On the estimation of average treatment effects with right-censored time to event outcome and competing risks

Brice Maxime Hugues Ozenne¹,² | Thomas Harder Scheike¹ | Laila Stærk³ | Thomas Alexander Gerds¹,⁴

¹Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark
²Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet, Copenhagen, Denmark
³Department of Cardiology, Copenhagen University Hospital Herlev and Gentofte, Hellerup, Denmark
⁴Danish Heart Foundation, Copenhagen, Denmark

Abstract
We are interested in the estimation of average treatment effects based on right-censored data of an observational study. We focus on causal inference of differences between \( r \)-year absolute event risks in a situation with competing risks. We derive doubly robust estimation equations and implement estimators for the nuisance parameters based on working regression models for the outcome, censoring, and treatment distribution conditional on auxiliary baseline covariates. We use the functional delta method to show that these estimators are regular asymptotically linear estimators and estimate their variances based on estimates of their influence functions. In empirical studies, we assess the robustness of the estimators and the coverage of confidence intervals. The methods are further illustrated using data from a Danish registry study.

Keywords
Cox regression model, hazard ratio, probabilistic index, relative risk, survival analysis

1 | INTRODUCTION

Average treatment effects (ATEs) are important parameters in epidemiology (Hernán & Robins, 2006; Robins, 1986). In observational studies, these parameters are interpreted in a suitable framework for causal inference (Hernán & Robins, 2019; Pearl, 2000) as what one would have observed had the treatment been randomized. Estimators of ATE include outcome regression model-based estimators, which standardize the expected outcome to a given distribution of the confounders (G-formula), inverse probability of treatment weighted (IPTW) estimators, which rely on a model for the propensity of treatment, and doubly robust estimators, which combine the two types of estimators with the aim to reduce bias (Glynn & Quinn, 2010; Kang & Schafer, 2007; van der Laan & Robins, 2003).

In this article, we are motivated by applications in pharmacoepidemiology, where the aim is to evaluate differences between alternative drug treatments based on large-scale registry data (Hernán & Robins, 2016). We are particularly interested in applications where the outcome is a right-censored time to event since a common time origin, and death without the outcome is a competing risk. To define our parameter of interest, we assume that, at the time origin, subjects are assigned to one of two treatment groups. We evaluate the risk of the event at a fixed prediction time horizon, say \( r \)-years after the time origin.
main target parameter is the ATE defined as the difference between the absolute \( \tau \)-year risks of the event of interest in the two treatment groups assigned at the time origin. For this parameter, we aim to achieve an interpretation in the counterfactual world where the treatment assignment is randomized. This setup is appropriate for one time treatments, such as aortic valve replacement with two alternative prostheses (Aasbjerg et al., 2019). The setup can also be used to evaluate differences in the initial treatment of dynamic treatment regimes, where subjects can switch between treatment groups after the time origin. In our illustrative example, we consider initiation with anticoagulation treatment using two alternative drugs. Our setup does not consider dynamic treatment regimes (Yavuz, Chng, & Wahed, 2018) or time-dependent covariates (Bekaert, Vansteelandt, & Mertens, 2010; Moodie, Stephens, & Klein, 2014).

We use working Cox regression models for the cause-specific hazard rates to estimate the absolute risk of the event (Benichou & Gail, 1990; Ozenne et al., 2017). Furthermore, we allow the censoring distribution to depend on baseline covariates via a separate Cox regression model and work with a logistic regression model for the propensity of treatment. We study the robustness of the proposed estimator to a possible misspecification of any of these working models. Our work relates and extends recent developments in survival analysis: Wang, Beste, Maier, and Zhou (2016) proposed a doubly robust estimator for right-censored survival data when using parametric working regression models for the outcome distribution and the treatment distribution, and a nonparametric model (Kaplan–Meier) for the censoring distribution. Using the semiparametric theory (Bickel, Klaassen, Ritov, & Wellner, 1993; van der Laan & Robins, 2003; Tsiatis, 2006), we derive an augmentation term that makes the estimator robust against misspecification of the censoring model. The augmentation term resembles the one in the survival case (Zhao et al., 2014). We also derive the influence function of the estimator and show that it can be greatly simplified when all working models are correctly specified.

Alternative estimators of the ATE were proposed based on pseudo-observations (Andersen, Syriopoulou, & Parner, 2017), relying on the assumption that censoring is independent of all other variables, by direct standardization (Zhang & Zhang, 2011), relying on a regression model for the subdistribution hazard function, and inverse probability weighting (Neumann & Billionnet, 2016), relying solely on a model for the propensity of treatment.

This paper is structured as follows: Section 2 formally introduces the competing risk setting, parameter of interest, and statistical models. Section 3 presents the G-formula, IPTW, and doubly robust estimators in a competing risk setting. We derive in Section 4 the asymptotic properties of the three estimators: consistency, asymptotic normality, and their influence function. Robustness of the estimators to model misspecification and coverage of confidence intervals based on the asymptotic distribution of the estimators is assessed in Section 5 using simulation studies. Finally, in Section 6, we apply the estimators to compare two anticoagulation treatments regarding their impact on the risk of bleeding (adverse endpoint) in patients with atrial fibrillation (AF). The data used for this illustration are a subset of the data of Stærk et al. (2018), where we applied Cox regression for the event hazard and the hazard of death without event in order to estimate average differences in \( \tau \)-year risk of stroke and bleeding between alternative drugs for anticoagulation therapy.

## 2 | COMPETING RISK SETTING

### 2.1 | Notation and parameter of interest

We consider a random sample of \( n \) individuals \( \{(\tilde{T}_i, \Delta_i, A_i, W_i)\}_{i=1}^n \), where \( A \) is a binary treatment variable assigned at baseline, \( W \) is a \( d \)-dimensional vector of auxiliary covariates measured at baseline, \( \tilde{T} \) is a right-censored event time, that is, \( \tilde{T} = T \wedge C \), where \( T \) is the event time, \( C \) is the censoring time, \( \Delta \) is the event type for which we assume that \( \Delta = 1 \) means that the event of interest occurred and \( \Delta = 2 \) that the competing event occurred, and \( \tilde{\Delta} = \Delta 1\{T \leq C\} \) indicates uncensored observation (we use \( 1\{-\} \) to denote the indicator function). We assume throughout that \( (T, \Delta) \) are conditionally independent of \( C \) given \( (W, A) \) and that in the case of tied event and censoring times, that is, \( C = T \), the event time is earlier. Also, for a fixed time point \( \tau \), we assume that the probability of right-censoring is bounded away from zero: \( \forall (a, w) \in [0,1] \times W, P[C > \tau | A = a, W = w] > \epsilon \), where \( \epsilon > 0 \). We denote \( Y(\tau) = 1\{T \leq \tau, \Delta = 1\} \) for the indicator for the event of interest at time \( \tau \) and note that its expected value is the absolute risk that the event of interest occurs before time \( \tau \).

To define the target parameter, we introduce the potential outcomes \( Y^a(\tau) \), that is, the response of a randomly selected individual had that individual, possibly contrary to the fact, been given treatment \( A = a \). The target parameter is the expected difference:

\[
\Psi(\tau) = E[Y^1(\tau) - Y^0(\tau)].
\]
We make the following assumptions: \( Y(\tau) = (1 - A)Y^0(\tau) + AY^1(\tau) \) (consistency assumption), \( \forall a \in \{0, 1\}, (Y^a(\tau), A) \) are conditionally independent given \( W \), (no unmeasured confounders), and \( \forall (a, w) \in \{0, 1\} \times \mathcal{W}, \ P[A = a|w] > 0 \) (positivity assumption), where \( \mathcal{W} \subset \mathbb{R}^d \) denotes the set of possible values for \( W \).

### 2.2 Modeling

To estimate the target parameter based on the observed data, we consider the following conditional distributions as nuisance parameters. The cumulative incidence function \( F_1 \) describes the absolute risk of the event of interest by time \( t \): \[
F_1(t|A, W) = P(T \leq t, \Delta = 1|A, W),
\]

\( G \) is the conditional probability of being uncensored \[
G(t|A, W) = P(C > t|A, W),
\]

and \( \pi \) describes the propensity of treatment conditional on \( W \) \[
\pi(W) = P(A = 1|W).
\]

Under the identifiability assumptions stated in Section 2.1, the likelihood of the observed variables \( O_i = (\tilde{T}_i, \tilde{\Delta}_i, A_i, W_i) \) factorizes (Begun, Hall, Huang, & Wellner, 1983; Gill, Van der Laan, & Robins, 1995) and the density of their joint probability distribution \( P \) with respect to a suitable dominating measure can be parameterized:

\[
P(dr, \delta, a, dw) = \begin{cases} 
   G(t - |a, w)F_1(dr|a, w)(a\pi(w) + (1 - a)(1 - \pi(w))H(dw))^{1{\delta=1}} \\
   (G(t - |a, w)F_2(dr|a, w)(a\pi(w) + (1 - a)(1 - \pi(w))H(dw))^{1{\delta=2}} \\
   (1 - F_1(t - |a, w) - F_2(t - |a, w))G(dr|a, w)(a\pi(w) + (1 - a)(1 - \pi(w))H(dw))^{1{\delta=0}}
\end{cases}
\]

where \( F_2(t|A, W) = P(T \leq t, \Delta = 2|A, W) \), \( H \) is the marginal distribution of \( W \), and \( t^- \) denotes the left-handed limit at time \( t \). Our working model for the joint probability distribution \( P \) leaves the \( H \) part completely nonparametric but for each of the other nuisance parameters, we specify a (semi-)parametric regression model as our working model and define a corresponding estimator. Our working model for \( F_1 \) uses the parameterization of Benichou and Gail (1990) in terms of the cumulative cause-specific hazard functions \( \Lambda_1 \) for the event of interest and \( \Lambda_2 \) for the competing event: \[
F_1(t|A, W) = \int_{0}^{t} S(s - |A, W)\Lambda_1(ds|A, W),
\]

where \( S(s|A, W) = \exp\{-(\Lambda_1(s|A, W) + \Lambda_2(s|A, W))\} \) is the event-free survival function. Specifically, we consider two separate Cox regression models for \( \Lambda_1 \) and \( \Lambda_2 \) such that the model is parameterized in terms of the cause-specific hazard ratios and baseline hazard functions. An alternative parameterization of \( F_1 \) can be obtained by binomial regression for competing risks (Scheike, Zhang, & Gerds, 2008), where the Fine–Gray regression model (Fine & Gray, 1999) is a special case. Our working models for the censoring mechanism and the propensity of treatment are a Cox regression model and a logistic regression model, respectively. Note that all these working models come with their regular asymptotically linear estimators for the
respective nuisance parameter based on the observed data. Thus, we assume that there exist regular asymptotically linear estimators $\hat{F}_n$, $\hat{\pi}_n$, $\hat{S}_n$, $\hat{G}_n$ with respective large sample limits in probability $F^*_1, \pi^*, S^*, G^*$ such that:

\[
\begin{align*}
\sqrt{n}(\hat{\pi}_n - \pi^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{IF}_{\pi^*}(O_i) + o_p(1), \\
\sqrt{n}(\hat{G}_n - G^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{IF}_{G^*}(O_i) + o_p(1), \\
\sqrt{n}(\hat{F}_n - F^*_1) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{IF}_{F^*_1}(O_i) + o_p(1), \\
\sqrt{n}(\hat{S}_n - S^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{IF}_{S^*}(O_i) + o_p(1),
\end{align*}
\]

where $\text{IF}_{F^*_1}, \text{IF}_{\pi^*}, \text{IF}_{S^*}, \text{IF}_{G^*}$ are the influence functions corresponding to the estimators that represent the first-order von Mises expansion of the corresponding statistical functional (van der Vaart, 1998). If our working model for $F_1$ is correctly specified, then the asymptotic bias is zero, $F^*_1 - F_1 = 0$, and the same holds for the working models for $\pi, S,$ and $G$. Note that since both $F_1$ and $S$ can be expressed as differentiable functionals of $\Lambda_j$ for $j = 1, 2$, a sufficient condition for the last two lines of Equation (2) is

\[
\sqrt{n}(\Lambda_j - \Lambda^*_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \text{IF}_{\Lambda^*_j}(O_i) + o_p(1),
\]

where $\text{IF}_{\Lambda^*_j}$ is the influence function of the Cox regression estimator of the cumulative hazard function $j$, and $\Lambda^*_j$ is the corresponding large sample limit.

In case of a misspecified model, an asymptotic linear expansion of the estimators as in Equation (2) still continues to hold under the usual regularity conditions around the least-false parameters $F^*_1, \pi^*, S^*, G^*$ (White, 1982; Hjort, 1992; Bickel, Klaassen, Ritov, & Wellner, 1993; Gerds & Schumacher, 2001). However, there would be a large sample bias.

3 | ESTIMATORS FOR THE ATE

We consider the three types of estimators for the estimand $\Psi(\tau)$. Each type of estimator is based on a different combination of estimators of the outcome, treatment, and censoring mechanisms. We start by defining the estimators in the case without censoring.

3.1 | Uncensored data

The first class of estimators is based on the G-causal parameter (Robins, 1986, p.1410), also called backdoor adjustment (Pearl, 2000, section 3.2), which yields the G-formula:

\[
\Psi(\tau) = E[F_1(\tau | A = 1, W) - F_1(\tau | A = 0, W)].
\]

The regression estimator is obtained by substituting $\hat{F}_n$ for $F_1$:

\[
\hat{\Psi}_{\text{G-formula}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \{ \hat{F}_n(\tau | A = 1, W_i) - \hat{F}_n(\tau | A = 0, W_i) \}
\]

and is a consistent estimator of $\Psi(\tau)$ if the outcome model is correct.

The second class of estimators uses IPTW and is based on the formula:

\[
\Psi(\tau) = E \left[ Y(\tau) \left( \frac{A}{\pi(W)} - \frac{1-A}{1-\pi(W)} \right) \right].
\]
The IPTW estimator is obtained by substituting \( \hat{\pi}_n \) for \( \pi \):

\[
\hat{\Psi}_{\text{IPTW}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i(\tau) \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) \right\}
\]

(4)

and is a consistent estimator of \( \Psi(\tau) \) if the treatment model is correct.

The third class of estimators combines the G-formula estimator and the IPTW estimator into a doubly robust estimator, an estimator that is consistent if either the outcome model or the treatment model is correctly specified (Hernán & Robins, 2019, chapter 13). Following Tsiatis (2006, section 13.5), we use the formula

\[
\Psi(\tau) = E \left[ \frac{Y(\tau)A}{\pi(W)} + F_1(\tau|A = 1, W)(1 - \frac{A}{\pi(W)}) - \frac{Y(\tau)(1 - A)}{1 - \pi(W)} - F_1(\tau|A = 0, W)(1 - \frac{1 - A}{1 - \pi(W)}) \right].
\]

The augmented IPTW estimator (denoted by AIPTW) substitutes \( \hat{\pi}_n \) for \( \pi \) and \( \hat{F}_1n \) for \( F_1 \):

\[
\hat{\Psi}_{\text{AIPTW}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ Y_i(\tau)A_i \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) - \frac{Y_i(\tau)(1 - A_i)}{1 - \hat{\pi}_n(W_i)} - \frac{\hat{F}_1n(\tau|A = 1, W)}{1 - \hat{\pi}_n(W_i)} \right\}.
\]

(5)

We refer to Glynn and Quinn (2010) and Kennedy (2016) for nice reviews of the doubly robust AIPTW estimator in uncensored data.

### 3.2 Right-censored data

In the presence of right-censoring, the binary outcome at the time point of interest \( Y(\tau) \) is not observed for all subjects, it is only observed in the event \( \{C > T \land \tau\} = \{T > \tau\} \cup \{T \leq \tau, \Delta \neq 0\} \). To construct estimators of the ATE based on the right-censored data, we combine the estimators of the previous section with inverse probability-of-censoring weighting (IPCW) and now also using the estimator \( \hat{\pi}_n \). Note that the G-formula estimator defined in Equation (3) does not explicitly involve \( Y(\tau) \). Hence, in order to use the G-formula, whatever method is used to estimate the outcome model, it must explicitly allow for censored data. Using that \( 1\{\tilde{T} > \tau\}Y(\tau) = 0 \), we define the following IPCW estimators:

\[
\hat{\Psi}_{\text{IPTW,IPCW}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1\{\tilde{T}_i \leq \tau, \Delta_i \neq 0\}}{\hat{\pi}_n(\tilde{T}_i|A_i, W_i)} Y_i(\tau) \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) \right\}
\]

(6)

\[
\hat{\Psi}_{\text{AIPTW,IPCW}}(\tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1\{\tilde{T}_i \leq \tau, \Delta_i \neq 0\}}{\hat{\pi}_n(\tilde{T}_i|A_i, W_i)} Y_i(\tau)A_i \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) + \hat{F}_1n(\tau|A = 1, W) \left( 1 - \frac{A_i}{\hat{\pi}_n(W_i)} \right) \right\}
\]

\[
- \frac{1\{\tilde{T}_i \leq \tau, \Delta_i \neq 0\}}{\hat{\pi}_n(\tilde{T}_i|A_i, W_i)} Y_i(\tau)(1 - A_i) \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) - \hat{F}_1n(\tau|A = 0, W) \left( 1 - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) \right\}
\]

(7)

Both estimators can now be augmented using semiparametric theory (see van der Laan & Robins, 2003). In Online Appendix A, we derive the set of observed-data estimating functions for \( \Psi \). These estimating equations include an augmentation term which, when set to 0, leads to the IPCW estimators (Equations 6 and 7). Alternatively, the augmentation term can be chosen in order to minimize the asymptotic variance of the corresponding estimator. This choice leads to the following estimators (see Online Appendix A for details):

\[
\hat{\Psi}_{\text{IPTW,AIPCCW}}(\tau) = \hat{\Psi}_{\text{IPTW,IPCW}}(\tau) + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\hat{I}(\tilde{T}_i, \tau|A_i, W_i)}{\hat{\pi}_n(W_i)} \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) \right\}
\]

(8)

\[
\hat{\Psi}_{\text{AIPTW,AIPCCW}}(\tau) = \hat{\Psi}_{\text{AIPTW,IPCW}}(\tau) + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\hat{I}(\tilde{T}_i, \tau|A_i, W_i)}{\hat{\pi}_n(W_i)} \left( \frac{A_i}{\hat{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \hat{\pi}_n(W_i)} \right) \right\}
\]

(9)
Here, \( N^C_i(t) = 1(\hat{T}_i \leq t, \Lambda_i = 0) \) denotes the censoring counting process of subject \( i \) and \( \Lambda^C \) is the cumulative hazard function of \( G \) such that \( M^C_i(t) = N^C_i(t) - \int_0^t 1(\hat{T}_i \geq s) \Lambda^C(ds|A_i, W_i) \) is a 0 mean process (a martingale with respect to the natural filtration, see, e.g., Andersen, Borgan, Gill, & Keiding, 1993, section II.4). We use the notation \( \hat{M}^C_i(t) = N^C_i(t) - \int_0^t 1(\hat{T}_i \geq s) \Lambda^C(ds|A_i, W_i) \) and \( M^{C,*}_i(t) = N^C_i(t) - \int_0^t 1(\hat{T}_i \geq s) \Lambda^{C,*}(ds|A_i, W_i) \), where \( \Lambda^{C,*} \) is the large sample limit of \( \Lambda^C \).

\section{Asymptotic Properties}

In this section, we study the asymptotic properties of the following estimators: \( \hat{\Psi}_{\text{G-formula}}(\tau) \), \( \hat{\Psi}_{\text{IPTW,IPCW}}(\tau) \), and \( \hat{\Psi}_{\text{AIPCW}}(\tau) \). For a random variable \( X_n \), we use \( \text{p-lim}_{n \to \infty} X_n = X \) to denote that \( X_n \) converges in probability toward the random variable \( X \).

\subsection{Consistency}

By Equation (2) and the law of large numbers, we have
\[
\text{p-lim}_{n \to \infty} \hat{\Psi}_{\text{G-formula}}(\tau) = \mathbb{E}[F^*_1(\tau|A = 1, W) - F^*_1(\tau|A = 0, W)].
\]

Thus, if the outcome model is correctly specified at \( \tau \), that is, for \( a \in [0, 1] \) and almost all \( w \), \( F_1(\tau|a, w) = F^*_1(\tau|a, w) \), then \( \hat{\Psi}_{\text{G-formula}} \) is a consistent estimator for \( \Psi(\tau) \). Similarly, we have under the assumptions of Section 2
\[
\text{p-lim}_{n \to \infty} \hat{\Psi}_{\text{IPTW,IPCW}}(\tau) = \mathbb{E}
\left[
\frac{G(\hat{T}|A, W)}{G^*(\hat{T}|A, W)} \left( \frac{F_1(\tau|A = 1, W)\pi(W)}{\pi^*(W)} - \frac{F_1(\tau|A = 0, W)(1 - \pi(W))}{1 - \pi^*(W)} \right)
\right].
\]

Hence, if the working models for the treatment and censoring mechanism are correctly specified, that is, \( \pi(w) = \pi^*(w) \) and \( G(s|a, w) = G^*(s|a, w) \) for all \( s \in [0, \tau], a \in [0, 1] \) and almost all \( w \), then \( \hat{\Psi}_{\text{IPTW,IPCW}}(\tau) \) is consistent. The following theorem states sufficient conditions under which \( \hat{\Psi}_{\text{AIPCW}} \) is consistent.

\textbf{Theorem 1.} Under the assumptions stated in Section 2, the estimator \( \hat{\Psi}_{\text{AIPCW}}(\tau) \) is consistent whenever one of the following conditions is satisfied for all \( s \in [0, \tau] \), \( a \in [0, 1] \) and almost all \( w \):

1. \( G^*(s|a, w) = G(s|a, w) \text{ and } F^*_1(\tau|a, w) = F_1(\tau, a, w) \)
2. \( G^*(s|a, w) = G(s|a, w) \text{ and } \pi^*(w) = \pi(w) \)
3. \( F^*_1(s|a, w) = F_1(s|a, w) \text{ and } S^*(s|a, w) = S(s|a, w) \)

\textbf{Proof:} We refer to Online Appendix B for the proof and here only give the key elements of the proof: When the censoring model is correctly specified, 1. and 2. follow from the fact that \( \hat{\Psi}_{\text{AIPCW}}(\tau) \) and \( \hat{\Psi}_{\text{AIPCW}}(\tau) \) have the same large sample limit. When the censoring model is misspecified but the outcome and survival models are correctly specified, then \( \hat{\Psi}_{\text{AIPCW}}(\tau) \) and \( \hat{\Psi}_{\text{G-formula}}(\tau) \) have the same large sample limit, which gives 3.

\subsection{Asymptotic Distribution}

All estimators described in the previous section can be written as averages of functions of the data indexed by estimated nuisance parameters:
\[
\hat{\Psi}_x(\tau) = \frac{1}{n} \sum_{i=1}^n h_x(\tau; O_i, \hat{F}_{1n}, \hat{R}_n, \hat{S}_n, \hat{G}_n),
\]

where \( x \in \{ \text{G-formula; IPTW,IPCW; AIPCW,IPCW; IPTW,AIPCW; AITW,AIPCW} \} \)
and a suitable function \( h_x \). For instance,
\[
h_{\text{G-formula}}(r; O_i; \hat{F}_{in}, \hat{\pi}_n, \hat{S}_n, \hat{G}_n) = \hat{F}_{in}(r| A = 1, W_i) - \hat{F}_{in}(r| A = 0, W_i).
\]

If the nuisance parameters were known, say equal to \((F^*_1, \pi^*, S^*, G^*)\), the correspondingly defined plug-in estimators would be simple averages of independent and identically distributed quantities with influence function:
\[
\tilde{I}_{F_x}(r; O_i) = h_x(r; O_i; F^*_1, \pi^*, S^*, G^*) - \Psi^*_x(r),
\]
(10)

where \( \Psi^*_x \) is the large sample limit of \( \tilde{\Psi}_x \). From the central limit theorem, we would get that the estimators are asymptotically normal with asymptotic variance equal to the variance of the influence function. However, in practice, the nuisance parameters are estimated with the same data and the asymptotic expansions of the estimators of the ATE may involve the influence functions of the estimators of the nuisance parameters given in Equation (2). The general idea to obtain the influence function is to apply the functional delta method (van der Vaart, 1998, chapter 20) to get a von Mises expansion of the form:
\[
\sqrt{n}(\tilde{\Psi}_x(r) - \Psi^*_x(r)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{IF}_x(r; O_i) + o_P(1).
\]

The influence function has two terms:
\[
\text{IF}_x(r; O_i) = \tilde{I}_{F_x}(r; O_i) + \phi_x(r; O_i; \text{IF}^{F_1}_x, \text{IF}^{\pi}_x, \text{IF}^S_x, \text{IF}^G_x),
\]
(11)

where a function \( \phi_x \) relates to the influence functions of the estimators of the nuisance parameters. In the case of the G-formula estimator,
\[
\phi_{\text{G-formula}}(r; O_i; \text{IF}^{F_1}_x, \text{IF}^{\pi}_x, \text{IF}^S_x, \text{IF}^G_x) = E[\text{IF}^{F_1}_x(r; A = 1, W; O_i)|O_i] - E[\text{IF}^{F_1}_x(r; A = 0, W; O_i)|O_i]
\]

and for the IPTW, IPCW estimator:
\[
\phi_{\text{IPTW, IPCW}}(r; O_i; \text{IF}^{F_1}_x, \text{IF}^{\pi}_x, \text{IF}^S_x, \text{IF}^G_x) = -E \left[ \text{IF}^{\pi}(W; O_i) \frac{1\{T \leq r, \tilde{\Delta} \neq 0\}}{G^*(T|A, W)} Y(r) \left( A \frac{1}{\pi^*(W)} - \frac{1 - A}{1 - \pi^*(W)} \right) | O_i \right]
\]
\[
- E \left[ \text{IF}^G(\tilde{T}, A, W; O_i) \frac{1\{\tilde{T} \leq r, \tilde{\Delta} \neq 0\}}{G^*(\tilde{T}|A, W')} Y(r) \left( A \frac{1}{\pi^*(W')} - \frac{1 - A}{1 - \pi^*(W')} \right) | O_i \right],
\]

where the expectation is taken relative to the random variables \((\tilde{T}, \tilde{\Delta}, A, W)\), which are independent of \( O_i \), their realization for individual \( i \). The formula for the influence function of \( \tilde{\Psi}_{\text{AIPTW, IPCW}}(r) \) is more complex and can be found in Online Appendix C.

Under the assumptions stated in Section 2, and in particular under Equation (2), the functional delta method yields that the asymptotic distribution of the estimator \( \tilde{\Psi}_x \) is a normal distribution with variance equal to the variance of the influence function. The variance of \( \tilde{\Psi}_x \) can then be estimated based on an estimate \( \tilde{I}_{\tilde{F}_x} \) of the influence function:
\[
\frac{1}{n} \sum_{i=1}^n \left( \tilde{F}_{in}(r| A = 1, W_i) - \tilde{F}_{in}(r| A = 0, W_i) - \tilde{\Psi}_{\text{AIPTW, IPCW}}(r) \right)
\]
\[
+ \left( \frac{A_i}{\tilde{\pi}_n(W_i)} - \frac{1 - A_i}{1 - \tilde{\pi}_n(W_i)} \right) \left( \frac{1\{\tilde{T}_i \leq r, \tilde{\Delta}_i \neq 0\}}{\tilde{G}_n(\tilde{T}_i|A_i, W_i)} \tilde{F}_{in}(r| A_i, W_i) \right) \right)^2.
\]

This result is a consequence of the orthogonality between the estimating function and the nuisance parameter tangent space, see also Tsiatis (2006, remark 4, section 3.3). In that case, the AITPW, AIPCW estimator is nonparametric efficient.
FIGURE 1  Illustration of the data-generating mechanism used in the simulation studies. Shown are the Aalen–Johansen estimates for the absolute risks of cause 1 in both treatment arms in two independently drawn datasets (nonrandomized and randomized) each of size \( n = 10,000 \).
Panel A: The treatment effect is zero. In the nonrandomized world, the Aalen–Johansen estimate of the 8-year risk difference is large. Panel B: The treatment has a protective effect. In the nonrandomized world, the Aalen–Johansen estimate of the 8-year risk difference is about zero.

5 | EMPIRICAL STUDIES

The following simulation studies investigate the bias-variance tradeoff of the various estimators under model misspecification and the small sample coverage based on the asymptotic variance formula.

5.1 | Simulation setting

In total, 12 auxiliary covariates are simulated, six having a standard normal distribution (\( W_1, \ldots, W_6 \)) and the remaining six having a Bernoulli distribution (\( W_7, \ldots, W_{12} \)). A binary treatment variable is drawn following a logistic regression model. We use three Cox–Weibull regression models (table II, Bender, Augustin, & Blettner, 2005) to simulate three latent times conditional on treatment and auxiliary covariates, one for the event of interest, one for the competing risk, and one for the right-censoring time. The observed time is then obtained as the minimum of the three latent times and the event status corresponds to the event with the smallest latent time. In the main analyses, the 12 auxiliary covariates are independent. The covariate effects on the treatment, hazard rate of the event of interest, the hazard rate of the competing risk, and the hazard rate of the censoring are controlled by including additive effects of the six binary variables, six continuous variables, and the squares of the six continuous variables into the linear predictors of the logistic regression and the Cox–Weibull regression models, respectively. The effect of treatment on the three hazard rates is controlled by three additional regression parameters. Note that the randomized world corresponds to setting all regression parameters of the logistic regression model to zero and deviations from the randomized world can be controlled by varying these covariate effects (Figure 1).

For various parameter settings, we report results of the estimators G-formula (Equation 3), IPTW, IPCW (Equation 6), and A IPTW, AIPCW (Equation 9) across 1,000 simulated datasets. These estimators are implemented in R (R Core Team, 2018) in the package riskRegression (Gerds & Ozenne, 2019, function ate). When estimating the variance of the A IPTW, AIPCW estimators, we consider two estimators for the influence function. The first, denoted by A IPTW, AIPCW(nonrobust CI), only estimates the first term of Equation (11) since the second term is 0 in correctly specified models. The second estimates both terms and is denoted by A IPTW, AIPCW(robust CI). The R-code of the simulation studies is available as Supplementary Material.
FIGURE 2 Simulation setting where there is no treatment effect (panel A of Figure 1). Boxplots show results of 1,000 simulated datasets (each with sample size 500) and each of four methods for estimating the average 10-year risk difference between treated and untreated subjects. Upper left panel: all regression models (treatment, event of interest, competing risk, and censoring) are correctly specified. Upper right panel: the treatment model is misspecified (missing covariates and missing quadratic effects). Lower left panel: the event of interest and the competing risk models are misspecified (missing covariates and missing quadratic effects). Lower right panel: the censoring model is misspecified (missing quadratic effects).

5.2 | Simulation results

We report results for a data-generating model without treatment effect (Panel A, Figure 1). The figure shows Aalen–Johansen estimates (Aalen & Johansen, 1978; Andersen et al., 1993) of the cumulative incidence functions. Similar results are obtained when considering a nonzero treatment effect, but then the "true" value needs to be obtained empirically. Model misspecification is simulated by omitting covariates and quadratic effects. We created four scenarios. In the first one, all models are correctly specified. In the three other scenarios, precisely one of the censoring, outcome, or treatment models is misspecified. As shown in Figure 2 (upper panel), the AIPW,AIPCW estimator is consistent even when one of the models (outcome, treatment, or censoring) is misspecified. The G-formula estimator and the IPTW,IPCW estimator need one or two models to be correctly specified to be consistent—the outcome model for the G-formula estimator and both the treatment and censoring models for the IPTW,IPCW estimator. The G-formula estimator appears to be less variable compared to the other estimators. The IPTW,IPCW estimator is at least as variable but often more variable than the AIPW,AIPCW estimator.

As shown in Figure 3, the coverage of the AIPW,AIPCW estimator was found satisfactory (i.e., close to 0.95), except in small samples where the coverage was slightly too low (e.g., 0.9 for n=100 in the case of a misspecified outcome model). In particular, the nonrobust estimator of the variance of AIPW,AIPCW estimator was found to appropriately control the type 1 error rate, even when one model was misspecified.

6 | REAL DATA APPLICATION

For the sole purpose of illustration, we consider a subset of the data presented in Stærk et al. (2018). This Danish registry study included n = 21,149 patients with a diagnosis of AF in the period 2012–2016 who initiated anticoagulation treatment with a
standard dose of dabigatran \((n = 7,078)\) or rivaroxaban \((n = 6,868)\) or apixaban \((n = 7,203)\). All three treatments belong to the group of nonvitamin K antagonist oral anticoagulants. Here, we consider only data from patients who initiated treatment with either dabigatran or rivaroxaban. The follow-up started at the date of treatment initiation. The original study by Stærk et al. (2018) presented results on several adverse endpoints, including thromboembolism/stroke and major bleeding, where death without the endpoint is the only competing risk. Here, we consider the analysis of the endpoint major bleeding, where death without major bleeding or shift or discontinuation of treatment is the competing risk. The treatment assignment is not randomized, but there are official guidelines and presumably also doctor preferences which most likely also depend on the patient characteristics. Note that the results presented here for G-formula are not directly comparable to those presented in Stærk et al. (2018), because we restrict here all Cox regression models to the subset of the dabigatran and rivaroxaban patients. Otherwise, we use the same covariate adjustment as described in detail in Stærk et al. (2018) for all Cox regression models and the logistic regression model of the treatment mechanism. Figure 4 displays the estimates absolute risk of major bleeding obtained with G-formula and AIPTW,AIPCW. Within the limitation of the available confounder information, the results can be interpreted as what one would have observed in a hypothetical world where all patients initiated dabigatran (or rivaroxaban), respectively.

The interpretation of these results is limited to the population of patients who initiated either dabigatran or rivaroxaban in the period 2012–2016. Based on the AIPTW,AIPCW estimate evaluated at 12 months, the interpretation could be as follows. If every patient had received dabigatran, the 1-year risk [95% confidence interval] of a major bleeding would have been 1.58% \([0.60;2.57]\) lower compared to when every patient had received rivaroxaban. Interestingly, the AIPTW,AIPCW estimates of the risk differences are larger in magnitude compared to the G-formula estimates. For example, the estimate of ATE (12-month) using G-formula is only 0.97% \([0.40;1.54]\).
7 | DISCUSSION

In the presence of completely observed outcomes, estimation of the ATE can be performed using estimators based on the G-formula, IPTW, or a combination of both (AIPTW). Although these are classical tools in causal inference (see Hernán & Robins, 2019), we review in this article their extension to right-censored data and to the presence of competing risks. Using results from semiparametric theory, we derive the augmentation term relative to the working model for the censoring distribution. We investigate the robustness of this new estimator against misspecification of the working models. We also show the asymptotic normality of this estimator and derive an analytical formula for its influence function, which can be used to estimate the variance of the estimator. The variance of the estimator may depend on the estimators of the nuisance parameters. Alternatively, nonparametric bootstrap could be used; it requires more computing time, but less memory. In our software implementation (Gerds & Ozenne, 2019), we focus on the use of cause-specific Cox regression models for the outcome model, a logistic regression for the treatment model, and a Cox regression model for the censoring model. An alternative would be to use a Fine–Gray regression model for the outcome. However, then, one would need an additional working regression model for the conditional event-free survival function $S(\cdot|A, W)$.

The simulations confirm the superiority of the AIPTW, AIPCW estimator over the IPTW, IPCW estimator. They also show that the G-formula estimator is less variable than the AIPTW, AIPCW estimator when the outcome model is correctly specified. However, the G-formula estimator has a bias that the AIPTW, AIPCW estimator does not have when the outcome model is misspecified. It is worth noting that the definition of the G-formula estimator is unchanged in the presence of censoring—only the outcome model has to properly handle censoring.

Competing risks essentially lead to a change in the definition of the outcome, where we use $1\{T \leq \tau, \Delta = 1\}$ instead of $1\{T \leq \tau\}$. However, one should not overlook that the presence of competing risks complexifies the assessment of the treatment effect, especially when the treatment has a positive effect on the cause of interest but a negative effect on the competing events. We refer to Young, Tchetgen, and Hernán (2018) for a detailed discussion on the implications of how the estimand is defined in the presence of competing risks.

Recently, Lesko and Lau (2017) pointed out that bias will occur if we do not have the correct models for the probability of the outcome of interest $F_1(t|A, W) = P[T \leq t, \Delta = 1|A, W]$, in particular when the model for the hazard rate of the competing risk $A_2$ is misspecified. In practice, this means that, if we estimate the outcome model via a cause-specific Cox regression models, both conditional hazard functions need to be correctly specified. While this approach relies on prior knowledge to define the working models, automated techniques and the use of cross-validation (Benkeser, Gilbert, & Carone, 2018) may be preferable when prior knowledge is sparse. Indeed, the Cox regression model makes the assumption of proportional hazards, which may...
not always be appropriate. This assumption can be relaxed, for example, by using stratified baseline hazard functions, time varying coefficients (Martinussen & Scheike, 2007), or an alternative approach that does not rely on this assumption (e.g., using pseudo-observations, Andersen et al., 2017).

We have focused on a binary treatment variable. In the case of a multi-valued treatment variable, the several estimands can be defined depending on the type of the treatment variable (ordinal vs. nominal), see Imbens (2000) for a nice discussion. One option is to compare each pair of treatments in the subpopulation of subjects treated with either of the treatments. This is what we have done in the real data analysis.

We have also focused on a single time point to evaluate the treatment effect. The methods presented here can be extended to multiple time points, perhaps at the cost of a multiple testing issue. However, as a reviewer has pointed out, there is no guarantee that the doubly robust estimating equations lead to estimates of the cumulative incidence functions that are monotone in time, that is, in the interval \((0; \tau)\).

Handling time-varying treatments and therefore possible time-varying confounding is more challenging and beyond the scope of this article; we refer the interested readers to Bekaert et al. (2010); Daniel et al. (2013); Moodie et al. (2014); and Hernán and Robins (2019).

ACKNOWLEDGMENTS

We thank the Associate Editor and two referees for constructive comments. B.M.H.O. was supported by Hjerteforeningen Forskningsstøtte (nr. 15-R99-A5954 015-S15) and the Lundbeck Foundation (R231-2016-3236). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 746850.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Brice Maxime Hugues Ozenne https://orcid.org/0000-0001-9694-2956

REFERENCES

Aalen, O., & Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. Scandinavian Journal of Statistics, 5, 141–150.

Aasbjerg, K., Mortensen, P. E., Nørgaard, M. A., Rytgard, H. C., Gerds, T. A., Søgaard, P., ..., Køber, L. (2019). Comparison of survival after aortic valve replacement with mitroflow or perimount prostheses. Seminars in Thoracic and Cardiovascular Surgery, 31, 350–358.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1993). Statistical models based on counting processes. Springer Series in Statistics. New York: Springer.

Andersen, P. K., Syriopoulou, E., & Parner, T. (2017). Causal inference in survival analysis using pseudo-observations. Statistics in Medicine, 36(17), 2669–2681.

Begun, J. M., Hall, W., Huang, W.-M., & Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. Annals of Statistics, 11(2), 432–452.

Bekaert, M., Vansteelandt, S., & Mertens, K. (2010). Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. Lifetime Data Analysis, 16(1), 45.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. Statistics in Medicine, 24, 1713–1723.

Benichou, J., & Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. Biometrics, 46(3), 813–826.

Benkeser, D., Gilbert, P. B., & Carone, M. (2019). Estimating and testing vaccine sieve effects using machine learning. Journal of the American Statistical Association, 114, 1038–1049.

Bickel, P. J., Klaassen, C. A., Ritov, Y., & Wellner, J. A. (1993). Efficient and adaptive estimation for semiparametric models. Baltimore, MD: Johns Hopkins.
Daniel, R. M., Cousens, S., De Stavola, B., Kenward, M. G., & Sterne, J. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine* 32(9), 1584–1618.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496–509.

Gerds, T., & Schumacher, M. (2001). On functional misspecification of covariates in the Cox regression model. *Biometrika*, 88(2), 572–580.

Gerds, T. A., & Ozenne, B. (2019). *riskRegression: Risk regression models and prediction scores for survival analysis with competing risks*. R package version 2019.03.03.

Gill, R. D., Van der Laan, M. J., & Robins, J. M. (1995). Coarsening at random: Characterizations, conjectures and counter-examples. In D. Y. Lin & T. R. Fleming (Eds.), *Proceedings of the 1st Seattle Symposium in Biostatistics* (pp. 255–294). Berlin: Springer Lecture Notes in Statistics.

Glynn, A. N., & Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, 18(1), 36–56.

Hernán, M. A., & Robins, J. M. (2006). Instruments for causal inference: An epidemiologist’s dream? *Epidemiology*, 17(4), 360–372.

Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American Journal of Epidemiology*, 183(8), 758–764.

Hernán, M. A., & Robins, J. M. (2019). *Causal inference*. Boca Raton, FL: Chapman & Hall/CRC.

Hjort, N. L. (1992). On inference in parametric survival models. *International Statistical Review*, 60, 355–387.

Imbens, G. W. (2000). The role of the propensity score in estimating dose–response functions. *Biometrika*, 87(3), 706–710.

Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4), 523–539.

Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research* (pp. 141–167). ICSA Book Series in Statistics. Cham: Springer.

Lesko, C. R., & Lau, B. (2017). Bias due to confounders for the exposure-competing risk relationship. *Epidemiology (Cambridge, Mass.*), 28(1), 20.

Martinussen, T., & Scheike, T. H. (2007). *Dynamic regression models for survival data*. Berlin: Springer Science & Business Media.

Moodie, E. E., Stephens, D. A., & Klein, M. B. (2014). A marginal structural model for multiple-outcome survival data: Assessing the impact of injection drug use on several causes of death in the Canadian Co-infection Cohort. *Statistics in Medicine*, 33(8), 1409–1425.

Neumann, A., & Billionnet, C. (2016). Covariate adjustment of cumulative incidence functions for competing risks data using inverse probability of treatment weighting. *Computer Methods and Programs in Biomedicine*, 129, 63–70.

Ozenne, B., Sørensen, A. L., Scheike, T., Torp-Pedersen, C., & Gerds, T. A. (2017). Riskregression: Predicting the risk of an event using Cox regression models. *R Journal*, 9(2), 440–460.

Pearl, J. (2000). *Causality, models, reasoning and inference*. Cambridge: Cambridge University Press.

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

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

Scheike, T., Zhang, M., & Gerds, T. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, 95(1), 205–220.

Sterk, L., Gerds, T., Lip, G. Y., Ozenne, B., Bonde, A., Lamberts, M., …, Olesen, J. (2018). Standard and reduced doses of dabigatran, rivaroxaban and apixaban for stroke prevention in atrial fibrillation: A nationwide cohort study. *Journal of Internal Medicine*, 283(1), 45–55.

Tsiatis, A. (2006). *Semiparametric theory and missing data*. Berlin: Springer Verlag.

van der Laan, M. J., & Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Berlin: Springer.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

Wang, X., Beste, L. A., Maier, M. M., & Zhou, X.-H. (2016). Double robust estimator of average causal treatment effect for censored medical cost data. *Statistics in Medicine*, 35(18), 3101–3116.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–25.

Yavuz, I., Chng, Y., & Wahed, A. S. (2018). Estimating the cumulative incidence function of dynamic treatment regimes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(1), 85–106.

Young, J. G., Tchetgen, E. J. T., & Hernán, M. A. (2018). The choice to define competing risk events as censoring events and implications for causal inference.

Zhang, X., & Zhang, M.-J. (2011). SAS macros for estimation of direct adjusted cumulative incidence curves under proportional subdistribution hazards models. *Computer Methods and Programs in Biomedicine*, 101(1), 87–93.

Zhao, Y.-Q., Zeng, D., Laber, E. B., Song, R., Yuan, M., & Kosorok, M. R. (2014). Doubly robust learning for estimating individualized treatment with censored data. *Biometrika*, 102(1), 151–168.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

---

**How to cite this article:** Ozenne BMH, Scheike TH, Sterek L, Gerds TA. On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biometrical Journal*. 2020:62:751–763. https://doi.org/10.1002/bimj.201800298