X)$ in this special case. Importantly, all the terms in the influence function involve the principal density, which further demonstrates the difficulty of avoiding it in identification formulas.

The absence of terms corresponding to the perturbation of the conditional distribution of Z given X implies that knowing the treatment probabilities would not help improve the semiparametric efficiency of our causal quantities. Therefore, the same EIF as in Theorem 2.3 applies to randomized experiments where the true treatment probabilities are known. A similar phenomenon appears in the estimation of the classic average treatment effect (Hahn, 1998) and has been studied in Hitomi et al. (2008).

The EIF motivates the following estimating equation for η_1 .

Lemma 2.3 (Estimating equation based on the EIF). *Suppose that Assumptions 2.1 and 2.2 hold, and $0 < \pi(x) < 1$ for all x in the support of X . Define*

$$D_{1,\text{eif}}(Y, S, Z, X; \eta_1, \pi, e, \mu_1) = \ell_1(X) + \ell_2(S, Z, X) + \ell_3(Y, S, Z, X). \quad (2.18)$$

We have

$$E\{D_{1,\text{eif}}(Y, S, Z, X; \eta_1, \pi, e, \mu_1)\} = 0. \quad (2.19)$$

This follows directly from the zero property of the EIF. Lemma 2.3 leads to the following estimation strategy.

Estimator 2.3 (Estimator based on EIF). *We provide a two-step estimator $\hat{\eta}_{1,\text{eif}}$ motivated by the form of EIF. It is based on all three nuisance parameters: the treatment probability, principal density, and outcome mean.*

- *Step 1: Estimate the treatment probability $\hat{\pi}_i$, principal density $\hat{e}_i$, and outcome mean $\hat{\mu}_{1i}$ for each individual i .*

- *Step 2: Obtain the estimator $\hat{\eta}_{1,\text{eif}}$ by plugging the estimated values of the nuisance parameters $(\hat{\pi}_i, \hat{e}_i, \hat{\mu}_{1i})$ into (2.18) and solving for the empirical analog of the estimating equation (2.19)*

$$n^{-1} \sum_{i=1}^n \hat{D}_{1,\text{eif}}(Y_i, S_i, Z_i, X_i; \hat{\eta}_{1,\text{eif}}, \hat{\pi}_i, \hat{e}_i, \hat{\mu}_{1i}) = 0. \quad (2.20)$$

We can similarly construct a two-step estimator $\hat{\eta}_{0,\text{eif}}$ based on the EIF for η_0 , details of which are relegated to Section A.1.4 in the Supplementary Material.

The estimation approach can be applied in both randomized experiments and observational studies. In randomized experiments, knowing the true treatment probability does not improve the semiparametric efficiency bound because the EIF does not have a term corresponding to the perturbation of the conditional distribution of Z given X . However, randomization ensures that the model for the treatment probability is always correctly specified, thereby enhancing the robustness of the estimators.

In the special case when the working model is linear in η , i.e., $f_1(s_1, s_0; \eta) = \eta^T g_1(s_1, s_0)$ for some known pre-specified function g_1 and unknown parameters η , the EIF for η_1 is also linear in η_1 . This structure enables a closed-form solution to the linear estimating equations, substantially reducing computational complexity in estimation. We present the EIF and the closed-form expression of the associated estimator in Section A.1.2 of the Supplementary Material. We implement our proposed estimation procedure based on this closed-form solution in an R package `continuousPCE`.

The function $g_1(s_1, s_0)$ can be tailored to suit various real-world applications. In Section A.1.3 of the Supplementary Material, we discuss some specific choices of $g_1(s_1, s_0)$, such as polynomial functions of (s_1, s_0) , and provide the corresponding closed-form EIFs.

When the working model is nonlinear in η_1 , the empirical estimating equation (2.20) often involves nonlinear functions in η_1 . Consequently, estimating η_1 requires the use of numerical methods such as Newton's method, posing computational challenges.

2.5 Asymptotic theory and statistical inference

In this section, we derive the asymptotic properties of the estimator based on the EIF. We show that it is conditionally doubly robust and exhibits a parametric convergence rate under certain regularity conditions, even when nuisance functions are estimated at slower rates. Similar to Section 2.4, we only present the result for $\hat{\eta}_{1,\text{eif}}$.

To study the estimator under model misspecification, we work with the $L_2(P)$ probability limits. For the estimated treatment probability function $\hat{\pi}$, define its probability limit $\bar{\pi}$ as the function such that $\|\hat{\pi} - \bar{\pi}\|_2$ converges in probability to 0 as n goes to infinity. Similarly, define $\bar{e}$ and $\bar{\mu}_1$ as the $L_2(P)$ probability limit of the estimated principal density $\hat{e}$ and outcome mean $\hat{\mu}_1$, respectively. Therefore, $\bar{\pi} = \pi$ indicates that the treatment probability model is correctly specified, $\bar{e} = e$ indicates that the principal density model is correctly specified, and $\bar{\mu}_1 = \mu_1$ indicates that the outcome mean model is correctly specified.

We first introduce the following three regularity conditions on the estimated nuisance parameters and the estimating equation.

Assumption 2.3 (Glivenko–Cantelli class). *Let η and ξ denote the parameter of interest and the nuisance parameters, respectively. Assume that the class of functions $\{D_{1,\text{eif}}(\cdot; \tilde{\eta}, \tilde{\xi}) : \|\tilde{\eta} - \eta\| < \delta, d(\tilde{\xi}, \bar{\xi}) < \delta\}$ is P -Glivenko–Cantelli for some $\delta > 0$.*

Assumption 2.4 (Donsker class). *Let η and ξ denote the parameter of interest and the nuisance parameters, respectively. The function $v = (y, s, z, x) \mapsto D_{1,\text{eif}}(v; \tilde{\eta}, \tilde{\xi})$ indexed by $(\tilde{\eta}, \tilde{\xi})$ satisfies*

- (a) *the class of functions $\{D_{1,\text{eif}}(\cdot; \tilde{\eta}, \tilde{\xi}) : \|\tilde{\eta} - \eta\| < \delta, d(\tilde{\xi}, \bar{\xi}) < \delta\}$ is Donsker for some $\delta > 0$, and*
- (b) *$E\{\|D_{1,\text{eif}}(\cdot; \tilde{\eta}, \tilde{\xi}) - D_{1,\text{eif}}(\cdot; \eta, \bar{\xi})\|^2\} \rightarrow 0$ as $(\tilde{\eta}, \tilde{\xi}) \rightarrow (\eta, \bar{\xi})$.*

Assumption 2.5 (Differentiability). *The mapping $\tilde{\eta} \mapsto E\{D_{1,\text{eif}}(V; \tilde{\eta}, \tilde{\xi})\}$ is differentiable at η uniformly in $\tilde{\xi}$ in a neighborhood of $\bar{\xi}$ with nonsingular derivative matrices $C(\eta, \bar{\xi}) \equiv \partial E\{D_{1,\text{eif}}(V; \tilde{\eta}, \tilde{\xi})\} / \partial \tilde{\eta}|_{\tilde{\eta}=\eta}$.*

Assumptions 2.3 and 2.4 restrict the model complexity of the nuisance parameters. Specifically, Assumption 2.3 is sufficient for a uniform law of large numbers, hence for the consistency of the Z-estimator $\hat{\eta}_{1,\text{eif}}$. Assumption 2.4 is a standard regularity condition for the asymptotic distribution of Z-estimators (Vaart, 2000). Moreover, we could use sample-splitting and cross-fitting methods to estimate the model parameters to relax the Donsker conditions (Koshevnik and Levit, 1977; Pfanzagl and Wefelmeyer, 1985; Klaassen, 1987; Zheng and Laan, 2011; Chernozhukov et al., 2018). Assumption 2.5 is another regularity condition that validates the use of the Delta method when we derive the asymptotic distribution of $\hat{\eta}_{1,\text{eif}}$.

The following theorem gives a general result for the asymptotic distribution of $\hat{\eta}_{1,\text{eif}}$.

Theorem 2.4 (Consistency). *Suppose that the probability limit of the estimated nuisance parameters satisfies that either $(\bar{\pi}, \bar{e}) = (\pi, e)$ or $(\bar{e}, \bar{\mu}_1) = (e, \mu_1)$, and the weight function $w_1(s_1, s_0)$ and the working model $f_1(s_1, s_0)$ are bounded.*

(a) Under Assumptions 2.1–2.3, $\hat{\eta}_{1,\text{eif}}$ is a consistent estimator of η_1 .

(b) Under Assumptions 2.1–2.5,

$$\|\hat{\eta}_{1,\text{eif}} - \eta_1\|_2 = O_P\{n^{-1/2} + \text{Rem}\},$$

where

$$\begin{aligned} \text{Rem} = & \|\hat{\pi} - \pi\|_2 \|\hat{\mu}_1 - \mu_1\|_2 + \|\hat{\pi} - \pi\|_2 (\|\hat{p}_1 - p_1\|_2 + \|\hat{p}_0 - p_0\|_2) \\ & + \|\hat{\mu}_1 - \mu_1\|_2 (\|\hat{p}_1 - p_1\|_2 + \|\hat{p}_0 - p_0\|_2) + \|\hat{p}_1 - p_1\|_2 \|\hat{p}_0 - p_0\|_2. \end{aligned}$$

Theorem 2.4 shows that the asymptotic bias of $\hat{\eta}_{1,\text{eif}}$ is determined by a function of the asymptotic biases of the nuisance parameters. While $\|\hat{\pi} - \pi\|_2 \|\hat{\mu}_1 - \mu_1\|_2$ converges to zero if either the treatment probability or the outcome mean model is correctly specified, the other terms rely heavily on the principal density model. In particular, the last term $\|\hat{p}_1 - p_1\|_2 \|\hat{p}_0 - p_0\|_2$ converges to zero only if the principal density model is correctly specified. As a result, the consistency of the estimator $\hat{\eta}_{1,\text{eif}}$ for η_1 is contingent on a correct specification of the principal density model, and either the treatment probability or outcome mean model. This establishes the conditional double robustness of $\hat{\eta}_{1,\text{eif}}$ when the principal density model is correctly specified. In contrast, the triply robust property demonstrated in Jiang et al. (2022) allows the estimator to remain consistent even if the principal score model is misspecified, provided the treatment probability and outcome mean models are correct. Our result echoes the discussion on the crucial role of the principal density in identifying the PCE surface with a continuous S .

Theorem 2.4 also implies the result on root- n consistency of $\hat{\eta}_{1,\text{eif}}$, provided the following assumption on the convergence rates of the treatment probability, principal density, and the outcome mean.

Assumption 2.6 (Convergence rates of nuisance parameters). *The convergence rates of the estimators of the nuisance parameters satisfy*

- (a) $\|\hat{\pi} - \pi\|_2 \|\hat{p}_z - p_z\|_2 = o_P(n^{-1/2})$ for $z = 0, 1$,
- (b) $\|\hat{p}_z - p_z\|_2 \|\hat{\mu}_1 - \mu_1\|_2 = o_P(n^{-1/2})$ for $z = 0, 1$,
- (c) $\|\hat{\pi} - \pi\|_2 \|\hat{\mu}_1 - \mu_1\|_2 = o_P(n^{-1/2})$, and
- (d) $\|\hat{p}_z - p_z\|_2 = o_P(n^{-1/4})$ for $z = 0, 1$.

Assumption 2.6 requires that the models on treatment probability, principal density, and outcome mean are all correctly specified. It imposes more stringent requirements on the

convergence rates of the models. If domain knowledge enables us to correctly specify parametric models to estimate the treatment probability and the outcome mean model, parametric estimation guarantees a root- n rate of nuisance parameters π and μ_z , thus Assumptions 2.6(a)–2.6(c) are achieved. However, parametric models might be restrictive in observational studies with multi-dimensional covariates. We can instead use flexible machine learning models for estimation and still achieve the required rates (Robins et al., 1994; Laan and Robins, 2003a; Chernozhukov et al., 2018). For the conditional density estimation of p_z , we can use nonparametric density estimation such as kernel density estimation when the covariates X is low-dimensional. With larger dimensions of the covariates, we need to impose more restrictions on the principal density to achieve the required convergence rate of $\hat{p}_z$ in Assumption 2.6(d).

Corollary 2.1 (Root- n consistency). *Under Assumptions 2.1–2.6, we have $\|\hat{\eta}_{1,\text{eif}} - \eta_1\|_2 = O_P(n^{-1/2})$.*

The next theorem shows the asymptotic Normality and semiparametric efficiency of $\hat{\eta}_{1,\text{eif}}$.

Theorem 2.5 (Asymptotic distribution). *Under Assumptions 2.1–2.6, we have*

$$\sqrt{n}(\hat{\eta}_{1,\text{eif}} - \eta_1) \xrightarrow{d} \mathcal{N}(0, E\{\varphi_1(V; \eta_1, \xi) \varphi_1(V; \eta_1, \xi)^T\})$$

where $\varphi_1(V; \eta_1, \xi)$ is the EIF for η_1 as in Theorem 2.3. That is, $\hat{\eta}_{1,\text{eif}}$ achieves the semiparametric efficiency bound.

If Assumption 2.6 holds, we can construct a consistent estimator of the asymptotic variance using the sample analog of the semiparametric efficiency bound, i.e., $n^{-1} \sum_{i=1}^n \varphi_1(V; \hat{\eta}_{1,\text{eif}}, \hat{\xi}) \varphi_1(V; \hat{\eta}_{1,\text{eif}}, \hat{\xi})^T$. This, however, requires all the nuisance models to be correctly specified. Under certain smoothness conditions on the nuisance models, we can estimate the asymptotic variance of $\hat{\eta}_{1,\text{eif}}$ using nonparametric bootstrap even if Assumption 2.6 fails (Kennedy et al., 2019). If neither Assumption 2.6 holds nor the requirements for valid nonparametric bootstrap are satisfied, the statistical inference needs more careful procedures. This is beyond the scope of this paper, and we leave it to future work.

2.6 Simulation

In this section, we evaluate the finite sample performance of various estimators for η_1 defined in Equation (2.6). We employ a linear projection working model $f_1(s_1, s_0; \eta_1) = \beta_1^1 s_1 + \beta_1^0 s_0 + \alpha_1$ with uniform weighting function $w_1(s_1, s_0) = 1$ for any s_1 and s_0 .

Generate the covariates $X = (X_1, X_2)^T \in \mathbb{R}^2$ from two independent standard Normal distributions and the treatment by $Z \mid X \sim \text{Bernoulli}\{\pi(X)\}$. Given a pre-specified

value of ρ , generate the potential values of the post-treatment variable by $S_1 = 0.6X_1 + 0.8\{\rho^{1/2}\epsilon_\rho + (1 - \rho)^{1/2}\epsilon_1\}$ and $S_0 = 0.6X_2 + 0.8\{\rho^{1/2}\epsilon_\rho + (1 - \rho)^{1/2}\epsilon_0\}$, where ϵ_ρ , ϵ_1 , and ϵ_0 are independently from standard normal distribution. Generate the potential outcomes by $Y_z \mid S_1, S_0, X \sim \mathcal{N}(\mu_z(S_1, S_0, X), 1)$ for $z = 0, 1$. The observed post-treatment variable and outcome are then determined by $S = ZS_1 + (1 - Z)S_0$ and $Y = ZY_1 + (1 - Z)Y_0$, respectively. We consider three different data-generating regimes. In the first regime, $\pi(X) = 0.5$, $\mu_1(S_1, S_0, X) = S_1 + X_1$ and $\mu_0(S_1, S_0, X) = S_0 + X_2$. In the second regime, $\pi(X) = \text{logit}\{(\tilde{X}_1 + \tilde{X}_2)/2\}$, where $\tilde{X}_j = \{(X_j + 0.25)^2 - 1\}/\sqrt{2}$ for $j = 1, 2$, and $\mu_z(S_1, S_0, X)$ is the same as in the first regime. In the third regime, $\pi(X)$ is the same as in the first regime, and $\mu_1(S_1, S_0, X) = S_1 + 2\sqrt{2}\tilde{X}_1$ and $\mu_0(S_1, S_0, X) = S_0 + 2\sqrt{2}\tilde{X}_2$.

We consider three estimators: the EIF based estimator $\hat{\eta}_{1,\text{eif}}$, and the estimators based on the estimating equations (2.11) and (2.14), denoted by $\hat{\eta}_{1,\text{pd+om}}$ and $\hat{\eta}_{1,\text{tp+pd}}$, respectively. We estimate the treatment probability by logistic regressions with linear predictors (X_1, X_2) , and outcome means by linear regressions with predictors (S, X_1, X_2) fitted separately in the treatment and control groups. Therefore, the treatment probability model is misspecified in the second regime and the outcome mean model is misspecified in the third regime. We estimate the principal density using a nonparametric kernel conditional density estimation and a Gaussian copula with the correctly specified correlation coefficient ρ . We use numerical integration when computing the single and double integrals.

For each data-generating regime, we generate a random sample of size n and compute the values of three estimators. We compare their performance by repeating the process over 500 Monte Carlo samples and computing the bias and the standard deviation of the three estimators. To avoid repetition, we only report the results for β_1^1 , which is the projection coefficient of s_1 in the working model $f_1(s_1, s_0; \eta_1) = \beta_1^1 s_1 + \beta_1^0 s_0 + \alpha_1$. To explore how the finite-sample performance varies across different sample sizes n , we choose four sample sizes $n = 200, 500, 1000, 2000$. We generate data following the aforementioned processes with various values of the correlation coefficient $\rho = 0, 0.2, 0.5$. R scripts to replicate all simulation results are available at https://github.com/sizhu-lu/pce_paper.

Table 2.1 shows the biases and the root mean squared errors (RMSEs) of the three estimators across various values of the sensitivity parameter $\rho = 0, 0.2, 0.5$ under each data generating regime. The performance is similar across different ρ values. The EIF estimator $\hat{\beta}_{1,\text{eif}}^1$ is consistent for the true value if either the treatment probability or the outcome mean model is correctly specified. The weighting estimator $\hat{\beta}_{1,\text{tp+pd}}^1$ performs well when the treatment probability model is correctly specified and is inconsistent otherwise. The finite-sample performance of $\hat{\beta}_{1,\text{pd+om}}^1$ is not satisfying under all three regimes. There exists a large finite-sample bias even when all models are correctly specified, although the bias decreases as the sample size increases. We conjecture the relatively poor performance is due to the double integration involved in estimating $\hat{\beta}_{1,\text{pd+om}}^1$. Overall, we recommend using the EIF estimator as it is generally robust to the misspecification of either the treatment probability or the

outcome mean model.

2.7 Application to a randomized experiment

In this section, we revisit Example 2.4 and analyze the data from a randomized experiment in China (Yin, 2026). The study aims to understand the causal effect of offering a credit limit increase on consumers' income expectations and economic decision-making. The randomized controlled trial is conducted in collaboration with a large commercial bank in China. In the trial, bank customers are randomly assigned to two groups: customers in the treatment group, denoted by $Z = 1$, are immediately offered an increase in their available credit, while customers in the control group, denoted by $Z = 0$, are not qualified for the credit limit increase until six months after the trial. The outcome of interest Y represents customers' consumption of non-durable goods in dollars 6 months after the treatment. The difference-in-means estimator of the average treatment effect on the outcome $E\{Y(1) - Y(0)\}$ is 678.487 with a 95% confidence interval (592.157, 764.818), which indicates that the treatment leads to an increase in future consumption.

In addition to the treatment and the outcome, the study also collected data on customers' expectations for their income in 6 months in both a pre-experiment survey right before the treatment and a post-experiment survey conducted afterward. We define the post-treatment variable of interest S as the difference in customers' income expectations between the pre- and post-treatment surveys. People's expectations change with time, and the treatment has a significant positive effect on people's expected future income on average. The difference-in-means estimator of $E\{S(1) - S(0)\}$ is 749.590 with a 95% confidence interval (635.246, 863.934). Based on the observed data, we aim to understand how the increase in credit limit affects consumption behavior through changes in income expectations. We explore how the PCE surface $\tau(s_1, s_0)$ varies with different values of s_1 and s_0 , which could provide insights into the causal mechanism of the treatment effect.

We specify our working models f_1 and f_0 to be linear in s_1 and s_0 with an intercept, i.e., $f_z(s_1, s_0; \eta_z) = \beta_z^1 s_1 + \beta_z^0 s_0 + \alpha_z$ for $z = 0, 1$, and thus the estimated PCE surface has a functional form of

$$f_\tau(s_1, s_0; \eta_\tau) = \beta_\tau^1 s_1 + \beta_\tau^0 s_0 + \alpha_\tau,$$

where $\beta_\tau^1 = \beta_1^1 - \beta_0^1$, $\beta_\tau^0 = \beta_1^0 - \beta_0^0$, and $\alpha_\tau = \alpha_1 - \alpha_0$. Under the linear working models f_1 and f_0 , we have $\tau(s_1, s_0) - \tau(s'_1, s_0) = \beta_\tau^1(s_1 - s'_1)$ for all s_0 . Therefore, β_τ^1 quantifies the extent to which the effect of credit limit increase on the consumption changes with varying potential values of income expectation under treatment. Similarly, β_τ^0 measures the changes in the treatment effect with varying potential values of income expectation under control. Moreover, when $s_1 = s_0 = s$, $\tau(s_1, s_0) = (\beta_\tau^1 + \beta_\tau^0)s + \alpha_\tau$, indicating the magnitude of the treatment effect not attributed to the change in expected future income.

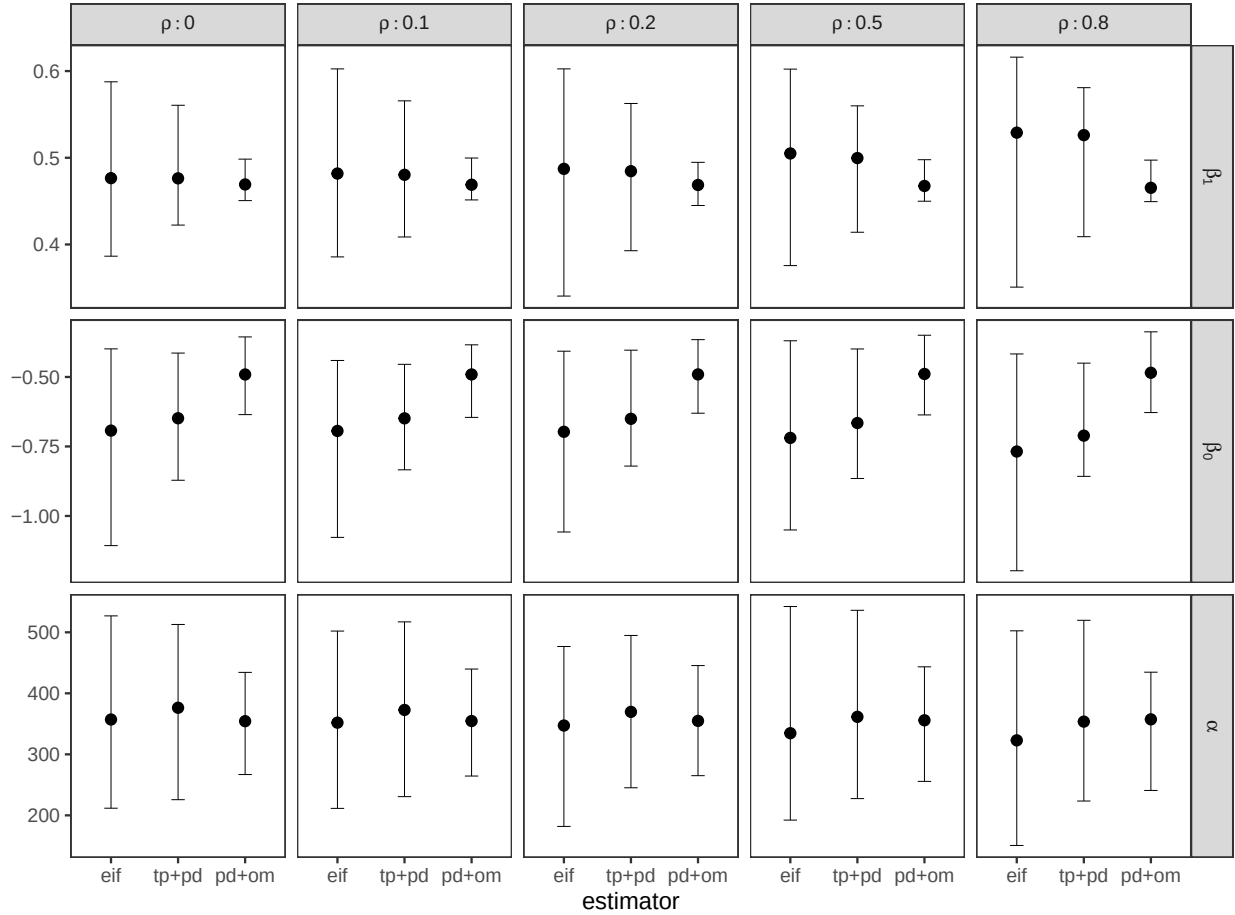


Figure 2.1: Estimated PCE surface of the credit limit increase on future consumption. The rows correspond to the three parameters in the working model and the columns correspond to different values of ρ . Each subplot shows the point estimates and the corresponding 95% confidence intervals for the three estimation strategies.

In this application, Assumption 2.1 is guaranteed by the design of the original randomized controlled trial. For Assumption 2.2, we condition on the gender, age, annual income, and credit score. We use a logistic regression for the treatment probability model $\pi(X)$, kernel density estimation for the conditional densities $p_z(s_z, X)$, and linear regressions for the outcome mean models $\mu_z(s_z, X)$, for $z = 0, 1$. To get the estimated principal density, we specify a Gaussian copula with different values of the correlation coefficient ρ . We compute point estimators of the coefficients of the EIF estimator and the two basic estimators, and construct 95% confidence intervals using the nonparametric bootstrap. Again, R scripts that reproduce our results are available at https://github.com/sizhu-lu/pce_paper.

The results of $(\hat{\beta}_{\tau,*}^1, \hat{\beta}_{\tau,*}^0, \hat{\alpha}_{\tau,*})$ ($*$ = eif, tp+pd, pd+om) are summarized in Figure 2.1 across various values of the sensitivity parameter ρ . The point estimators do not vary much across

different values of ρ and estimation strategies. By design of the randomized controlled trial, the treatment is randomly assigned, so the treatment probability model is guaranteed to have a correct specification, leading to similar point estimators of the weighting estimator and the EIF estimator. Overall, the estimated PCE surface exhibits an increasing trend in S_1 and a decreasing trend in S_0 , suggesting that the credit limit increase offer positively affects customers' consumption behavior with heterogeneity across principal strata. For example, consider the EIF-estimated PCE surface when $\rho = 0$. The estimated functional form is $f_\tau(s_1, s_0) = 0.476s_1 - 0.693s_0 + 357.067$. This indicates that the treatment effect on consumption is more pronounced for customers who experience a greater increase in their expected future income between the pre- and post-treatment surveys under the treatment condition, and a larger decrease in their expected future income between the two surveys under the control condition. For customers with $S_1 = S_0 = s$, the effect of the credit limit increase offer is estimated as $-0.217s + 357.067$. This suggests that for customers whose expected future income remains unchanged by the credit limit offer but increases significantly over time, the impact of the credit limit offer on their consumption is reduced.

2.8 Extensions

We outline several extensions of our framework in the Supplementary Material. First, we consider sensitivity analyses for the principal ignorability assumption, which is crucial for identification but inherently untestable from observed data. Second, we can extend the framework to settings with multiple continuous post-treatment variables, in which S is a random vector. By generalizing the principal ignorability assumption to the multivariate setting, we can derive analogous identification formulas for the projection parameters. However, this extension introduces challenges in specifying the copula model and complicates the derivation of the EIF. Third, we consider conditional PCEs, defined as $E\{Y_1 - Y_0 \mid U = (s_1, s_0), W = w\}$, where W is a subset of the observed covariates X . These conditional PCEs remain identifiable under the same set of assumptions, and projection parameters can be similarly defined using working models. We leave the comprehensive investigation of these extensions to future work.

Table 2.1: Biases and RMSEs of the three estimators $\hat{\beta}_{1,\text{eif}}^1$, $\hat{\beta}_{1,\text{tp+pd}}^1$, and $\hat{\beta}_{1,\text{pd+om}}^1$ with different values of $\rho = 0, 0.2, 0.5$. In Regime 1, both the treatment probability and outcome mean models are correct; In Regime 2, the treatment probability model is misspecified; In Regime 3, the outcome mean models are misspecified. The principal density model is correct in all three regimes.

ρ	regime	n	$\hat{\beta}_{1,\text{eif}}^1$		$\hat{\beta}_{1,\text{tp+pd}}^1$		$\hat{\beta}_{1,\text{pd+om}}^1$	
			Bias	RMSE	Bias	RMSE	Bias	RMSE
$\rho = 0$	1	200	0.003	0.150	0.010	0.138	-0.147	0.199
		500	-0.004	0.099	-0.003	0.092	-0.126	0.155
		1000	0.004	0.067	0.005	0.064	-0.096	0.115
		2000	0.001	0.046	0.003	0.045	-0.081	0.092
	2	200	0.008	0.156	0.099	0.165	-0.154	0.208
		500	0.003	0.100	0.096	0.130	-0.122	0.151
		1000	0.008	0.069	0.104	0.122	-0.095	0.114
		2000	0.006	0.051	0.102	0.111	-0.078	0.090
	3	200	0.024	0.464	0.027	0.440	-0.131	0.422
		500	-0.012	0.316	-0.011	0.302	-0.133	0.318
		1000	0.016	0.213	0.018	0.207	-0.082	0.215
		2000	0.014	0.148	0.015	0.145	-0.068	0.162
$\rho = 0.2$	1	200	-0.026	0.161	-0.042	0.154	-0.184	0.233
		500	-0.008	0.097	-0.023	0.094	-0.132	0.160
		1000	-0.002	0.066	-0.015	0.067	-0.102	0.120
		2000	-0.003	0.046	-0.014	0.048	-0.086	0.097
	2	200	-0.005	0.156	0.059	0.149	-0.170	0.220
		500	-0.009	0.100	0.064	0.111	-0.134	0.162
		1000	0.002	0.068	0.079	0.102	-0.100	0.120
		2000	0.002	0.049	0.078	0.091	-0.083	0.095
	3	200	0.030	0.490	0.011	0.464	-0.140	0.439
		500	-0.009	0.325	-0.024	0.311	-0.140	0.327
		1000	0.018	0.215	0.005	0.209	-0.087	0.217
		2000	0.015	0.151	0.005	0.147	-0.073	0.166
$\rho = 0.5$	1	200	-0.055	0.169	-0.101	0.185	-0.193	0.244
		500	-0.031	0.103	-0.064	0.115	-0.139	0.167
		1000	-0.015	0.070	-0.049	0.084	-0.104	0.123
		2000	-0.010	0.050	-0.036	0.061	-0.081	0.094
	2	200	-0.048	0.174	-0.017	0.151	-0.188	0.241
		500	-0.033	0.110	0.008	0.098	-0.142	0.172
		1000	-0.013	0.071	0.034	0.076	-0.102	0.123
		2000	-0.009	0.050	0.041	0.063	-0.083	0.095
	3	200	0.025	0.549	-0.023	0.517	-0.155	0.496
		500	-0.005	0.356	-0.043	0.344	-0.141	0.355
		1000	0.017	0.244	-0.013	0.239	-0.085	0.237
		2000	0.014	0.168	-0.010	0.164	-0.074	0.180

Chapter 3

Estimating treatment effects with competing intercurrent events in randomized controlled trials

3.1 Intercurrent events in randomized controlled trials

Randomized controlled trials (RCTs) are considered the gold standard for evaluating treatment efficacy, primarily because randomization supports assumption-lean inference of the treatment effect. However, after treatment initiation, various events, referred to as intercurrent events (IEs), can arise, impacting the interpretation or availability of outcome measurements and posing significant challenges to the analysis of RCTs. Examples of IEs include treatment discontinuation due to adverse events or lack of efficacy, patient relocation, and the use of additional medications. Carefully accounting for IEs is essential to ensure the validity and reliability of the causal conclusions drawn from RCTs.

Recognizing the critical need for clearly defined estimands in the presence of IEs, the International Council for Harmonization (ICH) issued the E9(R1) Addendum (ICH E9 (R1), 2019). This addendum introduces a structured estimand framework for clinical trials to obtain precisely defined treatment effects that align with the clinical questions of interest. It outlines strategies for addressing IEs during the formulation of the clinical question and emphasizes that careful specification of the treatment, population, and outcome variable often addresses many of the IEs raised in discussions between sponsors and regulators. Since its release, the ICH E9(R1) has been widely discussed, increasingly adopted in clinical drug development, and has sparked substantial interest in statistical research (Qu et al., 2021; Kang et al., 2022; Ionan et al., 2023; Han and Zhou, 2023; Olarte Parra et al., 2025b).

A widely accepted strategy is the *treatment policy strategy* (ICH E9 (R1), 2019), which

includes all participants in their originally assigned groups and uses the observed outcome values, regardless of whether or not an IE occurs. This approach aligns with the intention-to-treat principle and reflects the effect of a treatment policy in real clinical settings. However, it cannot address IEs that are terminal events, such as death, because such events preclude the existence of the outcome variable.

The *composite variable strategy* is an alternative approach that is well-suited for handling IEs that are informative of the patient’s outcome such as the terminal events. This strategy incorporates the occurrence of IEs directly into the outcome definition. Specifically, it defines a composite outcome: if no IE occurs, the outcome of interest is used as observed; if an IE occurs, the endpoint is deterministically set to a pre-specified, clinically meaningful value, typically indicating treatment failure. This strategy is widely used in practice across various types of outcomes, including binary outcomes (e.g., the non-responder rule), ordinal or continuous outcomes (e.g., Rosenbaum, 2006), and time-to-event outcomes (e.g., progression-free survival).

In addition to the treatment policy and composite variable strategies, ICH E9 (R1), 2019 outlines three additional strategies for handling IEs. The *hypothetical strategy* evaluates treatment effects under a hypothetical scenario in which the IE would not occur. The *while-on-treatment* and *principal stratification strategies* are also described but are less commonly used in current practice due to their reliance on strong assumptions and the potential to introduce bias in treatment comparisons. ICH E9 (R1), 2019 recommends using different strategies based on the specific type of IEs involved.

Several recent papers have begun to consider settings with multiple IE types and the use of multiple strategies. For example, Lipkovich et al. (2020) and Qu et al. (2021) discuss estimands that tailor strategies to different IEs without rigorously examining how to combine different strategies or rigorously addressing the potential pitfalls of doing so without careful consideration. More recently, Olarte Parra et al. (2025a) provides a careful conceptual discussion of estimands that combine the treatment policy and hypothetical strategies, using potential outcomes and causal diagrams to clarify alternative interpretations and the corresponding identification assumptions. Complementing this line of work, we study a different hybrid-strategy setting that combines the hypothetical strategy with the composite outcome strategy. Our setting also highlights a practically important feature in multi-IE settings that received limited explicit attention in the literature: multiple IEs can *compete* with each other.

3.1.1 Two phase-3 trials in systemic lupus erythematosus

To rigorously examine the challenges posed by IEs in clinical trials, we analyze data from two recent randomized studies for systemic lupus erythematosus (Morand et al., 2023; Petri et al., 2023), in which participants were randomly assigned to either the baricitinib treatment

group or the placebo control group. These twin trials were designed to provide substantiated evidence on the causal treatment effect of baricitinib versus placebo.

The primary outcome is a response index measured at 52 weeks after treatment initiation. Ideally, this outcome would be compared directly between the two treatment groups at week 52. However, 429 out of 1,535 patients (27.95%) experienced IEs during the follow-up period, resulting in unobserved outcome data. Figure 3.1 shows the types and proportions of these IEs. During the 52 weeks, various IEs occurred: some patients discontinued treatment due to adverse events or lack of efficacy; others discontinued treatment due to study withdrawal for unspecified reasons or unable to contact; and some were excluded from the study due to protocol deviations. We revisit this example in greater detail in Section 3.5.

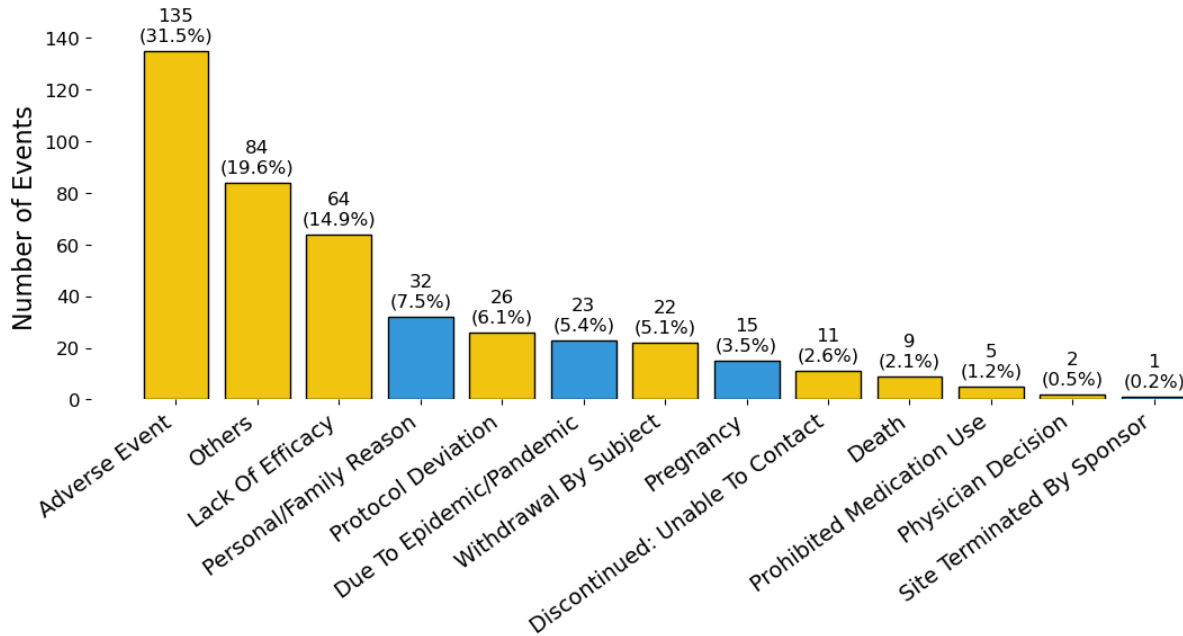


Figure 3.1: Bar plot showing the IE types and proportions. The yellow bars represent effect-informative IEs, and the blue bars represent effect-uninformative IEs.

3.1.2 Our proposal and contribution

We propose to classify IEs into two broad types: (1) effect-uninformative IEs, such as treatment discontinuation due to relocation or COVID-19 lockdown, which are assumed to be independent of treatment efficacy conditional on the observed covariates; and (2) effect-informative IEs, such as treatment discontinuation due to adverse events or lack of efficacy, use of rescue medication, and terminating events such as death, which are often informative about a patient's treatment effect.

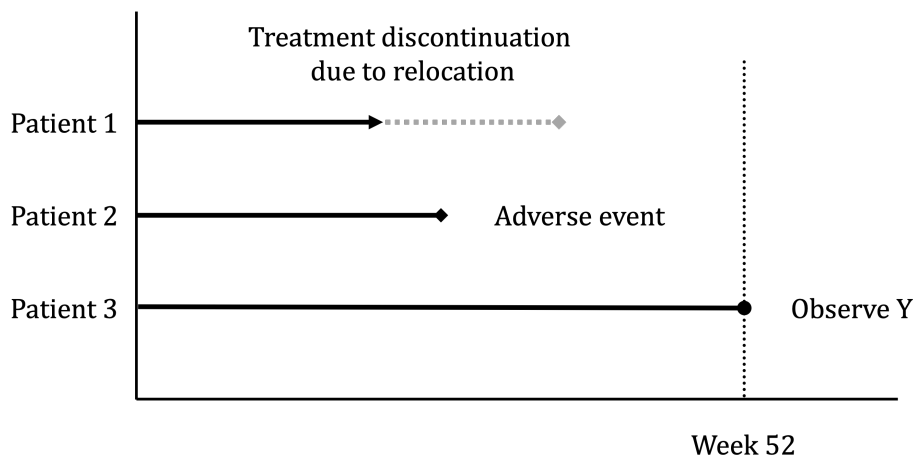


Figure 3.2: Illustration of three representative scenarios of intercurrent events in the immunology trial. The primary outcome is measured at week 52, and the solid lines indicate the observed follow-up period for each patient. Patient 1 experienced an effect-uninformative IE (represented by an arrow), Patient 2 experienced an effect-informative IE (represented by a diamond), and Patient 3 had no IE before week 52. This figure highlights the issue of competing IEs: for instance, had Patient 1 not discontinued early and remained on treatment (illustrated by the gray dashed line), they might have experienced an effect-informative IE later in the follow-up period.

In our application studies, the two types of IEs were classified based on a manual review of the detailed comments collected at the clinical sites. Figure 3.1 summarizes the main categories and illustrates our classification scheme: yellow bars represent effect-informative IEs, and blue bars represent effect-uninformative IEs. For instance, some patients withdrew from the study due to external factors such as relocation or the COVID-19 pandemic, which were classified as effect-uninformative. Others withdrew due to concerns about potential side effects or lack of efficacy, which were considered effect-informative. When the reason for withdrawal was unclear in the documentation, we defaulted to classifying the event as effect-informative and, under our composite outcome construction, assigned the corresponding failure value. This choice is intended to avoid overestimating arm-specific mean outcomes, while it does not necessarily make the estimated treatment effect conservative, since the net direction depends on how often such cases occur in each arm. Both types of IEs may occur during the follow-up period. A complete list of the IE categories observed, together with their assigned classification and rationale, is provided in Supplementary Section B.2.1.

Figure 3.2 depicts the timeline and three representative scenarios from our motivating example. Before week 52, patient 1 withdrew from the study due to relocation, which is plausibly unrelated to the biological effect of treatment. Patient 2 discontinued due to an adverse event, reflecting an effect-informative IE. Patient 3 completed the study through week 52, at which point the primary outcome was measured.

Scenarios in Figure 3.1 also highlight why a single, uniform estimand strategy can be misaligned with the clinical decision problem. Both types of IEs can result in an unmeasured primary outcome at the pre-specified landmark time: once an IE occurs, the outcome may be unobserved, and in some effect-informative cases such as death, it may not even be well-defined. If we apply the hypothetical strategy to all IEs, we implicitly ask an unrealistic question: “Does the drug work in a world where no one ever experienced clinically meaningful failures such as adverse events or lack of efficacy?” In reality, what decision-makers care about instead is: “Should we approve this drug, accounting for both possible benefit and clinically meaningful failures?” Conversely, if we apply a composite outcome strategy to all IEs, we would incorrectly penalize external disruptions such as relocation, and the resulting estimate would carry unnecessarily little information. Moreover, answering the unrealistic questions using a hypothetical strategy to all IEs typically requires untestable, strong, and thus hard-to-defend assumptions such as conditional independence between the time to an adverse event and the potential outcomes of interest given baseline covariates. Related concerns arise for principal stratification estimands, which target effects in latent subgroups and generally rely on untestable assumptions for identification. By contrast, our combined strategy is designed to align the estimand with regulatory decision questions while relying on more transparent assumptions. We formalize these assumptions and the corresponding identification results in Section 3.3.

Motivated by this mismatch and to ensure the estimand is clinically interpretable and relevant, we apply different strategies tailored to the nature of the IE. For IEs that are plausibly unrelated to the biological effect of the treatment (e.g., treatment discontinuation due to relocation or administrative withdrawal), we apply the *hypothetical strategy*, which imagines a scenario where the IE did not occur. This enables estimation of the treatment effect as if the patient had remained on treatment and in the study as planned, under the assumption that the IE time is conditionally independent of the outcome given treatment and baseline covariates. This assumption is reasonable in settings where the IE arises from external or administrative factors that are not influenced by post-randomization conditions. For example, if a patient relocates for personal reasons unrelated to their health status or treatment response, then, conditional on baseline characteristics and assigned treatment, the occurrence of this IE can be viewed as independent of the patient’s outcome. In contrast, for effect-informative IEs (e.g., treatment discontinuation due to adverse events or lack of efficacy), we apply the *composite variable strategy*, treating the IE itself as an indication of treatment failure, as such events signify that the patient is unable to continue treatment. A formal definition and discussion of this strategy are provided in Section 3.3. Combining the two strategies enhances the interpretability of the estimand and ensures it better reflects clinically relevant questions, as no single strategy can adequately address both types of IEs. Specifically, it would be inappropriate to classify a patient as a non-responder if they discontinue treatment due to relocation, and it is of limited clinical relevance to consider a hypothetical scenario in which adverse events do not occur. This approach is consistent with regulatory guidance, which recommends applying different strategies depending on the type

of IE involved (ICH E9 (R1), 2019; Kang et al., 2022; Kahan et al., 2024).

However, a central yet previously overlooked challenge is the presence of *competing IEs*, where the first IE censors the subsequent ones. For example, if a patient discontinues treatment due to relocation, as illustrated by patient 1 in Figure 3.2, we only observe that no effect-informative IE occurred before the discontinuation. However, it remains unknown whether an adverse event would have occurred before the final outcome measurement in the hypothetical scenario in which the patient had not relocated and had continued treatment as planned. Although such competing IEs are ubiquitous in practice, this issue has not, to our knowledge, been explicitly recognized or addressed in prior methodological work. A related recent paper, Olarte Parra et al. (2025a), also considers settings with two types of IEs whose causal ordering matters, and distinguishes different cases using directed acyclic graphs and single-world intervention graphs. Our focus is different. Rather than analyzing these through a graphical causal framework, we adopt a survival analysis and competing event perspective and explicitly model the event times of both IEs, which allows us to address settings in which one IE censors the observation of another.

Therefore, the central thesis of this paper is to address the challenge of competing IEs. We begin by developing a principled framework that clearly defines the estimand and establishes its nonparametric identification. Specifically, we derive two identification formulas, each relying on a different set of nuisance parameters. Building on these results, we propose two basic estimators corresponding to the two identification strategies. We then introduce an augmented estimator that combines the two, achieving double robustness. To further enhance both robustness and efficiency, we derive the efficient influence function (EIF) and construct an EIF-based estimator that attains the semiparametric efficiency bound under appropriate conditions. We apply our methods to the two systemic lupus erythematosus trials, both partially impacted by the COVID-19 pandemic, demonstrating the robustness and practical utility of the proposed framework.

3.1.3 Organization and notation

The remainder of the paper is organized as follows. In Section 3.2, we present a motivating example, introduce the basic setup of our research question, define the causal parameter of interest, and highlight the identification challenge. In Section 3.3, we state the key identification assumptions, establish the nonparametric identification of the causal parameter, and construct three estimators. In Section 3.4, we derive the EIF, propose an estimator that is both doubly robust and asymptotically efficient, and examine its asymptotic properties. In Section 3.5, we apply the methods to re-analyze the two systemic lupus erythematosus trials. Finally, in Section 3.6, we conclude with a discussion of two directions for future research. The supplementary material includes a simulation study and all proofs. We also provide publicly available R code implementing all four proposed estimators.

We use the following notation. Let $\|r\|_2 = \{\int r(o)^2 dP(o)\}^{1/2}$ denote the $L_2(P)$ norm where $P(\cdot)$ denotes the distribution of the observed data $O = o$. For the survival functions, let $\|r\|_2 = \{\iint r(t, o)^2 dP(o)dt\}^{1/2}$ denote the $L_2(P)$ norm. We write $b_n = O_P(a_n)$ if b_n/a_n is bounded in probability and $b_n = o_P(a_n)$ if b_n/a_n converges to 0 in probability.

3.2 Setup and estimand

3.2.1 Setup

Let A denote the binary treatment indicator, where $A = 1$ corresponds to assignment to the treatment group and $A = 0$ to the control group. The primary outcome, denoted by Y , is measured at a pre-specified time point k . In our motivating example, Y is the response index and k is 52 weeks. We adopt the potential outcomes framework, and write $Y(a)$ for the potential outcome under treatment assignment $A = a$, for $a \in \{0, 1\}$.

As discussed in Section 3.1, two types of IEs may occur after treatment initiation: effect-uninformative IEs and effect-informative IEs. Let C denote the time to an effect-uninformative IE and T denote the time to an effect-informative IE. Because both are post-treatment variables, we write $C(a)$ and $T(a)$ for their potential values under treatment assignment $A = a$, where $a \in \{0, 1\}$. Since our interest is restricted to events occurring before the outcome assessment time k , we set $C(a) = \infty$ if $C(a) > k$ and $T(a) = \infty$ if $T(a) > k$. The observed event times are therefore $C = C(A) = AC(1) + (1 - A)C(0)$ and $T = T(A) = AT(1) + (1 - A)T(0)$. In the observed data, we do not separately observe both post-treatment event times for the same individual. Rather, we only observe the earliest of the two event times before k , $C(A) \wedge T(A) \wedge k$. Let X denote the vector of collected baseline covariates.

Our estimand is motivated by the hypothetical strategy for effect-uninformative IEs, under which we ask what would have happened had such an IE not occurred. This indicates that C lies on the causal pathway from treatment to the outcome. We therefore introduce the nested potential outcome $Y(a, c)$, which denotes the outcome under treatment assignment $A = a$ if the effect-uninformative IE occurs at time c , with $c = \infty$ corresponding to no effect-uninformative IE before time k . For example, suppose a treated patient experiences an effect-uninformative IE at week 20 and consequently discontinues treatment. We will not observe their outcome of interest. However, even if we could send the patient back to the clinic at week 52 and get their primary outcome measured, that observed quantity would correspond to $Y(1, c = 20)$ rather than to the hypothetical outcome of interest, $Y(1, c = \infty)$, which represents the outcome that would have been observed had the effect-uninformative IE not occurred before time k .

3.2.2 Causal estimand and challenges in identification

We address the effect-informative IEs using the composite outcome strategy and address the effect-uninformative IEs using the hypothetical strategy. To combine them, we formally define a combined potential outcome as

$$Y^{\text{cp}}(a) = Y(a, c = \infty)1\{T(a) > k\} = \begin{cases} Y(a, c = \infty) & \text{if } T(a) > k, \\ 0 & \text{if } T(a) \leq k. \end{cases}$$

This construction defines the outcome as zero (i.e., failure) if an effect-informative IE occurs before the outcome is measured. We then define our causal estimand as the mean contrast:

$$\tau = E\{Y^{\text{cp}}(1) - Y^{\text{cp}}(0)\} = E[Y(1, c = \infty)1\{T(1) > k\}] - E[Y(0, c = \infty)1\{T(0) > k\}]. \quad (3.1)$$

This estimand addresses the effect-informative IEs using the *composite variable strategy*. It is widely used for binary outcomes to represent treatment success or failure, where effect-informative IEs such as adverse events or lack of efficacy are treated as failures (ICH E9 (R1), 2019). In our motivating example, $Y(a, c = \infty)$ indicates whether a patient would be a responder at week 52 under treatment assignment $A = a$, had no IE occurred. For patients who experience an effect-informative IE before week 52, $Y(a, c = \infty)$ is unobserved. Under the composite variable strategy, these patients are defined as non-responders.

The composite variable approach extends beyond binary outcomes, but its applicability depends on whether treatment failure can be represented by a clinically meaningful reference value. For ordinal or categorical outcomes, this is often feasible when there is a clearly defined worst category or failure category. For continuous outcomes, the approach is most compelling when a prespecified value v has a clear clinical interpretation as no benefit or treatment failure. Specifically, for outcomes such as chronic pain, physical functioning, and cognitive performance that are typically measured on an ordinal or continuous scale, if a predefined value v may be used to indicate treatment failure, then a composite outcome under treatment assignment $A = a$ can be constructed as $Y(a, c = \infty)1\{T(a) > k\} + v1\{T(a) \leq k\}$, where the outcome retains its actual value if no effect-informative IE occurs before time k , and takes the failure value v otherwise. The corresponding causal estimand can then be defined as in (3.1), with $Y(a, c = \infty)$ replaced by $Y(a, c = \infty) - v$ for $a = 0, 1$. For example, in settings where Y is percent weight loss from baseline, $v = 0$ could correspond to no weight loss, or where Y is improvement from baseline on a symptom score, $v = 0$ could correspond to no improvement.

At the same time, we acknowledge that for some continuous outcomes the choice of a single failure value may be less obvious and may not be universally acceptable. Rosenbaum, 2006 discussed a similar causal parameter in settings where the IE is death and the outcome of interest is a measure of quality of life assessed after a fixed period, where there may

be no unique value that fully captures the clinical meaning of an effect-informative IE. In such settings, the composite variable strategy remains a possible estimand choice, but its interpretability depends critically on whether the chosen reference value is clinically justified in advance. When no such value is well defined, the applicability of this strategy is more limited.

Identifying τ in the presence of both types of IEs presents a key methodological challenge. For clarity, in the remainder of this subsection, we use treatment discontinuation due to relocation and adverse events to represent effect-uninformative and effect-informative IEs, respectively. The causal estimand τ defined in (3.1) is based on the composite potential outcome $Y^{\text{cp}}(a)$, which takes the value 0 if an adverse event occurs before the outcome measurement, i.e., when $T(a) < k$. If adverse event were the only type of IE, and treatment is randomized, τ could be identified from observed data using the difference in means of the composite outcomes: $\tau = E\{Y1(T > k) \mid A = 1\} - E\{Y1(T > k) \mid A = 0\}$. However, this formula is infeasible in the presence of effect-uninformative IEs, which may occur before effect-informative IEs and thus censor both $T(a)$ and $Y(a, c = \infty)$. As illustrated in Figure 3.2, patient 1 discontinued due to relocation, which censored both the occurrence of a potential adverse event and the outcome, resulting in an unobserved composite outcome $Y1(T > k)$. Therefore, for patients who experience effect-uninformative IEs, the problem cannot be treated as standard censoring, and one cannot naïvely impute the outcome Y under a hypothetical scenario where the effect-uninformative IE did not occur. This is because such patients may still have experienced an effect-informative IE had they not discontinued due to relocation. In other words, a valid identification strategy must recover the expected value of the composite outcome $Y1(T > k)$, rather than the outcome Y alone.

Table 3.1 summarizes the observed IE types and corresponding composite outcomes, highlighting that competing IEs complicate the identification of τ . When no IE occurs, the outcome Y is observed and equals the composite outcome since $T > k$. When an adverse event is observed, the composite outcome is, by definition, 0. However, when a patient discontinues treatment due to relocation, both the outcome Y and the time to the adverse event T are unobserved, and so is the composite outcome. We address this challenge and present formal identification results in the following section.

Table 3.1: Summary of the observed IE types and outcome. C is the time to treatment discontinuation due to relocation, T is the time to adverse event, Y is the outcome of interest, and k is the pre-specified time point when the measurement of Y is taken. A question mark “?” indicates that the corresponding value is unobserved.

observed IE type	(T, C, k) -relationship	Y	$1(T > k)$	composite outcome Y^{cp}
no any type of IE	$C \wedge T > k$	Y	1	Y
adverse event	$C \wedge k > T$	?	0	0
treatment discontinuation due to relocation	$T \wedge k > C$	?	?	?

3.3 Nonparametric identification and basic estimators

3.3.1 Assumptions

We assume that the joint distribution of $\{X_i, A_i, T_i(1), T_i(0), C_i(1), C_i(0), Y_i(1, c = \infty), Y_i(0, c = \infty)\}$ for patient i is independently and identically distributed from a superpopulation. For notational simplicity, we omit the subscript i when there is no confusion.

We begin by stating the following assumption on treatment assignment.

Assumption 3.1 (Treatment assignment). *We assume the following conditional independence and overlap conditions:*

- (a) $A \perp\!\!\!\perp \{Y(a, c = \infty), T(a), C(a)\} \mid X$ for $a = 0, 1$.
- (b) For some constant $\eta \in (0, 0.5)$, $\eta < e(X) < 1 - \eta$ with probability 1, where $e(X) = \text{pr}(A = 1 \mid X)$ denotes the propensity score (Rosenbaum and Rubin, 1983).

Assumption 3.1 holds in RCTs by design, allowing our results to apply directly to settings such as our motivating example. However, our formulation is more general and can also accommodate cases where treatment assignment is not completely randomized, such as in stratified randomized experiments and observational studies.

Under Assumption 3.1, the causal estimand τ can be expressed as:

$$\begin{aligned} \tau &= E[E\{Y1(T > k) \mid A = 1, X\} - E\{Y1(T > k) \mid A = 0, X\}] \\ &= E\left\{ \frac{AY1(T > k)}{e(X)} - \frac{(1 - A)Y1(T > k)}{1 - e(X)} \right\}, \end{aligned}$$

which correspond to the standard outcome regression and inverse probability weighting identification formulas, respectively, when treating the combined potential outcome $Y^{\text{cp}}(a) = Y(a, c = \infty)1\{T(a) > k\}$ as the “new” outcome of interest. However, as discussed in Section 3.2.2, these expressions no longer serve as feasible identification formulas in the presence of competing effect-uninformative IEs. This is because $Y1(T > k)$ is not fully observed if an effect-uninformative IE occurs.

To address this challenge, we introduce additional assumptions on effect-uninformative IEs. Specifically, we assume that the potential time of an effect-uninformative IE under treatment a is conditionally independent of the potential outcomes and the potential time to effect-informative IEs, given baseline covariates. This assumption is analogous to the censoring-at-random assumption in survival analysis when treating effect-uninformative IEs as censoring

events (Robins et al., 1994; Tsiatis, 2006). Additionally, we also assume a standard positivity condition ensuring that the probability of no occurrence of effect-uninformative IE by time k , given observed covariates, is strictly positive. These conditions are summarized in the following assumption.

Assumption 3.2 (Effect-uninformative IE). *We assume the following conditional independence and positivity conditions for the time of effect-uninformative IE:*

- (a) $C(a) \perp\!\!\!\perp \{Y(a, c = \infty), T(a)\} \mid X$ for $a = 0, 1$.
- (b) For some constant $\eta_C > 0$, $\text{pr}\{C(a) > k \mid X\} > \eta_C$ with probability 1.

Assumption 3.2(a) states that the time to an effect-uninformative IE, such as treatment discontinuation due to relocation, is independent of both the potential outcome (i.e., whether a patient would be a responder at week 52) and the time to an effect-informative IE (e.g., an adverse event), conditional on observed baseline covariates. We also emphasize that this is not a blanket assumption for all IEs. Rather, it is intended only for events that, after clinical review, are judged to be plausibly unrelated to the patient’s underlying biological response to treatment. This assumption is most credible for IEs driven by external or administrative factors, such as relocation for personal or work reasons, pandemic-related disruptions, travel restrictions, or site closure for administrative reasons. In such settings, once we condition on treatment assignment and relevant baseline characteristics, the occurrence time of the event may reasonably be viewed as carrying little additional information about the patient’s latent outcome or latent time to an effect-informative IE. By contrast, the assumption is less credible for discontinuations related to adverse events, worsening symptoms, perceived lack of efficacy, or other evolving post-randomization clinical factors. Such events are therefore classified as effect-informative and incorporated into the composite outcome rather than handled under the hypothetical strategy. Thus, Assumption 3.2(a), while strong and untestable from observed data, provides a clear and clinically motivated criterion for classifying a subset of IEs as effect-uninformative.

Assumption 3.2(b) requires that every patient has a strictly positive probability of not experiencing an effect-uninformative IE before time k , given their covariates. This standard positivity condition rules out deterministic effect-uninformative IE happening in any subpopulation and is reasonable in our application.

If richer longitudinal covariates are available, Assumption 3.2(a) may be relaxed to a history conditional independence condition, in which $C(a)$ is assumed independent of $\{Y(a, c = \infty), T(a)\}$ conditional on the history of all time-varying covariates. In applications, it is also useful to assess robustness under alternative clinically meaningful classifications of ambiguous IE categories or through models that parameterize departures from conditional independence.

We revisit the former direction in Section 3.5.4 and leave detailed methodologies of the latter extension for future work.

3.3.2 Two identification formulas

Using the composite variable strategy to handle effect-informative IEs, recall that we define the causal estimand as in (3.1). Under Assumptions 3.1 and 3.2, we show that the estimand τ is nonparametrically identifiable and present two identification formulas in the following theorem.

Theorem 3.1 (Nonparametric identification of τ). *Under Assumptions 3.1 and 3.2, τ is nonparametrically identified by two distinct formulas. First,*

$$\tau = E \{ \mu_1(X) S_1(k | X) - \mu_0(X) S_0(k | X) \}, \quad (3.2)$$

where for $a = 0, 1$, $\mu_a(X) = E(Y | T \wedge C > k, X, A = a)$ is the conditional mean of observed outcome among those with no IE and $A = a$, and $S_a(t | X) = \text{pr}(T > t | X, A = a)$ is the conditional survival function of effect-informative IE in treatment arm a . Second,

$$\tau = E \left[\frac{AY1(T \wedge C > k)}{e(X)G_1(k | X)} - \frac{(1 - A)Y1(T \wedge C > k)}{\{1 - e(X)\}G_0(k | X)} \right], \quad (3.3)$$

where $e(X) = \text{pr}(A = 1 | X)$ is the propensity score and for $a = 0, 1$, $G_a(t | X) = \text{pr}(C > t | X, A = a)$ is the conditional survival function of effect-uninformative IE in treatment arm a .

The first identification formula (3.2) expresses τ using a standard outcome regression approach. It decomposes the conditional expectation of the composite outcome given covariates and treatment as $E\{Y1(T > k) | X, A\} = E(Y | T > k, X, A)\text{pr}(T > k | X, A) = E(Y | T \wedge C > k, X, A)\text{pr}(T > k | X, A)$. This formulation models the conditional mean outcome Y among individuals who do not have any IE by time k , and weights it by the probability of not experiencing an effect-informative IE by that time. The second identification formula (3.3) provides an alternative identification strategy based on inverse probability weighting. It identifies τ by reweighting observed outcomes among individuals who remain free of IEs up to time k , separately within each treatment arm. The weights adjust for the different probability of observing the composite outcome and consist of two components: $e(X)$, which accounts for differences in treatment assignment probabilities, and $G_a(k | X)$, which adjusts for the probability of remaining free of effect-uninformative IEs up to time k within each treatment group. Effectively, individuals with a lower probability of being observed are up-weighted, while those with a higher probability are down-weighted, ensuring an unbiased estimate of the treatment effect.

Both equations (3.2) and (3.3) are valid identification formulas, as the nuisance parameters involved are either functions of observed data or identifiable from observed data. Specifically, the propensity score $e(X)$ and the conditional outcome model $\mu_a(X)$ are functions of observed data. Although the survival functions for both types of IEs, $S_a(t | X)$ for effect-informative IEs and $G_a(t | X)$ for effect-uninformative IEs, are not direct functions of the observed data, they are identifiable under standard results in survival analysis with censoring at random (Robins and Rotnitzky, 1992; Robins and Finkelstein, 2000; Ebrahimi et al., 2003). For example, $S_1(t | X) = \text{pr}(T > t | X, A = 1)$ for $t \leq k$ can be estimated using data from the treatment group by viewing $T \wedge C \wedge k$ as the observed event time and using $1(C \wedge k \geq T)$ as the event indicator for the effect-informative IE. Estimation can then proceed using parametric or semiparametric methods, such as the Cox proportional hazards model (Cox, 1972), or more flexible approaches (Wolock et al., 2024). A similar approach can be used to estimate $G_1(t | X)$ by switching the roles of T and C . The corresponding functions $S_0(t | X)$ and $G_0(t | X)$ can be estimated in the same way using data from the control group.

3.3.3 Two basic estimators

The identification formulas (3.2) and (3.3) in Theorem 3.1 motivate two basic estimators, each relying on estimates of the corresponding nuisance functions. Let $\hat{S}_a(t | X)$ and $\hat{G}_a(t | X)$ denote the fitted survival models for effect-informative and effect-uninformative IEs, respectively, for $t \leq k$ and $a = 0, 1$. Let $\hat{\mu}_a(X)$ be the fitted outcome regression model for each treatment arm, and $\hat{e}(X)$ be the fitted propensity score model. By substituting these fitted values into the identification formulas and taking empirical analogs, we obtain the following two estimators:

$$\begin{aligned}\hat{\tau}^{\text{out}} &= n^{-1} \sum_{i=1}^n \hat{\mu}_1(X_i) \hat{S}_1(k | X_i) - n^{-1} \sum_{i=1}^n \hat{\mu}_0(X_i) \hat{S}_0(k | X_i), \\ \hat{\tau}^{\text{ipw}} &= n^{-1} \sum_{i=1}^n \frac{A_i Y_i 1(T_i \wedge C_i > k)}{\hat{e}(X_i) \hat{G}_1(k | X_i)} - n^{-1} \sum_{i=1}^n \frac{(1 - A_i) Y_i 1(T_i \wedge C_i > k)}{\{1 - \hat{e}(X_i)\} \hat{G}_0(k | X_i)}.\end{aligned}$$

The outcome regression estimator $\hat{\tau}^{\text{out}}$ is consistent if both $\hat{\mu}_a(X)$ and $\hat{S}_a(k | X)$ are consistently estimated. The inverse probability weighting estimator $\hat{\tau}^{\text{ipw}}$ is consistent if both $\hat{e}(X)$ and $\hat{G}_a(k | X)$ are consistently estimated.

3.3.4 An augmented estimator

Motivated by combining the two identification formulas (3.2) and (3.3), we introduce another estimator that augments the weighting estimator $\hat{\tau}^{\text{ipw}}$ using the estimated outcome models,

$$\hat{\tau}^{\text{aug}} = \hat{\tau}^{\text{ipw}} - n^{-1} \sum_{i=1}^n \left\{ \frac{A_i - \hat{e}(X_i)}{\hat{e}(X_i)} \hat{\mu}_1(X_i) \hat{S}_1(k | X_i) + \frac{A_i - \hat{e}(X_i)}{1 - \hat{e}(X_i)} \hat{\mu}_0(X_i) \hat{S}_0(k | X_i) \right\}.$$

It has a similar mathematical form to the classic augmented inverse probability weighting estimator for the average treatment effect (Bang and Robins, 2005). We provide the properties of $\hat{\tau}^{\text{aug}}$ in the following Proposition 3.1.

Proposition 3.1 (Double robustness of $\hat{\tau}^{\text{aug}}$). *Suppose Assumptions 3.1 and 3.2 hold, and assume that $G_a(k | X)$ is correctly specified for $a = 0, 1$. $\hat{\tau}^{\text{aug}}$ is a consistent estimator for τ if either $e(X)$ is correct, or both $\mu_a(X)$ and $S_a(k | X)$ are correct for $a = 0, 1$.*

The augmented estimator $\hat{\tau}^{\text{aug}}$ can be viewed as an intermediate bridge between the basic weighting estimator and the EIF-based estimator developed in the next section. Relative to $\hat{\tau}^{\text{ipw}}$, it improves robustness by incorporating the outcome model, but it still requires correct specification of the censoring model $G_a(k | X)$. We provide additional motivations and remarks in Section B.1.1 of the Supplementary Material.

3.4 A semiparametrically efficient and doubly robust estimator

The augmented weighting estimator $\hat{\tau}^{\text{aug}}$ improves robustness of $\hat{\tau}^{\text{ipw}}$. However, it is not robust to the misspecification of the survival function for effect-uninformative IE, nor does it achieve the semiparametric efficiency bound. In this section, we show that $\hat{\tau}^{\text{aug}}$ can be further improved. We derive the semiparametric efficient influence function (EIF) for τ to learn the best asymptotic efficiency a consistent estimator of τ can achieve and propose an asymptotically efficient and doubly robust estimator based on the EIF.

We first describe the full and observed data structures in our setting and introduce some additional notation. Ideally, we want to observe the full data $(X, Y^{\text{cp}}(1), Y^{\text{cp}}(0))$. The missingness of the full data comes from two strings. First, for a given treatment arm $a = 0, 1$, the effect-uninformative IE time $C(a)$ censors the full data because $Y^{\text{cp}}(a)$ is only observable if $C(a) \geq \{T(a) \wedge k\}$. Within each treatment group, we do not observe the full data due to censoring and only observe

$$(\Delta(a) = 1\{C(a) \geq T(a) \wedge k\}, \tilde{T}(a) = C(a) \wedge T(a) \wedge k, \Delta(a)Y^{\text{cp}}(a)).$$

Second, the treatment assignment generates another level of missingness, because, for each observation, we never simultaneously observe both composite potential outcomes $\{Y^{\text{cp}}(1), Y^{\text{cp}}(0)\}$ even without censoring. Therefore, the observed data is

$$O = (A, \Delta = \Delta(A), \tilde{T} = \tilde{T}(A), \Delta Y^{\text{cp}} = \Delta(A)Y^{\text{cp}}(A)).$$

3.4.1 EIF and EIF-based estimator

In the following theorem, we provide the EIF for τ .

Theorem 3.2 (EIF for τ). *Under the nonparametric model, the EIF for μ_1 is*

$$\begin{aligned} D_1(O) = & \frac{A}{e(X)} \left\{ \frac{Y1(T \wedge C > k)}{G_1(k | X)} + \mu_1(X)S_1(k | X) \int_0^{\tilde{T}} \frac{dM_{G_1}(t)}{S_1(t | X)G_1(t | X)} \right\} \\ & - \frac{A - e(X)}{e(X)} \mu_1(X)S_1(k | X) - \mu_1, \end{aligned} \quad (3.4)$$

the EIF for μ_0 is

$$\begin{aligned} D_0(O) = & \frac{1 - A}{1 - e(X)} \left\{ \frac{Y1(T \wedge C > k)}{G_0(k | X)} + \mu_0(X)S_0(k | X) \int_0^{\tilde{T}} \frac{dM_{G_0}(t)}{S_0(t | X)G_0(t | X)} \right\} \\ & - \frac{e(X) - A}{1 - e(X)} \mu_0(X)S_0(k | X) - \mu_0, \end{aligned} \quad (3.5)$$

and thus, the EIF for τ is $D_\tau(O) = D_1(O) - D_0(O)$, where $dM_{G_a}(t) = 1(C \in dt, \Delta = 0) - 1(\tilde{T} \geq t)d\Lambda_a(t | X)$ with $\Lambda_a(t | X)$ denoting the conditional cumulative hazard function for the effect-uninformative IE C in the treatment group $A = a$ for $a = 0, 1$.

The $M_{G_a}(t)$ in the EIF is the martingale constructed from the censoring counting process. Intuitively, $1(C \in dt, \Delta = 0)$ is the actual observed increment in the censoring counting process at time t , which records whether a censoring event has occurred, while $1(\tilde{T} \geq t)d\Lambda_a(t | X)$ represents the expected increment in the counting process, given the history up to time t . The martingale $M_{G_a}(t)$ captures the difference between the actual observed events and their expected occurrences. The EIF implies another identification formula for τ by the property that $E\{D_\tau(O)\} = 0$. Rearranging terms, we have

$$\begin{aligned} \tau = & E \left\{ \frac{AY1(T \wedge C > k)}{e(X)G_1(k | X)} - \frac{A - e(X)}{e(X)} \mu_1(X)S_1(k | X) \right\} \\ & - E \left[\frac{(1 - A)Y1(T \wedge C > k)}{\{1 - e(X)\}G_0(k | X)} - \frac{e(X) - A}{1 - e(X)} \mu_0(X)S_0(k | X) \right] \\ & + E \left[\frac{A}{e(X)} \mu_1(X)S_1(k | X) \int_0^{\tilde{T}} \frac{dM_{G_1}(t)}{S_1(t | X)G_1(t | X)} \right] \\ & - E \left[\frac{1 - A}{1 - e(X)} \mu_0(X)S_0(k | X) \int_0^{\tilde{T}} \frac{dM_{G_0}(t)}{S_0(t | X)G_0(t | X)} \right], \end{aligned} \quad (3.6)$$

where the first two lines are the same as in the augmented weighting identification formula. Intuitively, the last two terms in (3.6) provide additional correction to achieve robustness to misspecification of the survival function for effect-uninformative IE. These augmentation terms are zero at the population level under true $G_a(t | X)$. However, their empirical counterparts may deviate significantly from zero, providing diagnostic insight into possible

misspecification of $G_a(t | X)$. The martingale term in Theorem 3.2 is closely related to EIFs for right-censored survival data, such as Westling et al. (2024), because the nuisance tangent space for the censoring mechanism $C | A, X$ is the same as in standard censoring settings. The main difference is that our target is a treatment effect on a primary endpoint at a fixed horizon under a hybrid estimand with both effect-uninformative and effect-informative IEs, which introduces additional nuisance components, including the outcome regression and the survival model for the effect-informative IE.

The EIFs in Theorem 3.2 motivate the following estimator for τ ,

$$\begin{aligned} \hat{\tau}^{\text{eif}} &= \hat{\tau}^{\text{aug}} + n^{-1} \sum_{i=1}^n \frac{A_i}{\hat{e}(X_i)} \hat{\mu}_1(X_i) \hat{S}_1(k | X_i) \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_1}(t)}{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)} \\ &\quad - n^{-1} \sum_{i=1}^n \frac{1 - A_i}{1 - \hat{e}(X_i)} \hat{\mu}_0(X_i) \hat{S}_0(k | X_i) \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_0}(t)}{\hat{S}_0(t | X_i) \hat{G}_0(t | X_i)}. \end{aligned}$$

To gain insights, consider the first augmentation term in $\hat{\tau}^{\text{eif}}$. Since the observed time points are discrete, the integral for observation i can be estimated as follows:

$$\begin{aligned} \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_1}(t)}{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)} &= \sum_{t \leq \tilde{T}_i} \frac{1(\Delta_i = 0, C_i = t) - \hat{\lambda}_1(t | X_i)}{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)} \\ &= - \sum_{t \leq \tilde{T}_i} \frac{\hat{\lambda}_1(t | X_i)}{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)} + \frac{1(\Delta_i = 0)}{\hat{S}_1(\tilde{T}_i | X_i) \hat{G}_1(\tilde{T}_i | X_i)}, \end{aligned} \quad (3.7)$$

where $\hat{\lambda}_1(t | X_i)$ denotes the estimated conditional hazard function for the effect-uninformative IE at time t given covariates X_i . In (3.7), the first term is a sum of the ratio $-\hat{\lambda}_1(t | X_i)/\{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)\}$, where the summation is over all observed event time points before $\tilde{T}_i$. The second term is 0 for individuals who are not right-censored by the effect-uninformative IE, and is $1/\{\hat{S}_1(\tilde{T}_i | X_i) \hat{G}_1(\tilde{T}_i | X_i)\}$ for observations with an effect-uninformative IE occurred at time $\tilde{T}_i$.

To construct the estimator $\hat{\tau}^{\text{eif}}$, we need to estimate the following nuisance parameters: the propensity score model $e(X)$, the outcome model $\mu_a(X)$ for $a = 0, 1$, and the survival functions for effect-informative and effect-uninformative IEs $S_a(t | X)$ and $G_a(t | X)$, respectively, for $t \leq k$ and $a = 0, 1$. Importantly, as shown in Theorem 3.3, the consistency of $\hat{\tau}^{\text{eif}}$ does not require correct specification of all four nuisance parameters.

3.4.2 Asymptotic properties of the EIF-based estimator

We next discuss the asymptotic properties of $\hat{\tau}^{\text{eif}}$. We first introduce additional notation. For $t \leq k$ and $a = 0, 1$, let e^* , G_a^* , μ_a^* , and S_a^* denote the probability limit of the estimated

nuisance functions $\hat{e}$, $\hat{G}_a$, $\hat{\mu}_a$, $\hat{S}_a$, respectively, i.e., $\|\hat{e} - e^*\| = o_P(1)$, $\|\hat{G}_a - G_a^*\| = o_P(1)$, $\|\hat{\mu}_a - \mu_a^*\| = o_P(1)$, and $\|\hat{S}_a - S_a^*\| = o_P(1)$. If a given nuisance model is correctly specified, its corresponding limit equals the true function. For example, if the propensity score model is consistent, then $e^* = e$, and similar results hold for the other three nuisance components.

In the following theorem, we provide the double robustness of the EIF-based estimator $\hat{\tau}^{\text{eif}}$.

Theorem 3.3 (Double robustness of $\hat{\tau}^{\text{eif}}$). *Under Assumptions 3.1 and 3.2, $\hat{\tau}^{\text{eif}}$ is doubly robust in the sense that it is consistent for τ if either $\{e^*(X) = e(X), G_a^*(t | X) = G_a(t | X)\}$ or $\{\mu_a^*(X) = \mu_a(X), S_a^*(t | X) = S_a(t | X)\}$ for $t \leq k$ and $a = 0, 1$.*

Theorem 3.3 shows that $\hat{\tau}^{\text{eif}}$ is consistent if at least one of the following two sets of nuisance parameters are consistently estimated: (1) the outcome model $\mu_a(X)$ and the survival function for the effect-informative IE $S_a(t | X)$; (2) the propensity score model $e(X)$ and the survival function for the effect-uninformative IE $G_a(t | X)$, for $t \leq k$ and $a = 0, 1$. When our data is from an RCT, i.e., the propensity score $e(X)$ is known by design, so $\hat{e}(X)$ can be correctly specified. Consequently, the EIF-based estimator $\hat{\tau}^{\text{eif}}$ is a consistent estimator for τ if either $G_a(t | X)$ is consistently estimated, or both $\mu_a(X)$ and $S_a(t | X)$ are consistently estimated, for $t \leq k$ and $a = 0, 1$.

To conduct statistical inference, we provide the asymptotic distribution of $\hat{\tau}^{\text{eif}}$. We first introduce three technical conditions and then state the asymptotic result in a theorem.

Assumption 3.3 (Consistency of the nuisance parameters). *Assume that either $\{e^*(X) = e(X), G_a^*(t | X) = G_a(t | X)\}$ or $\{\mu_a^*(X) = \mu_a(X), S_a^*(t | X) = S_a(t | X)\}$ for $t \leq k$ and $a = 0, 1$ is satisfied.*

Assumption 3.4 (Donsker Condition). *The class of functions $\{(e, G_a, \mu_a, S_a) : \|e - e^*\| < \delta, \|G_a - G_a^*\| < \delta, \|\mu_a - \mu_a^*\| < \delta, \|S_a - S_a^*\| < \delta\}$ is Donsker for some $\delta > 0$.*

Assumption 3.5 (Convergence rates of nuisance parameters). *The convergence rate of the nuisance parameters estimation satisfies $\{\|\hat{e} - e^*\| + \|\hat{G}_a - G_a^*\|\}\{\|\hat{\mu}_a - \mu_a^*\| + \|\hat{S}_a - S_a^*\|\} = o_P(n^{-1/2})$ for $a = 0, 1$.*

Assumption 3.3 requires either the propensity score and survival function of the effect-uninformative IE or the outcome and survival function of the effect-informative IE to be consistently estimated. It guarantees the consistency of $\hat{\tau}^{\text{eif}}$. Assumption 3.4 imposes restrictions on the nuisance model complexity and is a standard regularity condition (Vaart, 2000). We can also employ flexible machine learning models with cross-fitting techniques in the estimation of nuisance parameters to relax the Donsker condition (Pfanzagl and Wefelmeyer, 1985; Klaassen, 1987; Zheng and Laan, 2011; Chernozhukov et al., 2018). We provide a more detailed procedure in Section B.1.4 of the Supplementary Material, leaving a

full theory and empirical study of implementations to future work. Finally, Assumption 3.5 imposes additional restrictions on the rate of convergence of the nuisance parameters, in addition to all of them being consistently estimated.

Theorem 3.4 (Asymptotic distribution). *Under Assumptions 3.1– 3.5, the EIF-based estimator satisfies*

$$n^{1/2}(\hat{\tau}^{\text{eif}} - \tau) = n^{-1/2} \sum_{i=1}^n D_{\tau}(O_i) + o_P(1).$$

The estimator $\hat{\tau}^{\text{eif}}$ is a consistent and asymptotically Normal estimator of τ with asymptotic variance equal to $E\{D_{\tau}^2(O)\}$, thus achieving the semiparametric efficiency bound.

Therefore, we can construct the variance estimator based on the semiparametric efficiency bound by taking the empirical analog of the plug-in estimation $n^{-1} \sum_{i=1}^n \hat{D}_{\tau}^2(O_i)$. We can also use a nonparametric bootstrap to estimate variance and conduct statistical inference if the nuisance estimators satisfy certain smoothness conditions.

3.4.3 Comparing the estimators

In real data analysis, we recommend implementing and comparing all four proposed estimators to assess whether they provide coherent scientific implications. Theoretical comparisons among these estimators primarily focus on two key dimensions: robustness to nuisance model misspecification and asymptotic efficiency. In practice, examining how the point estimates differ across these estimators can provide insight into the presence of nuisance model misspecification. Specifically, certain pairwise differences among the estimators can serve as informal diagnostics. For example, if $e(X)$ is correctly specified, the difference between $\hat{\tau}^{\text{ipw}}$ and $\hat{\tau}^{\text{aug}}$ should converge to zero. Similarly, if $\mu_a(X)$ and $S_a(k | X)$ are correctly specified, the difference between $\hat{\tau}^{\text{out}}$ and $\hat{\tau}^{\text{aug}}$ should converge to zero. Finally, if $G_a(t | X)$ is correctly specified, the difference between $\hat{\tau}^{\text{eif}}$ and $\hat{\tau}^{\text{aug}}$ should converge to zero. Therefore, when implementing all four estimators in practice, the observed difference among them may serve as an informal diagnostic tool to detect possible nuisance model misspecification. We provide more details in Section B.1.2 in the supplementary material.

Among the four estimators, a key comparison lies between the EIF-based estimator $\hat{\tau}^{\text{eif}}$ with the augmented weighting estimator $\hat{\tau}^{\text{aug}}$. On the one hand, $\hat{\tau}^{\text{eif}}$ is more robust and asymptotically efficient: it allows for misspecification of the censoring model $C | A, X$ and it achieves the semiparametric efficiency bound under correct nuisance specification. Continuing the intuition discussed in Section 3.3.4, $\hat{\tau}^{\text{aug}}$ only orthogonalizes the part corresponding to $A | X$ but not $C | X, A$, making it robust only to misspecification of the $A | X$ model, not the $C | X, A$ model. In contrast, $\hat{\tau}^{\text{eif}}$ accounts for both possible sources of misspecification, leading to greater robustness and efficiency. On the other hand, $\hat{\tau}^{\text{aug}}$ is more straightforward to

implement in practice, as it only requires the estimation of $S_a(k | X)$ and $G_a(k | X)$ at time k for $a = 0, 1$, while $\hat{\tau}^{\text{eif}}$ requires consistent estimation of the entire survival curves $S_a(t | X)$ and $G_a(t | X)$ for $t \leq k$ and $a = 0, 1$. We consider the choice between $\hat{\tau}^{\text{eif}}$ and $\hat{\tau}^{\text{aug}}$ as a trade-off between implementation simplicity and robustness and efficiency of the estimators.

Finally, we note the theoretical possibility of applying a martingale-based correction to the basic weighting estimator $\hat{\tau}^{\text{ipw}}$, in a manner analogous to the correction in the EIF-based estimator $\hat{\tau}^{\text{eif}}$ compared with $\hat{\tau}^{\text{aug}}$. Such a corrected estimator would be robust to misspecification of $G_a(t | X)$. However, it remains less attractive in practice: it is neither robust to misspecification of the propensity score nor asymptotically efficient, and it lacks the implementation simplicity of $\hat{\tau}^{\text{aug}}$. For these reasons, we do not recommend it, though we include its explicit form in Section B.1.3 of the supplementary material only for theoretical completeness.

3.4.4 Simulation results

We conduct Monte Carlo simulations to evaluate the finite-sample performance of four estimators under various combinations of correctly specified and misspecified nuisance models. Consistent with our theory, $\hat{\tau}^{\text{eif}}$ demonstrates the greatest robustness across all misspecification scenarios, maintaining low bias and valid coverage even when multiple nuisance models are incorrect. We relegate the detailed results to Section B.3 in the supplementary material.

3.5 Analyzing two phase-3 trials

3.5.1 Data analysis

In this section, we re-analyze data from two trials on systemic lupus erythematosus (Morand et al., 2023; Petri et al., 2023). Both trials are double-blinded, randomized, placebo-controlled phase-3 trials. There are three treatment arms: 2mg baritinib, 4mg baritinib, and placebo. The primary outcome in both studies is the Systemic lupus erythematosus Responder Index 4 (SRI4) at week 52, a binary composite endpoint that reflects clinical response. A participant is classified as a responder if they show a predefined improvement in disease activity, without overall worsening or the emergence of significant disease activity in new organ systems. Baseline covariates used for model fitting include geographic region, baseline corticosteroid use, and the Physician’s Global Assessment score measured at baseline. We provide detailed baseline characteristics in Section B.2.2 of the Supplementary Material. Although the two trials each have three treatment arms, our current framework is formulated for binary treatment contrasts. Accordingly, in the data analysis we conduct the two standard pairwise comparisons against placebo, namely 2mg versus placebo and 4mg versus placebo. A fully joint multi-arm formulation is a natural extension, but is beyond the

scope of the current paper.

A substantial proportion of participants in both trials experienced IEs: 218 out of 760 in one trial and 211 out of 775 in the other. As described in Section 3.1.2, we classify these IEs into two categories: effect-informative IEs (82.6% and 84.4% in the two trials, respectively) and effect-uninformative IEs (17.4% and 15.6%). We conduct separate comparisons of the 2mg and 4mg baricitinib treatment arms versus the placebo group, yielding four pairwise analyses with sample sizes ranging from 505 to 517. Among participants who experienced an effect-informative IE, the median time to IE ranges from 16 to 23 weeks. Effect-uninformative IEs are sparse (3–8% across arms) and tend to occur slightly later (median 18–25 weeks). We present the Kaplan–Meier-based cumulative incidence of each IE type by arm and trial in Section B.2.2 of the Supplementary Material.

For each comparison, we apply all four estimators: $\hat{\tau}^{\text{out}}$, $\hat{\tau}^{\text{ipw}}$, $\hat{\tau}^{\text{aug}}$, and $\hat{\tau}^{\text{EIF}}$. We use logistic regression to estimate the propensity score and outcome models, and use Cox proportional hazards regression to separately estimate the survival functions for the two types of IEs. We conduct variance estimation and inference using a nonparametric bootstrap with 500 replicates. Table 3.2 presents the resulting point estimates, bootstrap standard errors, p -values, and 95% confidence intervals. We did not apply a formal multiplicity adjustment in this data analysis because the comparisons are presented primarily for descriptive and methodological illustration rather than as a confirmatory family of hypothesis tests.

Table 3.2: Treatment effect of Baricitinib on the primary outcome SRI4. The first four columns report the results of Trial 1 and the last four columns report those of Trial 2. Within each trial, the first five rows report the point estimate, bootstrap standard error, p -value, and lower and upper 95% confidence limits for each estimator of the 2mg treatment arm, and the last five rows report those of the 4mg treatment arm. For each trial and each treatment arm, we report the results based on all four estimators.

		Trial 1 (Petri et al., 2023)				Trial 2 (Morand et al., 2023)			
		$\hat{\tau}^{\text{out}}$	$\hat{\tau}^{\text{ipw}}$	$\hat{\tau}^{\text{aug}}$	$\hat{\tau}^{\text{EIF}}$	$\hat{\tau}^{\text{out}}$	$\hat{\tau}^{\text{ipw}}$	$\hat{\tau}^{\text{aug}}$	$\hat{\tau}^{\text{EIF}}$
2mg	point	0.030	0.029	0.026	0.026	0.019	0.019	0.022	0.022
	se	0.042	0.043	0.042	0.043	0.042	0.043	0.043	0.042
	p -value	0.479	0.504	0.534	0.540	0.643	0.662	0.602	0.606
	CI lower	−0.053	−0.055	−0.057	−0.057	−0.062	−0.065	−0.061	−0.061
	CI upper	0.113	0.113	0.110	0.110	0.101	0.102	0.106	0.104
4mg	point	0.113	0.120	0.115	0.113	−0.002	−0.002	−0.002	−0.002
	se	0.046	0.046	0.046	0.046	0.042	0.042	0.042	0.042
	p -value	0.013	0.008	0.012	0.013	0.961	0.962	0.966	0.969
	CI lower	0.023	0.030	0.025	0.023	−0.084	−0.084	−0.084	−0.084
	CI upper	0.204	0.209	0.204	0.203	0.080	0.080	0.081	0.080

3.5.2 Interpretation of the statistical results

Across both trials and treatment arms, the point estimates and p -values from the four methods are generally consistent. Our results suggest a statistically significant positive effect of 4mg baricitinib on the SRI4 outcome in Trial 1, while in Trial 2, the effect is not significant and the point estimate is slightly negative. No significant treatment effect is observed for the 2 mg baricitinib dose in either trial. As discussed in Section 3.4.3, though not a formal statistical test, qualitatively, the robustness of estimates across the four estimators in Table 3.2 provides empirical evidence that our nuisance models may not be severely misspecified.

In these two immunology trials, our estimand has a direct clinical interpretation. For each baricitinib versus placebo comparison, it targets the difference in the probability of week-52 SRI4 response under a hybrid strategy for handling IEs. If a patient experiences an effect-informative IE before week 52, such as discontinuation due to adverse events or lack of efficacy, that patient is counted as a non-responder under the composite strategy. If a patient experiences an effect-uninformative IE, such as relocation, a pandemic-related disruption, or sponsor-driven site closure, the estimand instead asks what the week-52 SRI4 outcome would have been had that external disruption not occurred. Thus, the estimand addresses the following clinically relevant question: *What is the treatment effect on week-52 SRI4 if external disruptions had not occurred, while still counting clinically meaningful treatment failures as failures?* For example, the estimated effect of about 0.11 for the 4mg arm in Trial 1 implies that, under this hybrid strategy, 4mg baricitinib increases the probability of week-52 SRI4 response by approximately 11% points relative to placebo.

3.5.3 Comparison with ad-hoc methods

Next, we compare our proposed methods with two commonly used ad-hoc methods. The first method is the *non-responder rule* (often referred to as “non-responder imputation” in the applied literature). We adopt the term “rule” instead of “imputation” to emphasize that this approach redefines the endpoint to take the value 0 upon an IE, rather than treating the outcome as missing and imputing it. It is equivalent to using the composite outcome strategy that treats all IEs as effect-informative and assigns the failure value 0 upon any IE. The second method naively applies a hypothetical strategy to all IEs, assuming they are independent of potential outcomes conditional on observed covariates.

Table 3.3 reports the estimated treatment effects of both treatment arms across the two trials using the inverse probability weighting estimator within each approach. We would like to emphasize that this is *not* merely a comparison of different estimators for a common estimand. Instead, the three approaches correspond to different choices of estimand because they classify and handle IEs in different ways. Therefore, the results illustrate how the estimated treatment effect changes when different IE classifications, estimand strategies,

and identifying assumptions are adopted.

The non-responder rule method generally yields slightly smaller effect estimates than our proposed methods, although the differences are modest, likely due to the relatively low proportion of effect-uninformative IEs in both trials. Nevertheless, our proposed approach targets a clinically more meaningful causal estimand. The second ad hoc method naively applies the hypothetical strategy to all IEs, assuming that their occurrence is conditionally independent of the potential outcomes given baseline covariates. This assumption is untestable and likely violated in practice. For example, treatment discontinuation due to adverse events or lack of efficacy is plausibly related to a patient's potential outcome. As shown in Table 3.3, this method produces treatment effect estimates that differ substantially from those obtained using our proposed approach, potentially leading to misleading clinical conclusions. These empirical findings are consistent with the large-sample bias observed in the simulation study and underscore the importance of using appropriate strategies tailored to different types of IEs.

Table 3.3: Treatment effect of Baricitinib on the primary outcome SRI4 using non-responder rule (NRR) and hypothetical strategy (HS). The first two columns report the results of Trial 1 and the last two columns report those of Trial 2. The first five rows report the point estimate, bootstrap standard error, p -value, and lower and upper 95% confidence limits for the 2mg treatment arm, and the last five rows report those of the 4mg treatment arm. For each trial and each treatment arm, we report the results based on the inverse probability-weighting estimator.

		Trial 1		Trial 2	
		NRR	HS	NRR	HS
2mg	point	0.039	0.011	0.010	0.022
	se	0.043	0.049	0.042	0.050
	p -value	0.362	0.815	0.817	0.654
	CI lower	-0.045	-0.085	-0.072	-0.076
	CI upper	0.123	0.107	0.092	0.120
4mg	point	0.093	0.073	-0.008	0.044
	se	0.044	0.050	0.041	0.051
	p -value	0.036	0.147	0.847	0.382
	CI lower	0.007	-0.025	-0.088	-0.056
	CI upper	0.179	0.171	0.072	0.144

3.5.4 Robustness checks

As a robustness check to the proportional hazards assumption embedded in the Cox nuisance models, we replaced all four Cox models with log-normal accelerated failure time (AFT)

models, which impose no proportional hazards constraint. The AFT survival curves are evaluated analytically at all observed follow-up times and fed into the same martingale integral computation as the Cox-based EIF estimator. The AFT sensitivity results are nearly identical to the Cox-based results across all four comparisons: for the 4mg arm in Trial 1, the AFT EIF estimate is 0.113 (95% CI: [0.023, 0.203]), identical to the Cox-based result. Full results are reported in Section B.2.3 of the Supplementary Material. This close agreement suggests that the conclusions are robust to the choice of parametric survival model for the nuisance functions.

As a further robustness check on the IE classification and on how potential misspecifications affect the estimands and estimates, we reclassified IE categories under alternative clinically possible rules and re-estimated the treatment effects. This robustness check probes the sensitivity of the results to the classification decisions. Under all scenarios we tested, the estimated treatment effects and their confidence intervals remain close to the primary analysis results, and the statistical conclusion for the 4mg baricitinib arm in Trial 1 is preserved. Full results are reported in Section B.2.4 of the Supplementary Material.

3.5.5 Beyond τ : causal risk ratio and odds ratio

For binary composite outcomes, our framework also extends directly to other effect measures, such as the causal risk ratio and odds ratio, by first estimating the arm-specific mean composite outcomes and then applying the corresponding smooth transformation. The associated EIFs follow from those for the arm-specific means via the delta method. Although mainly focusing on the risk difference, τ , in the main paper, we also implement causal risk ratio and odds ratio estimates for all four comparisons and report the full results in Section B.2.5 of the Supplementary Material.

3.6 Future directions

In this study, we address the challenges posed by IEs in RCTs by proposing methods to handle competing IEs. We classify IEs into effect-informative and effect-uninformative events and apply different strategies to identify a clinically meaningful causal effect. For effect-informative IEs, which are often informative about a patient’s outcome, we use a composite variable strategy that assigns an outcome value indicative of treatment failure. For effect-uninformative IEs, we apply a hypothetical strategy, assuming their timing is conditionally independent of the outcome given treatment and baseline covariates, and envisioning a scenario in which such events do not occur. The central thesis of this paper is to address the challenge of competing IEs, where the first IE censors all subsequent ones. In this paper, we propose a principled framework that carefully formulates the estimand, establishes its non-parametric identification and semiparametric estimation theory, and introduces weighting,

outcome regression, and doubly robust estimators. While our proposed framework provides a rigorous and flexible approach for handling competing IEs in RCTs, several challenges and extensions remain open for future research.

3.6.1 Data collection and IE classification

Our proposed methods have broad applicability across various therapeutic areas, including immunology, oncology, and cardiology, where treatment discontinuation and other IEs frequently occur. However, the effectiveness of these approaches relies on the accurate classification of IEs, which should be performed in collaboration with clinicians to ensure clinical relevance in trial analyses. To support this, a modernized case report form is needed to enable more granular documentation of the timing, reasons, and magnitude of treatment discontinuation. For example, the reason for the event should be specified, such as discontinuation due to toxicity versus lack of efficacy. In some cases, the event may need to meet a threshold of magnitude, such as the use of additional medication exceeding a specified duration or dose. Additionally, the timing of the event may be relevant, particularly in relation to its proximity to outcome assessment. Fortunately, a cross-industry PHUSE working group is tackling this problem PHUSE Working Group, 2025.

3.6.2 Random assessment time K

In practice, the measurement of the outcome of interest Y does not necessarily happen at a fixed point k , but within a prespecified assessment window. For example, a protocol may define the primary endpoint at week 52 while allowing the final study visit to occur within a window such as weeks 50–54. In that case, the actual visit time at which the outcome is measured varies across patients. Let K denote this realized assessment time. The outcome remains the same clinical endpoint of interest, but it is recorded at the patient-specific visit time K rather than at a common fixed time k . If this variability in visit timing is driven by protocol logistics or other external factors that are unrelated to treatment assignment, potential outcomes, and IEs, then K can be treated as a random variable independent of treatment A , potential outcomes $Y(a, c = \infty)$, and both $T(a)$ and $C(a)$ for $a = 0, 1$. Under this extension, the analogous versions of the identification formulas (3.2) and (3.3) are obtained by replacing k with K , i.e.,

$$\tau = E \{ \mu_1(X) S_1(K | X) - \mu_0(X) S_0(K | X) \} \quad (3.8)$$

$$= E \left[\frac{AY1(T \wedge C > K)}{e(X)G_1(K | X)} - \frac{(1 - A)Y1(T \wedge C > K)}{\{1 - e(X)\}G_0(K | X)} \right]. \quad (3.9)$$

If K is a pre-treatment covariate that is observed, the proposed four estimators carry over by replacing the fixed k with the observed values of K . In contrast, if K is itself affected and is

a post-treatment variable, we can construct a weighting estimator following (3.9). However, (3.8) no longer provides a feasible identification formula unless $S_a(K | X)$ is identified for $a = 0, 1$. Considering the presence of three competing time-to-event random variables, it becomes necessary to employ competing risks models. We defer this analysis to future work.

Chapter 4

Combining treatment policy and hypothetical strategies to handle semi-competing intercurrent events

4.1 Introduction

Intercurrent events (IEs) are post-treatment events that affect either the interpretation or the existence of the outcome measurement relevant to a clinical question. In long-term randomized trials, such events are usually unavoidable. Patients may discontinue their assigned treatment, use rescue medication, change dose, receive additional background therapy, withdraw from follow-up, relocate, or experience events that make the final outcome unavailable or no longer directly interpretable. The ICH E9(R1) addendum formalized the estimand framework precisely to avoid mixing such post-treatment complications with missing data problems. Instead, the clinical question should specify how each relevant IE is handled when defining the treatment effect of interest. The addendum introduced five broad strategies for handling IEs: treatment policy, hypothetical, composite outcome, while-on-treatment, and principal stratum strategies (ICH E9 (R1), 2019). Two strategies are especially relevant for the setting considered in this paper: treatment policy and hypothetical strategies. Under a treatment policy strategy, the outcome is defined and interpreted regardless of whether the IE occurs, and the IE is treated as part of the post-randomization clinical practice induced by treatment assignment. Under a hypothetical strategy, we are interested in the potential outcome under a counterfactual scenario in which the IE is set not to occur before the outcome assessment.

This distinction is especially important when different types of IEs occur in the same trial. Some IEs are part of the treatment pathway that would naturally occur in clinical practice. For example, a patient may require dose adjustment, use rescue medication, or adopt

additional disease-management strategies because of insufficient response or tolerability concerns. These events are often informative about the treatment process and may be part of the real-world effect of assigning treatment, and therefore should be handled using a treatment policy strategy. Other IEs are external to the scientific question of interest. For example, a patient may relocate, withdraw for administrative reasons, or miss the final visit for reasons unrelated to treatment efficacy. For these events, it may be more meaningful to ask what the outcome would have been *had the event not occurred*, suggesting the use of a hypothetical strategy. Using a single strategy for all IEs can therefore make the clinical interpretation of the estimand unclear: a treatment policy strategy for all IEs may include noise from events unrelated to efficacy, while a hypothetical strategy for all IEs may remove post-treatment features that are actually part of real-world treatment use.

4.1.1 Motivating example

Consider a long-term trial of a pharmacological treatment for obesity, such as the SURMOUNT-1 trial of tirzepatide. SURMOUNT-1 was a phase 3, multicenter, double-blind, placebo-controlled trial in adults with obesity or overweight without diabetes, with body weight measured after a long follow-up period (Jastreboff et al., 2022). Such type of trials provide a useful setting for studying IEs because obesity is a chronic condition, treatment is administered over many months, and the endpoint is measured well after treatment initiation. During follow-up, patients may experience post-treatment events that affect either the interpretation or the observation of the final weight measurement.

The central issue is that not all IEs play the same scientific role. Some events, such as dose modification, rescue medication use, or additional disease-management strategies, are closely tied to treatment response, tolerability, and routine clinical management. These events are not merely nuisances. They are part of the clinical trajectory after treatment initiation and may affect the final outcome in ways that matter to patients, clinicians, and regulators. If the question is: “What weight loss should we expect when this treatment is prescribed in practice, allowing for treatment modification and patient-driven management during follow-up?”, then these events should be handled using a treatment policy strategy. Under this strategy, the outcome is interpreted regardless of whether such events occur, because their occurrence is considered part of the treatment condition being evaluated. Other events, such as relocation, withdrawal because of a family emergency, administrative site closure, or missingness of the final visit for reasons unrelated to obesity, are external to the scientific question. For these events, a hypothetical strategy is more natural: the target outcome is the one that would have been observed had the external event not occurred before the planned outcome assessment.

Figure 4.1 illustrates these ideas using five representative patient trajectories in the motivating long-term chronic weight management trial. After treatment assignment, patient 1

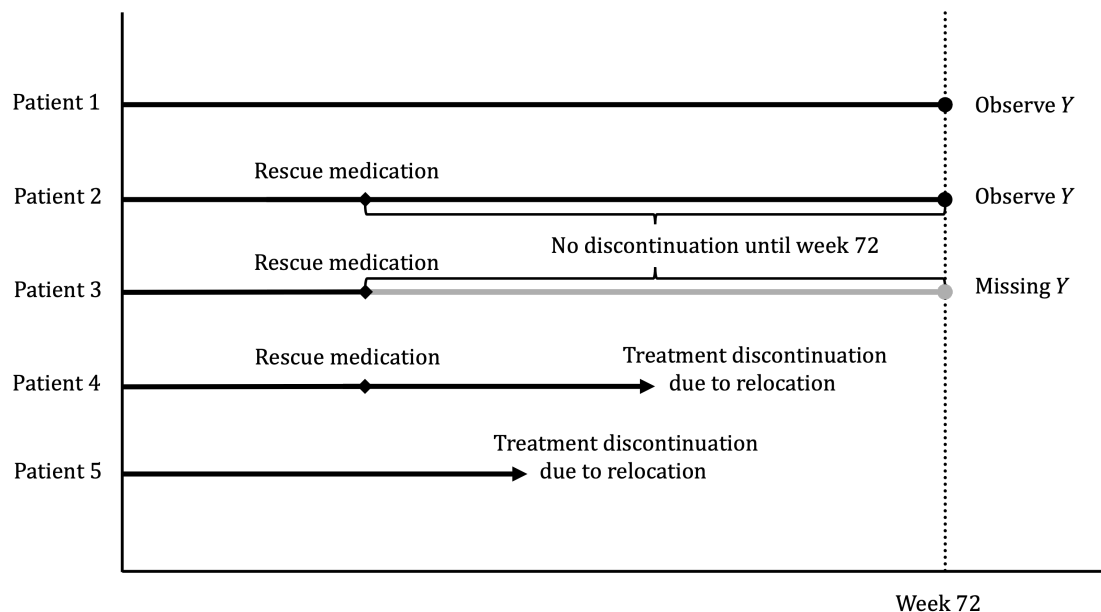


Figure 4.1: Illustration of five representative scenarios of intercurrent events in the long-term chronic weight management trial. Rescue medication use is an example of effect-informative IE and relocation-related treatment discontinuation is an example of effect-uninformative IE.

does not experience any IE before the outcome assessment at week 72, so the outcome is observed. Patients 2–5 experience different types of IEs at different times after treatment initiation, leading to different observed-data structures. Patients 2–4 experience insufficient response or tolerability concerns and therefore use rescue medication around week 24. Because rescue medication use is part of the post-treatment clinical trajectory and may be informative about treatment response or tolerability, it is treated as an effect-informative IE and handled using a treatment policy strategy. After using rescue medication, patient 2 remains in the trial through week 72 without any additional IE, so the outcome is observed. Patient 3 also remains free of any additional IE, but the week-72 outcome is not measured, and thus, the hypothetical outcome is missing. Patient 4 later relocates around week 48 for reasons unrelated to treatment efficacy, leading to treatment discontinuation and an unobserved week-72 outcome. Patient 5 also experiences relocation-related treatment discontinuation around the middle of follow-up. Because this effect-uninformative IE occurs before any subsequent effect-informative IE can be observed, both the later effect-informative IE process and the hypothetical outcome are unobserved. Because relocation is external to the scientific question, it is treated as an effect-uninformative IE and handled using a hypothetical strategy.

These trajectories illustrate the observed-data structure in our setup. For patients who do not experience any IE before outcome measurement, the hypothetical outcome of interest

is observed. For patients whose observed first IE is effect-informative, such as patients 2–4, the effect-informative IE is retained as part of the treatment-policy strategy. Their hypothetical outcome of interest is observed only if they do not further experience an effect-uninformative IE and their outcome is not missing. In contrast, for patients whose first IE is effect-uninformative, such as patient 5, the subsequent effect-informative IE process and the hypothetical outcome of interest are not observed.

This example motivates the combined-strategy estimand studied in this paper. We use a treatment policy strategy for effect-informative IEs and a hypothetical strategy for effect-uninformative IEs. The resulting estimand answers a clinically interpretable question: what is the treatment effect under real-world treatment evolution, after removing external IEs that are not part of the scientific question?

4.1.2 Related literature and contribution

The estimand framework has generated a large literature on how clinical trial objectives should be translated into precise treatment effects. The ICH E9 (R1) addendum introduced five broad strategies for addressing IEs, and subsequent papers have provided practical guidance for implementing this framework in trial protocols, statistical analysis plans, and regulatory discussions (Ratitch et al., 2020; Kahan et al., 2024; Polverejan et al., 2023). A central message of this literature is that the strategy should be driven by the clinical question, and that different IEs in the same trial may require different strategies.

Several papers study specific strategies in more detail. The treatment policy strategy has been advocated when post-treatment events are part of the treatment regimen or clinical practice pathway, because it preserves a pragmatic interpretation of the treatment effect (Fleming et al., 2025). The hypothetical strategy has been studied from causal inference, missing-data, and g-computation perspectives, with emphasis on defining the counterfactual scenario in which a particular IE would not occur (Olarte Parra et al., 2023). Other work discusses composite and principal-stratum strategies, especially when IEs are clinically meaningful outcomes or define relevant subpopulations (Ratitch et al., 2020). These papers provide important conceptual guidance, but most focus on one strategy at a time or on implementation rather than on nonparametric identification and efficiency theory for combined-strategy estimands. There is less work on estimands that combine different strategies for different IEs according to their types. Qu et al. (2021) propose an estimand using a mixture of strategies and illustrate its use in clinical trials. Recent work has also studied causal identification with multiple IEs (Olarte Parra et al., 2025a; Olarte Parra et al., 2025b). These contributions show the need for combined strategies, but general semiparametric theory is still limited, especially when one type of IE is handled by a treatment policy strategy and another by a hypothetical strategy.

This paper develops a causal and semiparametric framework for such a combined-strategy

estimand. The main statistical difficulty is that the two IE times have a semi-competing structure. If the effect-uninformative IE occurs first, then the effect-informative IE time and the outcome are not observed. If the effect-informative IE occurs first, later information on the effect-uninformative IE is available, and the outcome may still be observed. Thus, the event handled by the hypothetical strategy acts as a “censoring event” before the effect-informative IE is observed, but becomes part of the post-event observation process afterward. This differs from standard missing-data, right-censoring, and competing-risk problems.

We make the following contributions. We define a combined-strategy causal estimand using the potential outcomes framework. The estimand keeps the natural value of the effect-informative IE, corresponding to a treatment policy strategy, while intervening to prevent the effect-uninformative IE before the planned outcome assessment, corresponding to a hypothetical strategy. We then give transparent identification assumptions and provide two nonparametric identification formulas, one based on outcome modeling and one based on inverse-probability weighting. We also develop several estimators, including outcome-regression, weighting, augmented weighting, and efficient-influence-function(EIF)-based estimators. These estimators rely on different nuisance components and therefore offer flexible tools for implementation.

The rest of the paper proceeds as follows. Section 4.2 introduces the formal setup, states the identification assumptions, and derives nonparametric identification formulas. Section 4.3 presents basic plug-in, weighting, and augmented estimators. Section 4.4 develops the semi-parametric efficiency theory and the EIF-based estimator. Section 4.5 reports simulation study results. Section 4.6 concludes with practical implications and directions for future work.

4.2 Setup and nonparametric identification

4.2.1 Setup and estimand

Consider classifying IEs into two categories: effect-informative and effect-uninformative. We use a treatment policy strategy for the effect-informative IE and a hypothetical strategy for the effect-uninformative IE (ICH E9 (R1), 2019).

We introduce our formal setup and additional mathematical notation. Consider a super-population setting with the Neyman–Rubin potential outcomes framework. Let A denote the binary treatment variable, where $A = 1$ for treated units and $A = 0$ otherwise. Let T denote the time to an effect-informative IE and C denote the time to an effect-uninformative IE. Since both are post-treatment variables, we use $T(a)$ and $C(a)$ to denote their potential values under treatment arm a , respectively, for $a = 0, 1$. Let X denote the vector of baseline covariates.

The outcome of interest, Y , is measured at a fixed time k . Therefore, only IEs that occur no later than time k are of interest. We set $T = \infty$ or $C = \infty$ if the corresponding event does not occur before the outcome assessment. Thus, the support of T and C is $\mathcal{I} = [0, k] \cup \{\infty\}$, and we use $C > k$ and $C = \infty$ interchangeably.

For each treatment arm $a = 0, 1$, let $Y(a, T(a), c = \infty)$ denote the potential outcome under treatment a in the hypothetical world where the effect-uninformative IE never occurred before time k . This hypothetical intervention is meaningful as it is possible to intervene on C by, for example, preventing the administrative site closure or reducing barriers that would otherwise lead to treatment discontinuation after relocation. Therefore, setting the effect-uninformative IE time to infinity represents a scientifically interpretable intervention that removes events external to the treatment effect of interest. By contrast, the effect-informative IE is handled using a treatment policy strategy. Therefore, its natural value $T(a)$ is left unchanged in the estimand. Under the composition assumption, we have $Y(a, c = \infty) = Y(a, T(a), c = \infty)$. Our causal estimand of interest is the average treatment effect

$$\tau = E\{Y(1, c = \infty) - Y(0, c = \infty)\}. \quad (4.1)$$

This estimand preserves the real-world relevance of the effect-informative IE while removing the effect-uninformative IE from the target outcome. It answers the question of how effective the treatment is under real-world treatment evolution, after setting aside external events that are not part of the scientific question. Therefore, it gives a clinically meaningful estimand.

Compared with estimands that handle all IEs using a single strategy or with the estimand combining composite outcome and hypothetical strategies (Lu et al., 2025), our estimand in (4.1) requires more detailed follow-up information. In particular, even for patients who experience the effect-informative IE, we still need to observe whether the effect-uninformative IE occurs before time k and whether the final outcome is measured. This is needed to learn about $Y(a, c = \infty)$ after the effect-informative IE has occurred. In many long-term chronic disease trials, such follow-up is still collected after rescue medication, dose adjustment, or treatment modification.

The key feature of the setup is that the two post-treatment event times are *semi-competing*. If C occurs first, then the values of both T and Y in the hypothetical world we are interested in are censored. If T occurs first, then T is observed, and the subsequent value of C through time k is also observed. The final outcome Y may still be observed or missing after the occurrence of T . We write $R = 1$ if Y is observed after T and $R = 0$ otherwise. This semi-competing structure requires identification and estimation arguments beyond the standard settings.

Figure 4.1 also illustrates the observed-data structure induced by the semi-competing event process. If no IE occurs before k , $Y(A, c = \infty)$ is always observed. If an effect-informative IE

occurs first, the event is retained under the treatment-policy strategy, and the hypothetical outcome $Y(A, c = \infty)$ is observed only when (1) no subsequent effect-uninformative IE occurs before k , i.e., $C > k$, and (2) the outcome k is not missing, i.e., $R = 1$. If an effect-uninformative IE occurs first, subsequent information on the effect-informative IE process and the hypothetical outcome is no longer observed.

4.2.2 Identification assumptions

We now state the assumptions used to identify the arm-specific means $\mu_a = E\{Y(a, c = \infty)\}$ for $a = 0, 1$, and therefore for the average treatment effect $\tau = \mu_1 - \mu_0$.

Assumption 4.1 (Treatment ignorability). *We assume the following conditional independence and positivity conditions:*

(a) For each $a = 0, 1$,

$$A \perp\!\!\!\perp \{Y(a, c = \infty), T(a), C(a)\} \mid X.$$

(b) There exists $\eta_A > 0$ such that $\text{pr}(A = 1 \mid X = x) \in [\eta_A, 1 - \eta_A]$ for all $x \in \mathcal{X}$.

Assumption 4.1 is guaranteed by design in a randomized controlled trial. In observational studies, it is a standard but untestable condition.

Assumption 4.2 (Effect-uninformative IE). *We assume the following conditional independence and positivity conditions for the effect-uninformative IE time C :*

(a) For each $a = 0, 1$,

$$C \perp\!\!\!\perp \{T, Y(a, c = \infty)\} \mid A = a, X.$$

(b) There exists $\eta_C > 0$ such that $\text{pr}(C > k \mid A = a, X = x) > \eta_C$ for all $a = 0, 1$ and $x \in \mathcal{X}$.

Assumption 4.2 is the key condition that justifies treating an IE C as effect-uninformative. After conditioning on the treatment arm and baseline covariates, the timing of effect-uninformative IE carries no information about either the time to the effect-informative IE or the hypothetical outcome of interest. It thus carries no information about the treatment efficacy.

Assumption 4.3 (Post- T missingness of Y). *We assume the following conditional independence and positivity conditions for outcome observation:*

(a) For each $a = 0, 1$ and each $j \in \{0, \dots, k\}$,

$$R \perp\!\!\!\perp Y \mid C > k, T = j, A = a, X.$$

(b) There exists $\eta_R > 0$ such that $\text{pr}\{R = 1 \mid C > k, T = j, A = a, X = x\} > \eta_R$ for all $a = 0, 1, j \in \mathcal{I}$, and $x \in \mathcal{X}$.

Assumption 4.3 assumes that, among units not experiencing effect-uninformative IEs through time k , outcome observation is independent of the outcome itself after conditioning on the effect-informative IE time T , treatment arm A , and baseline covariates X . This is a missing-at-random condition for the post- T observation of Y . The assumption also highlights the role of T in our setup. Although T does not directly enter the estimand τ from the formula, it is part of the information needed to explain outcome missingness. Therefore, even when the target parameter is the average treatment effect on the hypothetical potential outcomes, T cannot be ignored because missingness of Y is assumed to be at random only after conditioning on the post-treatment information contained in T , and thus valid identification and estimation require proper adjustment for T .

4.2.3 Observed-data nuisance functions

We introduce the nuisance functions used in the identification formulas. Let $e_a(X) = \text{pr}(A = a \mid X)$ denote the probability of getting treatment a , so $e_1(X)$ denotes the propensity score. For each treatment arm $a = 0, 1$ and $j = 0, \dots, k$, let

$$f_a(j \mid X) = \text{pr}(T = j \mid A = a, X), \quad S_a(j \mid X) = \text{pr}(T > j \mid A = a, X)$$

denote the conditional mass function and survival function of the effect-informative IE time T . Similarly, let $G_a(j \mid X) = \text{pr}(C > j \mid A = a, X)$ denote the conditional mass function and survival function of the effect-uninformative IE time C .

For units with observed effect-informative IE time j and no effect-uninformative IE before time k , define the outcome-observation probability $r_{a,j}(X) = \text{pr}(R = 1 \mid C > k, T = j, A = a, X)$. Finally, define the subgroup outcome regression functions

$$\mu_{a,j}(X) = E(Y \mid R = 1, C > k, T = j, A = a, X), \quad j = 0, \dots, k,$$

and

$$\mu_{a,\infty}(X) = E(Y \mid T \wedge C > k, A = a, X).$$

4.2.4 Nonparametric identification

Under the assumptions above, the arm-specific means μ_a are nonparametrically identified. We give two identification formulas, where the first is an outcome-regression formula, and the second is a weighting formula.

Theorem 4.1 (Nonparametric identification). *Under Assumptions 4.1–4.3, we have the following two nonparametric identification formulas for the arm-specific means μ_1 and μ_0 .*

First, we have the outcome-regression identification formula

$$\mu_a = E \left\{ \mu_{a,\infty}(X) S_a(k | X) + \sum_{j=0}^k \mu_{a,j}(X) f_a(j | X) \right\}. \quad (4.2)$$

Second, we have the weighting identification formula

$$\mu_a = E \left[\frac{1(A = a)}{e_a(X)} \left\{ \frac{1(T \wedge C > k)Y}{G_a(k | X)} + \sum_{j=0}^k \frac{1(T = j, C > k, R = 1)Y}{G_a(k | X) r_{a,j}(X)} \right\} \right]. \quad (4.3)$$

Equation (4.2) separates the estimand into two parts. The functions $S_a(k | X)$ and $f_a(j | X)$ describe the distribution of the effect-informative IE time under treatment arm a . The functions $\mu_{a,\infty}(X)$ and $\mu_{a,j}(X)$ describe the corresponding conditional outcome means in the hypothetical world where C does not occur before time k . Equation (4.3) instead identifies μ_a by reweighting observed outcomes. The treatment weights account for treatment assignment. The factor $G_a(k | X)$ accounts for probability free of the effect-uninformative IE through k . The factors $r_{a,j}(X)$ account for missingness of the final outcome after the effect-informative IE has already been observed. Thus, the formula reflects the semi-competing structure: before T is observed, C acts as a censoring event; after T is observed, subsequent information on C and Y is part of the post-event observation process.

4.3 Basic and augmented estimators

4.3.1 Two basic estimators

The two formulas in Theorem 4.1 suggest two basic estimators. Let $\hat{\mu}_{a,\infty}(X)$, $\hat{\mu}_{a,j}(X)$, $\hat{S}_a(j | X)$, $\hat{f}_a(j | X)$, $\hat{G}_a(k | X)$, $\hat{r}_{a,j}(X)$, and $\hat{e}_a(X)$ denote fitted nuisance functions. The plug-in outcome estimator is

$$\hat{\mu}_a^{\text{out}} = n^{-1} \sum_{i=1}^n \left\{ \hat{\mu}_{a,\infty}(X_i) \hat{S}_a(k | X_i) + \sum_{j=0}^k \hat{\mu}_{a,j}(X_i) \hat{f}_a(j | X_i) \right\},$$

and the weighting estimator is

$$\hat{\mu}_a^{\text{ipw}} = n^{-1} \sum_{i=1}^n \frac{1(A_i = a)}{\hat{e}_a(X_i)} \left\{ \frac{1(T_i \wedge C_i > k)Y_i}{\hat{G}_a(k | X_i)} + \sum_{j=0}^k \frac{1(T_i = j, C_i > k, R_i = 1)Y_i}{\hat{G}_a(k | X_i) \hat{r}_{a,j}(X_i)} \right\}.$$

The treatment effect estimators are $\hat{\tau}^{\text{out}} = \hat{\mu}_1^{\text{out}} - \hat{\mu}_0^{\text{out}}$ and $\hat{\tau}^{\text{ipw}} = \hat{\mu}_1^{\text{ipw}} - \hat{\mu}_0^{\text{ipw}}$.

The estimator $\hat{\tau}^{\text{out}}$ is consistent if the outcome regressions $\{\mu_{a,\infty}, \mu_{a,j}\}$ and the T distribution $\{S_a, f_a\}$ are consistently estimated. The estimator $\hat{\tau}^{\text{ipw}}$ is consistent if the weighting components $\{e, G_a, r_{a,j}\}$ are consistently estimated.

4.3.2 An augmented estimator

Combining the two identification formulas yields

$$\begin{aligned} \mu_a = E \left[\frac{1(A=a)}{e_a(X)} \left\{ \frac{1(T \wedge C > k)Y}{G_a(k | X)} + \sum_{j=0}^k \frac{1(T=j, C > k, R=1)Y}{G_a(k | X)r_{a,j}(X)} \right\} \right. \\ \left. - \frac{1(A=a) - e_a(X)}{e_a(X)} \left\{ \mu_{a,\infty}(X)S_a(k | X) + \sum_{j=0}^k \mu_{a,j}(X)f_a(j | X) \right\} \right]. \quad (4.4) \end{aligned}$$

This suggests the augmented estimators

$$\hat{\mu}_a^{\text{aug}} = \hat{\mu}_a^{\text{ipw}} - n^{-1} \sum_{i=1}^n \frac{1(A_i=a) - \hat{e}_a(X_i)}{\hat{e}_a(X_i)} \left\{ \hat{\mu}_{a,\infty}(X_i)\hat{S}_a(k | X_i) + \sum_{j=0}^k \hat{\mu}_{a,j}(X_i)\hat{f}_a(j | X_i) \right\},$$

and $\hat{\tau}^{\text{aug}} = \hat{\mu}_1^{\text{aug}} - \hat{\mu}_0^{\text{aug}}$.

Proposition 4.1 (Double robustness of the augmented estimator). *Suppose Assumptions 4.1–4.3 hold. If $G_a(k | X)$ and $r_{a,j}(X)$ are correctly specified, then $\hat{\tau}^{\text{aug}}$ is consistent for τ if either the propensity score $e_a(X)$ is correctly specified, or $\{\mu_{a,\infty}(X), \mu_{a,j}(X), S_a(k | X), f_a(j | X)\}$ are correctly specified for all $j = 0, \dots, k$.*

Analogously to the standard augmented inverse-probability weighted estimator, the correction in (4.4) orthogonalizes the estimating equation with respect to the treatment assignment mechanism. However, it does not orthogonalize with respect to the full observed-data nuisance generated by the semi-competing event process. In the present setting that nuisance has two components: the post-event observation distribution of (C, R, Y) and the pre-event censoring of the latent event time T by C . Therefore, the estimator is not semiparametrically efficient.

4.4 Semiparametric efficiency theory

In this section, we derive an efficient influence function (EIF) for μ_1 and μ_0 , and thus for τ . In our setup, the observed-data nuisance generated by the semi-competing event process

splits into two pieces: the post-event distribution of (C, R, Y) after the effect-informative IE time, T , has already been observed, and the pre-event censoring of T by the effect-uninformative IE C . The EIF is obtained by first solving the post-event canonical gradient and then projecting the resulting score onto the orthocomplement of the pre-event censoring tangent space.

4.4.1 EIF representation

Define the reduced pre-event observed-data variables

$$\tilde{T} = T \wedge C \wedge k, \quad \delta = \begin{cases} 1, & T \leq C, T \leq k, \\ 2, & C < T, C \leq k, \\ 3, & T > k, C > k, \end{cases}$$

and let

$$\Delta = 1(\delta \in \{1, 3\}) = 1(C > T \wedge k).$$

Here, $\tilde{T}$ is the observed first event time, $\delta = 1$ indicates that the effect-informative IE time T is observed before C and happened before the outcome measurement time k , $\delta = 2$ indicates that an effect-uninformative IE occurs first and censors T and Y , and $\delta = 3$ indicates that no IE occurs through k . When $\delta = 1$ and $\tilde{T} = j$, the post-event observed-data block is $W_j = (C, R, RY)$, and when $\delta = 3$, the post-event observed-data is simply the outcome Y .

Before formally introducing the EIF form, we introduce some additional notation. For each arm $a = 0, 1$ and each $j = 0, \dots, k$, define

$$\omega_{a,j}(X) = \text{pr}(C > k \mid C > j, T = j, A = a, X) = \frac{G_a(k \mid X)}{G_a(j \mid X)}, \quad (4.5)$$

where the second equality follows from Assumption 4.2. The quantity $\omega_{a,j}(X)$ is the conditional probability that C not occurring before k once $T = j$ has already been observed. Next, define the post-event pseudo-outcome

$$\tilde{U}_a(O) = 1(\delta = 3)Y + \sum_{j=0}^k 1(\delta = 1, \tilde{T} = j) \left[\mu_{a,j}(X) + \frac{1(C > k, R = 1)}{\omega_{a,j}(X)r_{a,j}(X)} \{Y - \mu_{a,j}(X)\} \right] \quad (4.6)$$

Also define the full-data regression

$$m_a(T, X) = \mu_{a,\infty}(X)1(T > k) + \sum_{j=0}^k \mu_{a,j}(X)1(T = j)$$

and the continuation value $V_a(t, X) = E\{m_a(T, X) \mid T \geq t, A = a, X\}$. For discrete event times,

$$V_a(t, X) = \frac{\mu_{a,\infty}(X)S_a(k \mid X) + \sum_{j=t}^k \mu_{a,j}(X)f_a(j \mid X)}{S_a(t \mid X)}.$$

Then define $N_{C,a}(t) = 1(\tilde{T} \leq t, \delta = 2)$ as the stopped counting process for pre-event C in arm a , and let $\Lambda_a(t | X)$ denote the corresponding cumulative hazard of C before T is observed. Then the martingale increment for the stopped pre-event censoring process is

$$dM_{G_a}(t) = dN_{C,a}(t) - 1(\tilde{T} \geq t)d\Lambda_a(t | X)$$

Theorem 4.2 (EIF for μ_a and τ). *Under Assumptions 4.1–4.3, an efficient influence function for μ_a is*

$$\begin{aligned} D_a(O) = & \frac{1(A=a)}{e_a(X)} \left\{ \frac{\Delta \tilde{U}_a(O)}{G_a(\tilde{T} | X)} + \int_0^{\tilde{T}} \frac{V_a(t, X)}{G_a(t | X)} dM_{G_a}(t) \right\} \\ & - \frac{1(A=a) - e_a(X)}{e_a(X)} \left\{ \mu_{a,\infty}(X) S_a(k | X) + \sum_{j=0}^k \mu_{a,j}(X) f_a(j | X) \right\} - \mu_a \end{aligned} \quad (4.7)$$

Consequently, an efficient influence function for $\tau = \mu_1 - \mu_0$ is $D_\tau(O) = D_1(O) - D_0(O)$.

The EIF forms in Theorem 4.2 present a nested structure. The conditional probability $\omega_{a,j}(X) = G_a(k | X)/G_a(j | X)$ enters because in each subgroup where we observe $T = j$, we need to account for the possibility that C occurs between j and k . The conditional probability $r_{a,j}(X)$ accounts for the fact that outcome Y may still be missing even when $C > k$. After these post-event corrections are absorbed into $\tilde{U}_a(O)$, the remaining nuisance is the ordinary right-censoring of the effect-informative IE time T by the early part of the effective-uninformative IE time C process.

4.4.2 EIF-based estimator

Theorem 4.2 motivates the following estimators. Define

$$\begin{aligned} \hat{\mu}_a^{\text{eif}} = & n^{-1} \sum_{i=1}^n \frac{1(A_i=a)}{\hat{e}_a(X_i)} \left\{ \frac{\hat{\Delta}_i \hat{\tilde{U}}_{a,i}}{\hat{G}_a(\tilde{T}_i | X_i)} + \int_0^{\tilde{T}_i} \frac{\hat{V}_a(t, X_i)}{\hat{G}_a(t | X_i)} dM_{\hat{G}_a}(t) \right\} \\ & - n^{-1} \sum_{i=1}^n \frac{1(A_i=a) - \hat{e}_a(X_i)}{\hat{e}_a(X_i)} \hat{\mu}_a(X_i), \end{aligned}$$

where

$$\begin{aligned} \hat{\omega}_{a,j}(X_i) &= \frac{\hat{G}_a(k | X_i)}{\hat{G}_a(j | X_i)}, \\ \hat{\tilde{U}}_{a,i} &= 1(\delta_i = 3)Y_i + \sum_{j=0}^k 1(\delta_i = 1, \tilde{T}_i = j) \left[\hat{\mu}_{a,j}(X_i) + \frac{1(C_i > k, R_i = 1)}{\hat{\omega}_{a,j}(X_i) \hat{r}_{a,j}(X_i)} \{Y_i - \hat{\mu}_{a,j}(X_i)\} \right], \end{aligned}$$

$$\begin{aligned}\hat{\Delta}_i &= 1(\delta_i \in \{1, 3\}), \\ \hat{\mu}_a(X_i) &= \hat{\mu}_{a,\infty}(X_i)\hat{S}_a(k | X_i) + \sum_{j=0}^k \hat{\mu}_{a,j}(X_i)\hat{f}_a(j | X_i), \\ \hat{V}_a(t, X_i) &= \frac{\hat{\mu}_{a,\infty}(X_i)\hat{S}_a(k | X_i) + \sum_{j=t}^k \hat{\mu}_{a,j}(X_i)\hat{f}_a(j | X_i)}{\hat{S}_a(t | X_i)}.\end{aligned}$$

The corresponding treatment effect estimator is

$$\hat{\tau}^{\text{eif}} = \hat{\mu}_1^{\text{eif}} - \hat{\mu}_0^{\text{eif}}.$$

Theorem 4.3 (Double robustness of the EIF-based estimator). *Suppose Assumptions 4.1–4.3 hold. Then $\hat{\tau}^{\text{eif}}$ is consistent for τ if either of the following two sets of nuisance functions are consistently estimated: (i) $\{e_a(X), G_a(t | X), r_{a,j}(X)\}$ for $t, j \leq k$ or (ii) $\{\mu_{a,\infty}(X), \mu_{a,j}(X), f_a(j | X), S_a(k | X)\}$ for $j \leq k$.*

Under standard empirical process or rate conditions for the nuisance estimators, $\hat{\tau}^{\text{eif}}$ is asymptotically linear with influence function $D_\tau(O)$ and therefore achieves the semiparametric efficiency bound $E\{D_\tau^2(O)\}$.

Compared with the previous augmented weighting estimator, the EIF-based estimator adds an additional semiparametric ingredient: the martingale augmentation for pre-event censoring of the latent time T . In our semi-competing structure, this martingale correction is not directly applied to Y . We instead first build the post-event orthogonalized pseudo-outcome $\tilde{U}_a(O)$, and then project off the nuisance tangent space by the early part of C .

4.5 Simulation

In this section, we study the finite sample performance of the four proposed estimators: the outcome regression estimator $\hat{\tau}^{\text{out}}$, the weighting estimator $\hat{\tau}^{\text{ipw}}$, the augmented estimator $\hat{\tau}^{\text{aug}}$, and the EIF-based estimator $\hat{\tau}^{\text{eif}}$.

4.5.1 Data generating processes

We set the outcome measurement time $k = 4$ and generate baseline covariates $X = (X_1, X_2, X_3)^\top$ from independent standard Gaussian distributions. Treatment A is generated from a covariate-dependent propensity score. Conditional on treatment and covariates, we generate the latent effect-informative IE time T , the effect-uninformative IE time C , the hypothetical outcome $Y(a, c = \infty)$, and the post- T outcome-observation indicator R . The true nuisance functions are nonlinear functions of X and include interac-

tions and threshold effects. This creates a setting in which simple working models can be meaningfully misspecified.

To make nuisance functions misspecified in some data generating regimes, we make the true nuisance functions depend on nonlinear terms, interactions, and step functions of X : $\{\sin(1.4X_1), 1(X_2 > 0), 1(X_3 > 0.35), X_1X_2, X_1X_3, X_2X_3, X_2^2, X_3^2\}$. We generate the observed data following the semi-competing structure. If C occurs before T and before the outcome assessment time, then both T and Y are not observed. If T occurs first, then T is observed, and subsequent information on C is also observed. When $T = j \leq k$ and $C > k$, the final outcome may still be missing through R . If neither T nor C occurs before time k , the final outcome is observed.

We consider five nuisance-model regimes. In correctly specified components, we fit models using the true parametric forms. In misspecified components, we fit intentionally inadequate working models containing only the linear main effects (X_1, X_2, X_3) . Specifically, the five regimes are: all nuisance components correctly specified; only e misspecified; (e, G, r) misspecified and (μ, f) correctly specified; (μ, f) misspecified and (e, G, r) correctly specified; and all nuisance components misspecified.

These regimes are chosen to illustrate the robustness patterns of the four estimators. We expect the outcome-regression estimator to be consistent when (μ, f) are correctly specified, the weighting estimator to be consistent when (e, G, r) are correctly specified, the augmented estimator to be consistent when G and r are correctly specified and either e or (μ, f) is correctly specified, and the EIF-based estimator to be consistent when either the weighting nuisances (e, G, r) or the outcome nuisances (μ, f) is correctly specified. We use 500 Monte Carlo replications with sample size $n = 2000$. The true value of τ is computed by Monte Carlo integration using a large independently generated sample and is approximately 1.063. We report bias and the standard deviation of all estimators.

4.5.2 Simulation results

The simulation results are consistent with the theoretical robustness properties. When all nuisance functions are correctly specified, all four estimators are consistent. The weighting and augmented have larger standard deviations because of the weights, and the EIF estimator reduces it through the additional augmentation terms. When only the propensity score e is misspecified, the weighting estimator has non-negligible bias. When (e, G, r) are misspecified while (μ, f) are correct, the outcome regression estimator remains nearly unbiased. The weighting and augmented estimators perform poorly because their required weighting components are wrong. The EIF estimator is more robust than weighting and augmented weighting in this regime, although it retains a small finite-sample bias likely due to extreme weights. When (μ, f) are misspecified but (e, G, r) are correct, the outcome regression estimator has non-negligible bias. Finally, when all nuisance components are misspecified, all

Table 4.1: Finite-sample performance of the four estimators across nuisance-model regimes. For each estimator, we report Monte Carlo bias and standard deviation (SD) based on 500 Monte Carlo samples with $n = 2000$.

Regime	$\hat{\gamma}^{\text{out}}$		$\hat{\gamma}^{\text{ipw}}$		$\hat{\gamma}^{\text{aug}}$		$\hat{\gamma}^{\text{EIF}}$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
all correct	0.038	0.052	-0.015	0.426	-0.008	0.418	-0.011	0.311
e wrong	0.035	0.051	-0.449	0.300	-0.010	0.302	-0.010	0.235
(e, G, r) wrong	0.034	0.056	-0.264	0.193	0.172	0.191	0.065	0.156
(μ, f) wrong	-0.171	0.148	0.026	0.502	0.023	0.500	0.002	0.343
all wrong	-0.169	0.141	-0.260	0.198	-0.257	0.198	-0.267	0.174

estimators have non-negligible bias.

4.6 Discussion

We study a combined-strategy estimand for randomized trials with effect-informative and effect-uninformative IEs. The main idea is that different IEs play different scientific roles: some are part of the treatment pathway and should be preserved in the target estimand, while others are external to the clinical question and are more naturally handled through a hypothetical strategy. This event-specific view is consistent with the distinction between estimands and identification assumptions: the estimand should encode the clinical question, while assumptions determine whether it can be identified from observed data (Ding and Lu, 2025). We formalize this idea by combining a treatment policy strategy for effect-informative IEs with a hypothetical strategy for effect-uninformative IEs, and provide identification assumptions, two nonparametric identification formulas, and several estimators. The EIF-based estimator further shows the structure of the problem: post-event missingness is handled through an orthogonalized pseudo-outcome, while pre-event censoring by the effect-uninformative IE is handled through a martingale augmentation. The main limitation is that the hypothetical strategy relies on an untestable independence condition for the effect-uninformative IE, which should be justified by subject-matter knowledge and assessed through sensitivity analysis. Future work could extend the framework to multiple IE types, continuous event times, recurrent IEs, and longitudinal outcomes.

Bibliography

- Ai, C. et al. (2021). “A unified framework for efficient estimation of general treatment models”. *Quantitative Economics* 12.3, pp. 779–816.
- Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). “Identification of causal effects using instrumental variables (with discussion)”. *Journal of the American Statistical Association* 91.434, pp. 444–455.
- Bang, H. and J. M. Robins (2005). “Doubly robust estimation in missing data and causal inference models”. *Biometrics* 61.4, pp. 962–973.
- Bartolucci, F. and L. Grilli (2011). “Modeling partial compliance through copulas in a principal stratification framework”. *Journal of the American Statistical Association* 106.494, pp. 469–479.
- Ben-Michael, E. et al. (2021). “The balancing act in causal inference”. *arXiv preprint arXiv:2110.14831*.
- Bickel, P. J. et al. (1993). *Efficient and adaptive estimation for semiparametric models*. Vol. 4. Springer.
- Boileau, P. et al. (2025). “A nonparametric framework for treatment effect modifier discovery in high dimensions”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 87.1, pp. 157–185.
- Chattopadhyay, A., C. H. Hase, and J. R. Zubizarreta (2020). “Balancing vs modeling approaches to weighting in practice”. *Statistics in Medicine* 39.24, pp. 3227–3254.
- Cheng, C. et al. (2024). “Multiply robust estimation for causal survival analysis with treatment noncompliance”. *Clinical Trials*.
- Cheng, J. and D. S. Small (2006). “Bounds on causal effects in three-arm trials with non-compliance”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.5, pp. 815–836.
- Chernozhukov, V. et al. (2018). “Double/debiased machine learning for treatment and structural parameters”. *The Econometrics Journal* 21.1, pp. C1–C68.
- Comment, L. et al. (2019). “Survivor average causal effects for continuous time: a principal stratification approach to causal inference with semicompeting risks”. *arXiv preprint arXiv:1902.09304*.

- Conlon, A. S., J. M. Taylor, and M. R. Elliott (2014). “Surrogacy assessment using principal stratification when surrogate and outcome measures are multivariate normal”. *Biostatistics* 15.2, pp. 266–283.
- Cox, D. R. (1972). “Regression models and life-tables”. *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2, pp. 187–202.
- Daniels, M. J. et al. (2012). “Bayesian inference for the causal effect of mediation”. *Biometrics* 68.4, pp. 1028–1036.
- Ding, P. and J. Lu (2017). “Principal stratification analysis using principal scores”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.3, pp. 757–777.
- Ding, P. and S. Lu (2025). “The Roles of Estimands and Assumptions in Causal Inference: Comment on “Chasing Shadows: How Implausible Assumptions Skew Our Understanding of Causal Estimands””. *Statistics in Biopharmaceutical Research* 17.4, pp. 523–529.
- Ding, P. et al. (2011). “Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death”. *Journal of the American Statistical Association* 106, pp. 1578–1591.
- Ebrahimi, N., D. Molefe, and Z. Ying (2003). “Identifiability and censored data”. *Biometrika* 90.3, pp. 724–727.
- Efron, B. and D. Feldman (1991). “Compliance as an explanatory variable in clinical trials”. *Journal of the American Statistical Association* 86.413, pp. 9–17.
- Egleston, B. L., D. O. Scharfstein, and E. MacKenzie (2009). “On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death”. *Biometrics* 65, pp. 497–504.
- Elliott, M. R., T. E. Raghunathan, and Y. Li (2010). “Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes”. *Biostatistics* 11, pp. 353–372.
- Fan, J. and Q. Yao (2003). *Nonlinear time series: nonparametric and parametric methods*. Vol. 20. Springer.
- Feller, A., F. Mealli, and L. Miratrix (2017). “Principal score methods: Assumptions, extensions, and practical considerations”. *Journal of Educational and Behavioral Statistics* 42, pp. 726–758.
- Fleming, T. R. et al. (2025). “A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events”. *Statistics in Medicine*. DOI: 10.1002/sim.70104.
- Follmann, D. A. (2000). “On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial”. *Journal of the American Statistical Association* 95.452, pp. 1101–1109.
- Forastiere, L., A. Mattei, and P. Ding (2018). “Principal ignorability in mediation analysis: through and beyond sequential ignorability”. *Biometrika* 105.4, pp. 979–986.
- Forastiere, L., F. Mealli, and T. J. VanderWeele (2016). “Identification and estimation of causal mechanisms in clustered encouragement designs: disentangling bed nets using Bayesian principal stratification”. *Journal of the American Statistical Association* 111.514, pp. 510–525.

- Frangakis, C. E. and D. B. Rubin (1999). “Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes”. *Biometrika* 86, pp. 365–379.
- Frangakis, C. E. and D. B. Rubin (2002). “Principal stratification in causal inference”. *Biometrics* 58.1, pp. 21–29.
- Frumento, P. et al. (2012). “Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data”. *Journal of the American Statistical Association* 107, pp. 450–466.
- Gallop, R. et al. (2009). “Mediation analysis with principal stratification”. *Statistics in Medicine* 28, pp. 1108–1130.
- Gilbert, P. B. and M. G. Hudgens (2008). “Evaluating candidate principal surrogate endpoints”. *Biometrics* 64, pp. 1146–1154.
- Gilbert, P. B. et al. (2015). “Surrogate endpoint evaluation: Principal stratification criteria and the prentice definition”. *Journal of Causal Inference* 3.2, pp. 157–175.
- Gilbert, P. B. et al. (2024). “Four statistical frameworks for assessing an immune correlate of protection (surrogate endpoint) from a randomized, controlled, vaccine efficacy trial”. *Vaccine*.
- Gill, R. D. and S. Johansen (1990). “A survey of product-integration with a view toward application in survival analysis”. *The Annals of Statistics* 18.4, pp. 1501–1555.
- Hahn, J. (1998). “On the role of the propensity score in efficient semiparametric estimation of average treatment effects”. *Econometrica* 66.2, p. 315.
- Hall, P., R. C. Wolff, and Q. Yao (1999). “Methods for estimating a conditional distribution function”. *Journal of the American Statistical Association* 94.445, pp. 154–163.
- Han, S. and X.-H. Zhou (2023). “Defining estimands in clinical trials: a unified procedure”. *Statistics in Medicine* 42.12, pp. 1869–1887.
- Hill, J., J. Waldfogel, and J. Brooks-Gunn (2002). “Differential effects of high-quality child care”. *Journal of Policy Analysis and Management* 21.4, pp. 601–627.
- Hines, O. et al. (2022). “Demystifying statistical learning based on efficient influence functions”. *The American Statistician* 76.3, pp. 292–304.
- Hitomi, K., Y. Nishiyama, and R. Okui (2008). “A puzzling phenomenon in semiparametric estimation problems with infinite-dimensional nuisance parameters”. *Econometric Theory* 24.6, pp. 1717–1728.
- Huang, Y. and P. B. Gilbert (2011). “Comparing biomarkers as principal surrogate endpoints”. *Biometrics* 67, pp. 1442–1451.
- Hubbard, A. E., M. J. van der Laan, and J. M. Robins (2000). “Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies”. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer, pp. 135–177.
- ICH E9 (R1) (2019). “International council for harmonisation of technical requirements for pharmaceuticals for human use: addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials”, pp. 1–19.

- Imai, K. (2008). “Sharp bounds on the causal effects in randomized experiments with “truncation-by-death””. *Statistics and Probability Letters* 78, pp. 144–149.
- Imai, K. and M. Ratkovic (2015). “Robust estimation of inverse probability weights for marginal structural models”. *Journal of the American Statistical Association* 110.511, pp. 1013–1023.
- Ionan, A. C. et al. (2023). “Clinical and statistical perspectives on the ICH E9 (R1) estimand framework implementation”. *Statistics in Biopharmaceutical Research* 15.3, pp. 554–559.
- Jastreboff, A. M. et al. (2022). “Tirzepatide Once Weekly for the Treatment of Obesity”. *The New England Journal of Medicine* 387.3, pp. 205–216. DOI: 10.1056/NEJMoa2206038.
- Jiang, Z. and P. Ding (2021). “Identification of causal effects within principal strata using auxiliary variables”. *Statistical Science* 36.4, pp. 493–508.
- Jiang, Z., P. Ding, and Z. Geng (2016). “Principal causal effect identification and surrogate endpoint evaluation by multiple trials”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78, pp. 829–848.
- Jiang, Z., S. Yang, and P. Ding (2022). “Multiply robust estimation of causal effects under principal ignorability”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 84, pp. 1423–1445.
- Jin, H. and D. B. Rubin (2008). “Principal stratification for causal inference with extended partial compliance”. *Journal of the American Statistical Association* 103.481, pp. 101–111.
- Jo, B. and E. A. Stuart (2009). “On the use of propensity scores in principal causal effect estimation”. *Statistics in Medicine* 28, pp. 2857–2875.
- Jo, B. et al. (2011). “The use of propensity scores in mediation analysis”. *Multivariate Behavioral Research* 46, pp. 425–452.
- Joffe, M. M., D. Small, and C.-Y. Hsu (2007). “Defining and estimating intervention effects for groups that will develop an auxiliary outcome”. *Statistical Science* 22, pp. 74–97.
- Kahan, B. C. et al. (2024). “The estimands framework: a primer on the ICH E9 (R1) addendum”. *bmj* 384.
- Kang, M. et al. (2022). “Incorporating estimands into clinical trial statistical analysis plans”. *Clinical Trials* 19.3, pp. 285–291.
- Kennedy, E. H., S. Lorch, and D. S. Small (2019). “Robust causal inference with continuous instruments using the local instrumental variable curve”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81.1, pp. 121–143.
- Kim, C. et al. (2019). “Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls”. *The Annals of Applied Statistics* 13.3, pp. 1927–1956.
- Klaassen, C. A. (1987). “Consistent estimation of the influence function of locally asymptotically linear estimators”. *The Annals of Statistics* 15.4, pp. 1548–1562.
- Koshevnik, Y. A. and B. Y. Levit (1977). “On a non-parametric analogue of the information matrix”. *Theory of Probability and Its Applications* 21.4, pp. 738–753.
- Laan, M. J. and J. M. Robins (2003a). *Unified methods for censored longitudinal data and causality*. Springer.

- Laan, M. J. van der and J. M. Robins (2003b). *Unified methods for censored longitudinal data and causality*. Vol. 5. Springer.
- Laan, M. J. van der and D. Rubin (2007). “A note on targeted maximum likelihood and right censored data”.
- Li, Q. and J. S. Racine (2023). *Nonparametric econometrics: theory and practice*. Princeton University Press.
- Li, Y., J. M. G. Taylor, and M. R. Elliott (2010). “A Bayesian approach to surrogacy assessment using principal stratification in clinical trials”. *Biometrics* 66, pp. 523–531.
- Lipkovich, I., B. Ratitch, and C. H. Mallinckrodt (2020). “Causal inference and estimands in clinical trials”. *Statistics in Biopharmaceutical Research* 12.1, pp. 54–67.
- Lu, S., Z. Jiang, and P. Ding (2026). “Principal stratification with continuous post-treatment variables: Nonparametric identification and semiparametric estimation”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 88.1, pp. 239–260.
- Lu, S. et al. (2025). “Estimating treatment effects with competing intercurrent events in randomized controlled trials”. *arXiv preprint arXiv:2503.03049*.
- Mattei, A., L. Forastiere, and F. Mealli (2023). “Assessing principal causal effects using principal score methods”. In: *Handbook of Matching and Weighting Adjustments for Causal Inference*. Chapman and Hall/CRC, pp. 313–348.
- Mattei, A. and F. Mealli (2011). “Augmented designs to assess principal strata direct effects”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, pp. 729–752.
- Mattei, A., F. Mealli, and B. Pacini (2014). “Identification of causal effects in the presence of nonignorable missing outcome values”. *Biometrics* 70, pp. 278–288.
- Mealli, F. and B. Pacini (2013). “Using secondary outcomes to sharpen inference in randomized experiments with noncompliance”. *Journal of the American Statistical Association* 108, pp. 1120–1131.
- Mealli, F., B. Pacini, and E. Stanghellini (2016). “Identification of principal causal effects using additional outcomes in concentration graphs”. *Journal of Educational and Behavioral Statistics* 41, pp. 463–480.
- Morand, E. F. et al. (2023). “Baricitinib for systemic lupus erythematosus: a double-blind, randomised, placebo-controlled, phase 3 trial (SLE-BRAVE-I)”. *The Lancet* 401.10381, pp. 1001–1010.
- Neugebauer, R. and M. van der Laan (2007). “Nonparametric causal effects based on marginal structural models”. *Journal of Statistical Planning and Inference* 137.2, pp. 419–434.
- Nguyen, T. Q. et al. (2024). “Sensitivity analysis for principal ignorability violation in estimating complier and noncomplier average causal effects”. *Statistics in Medicine* 43, pp. 3664–3688.
- Ogburn, E. L., A. Rotnitzky, and J. M. Robins (2015). “Doubly robust estimation of the local average treatment effect curve”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77.2, pp. 373–396.
- Olarte Parra, C., R. M. Daniel, and J. W. Bartlett (2023). “Hypothetical Estimands in Clinical Trials: A Unification of Causal Inference and Missing Data Methods”. *Statistics*

- in *Biopharmaceutical Research* 15.2, pp. 421–432. DOI: 10.1080/19466315.2022.2081599.
- Olarte Parra, C., R. M. Daniel, and J. W. Bartlett (2025a). “Dealing with multiple intercurrent events using hypothetical and treatment policy strategies simultaneously”. *arXiv preprint arXiv:2502.03329*.
- Olarte Parra, C. et al. (2025b). “Estimating hypothetical estimands with causal inference and missing data estimators in a diabetes trial case study”. *Biometrics* 81.1, ujae167.
- Petri, M. et al. (2023). “Baricitinib for systemic lupus erythematosus: a double-blind, randomised, placebo-controlled, phase 3 trial (SLE-BRAVE-II)”. *The Lancet* 401.10381, pp. 1011–1019.
- Pfanzagl, J. and W. Wefelmeyer (1985). “Contributions to a general asymptotic statistical theory”. *Statistics & Risk Modeling* 3.3-4, pp. 379–388.
- PHUSE Working Group (2025). *Implementation of ICH E9(R1) Estimands Framework Using Data Standards*. White Paper WP-92. PHUSE Estimands Implementation Working Group. PHUSE. URL: [https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/WP-92+Implementation+of+Estimands+\(ICH+E9+\(R1\)\)+using+Data+Standards.pdf](https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/WP-92+Implementation+of+Estimands+(ICH+E9+(R1))+using+Data+Standards.pdf).
- Polverejan, E. et al. (2023). “Defining Clinical Trial Estimands: A Practical Guide for Study Teams with Examples Based on a Psychiatric Disorder”. *Therapeutic Innovation & Regulatory Science* 57.5, pp. 911–939. DOI: 10.1007/s43441-023-00524-2.
- Qu, Y., L. Shurzinske, and S. Sethuraman (2021). “Defining estimands using a mix of strategies to handle intercurrent events in clinical trials”. *Pharmaceutical Statistics* 20.2, pp. 314–323.
- Ratitch, B. et al. (2020). “Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice”. *Therapeutic Innovation & Regulatory Science* 54.2, pp. 324–341. DOI: 10.1177/2168479019838827.
- Robins, J. M. and D. M. Finkelstein (2000). “Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests”. *Biometrics* 56.3, pp. 779–788.
- Robins, J. M. and A. Rotnitzky (1992). “Recovery of information and adjustment for dependent censoring using surrogate markers”. In: *AIDS epidemiology: methodological issues*. Springer, pp. 297–331.
- Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). “Estimation of regression coefficients when some regressors are not always observed”. *Journal of the American Statistical Association* 89.427, pp. 846–866.
- Rosenbaum, P. R. (2006). “Comment: the place of death in the quality of life”. *Statistical Science* 21.3, pp. 313–316.
- Rosenbaum, P. R. and D. B. Rubin (1983). “The central role of the propensity score in observational studies for causal effects”. *Biometrika* 70.1, pp. 41–55.
- Roy, J., J. W. Hogan, and B. H. Marcus (2008). “Principal stratification with predictors of compliance for randomized trials with 2 active treatments”. *Biostatistics* 9.2, pp. 277–289.

- Rubin, D. B. (2004). “Direct and indirect causal effects via potential outcomes”. *Scandinavian Journal of Statistics* 31, pp. 161–170.
- (2006). “Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death”. *Statistical Science* 21.3, pp. 299–309.
- (2007). “The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials”. *Statistics in medicine* 26.1, pp. 20–36.
- Schwartz, S. L., F. Li, and F. Mealli (2011). “A Bayesian semiparametric approach to intermediate variables in causal inference”. *Journal of the American Statistical Association* 106.496, pp. 1331–1344.
- Sjölander, A. et al. (2009). “Sensitivity analysis for principal stratum direct effects, with an application to a study of physical activity and coronary heart disease”. *Biometrics* 65.2, pp. 514–520.
- Stuart, E. A. and B. Jo (2015). “Assessing the sensitivity of methods for estimating principal causal effects”. *Statistical Methods in Medical Research* 24, pp. 657–674.
- Sun, S., J. G. Nešlehová, and E. E. Moodie (2024). “Principal stratification for quantile causal effects under partial compliance”. *Statistics in Medicine* 43.1, pp. 34–48.
- Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Vol. 4. Springer.
- Vaart, A. W. Van der (2000). *Asymptotic statistics*. Vol. 3. Cambridge university press.
- Vander Weele, T. J. and I. Shpitser (2011). “A new criterion for confounder selection”. *Biometrics* 67.4, pp. 1406–1413.
- Wang, L., T. S. Richardson, and X.-H. Zhou (2017). “Causal analysis of ordinal treatments and binary outcomes under truncation by death”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79, pp. 719–735.
- Westling, T. et al. (2024). “Inference for treatment-specific survival curves using machine learning”. *Journal of the American Statistical Association* 119.546, pp. 1541–1553.
- Wolock, C. J. et al. (2024). “A framework for leveraging machine learning tools to estimate personalized survival curves”. *Journal of Computational and Graphical Statistics* 33.3, pp. 1098–1108.
- Yang, F. and P. Ding (2018). “Using survival information in truncation by death problems without the monotonicity assumption”. *Biometrics* 74.4, pp. 1232–1239.
- Yang, F. and D. S. Small (2016). “Using post-outcome measurement information in censoring-by-death problems”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78, pp. 299–318.
- Ye, T. et al. (2023). “Instrumented difference-in-differences (with discussion)”. *Biometrics* 79.2, pp. 569–581.
- Yin, X. (2026). “Learning in the Limit: Income Inference from Credit Extensions”. *Journal of Finance*.
- Zheng, W. and M. J. van der Laan (2011). “Cross-validated targeted minimum-loss-based estimation”. *Targeted Learning: Causal Inference for Observational and Experimental Data*, pp. 459–474.

- Zigler, C. M., F. Dominici, and Y. Wang (2012). “Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes”. *Biostatistics* 13.2, pp. 289–302.
- Zubizarreta, J. R. (2015). “Stable weights that balance covariates for estimation with incomplete outcome data”. *Journal of the American Statistical Association* 110.511, pp. 910–922.

Appendix A

Appendix to Chapter 2

Section A.1 provides additional results, including methodological extensions, discussions of linear working models with illustrative examples, the EIF for η_0 , and an alternative projection parameter that directly projects the PCE surface to a working model.

Section A.2 provides proofs of the theorems, lemmas, and examples.

A.1 Additional results

In this section, we provide additional results that support and extend the main text.

Section A.1.1 presents a sensitivity analysis for the principal ignorability assumption and extends the framework to two additional settings: multiple post-treatment variables and conditional PCEs.

Section A.1.2 derives the EIF for η_1 and its closed-form solution under linear working models.

Section A.1.3 presents three illustrative examples with specific functional forms of linear working models.

Section A.1.4 provides the EIF for η_0 and its closed-form solution under a linear working model.

Section A.1.5 discusses an alternative projection parameter defined by projecting the PCE surface to a working model, and presents the corresponding EIF along with results on its relationship to our primary projection approach.

A.1.1 Extensions

Sensitivity analysis

The identification of PCEs relies on the crucial principal ignorability assumption, which is untestable using observed data. This necessitates sensitivity analysis to assess the robustness of conclusions to the violation of principal ignorability. We introduce the following sensitivity parameters:

$$\varepsilon_1(s_0, X) = \frac{E\{Y_1 \mid U = (s_1, s_0), X\}}{E(Y_1 \mid S_1 = s_1, X)} \quad \text{and} \quad \varepsilon_0(s_1, X) = \frac{E\{Y_0 \mid U = (s_1, s_0), X\}}{E(Y_0 \mid S_0 = s_0, X)} \quad (\text{A.1})$$

Values of these two parameters characterize the extent of principal ignorability violation; both $\varepsilon_1(s_0, X)$ and $\varepsilon_0(s_1, X)$ are equal to 1 when principal ignorability holds. This sensitivity analysis setup generalizes that for a binary post-treatment variable in Jiang et al. (2022) and Nguyen et al. (2024). Under (A.1), we can obtain results for identification, estimation, and semiparametric theory for the projection parameters in parallel to those under principal ignorability.

Multiple post-treatment variables

We can extend the framework to settings with multiple continuous post-treatment variables. For simplicity, we focus on the case with two continuous post-treatment variables. Suppose the two post-treatment variables, denoted by S and T , are both continuous with potential outcomes (S_1, S_0) and (T_1, T_0) . Their joint value determines the principal strata $U = (s_1, s_0, t_1, t_0)$. Again, there are infinitely many principal strata when both post-treatment variables are continuous, and the PCE surface can be defined as a function of s_1, s_0, t_1 , and t_0 :

$$\tau(s_1, s_0, t_1, t_0) = E\{Y_1 - Y_0 \mid U = (s_1, s_0, t_1, t_0)\}.$$

We generalize the principal ignorability assumption as $E\{Y_1 \mid U = (s_1, s_0, t_1, t_0), X\} = E\{Y_1 \mid U = (s_1, s'_0, t_1, t'_0), X\}$ for any s_0, s'_0, t_0 , and t'_0 , and $E\{Y_0 \mid U = (s_1, s_0, t_1, t_0), X\} = E\{Y_0 \mid U = (s'_1, s_0, t'_1, t_0), X\}$ for any s_1, s'_1, t_1 , and t'_1 . Under treatment ignorability, the principal ignorability assumption leads to the identification of the mean surfaces:

$$\begin{aligned} E\{Y_1 \mid U = (s_1, s_0, t_1, t_0), X\} &= E(Y_1 \mid S = s_1, T = t_1, Z = 1, X) \\ &= E(Y \mid S = s_1, T = t_1, Z = 1, X), \\ E\{Y_0 \mid U = (s_1, s_0, t_1, t_0), X\} &= E(Y_0 \mid S = s_0, T = t_0, Z = 0, X) \\ &= E(Y \mid S = s_0, T = t_0, Z = 0, X). \end{aligned}$$

Similar to the case with a single post-treatment variable, the identification of the PCE surface requires that of the principal density, denoted by $e(s_1, s_0, t_1, t_0, X) = p\{U = (s_1, s_0, t_1, t_0) \mid$

$X\}$. The conditional joint distributions of (S_1, T_1) and (S_0, T_0) given X , $p(S_1 = s_1, T_1 = t_1 | X)$ and $p(S_0 = s_0, T_0 = t_0 | X)$ respectively, are identifiable under treatment ignorability. Therefore, we can specify a copula model to obtain the principal density from these two distributions. Let $\mu_1(s_1, t_1, X) = E(Y | S = s_1, T = t_1, Z = 1, X)$ and $\mu_0(s_0, t_0, X) = E(Y | S = s_0, T = t_0, Z = 0, X)$ denote the conditional outcome means. We can obtain the following identification formula for the average potential outcome surface:

$$E\{Y_z | U = (s_1, s_0, t_1, t_0)\} = E\left\{\frac{e(s_1, s_0, t_1, t_0, X)}{e(s_1, s_0, t_1, t_0)}\mu_z(s_z, t_z, X)\right\},$$

where $e(s_1, s_0, t_1, t_0) = E\{e(s_1, s_0, t_1, t_0, X)\}$. We can also generalize the projection approach with multiple post-treatment variables and derive the EIF for the projection parameters. However, the derivation of the EIF becomes considerably more complex due to the increased number of models to perturb when computing the derivatives.

Conditional PCE

In addition to the marginal PCEs, we can also consider the conditional PCE surface:

$$\tau(s_1, s_0, w) = E\{Y_1 - Y_0 | U = (s_1, s_0), W = w\},$$

where W is a subset of the observed covariates X . Let $m_z(s_1, s_0, w) = E\{Y_z | U = (s_1, s_0), W = w\}$ denote the conditional potential outcome surfaces for $z = 0, 1$. Under Assumptions 2.1 and 2.2, we have the following identification formula for $m_1(s_1, s_0, w)$:

$$\begin{aligned} m_1(s_1, s_0, w) &= E[E\{Y_1 | U = (s_1, s_0), W = w, X\} | U = (s_1, s_0), W = w] \\ &= E\left\{\frac{e(s_1, s_0, X | W = w)}{e(s_1, s_0 | W = w)}\mu_1(s_1, X) | W = w\right\}, \end{aligned}$$

where $e(s_1, s_0, X | W = w) = p\{U = (s_1, s_0) | X, W = w\}$ is the conditional principal density on the subset $W = w$, and $e(s_1, s_0 | W = w) = E\{e(s_1, s_0, X | W = w) | W = w\}$ is marginalized over $X | W = w$. The identification of $m_0(s_1, s_0, w)$ can be achieved similarly. Employing the same projection approach, we define our causal parameter of interest as the projection parameter

$$\eta_z = \arg \min_{\eta} E[w_z(S_1, S_0, W) \{m_z(S_1, S_0, W) - f_z(S_1, S_0, W; \eta)\}^2],$$

where the pre-specified weights $w_z(s_1, s_0, w)$ and working models $f_z(s_1, s_0, w)$ are functions of w as well. We can then develop the estimation strategies and derive the efficiency theory for η_z following a similar approach in the main text.

In the special case when $W = X$, $m_1(s_1, s_0, x) = E\{Y_1 | U = (s_1, s_0), X = x\} = E(Y_1 | S_1 = s_1, X = x)$ is not a function of s_0 . Similarly, $m_0(s_1, s_0, x)$ is not a function of s_1 . We can simplify the notation to $m_z(s_z, x)$ for $z = 0, 1$, and specify the working models as functions of (s_z, x) only.

A.1.2 Linear working model

We explore the use of a linear working model for η_1 that strikes a balance between reasonable approximation and clear interpretation.

Example A.1 (Working model is linear in η). *Suppose that the working model $f_1(s_1, s_0; \eta)$ is linear in η , i.e., $f_1(s_1, s_0; \eta) = \eta^T g_1(s_1, s_0)$ for some known pre-specified function g_1 but unknown η . We introduce the notation*

$$G_1(s_1, s_0, X; h) = w_1(s_1, s_0) e(s_1, s_0, X) g_1(s_1, s_0) h(s_1, s_0, X)^T,$$

where $h(s_1, s_0, X)$ can take various forms below as a scalar function or a vector function whose dimension is compatible with $g_1(s_1, s_0)$. We write it as $G_1(s_1, s_0, X; 1)$ when $h(s_1, s_0, X) = 1$ is a constant function. The EIF for η_1 is

$$\varphi_1(V) = H_1^{-1} \{A_1(Y, S, Z, X) - A_2(Y, S, Z, X) \eta_1\},$$

where

$$H_1 = \iint w_1(s_1, s_0) g_1(s_1, s_0) g_1(s_1, s_0)^T e(s_1, s_0) ds_1 ds_0$$

and

$$\begin{aligned} & A_1(Y, S, Z, X) \\ &= \frac{Z}{\pi(X)} \left[\frac{\int G_1(S, s_0, X; \mu_1) ds_0}{p_1(S, X)} - \iint G_1(s_1, s_0, X; \mu_1) r_u(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &+ \frac{1-Z}{1-\pi(X)} \left[\frac{\int G_1(s_1, S, X; \mu_1) ds_1}{p_0(S, X)} - \iint G_1(s_1, s_0, X; \mu_1) r_v(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &+ \iint G_1(s_1, s_0, X; \mu_1) ds_1 ds_0 + \frac{Z}{\pi(X)} \frac{\int G_1(S, s_0, X; 1) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\}, \\ & A_2(Y, S, Z, X) \\ &= \frac{Z}{\pi(X)} \left[\frac{\int G_1(S, s_0, X; g_1) ds_0}{p_1(S, X)} - \iint G_1(s_1, s_0, X; g_1) r_u(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &+ \frac{1-Z}{1-\pi(X)} \left[\frac{\int G_1(s_1, S, X; g_1) ds_1}{p_0(S, X)} - \iint G_1(s_1, s_0, X; g_1) r_v(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &+ \iint G_1(s_1, s_0, X; g_1) ds_1 ds_0. \end{aligned}$$

We can, therefore, construct an estimator for η_1 by solving for Equation (2.20). Define the following estimated values

$$\hat{G}_1(s_1, s_0, X_i; h) = w_1(s_1, s_0) g_1(s_1, s_0) \hat{e}(s_1, s_0, X_i) h(s_1, s_0, X_i)^T,$$

$$\begin{aligned}\hat{r}_u(s_1, s_0, S_i, X_i) &= 1 - \frac{\hat{c}_u(s_1, s_0, X_i)}{\hat{c}(s_1, s_0, X_i)} \left\{ 1(S_i \leq s_1) - \hat{F}_1(s_1, X_i) \right\}, \\ \hat{r}_v(s_1, s_0, S_i, X_i) &= 1 - \frac{\hat{c}_v(s_1, s_0, X_i)}{\hat{c}(s_1, s_0, X_i)} \left\{ 1(S_i \leq s_0) - \hat{F}_0(s_0, X_i) \right\},\end{aligned}$$

where $\hat{c}_*(s_1, s_0, X_i)/\hat{c}(s_1, s_0, X_i)$ is the value of $\dot{c}_*(s_1, s_0, X_i)/c(s_1, s_0, X_i)$ for $* = u, v$ evaluated at the estimated distribution functions $\hat{F}_1(s_1, X_i)$ and $\hat{F}_0(s_0, X_i)$. Then, the estimator is

$$\hat{\eta}_{1,\text{elf}} = \frac{\sum_{i=1}^n \hat{A}_{2i}}{\sum_{i=1}^n \hat{A}_{1i}},$$

where

$$\begin{aligned}\hat{A}_{1i} &= \frac{Z_i}{\hat{\pi}(X_i)} \left[\frac{\int \hat{G}_1(S_i, s_0, X_i; \hat{\mu}_1) ds_0}{\hat{p}_1(S_i, X_i)} - \iint \hat{G}_1(s_1, s_0, X_i; \hat{\mu}_1) \hat{r}_u(s_1, s_0, S_i, X_i) ds_1 ds_0 \right] \\ &+ \frac{1 - Z_i}{1 - \hat{\pi}(X_i)} \left[\frac{\int \hat{G}_1(s_1, S_i, X_i; \hat{\mu}_1) ds_1}{\hat{p}_0(S_i, X_i)} - \iint \hat{G}_1(s_1, s_0, X_i; \hat{\mu}_1) \hat{r}_v(s_1, s_0, S_i, X_i) ds_1 ds_0 \right] \\ &+ \iint \hat{G}_1(s_1, s_0, X_i; \hat{\mu}_1) ds_1 ds_0 + \frac{Z_i}{\hat{\pi}(X_i)} \frac{\int \hat{G}_1(S_i, s_0, X_i; 1) ds_0}{\hat{p}_1(S_i, X_i)} \{Y_i - \hat{\mu}_1(S_i, X_i)\}\end{aligned}$$

and

$$\begin{aligned}\hat{A}_{2i} &= \frac{Z_i}{\hat{\pi}(X_i)} \left[\frac{\int \hat{G}_1(S_i, s_0, X_i; g_1) ds_0}{\hat{p}_1(S_i, X_i)} - \iint \hat{G}_1(s_1, s_0, X_i; g_1) \hat{r}_u(s_1, s_0, S_i, X_i) ds_1 ds_0 \right] \\ &+ \frac{1 - Z_i}{1 - \hat{\pi}(X_i)} \left[\frac{\int \hat{G}_1(s_1, S_i, X_i; g_1) ds_1}{\hat{p}_0(S_i, X_i)} - \iint \hat{G}_1(s_1, s_0, X_i; g_1) \hat{r}_v(s_1, s_0, S_i, X_i) ds_1 ds_0 \right] \\ &+ \iint \hat{G}_1(s_1, s_0, X_i; g_1) ds_1 ds_0\end{aligned}$$

are the fitted values of A_1 and A_2 .

Under the linear working model, the EIF for η_1 exhibits linearity with respect to η_1 . This linearity allows closed-form solutions to the system of linear equations, significantly reducing computational complexity in estimation.

A.1.3 Three examples on linear working models

We present three illustrative examples based on specific choices of functional forms when the working model is linear: (1) a constant working model, (2) a linear function of s_1 , and (3) a polynomial function of (s_1, s_0) .

Example A.2 (Working model is constant). *Suppose that the weight function and the working model are both constant, i.e., $w_1(s_1, s_0) = 1$ and $f_1(s_1, s_0; \eta) = \eta$ for all s_1, s_0 on the support of S_1, S_0 . The parameter of interest is simplified to*

$$\eta_1 = \arg \min_{\eta} E \{m_1(S_1, S_0) - \eta\}^2 = E \{m_1(S_1, S_0)\} = E(Y_1).$$

From Theorem 2.3, we have $H_1 = 1$ and

$$\begin{aligned} \ell_1(X) &= E(Y | Z = 1, X) - \eta_1, \\ \ell_2(S, Z, X) &= \frac{Z}{\pi(X)} \{\mu_1(S, X) - E(Y | Z = 1, X)\}, \\ \ell_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \{Y - \mu_1(S, X)\}. \end{aligned}$$

Therefore, the EIF for η_1 is

$$\varphi_1(V) = \frac{Z}{\pi(X)} \{Y - E(Y | Z = 1, X)\} + E(Y | Z = 1, X) - \eta_1,$$

which is the EIF for the average potential outcome under treatment in observational studies (Hahn, 1998). Based on $\varphi_1(V)$, we can derive the canonical doubly robust estimator for $E(Y_1)$ (Bang and Robins, 2005).

Example A.2 shows that when we project the PCE surface onto a constant, the parameter of interest becomes the average potential outcome $E(Y_1)$ and the EIF reduces to that for $E(Y_1)$ in the classic setting. This example acts as a sanity check of Theorem 2.3.

For any function $r(Y, S, X)$, define

$$\psi_{r(Y_z, S_z, X)} = \frac{1(Z = z)}{P(Z = z | X)} [r(Y, S, X) - E\{r(Y, S, X) | X, Z = z\}] + E\{r(Y, S, X) | X, Z = z\}.$$

Under Assumption 2.1, $E\{\psi_{r(Y_z, S_z, X)}\} = E\{r(Y_z, S_z, X)\}$. Below, we provide the EIFs under the other two working models.

Example A.3 (Projection onto space of linear functions of s_1). *Suppose the weight function $w_1(s_1, s_0)$ is constant, i.e., $w_1(s_1, s_0) = 1$, $f_1(s_1, s_0; \eta) = \eta^1 s_1 + \alpha$ is a linear function of s_1 , and distributions of S_1 and S_0 are independent conditional on X . The EIF for $(\eta^1, \alpha)^T$ is*

$$\varphi_1(V) = H_1^{-1} (\phi_1^1(V), \phi^\alpha(V))^T,$$

where

$$H_1 = \begin{pmatrix} E\{E(S^2 | Z = 1, X)\} & E\{E(S | Z = 1, X)\} \\ E\{E(S | Z = 1, X)\} & 1 \end{pmatrix},$$

$$\begin{aligned}\phi_1^1(V) &= \psi_{Y_1 S_1} - \eta^1 \psi_{S_1^2} - \alpha \psi_{S_1}, \\ \phi^\alpha(V) &= \psi_{Y_1} - \eta^1 \psi_{S_1} - \alpha.\end{aligned}$$

Example A.4 (Projection onto space of polynomial functions of (s_1, s_0)). Suppose the weight function $w_1(s_1, s_0)$ is constant, i.e., $w_1(s_1, s_0) = 1$, $f_1(s_1, s_0; \eta) = \sum_{j=1}^p \eta_j^1 s_1^j + \sum_{j=1}^p \eta_j^0 s_0^j + \alpha$ is a polynomial function of s_1 and s_0 , and distributions of S_1 and S_0 are independent conditional on X . Let $\bar{S}$ denote the $(2p+1)$ -dimensional vector $(S_1, \dots, S_1^p, S_0, \dots, S_0^p, 1)^T$. The EIF for $\eta = (\eta_1^1, \dots, \eta_p^1, \eta_1^0, \dots, \eta_p^0, \alpha)^T$ is

$$\varphi_1(V) = H_1^{-1} (\phi_1^1(V), \dots, \phi_p^1(V), \phi_1^0(V), \dots, \phi_p^0(V), \phi_\alpha(V))^T$$

where $H_1 = E(\bar{S}\bar{S}^T)$ with $E(S_1^j S_0^k) = E\{E(S^j \mid Z = 1, X)E(S^k \mid Z = 0, X)\}$ for $j, k = 1, \dots, p$, and

$$\begin{aligned}\phi_k^1(V) &= \psi_{Y_1 S_1^k} - \sum_{j=1}^p \eta_j^1 \psi_{S_1^{j+k}} - \sum_{j=1}^p \eta_j^0 \psi_{S_1^k S_0^j} - \alpha \psi_{S_1^k}, \\ \phi_k^0(V) &= \psi_{Y_1 S_0^k} - \sum_{j=1}^p \eta_j^1 \psi_{S_1^j S_0^k} - \sum_{j=1}^p \eta_j^0 \psi_{S_0^{j+k}} - \alpha \psi_{S_0^k}, \\ \phi^\alpha(V) &= \psi_{Y_1} - \sum_{j=1}^p \eta_j^1 \psi_{S_1^j} - \sum_{j=1}^p \eta_j^0 \psi_{S_0^j} - \alpha.\end{aligned}$$

for $k = 1, \dots, p$.

Because working models are linear in the unknown parameter η_1 , we can construct closed-form estimators based on the EIFs.

A.1.4 EIF for η_0

In this section, we provide the EIF for η_0 and the closed-form solution of η_0 under a linear working model. Denote

$$\nu_0(s_1, s_0, X) = w_0(s_1, s_0) \dot{f}_0(s_1, s_0; \eta_0) e(s_1, s_0, X) \{\mu_0(s_0, X) - f_0(s_1, s_0; \eta_0)\}.$$

The following theorem presents the EIF for η_0 .

Theorem A.1. Under the nonparametric model, the EIF for η_0 is

$$\varphi_0(V) = H_0^{-1} \{k_1(X) + k_2(S, Z, X) + k_3(Y, S, Z, X)\},$$

where

$$H_0 = E \left(w_0(S_1, S_0) \left[\dot{f}_0(S_1, S_0; \eta_0) \dot{f}_0(S_1, S_0; \eta_0)^T - \{m_0(S_1, S_0) - f_0(S_1, S_0; \eta_0)\} \ddot{f}_0(S_1, S_0; \eta_0) \right] \right)$$

and

$$\begin{aligned} k_1(X) &= \iint \nu_0(s_1, s_0, X) ds_1 ds_0, \\ k_2(S, Z, X) &= \frac{Z}{\pi(X)} \left\{ \frac{\int \nu_0(S, s_0, X) ds_0}{p_1(S, X)} - \iint \nu_0(s_1, s_0, X) r_u(s_1, s_0, S, X) ds_1 ds_0 \right\} \\ &\quad + \frac{1-Z}{1-\pi(X)} \left\{ \frac{\int \nu_0(s_1, S, X) ds_1}{p_0(S, X)} - \iint \nu_0(s_1, s_0, X) r_v(s_1, s_0, S, X) ds_1 ds_0 \right\}, \\ k_3(Y, S, Z, X) &= \frac{1-Z}{1-\pi(X)} \frac{\int w_0(s_1, S) \dot{f}_0(s_1, S; \eta_0) e(s_1, S, X) ds_1}{p_0(S, X)} \{Y - \mu_0(S, X)\}. \end{aligned}$$

The following example provides the EIF under a linear working model.

Example A.5 (Working model is linear in η_0). Suppose that the working model $f_0(s_1, s_0; \eta_0)$ is linear in η_0 , i.e., $f_0(s_1, s_0; \eta_0) = \eta_0^T g_0(s_1, s_0)$ for some function g_0 . Denote

$$G_0(s_1, s_0, X; h) = w_0(s_1, s_0) g_0(s_1, s_0) e(s_1, s_0, X) h(s_1, s_0, X)^T,$$

where $h(s_1, s_0, X)$ can be a scalar function or vector function whose dimension is compatible with $g_0(s_1, s_0)$. The EIF for η_0 is

$$\varphi_0(V) = H_0^{-1} \{A_3(Y, S, Z, X) - A_4(Y, S, Z, X) \eta_0\},$$

where

$$\begin{aligned} H_0 &= \iint w_0(s_1, s_0) g_0(s_1, s_0) g_0(s_1, s_0)^T e(s_1, s_0) ds_1 ds_0, \\ A_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \left[\frac{\int G_0(S, s_0, X; \mu_0) ds_0}{p_1(S, X)} - \iint G_0(s_1, s_0, X; \mu_0) r_u(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &\quad + \frac{1-Z}{1-\pi(X)} \left[\frac{\int G_0(s_1, S, X; \mu_0) ds_1}{p_0(S, X)} - \iint G_0(s_1, s_0, X; \mu_0) r_v(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &\quad + \iint G_0(s_1, s_0, X; \mu_0) ds_1 ds_0 + \frac{1-Z}{1-\pi(X)} \frac{\int G_0(s_1, S, X; 1) ds_1}{p_0(S, X)} \{Y - \mu_0(S, X)\}, \\ A_4(Y, S, Z, X) &= \frac{Z}{\pi(X)} \left[\frac{\int G_0(S, s_0, X; g_0) ds_0}{p_1(S, X)} - \iint G_0(s_1, s_0, X; g_0) r_u(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &\quad + \frac{1-Z}{1-\pi(X)} \left[\frac{\int G_0(s_1, S, X; g_0) ds_1}{p_0(S, X)} - \iint G_0(s_1, s_0, X; g_0) r_v(s_1, s_0, S, X) ds_1 ds_0 \right] \\ &\quad + \iint G_0(s_1, s_0, X; g_0) ds_1 ds_0. \end{aligned}$$

We can solve for the closed-form solution of η_0 and construct the estimator, $\hat{\eta}_{0,\text{eif}}$, from the EIF.

A.1.5 Projection of $\tau(s_1, s_0)$ and EIF for η_τ

We can project the PCE surface $\tau(s_1, s_0; \eta)$ directly without the two separate projections for the average potential outcome means. This section presents the results for the projection parameter η_τ , defined as

$$\eta_\tau = \arg \min_{\eta \in \mathbb{R}^q} E \left[w_\tau(S_1, S_0) \{ \tau(S_1, S_0) - f_\tau(S_1, S_0; \eta) \}^2 \right]. \quad (\text{A.2})$$

The identification of η_τ also follows from the identification of $\tau(s_1, s_0)$. Denote

$$\begin{aligned} \nu_\tau(s_1, s_0, X) &= w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) e(s_1, s_0, X) \{ \mu_1(s_1, X) - \mu_0(s_0, X) - f_\tau(s_1, s_0; \eta_\tau) \}, \\ r_u(s_1, s_0, S, X) &= 1 - \frac{\dot{c}_u(s_1, s_0, X)}{c(s_1, s_0, X)} \{ 1(S \leq s_1) - F_1(s_1, X) \}, \\ r_v(s_1, s_0, S, X) &= 1 - \frac{\dot{c}_v(s_1, s_0, X)}{c(s_1, s_0, X)} \{ 1(S \leq s_0) - F_0(s_0, X) \}. \end{aligned}$$

The following theorem presents the EIF for η_τ .

Theorem A.2. *Under the nonparametric model, the EIF for η_τ is*

$$\varphi_\tau(V) = H_\tau^{-1} \{ t_1(X) + t_2(S, Z, X) + t_3(Y, S, Z, X) \},$$

where

$$H_\tau = E \left(w_\tau(S_1, S_0) \left[\dot{f}_\tau(S_1, S_0; \eta_\tau) \dot{f}_\tau(S_1, S_0; \eta_\tau)^T - \{ m_\tau(S_1, S_0) - f_\tau(S_1, S_0; \eta_\tau) \} \ddot{f}_\tau(S_1, S_0; \eta_\tau) \right] \right)$$

and

$$\begin{aligned} t_1(X) &= \iint \nu_\tau(s_1, s_0, X) ds_1 ds_0, \\ t_2(S, Z, X) &= \frac{Z}{\pi(X)} \left\{ \frac{\int_{\mathcal{S}_0} \nu_\tau(S, s_0, X) ds_0}{p_1(S, X)} - \iint \nu_\tau(s_1, s_0, X) r_u(s_1, s_0, S, X) ds_1 ds_0 \right\} \\ &\quad + \frac{1-Z}{1-\pi(X)} \left\{ \frac{\int_{\mathcal{S}_1} \nu_\tau(s_1, S, X) ds_1}{p_0(S, X)} - \iint \nu_\tau(s_1, s_0, X) r_v(s_1, s_0, S, X) ds_1 ds_0 \right\}, \\ t_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \frac{\int_{\mathcal{S}_0} w_\tau(S, s_0) \dot{f}_\tau(S, s_0; \eta_\tau) e(S, s_0, X) ds_0}{p_1(S, X)} \{ Y - \mu_1(S, X) \} \\ &\quad - \frac{1-Z}{1-\pi(X)} \frac{\int_{\mathcal{S}_1} w_\tau(s_1, S) \dot{f}_\tau(s_1, S; \eta_\tau) e(s_1, S, X) ds_1}{p_0(S, X)} \{ Y - \mu_0(S, X) \}. \end{aligned}$$

We can also construct an estimator based on the EIF for η_τ . The following corollary compares the EIF for η_τ and the EIF for $\eta_1 - \eta_0$ from the separate projections.

Corollary A.1. *Suppose that*

1. *the working models have the same linear form in the parameters of interest, i.e., $f_*(s_1, s_0; \eta_*) = \eta_*^T g(s_1, s_0)$;*
2. *the specified weights are the same, i.e., $w_1(s_1, s_0) = w_0(s_1, s_0) = w_\tau(s_1, s_0)$ for all values of s_1 and s_0 on the support of S_1 and S_0 .*

The EIF for $\eta_1 - \eta_0$ is the same as that for η_τ , which indicates that parameters defined by projecting $m_1(s_1, s_0)$ and $m_0(s_1, s_0)$ separately and then computing the difference have the same semiparametric efficiency bound as those obtained by projecting $\tau(s_1, s_0)$ directly.

The proof follows directly from the algebraic facts that $H_1 = H_0 = H_\tau$ and $t_j = \ell_j - k_j$ for $j = 1, 2, 3$ in the EIF formulas.

From Corollary A.1, if $\hat{\eta}_{1,\text{eif}}$ and $\hat{\eta}_{0,\text{eif}}$ are the EIF estimators under linear working models $f_1(s_1, s_0; \eta_1) = \eta_1^T g(s_1, s_0)$ and $f_0(s_1, s_0; \eta_0) = \eta_0^T g(s_1, s_0)$, respectively, then the EIF estimator for $\tau(s_1, s_0)$ under the linear working model $f_\tau(s_1, s_0) = \eta_\tau^T g(s_1, s_0)$ is $\hat{\eta}_{\tau,\text{eif}} = \hat{\eta}_{1,\text{eif}} - \hat{\eta}_{0,\text{eif}}$.

A.2 Proofs

In this section, we provide the proofs of all the results in the main text and the Supplementary Material. We first introduce the following lemma to simplify the proofs.

Lemma A.1. *Let X and Y be two random variables with densities $f_1(x)$ and $f_2(y)$. For any function $h(\cdot)$ with $E\{h(X)\} < \infty$, we have*

$$E\{h(X)\} = E\left\{\frac{f_1(Y)}{f_2(Y)}h(Y)\right\}.$$

The proof is straightforward and thus omitted.

A.2.1 Proof of Theorem 2.1

We have

$$\begin{aligned} E\{Y_1 \mid X, U = (s_1, s_0)\} &= E\{Y_1 \mid X, U = (s_1, s_0) \text{ or } (s_1, s'_0), \forall s'_0\} \\ &= E\{Y_1 \mid X, Z = 1, U = (s_1, s_0) \text{ or } (s_1, s'_0), \forall s'_0\} \\ &= E(Y_1 \mid X, Z = 1, S_1 = s_1) \end{aligned}$$

$$= \mu_1(s_1, X),$$

where the first equality follows from principal ignorability and the third equality follows from treatment ignorability. Therefore,

$$\begin{aligned} E\{Y_1 \mid U = (s_1, s_0)\} &= E[E\{Y_1 \mid X, U = (s_1, s_0)\} \mid U = (s_1, s_0)] \\ &= E\{\mu_1(s_1, X) \mid U = (s_1, s_0)\} \\ &= E\left[\frac{p\{X \mid U = (s_1, s_0)\}}{p(X)} \mu_1(s_1, X)\right] \\ &= E\left\{\frac{e(s_1, s_0, X)}{e(s_1, s_0)} \mu_1(s_1, X)\right\}, \end{aligned} \tag{A.3}$$

where the first equality follows from the law of iterated expectations and the third equality follows from Lemma A.1. The identification formulas for $E\{Y_0 \mid U = (s_1, s_0)\}$ and $\tau(s_1, s_0)$ can be obtained similarly. $\square$

A.2.2 Proof of Theorem 2.2

The derivation in (2.5) follows directly from the law of iterated expectations.

A.2.3 Proof of Lemma 2.1

From Theorem 2.1, we have

$$\begin{aligned} m_1(s_1, s_0) - f_1(s_1, s_0; \eta_1) &= E\{Y_1 - f_1(s_1, s_0; \eta_1) \mid U = (s_1, s_0)\} \\ &= E\left[\frac{e(s_1, s_0, X)}{e(s_1, s_0)} \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\}\right]. \end{aligned} \tag{A.4}$$

Plugging (A.4) into (2.10) yields $E\{D_{1, \text{pd}+\text{om}}(Y, S, Z, X; \eta_1, e, \mu_1)\} = 0$. Similarly, we can obtain $E\{D_{0, \text{pd}+\text{om}}(Y, S, Z, X; \eta_0, e, \mu_0)\} = 0$. $\square$

A.2.4 Proof of Lemma 2.2

We only prove the result for η_1 . Consider the expectation of the term within the integral in the estimating equation. We have

$$\begin{aligned} &E\left[w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) \frac{e(S, s_0, X)}{p_1(S, X)} \frac{Z}{\pi(X)} \{Y - f_1(S, s_0; \eta_1)\}\right] \\ &= E\left[E\left\{w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) \frac{Y - f_1(S, s_0; \eta_1)}{p_1(S, X)} \mid Z = 1, X\right\}\right] \end{aligned}$$

$$\begin{aligned}
&= E \left[E \left\{ w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) \frac{\mu_1(S, X) - f_1(S, s_0; \eta_1)}{p_1(S, X)} \mid Z = 1, X \right\} \right] \\
&= E \left[\int \left\{ w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \frac{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)}{p_1(s_1, X)} \right\} p(S = s_1 \mid Z = 1, X) ds_1 \right] \\
&= E \left[\int w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 \right] \\
&= \int E \left[w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} \right] ds_1,
\end{aligned}$$

where the first two equalities follow from the law of iterated expectation. Therefore,

$$\begin{aligned}
&E \{D_{1, \text{tp+pd}}(Y, S, Z, X; \eta_1, \pi, e)\} \\
&= \int E \left[w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) \frac{1}{p_1(S, X)} \frac{Z}{\pi(X)} \{Y - f_1(S, s_0; \eta_1)\} \right] ds_0 \\
&= E \left[\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \right] \\
&= 0,
\end{aligned}$$

where the last equality follows from Lemma 2.1.

Similarly, we can obtain $E \{D_{0, \text{tp+pd}}(Y, S, Z, X; \eta_0, e, \mu_0)\} = 0$. $\square$

A.2.5 Proof of Theorem 2.3

Nonparametric model setup

We follow the semiparametric theory in Bickel et al. (1993) to derive the EIF for our projection parameter η_1 . We denote the vector of all observed variables $V = (Y, S, Z, X)$ and factorize the likelihood as

$$p(V) = p(X)p(Z \mid X)p(S \mid Z, X)p(Y \mid S, Z, X).$$

To derive the EIF for η_1 , we consider a one-dimensional parametric submodel $p_\theta(V)$ which contains the true model $p(V)$ at $\theta = 0$. We use θ in the subscript to denote the quantity with respect to the submodel, e.g., $e_\theta(s_1, s_0)$ is the value of $e(s_1, s_0)$ with respect to the submodel. We use dot to denote the partial derivative with respect to θ , e.g., $\dot{e}_\theta(s_1, s_0) = \partial e_\theta(s_1, s_0) / \partial \theta$, and use $\mathfrak{s}_\theta(\cdot)$ to denote the score function with respect to the submodel. The decomposition of the score function under the submodel is

$$\mathfrak{s}_\theta(V) = \mathfrak{s}_\theta(X) + \mathfrak{s}_\theta(Z \mid X) + \mathfrak{s}_\theta(S \mid Z, X) + \mathfrak{s}_\theta(Y \mid S, Z, X),$$

where $\mathfrak{s}_\theta(X) = \partial \log p_\theta(X) / \partial \theta$, $\mathfrak{s}_\theta(Z \mid X) = \partial \log p_\theta(Z \mid X) / \partial \theta$, $\mathfrak{s}_\theta(S \mid Z, X) = \partial \log p_\theta(S \mid Z, X) / \partial \theta$, and $\mathfrak{s}_\theta(Y \mid S, Z, X) = \partial \log p_\theta(Y \mid S, Z, X) / \partial \theta$ are the score functions corresponding to the four components of the likelihood function. We write $\mathfrak{s}_\theta(\cdot)|_{\theta=0}$ as $\mathfrak{s}(\cdot)$, which is

the score function evaluated at the true model $\theta = 0$ under the one-dimensional submodel. The tangent space

$$\Lambda = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \mathcal{H}_4$$

is the direct sum of

$$\begin{aligned}\mathcal{H}_1 &= \{h(X) : \mathbb{E}\{h(X)\} = 0\}, \\ \mathcal{H}_2 &= \{h(Z, X) : \mathbb{E}\{h(Z, X) \mid X\} = 0\}, \\ \mathcal{H}_3 &= \{h(S, Z, X) : \mathbb{E}\{h(S, Z, X) \mid Z, X\} = 0\}, \\ \mathcal{H}_4 &= \{h(Y, Z, S, X) : \mathbb{E}\{h(Y, Z, S, X) \mid Z, S, X\} = 0\},\end{aligned}$$

where $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, and $\mathcal{H}_4$ are orthogonal to each other. Let $\varphi_1(V)$ denote the EIF for η_1 . It must satisfy the equation

$$\dot{\eta}_{1,\theta}|_{\theta=0} := \left. \frac{\partial \eta_{1,\theta}}{\partial \theta} \right|_{\theta=0} = \mathbb{E} \{ \varphi_1(V) \mathfrak{s}(V) \}.$$

To find such $\varphi_1(V)$, we need to calculate the derivative $\dot{\eta}_{1,\theta}|_{\theta=0}$. We first state several lemmas to simplify the proof.

Lemmas

Lemma A.2. *For $e(s_1, s_0, X)$ and $e(s_1, s_0)$, we have*

$$\begin{aligned}\dot{e}_\theta(s_1, s_0, X)|_{\theta=0} &= e(s_1, s_0, X) \{ \mathfrak{s}(S = s_1 \mid Z = 1, X) + \mathfrak{s}(S = s_0 \mid Z = 0, X) \} \\ &\quad + \dot{c}_u(F_1(s_1, X), F_0(s_0, X)) E \{ 1(S \leq s_1) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \} p_1(s_1, X) p_0(s_0, X) \\ &\quad + \dot{c}_v(F_1(s_1, X), F_0(s_0, X)) E \{ 1(S \leq s_0) \mathfrak{s}(S \mid Z = 0, X) \mid Z = 0, X \} p_1(s_1, X) p_0(s_0, X)\end{aligned}$$

and

$$\begin{aligned}\dot{e}_\theta(s_1, s_0)|_{\theta=0} &= E [e(s_1, s_0, X) \{ \mathfrak{s}(X) + \mathfrak{s}(S = s_1 \mid Z = 1, X) + \mathfrak{s}(S = s_0 \mid Z = 0, X) \}] \\ &\quad + E [\dot{c}_u(F_1(s_1, X), F_0(s_0, X)) E \{ 1(S \leq s_1) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \} p_1(s_1, X) p_0(s_0, X)] \\ &\quad + E [\dot{c}_v(F_1(s_1, X), F_0(s_0, X)) E \{ 1(S \leq s_0) \mathfrak{s}(S \mid Z = 0, X) \mid Z = 0, X \} p_1(s_1, X) p_0(s_0, X)].\end{aligned}$$

Proof of Lemma A.2. By the identification of the principal density in (2.1), we have the Gateaux derivative

$$\dot{e}_\theta(s_1, s_0, X)|_{\theta=0} = \left. \frac{\partial}{\partial \theta} e_\theta(s_1, s_0, X) \right|_{\theta=0}$$

$$= \left. \frac{\partial}{\partial \theta} \{c(F_{1,\theta}(s_1, X), F_{0,\theta}(s_0, X))p_{1,\theta}(s_1, X)p_{0,\theta}(s_0, X)\} \right|_{\theta=0}$$

$$= c(F_1(s_1, X), F_0(s_0, X)) \dot{p}_{1,\theta}(s_1, X)|_{\theta=0} p_0(s_0, X) \quad (\text{A.5})$$

$$+ c(F_1(s_1, X), F_0(s_0, X))p_1(s_1, X) \dot{p}_{0,\theta}(s_0, X)|_{\theta=0} \quad (\text{A.6})$$

$$+ \left. \frac{\partial}{\partial \theta} \{c(F_{1,\theta}(s_1, X), F_{0,\theta}(s_0, X))\} \right|_{\theta=0} p_1(s_1, X)p_0(s_0, X). \quad (\text{A.7})$$

We compute the three terms in (A.5)–(A.7) separately. First, for (A.5), we have

$$\begin{aligned} & c(F_1(s_1, X), F_0(s_0, X)) \dot{p}_{1,\theta}(s_1, X)|_{\theta=0} p_0(s_0, X) \\ &= c(F_1(s_1, X), F_0(s_0, X))p_1(s_1, X)p_0(s_0, X)\mathfrak{s}(S = s_1 \mid Z = 1, X) \\ &= e(s_1, s_0, X)\mathfrak{s}(S = s_1 \mid Z = 1, X). \end{aligned}$$

Similarly, the second term (A.6) is equal to $e(s_1, s_0, X)\mathfrak{s}(S = s_0 \mid Z = 0, X)$.

The derivative in (A.7) is

$$\begin{aligned} & \left. \frac{\partial}{\partial \theta} \{c(F_{1,\theta}(s_1, X), F_{0,\theta}(s_0, X))\} \right|_{\theta=0} \\ &= \dot{c}_u(F_1(s_1, X), F_0(s_0, X)) \left. \frac{\partial}{\partial \theta} F_{1,\theta}(s_1, X) \right|_{\theta=0} + \dot{c}_v(F_1(s_1, X), F_0(s_0, X)) \left. \frac{\partial}{\partial \theta} F_{0,\theta}(s_0, X) \right|_{\theta=0} \\ &= \dot{c}_u(F_1(s_1, X), F_0(s_0, X))E\{1(S \leq s_1)\mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X\} \\ & \quad + \dot{c}_v(F_1(s_1, X), F_0(s_0, X))E\{1(S \leq s_0)\mathfrak{s}(S \mid Z = 0, X) \mid Z = 0, X\}, \end{aligned}$$

where the last equality follows from the fact that for $z = 0, 1$,

$$\begin{aligned} \left. \frac{\partial}{\partial \theta} F_{z,\theta}(s_z, X) \right|_{\theta=0} &= \left. \frac{\partial}{\partial \theta} E_\theta\{1(S \leq s_z) \mid Z = z, X\} \right|_{\theta=0} \\ &= E\{1(S \leq s_z)\mathfrak{s}(S \mid Z = z, X) \mid Z = z, X\}. \end{aligned}$$

Combining all three terms, we have

$$\begin{aligned} & \dot{e}_\theta(s_1, s_0, X)|_{\theta=0} \\ &= e(s_1, s_0, X) \{\mathfrak{s}(S = s_1 \mid Z = 1, X) + \mathfrak{s}(S = s_0 \mid Z = 0, X)\} \\ & \quad + \dot{c}_u(F_1(s_1, X), F_0(s_0, X))E\{1(S \leq s_1)\mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X\} p_1(s_1, X)p_0(s_0, X) \\ & \quad + \dot{c}_v(F_1(s_1, X), F_0(s_0, X))E\{1(S \leq s_0)\mathfrak{s}(S \mid Z = 0, X) \mid Z = 0, X\} p_1(s_1, X)p_0(s_0, X). \end{aligned}$$

For $e(s_1, s_0)$, we have

$$\begin{aligned} & \dot{e}_\theta(s_1, s_0)|_{\theta=0} \\ &= \left. \frac{\partial}{\partial \theta} E_\theta\{e_\theta(s_1, s_0, X)\} \right|_{\theta=0} \end{aligned}$$

$$\begin{aligned}
&= E_\theta \{e_\theta(s_1, s_0, X) \mathfrak{s}_\theta(X)\}_{|\theta=0} + E \{\dot{e}_\theta(s_1, s_0, X)\}_{|\theta=0} \\
&= E[e(s_1, s_0, X) \{\mathfrak{s}(X) + \mathfrak{s}(S = s_1 | Z = 1, X) + \mathfrak{s}(S = s_0 | Z = 0, X)\}] \\
&\quad + E[\dot{e}_u(F_1(s_1, X), F_0(s_0, X)) E\{1(S \leq s_1) \mathfrak{s}(S | Z = 1, X) | Z = 1, X\} p_1(s_1, X) p_0(s_0, X)] \\
&\quad + E[\dot{e}_v(F_1(s_1, X), F_0(s_0, X)) E\{1(S \leq s_0) \mathfrak{s}(S | Z = 0, X) | Z = 0, X\} p_1(s_1, X) p_0(s_0, X)].
\end{aligned}$$

□

Lemma A.3. For $\mu_z(s_z, X)$, we have

$$\begin{aligned}
\dot{\mu}_{1,\theta}(s_1, X)|_{\theta=0} &= E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | S = s_1, Z = 1, X) | S = s_1, Z = 1, X], \\
\dot{\mu}_{0,\theta}(s_0, X)|_{\theta=0} &= E[\{Y - \mu_0(s_0, X)\} \mathfrak{s}(Y | S = s_0, Z = 0, X) | S = s_0, Z = 0, X].
\end{aligned}$$

Proof of Lemma A.3. We have

$$\begin{aligned}
\dot{\mu}_{1,\theta}(s_1, X)|_{\theta=0} &= \frac{\partial}{\partial \theta} E_\theta(Y | S = s_1, Z = 1, X) \Big|_{\theta=0} \\
&= E\{Y \mathfrak{s}(Y | S = s_1, Z = 1, X) | S = s_1, Z = 1, X\} \\
&= E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | S = s_1, Z = 1, X) | S = s_1, Z = 1, X],
\end{aligned}$$

where the last equality follows from $E\{\mu_1(s_1, X) \mathfrak{s}(Y | S = s_1, Z = 1, X) | S = s_1, Z = 1, X\} = 0$. The derivation of $\dot{\mu}_{0,\theta}(s_0, X)|_{\theta=0}$ is omitted due to similarity. □

Lemma A.4. Denote $\lambda_1(s_1, s_0) = E\{e(s_1, s_0, X) \mu_1(s_1, X)\}$, $\lambda_0(s_1, s_0) = E\{e(s_1, s_0, X) \mu_0(s_0, X)\}$, and $\lambda(s_1, s_0) = \lambda_1(s_1, s_0) - \lambda_0(s_1, s_0)$. We have

$$\dot{\lambda}_\theta(s_1, s_0) \Big|_{\theta=0} = \dot{\lambda}_{1,\theta}(s_1, s_0) \Big|_{\theta=0} - \dot{\lambda}_{0,\theta}(s_1, s_0) \Big|_{\theta=0},$$

where

$$\begin{aligned}
\dot{\lambda}_{z,\theta}(s_1, s_0) \Big|_{\theta=0} &= E\{e(s_1, s_0, X) \mu_z(s_z, X) \mathfrak{s}(X)\} + E\{\dot{e}_\theta(s_1, s_0, X) \Big|_{\theta=0} \mu_z(s_z, X)\} \\
&\quad + E\{e(s_1, s_0, X) \dot{\mu}_{z,\theta}(s_z, X) \Big|_{\theta=0}\}.
\end{aligned}$$

Proof of Lemma A.4. By the chain rule, we have

$$\begin{aligned}
\dot{\lambda}_{z,\theta}(s_1, s_0) \Big|_{\theta=0} &= \frac{\partial}{\partial \theta} E_\theta\{e_\theta(s_1, s_0, X) \mu_{1,\theta}(s_1, X)\} \Big|_{\theta=0} \\
&= E\{e(s_1, s_0, X) \mu_z(s_z, X) \mathfrak{s}(X)\} + E\{\dot{e}_\theta(s_1, s_0, X) \Big|_{\theta=0} \mu_z(s_z, X)\} \\
&\quad + E\{e(s_1, s_0, X) \dot{\mu}_{z,\theta}(s_z, X) \Big|_{\theta=0}\}.
\end{aligned}$$

□

EIF for η_1

We derive the EIF for η_1 . The EIFs for η_0 and η_τ can be obtained similarly. Because Equation (2.10) holds for any parametric submodel, we have

$$\begin{aligned} 0 &= E_\theta \left[w_1(S_1, S_0) \dot{f}_1(S_1, S_0; \eta_{1,\theta}) \{m_{1,\theta}(S_1, S_0) - f_1(S_1, S_0; \eta_{1,\theta})\} \right] \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_{1,\theta}) \{m_{1,\theta}(s_1, s_0) - f_1(s_1, s_0; \eta_{1,\theta})\} e_\theta(s_1, s_0) ds_1 ds_0 \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_{1,\theta}) \{\lambda_{1,\theta}(s_1, s_0) - e_\theta(s_1, s_0) f_1(s_1, s_0; \eta_{1,\theta})\} ds_1 ds_0, \end{aligned}$$

where the third equality follows from (A.3). Taking the first order derivative with respect to θ and evaluating at $\theta = 0$, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_{1,\theta}) \{\lambda_{1,\theta}(s_1, s_0) - e_\theta(s_1, s_0) f_1(s_1, s_0; \eta_{1,\theta})\} ds_1 ds_0 \Big|_{\theta=0} \\ &= \iint w_1(s_1, s_0) \ddot{f}_1(s_1, s_0; \eta_1) \dot{\eta}_{1,\theta}|_{\theta=0} \{\lambda_1(s_1, s_0) - e(s_1, s_0) f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \\ &\quad + \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \left\{ \dot{\lambda}_{1,\theta}(s_1, s_0) \Big|_{\theta=0} - \dot{e}_\theta(s_1, s_0) \Big|_{\theta=0} f_1(s_1, s_0; \eta_1) \right\} ds_1 ds_0 \\ &\quad - \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \dot{f}_1(s_1, s_0; \eta_1)^T e(s_1, s_0) \dot{\eta}_{1,\theta}|_{\theta=0} ds_1 ds_0, \end{aligned}$$

where the second equality follows from the chain rule. Therefore, we can write

$$\dot{\eta}_{1,\theta}|_{\theta=0} = H_1^{-1} \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \left\{ \dot{\lambda}_{1,\theta}(s_1, s_0) \Big|_{\theta=0} - \dot{e}_\theta(s_1, s_0) \Big|_{\theta=0} f_1(s_1, s_0; \eta_1) \right\} ds_1 ds_0, \quad (\text{A.8})$$

where

$$\begin{aligned} H_1 &= - \iint w_1(s_1, s_0) \ddot{f}_1(s_1, s_0; \eta_1) \{\lambda_1(s_1, s_0) - f_1(s_1, s_0; \eta_1) e(s_1, s_0)\} ds_1 ds_0 \\ &\quad + \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \dot{f}_1(s_1, s_0; \eta_1)^T e(s_1, s_0) ds_1 ds_0 \\ &= \iint w_1(s_1, s_0) \left[\dot{f}_1(s_1, s_0; \eta_1) \dot{f}_1(s_1, s_0; \eta_1)^T - \{m_1(s_1, s_0) - f_1(s_1, s_0; \eta_1)\} \ddot{f}_1(s_1, s_0; \eta_1) \right] \\ &\quad e(s_1, s_0) ds_1 ds_0 \\ &= E \left(w_1(S_1, S_0) \left[\dot{f}_1(S_1, S_0; \eta_1) \dot{f}_1(S_1, S_0; \eta_1)^T - \{m_1(S_1, S_0) - f_1(S_1, S_0; \eta_1)\} \ddot{f}_1(S_1, S_0; \eta_1) \right] \right), \end{aligned}$$

where the second equality follows from (A.3). Taking the second-order derivative of the objective function in the optimization problem (2.6), we obtain its Hessian matrix

$$2E \left(w_1(S_1, S_0) \left[\dot{f}_1(S_1, S_0; \eta_1) \dot{f}_1(S_1, S_0; \eta_1)^T - \{m_1(S_1, S_0) - f_1(S_1, S_0; \eta_1)\} \ddot{f}_1(S_1, S_0; \eta_1) \right] \right).$$

which is equal to $2H_1$.

We will calculate the two terms within the integral in (A.8) separately.

Part I. By Lemma A.4, we have

$$\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \dot{\lambda}_{1,\theta}(s_1, s_0) \Big|_{\theta=0} ds_1 ds_0 = B_1 + B_2 + B_3,$$

where

$$\begin{aligned} B_1 &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) E\{e(s_1, s_0, X) \mu_1(s_1, X) \mathfrak{s}(X)\} ds_1 ds_0, \\ B_2 &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) E\{\dot{e}_\theta(s_1, s_0, X)|_{\theta=0} \mu_1(s_1, X)\} ds_1 ds_0, \\ B_3 &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) E\{e(s_1, s_0, X) \dot{\mu}_{1,\theta}(s_1, X)|_{\theta=0}\} ds_1 ds_0. \end{aligned}$$

First, consider B_1 . We have

$$B_1 = E\left\{\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \mu_1(s_1, X) ds_1 ds_0 \cdot \mathfrak{s}(X)\right\}.$$

Because $E\{\mathfrak{s}(X)\} = 0$, we can write

$$B_1 = E\{h_1(X) \mathfrak{s}(X)\},$$

where

$$\begin{aligned} h_1(X) &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \mu_1(s_1, X) ds_1 ds_0 \\ &\quad - E\left\{\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \mu_1(s_1, X) ds_1 ds_0\right\} \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \mu_1(s_1, X) ds_1 ds_0 \\ &\quad - E\left\{\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) f_1(s_1, s_0; \eta_1) ds_1 ds_0\right\} \quad (\text{Lemma 2.1}) \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \mu_1(s_1, X) ds_1 ds_0 \\ &\quad - \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) f_1(s_1, s_0; \eta_1) e(s_1, s_0) ds_1 ds_0. \end{aligned}$$

Next, consider B_2 . By Lemma A.2, we can further write

$$B_2 = B_{21} + B_{22} + B_{23} + B_{24},$$

where

$$\begin{aligned}
B_{21} &= \iint w_1 \dot{f}_1 E \{ e(s_1, s_0, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) \mu_1(s_1, X) \} ds_1 ds_0, \\
B_{22} &= \iint w_1 \dot{f}_1 E \{ e(s_1, s_0, X) \mathfrak{s}(S = s_0 \mid Z = 0, X) \mu_1(s_1, X) \} ds_1 ds_0, \\
B_{23} &= \iint w_1 \dot{f}_1 E [\dot{c}_u E \{ 1(S \leq s_1) s(S \mid Z = 1, X) \mid Z = 1, X \} p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X)] ds_1 ds_0, \\
B_{24} &= \iint w_1 \dot{f}_1 E [\dot{c}_v E \{ 1(S \leq s_0) s(S \mid Z = 0, X) \mid Z = 0, X \} p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X)] ds_1 ds_0.
\end{aligned}$$

We suppress the dependence of $w_1(s_1, s_0)$, $\dot{f}_1(s_1, s_0; \eta_1)$, $\dot{c}_u(F_1(s_1, X), F_0(s_0, X))$, and $\dot{c}_v(F_1(s_1, X), F_0(s_0, X))$ on (s_1, s_0, X) and η_1 for ease of notation.

Denote

$$\begin{aligned}
\tilde{\mu}_{c11}(s_1, X) &= \int w_1 \dot{f}_1 \frac{e(s_1, s_0, X)}{p_1(s_1, X)} \mu_1(s_1, X) ds_0, \\
\tilde{\mu}_{c01}(s_0, X) &= \int w_1 \dot{f}_1 \frac{e(s_1, s_0, X)}{p_0(s_0, X)} \mu_1(s_1, X) ds_1.
\end{aligned}$$

For B_{21} , we have

$$\begin{aligned}
B_{21} &= \iint w_1 \dot{f}_1 E \{ e(s_1, s_0, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) \mu_1(s_1, X) \} ds_1 ds_0 \\
&= E \left[\int \left\{ \int w_1 \dot{f}_1 \frac{e(s_1, s_0, X)}{p_1(s_1, X)} \mu_1(s_1, X) ds_0 \right\} p_1(s_1, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) ds_1 \right] \\
&= E \left\{ \int \tilde{\mu}_{c11}(s_1, X) p_1(s_1, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) ds_1 \right\} \\
&= E [E \{ \tilde{\mu}_{c11}(S, X) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \}] \\
&= E \left\{ \frac{Z}{\pi(X)} \tilde{\mu}_{c11}(S, X) \mathfrak{s}(S \mid Z, X) \right\},
\end{aligned}$$

where the last equality follows from the law of iterated expectations. Similarly, for B_{22} , we have

$$B_{22} = E \left\{ \frac{1-Z}{1-\pi(X)} \tilde{\mu}_{c01}(S, X) \mathfrak{s}(S \mid Z, X) \right\}.$$

Denote

$$\begin{aligned}
\tilde{\mu}_{\dot{c}11}(t_1, X) &= \iint 1(s_1 \geq t_1) w_1 \dot{f}_1 \dot{c}_u p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X) ds_1 ds_0, \\
\tilde{\mu}_{\dot{c}01}(t_0, X) &= \iint 1(s_0 \geq t_0) w_1 \dot{f}_1 \dot{c}_v p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X) ds_1 ds_0,
\end{aligned}$$

For B_{23} , we have

$$\begin{aligned}
B_{23} &= \iint w_1 \dot{f}_1 E [\dot{c}_u E \{1(S \leq s_1) \mathfrak{s}(S | Z = 1, X) | Z = 1, X\} p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X)] ds_1 ds_0 \\
&= E \left[\iint w_1 \dot{f}_1 \dot{c}_u \left\{ \int 1(t_1 \leq s_1) \mathfrak{s}(t_1 | Z = 1, X) p_1(t_1, X) dt_1 \right\} p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X) ds_1 ds_0 \right] \\
&= E \left[\int \left\{ \iint 1(s_1 \geq t_1) w_1 \dot{f}_1 \dot{c}_u p_1(s_1, X) p_0(s_0, X) \mu_1(s_1, X) ds_1 ds_0 \right\} \mathfrak{s}(t_1 | Z = 1, X) p_1(t_1, X) dt_1 \right] \\
&= E \left\{ \int \tilde{\mu}_{c11}(t_1, X) \mathfrak{s}(t_1 | Z = 1, X) p_1(t_1, X) dt_1 \right\} \\
&= E [E \{ \tilde{\mu}_{c11}(S, X) \mathfrak{s}(S | Z = 1, X) | Z = 1, X \}] \\
&= E \left\{ \frac{Z}{\pi(X)} \tilde{\mu}_{c11}(S, X) \mathfrak{s}(S | Z, X) \right\},
\end{aligned}$$

where $\mathfrak{s}(t_1 | Z = 1, X)$ denotes $\mathfrak{s}(S = t_1 | Z = 1, X)$, and the last equality follows from the law of iterated expectations. Similarly, for B_{24} , we have

$$B_{24} = E \left\{ \frac{1-Z}{1-\pi(X)} \tilde{\mu}_{c01}(S, X) \mathfrak{s}(S | Z, X) \right\}.$$

Therefore, combining the four terms of B_2 , we have

$$\begin{aligned}
B_2 &= E \left(\left[\frac{Z}{\pi(X)} \{ \tilde{\mu}_{c11}(S, X) + \tilde{\mu}_{c11}(S, X) \} + \frac{1-Z}{1-\pi(X)} \{ \tilde{\mu}_{c01}(S, X) + \tilde{\mu}_{c01}(S, X) \} \right] \mathfrak{s}(S | Z, X) \right) \\
&= E \left[\left\{ \frac{Z}{\pi(X)} \tilde{\mu}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{\mu}_{01}(S, X) \right\} \mathfrak{s}(S | Z, X) \right],
\end{aligned}$$

where we denote $\tilde{\mu}_{z1}(s_z, X) = \tilde{\mu}_{cz1}(s_z, X) + \tilde{\mu}_{\dot{c}z1}(s_z, X)$ for $z = 0, 1$. As a result, we have

$$B_2 = E \{ h_2(S, Z, X) \mathfrak{s}(S | Z, X) \},$$

where

$$\begin{aligned}
h_2(S, Z, X) &= \frac{Z}{\pi(X)} \tilde{\mu}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{\mu}_{01}(S, X) \\
&\quad - E \left\{ \frac{Z}{\pi(X)} \tilde{\mu}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{\mu}_{01}(S, X) | Z, X \right\} \\
&= \frac{Z}{\pi(X)} [\tilde{\mu}_{11}(S, X) - E \{ \tilde{\mu}_{11}(S, X) | Z = 1, X \}] \\
&\quad + \frac{1-Z}{1-\pi(X)} [\tilde{\mu}_{01}(S, X) - E \{ \tilde{\mu}_{01}(S, X) | Z = 0, X \}].
\end{aligned}$$

Then, consider B_3 . Denote

$$\tilde{e}_1(s_1, x) = \int w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, x) ds_0.$$

From Lemma A.3, we have

$$\begin{aligned}
B_3 &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \\
&\quad \cdot E\{e(s_1, s_0, X) E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | s_1, Z = 1, X) | s_1, Z = 1, X]\} ds_1 ds_0 \\
&= \int E(\tilde{e}_1(s_1, X) E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | s_1, Z = 1, X) | s_1, Z = 1, X]) ds_1 \\
&= E\left(\int \tilde{e}_1(s_1, X) E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | s_1, Z = 1, X) | s_1, Z = 1, X] ds_1\right). \quad (\text{A.9})
\end{aligned}$$

By the law of iterated expectation, we have

$$\begin{aligned}
&E\left[\frac{Z}{\pi(X)} \frac{\tilde{e}_1(S, X)}{p_1(S, X)} \{Y - \mu_1(S, X)\} \mathfrak{s}(Y | S, Z, X)\right] \\
&= E\left(E\left[\frac{\tilde{e}_1(S, X)}{p_1(S, X)} \{Y - \mu_1(S, X)\} \mathfrak{s}(Y | S, Z, X) | X, Z = 1\right]\right) \\
&= E\left(\int E\left[\frac{\tilde{e}_1(s_1, X)}{p_1(s_1, X)} \{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | s_1, Z, X) | S = s_1, Z = 1, X\right] p(S = s_1 | Z = 1, X) ds_1\right) \\
&= E\left(\int \tilde{e}_1(s_1, X) E[\{Y - \mu_1(s_1, X)\} \mathfrak{s}(Y | s_1, Z, X) | S = s_1, Z = 1, X] ds_1\right),
\end{aligned}$$

which, coupled with (A.9), implies that

$$B_3 = E\{h_3(Y, S, Z, X) \mathfrak{s}(Y | S, Z, X)\},$$

where

$$h_3(Y, S, Z, X) = \frac{Z}{\pi(X)} \frac{\int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\}.$$

Part II. Consider the second part of $\dot{\eta}_{1,\theta}|_{\theta=0}$ within the integral in (A.8). By Lemma A.2, we have

$$\iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) f_1(s_1, s_0; \eta_1) \dot{e}_\theta(s_1, s_0)|_{\theta=0} ds_1 ds_0 = B_4 + B_5,$$

where

$$\begin{aligned}
B_4 &= \iint w_1 \dot{f}_1 f_1 E\{e(s_1, s_0, X) \mathfrak{s}(X)\} ds_1 ds_0, \\
B_5 &= \iint w_1 \dot{f}_1 f_1 E[e(s_1, s_0, X) \{\mathfrak{s}(S = s_1 | Z = 1, X) + \mathfrak{s}(S = s_0 | Z = 0, X)\}] ds_1 ds_0 \\
&\quad + \iint w_1 \dot{f}_1 f_1 E[\dot{c}_u E\{1(S \leq s_1) \mathfrak{s}(S | Z = 1, X) | Z = 1, X\} p_1(s_1, X) p_0(s_0, X)] ds_1 ds_0
\end{aligned}$$

$$+ \iint w_1 \dot{f}_1 f_1 E [\dot{c}_v E \{1(S \leq s_0) \mathfrak{s}(S | Z = 0, X) | Z = 0, X\} p_1(s_1, X) p_0(s_0, X)] ds_1 ds_0.$$

Again, we suppress the dependence of $w_1(s_1, s_0)$, $\dot{f}_1(s_1, s_0; \eta_1)$, $f_1(s_1, s_0; \eta_1)$, $\dot{c}_u(F_1(s_1, X), F_0(s_0, X))$ and $\dot{c}_v(F_1(s_1, X), F_0(s_0, X))$ on (s_1, s_0, X) and η_1 for ease of notation.

First, consider B_4 . We have

$$\begin{aligned} B_4 &= \iint w_1 \dot{f}_1 f_1 E \{e(s_1, s_0, X) \mathfrak{s}(X)\} ds_1 ds_0 \\ &= E \left[\left\{ \iint w_1 \dot{f}_1 f_1 e(s_1, s_0, X) ds_1 ds_0 \right\} \mathfrak{s}(X) \right] \\ &= E \{h_4(X) \mathfrak{s}(X)\}, \end{aligned}$$

where

$$\begin{aligned} h_4(X) &= \iint w_1 \dot{f}_1 f_1 e(s_1, s_0, X) ds_1 ds_0 - E \left\{ \iint w_1 \dot{f}_1 f_1 e(s_1, s_0, X) ds_1 ds_0 \right\} \\ &= \iint w_1 \dot{f}_1 f_1 \{e(s_1, s_0, X) - e(s_1, s_0)\} ds_1 ds_0. \end{aligned}$$

Next, we consider B_5 . We can further write B_5 as

$$B_5 = B_{51} + B_{52} + B_{53} + B_{54},$$

where

$$\begin{aligned} B_{51} &= \iint w_1 \dot{f}_1 f_1 E \{e(s_1, s_0, X) \mathfrak{s}(S = s_1 | Z = 1, X)\} ds_1 ds_0, \\ B_{52} &= \iint w_1 \dot{f}_1 f_1 E \{e(s_1, s_0, X) \mathfrak{s}(S = s_0 | Z = 1, X)\} ds_1 ds_0, \\ B_{53} &= \iint w_1 \dot{f}_1 f_1 E [\dot{c}_u E \{1(S \leq s_1) \mathfrak{s}(S | Z = 1, X) | Z = 1, X\} p_1(s_1, X) p_0(s_0, X)] ds_1 ds_0, \\ B_{54} &= \iint w_1 \dot{f}_1 f_1 E [\dot{c}_v E \{1(S \leq s_0) \mathfrak{s}(S | Z = 0, X) | Z = 0, X\} p_1(s_1, X) p_0(s_0, X)] ds_1 ds_0. \end{aligned}$$

Denote

$$\begin{aligned} \tilde{f}_{c11}(s_1, X) &= \int w_1 \dot{f}_1 f_1 \frac{e(s_1, s_0, X)}{p_1(s_1, X)} ds_0, \\ \tilde{f}_{c01}(s_0, X) &= \int w_1 \dot{f}_1 f_1 \frac{e(s_1, s_0, X)}{p_0(s_0, X)} ds_1. \end{aligned}$$

For B_{51} , we have

$$B_{51} = \iint w_1 \dot{f}_1 f_1 E \{e(s_1, s_0, X) \mathfrak{s}(S = s_1 | Z = 1, X)\} ds_1 ds_0$$

$$\begin{aligned}
&= E \left[\int \left\{ \int w_1 \dot{f}_1 f_1 \frac{e(s_1, s_0, X)}{p_1(s_1, X)} ds_0 \right\} p_1(s_1, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) ds_1 \right] \\
&= E \left\{ \int \tilde{f}_{c11}(s_1, X) p_1(s_1, X) \mathfrak{s}(S = s_1 \mid Z = 1, X) ds_1 \right\} \\
&= E \left[E \left\{ \tilde{f}_{c11}(S, X) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \right\} \right] \\
&= E \left\{ \frac{Z}{\pi(X)} \tilde{f}_{c11}(S, X) \mathfrak{s}(S \mid Z, X) \right\}
\end{aligned}$$

and similarly,

$$B_{52} = E \left\{ \frac{1-Z}{1-\pi(X)} \tilde{f}_{c01}(S, X) \mathfrak{s}(S \mid Z, X) \right\}.$$

Denote

$$\begin{aligned}
\tilde{f}_{\dot{c}11}(t_1, X) &= \iint 1(s_1 \geq t_1) w_1 \dot{f}_1 f_1 \dot{c}_u p_1(s_1, X) p_0(s_0, X) ds_1 ds_0, \\
\tilde{f}_{\dot{c}01}(t_0, X) &= \iint 1(s_0 \geq t_0) w_1 \dot{f}_1 f_1 \dot{c}_v p_1(s_1, X) p_0(s_0, X) ds_1 ds_0.
\end{aligned}$$

For B_{53} , we have

$$\begin{aligned}
B_{53} &= \iint w_1 \dot{f}_1 f_1 E [\dot{c}_u E \{ 1(S \leq s_1) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \} p_1(s_1, X) p_0(s_0, X)] ds_1 ds_0 \\
&= E \left[\iint w_1 \dot{f}_1 f_1 \dot{c}_u \left\{ \int 1(t_1 \leq s_1) \mathfrak{s}(t_1 \mid Z = 1, X) p_1(t_1, X) dt_1 \right\} p_1(s_1, X) p_0(s_0, X) ds_1 ds_0 \right] \\
&= E \left[\int \left\{ \iint 1(s_1 \geq t_1) w_1 \dot{f}_1 f_1 \dot{c}_u p_1(s_1, X) p_0(s_0, X) ds_1 ds_0 \right\} \mathfrak{s}(t_1 \mid Z = 1, X) p_1(t_1, X) dt_1 \right] \\
&= E \left\{ \int \tilde{f}_{\dot{c}11}(t_1, X) p_1(t_1, X) \mathfrak{s}(t_1 \mid Z = 1, X) dt_1 \right\} \\
&= E \left[E \left\{ \tilde{f}_{\dot{c}11}(S, X) \mathfrak{s}(S \mid Z = 1, X) \mid Z = 1, X \right\} \right] \\
&= E \left\{ \frac{Z}{\pi(X)} \tilde{f}_{\dot{c}11}(S, X) \mathfrak{s}(S \mid Z, X) \right\}
\end{aligned}$$

and similarly,

$$B_{54} = E \left\{ \frac{1-Z}{1-\pi(X)} \tilde{f}_{\dot{c}01}(S, X) \mathfrak{s}(S \mid Z, X) \right\}.$$

Therefore, we obtain

$$B_5 = E \left(\left[\frac{Z}{\pi(X)} \left\{ \tilde{f}_{c11}(S, X) + \tilde{f}_{\dot{c}11}(S, X) \right\} + \frac{1-Z}{1-\pi(X)} \left\{ \tilde{f}_{c01}(S, X) + \tilde{f}_{\dot{c}01}(S, X) \right\} \right] \mathfrak{s}(S \mid Z, X) \right)$$

$$= E \left[\left\{ \frac{Z}{\pi(X)} \tilde{f}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{f}_{01}(S, X) \right\} \mathfrak{s}(S | Z, X) \right],$$

where we denote $\tilde{f}_{z1}(s_z, X) = \tilde{f}_{cz1}(s_z, X) + \tilde{f}_{\dot{c}z1}(s_z, X)$ for $z = 0, 1$. As a result, we have

$$B_5 = E \{ h_5(S, Z, X) \mathfrak{s}(S | Z, X) \},$$

where

$$\begin{aligned} h_5(S, Z, X) &= \frac{Z}{\pi(X)} \tilde{f}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{f}_{01}(S, X) \\ &\quad - E \left\{ \frac{Z}{\pi(X)} \tilde{f}_{11}(S, X) + \frac{1-Z}{1-\pi(X)} \tilde{f}_{01}(S, X) \mid Z, X \right\} \\ &= \frac{Z}{\pi(X)} \left[\tilde{f}_{11}(S, X) - E \left\{ \tilde{f}_{11}(S, X) \mid Z = 1, X \right\} \right] \\ &\quad + \frac{1-Z}{1-\pi(X)} \left[\tilde{f}_{01}(S, X) - E \left\{ \tilde{f}_{01}(S, X) \mid Z = 0, X \right\} \right]. \end{aligned}$$

Summary of the proof. We have

$$\begin{aligned} H_1 \dot{\eta}_{1,\theta}|_{\theta=0} &= B_1 + B_2 + B_3 + B_4 + B_5 \\ &= E \{ h_1(X) \mathfrak{s}(X) \} + E \{ h_2(S, Z, X) \mathfrak{s}(S | Z, X) \} \\ &\quad + E \{ h_3(Y, S, Z, X) \mathfrak{s}(Y | S, Z, X) \} \\ &\quad - E \{ h_4(X) \mathfrak{s}(X) \} - E \{ h_5(S, Z, X) \mathfrak{s}(S | Z, X) \} \\ &= E \{ [h_1(X) - h_4(X)] \mathfrak{s}(X) \} \\ &\quad + E \{ [h_2(S, Z, X) - h_5(S, Z, X)] \mathfrak{s}(S | Z, X) \} \\ &\quad + E \{ h_3(Y, S, Z, X) \mathfrak{s}(Y | S, Z, X) \}. \end{aligned}$$

We can verify that

$$\begin{aligned} H_1^{-1} \{ h_1(X) - h_4(X) \} &\in \mathcal{H}_1, \\ H_1^{-1} \{ h_2(S, Z, X) - h_5(S, Z, X) \} &\in \mathcal{H}_3, \\ H_1^{-1} h_3(Y, S, Z, X) &\in \mathcal{H}_4. \end{aligned}$$

By the fact that $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, and $\mathcal{H}_4$ are orthogonal to each other, we have

$$\begin{aligned} \dot{\eta}_{1,\theta}|_{\theta=0} &= E \left[H_1^{-1} \{ h_1(X) - h_4(X) \} \mathfrak{s}(V) \right] \\ &\quad + E \left[H_1^{-1} \{ h_2(S, Z, X) - h_5(S, Z, X) \} \mathfrak{s}(V) \right] \\ &\quad + E \left\{ H_1^{-1} h_3(Y, S, Z, X) \mathfrak{s}(V) \right\}. \end{aligned}$$

Therefore, the EIF for η_1 is

$$\varphi_1(V) = H_1^{-1} \{ \ell_1(X) + \ell_2(S, Z, X) + \ell_3(Y, S, Z, X) \},$$

where

$$\begin{aligned}
\ell_1(X) &= h_1(X) - h_4(X) \\
&= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0, \\
\ell_2(S, Z, X) &= h_2(S, Z, X) - h_5(S, Z, X) \\
&= \frac{Z}{\pi(X)} [m_{11}(S, X) - E\{m_{11}(S, X) \mid Z = 1, X\}] \\
&\quad + \frac{1-Z}{1-\pi(X)} [m_{01}(S, X) - E\{m_{01}(S, X) \mid Z = 0, X\}], \\
\ell_3(Y, S, Z, X) &= h_3(Y, S, Z, X) \\
&= \frac{Z}{\pi(X)} \frac{\int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\},
\end{aligned}$$

with $m_{z1}(S, X) = \tilde{\mu}_{z1}(S, X) - \tilde{f}_{z1}(S, X)$ for $z = 0, 1$.

Finally, we simplify the expression of $\ell_2(S, Z, X)$. Denote $m_{11}(S, X) = m_{c11}(S, X) + m_{\dot{c}11}(S, X)$ with

$$\begin{aligned}
m_{c11}(S, X) &= \int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) \frac{e(S, s_0, X)}{p_1(S, X)} \{\mu_1(S, X) - f_1(S, s_0; \eta_1)\} ds_0, \\
m_{\dot{c}11}(S, X) &= \iint 1(s_1 \geq S) w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \dot{c}_u(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \\
&\quad \cdot \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0.
\end{aligned}$$

For $m_{c11}(S, X)$, we have

$$\begin{aligned}
&E\{m_{c11}(S, X) \mid Z = 1, X\} \\
&= \iint w_1 \dot{f}_1 \frac{e(s_1, s_0, X)}{p_1(s_1, X)} \{\mu_1(s_1, X) - f_1\} p_1(s_1, X) ds_1 ds_0 \\
&= \iint w_1 \dot{f}_1 e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1\} ds_1 ds_0.
\end{aligned}$$

For $m_{\dot{c}11}(S, X)$, we have

$$\begin{aligned}
&E\{m_{\dot{c}11}(S, X) \mid Z = 1, X\} \\
&= \iint 1(t_1 \geq s_1) w_1(t_1, s_0) \dot{f}_1 \dot{c}_u(t_1, s_0, X) p_1(s_1, X) p_1(t_1, X) p_0(s_0, X) \{\mu_1(t_1, X) - f_1\} dt_1 ds_0 ds_1 \\
&= \iint 1(t_1 \geq s_1) w_1(t_1, s_0) \dot{f}_1 \dot{c}_u(t_1, s_0, X) p_1(s_1, X) p_1(t_1, X) p_0(s_0, X) \{\mu_1(t_1, X) - f_1\} dt_1 ds_0 ds_1 \\
&= \iint F_1(t_1, X) w_1(t_1, s_0) \dot{f}_1 \dot{c}_u(t_1, s_0, X) p_1(t_1, X) p_0(s_0, X) \{\mu_1(t_1, X) - f_1\} dt_1 ds_0
\end{aligned}$$

$$= \iint F_1(s_1, X) w_1(s_1, s_0) \dot{f}_1(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \{\mu_1(s_1, X) - f_1\} ds_1 ds_0,$$

where we suppress the dependence of $\dot{f}_1$ and f_1 on the parameters and change t_1 to s_1 in the last equality. Therefore, we obtain

$$m_{11}(S, X) - E\{m_{11}(S, X) \mid Z = 1, X\} = m_{c11}(S, X) - m_{r11}(S, X),$$

where

$$\begin{aligned} m_{r11}(S, X) &= E\{m_{c11}(S, X) \mid Z = 1, X\} - [m_{c11}(S, X) - E\{m_{c11}(S, X) \mid Z = 1, X\}] \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \\ &\quad - \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \frac{\dot{c}_u(s_1, s_0, X)}{c(s_1, s_0, X)} e(s_1, s_0, X) \\ &\quad \cdot \{1(S \leq s_1) - F_1(s_1, X)\} \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} \\ &\quad \cdot \left[1 - \frac{\dot{c}_u(s_1, s_0, X)}{c(s_1, s_0, X)} \{1(S \leq s_1) - F_1(s_1, X)\}\right] ds_1 ds_0 \end{aligned}$$

Similarly, we can denote $m_{01}(S, X) = m_{c01}(S, X) + m_{c01}(S, X)$ with

$$\begin{aligned} m_{c01}(S, X) &= \int w_1(s_1, S) \dot{f}_1(s_1, S; \eta_1) \frac{e(s_1, S, X)}{p_0(S, X)} \{\mu_1(s_1, X) - f_1(s_1, S; \eta_1)\} ds_1, \\ m_{c01}(S, X) &= \iint 1(s_0 \geq S) w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \dot{c}_v(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \\ &\quad \cdot \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \end{aligned}$$

and obtain

$$m_{01}(S, X) - E\{m_{11}(S, X) \mid Z = 0, X\} = m_{c01}(S, X) - m_{r01}(S, X),$$

where

$$\begin{aligned} m_{r01}(S, X) &= E\{m_{c11}(S, X) \mid Z = 0, X\} - [m_{c11}(S, X) - E\{m_{c11}(S, X) \mid Z = 0, X\}] \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \\ &\quad - \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \frac{\dot{c}_v(s_1, s_0, X)}{c(s_1, s_0, X)} e(s_1, s_0, X) \\ &\quad \cdot \{1(S \leq s_0) - F_0(s_0, X)\} \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} ds_1 ds_0 \\ &= \iint w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) e(s_1, s_0, X) \{\mu_1(s_1, X) - f_1(s_1, s_0; \eta_1)\} \\ &\quad \cdot \left[1 - \frac{\dot{c}_v(s_1, s_0, X)}{c(s_1, s_0, X)} \{1(S \leq s_0) - F_0(s_0, X)\}\right] ds_1 ds_0. \end{aligned}$$

□

A.2.6 Proof of Example A.2

First, we have

$$\eta_1 = \arg \min_{\eta} E \{m_1(S_1, S_0) - \eta\}^2 = E \{m_1(S_1, S_0)\} = E(Y_1).$$

Because $w_1(s_1, s_0) = 1$ and $f_1(s_1, s_0; \eta) = \eta$ for all s_1, s_0 , we have $\dot{f}_1(s_1, s_0; \eta) = 1$ and $\ddot{f}_1(s_1, s_0; \eta) = 0$. Thus, $H_1 = 1$ and $\varphi_1(V) = \ell_1(X) + \ell_2(S, Z, X) + \ell_3(Y, S, Z, X)$. We will compute these three terms using results in Theorem 2.3 and the following properties of the copula density function $c(\cdot, \cdot)$:

$$\int c(s_1, s_0, X) p_0(s_0, X) ds_0 = \int \frac{e(s_1, s_0, X)}{p_1(s_1, X)} ds_0 = 1, \quad (\text{A.10})$$

$$\int \dot{c}_u(s_1, s_0, X) p_0(s_0, X) ds_0 = 0, \quad (\text{A.11})$$

where (A.10) follows from (2.1) and (A.11) follows from (A.10) by taking the first order derivative with respect to $F_1(s_1, X)$ on both sides.

For $\ell_1(X)$, we have

$$\begin{aligned} \ell_1(X) &= \iint e(s_1, s_0, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 ds_0 \\ &= E(Y \mid Z = 1, X) - \eta_1. \end{aligned}$$

For $\ell_2(S, Z, X)$, we compute $m_{c11}(S, X)$, $m_{r11}(S, X)$, $m_{c01}(S, X)$, and $m_{r01}(S, X)$ separately based on the expressions in the proof of Theorem 2.3. We have

$$m_{c11}(S, X) = \int \frac{e(S, s_0, X)}{p_1(S, X)} \{\mu_1(S, X) - \eta_1\} ds_0 = \mu_1(S, X) - \eta_1.$$

and

$$\begin{aligned} m_{\dot{c}11}(S, X) &= \iint \dot{c}_u(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) 1(S \leq s_1) \{\mu_1(s_1, X) - \eta_1\} ds_1 ds_0 \\ &= \int \left\{ \int \dot{c}_u(s_1, s_0, X) p_0(s_0, X) ds_0 \right\} p_1(s_1, X) 1(S \leq s_1) \{\mu_1(s_1, X) - \eta_1\} ds_1 \\ &= 0, \end{aligned}$$

where the last equality follows from (A.11). Therefore,

$$m_{r01}(S, X) = E\{m_{c11}(S, X) \mid Z = 1, X\} = E(Y \mid Z = 1, X) - \eta_1.$$

For $m_{c01}(S, X)$, we consider

$$\begin{aligned} m_{c01}(s_0, X) &= \int \frac{e(s_1, s_0, X)}{p_0(s_0, X)} \{\mu_1(s_1, X) - \eta_1\} ds_1 \\ &= \int c(s_1, s_0, X) p_1(s_1, X) \{\mu_1(s_1, X) - \eta_1\} ds_1. \end{aligned} \quad (\text{A.12})$$

From the chain rule, we have

$$\frac{\partial}{\partial t_0} c(s_1, t_0, X) = \dot{c}_v(s_1, t_0, X) p_0(t_0, X).$$

Integrating from $-\infty$ to s_0 on both sides, we obtain

$$c(s_1, s_0, X) = \int_{-\infty}^{s_0} \dot{c}_v(s_1, t_0, X) p_0(t_0, X) dt_0, \quad (\text{A.13})$$

where we use $c(s_1, -\infty, X) = 0$. Plugging (A.13) into (A.12) yields

$$\begin{aligned} m_{c01}(s_0, X) &= \int \left\{ \int_{-\infty}^{s_0} \dot{c}_v(s_1, t_0, X) p_0(t_0, X) dt_0 \right\} p_1(s_1, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 \\ &= \iint 1(t_0 < s_0) \dot{c}_v(s_1, t_0, X) p_0(t_0, X) p_1(s_1, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 dt_0. \end{aligned}$$

Therefore,

$$\begin{aligned} m_{c01}(S, X) &= \iint 1(t_0 < S) \dot{c}_v(s_1, t_0, X) p_0(t_0, X) p_1(s_1, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 dt_0 \\ &= \iint 1(s_0 < S) \dot{c}_v(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 ds_0, \end{aligned}$$

where we change t_0 to s_0 in the last equality. For $m_{c01}(S, X)$, we have

$$m_{c01}(S, X) = \iint 1(S \leq s_0) \dot{c}_v(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 ds_0.$$

Therefore,

$$m_{c01}(S, X) + m_{c01}(S, X) = \iint \dot{c}_v(s_1, s_0, X) p_1(s_1, X) p_0(s_0, X) \{\mu_1(s_1, X) - \eta_1\} ds_1 ds_0,$$

which does not depend on S . As a result,

$$\begin{aligned} m_{c01}(S, X) - m_{r01}(S, X) &= m_{c01}(S, X) + m_{c01}(S, X) - E \{m_{c01}(S, X) + m_{c01}(S, X) \mid Z = 0, X\} \\ &= 0. \end{aligned}$$

Combining results on $m_{c11}(S, X)$, $m_{r11}(S, X)$, $m_{c01}(S, X)$, and $m_{r01}(S, X)$, we obtain

$$\ell_2(S, Z, X) = \frac{Z}{\pi(X)} \{ \mu_1(S, X) - E(Y \mid Z = 1, X) \}.$$

For $\ell_3(Y, S, Z, X)$, we have

$$\begin{aligned} \ell_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \frac{\int e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\} \\ &= \frac{Z}{\pi(X)} \{Y - \mu_1(S, X)\}. \end{aligned}$$

Combining these three terms gives the result in Example A.2. $\square$

A.2.7 Proof of Example A.1

Under the linear working model $f_1(s_1, s_0; \eta_1) = \eta_1^T g_1(s_1, s_0)$, we have $\dot{f}_1(s_1, s_0; \eta_1) = g_1(s_1, s_0)$ and $\ddot{f}_1(s_1, s_0; \eta_1) = 0$. We can then obtain the EIF by plugging the expressions of $\dot{f}_1(s_1, s_0; \eta_1)$ and $\ddot{f}_1(s_1, s_0; \eta_1)$ into the general EIF given in Theorem 2.3. $\square$

A.2.8 Proof of Theorem 2.4

We first introduce the following lemma to simplify the proof.

Lemma A.5 (Valid estimating equation). *$E\{D_{1,\text{eif}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1)\} = 0$ if either $(\bar{\pi}, \bar{e}) = (\pi, e)$ or $(\bar{e}, \bar{\mu}_1) = (e, \mu_1)$.*

Proof of Lemma A.5. Denote

$$\begin{aligned} \bar{v}_1(s_1, s_0, X) &= w_1(s_1, s_0) \dot{f}_1(s_1, s_0; \eta_1) \bar{e}(s_1, s_0, X) \{ \bar{\mu}_1(s_1, X) - f_1(s_1, s_0; \eta_1) \}, \\ \bar{r}_u(s_1, s_0, S, X) &= 1 - \frac{\bar{\dot{c}}_u(s_1, s_0, X)}{\bar{c}(s_1, s_0, X)} \{ 1(S \leq s_1) - \bar{F}_1(s_1, X) \}, \\ \bar{r}_v(s_1, s_0, S, X) &= 1 - \frac{\bar{\dot{c}}_v(s_1, s_0, X)}{\bar{c}(s_1, s_0, X)} \{ 1(S \leq s_0) - \bar{F}_0(s_0, X) \}, \end{aligned}$$

where $\bar{F}_z(s_z, X)$ is the probability limit of the cumulative distribution function of S_z given X for $z = 0, 1$ and $\bar{\dot{c}}_*(s_1, s_0, X_i)/\bar{c}(s_1, s_0, X_i)$ equal to $\partial \log c(u, v)/\partial *$ for $*$ = u, v evaluated at the $\bar{F}_1(s_1, X_i)$ and $\bar{F}_0(s_0, X_i)$. We have

$$E\{D_{1,\text{eif}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1)\}$$

$$= E \{ \ell_1 (X; \eta_1, \bar{e}, \bar{\mu}_1) \} + E \{ \ell_2 (S, Z, X; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) \} + E \{ \ell_3 (Y, S, Z, X; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) \} \quad (\text{A.14})$$

where $V = (Y, S, Z, X)$ and

$$\begin{aligned} \ell_1 (X; \eta_1, \bar{e}, \bar{\mu}_1) &= \iint \bar{\nu}_1 (s_1, s_0, X) ds_1 ds_0, \\ \ell_2 (S, Z, X; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) &= \frac{Z}{\bar{\pi}(X)} \left\{ \int \frac{\bar{\nu}_1 (S, s_0, X)}{\bar{p}_1 (S, X)} ds_0 - \iint \bar{\nu}_1 (s_1, s_0, X) \bar{r}_u (s_1, s_0, S, X) ds_1 ds_0 \right\} \\ &\quad + \frac{1-Z}{1-\bar{\pi}(X)} \left\{ \int \frac{\bar{\nu}_1 (s_1, S, X)}{\bar{p}_0 (S, X)} ds_1 - \iint \bar{\nu}_1 (s_1, s_0, X) \bar{r}_v (s_1, s_0, S, X) ds_1 ds_0 \right\}, \\ \ell_3 (Y, S, Z, X; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) &= \frac{Z}{\bar{\pi}(X)} \frac{\int w_1 (S, s_0) \dot{f}_1 (S, s_0; \eta_1) \bar{e} (S, s_0, X) ds_0}{\bar{p}_1 (S, X)} \{Y - \bar{\mu}_1 (S, X)\}. \end{aligned}$$

For ease of notation, we suppress the dependence of the functions on (s_1, s_0, S, X) and η_1 when no confusion arises. We have

$$\begin{aligned} \bar{\nu}_1 &= w_1 \dot{f}_1 (\bar{e} - e) (\mu_1 - f_1) + w_1 \dot{f}_1 \bar{e} (\bar{\mu}_1 - \mu_1) + w_1 \dot{f}_1 e (\mu_1 - f_1), \\ \bar{r}_u &= r_u + \left(\frac{\dot{\bar{c}}_u}{\bar{c}} - \frac{\dot{c}_u}{c} \right) \{ \bar{F}_1 - 1(S \leq s_1) \} + \frac{\dot{c}_u}{c} (\bar{F}_1 - F_1), \\ \bar{r}_v &= r_v + \left(\frac{\dot{\bar{c}}_v}{\bar{c}} - \frac{\dot{c}_v}{c} \right) \{ \bar{F}_0 - 1(S \leq s_0) \} + \frac{\dot{c}_v}{c} (\bar{F}_0 - F_0). \end{aligned} \quad (\text{A.15})$$

We then compute the three terms in (A.14) separately. For the first term, we have

$$\begin{aligned} E \{ \ell_1 (X; \eta_1, \bar{e}, \bar{\mu}_1) \} &= E \left\{ \iint \bar{\nu}_1 ds_1 ds_0 \right\} \\ &= E \left\{ \iint w_1 \dot{f}_1 (\bar{e} - e) (\mu_1 - f_1) ds_1 ds_0 \right\} \\ &\quad + E \left\{ \iint w_1 \dot{f}_1 \bar{e} (\bar{\mu}_1 - \mu_1) ds_1 ds_0 \right\}, \end{aligned} \quad (\text{A.16})$$

where the second equality follows from

$$E \left\{ \iint w_1 \dot{f}_1 e (\mu_1 - f_1) ds_1 ds_0 \right\} = 0.$$

For the second term, we have

$$\begin{aligned} E \left\{ \frac{Z}{\bar{\pi}(X)} \int \frac{\bar{\nu}_1 (S, s_0, X)}{\bar{p}_1 (S, X)} ds_0 \right\} &= E \left[E \left\{ \frac{Z}{\bar{\pi}(X)} \int \frac{\bar{\nu}_1 (S, s_0, X)}{\bar{p}_1 (S, X)} ds_0 \right\} \mid X \right] \\ &= E \left[\frac{\pi(X)}{\bar{\pi}(X)} E \left\{ \int \frac{\bar{\nu}_1 (S, s_0, X)}{\bar{p}_1 (S, X)} ds_0 \mid Z = 1, X \right\} \right] \end{aligned}$$

$$\begin{aligned}
&= E \left(\frac{\pi}{\bar{\pi}} \iint \frac{\bar{\nu}_1 p_1}{\bar{p}_1} ds_1 ds_0 \right) \\
&= E \left(\frac{\pi}{\bar{\pi}} \iint \frac{\bar{\nu}_1 (p_1 - \bar{p}_1)}{\bar{p}_1} ds_1 ds_0 \right) + E \left(\frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 ds_1 ds_0 \right)
\end{aligned}$$

and

$$\begin{aligned}
&E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 \bar{r}_u ds_1 ds_0 \right\} \\
&= E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 (\bar{r}_u - r_u) ds_1 ds_0 \right\} + E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 r_u ds_1 ds_0 \right\} \\
&= E \left[\frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 \left(\frac{\bar{\dot{c}}_u}{\bar{c}} - \frac{\dot{c}_u}{c} \right) \{ \bar{F}_1 - 1(S \leq s_1) \} ds_1 ds_0 \right] \\
&\quad + E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 (\bar{F}_1 - F_1) \frac{\dot{c}_u}{c} ds_1 ds_0 \right\} + E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 r_u ds_1 ds_0 \right\} \\
&= E \left\{ \frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 \left(\frac{\bar{\dot{c}}_u}{\bar{c}} - \frac{\dot{c}_u}{c} \right) (\bar{F}_1 - F_1) ds_1 ds_0 \right\} \\
&\quad + E \left\{ \frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 (\bar{F}_1 - F_1) \frac{\dot{c}_u}{c} ds_1 ds_0 \right\} + E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 r_u ds_1 ds_0 \right\} \\
&= E \left\{ \frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 \frac{\bar{\dot{c}}_u}{\bar{c}} (\bar{F}_1 - F_1) ds_1 ds_0 \right\} + E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 r_u ds_1 ds_0 \right\}.
\end{aligned}$$

From the law of iterated expectations,

$$E \left\{ \frac{Z}{\bar{\pi}(X)} \iint \bar{\nu}_1 r_u ds_1 ds_0 \right\} = E \left\{ \frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 E(r_u | Z = 1, X) ds_1 ds_0 \right\} = E \left(\frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 ds_1 ds_0 \right),$$

where the second equality follows from $E(r_u | Z = 1, X) = 1$. Therefore, we obtain

$$\begin{aligned}
&E \left[\frac{Z}{\bar{\pi}(X)} \left\{ \int \frac{\bar{\nu}_1(S, s_0, X)}{\bar{p}_1(S, X)} ds_0 - \iint \bar{\nu}_1(s_1, s_0, X) \bar{r}_u(s_1, s_0, S, X) ds_1 ds_0 \right\} \right] \\
&= -E \left[\frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}} (\bar{F}_1 - F_1) \right\} ds_1 ds_0 \right]. \tag{A.17}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&E \left[\frac{1 - Z}{1 - \bar{\pi}(X)} \left\{ \int \frac{\bar{\nu}_1(s_1, S, X)}{\bar{p}_0(S, X)} ds_0 - \iint \bar{\nu}_1(s_1, s_0, X) \bar{r}_v(s_1, s_0, S, X) ds_1 ds_0 \right\} \right] \\
&= -E \left[\frac{1 - \pi}{1 - \bar{\pi}} \iint \bar{\nu}_1 \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}} (\bar{F}_0 - F_0) \right\} ds_1 ds_0 \right]. \tag{A.18}
\end{aligned}$$

For the third term, we have

$$E \left[\frac{Z}{\bar{\pi}(X)} \frac{\int w_1(S, s_0) \dot{f}_1(S, s_0) \bar{e}(S, s_0, X) ds_0}{\bar{p}_1(S, X)} \{Y - \bar{\mu}_1(S, X)\} \right]$$

$$\begin{aligned}
&= E \left(\frac{\pi(X)}{\bar{\pi}(X)} E \left[\frac{\int w_1(S, s_0) \dot{f}_1(S, s_0) \bar{e}(S, s_0, X) ds_0}{\bar{p}_1(S, X)} \{Y - \bar{\mu}_1(S, X)\} \mid X, Z = 1 \right] \right) \\
&= E \left(\frac{\pi(X)}{\bar{\pi}(X)} E \left[\frac{\int w_1(S, s_0) \dot{f}_1(S, s_0) \bar{e}(S, s_0, X) ds_0}{\bar{p}_1(S, X)} \{\mu_1(S, X) - \bar{\mu}_1(S, X)\} \mid X, Z = 1 \right] \right) \\
&= E \left[\frac{\pi(X)}{\bar{\pi}(X)} \int \frac{\int w_1(s_1, s_0) \dot{f}_1(s_1, s_0) \bar{e}(s_1, s_0, X) ds_0}{\bar{p}_1(s_1, X)} \{\mu_1(s_1, X) - \bar{\mu}_1(s_1, X)\} p_1(s_1, X) ds_1 \right] \\
&= -E \left\{ \frac{\pi}{\bar{\pi}} \iint w_1 \dot{f}_1 \bar{e} (\bar{\mu}_1 - \mu_1) \frac{p_1}{\bar{p}_1} ds_1 ds_0 \right\}, \tag{A.19}
\end{aligned}$$

where the first two equalities follow from the law of iterated expectations.

Combining (A.16) to (A.19), we have

$$\begin{aligned}
&E \{D_{1,\text{eif}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1)\} \\
&= E \left\{ \iint w_1 \dot{f}_1 (\bar{e} - e) (\mu_1 - f_1) ds_1 ds_0 \right\} + E \left\{ \iint w_1 \dot{f}_1 \bar{e} (\bar{\mu}_1 - \mu_1) ds_1 ds_0 \right\} \\
&\quad - E \left[\frac{\pi}{\bar{\pi}} \iint \bar{\nu}_1 \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}} (\bar{F}_1 - F_1) \right\} ds_1 ds_0 \right] \\
&\quad - E \left[\frac{1 - \pi}{1 - \bar{\pi}} \iint \bar{\nu}_1 \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}} (\bar{F}_0 - F_0) \right\} ds_1 ds_0 \right] - E \left\{ \frac{\pi}{\bar{\pi}} \iint w_1 \dot{f}_1 \bar{e} (\bar{\mu}_1 - \mu_1) \frac{p_1}{\bar{p}_1} ds_1 ds_0 \right\}.
\end{aligned}$$

Plugging (A.15) into the above equation, we obtain

$$E \{D_{1,\text{eif}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1)\} = E \left\{ \iint w_1 \dot{f}_1 D_{\text{integrand}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) ds_1 ds_0 \right\},$$

where

$$\begin{aligned}
&D_{\text{integrand}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) \\
&= (\bar{e} - e) (\mu_1 - f_1) + \bar{e} (\bar{\mu}_1 - \mu_1) - \frac{\pi}{\bar{\pi}} \bar{e} (\bar{\mu}_1 - \mu_1) \frac{p_1}{\bar{p}_1} \\
&\quad - \frac{\pi}{\bar{\pi}} \{ \bar{e} (\mu_1 - f_1) + \bar{e} (\bar{\mu}_1 - \mu_1) \} \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}} (\bar{F}_1 - F_1) \right\} \\
&\quad - \frac{1 - \pi}{1 - \bar{\pi}} \{ \bar{e} (\mu_1 - f_1) + \bar{e} (\bar{\mu}_1 - \mu_1) \} \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}} (\bar{F}_0 - F_0) \right\} \\
&= (\bar{e} - e) (\mu_1 - f_1) + \bar{e} (\bar{\mu}_1 - \mu_1) \left(1 - \frac{\pi p_1}{\bar{\pi} \bar{p}_1} \right) \\
&\quad + \frac{\bar{\pi} - \pi}{\bar{\pi}} \{ \bar{e} (\mu_1 - f_1) + \bar{e} (\bar{\mu}_1 - \mu_1) \} \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}} (\bar{F}_1 - F_1) \right\}
\end{aligned}$$

$$\begin{aligned}
& -\{\bar{e}(\mu_1 - f_1) + \bar{e}(\bar{\mu}_1 - \mu_1)\} \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}}(\bar{F}_1 - F_1) \right\} \\
& -\frac{\bar{\pi} - \pi}{1 - \bar{\pi}} \{\bar{e}(\mu_1 - f_1) + \bar{e}(\bar{\mu}_1 - \mu_1)\} \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\} \\
& -\{\bar{e}(\mu_1 - f_1) + \bar{e}(\bar{\mu}_1 - \mu_1)\} \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\}.
\end{aligned}$$

Rearranging the terms, we obtain

$$\begin{aligned}
& D_{\text{integrand}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) \\
& = (\bar{\pi} - \pi) \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}}(\bar{F}_1 - F_1) \right\} \frac{\bar{e}(\mu_1 - f_1)}{\bar{\pi}} - (\bar{\pi} - \pi) \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\} \frac{\bar{e}(\mu_1 - f_1)}{1 - \bar{\pi}} \\
& + (\bar{\pi} - \pi) \frac{\bar{\dot{c}}_u}{\bar{c}}(\bar{F}_1 - F_1)(\bar{\mu}_1 - \mu_1) \frac{\bar{e}}{\bar{\pi}} - (\bar{\pi} - \pi) \left\{ \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\} (\bar{\mu}_1 - \mu_1) \frac{\bar{e}}{1 - \bar{\pi}} \\
& - \bar{e}(\mu_1 - f_1) \left\{ \frac{\bar{p}_1 - p_1}{\bar{p}_1} + \frac{\bar{\dot{c}}_u}{\bar{c}}(\bar{F}_1 - F_1) + \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\} \\
& - \left\{ \frac{\bar{\dot{c}}_u}{\bar{c}}(\bar{F}_1 - F_1) + \frac{\bar{p}_0 - p_0}{\bar{p}_0} + \frac{\bar{\dot{c}}_v}{\bar{c}}(\bar{F}_0 - F_0) \right\} (\bar{\mu}_1 - \mu_1) \bar{e} \\
& + (\bar{e} - e)(\mu_1 - f_1) + (\bar{\pi} - \pi)(\bar{\mu}_1 - \mu_1) \frac{\bar{e}}{\bar{\pi}}.
\end{aligned}$$

When the principal density model is correctly specified, $(\bar{p}_1, \bar{p}_0, \bar{F}_1, \bar{F}_0, \bar{e}) = (p_1, p_0, F_1, F_0, e)$, and thus

$$D_{\text{integrand}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1) = (\bar{\pi} - \pi)(\bar{\mu}_1 - \mu_1) \frac{\bar{e}}{\bar{\pi}},$$

which is equal to zero if either $\bar{\pi} = \pi$ or $\bar{\mu}_1 = \mu_1$. As a result, we have $E\{D_{1,\text{eif}}(V; \eta_1, \bar{\pi}, \bar{e}, \bar{\mu}_1)\} = 0$ if either $(\bar{\pi}, \bar{e}) = (\pi, e)$ or $(\bar{e}, \bar{\mu}_1) = (e, \mu_1)$. $\square$

Proof of Theorem 2.4(a)

Our proposed estimator $\hat{\eta}_{1,\text{eif}}$ is a Z-estimator with the system of estimating equations $D_{1,\text{eif}}$. By Lemma A.5, the Glivenko-Cantelli Theorem, and the existence and uniqueness of η_1 , consistency of the Z-estimator $\hat{\eta}_{1,\text{eif}}$ to η_1 follows from Theorem 5.9 in Vaart (2000).

Proof of Theorem 2.4(b)

From Theorem 5.31 in Vaart (2000), under the conditions stated in the theorem, the convergence rate of $\hat{\eta}_{1,\text{eif}}$ is

$$O_P\{n^{-1/2} + \text{Rem}\},$$

where $\text{Rem} = \|E\{D_{1,\text{eif}}(V; \eta, \hat{\pi}, \hat{e}, \hat{\mu}_1)\}\|$. Following similar arguments as in the proof of Lemma A.5, we have

$$E\{D_{1,\text{eif}}(V; \eta, \hat{\pi}, \hat{e}, \hat{\mu}_1)\} = E\left\{\iint w_1 \dot{f}_1 D_{\text{integrand}}(V; \eta, \hat{\pi}, \hat{e}, \hat{\mu}_1) ds_1 ds_0\right\}, \quad (\text{A.20})$$

where

$$D_{\text{integrand}}(V; \eta, \hat{\pi}, \hat{e}, \hat{\mu}_1) = (\hat{\pi} - \pi)(\hat{\mu}_1 - \mu_1) \frac{\hat{e}}{\hat{\pi}} \quad (\text{A.21})$$

$$+ (\hat{\pi} - \pi) \left[\frac{1}{\hat{\pi}} \left\{ \frac{\hat{p}_1 - p_1}{\hat{p}_1} + \frac{\hat{c}_u}{\hat{c}} (\hat{F}_1 - F_1) \right\} - \frac{1}{1 - \hat{\pi}} \left\{ \frac{\hat{p}_0 - p_0}{\hat{p}_0} + \frac{\hat{c}_v}{\hat{c}} (\hat{F}_0 - F_0) \right\} \right] \hat{e} (\mu_1 - f_1) \quad (\text{A.22})$$

$$- \left\{ \frac{\hat{c}_u}{\hat{c}} (\hat{F}_1 - F_1) + \frac{\hat{p}_0 - p_0}{\hat{p}_0} + \frac{\hat{c}_v}{\hat{c}} (\hat{F}_0 - F_0) \right\} (\hat{\mu}_1 - \mu_1) \hat{e} \quad (\text{A.23})$$

$$+ (\hat{\pi} - \pi) \frac{\hat{c}_u}{\hat{c}} (\hat{F}_1 - F_1) (\hat{\mu}_1 - \mu_1) \frac{\hat{e}}{\hat{\pi}} - (\hat{\pi} - \pi) \left\{ \frac{\hat{p}_0 - p_0}{\hat{p}_0} + \frac{\hat{c}_v}{\hat{c}} (\hat{F}_0 - F_0) \right\} (\hat{\mu}_1 - \mu_1) \frac{\hat{e}}{1 - \hat{\pi}} \quad (\text{A.24})$$

$$+ (\mu_1 - f_1) \left[(\hat{e} - e) - \hat{e} \left\{ \frac{\hat{p}_1 - p_1}{\hat{p}_1} + \frac{\hat{c}_u}{\hat{c}} (\hat{F}_1 - F_1) + \frac{\hat{p}_0 - p_0}{\hat{p}_0} + \frac{\hat{c}_v}{\hat{c}} (\hat{F}_0 - F_0) \right\} \right]. \quad (\text{A.25})$$

Plugging the expression of $D_{\text{integrand}}(V; \eta, \hat{\pi}, \hat{e}, \hat{\mu}_1)$ into (A.20), the corresponding first term

$$\left\| E \left\{ \iint w_1 \dot{f}_1 (\hat{\pi} - \pi) (\hat{\mu}_1 - \mu_1) \frac{\hat{e}}{\hat{\pi}} ds_1 ds_0 \right\} \right\| \leq C \|\hat{\pi} - \pi\|_2 \|\hat{\mu}_1 - \mu_1\|_2$$

by the Cauchy-Schwarz inequality and the fact that $w_1 \hat{e} / \hat{\pi} \|\dot{f}_1\|$ is bounded. Similarly, we can bound the terms in (A.22)–(A.24) and simplify the bound using the fact that the convergence rate of $\hat{F}_z$ is no slower than $\hat{p}_z$ for $z = 0, 1$ in both nonparametric and parametric estimation of the conditional distribution function and conditional density function (Vaart, 2000; Hall et al., 1999; Fan and Yao, 2003; Li and Racine, 2023). For (A.25), we can write the term in the square brackets as

$$\begin{aligned} & c(\hat{F}_1, \hat{F}_0) \hat{p}_1 \hat{p}_0 - c(F_1, F_0) p_1 p_0 - c(\hat{F}_1, \hat{F}_0) (\hat{p}_1 - p_1) \hat{p}_0 - \dot{c}_u(\hat{F}_1, \hat{F}_0) \hat{p}_1 \hat{p}_0 (\hat{F}_1 - F_1) \\ & - c(\hat{F}_1, \hat{F}_0) (\hat{p}_0 - p_0) \hat{p}_1 - \dot{c}_v(\hat{F}_1, \hat{F}_0) \hat{p}_1 \hat{p}_0 (\hat{F}_0 - F_0) \\ = & \hat{p}_1 \hat{p}_0 \left\{ -c(F_1, F_0) + c(\hat{F}_1, \hat{F}_0) + \dot{c}_u(\hat{F}_1, \hat{F}_0) (F_1 - \hat{F}_1) + \dot{c}_v(\hat{F}_1, \hat{F}_0) (F_0 - \hat{F}_0) \right\} \\ & + \left\{ c(F_1, F_0) - c(\hat{F}_1, \hat{F}_0) \right\} \{(\hat{p}_1 - p_1) \hat{p}_0 + (\hat{p}_0 - p_0) \hat{p}_1\} \end{aligned}$$

$$\begin{aligned}
& -c(F_1, F_0) \{p_1 p_0 + (\hat{p}_1 - p_1) \hat{p}_0 + (\hat{p}_0 - p_0) \hat{p}_1 - \hat{p}_1 \hat{p}_0\} \\
= & -\hat{p}_1 \hat{p}_0 \left\{ \frac{1}{2} \ddot{c}_{uu}(\hat{F}_1, \hat{F}_0) (\hat{F}_1 - F_1)^2 + \frac{1}{2} \ddot{c}_{vv}(\hat{F}_1, \hat{F}_0) (\hat{F}_0 - F_0)^2 \right. \\
& \left. + \ddot{c}_{uv}(\hat{F}_1, \hat{F}_0) (\hat{F}_1 - F_1) (\hat{F}_0 - F_0) \right\} \\
& + \left\{ \dot{c}_u(\hat{F}_1, \hat{F}_0) (F_1 - \hat{F}_1) + \dot{c}_v(\hat{F}_1, \hat{F}_0) (F_0 - \hat{F}_0) \right\} \{(\hat{p}_1 - p_1) \hat{p}_0 + (\hat{p}_0 - p_0) \hat{p}_1\} \\
& -c(F_1, F_0) (\hat{p}_1 - p_1) (\hat{p}_0 - p_0) \\
& + o \left\{ (\hat{F}_1 - F_1)^2 + (\hat{F}_0 - F_0)^2 \right\} \\
& + o \left\{ (\hat{F}_1 - F_1) (\hat{p}_1 - p_1) + (\hat{F}_1 - F_1) (\hat{p}_0 - p_0) + (\hat{F}_0 - F_0) (\hat{p}_1 - p_1) + (\hat{F}_0 - F_0) (\hat{p}_0 - p_0) \right\},
\end{aligned}$$

where $\ddot{c}_{uu}(u, v) = \partial^2 c(u, v) / \partial u^2$, $\ddot{c}_{vv}(u, v) = \partial^2 c(u, v) / \partial v^2$, $\ddot{c}_{uv}(u, v) = \partial^2 c(u, v) / \partial u \partial v$, and the equality follows from Taylor expansion.

Therefore, (A.25) can be bounded by $\|\hat{p}_1 - p_1\|_2 \|\hat{p}_0 - p_0\|_2$ again due to the fact that $\hat{F}_z$ has a faster convergence rate than $\hat{p}_z$ for $z = 0, 1$. Combining all of these terms gives the results in Theorem 2.4(b). $\square$

A.2.9 Proof of Theorem 2.5

By Theorem 5.31 in Vaart (2000), under the conditions stated in the theorem, we have

$$\begin{aligned}
\sqrt{n}(\hat{\eta}_{1,\text{eif}} - \eta_1) &= -C_1^{-1} \sqrt{n} E \{D_{1,\text{eif}}(V; \eta_1, \hat{\pi}, \hat{e}, \hat{\mu}_1)\} \\
&\quad - C_1^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n D_{1,\text{eif}}(V_i; \eta_1, \pi, e, \mu_1) \right\} + o_P(1 + \sqrt{n} \text{Rem}), \quad (\text{A.26})
\end{aligned}$$

where $C_1 = C(\eta_1, \bar{\xi})$ denotes the derivative in Assumption 2.5. When Assumption 2.6 is satisfied, the first term in (A.26), also called the drift term, $E \{D_{1,\text{eif}}(V; \eta_1, \hat{\pi}, \hat{e}, \hat{\mu}_1)\} = o_P(n^{-1/2})$, and the remainder term is of higher order. Thus $\sqrt{n}(\hat{\eta}_{1,\text{eif}} - \eta_1)$ reduces to

$$-C_1^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n D_{1,\text{eif}}(V_i; \eta_1, \pi, e, \mu_1) \right\},$$

and

$$\sqrt{n}(\hat{\eta}_{1,\text{eif}} - \eta_1) \xrightarrow{d} \mathcal{N}(0, C_1^{-1} E \{D_{1,\text{eif}}(V; \eta_1, \xi) D_{1,\text{eif}}(V; \eta_1, \xi)^T\} C_1^{-1}),$$

where $C_1 = C(\eta_1, \bar{\xi})$ denotes the derivative in Assumption 2.5 (Vaart, 2000).

By the fact that $E \{\varphi_1(V; \eta_1, \xi) \varphi_1(V; \eta_1, \xi)^T\} = H_1^{-1} E \{D_{1,\text{eif}}(V; \eta_1, \xi) D_{1,\text{eif}}(V; \eta_1, \xi)^T\} H_1^{-1}$, it remains to show $H_1 = C_1$. Under Assumption 2.6, $\bar{\xi} = \xi$ and

$$C_1 = \frac{\partial}{\partial \eta_1} E \{D_{1,\text{eif}}(V; \eta_1, \bar{\xi})\}$$

$$= \frac{\partial}{\partial \eta_1} E \{ \ell_1 (X; \eta_1, \bar{\xi}) \} + \frac{\partial}{\partial \eta_1} E \{ \ell_2 (S, Z, X; \eta_1, \bar{\xi}) \} + \frac{\partial}{\partial \eta_1} E \{ \ell_3 (Y, S, Z, X; \eta_1, \bar{\xi}) \}.$$

For the first term, we have

$$\begin{aligned} & \frac{\partial}{\partial \eta_1} E \{ \ell_1 (X; \eta_1, \bar{\xi}) \} \\ &= E \left[\iint w_1 \left\{ \ddot{f}_1 e(s_1, s_0, X) \mu_1(s_1, X) - \ddot{f}_1 e(s_1, s_0) f_1 - \dot{f}_1 \dot{f}_1^T e(s_1, s_0) \right\} ds_1 ds_0 \right] \\ &= \iint w_1 \ddot{f}_1 E \{ e(s_1, s_0, X) \mu_1(s_1, X) \} ds_1 ds_0 - \iint w_1 e(s_1, s_0) \left(\ddot{f}_1 f_1 - \dot{f}_1 \dot{f}_1^T \right) ds_1 ds_0 \\ &= H_1. \end{aligned}$$

For the second term, we have

$$\begin{aligned} & E \left[\frac{Z}{\pi(X)} \{ m_{11}(S, X; \eta_1) - E(m_{11}(S, X; \eta_1) \mid Z = 1, X) \} \right] \\ &= E \left[E \left\{ \frac{Z}{\pi(X)} \{ m_{11}(S, X; \eta_1) - E(m_{11}(S, X; \eta_1) \mid Z = 1, X) \} \mid X \right\} \right] \\ &= E \left[\frac{E(Z \mid X)}{\pi(X)} E \{ m_{11}(S, X; \eta_1) - E(m_{11}(S, X; \eta_1) \mid Z = 1, X) \mid X, Z = 1 \} \right] \\ &= E [E \{ m_{11}(S, X; \eta_1) - E(m_{11}(S, X; \eta_1) \mid Z = 1, X) \mid X, Z = 1 \}] \\ &= 0, \end{aligned}$$

where the first equality is by the law of iterated expectation and the second equality is by treatment ignorability. Similarly, we have

$$E \left[\frac{1 - Z}{1 - \pi(X)} \{ m_{01}(S, X; \eta_1) - E(m_{01}(S, X; \eta_1) \mid Z = 0, X) \} \right] = 0.$$

As a result, $E \{ \ell_2 (S, Z, X; \eta_1, \bar{\xi}) \} = 0$ for all η_1 and hence $\partial E \{ \ell_2 (S, Z, X; \eta_1, \bar{\xi}) \} / \partial \eta_1 = 0$.

For the third term, we have

$$\begin{aligned} & \frac{\partial}{\partial \eta_1} E \{ \ell_3 (Y, S, Z, X; \eta_1, \bar{\xi}) \} \\ &= E \left[\frac{Z}{\pi(X)} \frac{\partial \int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0 / \partial \eta_1}{p_1(S, X)} \{ Y - \mu_1(S, X) \} \right] \\ &= E \left[E \left\{ \frac{\partial \int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0 / \partial \eta_1}{p_1(S, X)} (Y - \mu_1(S, X)) \mid X, Z = 1 \right\} \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[E \left\{ \frac{\partial \int w_1(S, s_0) \dot{f}_1(S, s_0; \eta_1) e(S, s_0, X) ds_0 / \partial \eta_1}{p_1(S, X)} (E(Y | S, X, Z = 1) - \mu_1(S, X)) | X, Z = 1 \right\} \right] \\
&= 0,
\end{aligned}$$

where the second and the third equalities follow from the law of total expectation. Therefore,

$$C_1 = \frac{\partial}{\partial \eta_1} E \{ \ell_1(X; \eta_1, \bar{\xi}) \} + \frac{\partial}{\partial \eta_1} E \{ \ell_2(S, Z, X; \eta_1, \bar{\xi}) \} + \frac{\partial}{\partial \eta_1} E \{ \ell_3(Y, S, Z, X; \eta_1, \bar{\xi}) \} = H_1.$$

As a result, $\sqrt{n}(\hat{\eta}_{\text{eif}} - \eta) \xrightarrow{d} \mathcal{N}(0, E \{ \varphi_1(V; \eta_1, \xi) \varphi_1(V; \eta_1, \xi)^T \})$. $\square$

A.2.10 Proof of Theorem A.2

To find the $\varphi_\tau(V)$ satisfies

$$\dot{\eta}_{\tau, \theta}|_{\theta=0} \equiv \left. \frac{\partial \eta_{\tau, \theta}}{\partial \theta} \right|_{\theta=0} = \mathbb{E} \{ \varphi_\tau(V) \mathfrak{s}(V) \},$$

we calculate the derivative $\dot{\eta}_{\tau, \theta}|_{\theta=0}$. Taking the first order derivative with respect to η_τ on the objective function in (A.2).

$$0 = E \left[w_\tau(S_1, S_0) \dot{f}_\tau(S_1, S_0; \eta_\tau) \{ \tau(S_1, S_0) - f_\tau(S_1, S_0; \eta_\tau) \} \right]. \quad (\text{A.27})$$

By the fact that (A.27) holds for any parametric submodel, we have

$$\begin{aligned}
0 &= E_\theta \left[w_\tau(S_1, S_0) \dot{f}_\tau(S_1, S_0; \eta_{\tau, \theta}) \{ \tau_\theta(S_1, S_0) - f_\tau(S_1, S_0; \eta_{\tau, \theta}) \} \right] \\
&= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_{\tau, \theta}) \{ \tau_\theta(s_1, s_0) - f_\tau(s_1, s_0; \eta_{\tau, \theta}) \} e_\theta(s_1, s_0) ds_1 ds_0 \\
&= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_{\tau, \theta}) \left\{ \frac{\lambda_\theta(s_1, s_0)}{e_\theta(s_1, s_0)} - f_\tau(s_1, s_0; \eta_{\tau, \theta}) \right\} e_\theta(s_1, s_0) ds_1 ds_0 \\
&= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_{\tau, \theta}) \{ \lambda_\theta(s_1, s_0) - e_\theta(s_1, s_0) f_\tau(s_1, s_0; \eta_{\tau, \theta}) \} ds_1 ds_0,
\end{aligned}$$

where $\lambda_\theta(s_1, s_0) = E_\theta[e_\theta(s_1, s_0, X) \{ \mu_{1, \theta}(s_1, X) - \mu_{0, \theta}(s_0, X) \}]$. Taking the first order derivative with respect to θ and evaluating at $\theta = 0$, we have

$$\begin{aligned}
0 &= \left. \frac{\partial}{\partial \theta} \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_{\tau, \theta}) \{ \lambda_\theta(s_1, s_0) - e_\theta(s_1, s_0) f_\tau(s_1, s_0; \eta_{\tau, \theta}) \} ds_1 ds_0 \right|_{\theta=0} \\
&= \iint w_\tau(s_1, s_0) \ddot{f}_\tau(s_1, s_0; \eta_\tau) \dot{\eta}_{\tau, \theta}|_{\theta=0} \{ \lambda(s_1, s_0) - e(s_1, s_0) f_\tau(s_1, s_0; \eta_\tau) \} ds_1 ds_0 \\
&\quad + \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \left\{ \left. \dot{\lambda}_\theta(s_1, s_0) \right|_{\theta=0} - \left. \dot{e}_\theta(s_1, s_0) \right|_{\theta=0} f_\tau(s_1, s_0; \eta_\tau) \right\} ds_1 ds_0
\end{aligned}$$

$$- \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \dot{f}_\tau(s_1, s_0; \eta_\tau)^\top e(s_1, s_0) \dot{\eta}_{\tau, \theta}|_{\theta=0} ds_1 ds_0.$$

Therefore, we can write

$$\dot{\eta}_{\tau, \theta}|_{\theta=0} = H_\tau^{-1} \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \left\{ \dot{\lambda}_\theta(s_1, s_0) \Big|_{\theta=0} - \dot{e}_\theta(s_1, s_0) \Big|_{\theta=0} f_\tau(s_1, s_0; \eta_\tau) \right\} ds_1 ds_0, \quad (\text{A.28})$$

where

$$\begin{aligned} H_\tau &= - \iint w_\tau(s_1, s_0) \ddot{f}_\tau(s_1, s_0; \eta_\tau) E[e(s_1, s_0, X) \{\mu_1(s_1, X) - \mu_0(s_0, X)\}] ds_1 ds_0 \\ &\quad + \iint w_\tau(s_1, s_0) \left\{ \ddot{f}_\tau(s_1, s_0; \eta_\tau) f_\tau(s_1, s_0; \eta_\tau) + \dot{f}_\tau(s_1, s_0; \eta_\tau) \dot{f}_\tau(s_1, s_0; \eta_\tau)^\top \right\} e(s_1, s_0) ds_1 ds_0. \end{aligned}$$

Again, we calculate the two parts separately.

Part I. By Lemma A.4, we have

$$\iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \dot{\lambda}_\theta(s_1, s_0) \Big|_{\theta=0} ds_1 ds_0 = B_1 + B_2 + B_3$$

where

$$\begin{aligned} B_1 &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) E[e(s_1, s_0, X) \{\mu_1(s_1, X) - \mu_0(s_0, X)\} \mathfrak{s}(X)] ds_1 ds_0 \\ &= E\{h_1(X) \mathfrak{s}(X)\}, \\ B_2 &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) E[\dot{e}_\theta(s_1, s_0, X) \Big|_{\theta=0} \{\mu_1(s_1, X) - \mu_0(s_0, X)\}] ds_1 ds_0 \\ &= E\{h_2(S, Z, X) \mathfrak{s}(S | Z, X)\}, \\ B_3 &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) E[e(s_1, s_0, X) \{\dot{\mu}_{1, \theta}(s_1, X) \Big|_{\theta=0} - \dot{\mu}_{0, \theta}(s_0, X) \Big|_{\theta=0}\}] ds_1 ds_0 \\ &= E\{h_3(Y, S, Z, X) \mathfrak{s}(Y | S, Z, X)\} \end{aligned}$$

with a little abuse of notation. Through similar arguments, we have

$$\begin{aligned} h_1(X) &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) e(s_1, s_0, X) \{\mu_1(s_1, X) - \mu_0(s_0, X)\} ds_1 ds_0 \\ &\quad - \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) f_\tau(s_1, s_0; \eta_\tau) e(s_1, s_0) ds_1 ds_0, \\ h_2(S, Z, X) &= \frac{Z}{\pi(X)} [\tilde{\mu}_{11}(S, X) - E\{\tilde{\mu}_{11}(S, X) | Z = 1, X\}] \\ &\quad + \frac{1-Z}{1-\pi(X)} [\tilde{\mu}_{01}(S, X) - E\{\tilde{\mu}_{01}(S, X) | Z = 0, X\}] \end{aligned}$$

$$\begin{aligned}
& -\frac{Z}{\pi(X)} [\tilde{\mu}_{10}(S, X) - E\{\tilde{\mu}_{10}(S, X) \mid Z = 1, X\}] \\
& -\frac{1-Z}{1-\pi(X)} [\tilde{\mu}_{00}(S, X) - E\{\tilde{\mu}_{00}(S, X) \mid Z = 0, X\}], \\
h_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \frac{\int_{S_0} w_\tau(S, s_0) \dot{f}_\tau(S, s_0; \eta_\tau) e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\} \\
& -\frac{1-Z}{1-\pi(X)} \frac{\int_{S_1} w_\tau(s_1, S) \dot{f}_\tau(s_1, S; \eta_\tau) e(s_1, S, X) ds_1}{p_0(S, X)} \{Y - \mu_0(S, X)\},
\end{aligned}$$

where $\tilde{\mu}_*$ are the same as previously defined for $* = 11, 01, 10, 00$ except that the weighting function is changed to $w_\tau(s_1, s_0)$ and the working model is changed to $f_\tau(s_1, s_0; \eta_\tau)$.

Part II. The second part of $\dot{\eta}_\theta|_{\theta=0}$ as in Equation (A.28) is exactly the same as in the proof of Theorem 2.3, so we omit the details here.

Summary of the proof. We have

$$\begin{aligned}
H_\tau \dot{\eta}_\theta|_{\theta=0} &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \dot{\lambda}_\theta(s_1, s_0) \Big|_{\theta=0} ds_1 ds_0 \\
& - \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) f_\tau(s_1, s_0; \eta_\tau) \dot{e}_\theta(s_1, s_0) \Big|_{\theta=0} ds_1 ds_0 \\
&= E\{(h_1(X) - h_4(X)) \mathfrak{s}(X)\} \\
& + E\{(h_2(S, Z, X) - h_5(S, Z, X)) \mathfrak{s}(S \mid Z, X)\} \\
& + E\{h_3(Y, S, Z, X) \mathfrak{s}(Y \mid S, Z, X)\}
\end{aligned}$$

where

$$\begin{aligned}
H_\tau^{-1}(h_1(X) - h_4(X)) &\in \mathcal{H}_1, \\
H_\tau^{-1}(h_2(S, Z, X) - h_5(S, Z, X)) &\in \mathcal{H}_3, \\
H_\tau^{-1}h_3(Y, S, Z, X) &\in \mathcal{H}_4.
\end{aligned}$$

By similar arguments, the EIF is

$$\varphi_\tau(V) = H_\tau^{-1} \{t_1(X) + t_2(S, Z, X) + t_3(Y, S, Z, X)\},$$

where

$$\begin{aligned}
t_1(X) &= \iint w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) e(s_1, s_0, X) \\
& \quad \{ \mu_1(s_1, X) - \mu_0(s_0, X) - f_\tau(s_1, s_0; \eta_\tau) \} ds_1 ds_0, \\
t_2(S, Z, X) &= \frac{Z}{\pi(X)} [\tilde{\mu}_{11}(S, X) - E\{\tilde{\mu}_{11}(S, X) \mid Z = 1, X\}]
\end{aligned}$$

$$\begin{aligned}
& + \frac{1-Z}{1-\pi(X)} [\tilde{\mu}_{01}(S, X) - E\{\tilde{\mu}_{01}(S, X) \mid Z=0, X\}] \\
& - \frac{Z}{\pi(X)} [\tilde{\mu}_{10}(S, X) - E\{\tilde{\mu}_{10}(S, X) \mid Z=1, X\}] \\
& - \frac{1-Z}{1-\pi(X)} [\tilde{\mu}_{00}(S, X) - E\{\tilde{\mu}_{00}(S, X) \mid Z=0, X\}] \\
& - \frac{Z}{\pi(X)} [\tilde{f}_1(S, X) - E\{\tilde{f}_1(S, X) \mid Z=1, X\}] \\
& - \frac{1-Z}{1-\pi(X)} [\tilde{f}_0(S, X) - E\{\tilde{f}_0(S, X) \mid Z=0, X\}], \\
t_3(Y, S, Z, X) &= \frac{Z}{\pi(X)} \frac{\int_{S_0} w_\tau(S, s_0) \dot{f}_\tau(S, s_0; \eta_\tau) e(S, s_0, X) ds_0}{p_1(S, X)} \{Y - \mu_1(S, X)\} \\
& - \frac{1-Z}{1-\pi(X)} \frac{\int_{S_1} w_\tau(s_1, S) \dot{f}_\tau(s_1, S; \eta_\tau) e(s_1, S, X) ds_1}{p_0(S, X)} \{Y - \mu_0(S, X)\},
\end{aligned}$$

where $\tilde{\mu}_{*z}$ denotes the weighted integration of the conditional outcome model $\mu_z(s_z, X)$,

$$\begin{aligned}
\tilde{\mu}_{11}(s_1, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_1(s_1, X)} \mu_1(s_1, X) ds_0 \\
&+ \iint 1(t_1 \geq s_1) w_\tau(t_1, s_0) \dot{f}_\tau(t_1, s_0; \eta_\tau) \dot{c}_u(t_1, s_0, X) p_1(t_1, X) p_0(s_0, X) \mu_1(t_1, X) dt_1 ds_0, \\
\tilde{\mu}_{01}(s_0, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_0(s_0, X)} \mu_1(s_1, X) ds_1 \\
&+ \iint 1(t_0 \geq s_0) w_\tau(s_1, t_0) \dot{f}_\tau(s_1, t_0; \eta_\tau) \dot{c}_v(s_1, t_0, X) p_1(s_1, X) p_0(t_0, X) \mu_1(s_1, X) ds_1 dt_0, \\
\tilde{\mu}_{10}(s_1, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_1(s_1, X)} \mu_0(s_0, X) ds_0 \\
&+ \iint 1(t_1 \geq s_1) w_\tau(t_1, s_0) \dot{f}_\tau(t_1, s_0; \eta_\tau) \dot{c}_u(t_1, s_0, X) p_1(t_1, X) p_0(s_0, X) \mu_0(s_0, X) dt_1 ds_0, \\
\tilde{\mu}_{00}(s_0, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_0(s_0, X)} \mu_0(s_0, X) ds_1 \\
&+ \iint 1(t_0 \geq s_0) w_\tau(s_1, t_0) \dot{f}_\tau(s_1, t_0; \eta_\tau) \dot{c}_v(s_1, t_0, X) p_1(s_1, X) p_0(t_0, X) \mu_0(t_0, X) ds_1 dt_0, \\
\tilde{f}_1(s_1, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) f_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_1(s_1, X)} ds_0 \\
&+ \iint 1(t_1 \geq s_1) w_\tau(t_1, s_0) \dot{f}_\tau(t_1, s_0; \eta_\tau) f_\tau(t_1, s_0; \eta_\tau) \dot{c}_u(t_1, s_0, X) p_1(t_1, X) p_0(s_0, X) dt_1 ds_0, \\
\tilde{f}_0(s_0, X) &= \int w_\tau(s_1, s_0) \dot{f}_\tau(s_1, s_0; \eta_\tau) f_\tau(s_1, s_0; \eta_\tau) \frac{e(s_1, s_0, X)}{p_0(s_0, X)} ds_1
\end{aligned}$$

$$+ \iint 1(t_0 \geq s_0) w_\tau(s_1, t_0) \dot{f}_\tau(s_1, t_0; \eta_\tau) f_\tau(s_1, t_0; \eta_\tau) \dot{c}_v(s_1, t_0, X) p_1(s_1, X) p_0(t_0, X) ds_1 dt_0.$$

Using similar notation and derivation as in the proof of Theorem 2.3 gives the results stated in the theorem. $\square$

Appendix B

Appendix to Chapter 3

Section B.1 provides additional technical results on details of the augmented estimator $\hat{\tau}^{\text{aug}}$, comparing the four proposed estimators, another martingale-corrected weighting estimator discussed in Section 3.4.3 of the main paper, and a cross-fitted debiased machine learning implementation procedure of the EIF-based estimator.

Section B.2 provides additional data analysis results, including classification of all IEs in the two immunology trials, exploratory data analysis for baseline covariates and IEs, additional results on robustness check, estimation results targeting causal risk ratio and odds ratios, and remarks on computational time of our proposed estimators.

Section B.3 reports detailed results from Monte Carlo simulations that assess the finite-sample performance of our proposed estimators.

Section B.4 provides proofs of all theorems and propositions.

B.1 Additional technical results

B.1.1 Additional details on $\hat{\tau}^{\text{aug}}$

In this subsection, we discuss additional motivations and remarks about the proposed estimator $\hat{\tau}^{\text{aug}}$. First, by combining the two identification formulas (3.2) and (3.3), we derive the following augmented identification formula:

$$\begin{aligned} \tau = & E \left\{ \frac{AY1(T \wedge C > k)}{e(X)G_1(k | X)} - \frac{A - e(X)}{e(X)} \mu_1(X)S_1(k | X) \right\} \\ & - E \left[\frac{(1 - A)Y1(T \wedge C > k)}{\{1 - e(X)\} G_0(k | X)} - \frac{e(X) - A}{1 - e(X)} \mu_0(X)S_0(k | X) \right]. \end{aligned} \quad (\text{B.1})$$

Equation (B.1) can be interpreted both as a modified form of the outcome regression identification formula (3.2), and as an augmented form of the weighting identification formula in (3.3). Under the true model at the population level, the correction terms have mean 0, and thus (B.1) holds by construction. However, it becomes meaningful in the presence of possible model misspecification, where the augmentation plays a crucial role in improving robustness.

Based on (B.1), we construct the estimator $\hat{\tau}^{\text{aug}}$ that augments the weighting estimator $\hat{\tau}^{\text{ipw}}$ using the estimated outcome models. The improvement in its robustness compared with $\hat{\tau}^{\text{ipw}}$ is summarized in Proposition 3.1, which shows that the consistency of $\hat{\tau}^{\text{aug}}$ depends on correctly specifying $G_a(k | X)$ for $a = 0, 1$, while it remains robust to misspecification of the propensity score $e(X)$. To build intuition, from the perspective of semiparametric efficiency theory, Equation (B.1) can be viewed as a projection of the weighting identification formula (3.3) onto the nuisance tangent space of the propensity score model $A | X$. However, because it is not further projected onto the nuisance tangent space of the $G_a(k | X)$ model, the resulting estimator does not retain robustness to misspecification of $G_a(k | X)$. We will further address the issue in Section 3.4, where we introduce an alternative estimator that is robust to the misspecification of both models.

The augmented estimator $\hat{\tau}^{\text{aug}}$ dominates the weighting estimator $\hat{\tau}^{\text{ipw}}$ in terms of robustness. The consistency of $\hat{\tau}^{\text{aug}}$ is contingent on less restrictive requirements on nuisance parameter estimation in the sense that $\hat{\tau}^{\text{aug}}$ is consistent whenever $\hat{\tau}^{\text{ipw}}$ is. Consistency of $\hat{\tau}^{\text{aug}}$ requires the correct specification of the survival function for effect-uninformative IE $G_a(k | X)$ for $a = 0, 1$. Given that $G_a(k | X)$ is correctly modeled, $\hat{\tau}^{\text{aug}}$ is doubly robust because it is consistent if either the propensity score model is correct, or the outcome regression and the survival model for effect-informative IEs are both correct. There is no clear dominance between $\hat{\tau}^{\text{aug}}$ and $\hat{\tau}^{\text{out}}$ as they require consistent estimation of different sets of nuisance parameters. For all three estimators $\hat{\tau}^{\text{out}}$, $\hat{\tau}^{\text{ipw}}$, and $\hat{\tau}^{\text{aug}}$, we can construct variance estimators using nonparametric bootstrap.

B.1.2 Pairwise differences between the estimators

As discussed in Section 3.4.3, the observed differences among the four estimators provide insight into the presence of nuisance model misspecification. To illustrate the idea, consider the treated arm as an example. Let $\hat{\mu}_1^\dagger$ denote the treated arm counterpart in $\hat{\tau}^\dagger$ for $\dagger \in \{\text{out}, \text{ipw}, \text{aug}, \text{EIF}\}$, and let $\xrightarrow{P}$ denote convergence in probability. The probability limits of the pairwise differences between these estimators are given by

$$\begin{aligned}\hat{\mu}_1^{\text{aug}} - \hat{\mu}_1^{\text{ipw}} &\xrightarrow{P} E \left\{ \frac{e^*(X) - e(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \right\}, \\ \hat{\mu}_1^{\text{aug}} - \hat{\mu}_1^{\text{out}} &\xrightarrow{P} E \left[\frac{e(X)}{e^*(X)} \{ \mu_1^*(X) S_1^*(k | X) - \mu_1(X) S_1(k | X) \} \right],\end{aligned}$$

$$\hat{\mu}_1^{\text{eif}} - \hat{\mu}_1^{\text{aug}} \xrightarrow{P} E \left[\frac{e(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \int_0^k \frac{S_1(t | X) G_1(t | X)}{S_1^*(t | X) G_1^*(t | X)} \{d\Lambda_1(t | X) - d\Lambda_1^*(t | X)\} \right].$$

These probability limits indicate the impact of nuisance model misspecification on the pairwise comparisons. For example, if $G_1(t | X)$ is correctly specified, the expected difference between $\hat{\mu}_1^{\text{eif}}$ and $\hat{\mu}_1^{\text{aug}}$ should be small. Similar logic applies to other comparisons. Therefore, when implementing all four estimators in practice, the observed difference among them may serve as an informal diagnostic tool to detect possible nuisance model misspecification.

B.1.3 The martingale-corrected weighting estimator

In this subsection, we present the explicit form of the martingale-corrected weighting estimator introduced in Section 3.4.3. Drawing from the proof of Theorem 3.2 in Section B.4.3, if we do not project onto the propensity score nuisance tangent space in Step 2, the following identification formula for μ_1 holds:

$$\mu_1 = E \left[\frac{A}{e(X)} \left\{ \frac{Y1(T \wedge C > k)}{G_1(k | X)} + \mu_1(X) S_1(k | X) \int_0^{\tilde{T}} \frac{dM_{G_1}(t)}{S_1(t | X) G_1(t | X)} \right\} \right],$$

which motivates the following estimator for μ_1 :

$$\hat{\mu}_1^{\text{mc-ipw}} = n^{-1} \sum_{i=1}^n \frac{A_i Y_i 1(T_i \wedge C_i > k)}{\hat{e}(X_i) \hat{G}_1(k | X_i)} + n^{-1} \sum_{i=1}^n \frac{A_i}{\hat{e}(X_i)} \hat{\mu}_1(X_i) \hat{S}_1(k | X_i) \int_0^{\tilde{T}_i} \frac{dM_{\hat{G}_1}(t)}{\hat{S}_1(t | X_i) \hat{G}_1(t | X_i)}.$$

The control counterpart, $\hat{\mu}_0^{\text{mc-ipw}}$, is defined analogously, and the estimator for τ is given by $\hat{\tau}^{\text{mc-ipw}} = \hat{\mu}_1^{\text{mc-ipw}} - \hat{\mu}_0^{\text{mc-ipw}}$.

Under Assumptions 3.1 and 3.2, and assuming correct specification of $e(X)$, the estimator $\hat{\tau}^{\text{mc-ipw}}$ is consistent for τ if either $G_a(t | X)$ is correct, or both $\mu_a(X)$ and $S_a(t | X)$ are correct for $t \leq k$ and $a = 0, 1$. Although the martingale correction enhances robustness to misspecification of $G_a(t | X)$, $\hat{\tau}^{\text{mc-ipw}}$ is still less attractive than $\hat{\tau}^{\text{eif}}$: it does not achieve the efficiency bound and is not robust to misspecification of $e(X)$. Moreover, it does not inherit the implementation simplicity of $\hat{\tau}^{\text{aug}}$. In particular, it requires estimating the full survival curves for both the effect-uninformative IEs $G_a(t | X)$ and effect-informative IEs $S_a(t | X)$ over $t \leq k$, making it computationally demanding without offering clear practical advantages.

B.1.4 A cross-fitted debiased machine learning implementation of the EIF-based estimator

The EIF-based estimator in Section 3.4 admits a natural cross-fitted implementation in the spirit of double/debiased machine learning (Chernozhukov et al., 2018). We briefly outline

one possible algorithm here.

- Step 1: Randomly partition the units into L approximately equal-sized folds $\mathcal{I}_1, \dots, \mathcal{I}_L$.
- Step 2: For each fold $\ell = 1, \dots, L$, use the units outside $\mathcal{I}_\ell$ to estimate the nuisance functions $e(X), \mu_a(X), S_a(t | X), G_a(t | X)$ for $a=0,1$. Denote the resulting estimators by $\hat{e}^{(-\ell)}, \hat{\mu}_a^{(-\ell)}, \hat{S}_a^{(-\ell)}$, and $\hat{G}_a^{(-\ell)}$.
- Step 3: For each unit $i \in \mathcal{I}_\ell$, evaluate the EIF using the nuisance estimators trained on the complement of fold ℓ . That is, compute the cross-fitted version of the EIF contribution $\hat{D}_{\tau,i}^{(-\ell)}$ by substituting $\hat{e}^{(-\ell)}, \hat{\mu}_a^{(-\ell)}, \hat{S}_a^{(-\ell)}$, and $\hat{G}_a^{(-\ell)}$ into the expression of $D_\tau(O_i)$ in Theorem 3.2.
- Step 4: Define the cross-fitted EIF-based estimator by averaging the fold-specific scores: $\hat{\tau}^{\text{cf}} = \frac{1}{n} \sum_{i=1}^n \hat{\psi}_i$, where $\hat{\psi}_i$ denotes the plug-in estimating expression corresponding to unit i with nuisance functions estimated out-of-fold.
- Step 5: An asymptotic variance estimator can be constructed using bootstrap or the empirical variance of the cross-fitted EIF contributions, $n^{-1} \sum_{i=1}^n (\hat{D}_{\tau,i}^{(-\ell(i))})^2$, where $\ell(i)$ is the fold containing unit i . Under regularity conditions, this yields a valid inference in the usual way.

This cross-fitted construction follows the same general principle as standard double/debiased machine learning: nuisance functions are estimated on auxiliary subsamples, while the orthogonal score is evaluated on held-out units. In our setting, the main additional practical burden is that the nuisance functions include not only the outcome regression and propensity score, but also the survival functions for the effect-informative and effect-uninformative IEs.

B.2 Additional data analysis results

B.2.1 Detailed classification of IEs in the immunology trials

In the two immunology trials, IE categories were classified through manual review of the site-level documentation. We use the terms *effect-informative* and *effect-uninformative* to emphasize whether the event is judged to carry information about the patient's underlying treatment response. Table B.1 provides the complete list of IE categories observed in the two trials, together with their frequencies, assigned classifications, and the rationale used in our review.

Guided by the recommendations in the PHUSE white paper on estimand implementation and data standards (PHUSE Working Group, 2025), we use the following general rule for

classification. We classified an IE as effect-uninformative only when the documentation suggested that it arose from external or administrative reasons plausibly unrelated to the biological effect of treatment, such as relocation, pandemic-related disruption, or sponsor-driven site closure. We classified an IE as effect-informative when it was directly related to adverse events, lack of efficacy, patient health status, treatment management, or when the available documentation was too limited to justify an effect-uninformative classification. In ambiguous cases, we classified the IE as effect-informative to avoid overestimating arm-specific mean outcomes. Several categories required case-by-case review because the high-level label alone was not always sufficiently informative. In particular, categories such as personal or family reason, withdrawal by subject, protocol deviation, and unable to contact were reviewed using the site-level records. This review was used to determine whether the event was plausibly external to the patient’s treatment response or whether it was more credible to regard the event as carrying information about efficacy, tolerability, or patient status.

Table B.1: Classification of IEs in the two immunology trials

IE type	Count	Proportion	Informative?	Rationale
Adverse event	135	31.5%	Y	Classified as effect-informative because adverse events are plausibly related to treatment tolerability and the patient’s underlying treatment response.
Others	84	19.6%	Y	For discontinuations with insufficiently detailed documentation, we classified the event as effect-informative because the absence of clear information makes an effect-uninformative interpretation difficult to justify. We adopted this default to avoid overstating arm-specific mean outcomes.
Lack of efficacy	64	14.9%	Y	Classified as effect-informative because lack of efficacy is directly informative about the patient’s response to treatment.

Table B.1 continued

IE type	Count	Proportion	Informative?	Rationale
Personal or family reason	32	7.5%	N	Review of the detailed site-level comments suggested that these discontinuations were driven by external personal circumstances rather than health status or treatment response. Examples included relocation of residence and other family-related reasons unrelated to disease progression or adverse events.
Protocol deviation	26	6.1%	Y	Classified as effect-informative because these deviations were typically linked to treatment management or patient status in ways that could reflect the patient's response trajectory. Examples included changes in corticosteroid use or prohibited dose changes in concomitant medication.
Due to epidemic or pandemic	23	5.4%	N	Classified as effect-uninformative because these discontinuations were driven by external disruptions, such as pandemic-related restrictions, that are more plausibly unrelated to treatment efficacy.
Withdrawal by subject	22	5.1%	Y	Classified as effect-informative because many such withdrawals were associated with concerns about side effects, unwillingness to continue treatment, or other reasons potentially related to treatment experience. When documentation was limited, we again defaulted to effect-informative classification.
Pregnancy	15	3.5%	N	Classified as effect-uninformative when the documentation suggested that treatment discontinuation was driven by pregnancy planning, suspected pregnancy, or pregnancy itself, rather than by the patient's disease course or treatment response.

Table B.1 continued

IE type	Count	Proportion	Informative?	Rationale
Discontinued: unable to contact	11	2.6%	Y	Classified as effect-informative because the available documentation did not allow us to rule out discontinuation related to efficacy, adverse events, or other treatment-related concerns.
Death	9	2.1%	Y	Classified as effect-informative because death is directly informative about the patient's health status and precludes meaningful continuation of treatment.
Prohibited medication use	5	1.2%	Y	Classified as effect-informative because the need for prohibited medication may reflect inadequate disease control, toxicity management, or other clinically relevant aspects of treatment response.
Physician decision	2	0.5%	Y	Classified as effect-informative because physician-directed discontinuation was based on patient behavior or clinical status and therefore plausibly related to treatment experience or outcome risk.
Site terminated by sponsor	1	0.2%	N	Classified as effect-uninformative because sponsor-driven site closure is an external administrative disruption unrelated to the patient's biological treatment response.

B.2.2 Exploratory data analysis: baseline covariates and IE summaries

Table B.2 reports summary statistics of baseline covariates and IEs for all six arm-by-trial combinations. Figure B.1 shows the Kaplan–Meier-based cumulative incidence of effect-informative and effect-uninformative IEs by treatment arm and trial.

Table B.2: Baseline characteristics and IE summary by treatment arm and trial. PGA denotes Physician’s Global Assessment score (scale 0–100); SLEDAI denotes the SLE Disease Activity Index. IE counts and percentages are relative to the full arm n . Median time to IE (in weeks) and interquartile range [Q1, Q3] are reported for participants who experienced the respective IE type.

	Trial 1			Trial 2		
	Placebo	2mg	4mg	Placebo	2mg	4mg
n	253	255	252	256	261	258
<i>Baseline covariates</i>						
PGA, mean (SD)	59.9 (15.0)	60.5 (14.4)	59.2 (15.8)	60.0 (14.6)	61.0 (12.5)	58.7 (14.6)
SLEDAI, mean (SD)	10.0 (3.2)	10.3 (3.4)	10.0 (2.9)	10.1 (3.2)	10.1 (3.4)	10.1 (3.0)
<i>Region, n (%)</i>						
Asia	31 (12.3)	32 (12.5)	31 (12.3)	31 (12.1)	33 (12.6)	31 (12.0)
C/S America & Mexico	74 (29.2)	73 (28.6)	72 (28.6)	59 (23.0)	58 (22.2)	57 (22.1)
Europe	64 (25.3)	64 (25.1)	63 (25.0)	55 (21.5)	57 (21.8)	57 (22.1)
North America	47 (18.6)	49 (19.2)	50 (19.8)	50 (19.5)	50 (19.2)	49 (19.0)
Rest of World	37 (14.6)	37 (14.5)	36 (14.3)	61 (23.8)	63 (24.1)	64 (24.8)
<i>Intercurrent events</i>						
Effect-informative, n (%)	69 (27.3)	65 (25.5)	46 (18.3)	53 (20.7)	55 (21.1)	70 (27.1)
median (weeks) [Q1, Q3]	20 [8, 31]	23 [12, 32]	17 [12, 29]	21 [11, 40]	16 [11, 32]	20 [11, 33]
Effect-uninformative, n (%)	10 (4.0)	7 (2.7)	21 (8.3)	10 (3.9)	11 (4.2)	12 (4.7)
median (weeks) [Q1, Q3]	23 [7, 32]	18 [11, 28]	20 [11, 31]	25 [15, 36]	24 [8, 41]	22 [8, 36]
No IE, n (%)	174 (68.8)	183 (71.8)	185 (73.4)	193 (75.4)	195 (74.7)	176 (68.2)

B.2.3 Robustness check: accelerated failure time model for nuisance estimation

Table B.3 replicates Table 3.2 with all four Cox proportional hazards nuisance models replaced by log-normal accelerated failure time (AFT) models. All other components (propensity score logistic regression, outcome logistic regression, and martingale integral computation) are identical to the main analysis. Standard errors and confidence intervals are based on 500 bootstrap replicates.

B.2.4 Robustness check: alternative IE classifications

As a further robustness check on the IE classification, we tested three alternative classification rules for borderline discontinuation categories. The primary analysis treats “Withdrawal by Subject” and “Discontinued: Unable to Contact” as effect-informative, and “Pregnancy” as effect-uninformative. The three robustness check scenarios are: (S1) both “Withdrawal by Subject” and “Discontinued: Unable to Contact” reclassified as effect-uninformative; (S2) “Withdrawal by Subject” reclassified as effect-uninformative; (S3) “Pregnancy” reclassified as effect-informative. Table B.4 reports the treatment effect estimates for the primary SRI4 outcome under the three alternative IE classification rules. Each scenario was analyzed using Cox proportional hazards nuisance models, with 200 nonparametric bootstrap repli-

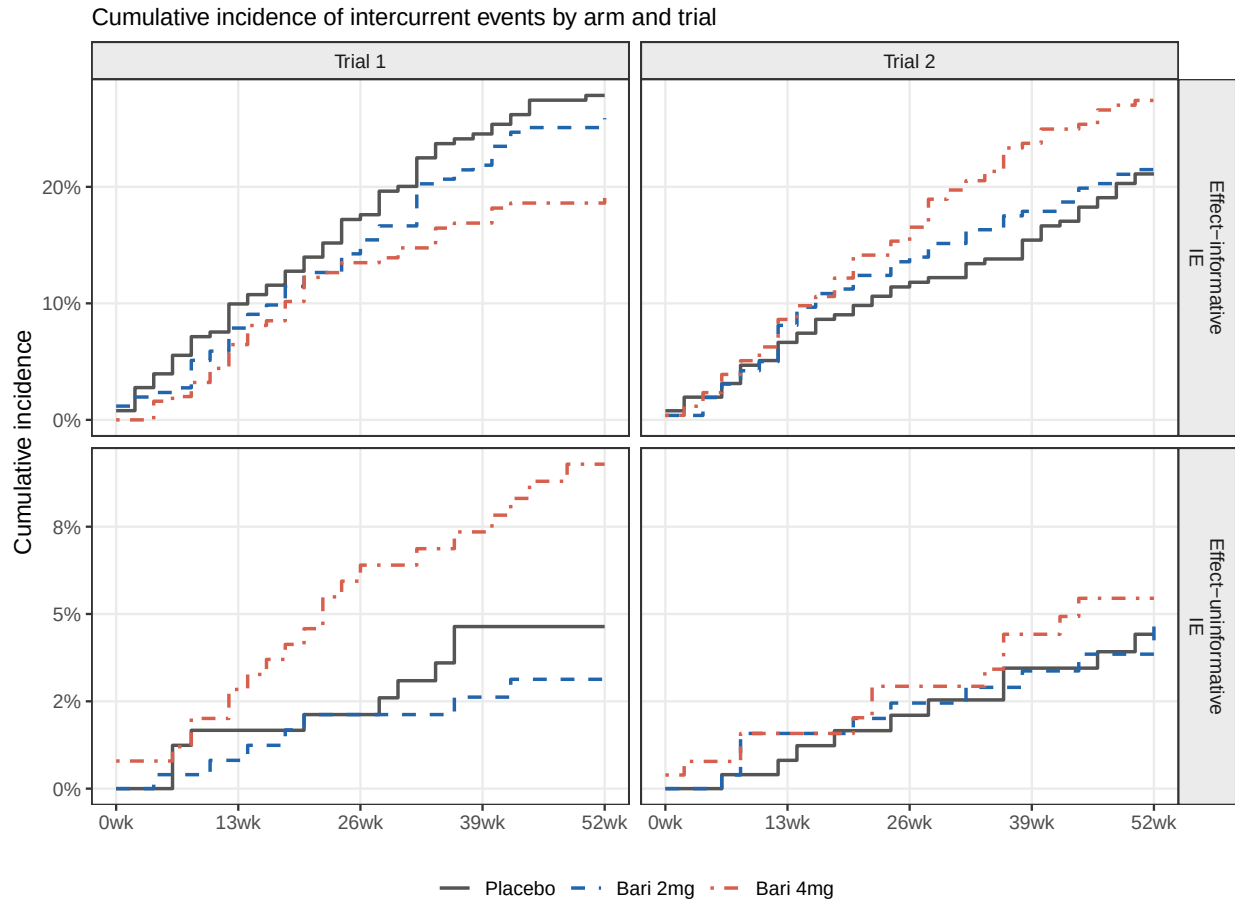


Figure B.1: Kaplan–Meier-based cumulative incidence of effect-informative IEs (top row) and effect-uninformative IEs (bottom row) by treatment arm, for Trial 1 (left column) and Trial 2 (right column). The x -axis shows weeks from treatment initiation and the y -axis shows cumulative incidence.

cates. All four estimators are shown alongside the bootstrap standard error, p -value, and 95% confidence interval, for the EIF estimator.

B.2.5 Results on causal risk ratio and odds ratio

Table B.5 reports the causal risk ratio (RR) and odds ratio (OR) for the primary SRI4 outcome under the proposed hybrid estimand. Both are estimated using the EIF-based estimator of the arm-specific mean composite outcomes $\hat{\mu}_1$ and $\hat{\mu}_0$. Confidence intervals for RR and OR are constructed on the log scale using the normal approximation and then exponentiated.

Table B.3: Treatment effect of Baricitinib on the primary outcome SRI4 under the proposed hybrid estimand, using log-normal AFT models in place of Cox proportional hazards regression for all survival nuisance functions. The first four columns report results for Trial 1 and the last four columns for Trial 2. Within each trial, the first four rows report the point estimate, bootstrap standard error, p -value, and lower and upper 95% confidence limits for each estimator of the 2mg treatment arm, and the last four rows report those of the 4mg treatment arm. For each trial and each treatment arm, we report the results based on all four estimators.

		Trial 1				Trial 2			
		$\hat{\tau}_{\text{out}}$	$\hat{\tau}_{\text{ipw}}$	$\hat{\tau}_{\text{aug}}$	$\hat{\tau}_{\text{EIF}}$	$\hat{\tau}_{\text{out}}$	$\hat{\tau}_{\text{ipw}}$	$\hat{\tau}_{\text{aug}}$	$\hat{\tau}_{\text{EIF}}$
2mg	point	0.032	0.028	0.025	0.026	0.017	0.019	0.023	0.022
	se	0.042	0.049	0.049	0.042	0.042	0.069	0.069	0.042
	p -value	0.449	0.571	0.602	0.537	0.691	0.780	0.741	0.608
	CI lower	-0.050	-0.068	-0.070	-0.057	-0.066	-0.116	-0.112	-0.061
	CI upper	0.114	0.124	0.121	0.109	0.099	0.155	0.158	0.104
4mg	point	0.108	0.119	0.114	0.113	-0.005	-0.004	-0.004	-0.002
	se	0.046	0.046	0.046	0.046	0.041	0.045	0.045	0.042
	p -value	0.019	0.010	0.014	0.013	0.911	0.933	0.938	0.969
	CI lower	0.018	0.029	0.023	0.023	-0.086	-0.092	-0.092	-0.083
	CI upper	0.198	0.209	0.204	0.203	0.076	0.084	0.085	0.080

B.2.6 Computational time

Table B.6 reports the time for a single full estimator evaluation on each of the four pairwise comparison datasets, measured on a MacBook Air with Apple M4 processor. The full procedure includes propensity score logistic regression, four Cox proportional hazards survival models (two for effect-informative IE, two for effect-uninformative IE), six outcome logistic regression models (one per outcome per treatment arm), and the discrete-time martingale integral required by the EIF-based estimator. The estimated bootstrap time is the single-run time times 500 replicates.

B.3 Simulation

In this section, we conduct Monte Carlo simulations to study the finite sample performance of our proposed estimators.

Table B.4: IE classification robustness check for the primary outcome SRI4. Rows within each scenario panel correspond to the four pairwise comparisons. Columns show all four point estimators $\hat{\tau}^{\text{out}}$, $\hat{\tau}^{\text{ipw}}$, $\hat{\tau}^{\text{aug}}$, and $\hat{\tau}^{\text{EIF}}$, together with the bootstrap standard error, 95% confidence interval, and p -value for $\hat{\tau}^{\text{EIF}}$.

	Treatment arm	$\hat{\tau}^{\text{out}}$	$\hat{\tau}^{\text{ipw}}$	$\hat{\tau}^{\text{aug}}$	$\hat{\tau}^{\text{EIF}}$	se	95% CI	p -value
S1	Trial 1, 2mg	0.031	0.028	0.026	0.026	0.045	[−0.062, 0.113]	0.565
	Trial 1, 4mg	0.107	0.112	0.107	0.105	0.047	[0.012, 0.198]	0.027
	Trial 2, 2mg	0.015	0.013	0.017	0.016	0.045	[−0.071, 0.104]	0.712
	Trial 2, 4mg	0.004	0.004	0.004	0.005	0.043	[−0.079, 0.089]	0.908
S2	Trial 1, 2mg	0.028	0.026	0.024	0.023	0.044	[−0.064, 0.109]	0.608
	Trial 1, 4mg	0.103	0.109	0.104	0.101	0.047	[0.009, 0.194]	0.031
	Trial 2, 2mg	0.018	0.017	0.021	0.020	0.045	[−0.068, 0.108]	0.656
	Trial 2, 4mg	0.003	0.004	0.004	0.005	0.043	[−0.080, 0.090]	0.911
S3	Trial 1, 2mg	0.031	0.030	0.028	0.028	0.044	[−0.059, 0.115]	0.525
	Trial 1, 4mg	0.105	0.111	0.106	0.105	0.046	[0.014, 0.195]	0.023
	Trial 2, 2mg	0.013	0.012	0.016	0.015	0.044	[−0.071, 0.102]	0.727
	Trial 2, 4mg	−0.006	−0.005	−0.005	−0.005	0.044	[−0.090, 0.081]	0.910

Table B.5: Causal risk ratio and odds ratio for the primary outcome SRI4 under the proposed hybrid estimand. $\hat{\tau}$, $\hat{\text{RR}}$, and $\hat{\text{OR}}$ denote the estimated original estimand τ (risk difference), risk ratio, and odds ratio, respectively. Confidence intervals for RR and OR are constructed on the log scale and exponentiated.

Treatment arm	$\hat{\tau}$ (95 % CI)	$\hat{\text{RR}}$ (95 % CI)	$\hat{\text{OR}}$ (95 % CI)
Trial 1, 2mg	0.026 [−0.057, 0.110]	1.06 [0.88, 1.27]	1.11 [0.79, 1.56]
Trial 1, 4mg	0.113 [0.023, 0.203]	1.25 [1.04, 1.49]	1.58 [1.09, 2.27]
Trial 2, 2mg	0.022 [−0.061, 0.104]	1.05 [0.88, 1.25]	1.09 [0.78, 1.52]
Trial 2, 4mg	−0.002 [−0.084, 0.080]	1.00 [0.83, 1.19]	0.99 [0.71, 1.39]

B.3.1 Data generating processes

We first generate the covariates $X = (X_1, X_2, X_3)^\top \in \mathbb{R}^3$ from three independent standard Gaussian distributions and denote $\tilde{X}_j = \{(X_j + 2)^2 - 1\} / \sqrt{12}$ for $j = 1, 2, 3$. Next, we generate the treatment assignment following $A \mid X \sim \text{Bern}(\{e(X)\})$ with $e(X)$ being the propensity score model, and generate the potential outcomes following $Y(a, c = \infty) \mid X \sim \mathcal{N}(\mu_a(X), \sigma_a^2)$ with $\mu_a(X)$ being the outcome model for $a = 0, 1$. We then generate the potential values of the effect-informative IE survival time $T(a)$ from the distribution with a survival function

$$S_a(t \mid X) = \exp[-0.002t^{1.2} \exp\{\gamma_a(X)\}]$$

Table B.6: Wall-clock time for a single full-estimator evaluation and the estimated time for 500 nonparametric bootstrap replicates. Sample size n includes both the treatment and placebo arms for each pairwise comparison.

Treatment arm	n	Single run (sec)	Bootstrap (min)
Trial 1, 2mg	508	0.32	2.7
Trial 1, 4mg	505	0.31	2.6
Trial 2, 2mg	517	0.41	3.4
Trial 2, 4mg	514	0.34	2.8

and the potential values of the effect-uninformative IE survival time $C(a)$ from the distributions with a survival function

$$G_a(t \mid X) = \exp[\varrho_a(t) \exp\{\delta_a(X)\}],$$

for $a = 0, 1$. For each of the nuisance parameters, we consider two scenarios when it will be correctly modeled and misspecified, therefore, we generate data following two different choices of each. We summarize the detailed data-generating choices in Table B.7.

Table B.7: Model choices for the nuisance parameters

	Correctly specified	Misspecified
$\text{logit}\{e(X)\}$	$(X_1 + X_2 + X_3)/5$	$1(X_1 \geq 0)\{\exp(\tilde{X}_2) - X_2(1 + \tilde{X}_3)\} - \exp(\tilde{X}_2)$
$\mu_1(X)$	$2(X_1 + X_2 + X_3)$	$1(X_1 \geq 0)\{X_2 + \exp(X_2)\tilde{X}_3 - \tilde{X}_2\} + \tilde{X}_2$
$\mu_0(X)$	$X_1 + X_2 + X_3$	$-\tilde{X}_1 - 1(X_1 > 0.5)\tilde{X}_2 + 1(X_1 < -0.5)X_2^2 \log(X_3 + 1)$
σ_a	$0.1(a + 1)$	1
$\gamma_1(X)$	$0.1(X_1 + 2X_2 - 2X_3)$	$0.1(X_1^2 X_2 - X_2 - 1) + X_2 \log(10X_3^2)$
$\gamma_0(X)$	$0.1(X_1 - 2X_2 + 2X_3)$	$0.01(-\tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3)$
$\delta_1(X)$	0	$0.1(X_1^2 X_2 - X_2 - 1) + X_2 \log(10X_3^2)$
$\delta_0(X)$	0	0
$\varrho_a(t)$	$-0.01t^{1.2}$	$-0.01at^{1.2} + 0.6 * 0.01^{1/1.2}(1 - a)t$

We consider five data-generating regimes: all four models are correctly specified, misspecified e and correct (G, μ, S) , misspecified (e, G) and correct (μ, S) , misspecified (μ, S) and correct (e, G) , and all four models are misspecified. Let $Y = AY(1, c = \infty) + (1 - A)Y(0, c = \infty)$, $T = AT(1) + (1 - A)T(0)$, and $C = AC(1) + (1 - A)C(0)$. Generate the observed event time as $T \wedge C \wedge k$, the observed event type indicator, and the observed outcome Y if $T \wedge C > k$.

B.3.2 Simulation results

We compare the finite sample performance of the outcome estimator $\hat{\tau}^{\text{out}}$, the weighting estimator $\hat{\tau}^{\text{ipw}}$, the augmented weighting estimator $\hat{\tau}^{\text{aug}}$, and the EIF estimator $\hat{\tau}^{\text{EIF}}$ based

on 1000 Monte Carlo samples with a sample size of $n = 1000$. When fitting the nuisance models, we use logistic regression of A on X to estimate the propensity score model, the linear regression of Y on X on the subsample $A = a, T \wedge C > k$ to estimate the outcome model, and the Cox proportional hazard regression of T and C on X on the subsample $A = a$ to estimate the effect-informative IE survival model and the censoring model, respectively, for $a = 0, 1$. We use X in all regression models, therefore, when the true data-generating processes involve non-linear functions of X , the fitted nuisance models suffer from misspecification. We evaluate the performance of the four estimators with reported bias, standard deviation, and coverage probability using a nonparametric bootstrap with 200 bootstrap samples in Table B.8.

Table B.8: Finite sample performance of the four estimators. For each estimator, we report the finite-sample bias, standard deviation (SD), and the coverage rate (CR) of a 95% confidence interval, which is constructed using a nonparametric bootstrap with 200 bootstrap iterations. Each row corresponds to a different data-generating regime.

	$\hat{\tau}^{\text{out}}$			$\hat{\tau}^{\text{ipw}}$			$\hat{\tau}^{\text{aug}}$			$\hat{\tau}^{\text{EIF}}$		
	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR	Bias	SD	CR
all correct	0.003	0.100	0.989	0.005	0.154	0.977	0.004	0.152	0.980	0.004	0.110	0.982
e wrong	-0.002	0.134	0.978	-0.338	0.527	0.916	-0.003	0.479	0.961	0.000	0.234	0.979
(e, G) wrong	-0.002	0.130	0.985	-0.194	1.591	0.831	0.141	1.559	0.972	-0.013	0.354	0.969
(μ, S) wrong	-0.085	0.170	0.936	-0.009	0.186	0.974	-0.009	0.186	0.974	-0.012	0.176	0.972
all wrong	0.230	0.233	0.837	0.660	5.799	0.898	0.850	5.782	0.967	0.573	3.022	0.919

Consistent with our theoretical results, all estimators have a finite sample bias close to zero when all models are correctly specified. When e is misspecified and (G, μ, S) are correct, the weighting estimator $\hat{\tau}^{\text{ipw}}$ is inconsistent while the other three estimators have small finite sample bias. Under the regime when e and G are misspecified, both $\hat{\tau}^{\text{ipw}}$ and $\hat{\tau}^{\text{aug}}$ perform poorly as they have larger finite sample biases, $\hat{\tau}^{\text{out}}$ and $\hat{\tau}^{\text{EIF}}$ show small finite sample biases as expected. When (μ, S) are misspecified, the outcome estimator $\hat{\tau}^{\text{out}}$ has a relatively large bias, while all other three estimators have near-zero biases. All estimators have non-negligible finite sample bias when all nuisance models are misspecified. $\hat{\tau}^{\text{EIF}}$ is most robust to model misspecifications across all different regimes.

$\hat{\tau}^{\text{EIF}}$ shows a relatively smaller standard deviation compared to $\hat{\tau}^{\text{ipw}}$ and $\hat{\tau}^{\text{aug}}$ in the regimes where all estimators are consistent. The outcome estimator $\hat{\tau}^{\text{out}}$ always has the smallest standard deviation when (μ, S) is correctly specified. Both $\hat{\tau}^{\text{ipw}}$ and $\hat{\tau}^{\text{aug}}$ have larger standard deviations across all regimes in which they are consistent. Furthermore, confidence intervals constructed using the nonparametric bootstrap yield valid coverage rates for the corresponding consistent estimators in their respective regimes.

Next, we evaluate two ad-hoc methods commonly used in clinical trials with IEs, demonstrating through simulations that they are biased in estimating the clinically relevant causal parameter. The first method is the non-responder rule method that assigns an outcome

as 0 whenever an IE occurs. It is equivalent to using the composite outcome strategy that treats all IEs as effect-informative, leading to a composite outcome equal to 0. The second method naively applies a hypothetical strategy to all IEs, assuming they are independent of potential outcomes conditional on observed covariates. We generate data under the “ μ_S_wrong ” regime, ensuring correct specification of both the propensity score $e(X)$ and the effect-uninformative IE survival model $G_a(t | X)$ for $a = 0, 1$. Based on 1000 Monte Carlo simulations with a sample size of $n = 1000$, the inverse probability-weighting estimator using the non-responder rule method exhibits a finite sample bias of -1.195 with a standard deviation of 0.077 , and that using the hypothetical strategy for all IEs has a bias of 1.188 with a standard deviation of 0.388 . Both ad-hoc methods yield inconsistent estimators.

B.4 Proofs

B.4.1 Proof of Theorem 3.1

First, equation (3.2) holds because its left-hand side

$$\begin{aligned}
 & E\{Y(1, c = \infty)1(T(1) > k)\} \\
 &= E[E\{Y(1, c = \infty)1(T(1) > k) | X\}] \\
 &= E[E\{Y(1, c = \infty)1(T > k) | X, A = 1\}] \\
 &= E\{E(Y(1, c = \infty) | T > k, X, A = 1)\text{pr}(T > k | X, A = 1)\} \\
 &= E\{E(Y(1, c = \infty) | C > k, T > k, X, A = 1)\text{pr}(T > k | X, A = 1)\} \\
 &= E\{E(Y | C > k, T > k, X, A = 1)\text{pr}(T > k | X, A = 1)\},
 \end{aligned}$$

which is equal to the right-hand side of (3.2), where the first equality is by the law of iterated expectations, the second equality is by the randomization assumption 3.1, and the fourth equality is by the censoring at random assumption 3.2.

Next, equation (3.3) holds because its right-hand side equals

$$\begin{aligned}
 & E \left[E \left\{ \frac{AY1(T \wedge C > k)}{e(X)\text{pr}(C > k | X, A = 1)} | X \right\} \right] \\
 &= E \left(\frac{1}{e(X)\text{pr}(C > k | X, A = 1)} E[AY(1, c = \infty)1\{T(1) \wedge C(1) > k\} | X] \right) \\
 &= E \left(\frac{1}{e(X)\text{pr}(C > k | X, A = 1)} E(A | X) E[Y(1, c = \infty)1\{T(1) > k\}1\{C(1) > k\} | X] \right) \\
 &= E \left(\frac{1}{\text{pr}(C > k | X, A = 1)} E[1\{C(1) > k\} | X] E[Y(1, c = \infty)1\{T(1) > k\} | X] \right) \\
 &= E[E\{Y(1, c = \infty)1(T(1) > k) | X\}],
 \end{aligned}$$

which is equal to the left-hand side of (3.3), where the first equality is by the law of iterated expectations, the third equality is by the randomization assumption 3.1, the fourth equality is by the censoring at random assumption 3.2, and the last equality is again by the law of iterated expectations. $\square$

B.4.2 Proof of Proposition 3.1

We prove that under the correct specification of the censoring model, $\hat{\tau}^{\text{aug}}$ is doubly robust in the sense that it is consistent for τ if either $e^* = e$ or $(\mu_1^* = \mu_1, S_1^* = S_1)$. We have

$$\begin{aligned} & E \left\{ \frac{AY1(T \wedge C > k)}{e^*(X)G_1(k | X)} - \frac{A - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \right\} \\ = & E \left[\frac{E\{AY1(T \wedge C > k) | X\}}{e^*(X)G_1(k | X)} - \frac{e(X) - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \right] \\ = & E \left\{ \frac{e(X)}{e^*(X)} \mu_1(X) S_1(k | X) - \frac{e(X) - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \right\}, \end{aligned} \quad (\text{B.2})$$

where the first equality follows from the law of integrated expectations and the second equality follows from similar derivations as in the proof of Theorem 3.1. Now observe that the final expression of (B.2) is equal to μ_1 if either $e^* = e$ or $(\mu_1^* = \mu_1, S_1^* = S_1)$. Therefore, we have

$$E \left\{ \frac{AY1(T \wedge C > k)}{e^*(X)G_1(k | X)} - \frac{A - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) - \mu_1 \right\} = 0$$

if either $e^* = e$ or $(\mu_1^* = \mu_1, S_1^* = S_1)$ holds. A parallel argument applies to the control counterpart. Therefore, combining the treated and control components, we conclude that $\hat{\tau}^{\text{aug}}$ is consistent for τ if either the propensity score model is correct, or the outcome and survival models for the effect-informative IE are both correct. This establishes the double robustness of $\hat{\tau}^{\text{aug}}$. $\square$

B.4.3 Proof of Theorem 3.2

We follow the semiparametric theory in Bickel et al., 1993 to derive the EIF for μ_1 . As discussed in Section 3.4, we have two levels of coarsening of the full data $(X, Y^{\text{cp}}(1), Y^{\text{cp}}(0))$, one is because the effect-uninformative IE time $C(a)$ is censoring $T(a) \wedge k$, and the other is due to the treatment assignment. Following the steps in Hubbard et al., 2000, we first consider the case when every observation is assigned to the treatment group, i.e., when there is no missingness generated by the treatment assignment, and then project the derived EIF onto the nuisance tangent space of the propensity score to get the final form of EIF for μ_1 .

Step 1. With full data $(X, Y^{\text{cp}}(1), Y^{\text{cp}}(0))$, the estimating equation for μ_1 is $D_{\text{full}} = Y^{\text{cp}}(1) - \mu_1$. Since $Y^{\text{cp}}(1)$ is only observable if $C(1) > T(1) \wedge k$, the full data is not available even if all corresponding potential outcomes under treatment $A = 1$ are observed, and the observed data is

$$O(1) = (X, \Delta(1), \tilde{T}(1), \Delta(1)Y^{\text{cp}}(1)) \sim P_1,$$

where $\Delta(1) = 1(C(1) > T(1) \wedge k)$ is the missing indicator with $\Delta(1) = 1$ if $Y^{\text{cp}}(1)$ is observed and $\Delta(1) = 0$ otherwise, and $\tilde{T}(1) = T(1) \wedge C(1) \wedge k$ is the observed event time with $\tilde{T}(1) = T(1) \wedge k$ if $\Delta(1) = 1$, $Y^{\text{cp}}(1)$ is observed, and $\tilde{T}(1) = C(1)$ if $\Delta(1) = 0$, $Y^{\text{cp}}(1)$ is missing. An identification formula for μ_1 is

$$\mu_1 = E \left\{ \frac{\Delta(1)Y^{\text{cp}}(1)}{G_1(\tilde{T}(1) | X)} \right\},$$

where $G_1(t | X) = \text{pr}(C(1) > t | X)$ is the probability of not censoring up to time t conditioning on the covariates. We can write the IPCW estimating equation as

$$D_{\text{IPCW}} = \frac{\Delta(1)\{Y^{\text{cp}}(1) - \mu_1\}}{G_1(\tilde{T}(1) | X)},$$

and have $E(D_{\text{IPCW}}) = 0$ by the law of iterated expectations and the censoring at random assumption.

Next, we follow the steps in Laan and Rubin, 2007 to derive the EIF for μ_1 when the observed data is $O(1) = (X, \Delta(1), \tilde{T}(1), \Delta(1)Y^{\text{cp}}(1))$. For ease of notation, we omit the subscripts, superscripts, or numbers in parenthesis (1) that indicate the potential outcomes under the treatment $A = 1$ in the following derivation, so the dependence on the treatment arm is implicit. For ease of notation, denote $T^1 = T \wedge k$.

Let $T(P)$ denote the tangent space which is the whole Hilbert space $L_0^2(P)$ since our model is nonparametric. The tangent space can be written as a direct sum of three components $T = T_X \oplus T_F \oplus T_{\text{CAR}}$ with

$$\begin{aligned} T_X &= \{h(X) \in L_0^2(P) : E\{h(X)\} = 0\}, \\ T_F &= \{h(O) \in L_0^2(P) : E\{h(O) | C, X\} = 0\}, \\ T_{\text{CAR}} &= \{h(O) \in L_0^2(P) : E\{h(O) | Y^{\text{cp}}, T^1, X\} = 0\}, \end{aligned}$$

where T_X , T_F , and T_{CAR} are orthogonal to each other due to the censoring at random assumption and the factorization of the observed data. These tangent spaces are generated from scores of submodels that perturb the marginal distribution of X , the conditional distribution of $Y^1 | X$, and the conditional censoring probability $C | X$, respectively. Due to the orthogonality, we have the decomposition of any $h(O) \in L_0^2(P)$ as

$$h(O) = \Pi\{h | T(P)\}$$

$$= \Pi\{h \mid T_X(P)\} + \Pi\{h \mid T_F(P)\} + \Pi\{h \mid T_{\text{CAR}}(P)\}.$$

Let $D(O)$ denote the efficient influence function for μ_1 . By the results in Chapter 1.4 of Laan and Robins, 2003b, we have: (1) $D(O)$ should be orthogonal to the nuisance tangent space T_{CAR} , thus $\Pi\{D(O) \mid T_{\text{CAR}}(P)\} = 0$; (2) $D(O)$ can be rewritten as

$$D = D_{\text{IPCW}} - \Pi\{D_{\text{IPCW}} \mid T_{\text{CAR}}(P)\},$$

thus

$$\begin{aligned} \Pi\{D \mid T_F(P)\} &= \Pi\{D_{\text{IPCW}} \mid T_F(P)\} - \Pi\{\Pi\{D_{\text{IPCW}} \mid T_{\text{CAR}}(P)\} \mid T_F(P)\} \\ &= \Pi\{D_{\text{IPCW}} \mid T_F(P)\} \end{aligned}$$

where the last equality is by the fact that T_F and T_{CAR} are orthogonal to each other, and similarly, $\Pi\{D \mid T_X(P)\} = \Pi\{D_{\text{IPCW}} \mid T_X(P)\}$.

These projections provide us with two different ways to compute the EIF D : (1) directly compute the projections $\Pi\{D_{\text{IPCW}} \mid T_X(P)\}$ and $\Pi\{D_{\text{IPCW}} \mid T_F(P)\}$ and sum them up; (2) compute the projection $\Pi\{D_{\text{IPCW}} \mid T_{\text{CAR}}(P)\}$ and subtract it from D_{IPCW} . In the classic survival outcome problem, these two methods are symmetric since T and C are censoring each other thus the projections on $T_F(P)$ and $T_{\text{CAR}}(P)$ are very similar. However, due to the complication generated by Y in our setting, the second approach is easier, since the tangent space $T_F(P)$ is hard to compute.

We next compute $\Pi\{D_{\text{IPCW}} \mid T_{\text{CAR}}(P)\}$. By Theorem 1.1 in Laan and Robins, 2003b,

$$T_{\text{CAR}}(P) = \overline{\left\{ \int H(t, \mathcal{F}(t)) dM_G(t) \text{ for all functions } H(t, \mathcal{F}(t)) \right\}} \cap L_0^2(P),$$

and the projection of a function $h(O)$ onto $T_{\text{CAR}}(P)$ is

$$\Pi\{h(O) \mid T_{\text{CAR}}(P)\} = \int_0^{\tilde{T}} \{E(h(O) \mid dA(t) = 1, \mathcal{F}(t)) - E(h(O) \mid dA(t) = 0, \mathcal{F}(t))\} dM_G(t),$$

where $A(t) = 1(C \leq t)$ is the indicator of whether censoring happens up until time t (define $C = \infty$ if $C > T^1$ so that C is always observed), $\mathcal{F}(t) = (\bar{A}(t-), X)$ is the history observed up to time t , and $dM_G(t) = 1(C \in dt, \Delta = 0) - 1(\tilde{T} \geq t)d\Lambda(t \mid X)$ is the Doob–Meyer martingale of the counting process of censoring C . $E(D_{\text{IPCW}} \mid dA(t) = 1, \mathcal{F}(t)) = 0$ by the definition of D_{IPCW} , thus we only need to compute $E(D_{\text{IPCW}} \mid dA(t) = 0, \mathcal{F}(t))$, which plus μ_1 is equal to

$$\begin{aligned} & E \left\{ \frac{\Delta Y^{\text{cp}}}{G(\tilde{T} \mid X)} \mid dA(t) = 0, \mathcal{F}(t) \right\} \\ &= E \left\{ \frac{\Delta Y^{\text{cp}}}{G(\tilde{T} \mid X)} \mid T^1 \geq t, C \geq t, X \right\} \end{aligned}$$

$$\begin{aligned}
&= E \left\{ \frac{\Delta Y^{\text{cp}}}{G(\tilde{T} | X)} \mid T^1 \geq t, X \right\} / G(t | X) \\
&= E \left\{ \frac{Y^{\text{cp}}}{G(T \wedge k | X)} \text{pr}(C > T \wedge k \mid T^1 \geq t, T, Y^1, X) \mid T^1 \geq t, X \right\} / G(t | X) \\
&= E \left\{ \frac{Y^{\text{cp}}}{G(T \wedge k | X)} \text{pr}(C > T \wedge k \mid X) \mid T^1 \geq t, X \right\} / G(t | X) \\
&= E \{ Y^{\text{cp}} \mid T^1 \geq t, X \} / G(t | X).
\end{aligned}$$

By the fact that the integration is over $t : t < \tilde{T}$, we have $t < T \wedge k \wedge C$ thus $t > k$. The conditional expectation

$$\begin{aligned}
E \{ Y^{\text{cp}} \mid T^1 \geq t, X \} &= E \{ Y^{\text{cp}} \mid T^1 \geq t, T^1 = T, X \} \text{pr}(T^1 = T \mid T^1 \geq t, X) \\
&\quad + E \{ Y^{\text{cp}} \mid T^1 \geq t, T^1 = k, X \} \text{pr}(T^1 = k \mid T^1 \geq t, X) \\
&= E(Y \mid T > k \geq t, X) \text{pr}(T > k \mid T^1 \geq t, X) \\
&= E(Y \mid T > k, X) \text{pr}(T > k \mid T \wedge k \geq t, X) \\
&= E(Y \mid T > k, X) \text{pr}(T > k \mid T \geq t, X) \\
&= \frac{E(Y 1(T > k) \mid X) \text{pr}(T > k \mid X)}{\text{pr}(T > k \mid X) \text{pr}(T \geq t \mid X)} \\
&= \frac{E(Y 1(T > k) \mid X)}{\text{pr}(T \geq t \mid X)}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\Pi\{D_{\text{IPCW}} \mid T_{\text{CAR}}(P)\} + \mu_1 &= - \int_0^{\tilde{T}} \frac{E(Y 1(T > k) \mid X)}{\text{pr}(T \geq t \mid X)} \frac{dM_G(t)}{G(t \mid X)} \\
&= -E(Y 1(T > k) \mid X) \int_0^{\tilde{T}} \frac{dM_G(t)}{S(t \mid X)G(t \mid X)},
\end{aligned}$$

and thus the EIF assuming $O(1)$ is the observed data is

$$D = \frac{\Delta Y^{\text{cp}}}{G(\tilde{T} | X)} + E(Y^{\text{cp}} \mid X) \int_0^{\tilde{T}} \frac{dM_G(t)}{S(t \mid X)G(t \mid X)} - \mu_1 \quad (\text{B.3})$$

Step 2. Next, we follow steps in Section 3 of Hubbard et al., 2000 and compute the EIF when the real observed data is $O = (X, A, \Delta, \tilde{T}, \Delta Y^{\text{cp}})$. We need to construct a weighting estimating equation and then subtract its projection onto the nuisance tangent space of the propensity score to get the final form of the EIF. To be clear on the distinction between potential outcomes and the observed values, we add back the dependence on the treatment assignment in (B.3) and write it as

$$D(1) = \frac{\Delta(1)Y^{\text{cp}}(1)}{G_1(\tilde{T}(1) \mid X)} + E(Y^{\text{cp}}(1) \mid X) \int_0^{\tilde{T}(1)} \frac{dM_{G_1}(t)}{S_1(t \mid X)G_1(t \mid X)} - \mu_1.$$

A valid weighting estimating equation is

$$D_{\text{IPW}} = \frac{A}{e(X)} \left\{ \frac{\Delta Y^{\text{cp}}}{G_1(\tilde{T} | X)} + E(Y^{\text{cp}} | X, A = 1) \int_0^{\tilde{T}} \frac{dM_{G_1}(t)}{S_1(t | X)G_1(t | X)} \right\} - \mu_1.$$

Further, project this onto the nuisance tangent space of the propensity score, the projection is

$$\begin{aligned} \Pi\{D_{\text{IPW}} | T_{\text{pscore}}\} &= E\{D_{\text{IPW}} | A, X\} - E\{D_{\text{IPW}} | X\} \\ &= \frac{A - e(X)}{e(X)} E\left\{ \frac{\Delta Y^{\text{cp}}}{G_1(\tilde{T} | X)} | X, A = 1 \right\} \\ &= \frac{A - e(X)}{e(X)} E(Y^{\text{cp}} | X, A = 1), \end{aligned}$$

where the second equality is by the fact that $E\{dM_{G_1}(t) | X, A = 1\} = E\{dM_{G_1}(t) | X\} = 0$, and the last equality is by the censoring at random assumption. Therefore, the EIF

$$\begin{aligned} D_1 &= D_{\text{IPW}} - \Pi\{D_{\text{IPW}} | T_{\text{pscore}}\} \\ &= \frac{A}{e(X)} \left\{ \frac{\Delta Y^{\text{cp}}}{G_1(\tilde{T} | X)} + E(Y1(T > k) | X, A = 1) \int_0^{\tilde{T}} \frac{dM_{G_1}(t)}{S_1(t | X)G_1(t | X)} \right\} \\ &\quad - \frac{A - e(X)}{e(X)} E(Y1(T > k) | X, A = 1) - \mu_1 \end{aligned}$$

has the given form in Theorem 3.2. □

B.4.4 Proof of Theorem 3.3

We prove the result by showing

$$\begin{aligned} \mu_1 &= E \left[\frac{A}{e^*(X)} \left\{ \frac{Y1(T \wedge C > k)}{G_1^*(k | X)} + \mu_1^*(X) S_1^*(k | X) \int_0^{\tilde{T}} \frac{dM_{G_1^*}(t)}{S_1^*(t | X)G_1^*(t | X)} \right\} \right. \\ &\quad \left. - \frac{A - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \right] \end{aligned} \quad (\text{B.4})$$

if either $\{e^*(X) = e(X), G_1^*(t | X) = G_1(t | X)\}$ or $\{\mu_1^*(X) = \mu_1(X), S_1^*(t | X) = S_1(t | X)\}$ for $t \leq k$. Define

$$\begin{aligned} \mathcal{T}_1 &= \frac{A}{e^*(X)} \frac{Y1(T \wedge C > k)}{G_1^*(k | X)} - \frac{A - e^*(X)}{e^*(X)} \mu_1^*(X) S_1^*(k | X) - \mu_1(X) S_1(k | X), \\ \mathcal{T}_2 &= \frac{A}{e^*(X)} \mu_1^*(X) S_1^*(k | X) \int_0^{\tilde{T}} \frac{dM_{G_1^*}(t)}{S_1^*(t | X)G_1^*(t | X)}. \end{aligned}$$

By the identification formula (3.2), to prove (B.4), it suffices to show $E(\mathcal{T}_1 + \mathcal{T}_2) = 0$ if either $\{e^*(X) = e(X), G_1^*(t | X) = G_1(t | X)\}$ or $\{\mu_1^*(X) = \mu_1(X), S_1^*(t | X) = S_1(t | X)\}$ for $t \leq k$.

First, we have

$$\begin{aligned} E\{dM_{G_1^*}(t) | X\} &= E\{1(C \in dt, \Delta = 0) | X\} - E\{1(\tilde{T} \geq t)d\Lambda_1^*(t | X) | X\} \\ &= S_1(t | X)G_1(t | X)d\Lambda_1(t | X) - S_1(t | X)G_1(t | X)d\Lambda_1^*(t | X) \\ &= S_1(t | X)G_1(t | X)\{d\Lambda_1(t | X) - d\Lambda_1^*(t | X)\}, \end{aligned}$$

and therefore,

$$\begin{aligned} E(\mathcal{T}_2 | X) &= \frac{e(X)}{e^*(X)}\mu_1^*(X)S_1^*(k | X)E\left\{\int_0^k \frac{dM_{G_1^*}(t)}{S_1^*(t | X)G_1^*(t | X)} | X\right\} \\ &= \frac{e(X)}{e^*(X)}\mu_1^*(X)S_1^*(k | X) \int_0^k \frac{S_1(t | X)G_1(t | X)}{S_1^*(t | X)G_1^*(t | X)}\{d\Lambda_1(t | X) - d\Lambda_1^*(t | X)\}, \end{aligned}$$

where the first equality follows from Assumption 3.1. Next, we have

$$\begin{aligned} E(\mathcal{T}_1 | X) &= \frac{e(X)G_1(k | X)}{e^*(X)G_1^*(k | X)}\mu_1(X)S_1(k | X) - \frac{e(X) - e^*(X)}{e^*(X)}\mu_1^*(X)S_1^*(k | X) - \mu_1(X)S_1(k | X) \\ &= \frac{e(X)}{e^*(X)}\left\{\frac{G_1(k | X)}{G_1^*(k | X)} - 1\right\}\mu_1^*(X)S_1^*(k | X) \\ &\quad + \left\{\frac{e(X)G_1(k | X)}{e^*(X)G_1^*(k | X)} - 1\right\}\{\mu_1(X)S_1(k | X) - \mu_1^*(X)S_1^*(k | X)\} \\ &= -\frac{e(X)}{e^*(X)}\mu_1^*(X)S_1^*(k | X) \int_0^k \frac{G_1(t | X)}{G_1^*(t | X)}\{d\Lambda_1(t | X) - d\Lambda_1^*(t | X)\} \\ &\quad + \left\{\frac{e(X)G_1(k | X)}{e^*(X)G_1^*(k | X)} - 1\right\}\{\mu_1(X)S_1(k | X) - \mu_1^*(X)S_1^*(k | X)\}, \end{aligned}$$

where the last equality follows from the Duhamel equation (Gill and Johansen, 1990; Westling et al., 2024). Combing $\mathcal{T}_1$ and $\mathcal{T}_2$, we have

$$\begin{aligned} E(\mathcal{T}_1 + \mathcal{T}_2) &= E\{E(\mathcal{T}_1 + \mathcal{T}_2 | X)\} \\ &= \frac{e(X)}{e^*(X)}\mu_1^*(X)S_1^*(k | X) \int_0^k \left\{\frac{S_1(t | X)}{S_1^*(t | X)} - 1\right\} \frac{G_1(t | X)}{G_1^*(t | X)}\{d\Lambda_1(t | X) - d\Lambda_1^*(t | X)\} \\ &\quad + \left\{\frac{e(X)G_1(k | X)}{e^*(X)G_1^*(k | X)} - 1\right\}\{\mu_1(X)S_1(k | X) - \mu_1^*(X)S_1^*(k | X)\}, \end{aligned}$$

which is equal to 0 if either $\{e^*(X) = e(X), G_1^*(t | X) = G_1(t | X)\}$ or $\{\mu_1^*(X) = \mu_1(X), S_1^*(t | X) = S_1(t | X)\}$ for $t \leq k$.

By symmetry, the part corresponding to the control arm $a = 0$ also holds, thus the double robustness in Theorem 3.3 holds. $\square$

B.4.5 Proof of Theorem 3.4

Following the von Mises expansion (Hines et al., 2022), let P and P_n denote the true and empirical distribution of the observed data, respectively, and $\hat{P}_n$ denotes an estimated P , we have

$$n^{1/2}(\hat{\tau}^{\text{eif}} - \tau) = n^{-1/2} \sum_{i=1}^n D_{\tau}(O_i) + n^{1/2}(P_n - P)\{D_{\tau}(O, \hat{P}_n) - D_{\tau}(O)\} - n^{1/2}R(P, \hat{P}_n),$$

where $D_{\tau}(O, \hat{P}_n)$ is the EIF for τ when plugging in the estimated values of the nuisance parameters and $R(P, \hat{P}_n)$ denotes the higher order remainder term. Under Assumption 3.4, the empirical process term $n^{1/2}(P_n - P)\{D_{\tau}(O, \hat{P}_n) - D_{\tau}(O)\} = o_P(1)$ (Vaart, 2000). The remainder term corresponding to the treatment arm $a = 1$ satisfies

$$\begin{aligned} R_1(P, \hat{P}_n) &= -E\{D_{\tau}(O, \hat{P}_n)\} - E\{\hat{\mu}_1(X)\hat{S}_1(k | X) - \mu_1(X)S_1(k | X)\} \\ &= -E \left[\frac{A}{\hat{e}(X)} \left\{ \frac{Y1(T \wedge C > k)}{\hat{G}_1(k | X)} + \hat{\mu}_1(X)\hat{S}_1(k | X) \int_0^k \frac{dM_{\hat{G}_1}(t)}{\hat{G}_1(t | X)\hat{S}_1(t | X)} \right\} \right. \\ &\quad \left. - \frac{A - \hat{e}(X)}{\hat{e}(X)} \hat{\mu}_1(X)\hat{S}_1(k | X) - \hat{\mu}_1(X)\hat{S}_1(k | X) \right. \\ &\quad \left. + \hat{\mu}_1(X)\hat{S}_1(k | X) - \mu_1(X)S_1(k | X) \right] \\ &= -E \left[\frac{e(X)}{\hat{e}(X)} \hat{\mu}_1(X)\hat{S}_1(k | X) \int_0^k \left\{ \frac{S_1(t | X)}{\hat{S}_1(t | X)} - 1 \right\} \frac{G_1(t | X)}{\hat{G}_1(t | X)} \{d\Lambda_1(t | X) - d\hat{\Lambda}_1(t | X)\} \right] \\ &\quad + E \left[\left\{ \frac{e(X)G_1(k | X)}{\hat{e}(X)\hat{G}_1(k | X)} - 1 \right\} \left\{ \mu_1(X)S_1(k | X) - \hat{\mu}_1(X)\hat{S}_1(k | X) \right\} \right] \\ &= E \left\{ \frac{e(X)}{\hat{e}(X)} \hat{\mu}_1(X)\hat{S}_1(k | X) \int_0^k \frac{S_1(t | X) - \hat{S}_1(t | X)}{\hat{S}_1(t | X)} \frac{G_1(t | X) - \hat{G}_1(t | X)}{\hat{G}_1(t | X)} dt \right\} \\ &\quad + E \left[\frac{e(X)G_1(k | X) - \hat{e}(X)\hat{G}_1(k | X)}{\hat{e}(X)\hat{G}_1(k | X)} \left\{ \mu_1(X)S_1(k | X) - \hat{\mu}_1(X)\hat{S}_1(k | X) \right\} \right], \end{aligned}$$

where the third equality follows from a similar derivation as in the Proof of Theorem 3.3 and the fourth equality follows from the Duhamel equation (Gill and Johansen, 1990; Westling et al., 2024). Let R_{11} and R_{12} denote the two terms in the last two lines, respectively. For the remainder term to be of small order, we need $R_{11} = o_P(n^{-1/2})$ and $R_{12} = o_P(n^{-1/2})$.

Next, we show that Assumption 3.5 is a sufficient condition for $R_{11} = o_P(n^{-1/2})$ and $R_{12} = o_P(n^{-1/2})$. In the following discussion, we suppress the dependency of the nuisance functions and their estimated values on X for ease of notation. By the facts that $e\hat{\mu}_1\hat{S}_1(k)/\hat{e}$ is bounded and $1/\{\hat{S}_1(t)\hat{G}_1(t)\}$ is bounded for any $t \leq k$, we employ the Cauchy-Schwarz inequality to

upper bound the term R_{11} by a constant times

$$\int_0^k \|S_1(t) - \hat{S}_1(t)\|_2 \|G_1(t) - \hat{G}_1(t)\|_2 dt,$$

where $\|\cdot\|_2$ denotes the $L_2(P)$ norm. It follows that Assumption 3.5 guarantees $R_{11} = o_P(n^{-1/2})$. For R_{12} , we further suppress the dependency of $S_1(k | X)$, $G_1(k | X)$ and their estimators on both k and X , and rewrite the term as

$$R_{12} = E[\{\hat{e}(G_1 - \hat{G}_1) + G_1(e - \hat{e})\}\{\hat{\mu}_1(S_1 - \hat{S}_1) + S_1(\mu_1 - \hat{\mu}_1)\}/(\hat{e}\hat{G}_1)].$$

Because $\hat{e}/\hat{G}_1$ is bounded, by the Cauchy-Schwarz inequality, the first term in R_{12} is upper bounded by $\|G_1 - \hat{G}_1\|_2 \|S_1 - \hat{S}_1\|_2 = o_P(n^{-1/2})$ by Assumption 3.5. The other three terms in R_{12} can be similarly bounded. Thus, $R_1(P, \hat{P}_n) = o_P(n^{-1/2})$.

Symmetric arguments imply the analogous term for the control arm satisfies $R_0(P, \hat{P}_n) = o_P(n^{-1/2})$. Therefore, the results in Theorem 3.4 follow from the central limit theorem. $\square$

Appendix C

Appendix to Chapter 4

C.1 Proofs

C.1.1 Proof of Theorem 4.1

We derive the identification formula for μ_1 , and that of μ_0 follows from similar arguments. We have

$$\begin{aligned}
 & E\{Y(1, c = \infty)\} \\
 &= E[E\{Y(1, c = \infty) \mid X\}] \\
 &= E[E\{Y(1, c = \infty) \mid A = 1, X\}] \\
 &= E \left[\sum_{k=1}^K E\{Y(1, c = \infty) \mid T = j, A = 1, X\} f_1(j \mid X) + E\{Y(1, c = \infty) \mid T > k, A = 1, X\} S_1(k \mid X) \right],
 \end{aligned}$$

where the first two equalities follow from the law of iterated expectations and Assumption 4.1. For the conditional mean of $Y(1, c = \infty)$, we have

$$\begin{aligned}
 E\{Y(1, c = \infty) \mid T = j, A = 1, X\} &= E\{Y(1, c = \infty) \mid C > k, T = j, A = 1, X\} \\
 &= E\{Y(1, c = \infty) \mid R = 1, C > k, T = j, A = 1, X\} \\
 &= E(Y \mid R = 1, C > k, T = j, A = 1, X) = \mu_{1,j}(X),
 \end{aligned}$$

where the first two equalities follow from Assumptions 4.2 and 4.3, respectively.

$$\begin{aligned}
 E\{Y(1, c = \infty) \mid T > k, A = 1, X\} &= E\{Y(1, c = \infty) \mid C > k, T > k, A = 1, X\} \\
 &= E(Y \mid C > k, T > k, A = 1, X) = \mu_{1,\infty}(X),
 \end{aligned}$$

where the first equality follows from Assumptions 4.2. Therefore, combining terms gives us the outcome regression identification formula (4.2):

$$\mu_1 = E \left\{ \sum_{k=1}^K \mu_{1,j}(X) f_1(j | X) + \mu_{1,\infty}(X) S_1(k | X) \right\}.$$

For the weighting formula, for any $j \leq k$, we have

$$\begin{aligned} & E \left\{ \frac{A 1(T = j, C > k, R = 1) Y}{e_1(X) G_1(k | X) r_{1,j}(X)} \right\} \\ = & E \left[E \left\{ \frac{1(T = j, C > k, R = 1) Y}{G_1(k | X) r_{1,j}(X)} \mid A = 1, X \right\} \right] \\ = & E \left\{ E(Y \mid R = 1, C > k, T = j, A = 1, X) \text{pr}(T = j \mid A = 1, X) \right. \\ & \quad \times \left. \frac{\text{pr}(C > k \mid A = 1, X) \text{pr}(R = 1 \mid C > k, T = j, A = 1, X)}{G_1(k | X) r_{1,j}(X)} \right\} \\ = & E \{ \mu_{1,j}(X) f_1(j | X) \}, \end{aligned}$$

where the first equality follows from the law of iterated expectations and Assumption 4.1 and the second equality follows from Assumption 4.2. Similarly, we have

$$\begin{aligned} E \left\{ \frac{A 1(T \wedge C > k) Y}{e_1(X) G_1(k | X)} \right\} &= E \left[E \left\{ \frac{1(T \wedge C > k) Y}{G_1(k | X)} \mid A = 1, X \right\} \right] \\ &= E \left\{ \frac{1}{G_1(k | X)} E(Y \mid C \wedge T > k, A = 1, X) \text{pr}(C \wedge T > k \mid A = 1, X) \right\} \\ &= E \{ \mu_{1,\infty}(X) S_1(k | X) \}. \end{aligned}$$

Therefore, the weighting formula in (4.3) is also an identification formula.

C.1.2 Proof to Theorem 4.2

Roadmap. We derive the EIF for μ_1 in this proof, and the EIF for μ_0 follows by the same argument. The proof has five steps. First, we factor the observed-data likelihood into a pre-event observation component and conditional post-event components. Second, we identify the stopped pre-event censoring tangent space and show that it is orthogonal to the conditional post-event score space. Third, within each post-event block, we derive the canonical gradient for the relevant conditional outcome mean and combine these gradients to construct the pseudo-outcome $\tilde{U}_1$. Fourth, using $\tilde{U}_1$, we reduce the remaining problem to the standard right-censoring problem, yielding the inverse-probability-of-censoring weighting term plus the pre-event censoring martingale correction. Finally, we consider the layer of treatment assignment and write the full observed-data model from the treated-arm conditional gradient using $A/e_1(X)$ and add the marginal covariate contribution.

Notation. Define

$$\tilde{T} = T \wedge C \wedge k, \quad \delta = \begin{cases} 1, & T \leq k, T < C, \quad T \text{ observed before censoring,} \\ 2, & C < T \wedge k, \quad \text{early censoring of } T, \\ 3, & T > k, C > k, \quad \text{no event and no censoring through } k, \end{cases}$$

and

$$\Delta = 1\{C > T \wedge k\} = 1(\delta = 1 \text{ or } \delta = 3),$$

and thus $\Delta = 1$ means that T is observed. Again, $G_1(t | X) = \text{pr}(C > t | A = 1, X)$ denotes the survival function of the effect-uninformative IE time, and $\Lambda_{T,1}(\cdot | X)$ and $\Lambda_1(\cdot | X)$ denote the conditional cumulative hazards of T and C , respectively, in arm 1. For $j \leq k$, write the conditional mean of Y on the block $(C > k, T = j)$ $\mu_{1,j}(X) = E\{Y | T = j, C > k, R = 1, A = 1, X\}$ and the no-event-through- k mean $\mu_{1,\infty}(X) = E\{Y | T > k, C > k, A = 1, X\}$.

Define the stratum-specific terminal regression

$$m_1(T, X) = \mu_{1,\infty}(X)1(T > k) + \sum_{j=0}^k \mu_{1,j}(X)1(T = j).$$

Intuitively, $m_1(T, X)$ is the target-relevant full-data “terminal value” once the effect-informative IE time history through time k is known. If the subject has $T = j \leq k$, then the relevant mean outcome is $\mu_{1,j}(X)$; if the subject has no effect-informative IE through k , the relevant mean outcome is $\mu_{1,\infty}(X)$. We will construct a post-event pseudo-outcome $\tilde{U}_1(O)$ such that, among units whose effect-informative IE status through k is observed, its conditional mean equals this terminal value. We also define $\mu_1(X) = E\{m_1(T, X) | A = 1, X\}$, with a little abuse of notation. Thus $\mu_1(X)$ averages the terminal regression $m_1(T, X)$ over the conditional distribution of T in treatment arm a , given baseline covariates X . It is the covariate-specific mean outcome under treated arm and no effect-uninformative IE, after integrating over the possible T strata. The marginal arm-specific estimand is therefore $\mu_1 = E\{\mu_1(X)\}$.

Finally, define the continuation value

$$V_1(t, X) = E\{m_1(T, X) | T \geq t, A = 1, X\}.$$

The quantity $V_1(t, X)$ is the expected remaining terminal value among units who still do not experience effect-informative IE at time t , conditional on treated arm $A = 1$ and covariates X . It is the value that is predictable at a possible pre-event censoring time t , and therefore it is the quantity that appears in the censoring martingale correction. In the following proof, for simplicity, we also use “event time” to refer to the effect-informative IE time T and use “censoring time” to refer to the effect-uninformative IE time C .

Step 1: observed-data likelihood and score decomposition in the treated arm.
Condition on $(A = 1, X = x)$. The observed data in the treated arm can be written as

$$O_1 = (U, \delta, W),$$

where $W = W_j = (C, R, RY)$ when $\delta = 1$ and $\tilde{T} = j$, there is no post-event block when $\delta = 2$, and $W = Y$ when $\delta = 3$. In the block $\delta = 1, \tilde{T} = j$, we observe the event time $T = j$, and thus we also have the information that $C > j$. The variable C in W_j records the subsequent truncated value of C through time k . The indicator R is only relevant for observing Y when $C > k$.

The observed-data likelihood is

$$p_1(o \mid x) = p_{\text{pre},1}(\tilde{t}, \delta \mid x) \left\{ \prod_{j=0}^k p_{1,j}(w \mid x)^{1(\delta=1, \tilde{T}=j)} \right\} p_{1,\infty}(y \mid x)^{1(\delta=3)}, \quad (\text{C.1})$$

where

$$p_{1,j}(w \mid x) = p(W_j = w \mid \delta = 1, \tilde{T} = j, A = 1, X = x),$$

and

$$p_{1,\infty}(y \mid x) = p(Y = y \mid \delta = 3, A = 1, X = x).$$

The pre-event factor is

$$p_{\text{pre},1}(\tilde{t}, \delta \mid x) = \begin{cases} \text{pr}(T = \tilde{t}, C > \tilde{t} \mid A = 1, X = x), & \delta = 1, \\ \text{pr}(C = \tilde{t}, T > \tilde{t} \mid A = 1, X = x), & \delta = 2, \\ \text{pr}(T > k, C > k \mid A = 1, X = x), & \delta = 3. \end{cases}$$

For any regular parametric submodel of (C.1), the conditional score decomposes as

$$S_1(O \mid X) = S_{\text{pre},1}(U, \delta \mid X) + \sum_{j=0}^k 1(\delta = 1, \tilde{T} = j) s_{1,j}(W_j, X) + 1(\delta = 3) s_{1,\infty}(Y, X), \quad (\text{C.2})$$

where

$$E\{s_{1,j}(W_j, X) \mid \delta = 1, \tilde{T} = j, A = 1, X\} = 0, \quad E\{s_{1,\infty}(Y, X) \mid \delta = 3, A = 1, X\} = 0.$$

Thus the post-event observed-data tangent space in arm 1 is

$$\begin{aligned} \mathcal{T}_{\text{post},1} = & \left\{ A \left[\sum_{j=0}^k 1(\delta = 1, \tilde{T} = j) v_j(W_j, X) + 1(\delta = 3) v_\infty(Y, X) \right] : \right. \\ & \left. E(v_j \mid \delta = 1, \tilde{T} = j, A = 1, X) = 0, E(v_\infty \mid \delta = 3, A = 1, X) = 0 \right\} \quad (\text{C.3}) \end{aligned}$$

Step 2: the pre-event censoring tangent space. Perturb the cumulative hazard of the stopped pre-event censoring process along

$$d\Lambda_{1,\varepsilon}(t \mid X) = \{1 + \varepsilon h(t, X)\}d\Lambda_1(t \mid X).$$

The pre-event censoring behaves the same as in the standard survival analysis with censoring literature, and therefore differentiating the pre-event likelihood gives the score

$$S_{C,\text{pre},h}(O) = A \int_0^{\tilde{T}} h(t, X) dM_{G_1}(t).$$

Therefore the pre-event censoring tangent space is

$$\mathcal{T}_{C,\text{pre},1} = \overline{\left\{ 1(A=1) \int_0^{\tilde{T}} h(t, X) dM_{G_1}(t) : E \int_0^{\tilde{T}} h^2(t, X) d\Lambda_1(t \mid X) < \infty \right\}}.$$

This is the same right-censoring tangent space that appears in the competing-event case. What is different here is the additional post-event space in (C.3). Accordingly, the C -related observed-data tangent space in the present semi-competing model decomposes as

$$\mathcal{T}_{C,1}^{\text{semi}} = \mathcal{T}_{C,\text{pre},1} \oplus \mathcal{T}_{C,\text{post},1}, \quad (\text{C.4})$$

where $\mathcal{T}_{C,\text{post},1}$ is the part of $\mathcal{T}_{\text{post},1}$ generated by the post-event distribution of (C, R, Y) .

The direct sum in (C.4) requires an additional orthogonality argument in the following. Let $\mathcal{F}_{\text{pre}} = \sigma(A, X, \tilde{T}, \delta)$ be the information contained in the pre-event observation pattern. For a stopped pre-event censoring score $S_{C,\text{pre},h}(O)$, we can write

$$S_{C,\text{pre},h}(O) = 1(A=1) \left[1(\delta=2)h(\tilde{T}, X) - \int_0^{\tilde{T}} h(t, X) d\Lambda_1(t \mid X) \right],$$

and hence $S_{C,\text{pre},h}$ is measurable with respect to $\mathcal{F}_{\text{pre}}$. By contrast, every conditional post-event score has mean zero given the pre-event pattern. For

$$S_{\text{post},v}(O) = 1(A=1) \left[\sum_{j=0}^k 1(\delta=1, \tilde{T}=j)v_j(W_j, X) + 1(\delta=3)v_\infty(Y, X) \right],$$

by definition we have $E\{S_{\text{post},v}(O) \mid \mathcal{F}_{\text{pre}}\} = 0$. Therefore, for every stopped pre-event censoring score and every conditional post-event score,

$$E\{S_{C,\text{pre},h}(O)S_{\text{post},v}(O)\} = E[S_{C,\text{pre},h}(O)E\{S_{\text{post},v}(O) \mid \mathcal{F}_{\text{pre}}\}] = 0.$$

Thus the stopped pre-event censoring score space is orthogonal to the conditional post-event score space, and this orthogonality also justifies deriving the post-event pseudo-outcome separately from the pre-event censoring martingale projection.

Step 3: post-event canonical gradients and construction of $\tilde{U}_1$. Fix $j \leq k$ and condition on the block $\{\delta = 1, \tilde{T} = j\}$. This block is the same as conditioning on $\{T = j, C > j\}$. We further define $B_j = 1(C > k, R = 1)$ for ease of notation. B_j denotes whether an outcome Y is observed within each block $\{\delta = 1, \tilde{T} = j\}$. We have

$$\begin{aligned} E(B_j \mid \delta = 1, \tilde{T} = j, A = 1, X) &= \text{pr}(C > k, R = 1 \mid C > j, T = j, A = 1, X) \\ &= \omega_{1,j}(X)r_{1,j}(X), \end{aligned} \quad (\text{C.5})$$

where

$$\omega_{1,j}(X) = \text{pr}(C > k \mid C > j, T = j, A = 1, X) = \frac{G_1(k \mid X)}{G_1(j \mid X)},$$

and

$$r_{1,j}(X) = \text{pr}(R = 1 \mid C > k, T = j, A = 1, X).$$

We can write the post-event target mean in this block as

$$\mu_{1,j}(X) = \frac{E(B_j Y \mid \delta = 1, \tilde{T} = j, A = 1, X)}{E(B_j \mid \delta = 1, \tilde{T} = j, A = 1, X)}.$$

Because the full tangent space is a direct sum of the component score spaces, we now perturb only the conditional distribution of the observed post-event block in $\{\delta = 1, \tilde{T} = j, A = 1, X\}$, and hold all other likelihood factors fixed to derive the post-event canonical gradients. Let the score for this conditional post-event submodel be $s_{1,j}(W_j, X)$, with

$$E\{s_{1,j}(W_j, X) \mid \delta = 1, \tilde{T} = j, A = 1, X\} = 0.$$

Write the perturbed numerator and denominator as

$$N_\varepsilon = E_\varepsilon(B_j Y \mid \delta = 1, \tilde{T} = j, A = 1, X), \quad D_\varepsilon = E_\varepsilon(B_j \mid \delta = 1, \tilde{T} = j, A = 1, X).$$

Then $\mu_{1,j,\varepsilon} = N_\varepsilon / D_\varepsilon$. Differentiating, we have

$$\dot{N} = E(B_j Y s_{1,j} \mid \delta = 1, \tilde{T} = j, A = 1, X), \quad \dot{D} = E(B_j s_{1,j} \mid \delta = 1, \tilde{T} = j, A = 1, X).$$

By the quotient rule,

$$\begin{aligned} \dot{\mu}_{1,j}(X) &= \frac{\dot{N} - \mu_{1,j}(X)\dot{D}}{D} \\ &= E \left[\frac{B_j}{\omega_{1,j}(X)r_{1,j}(X)} \{Y - \mu_{1,j}(X)\} s_{1,j}(W_j, X) \mid \delta = 1, \tilde{T} = j, A = 1, X \right]. \end{aligned}$$

Therefore the canonical gradient in the $T = j$ post-event block is

$$\phi_{1,j}^{\text{post}}(O) = \frac{1(C > k, R = 1)}{\omega_{1,j}(X)r_{1,j}(X)} \{Y - \mu_{1,j}(X)\}.$$

In the no-event-through- k block, $\delta = 3$ implies $T > k$ and $C > k$, and thus Y is fully observed, so the canonical gradient is just

$$\phi_{1,\infty}^{\text{post}}(O) = Y - \mu_{1,\infty}(X).$$

Combining these two parts of the canonical gradient and define the following post-event pseudo-outcome

$$\tilde{U}_1(O) = 1(\delta = 3)Y + \sum_{j=0}^k 1(\delta = 1, \tilde{T} = j) \left[\mu_{1,j}(X) + \frac{1(C > k, R = 1)}{\omega_{1,j}(X)r_{1,j}(X)} \{Y - \mu_{1,j}(X)\} \right]. \quad (C.6)$$

The blockwise construction also gives the correct derivative contribution to $\mu_1(X)$. For a post-event score $s_{1,j}$ in the block $\{\delta = 1, \tilde{T} = j\}$, we have

$$E \left[\frac{\Delta \tilde{U}_1}{G_1(\tilde{T} | X)} 1(\delta = 1, \tilde{T} = j) s_{1,j}(W_j, X) \mid A = 1, X \right] = \text{pr}(T = j \mid A = 1, X) \dot{\mu}_{1,j}(X),$$

$\text{pr}(\delta = 1, \tilde{T} = j \mid A = 1, X) = \text{pr}(T = j, C > j \mid A = 1, X) = \text{pr}(T = j \mid A = 1, X) G_1(j \mid X)$, and $\tilde{T} = j$ on this block. Therefore the factor $G_1(j \mid X)^{-1}$ converts the observed block probability into the full event-time probability. Similarly, for the no-event block,

$$E \left[\frac{\Delta \tilde{U}_1}{G_1(\tilde{T} | X)} 1(\delta = 3) s_{1,\infty}(Y, X) \mid A = 1, X \right] = \text{pr}(T > k \mid A = 1, X) \dot{\mu}_{1,\infty}(X).$$

Thus the pseudo-outcome term has the correct pathwise derivative with respect to all conditional post-event distributions. Intuitively, $\tilde{U}_1$ has absorbed the post-event censoring structure. The remaining problem is that T itself is observed only when $\Delta = 1$.

Step 4: reduction to an ordinary right-censoring problem. After Step 3, the relevant treated-arm full data reduce to $(X, T, m_1(T, X))$. The corresponding observed-data representer is $\Delta \tilde{U}_1(O)$, with inverse-probability-of-censoring correction. The remaining nuisance is therefore only the stopped pre-event censoring process. Let

$$\phi_1(O) = \frac{\Delta \tilde{U}_1(O)}{G_1(\tilde{T} | X)} + \int_0^{\tilde{T}} \frac{V_1(t, X)}{G_1(t | X)} dM_{G_1}(t) - \mu_1(X),$$

where

$$\mu_1(X) = E\{m_1(T, X) \mid A = 1, X\}.$$

The reduction relies on two conditional expectation identities. First, by the definition of $\tilde{U}_1$,

$$E\{\tilde{U}_1 \mid T, \Delta = 1, A = 1, X\} = m_1(T, X).$$

Second, by independent censoring,

$$E \left\{ \frac{\Delta \tilde{U}_1}{G_1(\tilde{T} | X)} \mid T, A = 1, X \right\} = m_1(T, X).$$

Therefore,

$$E \left\{ \frac{\Delta \tilde{U}_1}{G_1(\tilde{T} | X)} \mid A = 1, X \right\} = E\{m_1(T, X) \mid A = 1, X\} = \mu_1(X).$$

Also,

$$E \left\{ \int_0^{\tilde{T}} \frac{V_1(t, X)}{G_1(t | X)} dM_{G_1}(t) \mid T, A = 1, X \right\} = 0,$$

so the martingale augmentation does not alter the conditional mean but changes the element to make it orthogonal to the pre-event censoring tangent space. Therefore, the reduced gradient $\phi_1(O)$ satisfies

$$E\{\phi_1(O)S_h(O) \mid A = 1, X\} = 0$$

for every $S_h \in \mathcal{T}_{C, \text{pre}, 1}$. Thus $\phi_1(O)$ belongs to the orthocomplement of the pre-event censoring tangent space. Since it also represents the pathwise derivative of the reduced parameter $\mu_1(X)$, it is the efficient influence function for the treated-arm reduced problem.

Step 5: projection through treatment assignment. We finally return to the full observed-data model. Since $E\{\phi_1(O) \mid A = 1, X\} = 0$, the arm-specific observed-data gradient lifts to the full model as

$$D_1(O) = \frac{A}{e_1(X)} \phi_1(O) + \mu_1(X) - \mu_1.$$

The term $A/e_1(X)$ restricts the conditional gradient to the treated arm, and $\mu_1(X) - \mu_1$ is the contribution from the marginal distribution of X . Moreover, D_1 is orthogonal to the treatment score space

$$\mathcal{T}_A = \{(A - e_1(X))b(X) : E\{b^2(X)\} < \infty\},$$

because $E\{\phi_1(O) \mid A = 1, X\} = 0$ and $E\{A - e_1(X) \mid X\} = 0$.

This proves the EIF (4.7) for the treated arm $a = 1$. The same argument applies to $a = 0$, and subtracting the two arm-specific EIFs yields $D_\tau(O) = D_1(O) - D_0(O)$.