

Doubly valid and doubly sharp sensitivity analysis to unobserved confounding for survival outcomes

Jean-Baptiste Baitairian

JEAN-BAPTISTE.BAITAIRIAN@INRIA.FR

Sanofi R&D, Gentilly, France

Inria, Inserm, Université Paris Cité, HeKA, F-75015 Paris, France

Bernard Sebastien

BERNARD.SEBASTIEN@SANOFI.COM

Sanofi R&D, Gentilly, France

Rana Jreich

RANA.JREICH@SANOFI.COM

Sanofi R&D, Gentilly, France

Sandrine Katsahian

SANDRINE.KATSAHIAN@APHP.FR

Inria, Inserm, Université Paris Cité, HeKA, F-75015 Paris, France

CIC-EC 1418 - Paris HEGP, Paris, France

Agathe Guilloux

AGATHE.GUILLOUX@INRIA.FR

Inria, Inserm, Université Paris Cité, HeKA, F-75015 Paris, France

Abstract

Time-to-event outcomes are central in oncology and rare diseases, where treatment effects are often summarized by differences in survival curves or Restricted Mean Survival Time (RMST). In real-world data, estimating these causal effects relies on the absence of unobserved confounding, an assumption that is rarely satisfied. We develop a sensitivity analysis framework for causal treatment effects with survival outcomes under the Marginal Sensitivity Model (MSM). We introduce doubly valid and doubly sharp (DVDS) bounds for differences in survival functions and RMST, extending recent DVDS results to the time-to-event setting while accounting for informative censoring. In practice, our method yields tighter bounds and improved computational efficiency compared to a previous approach from the literature, on simulated and real data. For tractability, we assume independence between censoring and unobserved confounding, a limit that should be addressed in future works.

1. Introduction

Observational data, more generally referred to as real-world data (RWD), offer substantial opportunities to enhance clinical drug development and inform regulatory decision-making. Such data sources include Electronic Health Records (EHR), administrative claims databases, and, increasingly, measurements from wearable devices. In this context, the *21st Century Cures Act* mandates the *U.S. Food and Drug Administration* (FDA) to evaluate the potential use of real-world evidence (RWE) in the approval of medical products (Fang et al., 2025).

Time-to-event outcomes are central endpoints in many clinical trials, particularly in oncology and rare diseases. Overall Survival (OS) is often regarded as the gold-standard endpoint; however, its long follow-up time motivates the use of surrogate endpoints such

as Progression-Free Survival (PFS) or Disease-Free Survival (DFS) (U.S. Food and Drug Administration, 2018; Delgado and Guddati, 2021).

From these survival endpoints, several treatment effect measures can be defined. The hazard ratio remains a widely used summary measure (Barracough et al., 2011), yet its causal interpretation is limited (Lee et al., 2024; Voinot et al., 2025). As a result, alternative estimands have been advocated, including differences in survival functions and Restricted Mean Survival Time (RMST) (Royston and Parmar, 2013), which provide more causally interpretable treatment effects over a fixed time horizon.

Estimating such treatment effects from RWD raises fundamental challenges due to confounding and selection bias, as treatment assignment is no longer randomized. These issues are typically addressed within the framework of causal inference. In particular, causal estimands such as survival curve or RMST differences can be identified under the *ignorability* (or *unconfoundedness*) assumption (Rubin, 1974; Hirano and Imbens, 2004), which posits that all confounders of the treatment–outcome relationship are observed. In practice, this assumption is neither testable nor plausible in many clinical settings, especially when relying on routinely collected data.

Sensitivity analysis provides an approach to assess the impact of unobserved confounding by relaxing point identification to partial identification. Rather than returning a single estimate, these methods characterize a set of plausible values for the causal effect under a sensitivity model that quantifies the magnitude of hidden bias. Among such frameworks, the *Marginal Sensitivity Model* (MSM) (Tan, 2006) has emerged as a flexible and promising approach.

While sensitivity analysis under the MSM is now well developed for binary and continuous outcomes, extensions to causal survival estimands remain limited. In particular, accommodating both robust sensitivity bounds and informative censoring has received little attention in the literature. We address this gap by developing sensitivity bounds for survival function and RMST that inherit strong robustness properties while remaining computationally tractable.

Contributions. We build upon the DVDS framework of Dorn et al. (2024) to derive novel sensitivity bounds for survival outcomes, including survival functions and RMST. Our approach relies on two survival function estimators proposed by Cheng et al. (2022), which admit closed-form expressions. This enables us to obtain analytical bounds that inherit the double validity and double sharpness properties.

In contrast to Lee et al. (2024), whose bounds are defined through optimization problems, we give analytical expressions of our estimators, resulting in improved computational efficiency. Moreover, we work under the assumption of conditionally independent censoring (or informative censoring), which is weaker and more realistic than the independent censoring assumption adopted in this prior work.

Through simulations and real-data applications, we demonstrate that our bounds achieve improved robustness and validity properties, while significantly reducing computational cost, particularly in the case of survival curves.

Outline. In Section 3, we introduce notations and briefly review causal survival analysis under the ignorability assumption. Then, we present our novel sensitivity bounds for (difference of) survival functions; extensions to the RMST are provided in the appendix. In

Section 4, we compare our method with Lee et al. (2024) on simulated data and illustrate its performance on the Right Heart Catheterization (RHC) and German Breast Cancer Study Group (GBCSG) datasets. Section 5 concludes with a discussion and future directions.

Generalizable Insights about Machine Learning in the Context of Healthcare

- In observational healthcare data, ignorability is often unrealistic. Partial identification and sensitivity analysis provide a principled way to quantify uncertainty due to hidden confounding.
- Robustness properties (e.g., double robustness and sharpness) are essential when combining machine learning with causal inference under model misspecification.
- Time-to-event settings require to handle censoring, which is a key source of bias in survival analysis.
- Computationally efficient methods are critical for large-scale healthcare data, especially when sensitivity analyses are required for regulatory use of real-world evidence.

2. Related Work

Sensitivity Analysis under Unobserved Confounding. In observational studies, the ignorability assumption is generally not testable and may be violated due to unobserved confounders. Sensitivity analysis addresses this limitation by replacing point identification with partial identification. Instead of returning a single estimate, these methods characterize a *partially identified set* of causal effects compatible with a sensitivity model that quantifies the degree of departure from ignorability through a sensitivity parameter.

A framework for such analyses is the *Marginal Sensitivity Model* (MSM) introduced by Tan (2006). Within this framework, Zhao et al. (2019) proposed Inverse-Probability-Weighting-based (IPW) (Hirano et al., 2003) sensitivity bounds for Average Treatment Effects (ATE) under unobserved confounding in the case of binary treatments. Another previous model, *Rosenbaum’s Sensitivity Model* (RSM) (Rosenbaum, 2002), is more appropriate when dealing with matched cohorts.

Sharp and Doubly Robust Sensitivity Bounds. Subsequent work has focused on improving the statistical properties of the bounds proposed by Zhao et al. (2019). In particular, Dorn and Guo (2022) introduced *sharp* sensitivity bounds, ensuring that the limiting bounds coincide with the set of causal effects compatible with the sensitivity model and no others. More recently, Dorn et al. (2024) proposed *doubly valid and doubly sharp* (DVDS) bounds, which combine propensity score and outcome regression estimators in the same way as the *Augmented Inverse Probability Weighting* (AIPW) estimator (Robins et al., 1994).

These bounds enjoy several desirable robustness properties. *Double sharpness* ensures that the bounds remain consistent and sharp when either the propensity score model or the outcome regression model is correctly specified, together with an additional nuisance parameter. *Double validity* guarantees valid coverage of the true treatment effect even when this additional nuisance parameter is misspecified, at the cost of potentially conservative bounds. Related refinements have also been explored by Tan (2024).

Extensions to Survival Outcomes. Compared with binary or continuous outcomes, sensitivity analysis for time-to-event data remains relatively underdeveloped. A recent contribution by Lee et al. (2024) extended the MSM-based sensitivity analysis to survival outcomes, with a focus on differences in Restricted Mean Survival Time (RMST). While this work represents an important step toward handling survival estimands under unobserved confounding, the resulting bounds are defined through optimization procedures and rely on stronger censoring assumptions.

Moreover, estimations are biased when right censoring occurs. Right censoring can be corrected using weighting techniques such as the *Inverse Probability of Censoring Weighting* (IPCW) approach (Koul et al., 1981) or the partial likelihood (Cox, 1972).

Positioning of our Work. Our work builds on these developments by extending DVDS sensitivity bounds to causal survival estimands. In particular, we derive analytical bounds for survival functions and RMST that inherit strong robustness properties while accommodating conditionally independent censoring, thereby addressing limitations of existing approaches.

3. Methods

3.1. Notations and Estimands

We denote by $\mathbf{X} \in \mathcal{X} \subseteq \mathbb{R}^{p_{\mathbf{X}}}$ ($p_{\mathbf{X}} \geq 1$) the vector of observed confounders, and by $A \in \{0, 1\}$ the binary treatment (or exposure). Let $T \in \mathbb{R}_+$ be the event time and $C \in \mathbb{R}_+$ the censoring time. The observed time is $\tilde{T} = \min(T, C) = T \wedge C$, and we define $\Delta = \mathbf{1}(T \leq C)$, indicating whether the event is observed ($\Delta = 1$) or censored ($\Delta = 0$). Since observational data typically involve hidden variables, we assume the existence of unobserved confounders $\mathbf{U} \in \mathcal{U} \subseteq \mathbb{R}^{p_{\mathbf{U}}}$ ($p_{\mathbf{U}} \geq 1$).

We adopt the *Neyman–Rubin potential outcomes framework* (Rubin, 1974; Splawa-Neyman, 1990), adapted to survival data as in Lee et al. (2024) or Voinot et al. (2025). We denote by $T^{(a)}$ and $C^{(a)}$ the potential event and censoring times under treatment $a \in \{0, 1\}$. The following SUTVA assumption links observed and potential outcomes.

ASSUMPTION 1 (SUTVA FOR SURVIVAL OUTCOMES):

For $a \in \{0, 1\}$, if $A = a$, then $T = T^{(a)}$ and $C = C^{(a)}$ (consistency), and there is no interference between units. Equivalently, $T = AT^{(1)} + (1 - A)T^{(0)}$ and $C = AC^{(1)} + (1 - A)C^{(0)}$.

We define the *survival function* $S(t) = \mathbb{P}(T > t)$, with potential counterparts $S^{(a)}(t) = \mathbb{P}(T^{(a)} > t)$ for $a \in \{0, 1\}$. The *Restricted Mean Survival Time* (RMST) at horizon τ (Royston and Parmar, 2013) is defined as

$$\text{RMST}(\tau) = \int_0^\tau S(t) dt = \mathbb{E}[T \wedge \tau].$$

An important measure of treatment effect in survival analysis, and therefore our first estimand, is the *difference of survival functions at time t* (Mao et al., 2018; Cheng et al., 2022):

$$\Delta S(t) = S^{(1)}(t) - S^{(0)}(t) = \mathbb{E}[\mathbf{1}(T^{(1)} > t)] - \mathbb{E}[\mathbf{1}(T^{(0)} > t)].$$

A positive value of $\Delta S(t)$ indicates a higher probability of surviving beyond time t under treatment (assuming the event is death). Second, the *difference in RMST*

$$\Delta \text{RMST}(\tau) = \int_0^\tau S^{(1)}(t) dt - \int_0^\tau S^{(0)}(t) dt$$

corresponds to the difference in areas under the survival curves up to time τ , and can be interpreted as an average gain or loss of survival time up to τ (Han and Jung, 2022; Voinot et al., 2025). Figure 4 in Appendix A gives a representation of the different quantities at stake.

These causal estimands cannot be directly computed from the observed data, since both potential outcomes and censoring are present. However, they can be estimated from an i.i.d. sample $\mathcal{D} = \{\mathbf{X}_i, A_i, \tilde{T}_i, \Delta_i\}_{i=1}^n$ of size $n \geq 1$, with individual index i , under appropriate causal assumptions.

3.2. Survival Analysis and Causal Assumptions with Unobserved Confounders

We consider that, had we observed $\mathbf{U}$, we would have captured all common causes of A and $T^{(a)}$, and A and $C^{(a)}$ not included in $\mathbf{X}$ (Imbens, 2003; Kallus et al., 2021). This translates into the following $(\mathbf{X}, \mathbf{U})$ -ignorability assumption.

ASSUMPTION 2 (($\mathbf{X}, \mathbf{U}$)-IGNORABILITY):

For $a \in \{0, 1\}$, $(T^{(a)}, C^{(a)}) \perp\!\!\!\perp A | \mathbf{X}, \mathbf{U}$.

We also assume *conditional independent censoring*, or *informative censoring* (Mao et al., 2018).

ASSUMPTION 3 (CONDITIONAL INDEPENDENT CENSORING WITH $\mathbf{U}$):

For $a \in \{0, 1\}$, $T^{(a)} \perp\!\!\!\perp C^{(a)} | \mathbf{X}, \mathbf{U}, A$.

We define the *nominal propensity score* $e(\mathbf{X}) = \mathbb{P}(A = 1 | \mathbf{X})$ and the *true propensity score* $e(\mathbf{X}, \mathbf{U}) = \mathbb{P}(A = 1 | \mathbf{X}, \mathbf{U})$, and assume *positivity* (Rosenbaum and Rubin, 1983).

ASSUMPTION 4 (POSITIVITY):

$\forall \mathbf{x} \in \mathcal{X}, \mathbf{u} \in \mathcal{U}, 0 < e(\mathbf{x}) < 1$ and $0 < e(\mathbf{x}, \mathbf{u}) < 1$.

We next introduce $S(t | \mathbf{X} = \mathbf{x}, A = a)$ the *conditional survival functions*, and the *conditional censoring function* $\bar{G}(t | \mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}, A = a) = \mathbb{P}(C > t | \mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}, A = a)$. We also define the time $\tau_{\max} > 0$, which is the maximum time such that $S(\tau_{\max} | \mathbf{X} = \mathbf{x}, A = a) > 0$ and $\bar{G}(\tau_{\max} | \mathbf{X} = \mathbf{x}, A = a) > 0$, for all $\mathbf{x}$ in $\mathcal{X}$ and a in $\{0, 1\}$.

For simplicity, we assume that the censoring mechanism does not depend on $\mathbf{U}$.

ASSUMPTION 5 (INDEPENDENCE OF THE CENSORING FUNCTION WITH RESPECT TO $\mathbf{U}$):

For all $t \in [0, \tau_{\max}]$, $\bar{G}(t | \mathbf{X}, \mathbf{U}, A) = \bar{G}(t | \mathbf{X}, A)$.

This assumption is weaker than that of Lee et al. (2024), who impose complete independence, i.e., $C \perp\!\!\!\perp (A, \mathbf{X}, \mathbf{U})$ (non-informative censoring).

Under these assumptions, the potential survival function admits the forms

$$S_I^{(1)}(t) = 1 - \mathbb{E} \left[\frac{A \Delta \mathbf{1}(\tilde{T} \leq t)}{e(\mathbf{X}, \mathbf{U}) \bar{G}(\tilde{T}|\mathbf{X}, A = 1)} \right] \quad \text{and} \quad S_{II}^{(1)}(t) = \mathbb{E} \left[\frac{A \mathbf{1}(\tilde{T} > t)}{e(\mathbf{X}, \mathbf{U}) \bar{G}(t|\mathbf{X}, A = 1)} \right].$$

See Appendix B for the derivation of Forms I and II of the potential survival function. Intuitively, the two forms resemble a traditional Inverse Probability of Treatment Weighting (IPTW) approach where an observed outcome is weighted by the assigned treatment and the inverse of the propensity score to correct for selection bias. Moreover, the division by the censoring function corrects for censoring bias in both forms (IPCW). Note also that Form I uses only uncensored observations (because of Δ) whereas Form II leverages all observations. Finally, Form I may be more stable because the censoring function is evaluated at $\tilde{T}$ whereas it is evaluated at t in Form II: when the positivity of the censoring function is violated for high values of time, small values of the censoring function are “diluted” by the expectancy in Form I but not in Form II. In practice, we recommend to work with Form I, as it is more stable and as we did not identify settings where Form II performed better. Nevertheless, in the following, we keep explanations and results for both forms.

As $\mathbf{U}$ is, by definition, not observed, $e(\mathbf{X}, \mathbf{U})$ cannot be estimated from the data. Our goal is therefore not to get a point estimate but rather to derive *sets of estimates* by controlling the deviation between $e(\mathbf{X})$ and $e(\mathbf{X}, \mathbf{U})$.

3.3. Marginal Sensitivity Model and DVDS Bounds

To this end, we adopt the *Marginal Sensitivity Model* (MSM) of Tan (2006), which bounds the odds ratio between the true and nominal propensity scores by a parameter $\Gamma \geq 1$:

$$\Gamma^{-1} \leq \frac{\mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u}) / \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x}, \mathbf{U} = \mathbf{u})}{\mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}) / \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x})} \leq \Gamma. \quad (1)$$

When $\Gamma = 1$, this reduces to $\mathbf{X}$ -ignorability (or simply ignorability): $\forall a \in \{0, 1\}$, $(T^{(a)}, C^{(a)}) \perp\!\!\!\perp A|\mathbf{X}$ (because the nominal and true propensity scores are equal). Larger values of Γ represent a departure from this assumption. For ways to calibrate the sensitivity parameter Γ , we refer the reader to Cinelli and Hazlett (2020) and Baitairian et al. (2025a).

For convenience, we reformulate the MSM in terms of the potential outcome $Y_t(a) = \mathbf{1}(T^{(a)} \leq t)$ (because $S^{(a)}(t) = 1 - \mathbb{E}[\mathbf{1}(T^{(a)} \leq t)]$). By Proposition 8, Equation (1) is equivalent to

$$\Gamma^{-1} \leq \frac{f(Y_t(a) = y|\mathbf{X} = \mathbf{x}, A = 0)}{f(Y_t(a) = y|\mathbf{X} = \mathbf{x}, A = 1)} \leq \Gamma, \quad (2)$$

for $a \in \{0, 1\}$, $t \in [0, \tau_{\max}]$, and $\mathbf{x} \in \mathcal{X}$.

To derive the sensitivity bounds on $S_I^{(1)}(t)$, we need to define several quantities:

- a modified outcome $Y_{I,t,1} = \Delta \mathbf{1}(\tilde{T} \leq t) / \bar{G}(\tilde{T}|\mathbf{X}, A = 1)$;
- the conditional quantile $Q_\nu(\mathbf{X} = \mathbf{x}, A = 1)$ of order ν of the distribution of $Y_{I,t,1}$ conditionally on $\mathbf{X} = \mathbf{x}$ and $A = 1$;

- the regression functions ρ_+ and ρ_- :

$$\begin{aligned} \rho_{\pm}(t, \mathbf{X}, A = 1, \Gamma) &= \Gamma^{-1} \mathbb{E}[Y_{I,t,1} | \mathbf{X}, A = 1] \\ &\quad + (1 - \Gamma^{-1}) \mathbb{E} \left[Q_{\pm}(\mathbf{X}, 1) + \frac{\{Y_{I,t,1} - Q_{\pm}(\mathbf{X}, 1)\}_{\pm}}{1 - \gamma} \middle| \mathbf{X}, A = 1 \right] \end{aligned}$$

with $Q_+(\mathbf{X}, 1) = Q_{\gamma}(\mathbf{X}, A = 1)$, $Q_-(\mathbf{X}, 1) = Q_{1-\gamma}(\mathbf{X}, A = 1)$ and $\gamma = \Gamma/(1 + \Gamma)$. We write $\pm$ when we must read the formula first with a $+$ sign and then with a $-$ sign, and write $\mp$ for the opposite.

Doubly valid and doubly sharp bounds combine two types of bounds: some based on modified outcome regressions and Equation (2), and some based on propensity scores and Equation (1). Details are provided in Appendix B.4. The next proposition gives these two different upper bounds derived for $S_I^{(1)}(t)$.

Proposition 6 *Under the MSM of Equation (1), both*

$$1 - \mathbb{E}[AY_{I,t,1} + (1 - A)\rho_-(t, \mathbf{X}, 1, \Gamma)] \quad (3)$$

and

$$1 - \mathbb{E} \left[\frac{AQ_{1-\gamma}(\mathbf{X}, 1)}{e(\mathbf{X})} + A(Y_{I,t,1} - Q_{1-\gamma}(\mathbf{X}, 1)) \left(1 + \frac{1 - e(\mathbf{X})}{e(\mathbf{X})} \Gamma^{-\text{sign}(Y_{I,t,1} - Q_{1-\gamma}(\mathbf{X}, 1))} \right) \right] \quad (4)$$

are sharp upper bounds for $S_I^{(1)}(t)$.

The corresponding lower bounds and details on the derivation are given in Appendix B.4. Similar ideas can be used for $S_{II}^{(1)}(t)$ by working with $\mathbb{E}[\mathbb{1}(T^{(1)} > t) | \mathbf{X} = \mathbf{x}]$ instead of $1 - \mathbb{E}[\mathbb{1}(T^{(1)} \leq t) | \mathbf{X} = \mathbf{x}]$.

Combining these two bounds entails the following theorem.

Theorem 7 (DVDS bounds on $S_I^{(1)}(t)$) *Define the lower $S_{I,-}^{(1)}(t, \Gamma)$ and upper $S_{I,+}^{(1)}(t, \Gamma)$ bounds for $S_I^{(1)}(t)$ as*

$$\begin{aligned} S_{I,\pm}^{(1)}(t, \Gamma) &= 1 - \mathbb{E}[AY_{I,t,1} + (1 - A)\rho_{\mp}(t, \mathbf{X}, 1, \Gamma)] \\ &\quad - \mathbb{E} \left[A \frac{1 - e(\mathbf{X})}{e(\mathbf{X})} \left(Q_{\mp}(\mathbf{X}, 1) + \Gamma^{\mp \text{sign}(Y_{I,t,1} - Q_{\mp}(\mathbf{X}, 1))} (Y_{I,t,1} - Q_{\mp}(\mathbf{X}, 1)) - \rho_{\mp}(t, \mathbf{X}, 1, \Gamma) \right) \right], \end{aligned} \quad (5)$$

or equivalently

$$\begin{aligned} S_{I,\pm}^{(1)}(t, \Gamma) &= 1 - \mathbb{E} \left[\frac{AQ_{\mp}(\mathbf{X}, 1)}{e(\mathbf{X})} + A(Y_{I,t,1} - Q_{\mp}(\mathbf{X}, 1)) \left(1 + \frac{1 - e(\mathbf{X})}{e(\mathbf{X})} \Gamma^{\mp \text{sign}(Y_{I,t,1} - Q_{\mp}(\mathbf{X}, 1))} \right) \right] \\ &\quad - \mathbb{E} \left[\left(1 - \frac{A}{e(\mathbf{X})} \right) \rho_{\mp}(t, \mathbf{X}, 1, \Gamma) \right]. \end{aligned} \quad (6)$$

Under Assumptions 1 to 5, and when the censoring function $\bar{G}$ is consistently estimated, these bounds are doubly valid and doubly sharp with respect to the propensity scores and modified outcome regressions, as defined in Dorn et al. (2024).

An equivalent theorem for DVDS bounds on $S_{\text{II}}^{(1)}(t)$ is given in Appendix B.5. On the first line of Equation (5), we recognize the formulation of the bounds via modified outcome regression (Equation (3)). The second line has mean zero and introduces robustness in the sense that, if $\hat{\rho}_{\pm}$ was estimated consistently, the second line would be null, even if $\hat{e}(\mathbf{X})$ was misspecified, and the sharpness property from the formulation of the bounds via modified outcome regression would apply. On the first line of Equation (6), we recognize the formulation of the bounds via propensity scores (Equation (4)) and a similar line of reasoning applies.

Remark that the regression function $\rho_{\pm}$ depends on the sensitivity parameter Γ and on the Conditional Value at Risk (CVaR), also called Expected Shortfall (the second and longer conditional expectation). See, for example, Rockafellar et al. (2000) for a related literature on the CVaR. In our case, the CVaR is an expected value measured in the tail of the distribution of Y conditionally on $\mathbf{X}$ and A . When $\Gamma = 1$ (no unobserved confounding), $\rho_{\pm}$ is simply the expected value of Y conditionally on $\mathbf{X}$ and A . When Γ tends to infinity, $\rho_{\pm}$ gets closer to the CVaR, so to the tail of the distribution of the outcome.

Observe that the results from Dorn et al. (2024) are also compatible with survival outcomes and we can replace their continuous outcome “ Y ” with

$$Y_{\text{I},t,1} = \Delta \mathbf{1}(\tilde{T} \leq t) / \bar{G}(\tilde{T} | \mathbf{X}, A = 1) \quad \text{and} \quad Y_{\text{II},t,1} = \mathbf{1}(\tilde{T} > t) / \bar{G}(t | \mathbf{X}, A = 1)$$

to recognize, respectively, our DVDS bounds on Form I and Form II. However, our Theorems 7 and 10 are based on different proofs of the bounds via modified outcome regression. Note also that our bounds inherit the efficiency property of the bounds from Dorn et al. (2024).

Bounds for the unexposed group, $S_{\pm}^{(0)}(t, \Gamma)$, are obtained by replacing A with $1 - A$ (therefore, conditioning on $A = 1$ also becomes conditioning on $A = 0$) and replacing $e(\mathbf{X})$ with $1 - e(\mathbf{X})$. Bounds for the difference of survival functions are obtained by taking $S_{-}^{(1)}(t, \Gamma) - S_{+}^{(0)}(t, \Gamma)$ (lower bound) and $S_{+}^{(1)}(t, \Gamma) - S_{-}^{(0)}(t, \Gamma)$ (upper bound).

With an i.i.d. sample $\mathcal{D}$ as defined earlier, empirical estimators of Equations (5) and (19) can be obtained by replacing the nuisance parameters with plug-in estimators $\hat{e}$, $\hat{\bar{G}}$, $\hat{\rho}_{\pm}$, and $\hat{Q}_{\pm}$ using K -fold cross-fitting, and by replacing the expectations with empirical averages. In particular, the modified outcome regression $\rho_{\pm}$ can be achieved by regressing $\Gamma^{-1}Y_{\text{I},t,1} + (1 - \Gamma^{-1})(Q_{\pm}(\mathbf{X}, 1) + (1 - \gamma)^{-1}\{Y_{\text{I},t,1} - Q_{\pm}(\mathbf{X}, 1)\}_{\pm})$ on $\mathbf{X}$ and A for Form I (and similarly for Form II). The resulting intervals are called *Point Estimate Intervals* (PEI). Confidence intervals (CIs) can be obtained via the percentile bootstrap method (Zhao et al., 2019) or with less computationally intensive Wald-type CIs (Dorn et al., 2024).

4. Experiments

In this section, we compare the performance of our DVDS bounds for Forms I and II to the bounds of Lee et al. (2024) for the survival functions and difference of survival functions on simulated and real-world data. Results for bounds on the RMST and difference in RMST are given in Appendix D.3. The R repository [CSM.dvds.bound.survival.outcomes](#) to reproduce the results is publicly available on the Sanofi-Public GitHub under a non-commercial license.

4.1. Implementation Details

All experiments were performed under the R software (version 4.3.2) (R Core Team, 2023) using parallel computing on 32 CPUs (Amazon EC2 c6i.8xlarge instances). We parallelized computations between different values of time t (survival function) with our DVDS bounds and parallelized the optimization step in the method of Lee et al. (2024).

5-fold cross-fitting was used to estimate the nuisance parameters and avoid overfitting. The nominal propensity score $\hat{e}$ was estimated via logistic regression (`glm` function from the `stats` package, version 4.3.2), the conditional quantile regressions $\hat{Q}_\pm$ via linear quantile regression (`rq` function from the `quantreg` package, version 5.97) with the Frisch–Newton interior point method, the modified outcome regressions $\hat{\rho}_\pm$ via random forests (`ranger` function from the `ranger` package, version 0.16.0), and the censoring function $\hat{\tilde{G}}$ via a Cox proportional hazard model (`coxph` function from the `survival` package, version 3.5-7) with the Breslow method for dealing with ties. Except for the `rq` and `coxph` functions, we used the default hyperparameter values. In the second application on real data, 95%-confidence intervals were computed via the percentile bootstrap method with $B = 200$ bootstrap samples.

We reimplemented the direct optimization method from Lee et al. (2024) for the difference in RMST (Equations (9) to (12) from their paper) and adapted it for the difference of survival functions by removing the integration step. Even if it is faster, we did not implement the approximated method presented in their Section 3.3 as it relies on additional assumptions on the censoring process. As they suggest in Section 3.4, we used the `optim` function from the `stats` package (more precisely, we used the `optimParallel` function, a parallelized implementation of `optim`, from the `optimParallel` package, version 1.0-2) with the L-BFGS-B optimization method, a maximum number of iterations of 100, and a numeric forward difference approximation of the gradient.

Following Dorn et al. (2024), we do not normalize (or stabilize) our estimators as in Appendix B.1 and B.2. Experimental results support this choice by showing good stability properties, even without normalization.

4.2. Results on Synthetic Experiments

We applied our methodology to a simulated dataset whose construction was inspired by Lee et al. (2024) and Baitairian et al. (2025a). In our experiments, we considered $p_{\mathbf{X}} = 2$ observed confounders and $p_{\mathbf{U}} = 2$ unobserved confounders. We chose $\mathbf{X} \sim \mathcal{U}(-1, 1)^{p_{\mathbf{X}}}$, where $\mathcal{U}(a, b)^p$ is the uniform distribution on (a, b) of dimension p . For j in $\llbracket 1, p_{\mathbf{U}} \rrbracket$, conditionally on $\mathbf{X} = \mathbf{x}$, we designed each unobserved confounder U_j as

$$U_j | \mathbf{X} = \mathbf{x} \sim \lambda G + (1 - \lambda) \beta_j^T \mathbf{x} = \mathcal{N}((1 - \lambda) \beta_j^T \mathbf{x}, \lambda^2),$$

with $G \sim \mathcal{N}(0, 1)$, $0 < \lambda < 1$, and $\beta_j = (\beta_{j,1}, \dots, \beta_{j,p_{\mathbf{X}}}) \in \mathbb{R}_+^{p_{\mathbf{X}}}$. We chose the distribution of A conditionally on $\mathbf{X}$ and $\mathbf{U}$ to be a Bernoulli satisfying the MSM with a true sensitivity parameter $\Gamma^* = 3$. We designed the true propensity score $e(\mathbf{X}, \mathbf{U}) = \mathbb{P}(A = 1 | \mathbf{X}, \mathbf{U})$ such that it marginalizes to the nominal propensity score $e(\mathbf{X}) = \mathbb{P}(A = 1 | \mathbf{X}) = \text{logistic}(\delta^T \mathbf{X} + 0.5)$, with $\delta \in \mathbb{R}^{p_{\mathbf{X}}}$. We explain how this is done in Appendix D.1.

The potential outcome $T^{(a)}$, with a in $\{0, 1\}$, was randomly generated using the inverse transform sampling method via a proportional hazards model with Weibull-distributed base-

line hazard (Bender et al., 2005). In particular,

$$T^{(a)} = 10 \cdot \left(\frac{-\log(U^{(a)})}{0.95 \cdot \exp(a \log(5) + \beta_{\mathbf{X}}^T \mathbf{X} + \beta_{\mathbf{U}}^T \mathbf{U})} \right)^{1/1.8},$$

where $U^{(a)} \sim \mathcal{U}(0, 1)$. The event time T was then defined as $T = \mathbb{1}(A = 1) \cdot T^{(1)} + \mathbb{1}(A = 0) \cdot T^{(0)}$ and the censoring time C followed a Weibull distribution of shape 6 and scale 10. Note that, not to disadvantage the method of Lee et al. (2024), T and C are independent and C is independent from $\mathbf{X}$ and A , which corresponds to non-informative censoring. We provide experimental results under informative censoring in Appendix D.3. We generated 20 Monte-Carlo samples, each of size $n = 1000$, and computed the sensitivity bounds for 10 values of time equally spaced between 0.1 and 9. Finally, administrative censoring was applied after the 0.95-quantile of the observed times. The complete simulation setup is given in Appendix D.1. See also Appendix D.3 for additional results (higher number of Monte-Carlo samples, larger sample sizes, high dimensional dataset...).

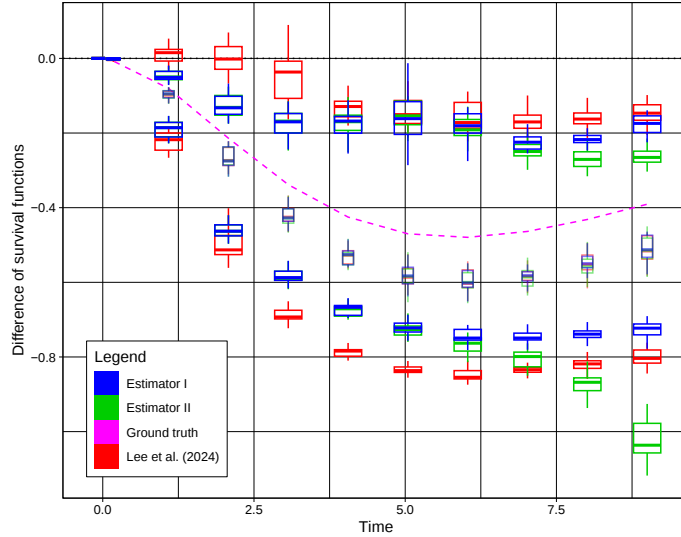


Figure 1: Sensitivity analysis for the difference of survival functions on the simulated data. The dotted lines in magenta are the true difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

Results of sensitivity analysis for the difference of survival functions are given in Figure 1 for our DVDS bounds via Forms I (in blue), II (in green) and the bounds from Lee et al. (2024) (in red). Under ignorability (narrow and transparent boxplots), the estimations deviate slightly from the ground truth (dotted lines in magenta) because of unobserved confounders, but are still close to the true value. Note that the three methods give approximately the same estimation.

When $\Gamma = 3$, that is, the same value Γ^* that generated the data, the intervals almost always contain the true value of the difference of survival functions. Remark that the DVDS bounds for Form II are unstable for high values of time t . This behavior was also identified by [Cheng et al. \(2022\)](#) and can be explained by near violations of the positivity of the censoring process. On the contrary, the DVDS bounds for Form I remain stable.

Notice that the intervals formed by the medians of our DVDS bounds for Form I are also practically always included in the analogous intervals of the method from [Lee et al. \(2024\)](#). These bounds show good coverage of the real difference of survival functions, and sharpness with respect to the bounds from [Lee et al. \(2024\)](#).

Finally, our bounds also enjoy shorter computation times. To obtain Figure 1, 12 hours and 48 minutes were necessary for the method of [Lee et al. \(2024\)](#). In comparison, only 1 minute and 48 seconds were required by our DVDS bounds for Form I *and* Form II *simultaneously*. This represents a reduction of computation time by a factor of 427.

4.3. Results on Real Data

4.3.1. RIGHT HEART CATHETERIZATION DATASET

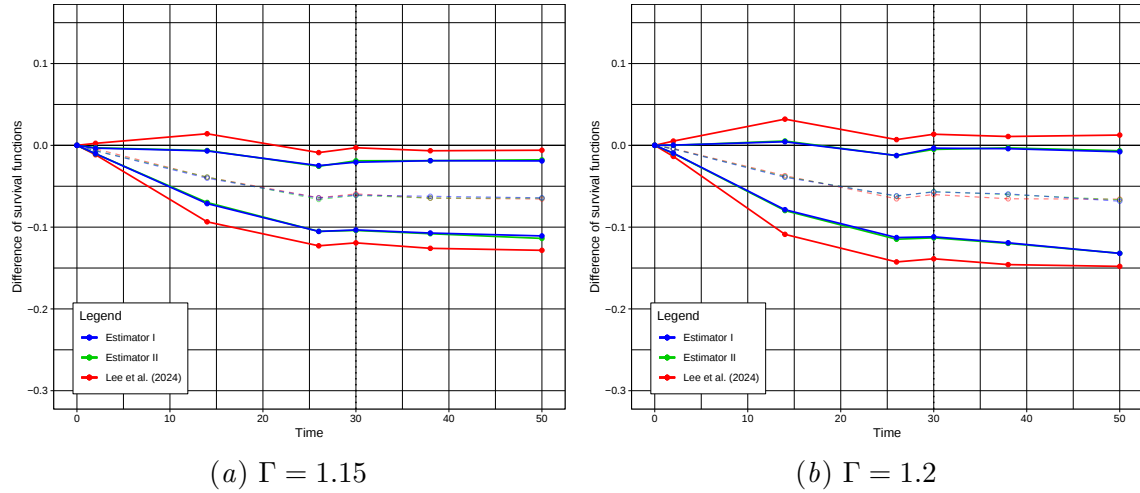


Figure 2: Sensitivity analysis for the difference of survival functions for $\Gamma = 1.15$ (2(a)subfigure) and 1.2 (2(b)subfigure) on the RHC data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from [Lee et al. \(2024\)](#) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical dotted black line indicates a time of 30 days. Note that the bounds with Form II are almost coincident with the bounds obtained with Form I.

The **Right Heart Catheterization** (RHC) dataset consists in 5735 observations from 5 different US hospitals and was used by [Connors et al. \(1996\)](#) to explore the link between survival and the use of RHC during the first 24 hours in an intensive care unit. It was

studied again later in [Lin et al. \(1998\)](#). [Connors et al. \(1996\)](#) showed that RHC could lead to decreased survival rates in critically ill patients. A binary exposition signals if RHC was used ($A = 1$) or not ($A = 0$). The study admission date, hospital discharge date, date of last contact and date of death are given when available. Patients are considered uncensored ($\Delta = 1$) when a date of death is provided. In that case, the observed time $\tilde{T}$ is the period between study admission and death. When no date of death is provided, patients are considered censored ($\Delta = 0$). In that case, the observed time is the period between study admission and last contact, or hospital discharge if it corresponds to the last contact. Finally, the dataset provides 47 continuous and categorical variables at baseline (demographic and medical covariates), all included in our analysis. We applied administrative censoring on observed times that were greater than 180 days and binary encoded categorical variables.

The results of the sensitivity analysis on the difference of survival functions are given in Figure 2, for a sensitivity parameter Γ equal to 1.15 and 1.2. Additional graphs as well as computation times are provided in Appendix D.3. The PEIs obtained via the DVDS bounds for Form I (in blue), II (in green) and via the method of [Lee et al. \(2024\)](#) (in red) were computed for 5 equally spaced values of time between 2 and 50 days, to which we also added 30 days. The estimates under ignorability ($\Gamma = 1$) were represented with dotted lines.

Observe that, under ignorability, the difference of survival functions is always negative, regardless of the time point between 0 and 50 days. This is consistent with findings from [Connors et al. \(1996\)](#) on the negative effect of RHC on survival.

We highlight day 30 with a vertical dotted line to compare the current results of sensitivity analysis on survival outcomes to the ones obtained in [Dorn et al. \(2024\)](#) on a binary outcome (survival rate at day 30). We designate by *critical value*, the lowest value of Γ for which the sensitivity bounds contain the null effect. We recall that, with a binary outcome, the critical value of Γ with the method from [Zhao et al. \(2019\)](#) is close to 1.15, whereas it is slightly larger and closer to 1.2 with the method from [Dorn et al. \(2024\)](#). Similar critical values are obtained with survival outcomes. As the method from [Lee et al. \(2024\)](#) is equivalent to the bounds from [Zhao et al. \(2019\)](#) adapted to survival outcomes, the red PEI from Figure 2(a)subfigure when $\Gamma_{c, Lee} = 1.15$ touches 0 at day 30. Our DVDS bounds for Form I and Form II give results that are closer to the ones obtained with the method from [Dorn et al. \(2024\)](#). Indeed, when $\Gamma_{c, DVDS} = 1.2$, the blue and green PEIs from Figure 2(b)subfigure touch 0 at day 30. Note also that our DVDS bounds are contained in the red bounds of our comparator for all time points between 0 and 50 days.

We also applied an *informal benchmarking* approach in the Appendix where Γ was estimated by taking the maximum odds ratio of the MSM with and without each observed covariate, one after another, as described by [Cinelli and Hazlett \(2020\)](#) or [Baitairian et al. \(2025a\)](#). The minimum, median and maximum estimated values are 1.07, 1.63, and 35.6. Among the observed covariates, 8 have an associated confounding strength $\hat{\Gamma}$ lower than $\Gamma_{c, Lee}$ and 9 have an associated confounding strength lower than $\Gamma_{c, DVDS}$.

In the end, we can conclude that, even if RHC appears to be detrimental to survival at 30 days in intensive care units, “small” confounders would be enough to explain away a causal association.

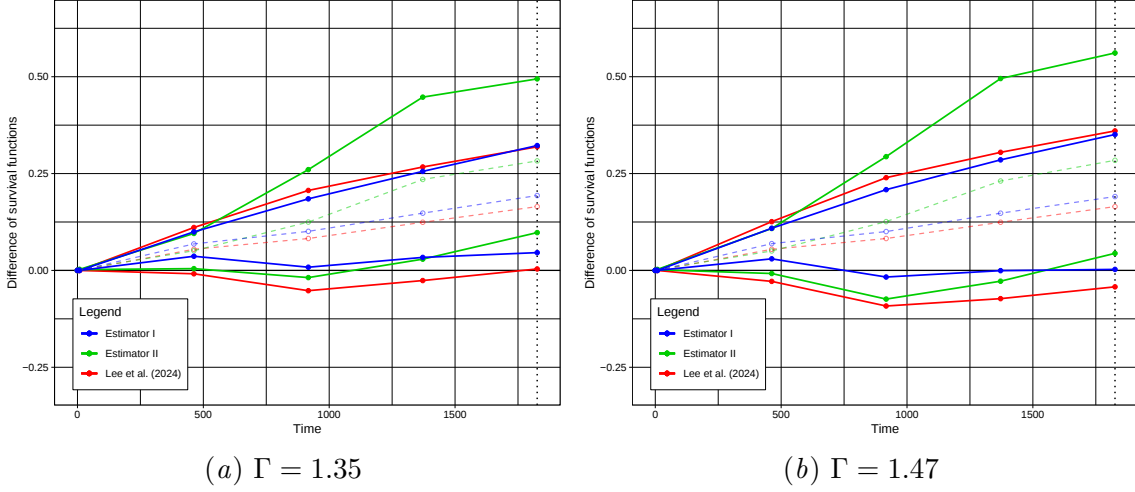


Figure 3: Sensitivity analysis for the difference of survival functions for $\Gamma = 1.35$ (3(a)subfigure) and 1.47 (3(b)subfigure) on the GBCSG data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical black dotted line indicates a time of 1826.25 days (5 years).

4.3.2. GERMAN BREAST CANCER STUDY GROUP DATASET

The **German Breast Cancer Study Group** (GBCSG) dataset (Schumacher et al., 1994) consists in 686 observations that relate the use of tamoxifen, a selective estrogen receptor modulator, to recurrence-free survival in breast cancer. We refer to Lee et al. (2024) for more details on the GBCSG dataset. A binary exposition indicates the treatment group ($A = 1$ if chemotherapy and tamoxifen were used, and $A = 0$ if only chemotherapy was used). The observed time $\tilde{T}$ is the number of days between the date of mastectomy and the date of the first occurrence of either locoregional recurrence, distant recurrence, contralateral tumor, secondary tumor, or death (Schumacher et al., 1994; Lee et al., 2024). Patients with no occurrence were censored at the end of follow-up ($\Delta = 0$). Finally, 7 continuous and categorical variables at baseline are provided and were included in our analysis. See Appendix D.2 for more details. We used a similar setup as with the RHC data except that the PEIs were computed for 5 equally spaced values of time between 8 and 1826.25 days (5 years). Results are summarized in Figure 3. Additional results as well as computation times are given in Appendix D.3.

Observe that, under ignorability, the difference of survival functions is always positive. This is consistent with studies that showed a positive effect of the combination of tamoxifen with chemotherapy when compared with chemotherapy alone (Early Breast Cancer Trialists' Collaborative Group et al., 1998). The slope is slightly higher when the bounds are based on Form II. At 5 years, the critical value for the method of Lee et al. (2024) is $\Gamma_{c, \text{Lee}} = 1.35$, whereas it is equal to $\Gamma_{c, \text{I}} = 1.47$ and $\Gamma_{c, \text{II}} = 1.58$ with, respectively, our bounds with Form I

and Form II (see Figure 14(d)subfigure in Appendix for the critical value with Form II). Lee et al. (2024) indicate that “according to an expert’s opinion [...], the degree of possible unmeasured confounding is expected to be weaker than tumor size”. As the maximum value of odds ratio of propensity scores with and without tumor size is approximately 2.05 (see Appendix D.3), our bounds do not allow to draw a conclusion about the sign of the difference of survival functions ($\Gamma_{c,I} < \Gamma_{c,II} \leq 2.05$) as well as the bounds of Lee et al. (2024) ($\Gamma_{c, Lee} \leq 2.05$).

Finally, Lee et al. (2024) identified a critical value of 1.08 with 95%-CIs for the RMST. In our experiments (see Figures 14 and 15 in Appendix D.3), we also identified the same critical value for survival functions with their method, and a slightly larger critical value with our DVDS bounds, which would not allow us to conclude about a positive effect of tamoxifen on recurrence-free survival in breast cancer if unobserved confounders of such a strength existed.

5. Conclusion and Discussion

In this manuscript, we have extended the sensitivity analysis to unobserved confounders proposed by Dorn et al. (2024) to survival outcomes and proposed two estimators of sensitivity bounds under an informative censoring assumption for the difference in survival curves and RMST. In particular, our bounds enjoy the double validity and double sharpness properties of the bounds of Dorn et al. (2024). We have shown on simulated and real data that our bounds outperformed the proposition of Lee et al. (2024)—an adaptation of the bounds by Zhao et al. (2019) to survival outcomes—in terms of sharpness and computation time.

Note that we used estimators of the survival function given in Cheng et al. (2022) and not the traditional Kaplan Meier estimator, as in Lee et al. (2024). This is because defining sensitivity bounds with a Kaplan Meier estimator leads to solving a fractional problem since the true propensity score appears in the numerator *and* the denominator of the Kaplan Meier, whereas it only appears once in Forms I and II of the survival function.

As compared to Lee et al. (2024), we made the weaker assumption of informative censoring. However, this led us to suppose that the censoring function is independent from the unobserved confounders to avoid another sensitivity analysis to the violation of informative censoring. Future works should tackle this issue.

Additionally, our bounds have robustness properties with respect to the propensity score and modified outcome regression, given that the conditional quantiles are consistently estimated, but lack robustness with respect to the censoring function. Voinot et al. (2025) present a potential solution called double augmented correction. This should also be addressed in future works.

Finally, in the appendix, we derived DVDS bounds for the RMST by directly considering an outcome that is proportional to the minimum between the event time T and the time horizon τ , instead of integrating bounds for the survival function. Even though this approach suffers from instabilities for high values of τ and when censoring is important, it could pave the way for a promising and faster method to compute sensitivity bounds for the RMST.

Acknowledgments

The authors want to thank Charlotte Voinot (Sanofi, Gentilly, France, and PreMeDICAL, UMR UA11, Montpellier, France) for open discussions about causal survival analysis. The authors also want to thank the reviewers for their valuable feedback and suggestions, which helped improving the quality of this work. The research in this paper was supported by Sanofi and INRIA (Institut National de Recherche en Informatique et en Automatique) through the CIFRE program (Convention Industrielle de Formation par la Recherche).

Conflicts of interest

Jean-Baptiste Baitairian, Bernard Sebastien and Rana Jreich are or were Sanofi employees and they may hold shares and/or stock options in the company. Agathe Guilloux is employed by Inria. Sandrine Katsahian is employed by Université Paris-Cité and Assistance Publique - Hôpitaux de Paris (AP-HP).

References

- Jean-Baptiste Baitairian, Bernard Sebastien, Rana Jreich, Sandrine Katsahian, and Agathe Guilloux. Sensitivity analysis to unobserved confounders: A comparative review to estimate confounding strength in sensitivity models. *arXiv preprint arXiv:2510.16560*, 2025a.
- Jean-Baptiste Baitairian, Bernard Sebastien, Rana Jreich, Sandrine Katsahian, and Agathe Guilloux. Sharp bounds for continuous-valued treatment effects with unobserved confounders. *Biometrical Journal*, 67(5):e70084, 2025b.
- Helen Barraclough, Lorinda Simms, and Ramaswamy Govindan. Biostatistics primer: what a clinician ought to know: hazard ratios. *Journal of Thoracic Oncology*, 6(6):978–982, 2011.
- Ralf Bender, Thomas Augustin, and Maria Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723, 2005.
- Chao Cheng, Fan Li, Laine E Thomas, and Fan Li. Addressing extreme propensity scores in estimating counterfactual survival functions via the overlap weights. *American journal of epidemiology*, 191(6):1140–1151, 2022.
- Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):39–67, 2020.
- Alfred F Connors, Theodore Speroff, Neal V Dawson, Charles Thomas, Frank E Harrell, Douglas Wagner, Norman Desbiens, Lee Goldman, Albert W Wu, Robert M Califf, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. *Jama*, 276(11):889–897, 1996.
- David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

- Amanda Delgado and Achuta Kumar Guddati. Clinical endpoints in oncology-a primer. *American journal of cancer research*, 11(4):1121, 2021.
- Jacob Dorn and Kevin Guo. Sharp sensitivity analysis for inverse propensity weighting via quantile balancing. *Journal of the American Statistical Association*, pages 1–13, 2022.
- Jacob Dorn, Kevin Guo, and Nathan Kallus. Doubly-valid/doubly-sharp sensitivity analysis for causal inference with unmeasured confounding. *Journal of the American Statistical Association*, pages 1–12, 2024.
- Early Breast Cancer Trialists’ Collaborative Group et al. Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet*, 351(9114):1451–1467, 1998.
- Yixin Fang, Pallavi Mishra-Kalyani, Xiang Zhang, Susan Gruber, Shu Yang, Peng Ding, Mingyang Shan, Joo-Yeon Lee, Mark van der Laan, Douglas Faries, et al. Sensitivity analysis for unmeasured confounding in medical product development and evaluation using real world evidence. *Statistics in Biopharmaceutical Research*, pages 1–12, 2025.
- Kyunghwa Han and Inkyung Jung. Restricted mean survival time for survival analysis: a quick guide for clinical researchers. *Korean Journal of Radiology*, 23(5):495, 2022.
- Martin Herdegen and Cosimo Munari. An elementary proof of the dual representation of expected shortfall. *Mathematics and Financial Economics*, 17(4):655–662, 2023.
- Keisuke Hirano and Guido W Imbens. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, 226164:73–84, 2004.
- Keisuke Hirano, Guido W Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189, 2003.
- Guido W Imbens. Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2):126–132, 2003.
- Nathan Kallus, Xiaojie Mao, and Masatoshi Uehara. Causal inference under unmeasured confounding with negative controls: A minimax learning approach. *arXiv preprint arXiv:2103.14029*, 2021.
- H Koul, Vyaghreswarudu Susarla, and John Van Ryzin. Regression analysis with randomly right-censored data. *The Annals of statistics*, pages 1276–1288, 1981.
- Seungjae Lee, Ji Hoon Park, and Woojoo Lee. Sensitivity analysis for unmeasured confounding in estimating the difference in restricted mean survival time. *Statistical Methods in Medical Research*, 33(11-12):1979–1992, 2024.
- Danyu Y Lin, Bruce M Psaty, and Richard A Kronmal. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, pages 948–963, 1998.

- Huzhang Mao, Liang Li, Wei Yang, and Yu Shen. On the propensity score weighting analysis with survival outcome: Estimands, estimation, and inference. *Statistics in medicine*, 37(26):3745–3763, 2018.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL <https://www.R-project.org/>.
- James M Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.
- R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.
- Paul R Rosenbaum. Covariance adjustment in randomized experiments and observational studies. *Statistical Science*, 17(3):286–327, 2002.
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- Patrick Royston and Mahesh KB Parmar. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC medical research methodology*, 13(1):152, 2013.
- Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.
- Martin Schumacher, G Bastert, H Bojar, K Hübner, M Olschewski, W Sauerbrei, C Schmoor, C Beyerle, RL Neumann, and HF Rauschecker. Randomized 2 x 2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. german breast cancer study group. *Journal of Clinical Oncology*, 12(10):2086–2093, 1994.
- Jerzy Splawa-Neyman. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465 – 472, 1990. doi: 10.1214/ss/1177012031. URL <https://doi.org/10.1214/ss/1177012031>.
- Zhiqiang Tan. A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101(476):1619–1637, 2006.
- Zhiqiang Tan. Model-assisted sensitivity analysis for treatment effects under unmeasured confounding via regularized calibrated estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae034, 2024.
- U.S. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics, 12 2018. URL <https://www.fda.gov/media/71195/download>.
- Charlotte Voinot, Clément Berenfeld, Imke Mayer, Bernard Sebastien, and Julie Josse. Treatment effect estimation in causal survival analysis: Practical recommendations. *arXiv preprint arXiv:2501.05836*, 2025.

Qingyuan Zhao, Dylan S Small, and Bhaswar B Bhattacharya. Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(4):735–761, 2019.

Appendix A. Additional Background Information

Figure 4 gives a visual representation of different quantities at stake: the survival functions and the difference in RMST.

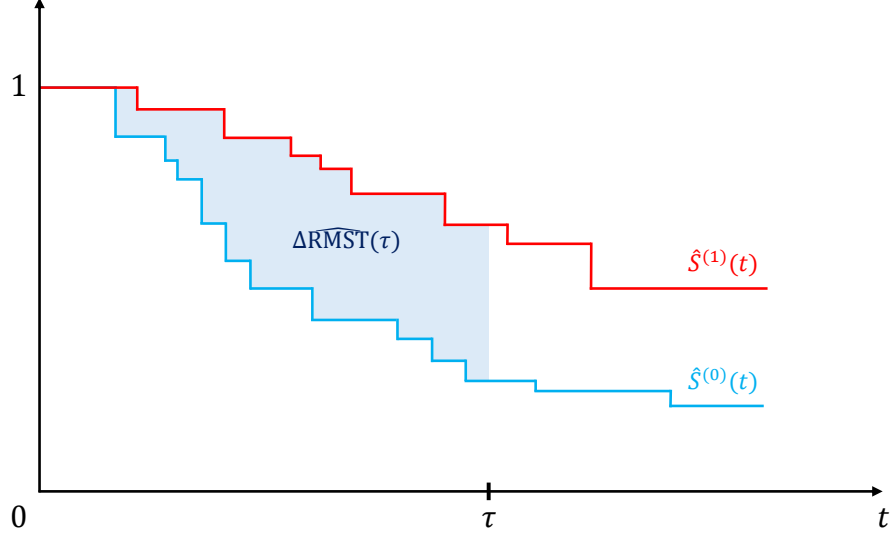


Figure 4: Illustration of the estimated difference in RMST (light blue area), from time 0 up to time τ , using the estimated survival functions among the exposed group (in red) and the unexposed group (in blue).

Appendix B. Proofs

B.1. Derivation of Form I

We refer the reader to the appendix of [Cheng et al. \(2022\)](#) for a proof of the consistency of the estimator of Form I under ignorability and we will show that, under Assumptions 1 to 5, $S_I^{(1)}(t) = S^{(1)}(t)$.

$$\begin{aligned}
 S^{(1)}(t) &= \mathbb{P}(T^{(1)} > t) \\
 &= \mathbb{E}[\mathbb{1}(T^{(1)} > t)] \\
 &= \mathbb{E}[\mathbb{E}[\mathbb{1}(T^{(1)} > t)|\mathbf{X}, \mathbf{U}]] \quad \text{by tower property} \\
 &= \mathbb{E}\left[\frac{e(\mathbf{X}, \mathbf{U})}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[\mathbb{1}(T^{(1)} > t)|\mathbf{X}, \mathbf{U}]\right] \quad \text{by positivity (Assumption 4)} \\
 &= \mathbb{E}\left[\frac{\mathbb{E}[A|\mathbf{X}, \mathbf{U}]}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[\mathbb{1}(T^{(1)} > t)|\mathbf{X}, \mathbf{U}]\right] \quad \text{because } \mathbb{E}[A|\mathbf{X}, \mathbf{U}] = \mathbb{P}(A = 1|\mathbf{X}, \mathbf{U}) = e(\mathbf{X}, \mathbf{U}) \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A \mathbb{1}(T^{(1)} > t)}{e(\mathbf{X}, \mathbf{U})} \middle| \mathbf{X}, \mathbf{U}\right]\right] \quad \text{by } (\mathbf{X}, \mathbf{U})\text{-ignorability (Assumption 2)} \\
 &= 1 - \mathbb{E}\left[\frac{\mathbb{E}[A \mathbb{1}(T \leq t)|\mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U})}\right] \quad \text{by tower property and consistency} \\
 &= 1 - \mathbb{E}\left[\frac{\mathbb{E}[A \mathbb{1}(T \leq t)|\mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\mathbb{E}[\mathbb{1}(C \geq T)|\mathbf{X}, \mathbf{U}, A]}{\mathbb{E}[\mathbb{1}(C \geq T)|\mathbf{X}, \mathbf{U}, A]}\right] \\
 &= 1 - \mathbb{E}\left[\frac{\mathbb{E}[A \mathbb{1}(T \leq t) \mathbb{1}(C \geq T)|\mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U}) \mathbb{P}(C \geq T|\mathbf{X}, \mathbf{U}, A)}\right] \quad \text{by Assumption 3} \\
 &= 1 - \mathbb{E}\left[\frac{A \mathbb{1}(T \leq t) \mathbb{1}(C \geq T)}{e(\mathbf{X}, \mathbf{U}) \mathbb{P}(C \geq T|\mathbf{X}, \mathbf{U}, A = 1)}\right] \quad \text{by Assumption 1}^*, \text{ tower prop., and } A \in \{0, 1\} \\
 &= 1 - \mathbb{E}\left[\frac{A \Delta \mathbb{1}(\tilde{T} \leq t)}{e(\mathbf{X}, \mathbf{U}) \mathbb{P}(C \geq \tilde{T}|\mathbf{X}, A = 1)}\right] \quad \text{by definition of } \Delta \text{ and } \tilde{T}, \text{ and by Assumption 5} \\
 &= 1 - \mathbb{E}\left[\frac{A \Delta \mathbb{1}(\tilde{T} \leq t)}{e(\mathbf{X}, \mathbf{U}) \tilde{G}(\tilde{T}|\mathbf{X}, A = 1)}\right] \quad \text{Unstabilized form} \\
 &= 1 - \mathbb{E}\left[\frac{A \Delta \mathbb{1}(\tilde{T} \leq t)}{e(\mathbf{X}, \mathbf{U}) \tilde{G}(\tilde{T}|\mathbf{X}, A = 1)}\right] \bigg/ \mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})}\right] \quad \text{Stabilized form}
 \end{aligned}$$

*We use SUTVA: $\mathbb{1}(T > t) = A \mathbb{1}(T^{(1)} > t) + (1 - A) \mathbb{1}(T^{(0)} > t)$ because $A \mathbb{1}(T > t) = A^2 \mathbb{1}(T^{(1)} > t) + A(1 - A) \mathbb{1}(T^{(0)} > t) = A \mathbb{1}(T^{(1)} > t)$. The same is true for C : $A \mathbb{1}(C > t) = A \mathbb{1}(C^{(1)} > t)$.

The stabilized form comes from the fact that $\mathbb{E}[A/e(\mathbf{X}, \mathbf{U})] = \mathbb{E}[\mathbb{E}[A/e(\mathbf{X}, \mathbf{U})|\mathbf{X}, \mathbf{U}]] = \mathbb{E}[\mathbb{E}[A|\mathbf{X}, \mathbf{U}]/e(\mathbf{X}, \mathbf{U})] = 1$.

B.2. Derivation of Form II

We refer the reader to the appendix of [Cheng et al. \(2022\)](#) for a proof of the consistency of the estimator of Form II and we will show that, under Assumptions 1 to 5, $S_{\text{II}}^{(1)}(t) = S^{(1)}(t)$.

$$\begin{aligned}
 S^{(1)}(t) &= \mathbb{P}(T^{(1)} > t) \\
 &= \mathbb{E}[\mathbb{1}(T^{(1)} > t)] \\
 &= \mathbb{E}[\mathbb{E}[\mathbb{1}(T^{(1)} > t) | \mathbf{X}, \mathbf{U}]] \quad \text{by tower property} \\
 &= \mathbb{E}\left[\frac{e(\mathbf{X}, \mathbf{U})}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[\mathbb{1}(T^{(1)} > t) | \mathbf{X}, \mathbf{U}]\right] \quad \text{by positivity (Assumption 4)} \\
 &= \mathbb{E}\left[\frac{\mathbb{E}[A | \mathbf{X}, \mathbf{U}]}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[\mathbb{1}(T^{(1)} > t) | \mathbf{X}, \mathbf{U}]\right] \quad \text{because } \mathbb{E}[A | \mathbf{X}, \mathbf{U}] = \mathbb{P}(A = 1 | \mathbf{X}, \mathbf{U}) = e(\mathbf{X}, \mathbf{U}) \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A \mathbb{1}(T^{(1)} > t)}{e(\mathbf{X}, \mathbf{U})} \middle| \mathbf{X}, \mathbf{U}\right]\right] \quad \text{by } (\mathbf{X}, \mathbf{U})\text{-ignorability (Assumption 2)} \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A \mathbb{1}(T > t)}{e(\mathbf{X}, \mathbf{U})} \middle| \mathbf{X}, \mathbf{U}\right]\right] \quad \text{by consistency}^* \text{ (Assumption 1)} \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A \mathbb{1}(T > t)}{e(\mathbf{X}, \mathbf{U})} \middle| \mathbf{X}, \mathbf{U}, A\right]\right] \quad \text{by tower property} \\
 &= \mathbb{E}\left[\frac{A \mathbb{E}[\mathbb{1}(T > t) | \mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U})}\right] \\
 &= \mathbb{E}\left[\frac{A \mathbb{E}[\mathbb{1}(T > t) | \mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\mathbb{E}[\mathbb{1}(C > t) | \mathbf{X}, \mathbf{U}, A]}{\mathbb{E}[\mathbb{1}(C > t) | \mathbf{X}, \mathbf{U}, A]}\right] \\
 &= \mathbb{E}\left[\frac{A \mathbb{E}[\mathbb{1}(\tilde{T} > t) | \mathbf{X}, \mathbf{U}, A]}{e(\mathbf{X}, \mathbf{U}) \mathbb{P}(C > t | \mathbf{X}, \mathbf{U}, A)}\right] \quad \text{by Assumption 3 and } \mathbb{1}(\tilde{T} > t) = \mathbb{1}(T > t) \mathbb{1}(C > t) \\
 &= \mathbb{E}\left[\frac{A \mathbb{1}(\tilde{T} > t)}{e(\mathbf{X}, \mathbf{U}) \mathbb{P}(C > t | \mathbf{X}, A = 1)}\right] \quad \text{by tower property, } A \in \{0, 1\}, \text{ and Assumption 5} \\
 &= \mathbb{E}\left[\frac{A \mathbb{1}(\tilde{T} > t)}{e(\mathbf{X}, \mathbf{U}) \tilde{G}(t | \mathbf{X}, A = 1)}\right] \quad \text{Unstabilized form} \\
 &= \mathbb{E}\left[\frac{A \mathbb{1}(\tilde{T} > t)}{e(\mathbf{X}, \mathbf{U}) \tilde{G}(t | \mathbf{X}, A = 1)}\right] \bigg/ \mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})}\right] \quad \text{Stabilized form}
 \end{aligned}$$

*We use SUTVA: $\mathbb{1}(T > t) = A \mathbb{1}(T^{(1)} > t) + (1 - A) \mathbb{1}(T^{(0)} > t)$ because $A \mathbb{1}(T > t) = A^2 \mathbb{1}(T^{(1)} > t) + A(1 - A) \mathbb{1}(T^{(0)} > t) = A \mathbb{1}(T^{(1)} > t)$. The same is true for C : $A \mathbb{1}(C > t) = A \mathbb{1}(C^{(1)} > t)$.

The stabilized form comes from the fact that $\mathbb{E}[A/e(\mathbf{X}, \mathbf{U})] = \mathbb{E}[\mathbb{E}[A/e(\mathbf{X}, \mathbf{U}) | \mathbf{X}, \mathbf{U}]] = \mathbb{E}[\mathbb{E}[A | \mathbf{X}, \mathbf{U}]/e(\mathbf{X}, \mathbf{U})] = 1$.

B.3. Proposition 8 and Proof

Proposition 8 *For all $a \in \{0, 1\}$, $\mathbf{x} \in \mathcal{X}$, and $y \in \mathcal{Y}$, the ratio in the MSM from Equation (1) can be equivalently written*

$$\Gamma^{-1} \leq \frac{f(Y_t(a) = y | \mathbf{X} = \mathbf{x}, A = 0)}{f(Y_t(a) = y | \mathbf{X} = \mathbf{x}, A = 1)} \leq \Gamma. \quad (7)$$

Proof By Lemma 1 from Tan (2024), the formulation of the MSM in terms of unobserved confounders (1) is equivalent to the formulation in terms of potential outcomes

$$\Gamma^{-1} \leq \frac{\mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}, Y_t(a) = y) / \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x}, Y_t(a) = y)}{\mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}) / \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x})} \leq \Gamma, \quad (8)$$

for $a \in \{0, 1\}$, $t \in [0, \tau_{\max}]$, $(y, \mathbf{x}, \mathbf{u}) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{U}$, and $\Gamma \geq 1$. By Bayes' formula,

$$\begin{aligned} \mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}, Y_t(a) = y) &= \frac{f(Y_t(a) = y|\mathbf{X} = \mathbf{x}, A = 1) \mathbb{P}(A = 1|\mathbf{X} = \mathbf{x})}{f(Y_t(a) = y|\mathbf{X} = \mathbf{x})} \quad \text{and} \\ \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x}, Y_t(a) = y) &= \frac{f(Y_t(a) = y|\mathbf{X} = \mathbf{x}, A = 0) \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x})}{f(Y_t(a) = y|\mathbf{X} = \mathbf{x})}. \end{aligned}$$

By plugging these formula into Equation (8), we obtain the desired result. The converse is also true by Bayes' formula. $\blacksquare$

B.4. Derivation of the DVDS Bounds for $S_I^{(1)}(t)$

We denote $\{\cdot\}_- = \min(0, \cdot)$, the negative part function, and $\{\cdot\}_+ = \max(0, \cdot)$, the positive part function.

B.4.1. BOUNDS VIA MODIFIED OUTCOME REGRESSION

We start by deriving bounds for $S^{(1)}(t)$ using a modified outcome regression approach. Observe that, for all $\mathbf{x} \in \mathcal{X}$,

$$S^{(1)}(t|\mathbf{X} = \mathbf{x}) = 1 - \mathbb{E}[Y_t(1)|\mathbf{X} = \mathbf{x}] = 1 - \int_{\mathcal{Y}} y f(Y_t(1) = y|\mathbf{X} = \mathbf{x}) dy. \quad (9)$$

However, the density $y \mapsto f(Y_t(1) = y|\mathbf{X} = \mathbf{x})$ is not directly identifiable from the data. Therefore, we rewrite it as

$$\begin{aligned} f(Y_t(1) = y|\mathbf{X} = \mathbf{x}) &= f(Y_t(1) = y, A = 1|\mathbf{X} = \mathbf{x}) + f(Y_t(1) = y, A = 0|\mathbf{X} = \mathbf{x}) \\ &= f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 1) \mathbb{P}(A = 1|\mathbf{X} = \mathbf{x}) \\ &\quad + f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 0) \mathbb{P}(A = 0|\mathbf{X} = \mathbf{x}). \end{aligned} \quad (10)$$

The only non-identifiable term is $f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 0)$. By introducing the density ratio

$$w^*(Y_t(1) = y, \mathbf{X} = \mathbf{x}) = \frac{f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 0)}{f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 1)},$$

we finally get

$$\begin{aligned} S^{(1)}(t|\mathbf{X} = \mathbf{x}) &= 1 - \left(e(\mathbf{x}) \int_{\mathcal{Y}} y f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 1) dy \right. \\ &\quad \left. + (1 - e(\mathbf{x})) \int_{\mathcal{Y}} y w^*(Y_t(1) = y, \mathbf{X} = \mathbf{x}) f(Y_t(1) = y|\mathbf{X} = \mathbf{x}, A = 1) dy \right) \\ &= 1 - e(\mathbf{x}) \mathbb{E}[Y_t(1)|\mathbf{X} = \mathbf{x}, A = 1] \\ &\quad - (1 - e(\mathbf{x})) \mathbb{E}[Y_t(1) w^*(Y_t(1), \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1], \end{aligned}$$

after plugging Equation (10) into Equation (9).

As in the proof of Appendix B.1, we further introduce the censoring function $\bar{G}$ by noting that

$$\mathbb{E}[Y_t(1)|\mathbf{X} = \mathbf{x}, A = 1] = \mathbb{E}\left[\frac{\Delta \mathbb{1}(\tilde{T} \leq t)}{\bar{G}(\tilde{T}|\mathbf{X}, A = 1)} \middle| \mathbf{X} = \mathbf{x}, A = 1\right].$$

By denoting $Y_{I,t,1} = \Delta \mathbb{1}(\tilde{T} \leq t)/\bar{G}(\tilde{T}|\mathbf{X}, A = 1)$, where the dependence in $\mathbf{X}$, $\tilde{T}$, and Δ is dropped for simplicity, we can write

$$\begin{aligned} S^{(1)}(t|\mathbf{X} = \mathbf{x}) &= 1 - e(\mathbf{x}) \mathbb{E}[Y_{I,t,1}|\mathbf{X} = \mathbf{x}, A = 1] \\ &\quad - (1 - e(\mathbf{x})) \mathbb{E}[Y_{I,t,1}w^*(Y_t(1), \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1]. \end{aligned}$$

Notice that, by Equation (2), w^* takes its values in $[\Gamma^{-1}, \Gamma]$ and that, for all $\mathbf{x} \in \mathcal{X}$,

$$\mathbb{E}[w^*(Y_t, \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1] = 1, \quad (11)$$

with $Y_t = \mathbb{1}(T > t)$. The sensitivity bounds for $S^{(1)}(t|\mathbf{X} = \mathbf{x})$ are therefore defined as

$$\begin{aligned} S_-^{(1)}(t, \mathbf{x}, \Gamma) &= 1 - e(\mathbf{x}) \mathbb{E}[Y_{I,t,1}|\mathbf{X} = \mathbf{x}, A = 1] \\ &\quad - (1 - e(\mathbf{x})) \max_{w \in \mathcal{W}(\Gamma)} \mathbb{E}[Y_{I,t,1}w(Y_t, \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1] \quad \text{and} \end{aligned} \quad (12)$$

$$\begin{aligned} S_+^{(1)}(t, \mathbf{x}, \Gamma) &= 1 - e(\mathbf{x}) \mathbb{E}[Y_{I,t,1}|\mathbf{X} = \mathbf{x}, A = 1] \\ &\quad - (1 - e(\mathbf{x})) \min_{w \in \mathcal{W}(\Gamma)} \mathbb{E}[Y_{I,t,1}w(Y_t, \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1], \end{aligned} \quad (13)$$

where the set $\mathcal{W}(\Gamma)$ is

$$\begin{aligned} \mathcal{W}(\Gamma) &= \{w : \{0, 1\} \times \mathcal{X} \mapsto [\Gamma^{-1}, \Gamma]; \\ &\quad \forall (\mathbf{x}, t) \in \mathcal{X} \times [0, \tau_{\max}], \mathbb{E}[w(Y_t, \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1] = 1\}. \end{aligned}$$

As w^* is unknown, the intuition behind Equations (12) and (13) is to search the minimizer and maximizer in a set $\mathcal{W}(\Gamma)$ of functions w that “look like” w^* .

Without loss of generality, let us consider the upper bound, $S_+^{(1)}(t, \mathbf{x}, \Gamma)$. As $w \in \mathcal{W}(\Gamma)$,

$$(w(Y_t, \mathbf{x}) - \Gamma^{-1})/(1 - \Gamma^{-1}) \in [0, (1 - \gamma)^{-1}],$$

with $\gamma = \Gamma/(1 + \Gamma)$. Using this fact and the normalization constraint from Equation (11), we rewrite the minimization problem as

$$\begin{aligned} &\min_{w \in \mathcal{W}(\Gamma)} \mathbb{E}[Y_{I,t,1}w(Y_t, \mathbf{x})|\mathbf{X} = \mathbf{x}, A = 1] \\ &= \Gamma^{-1} \mathbb{E}[Y_{I,t,1}|\mathbf{X} = \mathbf{x}, A = 1] + (1 - \Gamma^{-1}) \min_{w \in \mathcal{W}(\Gamma)} \mathbb{E}\left[Y_{I,t,1} \frac{w(Y_t, \mathbf{x}) - \Gamma^{-1}}{1 - \Gamma^{-1}} \middle| \mathbf{X} = \mathbf{x}, A = 1\right]. \end{aligned} \quad (14)$$

A similar reasoning as in the proof of Theorem 3.1 from Baitairian et al. (2025b) allows for the identification of the minimization problem as a Conditional Value at Risk (CVaR) of

level $1 - \gamma$, thanks to the ‘‘Fenchel–Moreau–Rockafellar’’ dual representation of the CVaR (Herdegen and Munari, 2023), the CVaR being defined as

$$\text{CVaR}_{1-\gamma}(t, \mathbf{X} = \mathbf{x}, A = 1) = \mathbb{E}[-Y_{I,t,1} | Y_{I,t,1} \leq Q_{1-\gamma}(\mathbf{x}, 1), \mathbf{X} = \mathbf{x}, A = 1],$$

where $Q_\nu(\mathbf{X} = \mathbf{x}, A = 1)$ is the quantile of order ν of $Y_{I,t,1}$ conditionally on $\mathbf{X} = \mathbf{x}$ and $A = 1$.

As pointed out by Rockafellar et al. (2000), the CVaR can be expressed directly in terms of the conditional quantiles of the outcome. By transposing this result to $Q_{1-\gamma}(\mathbf{X} = \mathbf{x}, A = 1)$, we get

$$\text{CVaR}_{1-\gamma}(t, \mathbf{X} = \mathbf{x}, A = 1) = \mathbb{E} \left[Q_{1-\gamma}(\mathbf{x}, 1) + \frac{\{Y_{I,t,1} - Q_{1-\gamma}(\mathbf{x}, 1)\}_-}{1 - \gamma} \middle| \mathbf{X} = \mathbf{x}, A = 1 \right], \quad (15)$$

where $\{\cdot\}_- = \min(0, \cdot)$.

After plugging Equation (15) into Equation (14), and combining the result with Equation (12), we obtain

$$S_+^{(1)}(t, \mathbf{x}, \Gamma) = 1 - e(\mathbf{x}) \mathbb{E}[Y_{I,t,1} | \mathbf{X} = \mathbf{x}, A = 1] - (1 - e(\mathbf{x})) \rho_-(t, \mathbf{x}, 1, \Gamma),$$

where

$$\begin{aligned} \rho_-(t, \mathbf{X} = \mathbf{x}, A = 1, \Gamma) &= \Gamma^{-1} \mathbb{E}[Y_{I,t,1} | \mathbf{x}, A = 1] \\ &\quad + (1 - \Gamma^{-1}) \mathbb{E} \left[Q_{1-\gamma}(\mathbf{x}, 1) + \frac{\{Y_{I,t,1} - Q_{1-\gamma}(\mathbf{x}, 1)\}_-}{1 - \gamma} \middle| \mathbf{x}, A = 1 \right]. \end{aligned}$$

Finally, taking the expectancy over $\mathbf{X}$ and using the tower rule gives

$$S_+^{(1)}(t, \Gamma) = 1 - \mathbb{E}[AY_{I,t,1} + (1 - A)\rho_-(t, \mathbf{x}, 1, \Gamma)]. \quad (16)$$

This result is similar to the formulation using adversarial regressions of Dorn et al. (2024), where the outcome ‘‘Y’’ is replaced with $Y_{I,t,1}$.

By results given in the proof of Theorem 3.1 in Baitairian et al. (2025b), these bounds are sharp when both the conditional quantile and censoring function are consistently estimated. Similar arguments to those given in Dorn et al. (2024) (Proposition 4) or in Baitairian et al. (2025b) (proof of Proposition 3.5) help us to show that the resulting bound is also singly-valid with respect to the modified outcome regression, that is, if the modified outcome regression correctly learns the conditional expectation of the *estimated* modified outcome but the conditional quantile is misspecified, the bound is still valid but possibly conservative.

B.4.2. BOUNDS VIA PROPENSITY SCORES

We now express our bounds for $S^{(1)}(t)$ in terms of propensity scores and will combine them to the formulation using modified outcome regression in the next section. Under $\mathbf{X}$ -ignorability, Form I writes

$$S_1^{(1)}(t) = 1 - \mathbb{E} \left[\frac{A\Delta \mathbb{1}(\tilde{T} \leq t)}{e(\mathbf{X})\bar{G}(\tilde{T} | \mathbf{X}, A = 1)} \right] = 1 - \mathbb{E} \left[\frac{AY_{I,t,1}}{e(\mathbf{X})} \right],$$

where $Y_{I,t,1} = \Delta \mathbb{1}(\tilde{T} \leq t) / \tilde{G}(\tilde{T} | \mathbf{X}, A = 1)$, as previously.

Using the formulation of the MSM with unobserved confounders (Equation (1)), [Dorn and Guo \(2022\)](#) (Proposition 2) introduced sharp sensitivity bounds when $e(\mathbf{X})$ is replaced with worst-case propensity scores $e_{\pm}(\mathbf{X}, Y)$ satisfying the MSM, where Y is the (continuous) outcome, under the constraint that $\mathbb{E}[A/e_{\pm}(\mathbf{X}, Y) | \mathbf{X}] = 1$. We reformulate their result in terms of survival outcomes.

Proposition 9 *There exist two worst-case propensity scores, $e_{-}(\mathbf{x}, Y_{I,t,1}, \Gamma)$ and $e_{+}(\mathbf{x}, Y_{I,t,1}, \Gamma)$, satisfying the MSM (Equation (1)) and*

$$\mathbb{E}[A/e_{\pm}(\mathbf{X}, Y_{I,t,1}, \Gamma) | \mathbf{X} = \mathbf{x}] = 1 \quad (17)$$

such that

$$\begin{aligned} 1/e_{-}(\mathbf{x}, Y_{I,t,1}, \Gamma) &= 1 + \frac{1 - e(\mathbf{x})}{e(\mathbf{x})} \Gamma^{-\text{sign}(Y_{I,t,1} - Q_{1-\gamma}(\mathbf{x}, 1))} \quad \text{and} \\ 1/e_{+}(\mathbf{x}, Y_{I,t,1}, \Gamma) &= 1 + \frac{1 - e(\mathbf{x})}{e(\mathbf{x})} \Gamma^{\text{sign}(Y_{I,t,1} - Q_{\gamma}(\mathbf{x}, 1))}, \end{aligned}$$

where $Q_{\nu}(\mathbf{X} = \mathbf{x}, A = 1)$ is the quantile of order ν of $Y_{I,t,1}$ conditionally on $\mathbf{X} = \mathbf{x}$ and $A = 1$. The sharp lower and upper bounds on $S_I^{(1)}(t)$ are then

$$S_{I,-}^{(1)}(t, \Gamma) = 1 - \mathbb{E} \left[\frac{AY_{I,t,1}}{e_{+}(\mathbf{X}, Y_{I,t,1}, \Gamma)} \right] \quad \text{and} \quad S_{I,+}^{(1)}(t, \Gamma) = 1 - \mathbb{E} \left[\frac{AY_{I,t,1}}{e_{-}(\mathbf{X}, Y_{I,t,1}, \Gamma)} \right]. \quad (18)$$

As in [Dorn et al. \(2024\)](#), we can introduce $Q_{\gamma}(\mathbf{X}, 1)$ or $Q_{1-\gamma}(\mathbf{X}, 1)$ in Equation (18) using the constraint from Equation (17). The resulting bounds are

$$\begin{aligned} S_{I,-}^{(1)}(t, \Gamma) &= 1 - \mathbb{E} \left[\frac{AQ_{\gamma}(\mathbf{X}, 1)}{e(\mathbf{X})} + A(Y_{I,t,1} - Q_{\gamma}(\mathbf{X}, 1)) \left(1 + \frac{1 - e(\mathbf{x})}{e(\mathbf{x})} \Gamma^{\text{sign}(Y_{I,t,1} - Q_{\gamma}(\mathbf{x}, 1))} \right) \right] \\ \text{and} \quad S_{I,+}^{(1)}(t, \Gamma) &= 1 - \mathbb{E} \left[\frac{AQ_{1-\gamma}(\mathbf{X}, 1)}{e(\mathbf{X})} + A(Y_{I,t,1} - Q_{1-\gamma}(\mathbf{X}, 1)) \left(1 + \frac{1 - e(\mathbf{x})}{e(\mathbf{x})} \Gamma^{-\text{sign}(Y_{I,t,1} - Q_{1-\gamma}(\mathbf{x}, 1))} \right) \right]. \end{aligned}$$

They are sharp when the propensity score, conditional quantile and censoring function are all consistently estimated. When the censoring function is consistently estimated, they are singly-valid with respect to the propensity score, that is, if the propensity score is correctly estimated but not the conditional quantile, the bound is still valid but possibly conservative. Proof of single validity rests upon the same arguments used in [Dorn et al. \(2024\)](#) (Proposition 2).

B.5. Theorem 10

The next theorem establishes the DVDS bounds on the second form of the survival function $S_{II}^{(1)}(t)$ and their properties, where we denote $Y_{II,t,1} = \mathbb{1}(\tilde{T} > t) / \tilde{G}(t | \mathbf{X}, A = 1)$. The proof is based on the same elements as those in Theorem 7.

Theorem 10 (DVDS bounds on $S_{\text{II}}^{(1)}(t)$) Define the lower $S_{\text{II},-}^{(1)}(t, \Gamma)$ and upper $S_{\text{II},+}^{(1)}(t, \Gamma)$ bounds for $S_{\text{II}}^{(1)}(t)$ as

$$\begin{aligned} S_{\text{II},\pm}^{(1)}(t, \Gamma) = & \mathbb{E}[AY_{\text{II},t,1} + (1 - A)\rho_{\pm}(t, \mathbf{X}, 1, \Gamma)] \\ & + \mathbb{E}\left[A \frac{1 - e(\mathbf{X})}{e(\mathbf{X})} \right. \\ & \times \left. \left(Q_{\pm}(\mathbf{X}, 1) + \Gamma^{\pm \text{sign}(Y_{\text{II},t,1} - Q_{\pm}(\mathbf{X}, A=1))} (Y_{\text{II},t,1} - Q_{\pm}(\mathbf{X}, 1)) - \rho_{\pm}(t, \mathbf{X}, 1, \Gamma) \right) \right], \quad (19) \end{aligned}$$

where $Q_+(\mathbf{X}, 1) = Q_{\gamma}(\mathbf{X}, A = 1)$, $Q_-(\mathbf{X}, 1) = Q_{1-\gamma}(\mathbf{X}, A = 1)$, with $\gamma = \Gamma/(1 + \Gamma)$, $Q_{\nu}(\mathbf{X}, A = 1)$, the quantile of order ν of the distribution of $Y_{\text{II},t,1}$ conditionally on $\mathbf{X}$ and $A = 1$, and

$$\begin{aligned} \rho_{\pm}(t, \mathbf{X}, A = 1, \Gamma) = & \Gamma^{-1} \mathbb{E}[Y_{\text{II},t,1} | \mathbf{X}, A = 1] \\ & + (1 - \Gamma^{-1}) \mathbb{E}\left[Q_{\pm}(\mathbf{X}, 1) + \frac{\{Y_{\text{II},t,1} - Q_{\pm}(\mathbf{X}, 1)\}_{\pm}}{1 - \gamma} \middle| \mathbf{X}, A = 1\right]. \end{aligned}$$

Under Assumptions 1 to 5, and when the censoring function $\hat{G}$ is consistently estimated, these bounds are doubly valid and doubly sharp with respect to the propensity scores and modified outcome regressions, as defined in Dorn et al. (2024).

Appendix C. Bounds for the RMST and Difference in RMST

We propose two methods to derive bounds for the RMST: by integrating the bounds for the survival function, or by deriving directly the bounds for the RMST.

C.1. Bounds via Integration by the Rectangle Method

Bounds for the RMST can be obtained by integrating the bounds for the survival functions with respect to t between 0 and a time horizon τ . The lower and upper bounds of the partially identified set for the difference in RMST are therefore, respectively,

$$\Delta\text{RMST}(\tau)_- = \int_0^{\tau} S_-^{(1)}(t, \Gamma) dt - \int_0^{\tau} S_+^{(0)}(t, \Gamma) dt \quad \text{and} \quad (20)$$

$$\Delta\text{RMST}(\tau)_+ = \int_0^{\tau} S_+^{(1)}(t, \Gamma) dt - \int_0^{\tau} S_-^{(0)}(t, \Gamma) dt, \quad (21)$$

where the information on the form (Form I or II) in the index is dropped.

With an i.i.d. sample $\mathcal{D}$ as defined earlier in Section 3.1, empirical estimators of Equations (20) and (21) can be obtained by integrating empirical estimators of Equations (5) and (19) using the rectangle method.

C.2. Faster Bounds via Analytical Expression

It is possible to show that, under Assumptions 1 to 5, the RMST among the treated can be written

$$\text{RMST}^{(1)}(\tau) = \mathbb{E}\left[\frac{A\Delta^{\tau}}{e(\mathbf{X}, \mathbf{U})\bar{G}(\tilde{T} \wedge \tau | \mathbf{X}, A = 1)} \tilde{T} \wedge \tau\right],$$

with $\Delta^\tau = \mathbb{1}(T \wedge \tau \leq C) = \Delta + (1 - \Delta) \mathbb{1}(\tilde{T} \geq \tau) = \max(\Delta, \mathbb{1}(\tilde{T} \geq \tau))$.

Proof

$$\begin{aligned}
 \text{RMST}^{(1)}(\tau) &= \mathbb{E}[T^{(1)} \wedge \tau] \\
 &= \mathbb{E}[\mathbb{E}[T^{(1)} \wedge \tau | \mathbf{X}, \mathbf{U}]] \quad \text{by tower property} \\
 &= \mathbb{E}\left[\frac{e(\mathbf{X}, \mathbf{U})}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[T^{(1)} \wedge \tau | \mathbf{X}, \mathbf{U}]\right] \quad \text{by positivity (Assumption 4)} \\
 &= \mathbb{E}\left[\frac{\mathbb{E}[A | \mathbf{X}, \mathbf{U}]}{e(\mathbf{X}, \mathbf{U})} \mathbb{E}[T^{(1)} \wedge \tau | \mathbf{X}, \mathbf{U}]\right] \quad \text{because } \mathbb{E}[A | \mathbf{X}, \mathbf{U}] = \mathbb{P}(A = 1 | \mathbf{X}, \mathbf{U}) = e(\mathbf{X}, \mathbf{U}) \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})} T^{(1)} \wedge \tau \middle| \mathbf{X}, \mathbf{U}\right]\right] \quad \text{by } (\mathbf{X}, \mathbf{U})\text{-ignorability (Assumption 2)} \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\mathbb{E}[\mathbb{1}(T^{(1)} \wedge \tau \leq C) | \mathbf{X}, \mathbf{U}, A, T^{(1)}]}{\mathbb{E}[\mathbb{1}(T^{(1)} \wedge \tau \leq C) | \mathbf{X}, \mathbf{U}, A, T^{(1)}]} T^{(1)} \wedge \tau \middle| \mathbf{X}, \mathbf{U}\right]\right] \\
 &= \mathbb{E}\left[\mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\mathbb{1}(T^{(1)} \wedge \tau \leq C)}{\tilde{G}(T^{(1)} \wedge \tau | \mathbf{X}, \mathbf{U}, A)} T^{(1)} \wedge \tau \middle| \mathbf{X}, \mathbf{U}\right]\right] \quad \text{by tower property} \\
 &= \mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\mathbb{1}(T \wedge \tau \leq C)}{\tilde{G}(T \wedge \tau | \mathbf{X}, \mathbf{U}, A)} T \wedge \tau\right] \quad \text{by consistency and tower property} \\
 &= \mathbb{E}\left[\frac{A}{e(\mathbf{X}, \mathbf{U})} \cdot \frac{\Delta^\tau}{\tilde{G}(T \wedge \tau | \mathbf{X}, A)} T \wedge \tau\right] \quad \text{by Assumption 5}
 \end{aligned}$$

If $A = 1$ and $\Delta^\tau = 1$, $\tilde{T} \wedge \tau = \min(T, C) \wedge \tau = T \wedge \tau$. Therefore,

$$\text{RMST}^{(1)}(\tau) = \mathbb{E}\left[\frac{A\Delta^\tau}{e(\mathbf{X}, \mathbf{U})\tilde{G}(\tilde{T} \wedge \tau | \mathbf{X}, A = 1)} \tilde{T} \wedge \tau\right].$$

■

DVDS bounds for the RMST can be obtained by replacing $Y_{\text{II},t,1}$ with

$$Y_{\text{RMST},\tau,1} = \frac{\Delta^\tau}{\tilde{G}(\tilde{T} \wedge \tau | \mathbf{X}, A = 1)} \tilde{T} \wedge \tau$$

in Theorem 10. The resulting lower and upper bounds are respectively denoted $\text{RMST}_-^{(1)}(\tau, \Gamma)$ and $\text{RMST}_+^{(1)}(\tau, \Gamma)$. These bounds enjoy the same robustness properties as those cited in Theorem 10 and can be computed the same way as the bounds for Form II.

Appendix D. Experiments

D.1. Complete Simulation Setup

In our experiments, we considered $p_{\mathbf{X}} = 2$ observed confounders and $p_{\mathbf{U}} = 2$ unobserved confounders. We chose $\mathbf{X} \sim \mathcal{U}(-1, 1)^{p_{\mathbf{X}}}$, where $\mathcal{U}(a, b)^p$ is the uniform distribution on (a, b) of dimension p . For j in $\llbracket 1, p_{\mathbf{U}} \rrbracket$, conditionally on $\mathbf{X} = \mathbf{x}$, we designed each unobserved confounder U_j as

$$U_j | \mathbf{X} = \mathbf{x} \sim \lambda G + (1 - \lambda) \beta_j^T \mathbf{x} = \mathcal{N}((1 - \lambda) \beta_j^T \mathbf{x}, \lambda^2),$$

with $G \sim \mathcal{N}(0, 1)$, $0 < \lambda < 1$, and $\beta_j = (\beta_{j,1}, \dots, \beta_{j,p_{\mathbf{X}}}) \in \mathbb{R}_+^{p_{\mathbf{X}}}$. We chose the distribution of A conditionally on $\mathbf{X}$ and $\mathbf{U}$ to be a Bernoulli satisfying the MSM with a true sensitivity parameter Γ^* . We designed the true propensity score $e(\mathbf{X}, \mathbf{U}) = \mathbb{P}(A = 1 | \mathbf{X}, \mathbf{U})$ such that it marginalizes to the nominal propensity score $e(\mathbf{X}) = \mathbb{P}(A = 1 | \mathbf{X}) = \text{logistic}(\delta^T \mathbf{X} + 0.5)$, with $\delta \in \mathbb{R}^{p_{\mathbf{X}}}$. For simplicity, we considered that $\mathbf{U} = \sum_{j=1}^{p_{\mathbf{U}}} U_j / p_{\mathbf{U}}$.

To do so, we chose a true propensity score of the form

$$e(\mathbf{X}, \mathbf{U}) = l(\mathbf{X}) \cdot \mathbb{1}(\mathbf{U} > t(\mathbf{X})) + u(\mathbf{X}) \cdot \mathbb{1}(\mathbf{U} \leq t(\mathbf{X})), \quad (22)$$

where

$$\begin{aligned} l(\mathbf{X}) &= \frac{e(\mathbf{X})}{e(\mathbf{X}) + (1 - e(\mathbf{X}))\Gamma^*} \quad \text{and} \\ u(\mathbf{X}) &= \frac{e(\mathbf{X})}{e(\mathbf{X}) + (1 - e(\mathbf{X}))/\Gamma^*}. \end{aligned}$$

We want to find the threshold $t(\mathbf{X})$ such that $\mathbb{E}[e(\mathbf{X}, \mathbf{U}) | \mathbf{X}] = e(\mathbf{X})$. Therefore, by applying the expectancy conditionally on $\mathbf{X}$ on Equation (22) and using the previous constraint, we get

$$\mathbb{P}(\mathbf{U} \leq t(\mathbf{X}) | \mathbf{X}) = \frac{e(\mathbf{X}) - l(\mathbf{X})}{u(\mathbf{X}) - l(\mathbf{X})}.$$

As we assumed that $\mathbf{U} = \sum_{j=1}^{p_{\mathbf{U}}} U_j / p_{\mathbf{U}}$, the previous equation becomes

$$\mathbb{P}\left(\sum_j U_j / p_{\mathbf{U}} \leq t(\mathbf{X}) \middle| \mathbf{X}\right) = \frac{e(\mathbf{X}) - l(\mathbf{X})}{u(\mathbf{X}) - l(\mathbf{X})}.$$

As the distribution of $U_j | \mathbf{X} = \mathbf{x}$ is Gaussian, if we assume that the U_j are mutually independent conditionally on $\mathbf{X} = \mathbf{x}$, the same applies to $\sum_j U_j / p_{\mathbf{U}}$ conditionally on $\mathbf{X} = \mathbf{x}$, so

$$\sum_j U_j / p_{\mathbf{U}} | \mathbf{X} = \mathbf{x} \sim \mathcal{N}\left(\sum_j (1 - \lambda) \beta_j^T \mathbf{x} / p_{\mathbf{U}}, \lambda^2 / p_{\mathbf{U}}\right).$$

Therefore, the threshold $t(\mathbf{X})$ is the quantile of a normal distribution, which can be found by using the `qnorm` function:

$$t(\mathbf{X}) = \text{qnorm}\left(\frac{e(\mathbf{X}) - l(\mathbf{X})}{u(\mathbf{X}) - l(\mathbf{X})}, \text{mean} = \sum_j (1 - \lambda) \beta_j^T \mathbf{x} / p_{\mathbf{U}}, \text{sd} = \lambda / \sqrt{p_{\mathbf{U}}}\right).$$

The potential outcome $T^{(a)}$, with a in $\{0, 1\}$, was randomly generated using the inverse transform sampling method via a proportional hazards model with Weibull-distributed baseline hazard (Bender et al., 2005). In particular,

$$T^{(a)} = 10 \cdot \left(\frac{-\log(U^{(a)})}{0.95 \cdot \exp(a \log(5) + \beta_{\mathbf{X}}^T \mathbf{X} + \beta_{\mathbf{U}}^T \mathbf{U})} \right)^{1/1.8},$$

where $U^{(a)} \sim \mathcal{U}(0, 1)$. The event time T was then defined as $T = \mathbb{1}(A = 1) \cdot T^{(1)} + \mathbb{1}(A = 0) \cdot T^{(0)}$ and the censoring time C followed a Weibull distribution of shape 6 and scale 10. Finally, we took the minimum between T and C to get the observed time $\tilde{T}$. We generated 20 Monte-Carlo samples, each of size $n = 1000$, and computed the sensitivity bounds for 10 values of time equally spaced between 0.1 and 9. Administrative censoring was applied after the 0.95-quantile of the observed times. Other parameter values used in the simulations are summarized in Table 1.

Parameter	Value
$p_{\mathbf{X}}$	2
$p_{\mathbf{U}}$	2
n	1000
Γ^*	3
λ	0.9
β_j	$\mathcal{U}(0, 1)^{p_{\mathbf{X}}}$
$\beta_{\mathbf{X}}$	$\log(\mathcal{U}(1.1, 1.3)^{p_{\mathbf{X}}})$
$\beta_{\mathbf{U}}$	$\log(\mathcal{U}(0.5, 0.8)^{p_{\mathbf{U}}})$
δ	$\mathcal{U}(-0.5, 0.5)^{p_{\mathbf{X}}}$

Table 1: Parameter values used in the simulations. $\mathcal{U}(a, b)^p$ means that the vector of size p was generated randomly according to a uniform distribution on (a, b) . $\log(\mathcal{U}(a, b)^p)$ means that the logarithm of each of the p components in the random sample was taken.

D.2. Additional Information on Real Data

D.2.1. GERMAN BREAST CANCER STUDY GROUP DATASET

We included 7 covariates in our analysis: **age** (the age, in years), **size** (the tumor size, in mm), **nodes** (the number of positive lymph nodes), **pgr** (the concentration of progesterone receptors, in fmol/L), **er** (the concentration of estrogen receptors, in fmol/L), **meno** (the menopausal status, i.e. 0 if premenopausal, and 1 if postmenopausal), and **grade** (the tumor grade, i.e. 1, 2, or 3). Categorical variables were binary encoded. The treatment A is the **hormon** variable, the observed time $\tilde{T}$ is the **rfstime** variable, and the censoring status variable Δ is the **status** variable. More details on the dataset can be found on <https://www.kaggle.com/datasets/utkarshx27/breast-cancer-dataset-used-royston-and-altman>.

D.2.2. RIGHT HEART CATHETERIZATION DATASET

The list and meaning of each variable in the original data can be found on <https://hbiostat.org/data/repo/rhc>.

D.3. Additional Results

D.3.1. RESULTS ON SYNTHETIC EXPERIMENTS

Survival Function. Results of sensitivity analysis for the survival functions and the difference of survival functions are given in Figure 5 for our DVDS bounds with Form I (in blue) and II (in green), and for the bounds from Lee et al. (2024) (in red). These results were obtained with the same setup as described in Section 4.2.

When $\Gamma = 3$, that is, the same value Γ^* that generated the data, the intervals almost always contain the true value of the survival or difference of survival functions. As pointed out in Section 4, remark that the DVDS bounds for Form II of the survival function are unstable for high values of time t . In particular, the upper bound obtained on the control group increases even if this should not happen, because the survival function is a decreasing function over time (Figure 5(b)subfigure). This can be explained by near violations of the positivity of the censoring process. On the contrary, the DVDS bounds for Form I remain stable.

To check that our bounds are scalable to high dimensional datasets, we applied our approach to simulated data of sample size 5000 with 2 observed confounders and 2 unobserved confounders (Figure 6), and sample size 3000, with 40 observed confounders and 20 unobserved confounders (Figure 7). In both cases, we used 100 Monte Carlo samples. We parallelized the implementation on 9 CPUs and obtained an execution time of 9.74 hours for the difference of survival function in the second experiment with high dimensional variables. Our results show that the method is still valid, with conclusions that are similar to what we presented in Section 4.2 and with boxplots of a similar width. Therefore, we conclude that our method is scalable to high dimensional datasets. Of course, as there are four nuisance parameters to estimate (propensity score, modified outcome regression, conditional quantile, conditional censoring function), each depending on the observed confounders $\mathbf{X}$, the higher the dimension of $\mathbf{X}$ is, the longer the execution time will be.

We also added two experiments where the true sensitivity parameter Γ^* is equal to 1.01 (near unconfoundedness) and 5 (higher confounding strength), and we kept the sensitivity parameter Γ given to our method equal to 3. As expected, under near unconfoundedness ($\Gamma^* = 1.01$), our methodology and the one from Lee et al. (2024) are less biased: in Figure 8, the boxplots of the values under unconfoundedness are closer to the true estimand and the bounds are centered around these true values. When the true sensitivity parameter is higher ($\Gamma^* = 5$) but $\Gamma = 3$ (Figure 9), the boxplots of the values under unconfoundedness move away from the true value. If Γ was too small (for instance, 1.5), many intervals would not contain the true estimand anymore. Therefore, Γ should be large enough for the bounds to contain the true treatment effect.

Finally, we compared the methods in the case of informative censoring, where the censoring time C depends on $\mathbf{X}$ and A (and we could also add $\mathbf{U}$). We do not present the bounds obtained with Form II because they diverged. C was generated the same way as the event time T , with

$$C^{(a)} = 7 \cdot \left(\frac{-\log(U^{(a)})}{0.95 \cdot \exp(a \log(5) + \beta_{\mathbf{X},C}^T \mathbf{X})} \right)^{1/6},$$

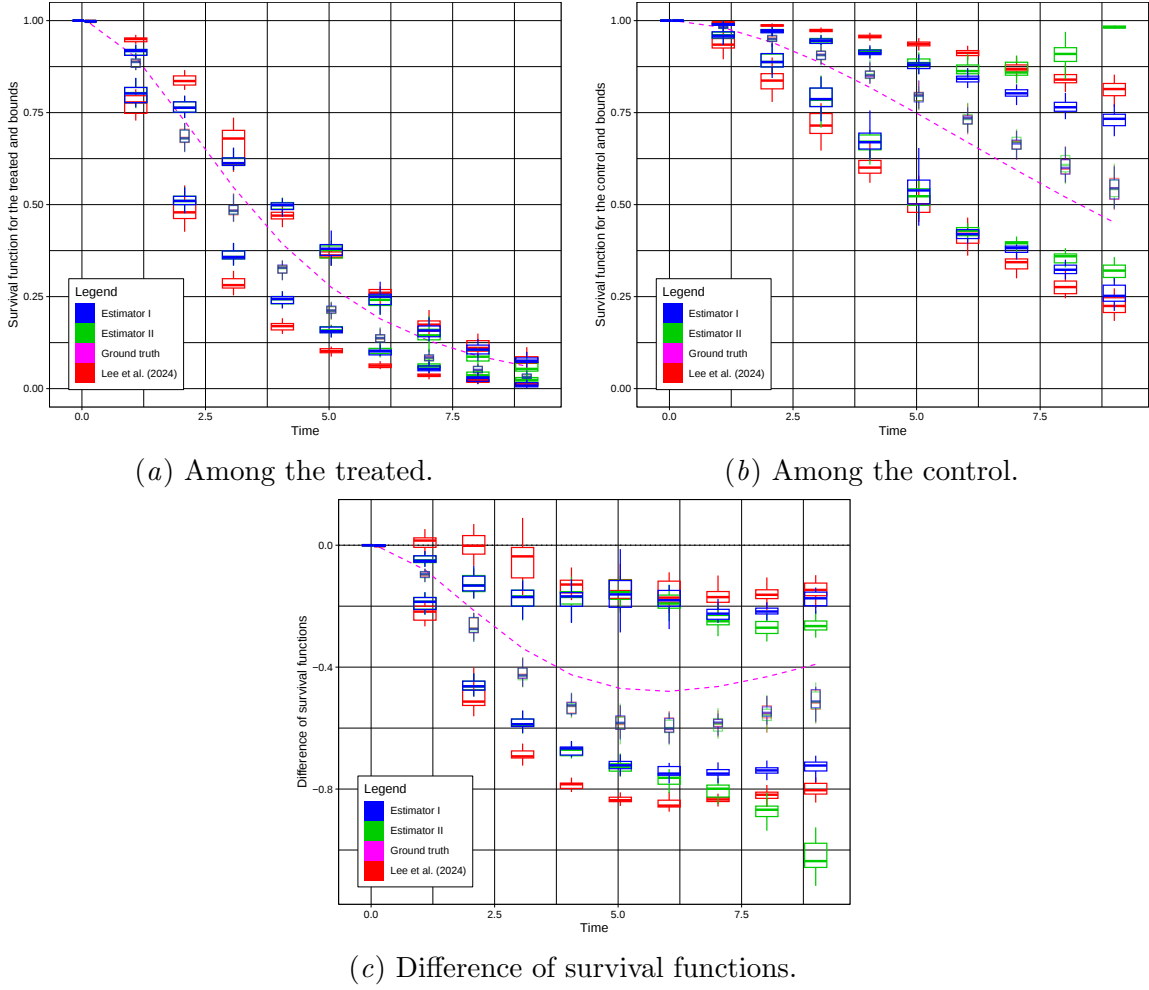


Figure 5: Sensitivity analysis for the survival function among the treated (5(a)subfigure), the control (5(b)subfigure) and the difference of survival functions (5(c)subfigure) **on the simulated data, as described in Section 4.2**. The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue), Form II (in green), and the estimator from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

for $a \in \{0, 1\}$, where $U^{(a)} \sim \mathcal{U}(0, 1)$ and $\beta_{\mathbf{X}, C} = \mathcal{U}(5.0, 5.5)^{p \times}$. Figure 10 shows that both methods stay valid when $\Gamma^* = \Gamma = 3$, but the boxplots of the two methods do not overlap under ignorability, as expected. Note that our method is less biased when the estimand is the difference of survival functions (Figure 10(c)subfigure), but at the cost of potentially larger intervals in practice. Indeed, we can see that, for high values of time, such as around

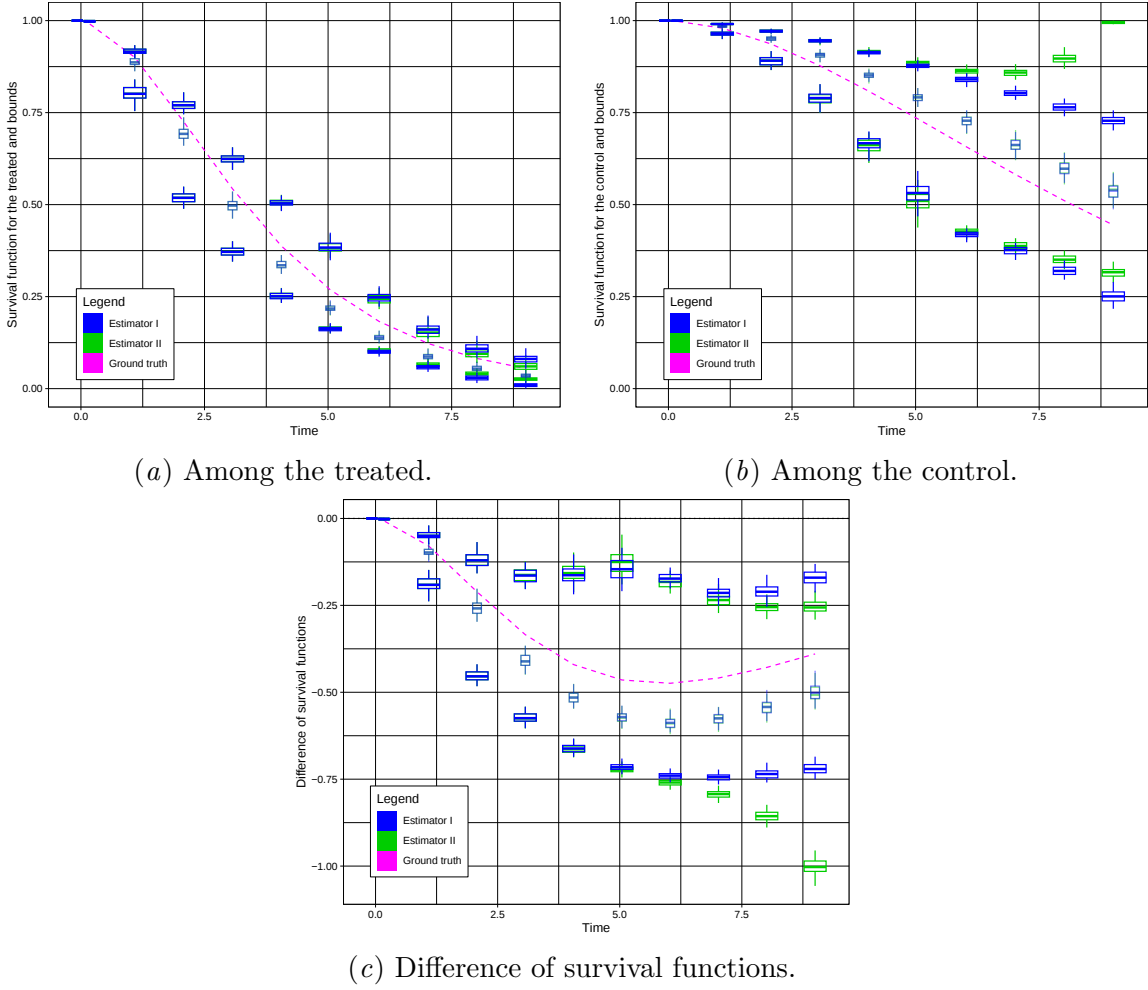


Figure 6: Sensitivity analysis for the survival function among the treated (6(a)subfigure), the control (6(b)subfigure) and the difference of survival functions (6(c)subfigure) **on high dimensional simulated data** ($n = 5000$, $p_X = 2$, and $p_U = 2$). The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue) and Form II (in green) on 100 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

7.5 time units, our method yields larger intervals than the method of Lee et al. (2024). This can be due to the fact that our bounds do not include a stabilization with respect to the censoring function $\bar{G}$, while the bounds of Lee et al. (2024) do not include $\bar{G}$ at all.

RMST and Difference in RMST. We parallelized computations between different values of time horizon τ for our DVDS bounds.

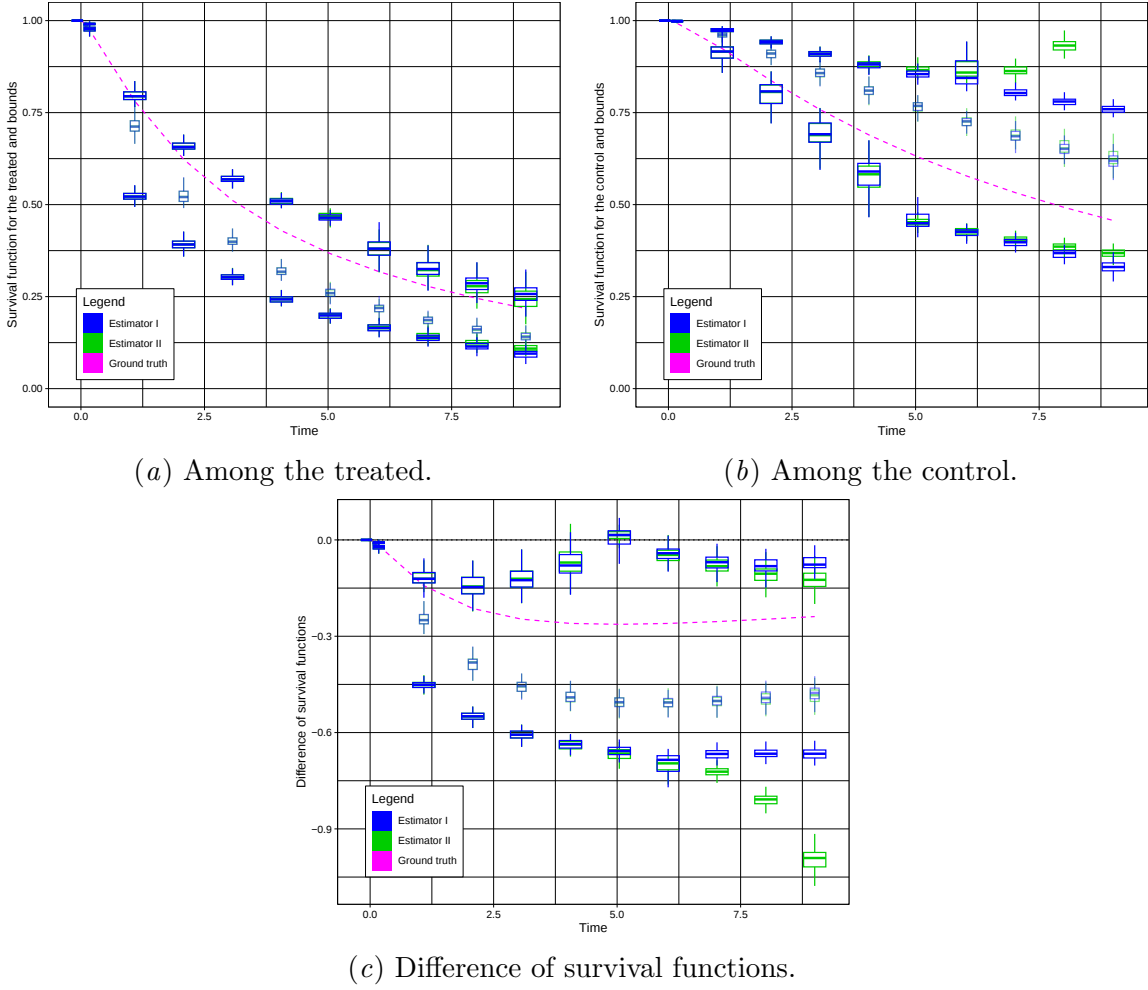


Figure 7: Sensitivity analysis for the survival function among the treated (7(a)subfigure), the control (7(b)subfigure) and the difference of survival functions (7(c)subfigure) **on high dimensional simulated data** ($n = 3000$, $p_X = 40$, and $p_U = 20$). The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue) and Form II (in green) on 100 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

For the RMST, Lee et al. (2024) originally computed the area under the survival curve via the left point rectangle method whereas we implemented the middle point rectangle method because it converges slightly faster towards the true integral. To compute the area under the curve, we used all observed time points between 0 and the time horizon τ .

Results of sensitivity analysis for the RMST and the difference in RMST are given in Figure 11 for our DVDS bounds (Form I in blue and Form II in green) and the bounds

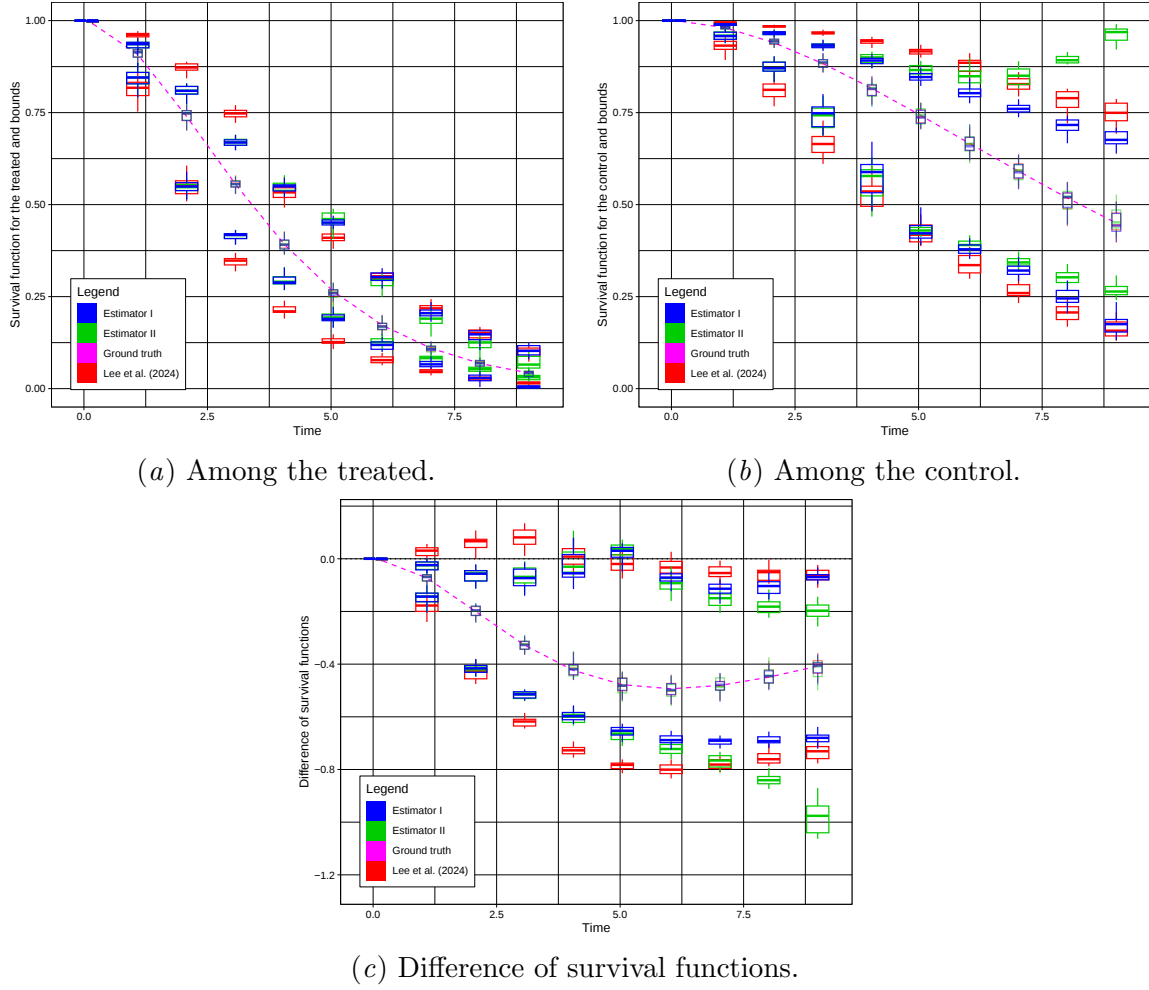


Figure 8: Sensitivity analysis for the survival function among the treated (8(a)subfigure), the control (8(b)subfigure) and the difference of survival functions (8(c)subfigure) **on the simulated data when the true sensitivity parameter is $\Gamma^* = 1.01$.** The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

from Lee et al. (2024) (in red). Under ignorability (narrow and transparent boxplots), the estimations deviate slightly from the ground truth (dotted lines in magenta) because of unobserved confounders, but are still close to the true value. Note that the three methods give approximately the same estimation.

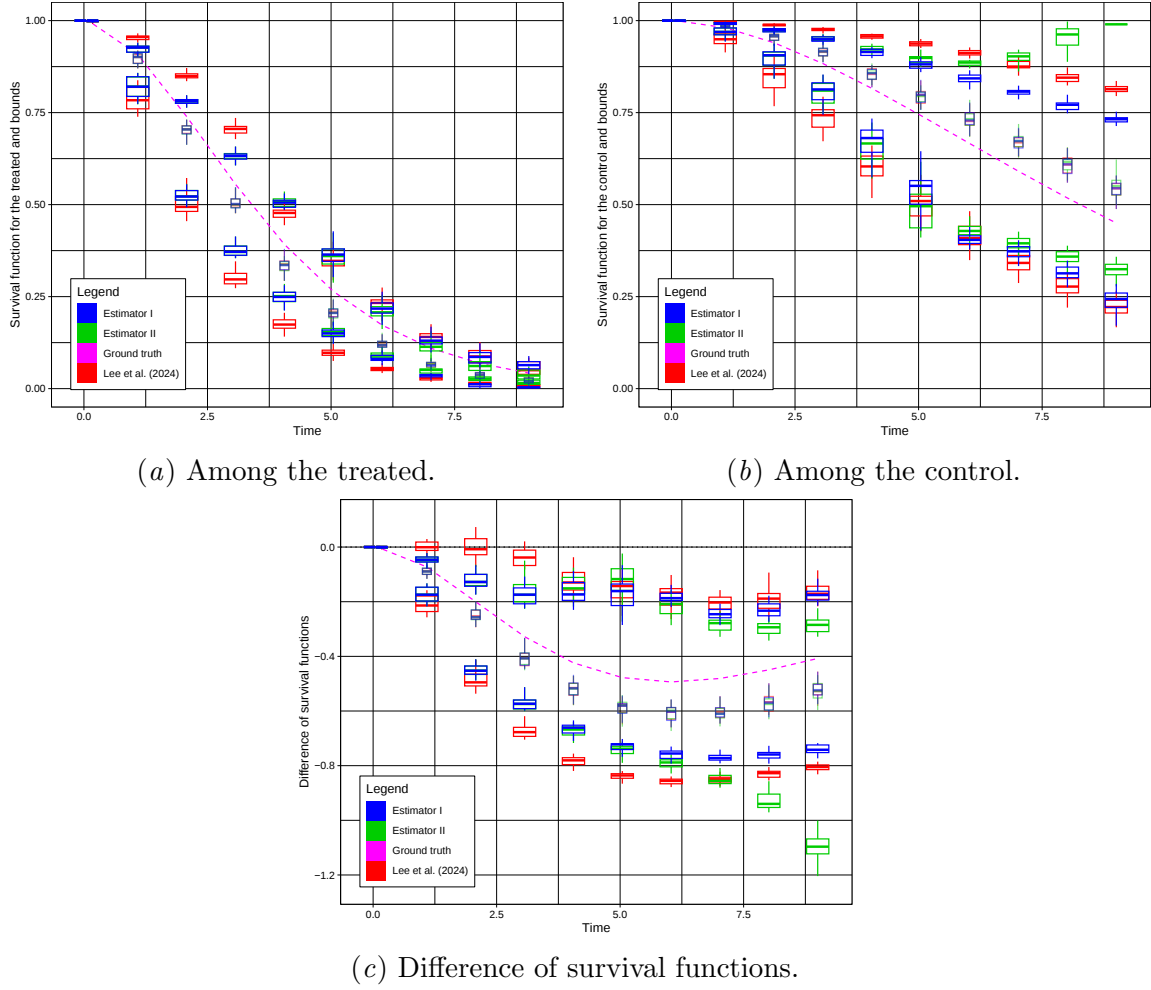


Figure 9: Sensitivity analysis for the survival function among the treated (9(a)subfigure), the control (9(b)subfigure) and the difference of survival functions (9(c)subfigure) **on the simulated data when the true sensitivity parameter is $\Gamma^* = 5$** . The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

When $\Gamma = 3$, that is, the same value Γ^* that generated the data, the intervals almost always contain the true value of the RMST or difference in RMST.

The sharpness property is less apparent with the difference in RMST (Figure 11(c)subfigure) than with the difference in survival functions (Figure 5(c)subfigure), especially with high values of time t , where an offset between our bounds and the bounds of the comparator

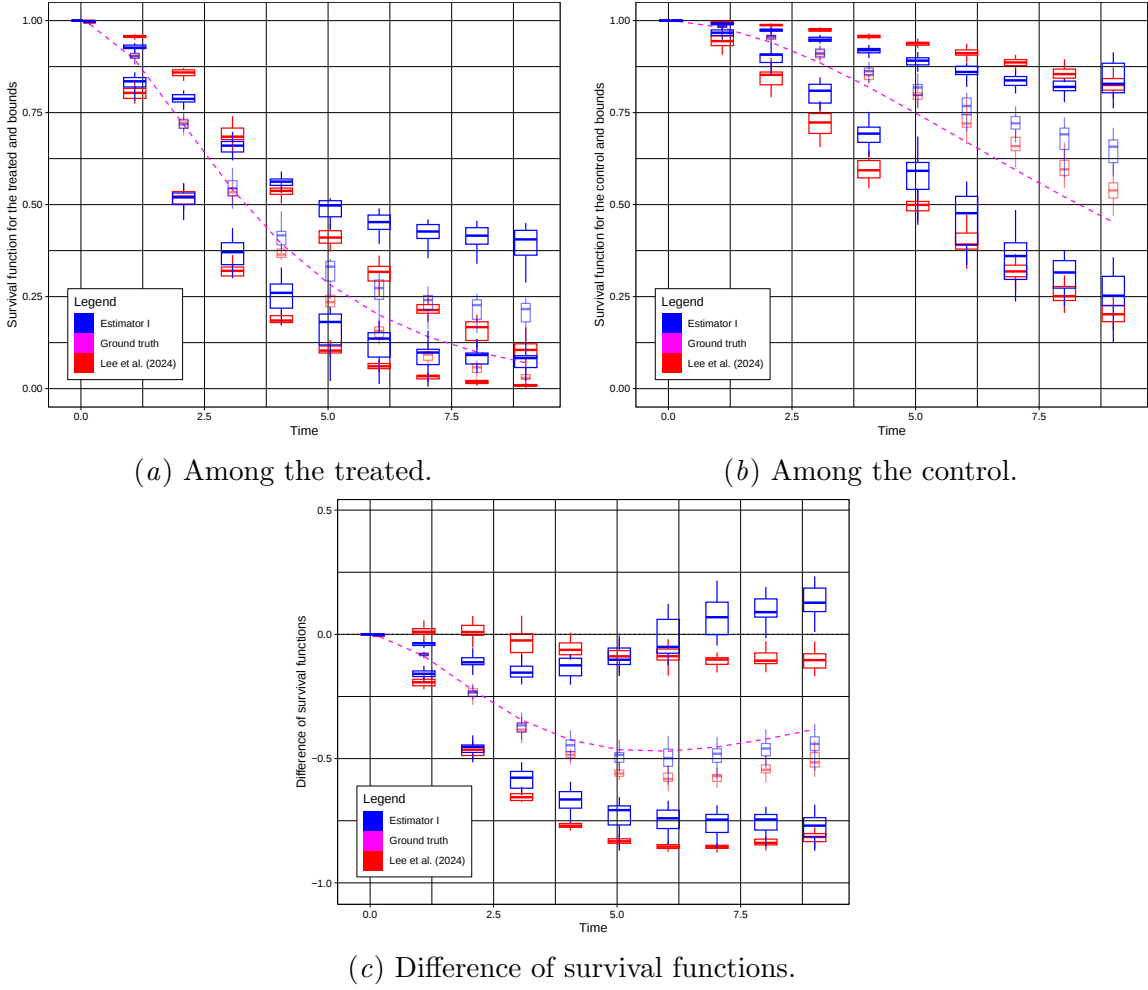


Figure 10: Sensitivity analysis for the survival function among the treated (10(a)subfigure), the control (10(b)subfigure) and the difference of survival functions (10(c)subfigure) **on the simulated data in the presence of informative censoring**. The dotted lines in magenta are the true survival functions and difference of survival functions. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) with Form I (in blue) and the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

appears. This could be explained by two facts: our DVDS bounds were written for the survival functions and then integrated with a rectangle method, whereas Lee et al. (2024) directly derived bounds for the RMST; moreover, to simplify, the simulation setup was based on the assumption of non-informative censoring, which was supposed by Lee et al. (2024), whereas we assumed informative censoring and introduced $\bar{G}$.

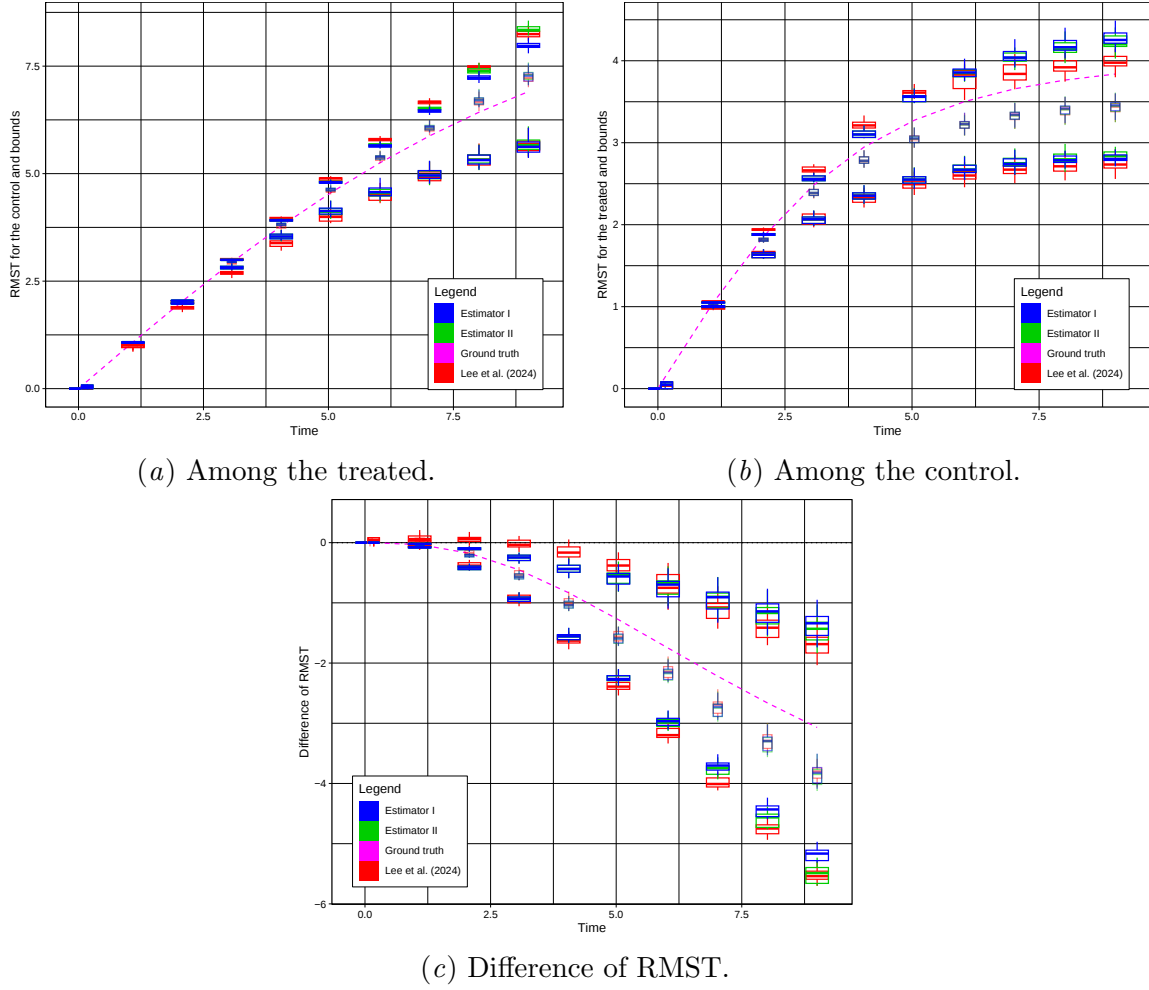


Figure 11: Sensitivity analysis for the RMST among the treated (11(a)subfigure), the control (11(b)subfigure) and the difference in RMST (11(c)subfigure) on the simulated data. The dotted lines in magenta are the true RMST and difference in RMST. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) **via integration by the rectangle method** with Form I (in blue) and Form II (in green), and with the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

Our DVDS bounds, and especially the ones for Form I, show good coverage of the real (difference in) RMST, and sharpness with respect to the bounds from Lee et al. (2024). Moreover, our bounds also enjoy shorter computation times, but, as compared with bounds for the survival function, the computation time gap is less important because of the integration by the rectangle method and the need to compute the PEIs *for each observed time point* between 0 and the time horizon τ with our method. The method by Lee et al. (2024)

took 9 hours and 54 minutes to complete, whereas our DVDS bounds (for Forms I and II simultaneously) required 8 hours and 48 minutes. Our method is still more rapid but the gain factor is only 1.1. Note that Lee et al. (2024) developed an approximate method which is faster than direct optimization, but at the cost of simplifying assumptions.

As seen in Appendix C.2, it is possible to implement faster DVDS bounds for the RMST (a few minutes with our bounds versus several hours with the direct optimization method from Lee et al. (2024)). However, in practice, they can perform poorly because they are more conservative than the bounds obtained via integration of the survival function. Indeed, in Figure 12, we can see that our bounds are tighter than the bounds from Lee et al. (2024) until a certain time horizon, from which several observations start to be censored. We tried to stabilize the bounds, with no improvement.

D.3.2. RESULTS ON REAL DATA

Calibration of Γ by informal benchmarking. We calibrated the sensitivity parameter Γ using an informal benchmarking approach where we took the maximum between the maximum odds ratio of the MSM and the inverse of the minimum odds ratio of the MSM with and without each observed covariate. The results are summarized in Tables 3 and 2 for, respectively, the RHC and the GBCSG data.

In the RHC data, there were 47 covariates that were included, among which some were binary encoded. The variable “immunhx” (immunosuppression) corresponds to the minimum estimated value of Γ , $\hat{\Gamma}_{\min} = 1.07$, whereas “paf1” (PaO2/FIO2 ratio) corresponds to the maximum estimated value of Γ , $\hat{\Gamma}_{\max} = 35.6$. The median value is 1.63. We already identified a critical value $\Gamma_{c,DVDS} = 1.2$ at day 30 with our method and a critical value $\Gamma_{c,Lee} = 1.15$ with the method of Lee et al. (2024). 8 covariates have a corresponding $\hat{\Gamma}$ lower than $\Gamma_{c,Lee}$ (covariates on income, race, immunosuppression, sex, or myocardial infarction) and 9 covariates have a corresponding $\hat{\Gamma}$ lower than $\Gamma_{c,DVDS}$ (the same as before and age). As the majority of observed covariates has a corresponding $\hat{\Gamma}$ that is greater than the critical values, the results are not robust to unobserved confounders and we cannot conclude on a negative effect of RHC on survival at day 30 because small confounders would be enough to explain away a causal association.

In the GBCSG data, there were 7 covariates that were included, among which some were binary encoded. The variable “grade_2” (tumor of grade 2) has the lowest associated $\hat{\Gamma}$, $\hat{\Gamma}_{\min} = 1.26$, and the variable “er” (estrogen receptor, in fmol/L) has the highest associated $\hat{\Gamma}$, $\hat{\Gamma}_{\max} = 2.75$. The median value is 2.03. The estimated $\hat{\Gamma}$ associated with tumor size (“size” variable) reported by Lee et al. (2024) is equal to 1.17 whereas we estimated a value of 2.05 because we used 5-fold cross-fitting. With 95%-level confidence intervals, we identified a critical value $\Gamma_{c,I} = 1.09$ at 5 years with our Form I estimator and a critical value $\Gamma_{c,Lee} = 1.08$ with the method of Lee et al. (2024). These two critical values are lower than 2.05, which means that we cannot conclude about a positive effect of tamoxifen and chemotherapy on recurrence-free survival at 5 years when compared to chemotherapy only.

Survival Function. Figures 13 and 14 were obtained with the same setup as described in Section 4.3, respectively, on the RHC and GBCSG data.

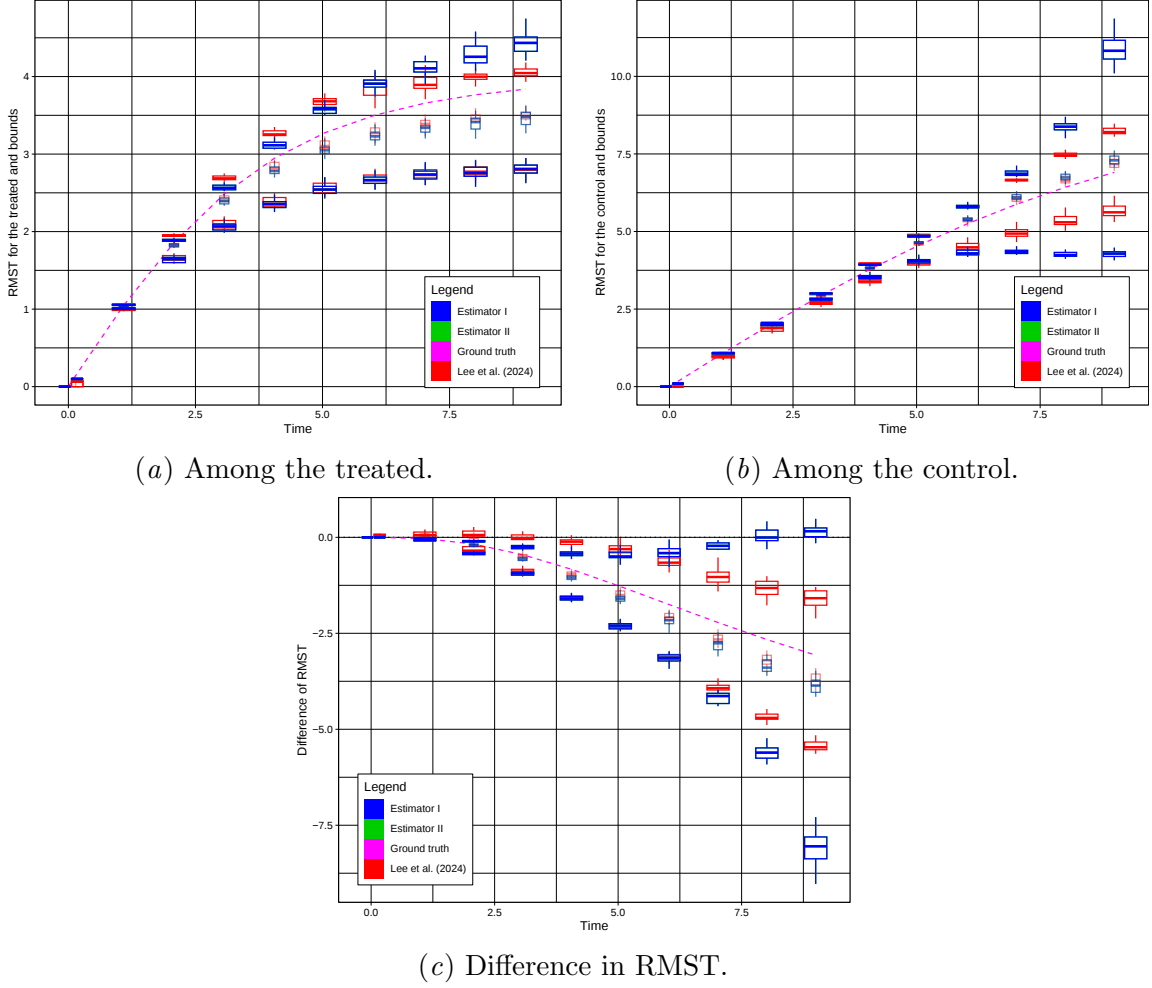


Figure 12: Sensitivity analysis for the RMST among the treated (12(a)subfigure), the control (12(b)subfigure) and the difference in RMST (12(c)subfigure) on the simulated data. The dotted lines in magenta are the true RMST and difference in RMST. The large boxplots correspond to the estimated upper and lower sensitivity bounds (PEI) **with our “fast RMST” method** on Form I (in blue) and Form II (in green), and with the method from Lee et al. (2024) (in red) on 20 Monte-Carlo samples, for $\Gamma = 3$. The narrower and transparent boxplots correspond to the value under ignorability ($\Gamma = 1$) for each method.

RMST and Difference in RMST. The results of the sensitivity analyses on the difference in RMST in the RHC and GBCSG datasets are given, respectively, in Figures 16 and 17, for a sensitivity parameter Γ equal to 1.15, 1.2, and 1.5 (and an additional value of 1.08 for the GBCSG dataset). For the RHC dataset, the PEIs obtained via the DVDS bounds for Form I (in blue), II (in green) and via the method of Lee et al. (2024) (in red) were computed for 5 equally spaced values of time between 2 and 50 days, to which we also

Covariate	$\hat{\Gamma}$	Covariate	$\hat{\Gamma}$
immunhx	1.07	seps_Yes	1.63
sex_Male	1.08	gastr_Yes	1.65
income_\$25-\$50k	1.11	psychhx	1.71
income_\$11-\$25k	1.11	hema_Yes	1.76
amihx	1.12	sod1	1.79
race_other	1.13	resp_Yes	1.81
income_Under \$11k	1.14	liverhx	1.99
race_white	1.15	dnr1_Yes	1.99
age	1.19	gibledhx	2.04
chrpulhx	1.24	ca_No	2.20
ca_Yes	1.24	bili1	2.24
scoma1	1.24	neuro_Yes	2.25
temp1	1.35	hema1	2.31
meta_Yes	1.37	hrt1	2.37
ninsclas_Medicare	1.38	wtkilo1	2.47
malighx	1.38	crea1	2.74
cardiohx	1.38	ph1	2.98
chfhx	1.39	pot1	3.10
ninsclas_Medicare & Medicaid	1.48	card_Yes	3.59
edu	1.50	aps1	3.80
transhx	1.56	trauma_Yes	4.18
wble1	1.57	resp1	4.78
ninsclas_No insurance	1.57	meanbp1	4.86
renal_Yes	1.58	ortho_Yes	8.38
ninsclas_Private	1.59	paco21	8.49
renalhx	1.60	alb1	19.51
ninsclas_Private & Medicare	1.61	pafi1	35.58
dementhx	1.63		

Table 2: Informal benchmarking on the RHC data. Variables with an underscore in their name correspond to binary encoded variables.

added 30 days. For the GBCSG dataset, we used 5 equally spaced values between 8 and 1826.25 days. The estimates under ignorability ($\Gamma = 1$) were represented with dotted lines.

For the RHC data, observe that, under ignorability, the difference in RMST is always negative, regardless of the time point between 0 and 50 days. This is consistent with findings from [Connors et al. \(1996\)](#) on the negative effect of RHC on survival. In particular, if no confounders were hidden, Figure 16 would suggest that RHC reduces survival by approximately 1 day on a time horizon of 30 days. For the GBCSG data, under ignorability, the difference in RMST is almost always positive (except at time 463 days for [Lee et al. \(2024\)](#)).

For the RHC data, at day 30, we reach similar conclusions to the ones obtained with survival functions. For the GBCSG data, at 5 years, the critical value is 1.2 with Form II,

Covariate	$\hat{\Gamma}$
grade_2	1.26
grade_3	1.37
pgr	1.50
age	2.00
size	2.05
meno	2.18
nodes	2.50
er	2.75

Table 3: Informal benchmarking on the GBCSG data

1.15 with Form I, and even lower with the method from [Lee et al. \(2024\)](#). Bounds with Form I are tighter than our comparator, and bounds with Form II are shifted toward higher values.

Computation time. Tables 4 and 5 provide the execution times for each experiment on real-world data (RHC and GBCSG data).

With the RHC data ($n = 5735$ and $p_{\mathbf{X}} = 47$), our method is faster than the method of [Lee et al. \(2024\)](#) by at least a factor of 1.19 for the difference of survival function, and by at least a factor of 1.87 for the difference in RMST. Moreover, the execution time stays consistent with our method when Γ changes whereas it varies more with the method of [Lee et al. \(2024\)](#) because of the optimization step.

With the GBCSG data ($n = 686$ and $p_{\mathbf{X}} = 7$), our method is faster than the method of [Lee et al. \(2024\)](#) for the difference of survival function by at least a factor of 6.90 but is longer for the difference in RMST by at least a factor of 20.0. This longer execution time can be explained by the integration step in our method, because we computed Forms I and II simultaneously, and because the optimization step of [Lee et al. \(2024\)](#) is well suited to the RMST when the sample size is not too large (a few hundreds in the GBCSG data vs. a few thousands in the RHC data).

Estimand	Γ	Exec. time for Estim. I and II	Exec. time for Lee et al. (2024)
Survival function	1.15	1.26	1.50
Survival function	1.2	1.21	1.60
Survival function	1.5	1.20	1.85
RMST	1.15	0.53	0.99
RMST	1.2	0.51	1.15
RMST	1.5	0.51	1.33

Table 4: Comparison of execution times (in hours) on the RHC data for the difference in survival functions and the difference in RMST.

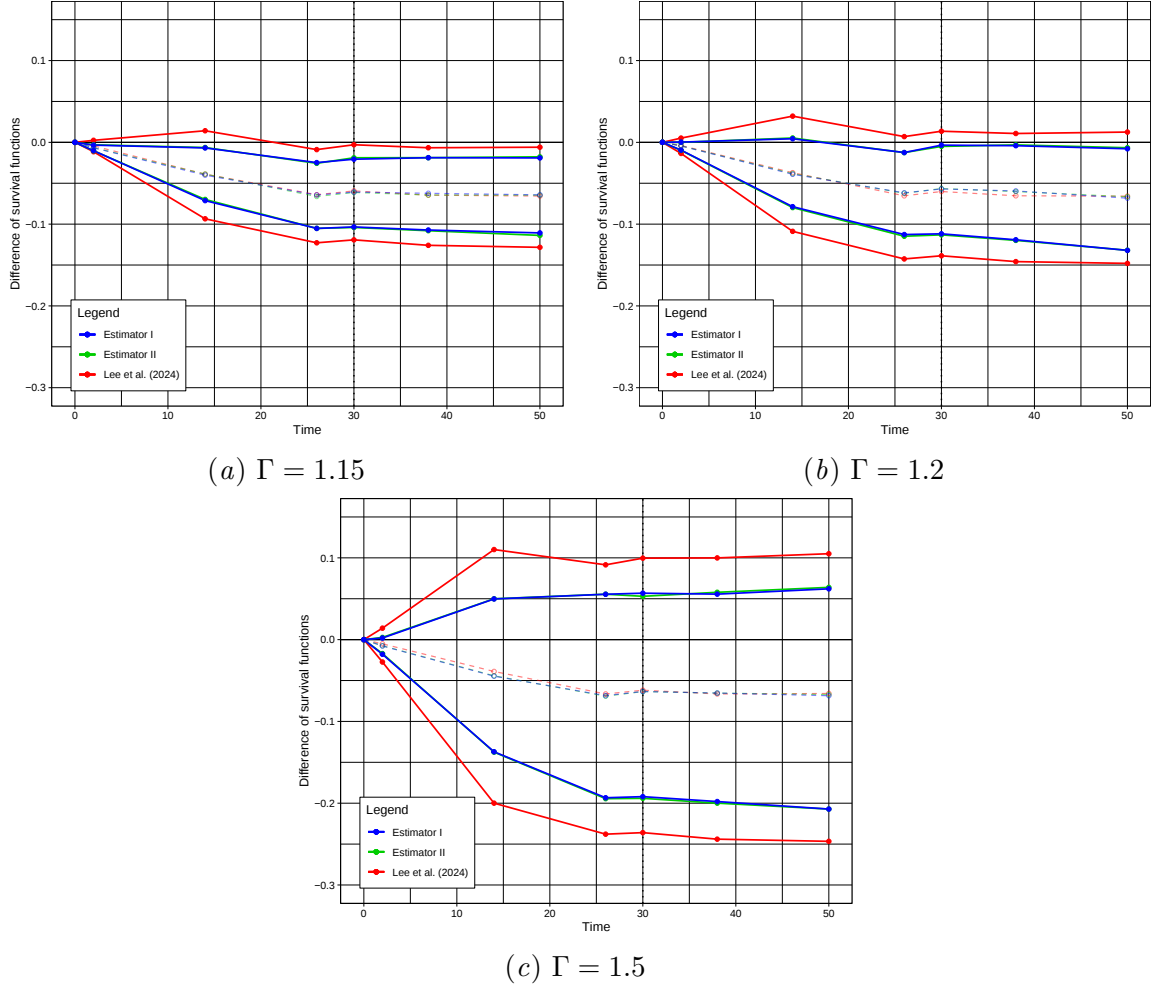


Figure 13: Sensitivity analysis for the difference of survival functions for $\Gamma = 1.15$ (13(a)subfigure), 1.2 (13(b)subfigure), and 1.5 (13(c)subfigure) on the RHC data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical dotted black line indicates a time of 30 days. Note that the bounds with Form II are almost coincident with the bounds with Form I.

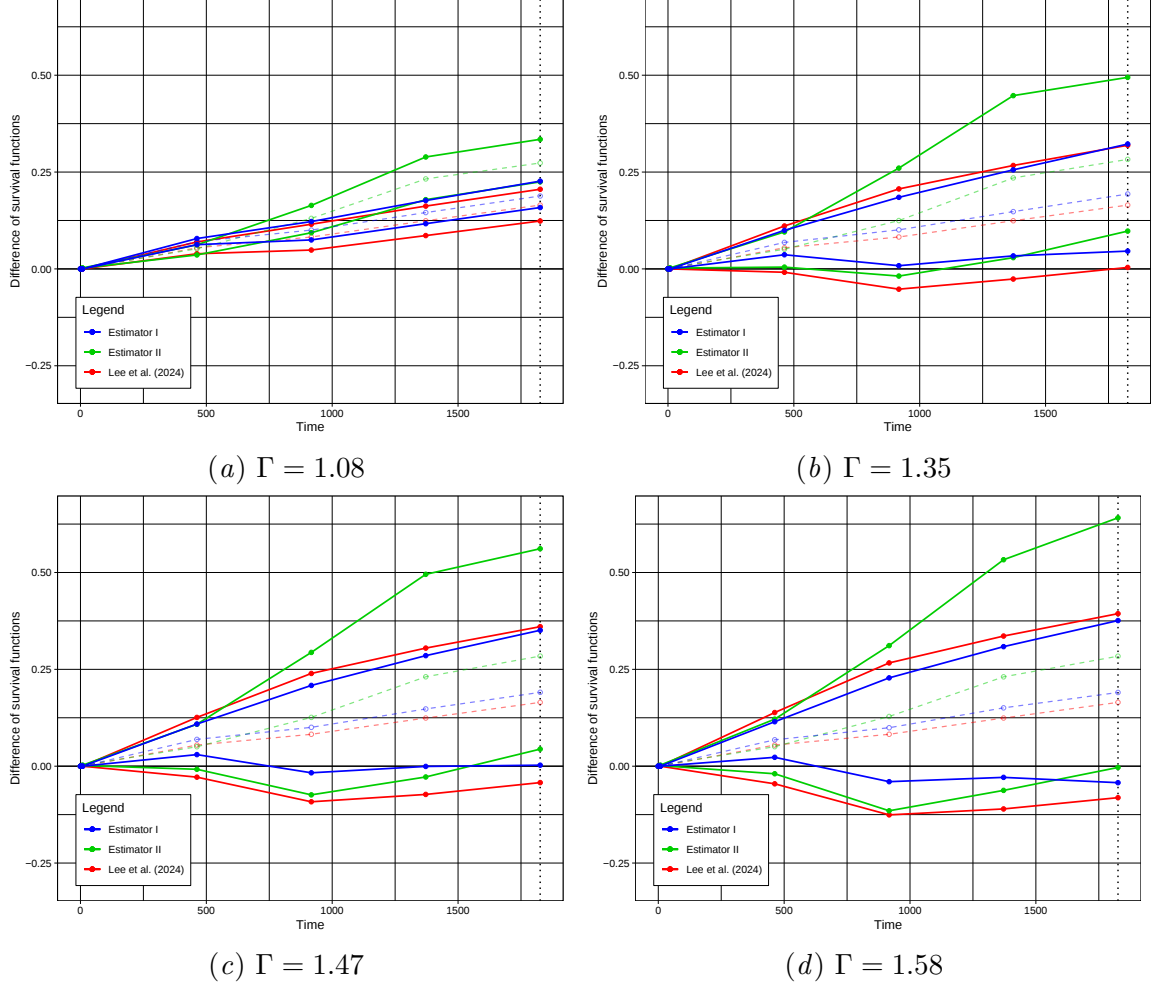


Figure 14: Sensitivity analysis for the difference of survival functions for $\Gamma = 1.08$ (14(a)subfigure), $\Gamma = 1.35$ (14(b)subfigure), $\Gamma = 1.47$ (14(c)subfigure), and $\Gamma = 1.58$ (14(d)subfigure) on the GBCSG data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical black dotted line indicates a time of 1826.25 days (5 years).

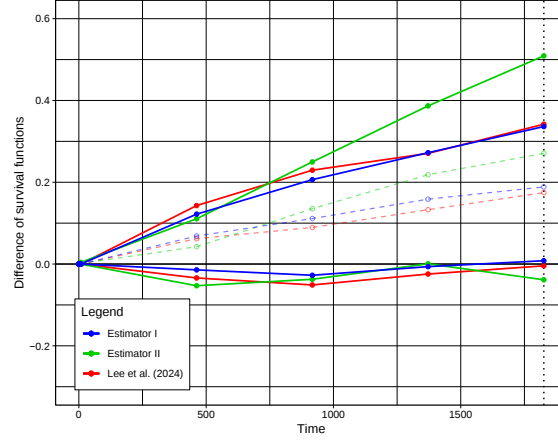


Figure 15: Sensitivity analysis for the difference of survival functions for $\Gamma = 1.08$ on the GBCSG data. The curves correspond to the 95%-CIs obtained with Form I (in blue), Form II (in green), and the method from [Lee et al. \(2024\)](#) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical black dotted line indicates a time of 1826.25 days (5 years).

Estimand	Γ	Exec. time for Estim. I and II	Exec. time for Lee et al. (2024)
Survival function	1.15	25.9 s	2.98 min
Survival function	1.2	22.2 s	3.24 min
Survival function	1.5	20.9 s	3.68 min
RMST	1.15	11.2 min	27.4 s
RMST	1.2	11.1 min	27.3 s
RMST	1.5	11.0 min	33.2 s

Table 5: Comparison of execution times on the GBCSG data for the difference in survival functions and the difference in RMST.

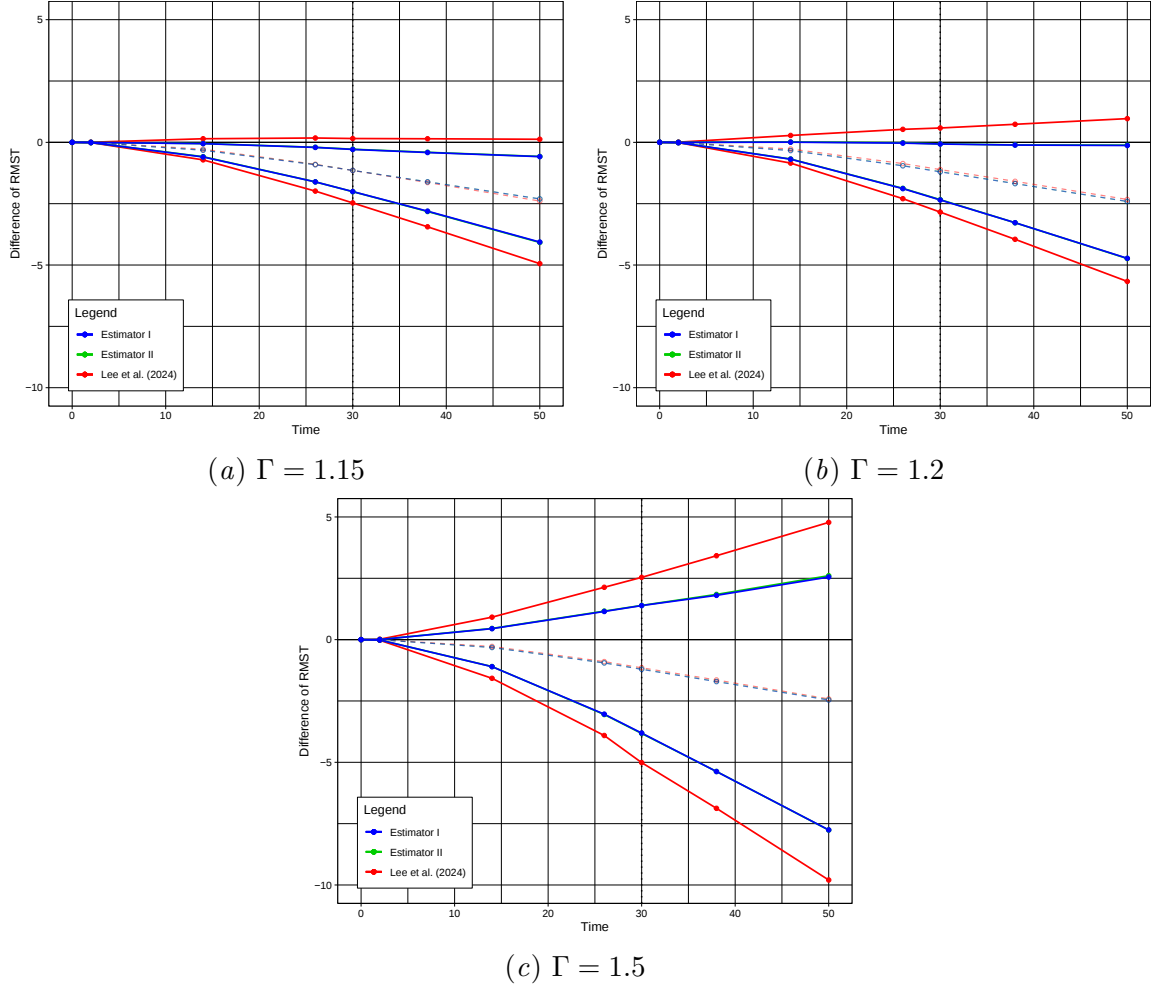


Figure 16: Sensitivity analysis for the difference in RMST for $\Gamma = 1.15$ (16(a)subfigure), 1.2 (16(b)subfigure), and 1.5 (16(c)subfigure) on the RHC data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical black dotted line indicates a time horizon of 30 days. Note that the bounds with Form II are almost coincident with the bounds with Form I.

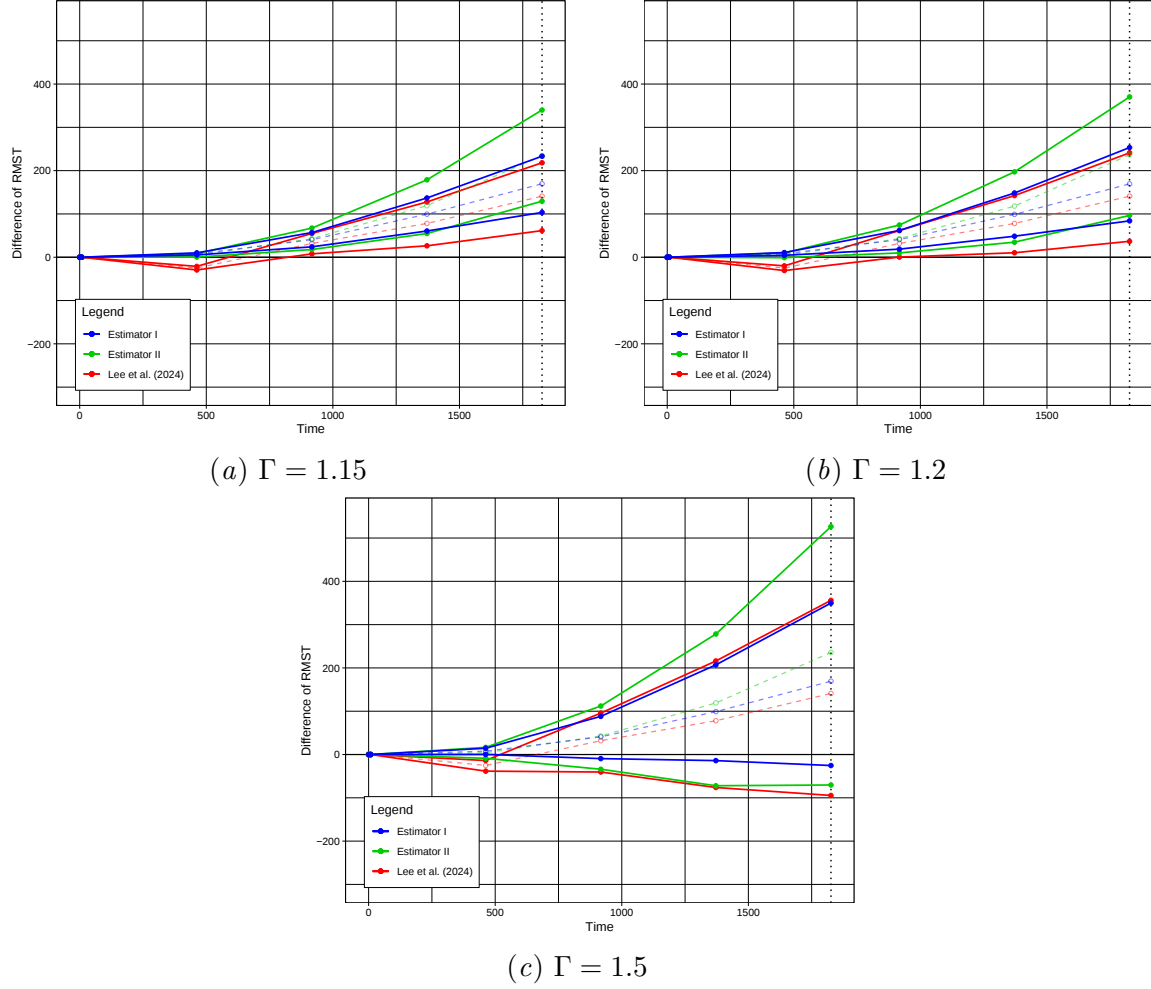


Figure 17: Sensitivity analysis for the difference in RMST for $\Gamma = 1.15$ (17(a)subfigure), 1.2 (17(b)subfigure), and 1.5 (17(c)subfigure) on the GBCSG data. The curves correspond to the lower and upper sensitivity bounds (PEI) obtained with Form I (in blue), Form II (in green), and the method from Lee et al. (2024) (in red). The dotted and transparent lines correspond to the value under ignorability ($\Gamma = 1$) for each method. A vertical black dotted line indicates a time horizon of 1826.25 days (5 years).