



On doubly robust estimation of the hazard difference

Dukes, Oliver; Martinussen, Torben; Tchetgen Tchetgen, Eric J; Vansteelandt, Stijn

Published in:
Biometrics

DOI:
[10.1111/biom.12943](https://doi.org/10.1111/biom.12943)

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
[Unspecified](#)

Citation for published version (APA):
Dukes, O., Martinussen, T., Tchetgen Tchetgen, E. J., & Vansteelandt, S. (2019). On doubly robust estimation of the hazard difference. *Biometrics*, 75(1), 100-109. <https://doi.org/10.1111/biom.12943>

On Doubly Robust Estimation of the Hazard Difference

Oliver Dukes^{1,*}, Torben Martinussen^{2,**}, Eric J. Tchetgen Tchetgen^{3,***}
and Stijn Vansteelandt^{1,****}

¹Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281 S9, Ghent 9000, Belgium

²Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B, 1014 Copenhagen K, Denmark

³Department of Statistics, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Pennsylvania 19104, U.S.A.

**email:* oliver.dukes@ugent.be

***email:* tma@sund.ku.dk

****email:* etchetge@hsph.harvard.edu

*****email:* stijn.vansteelandt@ugent.be

SUMMARY. The estimation of conditional treatment effects in an observational study with a survival outcome typically involves fitting a hazards regression model adjusted for a high-dimensional covariate. Standard estimation of the treatment effect is then not entirely satisfactory, as the misspecification of the effect of this covariate may induce a large bias. Such misspecification is a particular concern when inferring the hazard difference, because it is difficult to postulate additive hazards models that guarantee non-negative hazards over the entire observed covariate range. We therefore consider a novel class of semiparametric additive hazards models which leave the effects of covariates unspecified. The efficient score under this model is derived. We then propose two different estimation approaches for the hazard difference (and hence also the relative chance of survival), both of which yield estimators that are doubly robust. The approaches are illustrated using simulation studies and data on right heart catheterization and mortality from the SUPPORT study.

KEY WORDS: Additive hazards model; Causal inference; Doubly robust estimation; Lifetime and survival analysis; Semiparametric inference

1. Introduction

In the analysis of time-to-event data, one is often interested in the effect of an exposure A on a survival outcome T , subject to a censoring time C , and conditional on a set of variables L . This adjusted association may be summarized by the *hazard difference*, which can be estimated by fitting a multivariable additive hazards model, or the *hazard ratio*, commonly estimated via the Cox proportional hazards model. For example, in the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT), Connors et al. (1996) investigated the effect of right heart catheterization (RHC), a binary exposure, on patient mortality. The exposure could also be continuous, for example, particulate air pollution. In observational studies, the dimension of covariates to adjust for is often high. The SUPPORT investigators used expert input from clinicians to identify 72 variables that could affect the decision of whether to use RHC or not, which they wished to adjust for in the analysis. The problem of bias resulting from misspecification of the hazards regression model then becomes a dominant consideration.

Such concerns have prompted the development of *doubly robust* estimators of treatment effects (Robins and Rotnitzky, 2001). These estimators require two working models, one of which is a regression model for the outcome,

and another that relates to the treatment selection mechanism. Only one of these models needs to be correct in order to consistently estimate the treatment effect. Doubly robust estimators are now well established for parameters in linear, log-linear, and logistic conditional mean models (Robins, 1994; Tchetgen Tchetgen et al., 2010), and are particularly appealing when evaluating static treatment regimes or estimating optimal dynamic regimes in longitudinal studies. This is because it is challenging to specify a series of sequential regression models for the same outcome that are all simultaneously correct. More recently, the usefulness of doubly robust procedures has also been recognized in the context of data-adaptive selection or regularization. In particular, a number of common doubly robust estimators have turned out to be less susceptible to regularization bias than popular alternative estimators that do not possess the doubly robust property (Farrell, 2015). Standard confidence intervals for these doubly robust estimators are moreover uniformly valid, even when they ignore the use of such data-adaptive procedures (assuming the estimators for both working models converge sufficiently fast to the truth).

There has however been limited development of doubly robust estimators of the parameters indexing the hazard

regression models (additive or multiplicative) popular in survival analysis. For semiparametric proportional hazards models we conjecture that, partly due to the non-collapsibility of the hazard ratio, no estimators of the treatment effect hazard ratio exist that are consistent whenever the treatment selection mechanism (more precisely, the distribution of A given L) is known (Tchetgen Tchetgen et al., 2010). Double robustness with respect to the treatment selection mechanism is therefore not attainable under such models. In contrast, as we will show in this article, doubly robust estimation strategies do exist for the hazard difference under semiparametric additive hazards models. The additional robustness is particularly advantageous for additive models; these are prone to misspecification since they do not impose the constraint that hazards are non-negative.

In Section 2, we introduce the new class of semiparametric additive hazards models. A theory of estimation for these models is developed in Section 3, and the efficient score function is identified. Because it requires specification of the entire conditional distribution of the treatment given the covariates, we also describe a subclass of estimators which only requires this conditional distribution to be correctly specified up to the mean. Drawing from these results, two practical strategies for the estimation of the treatment effect are proposed in Section 4; both of which yield doubly robust estimators. In Section 5, our estimators are compared in simulations with standard estimators of additive hazards models, and we re-analyze data from the SUPPORT study in Section 6.

2. The Model

We begin with some notation. The counting process corresponding to the survival time T is denoted by $N(t) = I(T \geq t)$; \mathcal{F}_t is the history spanned by $N(t)$, with $R(t) = I(T \geq t)$. Let L include 1 for the intercept. For the moment, we assume there is no censoring.

Consider an additive hazards model of the form

$$E \{dN(t)|\mathcal{F}_t, L, A\} = \{d\Gamma(t)^T L + \psi A dt\} R(t),$$

where $d\Gamma(t)^T$ is a vector of coefficients that are allowed to depend on time (McKeague and Sasieni, 1994). This model imposes restrictions on the effect of L on the hazard at any time point t ; such restrictions are undesirable, because misspecification of an additive hazards model may be inevitable when L is high-dimensional and has continuous components. Incorrectly specifying the effect of L can then induce bias in estimation of ψ . We therefore further relax the model restrictions by developing inference for the semiparametric additive hazards model \mathcal{M} , defined by

$$E \{dN(t)|\mathcal{F}_t, L, A\} = \{d\Omega(t, L) + \psi A dt\} R(t), \quad (1)$$

where $d\Omega(t, L)$ denotes the effect of time and the covariates on the hazard and is left unspecified. Restrictions are now only imposed on the association between A and the hazard. A further relaxation of model \mathcal{M} is

$$E \{dN(t)|\mathcal{F}_t, L, A\} = \{d\Omega(t, L) + d\Psi(t)A\} R(t), \quad (2)$$

where $d\Psi(t)$ is an unknown locally integrable function of time. In Web Appendix A, we extend our results on estimation to model (2) (denoted by \mathcal{M}_{TV}), but otherwise assume that the effect of A is constant. To simplify the exposition, we will also assume there is no effect modification by Z , where $Z = z(L)$ is a vector function of L , and give details on extensions for multivariate ψ in the discussion.

By the equality

$$\frac{\text{pr}(T > t|A, L)}{\text{pr}(T > t|A = 0, L)} = \exp(-\psi A t),$$

which is implied by (1), it follows that $\exp(-\psi A t)$ can be interpreted as the adjusted relative change in the probability of surviving time t per unit increase in the exposure. This relative chance of survival is potentially easier to communicate than the hazard difference (or ratio). The reason is that contrasts between hazards lack a causal interpretation because they compare, at each time t , individuals who have not yet failed at that time. These individuals may not be exchangeable between treatment arms, even when the treatment is randomly assigned (Hernán, 2010).

3. Theory of Semiparametric Estimation

3.1. The Efficient Score for ψ

In this section, we develop a theory of estimation for ψ . We first give the semiparametric efficient score for ψ under model \mathcal{M} and discuss the properties of an efficient estimator. The derivation of all results is left to Web Appendix A.

Let $\Omega(t, L) = \int_0^t d\Omega(s, L)$, $\lambda(t|A, L; \psi) = d\Omega(t, L)/dt + \psi A$ and $dM(t; \psi) = dN(t) - \lambda(t|A, L; \psi)R(t)dt$ be the increment at time t of a local square-integrable martingale. Then the locally efficient score for ψ under model \mathcal{M} is

$$S_{\text{eff}} = \int_0^\infty \left[A - \frac{E\{\lambda^{-1}(t|A, L; \psi)A \exp(-\psi A t)|L\}}{E\{\lambda^{-1}(t|A, L; \psi) \exp(-\psi A t)|L\}} \right] \frac{dM(t; \psi)}{\lambda(t|A, L; \psi)} \quad (3)$$

The solution $\hat{\psi}$ for ψ to the equation

$$0 = \sum_{i=1}^n \int_0^\infty \left[A_i - \frac{E\{\lambda^{-1}(t|A_i, L_i; \psi)A_i \exp(-\psi A_i t)|L_i\}}{E\{\lambda^{-1}(t|A_i, L_i; \psi) \exp(-\psi A_i t)|L_i\}} \right] \frac{dM_i(t; \psi)}{\lambda(t|A_i, L_i; \psi)}$$

thus has an asymptotic variance which attains the semiparametric efficiency bound for ψ under model \mathcal{M} , when $f(A|L)$ is known and $\Omega(t, L)$ is correctly specified (Bickel et al., 1993).

In practice, the law $f(A|L)$ will usually be unknown, and thus so will $E\{\lambda^{-1}(t, A, L)A \exp(-\psi A t)|L\}$ and $E\{\lambda^{-1}(t, A, L) \exp(-\psi A t)|L\}$. One may then postulate a parametric model \mathcal{A}_D for the population distribution $f(A|L) = f(A|L; \alpha)$, where $f(A|L; \alpha)$ is a known function smooth in an unknown finite-dimensional parameter α . In practice, α can be estimated using maximum likelihood. Since $f(A|L)$ is ancillary to ψ , the efficiency bound for ψ is the same whether $f(A|L)$ is estimated or known. The score (3) is thus efficient under the intersection model $\mathcal{M} \cap \mathcal{A}_D$.

Implementation of an efficient estimator also requires knowledge of $d\Omega(t, L)$. In the semiparametric model \mathcal{M} , this function is left unspecified and is unknown to the data analyst. One option then is to estimate it via a working model \mathcal{B} , such as $d\Omega(t, L) = d\Omega\{L; \Gamma(t)\}$ where $d\Omega\{L; \Gamma(t)\}$ is a known function, smooth at each time point t in an unknown finite-dimensional parameter $\Gamma(t)$. With a slight abuse of notation, let \bar{L} denote the set of covariates that are included in the model \mathcal{B} ; for example, if $L = (1, L_1)^T$, then a potential choice could be $\bar{L} = (1, L_1, L_1^2)^T$. We will postulate a linear model $d\Omega\{L; \Gamma(t)\} = d\Gamma(t)^T \bar{L}$. The parameters $\Gamma(t)$ can then be consistently estimated using Aalen least-squares (Aalen, 1980), upon changing the increments $dN(t)$ to $dN(t) - d\hat{\Psi}_{init}(t)A$, with $d\hat{\Psi}_{init}(t)$ being a consistent estimator of the time varying effect $d\Psi_{init}(t)$ under the initial model $E\{dN(t)|\mathcal{F}_t, L, A\} = [d\Omega\{L; \Gamma(t)\} + d\Psi_{init}(t)A] R(t)$.

We therefore arrive at the estimating function

$$\int_0^\infty \left(A - \frac{E[\lambda^{-1}\{\Gamma(t), \psi\} A \exp(-\psi At)|L; \alpha]}{E[\lambda^{-1}\{\Gamma(t), \psi\} \exp(-\psi At)|L; \alpha]} \right) \times \frac{[dN(t) - d\Omega\{L; \Gamma(t)\} - \psi A dt] R(t)}{\lambda\{\Gamma(t), \psi\}}, \quad (4)$$

where $\lambda\{\Gamma(t), \psi\} = d\Omega\{L; \Gamma(t)\}/dt + \psi A$. Then the population expectation of (4) converges to

$$\int_0^\infty E \left[E \left\{ \left(A - \frac{E[\lambda^{-1}\{\Gamma(t), \psi\} A \exp(-\psi At)|L; \alpha]}{E[\lambda^{-1}\{\Gamma(t), \psi\} \exp(-\psi At)|L; \alpha]} \right) \times \lambda^{-1}\{\Gamma(t), \psi\} \exp(-\psi At) \middle| L \right\} \times \exp\{-\Omega(t, L)\} [d\Omega(t, L) - d\Omega\{L; \Gamma(t)\}] \right]$$

It follows that the function in (4) has mean zero when, in addition to \mathcal{M} , either model \mathcal{A}_D or \mathcal{B} is correct. The efficient estimator under model $\mathcal{M} \cap \mathcal{A}_D$ is therefore *doubly robust* (Robins and Rotnitzky, 2001). The semiparametric efficiency bound under model $\mathcal{M} \cap \mathcal{A}_D$ is only met when model \mathcal{B} is correctly specified and thus attained locally.

We have argued that the score (3) is efficient under $\mathcal{M} \cap \mathcal{A}_D$. However, the same score in (3) also delivers an efficient doubly robust estimator; specifically, it is efficient under the union model $\mathcal{M} \cap (\mathcal{A}_D \cup \mathcal{B})$ at the intersection submodel $\mathcal{M} \cap \mathcal{A}_D \cap \mathcal{B}$. This follows from a general result in Robins and Rotnitzky (2001); however, the efficiency bound under $\mathcal{M} \cap \mathcal{B}$ may be lower than the bound under the union model.

3.2. Efficiency in a Subclass of Estimators

A drawback of the efficient score derived in the previous section is that it requires postulation of a model for the entire conditional distribution $f(A|L)$. In Web Appendix A, we therefore derive the subclass of influence functions which have mean zero when the conditional mean $E(A|L)$ is known. When A is binary, this conditional mean is known as the *propensity score* (Rosenbaum and Rubin, 1983). We are then

lead to the class of estimating functions

$$\int_0^\infty d(t, L) \{A - E(A|L)\} R(t) \exp(\psi At) \times \{dN(t) - d\Omega^*(t, L) - \psi A dt\}, \quad (5)$$

where $d(t, L)$ and $d\Omega^*(t, L)$ are arbitrary functions of t and L with finite variance. Note indeed that (5) no longer depends on $f(A|L)$ but only on the conditional mean $E(A|L)$. The term $R(t) \exp(\psi At)$ can be interpreted as the removal of the treatment effect (in expectation) from the at-risk indicators (Martinussen et al., 2011). That (5) has mean zero under model \mathcal{M} when $E(A|L)$ is known can be seen as follows:

$$E[d(t, L) \{A - E(A|L)\} R(t) \exp(\psi At) \{dN(t) - d\Omega^*(t, L) - \psi A dt\}] = E[d(t, L) \exp\{-\Omega(t, L)\} \{d\Omega(t, L) - d\Omega^*(t, L)\} E\{A - E(A|L)|L\}] = 0$$

for all $d(t, L)$. It is shown in Web Appendix A that the optimal choice of $d(t, L)$ for efficiency is

$$d_{\text{eff}}(t, L) = \frac{\text{var}(A|L)}{E[\{A - E(A|L)\}^2 \exp(\psi At) \lambda(t|A, L)|L]}; \quad (6)$$

for $d\Omega^*(t, L)$, it is equal to $d\Omega(t, L)$. The efficient estimator within this subclass is obtained by solving the equations

$$0 = \sum_{i=1}^n \int_0^\infty \frac{\text{var}(A_i|L_i)}{E[\{A_i - E(A_i|L_i)\}^2 \exp(\psi A_i t) \lambda(t|A_i, L_i)|L_i]} \times \{A_i - E(A_i|L_i)\} R_i(t) \exp(\psi A_i t) \times \{dN_i(t) - d\Omega(t, L_i) - \psi A_i dt\} \quad (7)$$

The conditional expectation $E(A|L)$ is typically unknown. It can be estimated under a parametric model \mathcal{A}_E for the conditional mean $E(A|L) = E(A|L; \beta)$, where $E(A|L; \beta)$ is a known function, smooth in an unknown finite-dimensional parameter β . Under model \mathcal{A}_E , β can be estimated using maximum likelihood.

Estimation of model \mathcal{A}_E does not affect the efficiency bound for the class of estimators identified by (5). Furthermore, misspecification does not induce bias when $d\Omega^*(t, L) = d\Omega(t, L)$ is consistently estimated, since the estimating function in (7) is unbiased under the union model $\mathcal{M} \cap (\mathcal{A}_E \cup \mathcal{B})$ and therefore doubly robust.

4. Implementation

4.1. Estimation via $f(A|L)$

In this section, we will build on the efficiency theory of the previous section and outline two potential estimation strategies for ψ . All of the estimators are consistent and asymptotically normal and accompanying variance estimators, unless stated otherwise, are given in Web Appendix B.

From the perspective of maximizing efficiency, a reasonable approach to take is to construct an estimator based on the efficient score (3). We note first that the score requires inverse weighting by the hazard function; this is also the case for efficient estimators of the parameters indexing other additive

hazards models (McKeague and Sasieni, 1994). In practice, this can lead to estimators with unstable performance in small-samples. If the hazard weights are removed from (3), as is common in standard fitting strategies for additive hazards models, we are left with the estimating function

$$\int_0^\infty \left[A - \frac{E\{A \exp(-\psi At) | L\}}{E\{\exp(-\psi At) | L\}} \right] R(t) \{dN(t) - d\Omega(t, L) - \psi A dt\} \tag{8}$$

The ratio

$$\frac{E\{A \exp(-\psi At) | L\}}{E\{\exp(-\psi At) | L\}} \tag{9}$$

is the first order derivative of the cumulant generating function for $f(A|L)$, evaluated at $-\psi t$. Under some model \mathcal{A}_D , one may evaluate this ratio directly or using Monte Carlo integration. Estimators of the asymptotic variance can be derived following standard M-estimation arguments.

However, specifying a correct model for a distribution in this fashion is unappealing, as it is difficult to formulate plausible models and any resulting misspecification may then have a potentially large impact on subsequent inference. Also, Monte Carlo integration may be computationally inconvenient as the integration needs to be done at all parameter values through which one iterates when numerically solving the equation. In light of these limitations, we will pursue alternative strategies in the remainder of this section.

We return to the estimating function (8). By Bayes rule, it follows that under model \mathcal{M} ,

$$f(A|T \geq t, L = l) = \frac{\exp(-\psi At) f(A|L = l)}{E\{\exp(-\psi At) | L = l\}}$$

for all l . Therefore,

$$E(A|T \geq t, L) = \frac{E\{A \exp(-\psi At) | L\}}{E\{\exp(-\psi At) | L\}} \tag{10}$$

which suggests that the unbiased estimating function (8) can also be written as

$$\int_0^\infty \{A - E(A|T \geq t, L)\} R(t) \{dN(t) - d\Omega(t, L) - \psi A dt\} \tag{11}$$

Rather than modeling (8) indirectly via a model for the distribution of A given L , we may choose to instead specify a model \mathcal{A}_{TV} for the time-varying propensity score $E(A|T \geq t, L)$. A question then is how to specify a model at each time t that is congenial with model \mathcal{M} . Specifically, a parameterization of \mathcal{A}_{TV} is congenial with model \mathcal{M} if for each element in \mathcal{A}_{TV} and \mathcal{M} , there exists a distribution $f(A|L)$ such that the equality (10) holds. If no such distribution exists, then we know before even seeing the data that the proposed models for \mathcal{M} and \mathcal{A}_{TV} cannot both be correct. In Web Appendix A, we show that

the following generalized linear model

$$E(A|T \geq t, L) = E\{A|T \geq t, L; \theta(t)\} = g^{-1}\{\theta(t)^T \tilde{L}\}$$

is *always* congenial with \mathcal{M} when the dispersion parameter for $f(A|L)$ does not depend on L . Here, $g(\cdot)$ is a canonical link function; and \tilde{L} is the vector of covariates that are included in the model \mathcal{A}_{TV} . A similar estimating function to (11) appears in Kang et al. (2018); however, they use a different parameterization of \mathcal{A}_{TV} to the one we give above, which requires estimation of $P(T \geq t|A, L)$, $P(T \geq t|L)$, and $E(A|L)$. Parametric models for $P(T \geq t|A, L)$ and $P(T \geq t|L)$ may not be congenial with model $\mathcal{M} \cap \mathcal{B}$, which undermines the feasibility of doubly robust inference. Their proposal therefore relies on kernel density estimators, which are not suitable when L is high-dimensional.

An advantage of our parameterization of \mathcal{A}_{TV} is that it admits a closed-form estimator of ψ , which is defined as

$$\hat{\psi}_{TVPS-DR} = \frac{\sum_{i=1}^n \int_0^\infty \Lambda_i\{\hat{\theta}(t)\} R_i(t) J(t) [dN_i(t) - d\Omega\{L_i; \hat{\Gamma}(t)\}]}{\sum_{i=1}^n \int_0^\infty \Lambda_i\{\hat{\theta}(t)\} R_i(t) A_i dt} \tag{12}$$

Here, $\Lambda\{\theta(t)\} = A - E\{A|T \geq t, L; \theta(t)\}$ and $J(t) = 1$ if both $Y(t)$ and $\tilde{L}(t)$ have full rank and zero otherwise, where $Y(t)$ denotes a matrix with i th row $R_i(t)(\tilde{L}_i^T, A_i)$ and similarly for $\tilde{L}(t)$. It follows from the theory of M-estimation that this estimator is consistent and asymptotically normal under model $\mathcal{M} \cap (\mathcal{A}_{TV} \cup \mathcal{B})$.

When $\tilde{L} = \dot{L}$ (we hereby denote the common set of covariates by \dot{L}), the previous expressions can be further simplified. Given the use of the canonical link function, $\theta(t)$ can be estimated at time t as the solution to the estimating equations

$$0 = \sum_{i=1}^n \dot{L}_i^T R_i(t) [A_i - E\{A_i | T_i \geq t, L_i; \hat{\theta}(t)\}]$$

By estimating $\theta(t)$ in this way, we ensure that the estimating equations for ψ reduce to

$$0 = \sum_{i=1}^n \int_0^\infty \Lambda_i\{\theta(t)\} R_i(t) \{dN_i(t) - \psi A_i dt\} \tag{13}$$

and ψ can be estimated in closed-form as

$$\hat{\psi}_{TVPS-DR} = \frac{\sum_{i=1}^n \int_0^\infty \Lambda_i\{\hat{\theta}(t)\} dN_i(t) J(t)}{\sum_{i=1}^n \int_0^\infty \Lambda_i\{\hat{\theta}(t)\} R_i(t) A_i dt} \tag{14}$$

Surprisingly, estimation of $\Gamma(t)$ is no longer required, yet the doubly robust property is retained. To see why, note that when model \mathcal{A}_{TV} is misspecified, the expectation of (13) will converge to

$$\int_0^\infty E\{([E(A|T \geq t, L) - E\{A|T \geq t, L; \theta^*(t)\}]) R(t) d\Omega(t, L)\},$$

where $\theta^*(t)$ is the limiting value of $\hat{\theta}(t)$. Because of how $\theta(t)$ is estimated, the above display will equal zero when $d\Omega(t, L) = d\Omega\{L; \Gamma(t)\}$, thus demonstrating double robustness. This strategy is related to bias-reduced doubly robust estimation, as proposed by Vermeulen and Vansteelandt (2015); further discussion is given in Web Appendix A.

Vansteelandt et al. (2014) showed that Aalen least-squares estimators are robust to misspecification of the additive hazards model when A is normal with a mean that is linear in $\bar{L} = \tilde{L}$ and constant variance. Indeed, in this scenario, it follows from the Appendix of Vansteelandt et al. (2014) that the Aalen least-squares estimator is equivalent to the estimator given in (14). Using a Taylor expansion around (9),

$$\begin{aligned} \frac{E\{A \exp(-\psi A t) | L\}}{E\{\exp(-\psi A t) | L\}} &= E(A|L) + \text{var}(A|L)\psi t \\ &+ \frac{E(A^3|L) - 3E(A|L)E(A^2|L) + 2E(A|L)^3}{2!}\psi^2 t^2 + \dots, \end{aligned}$$

it follows that this robustness holds more generally so long as the mean and central moments of $f(A|L)$ are linear in L . This assumption would not generally hold if A is binary; however, our estimator given in (12) generalizes the robustness properties of Aalen least-squares to arbitrary exposure distributions. Furthermore, if the true treatment effect $\psi^*(t, L)$ is a function of t and L , such that model \mathcal{M} no longer holds, then the estimator defined in (12) continues to have a useful interpretation. Assuming model \mathcal{A}_{TV} is correct, then the estimator converges to

$$\frac{\int_0^\infty E\{\text{var}(A|T \geq t, L)R(t)\psi^*(t, L)\}dt}{\int_0^\infty E\{\text{var}(A|T \geq t, L)R(t)\}dt},$$

which is a weighted average of the treatment effects at different times and covariate values. In contrast, the Aalen least-squares estimator of ψ in the corresponding additive hazards model is not generally a convex combination of the time/covariate-specific treatment effects, even when L is correctly modeled. It is in particular not guaranteed to lie within the range of time/covariate-specific treatment effects.

4.2. Estimation via $E(A|L)$

In Section 3.2, we identified a subclass of estimators that are consistent under a correctly specified model of the conditional mean of the exposure. We now exploit these results in order to develop inference for ψ .

Consider the estimating equations for ψ suggested by the function (5):

$$\begin{aligned} 0 &= \sum_{i=1}^n \int_0^\infty d(t, L_i) \{A_i - E(A_i|L_i)\} R_i(t) \exp(\psi A_i t) \\ &\times \{dN_i(t) - d\Omega(t, L_i) - \psi A_i dt\} \end{aligned} \quad (15)$$

In evaluating the integral in (15), note that the function $d(t, L)$ impacts only the variance, rather than the unbiasedness of the estimating equations. The efficient choice (6) depends on the conditional distribution of the treatment A ,

and in certain cases may lead to the integral becoming analytically intractable. We therefore set $d(t, L)$ to 1, leaving the search for efficient yet computationally feasible choices to future work.

Letting $\Delta(\beta)$ denote $\{A - E(A|L; \beta)\}$, the estimating equation becomes

$$0 = \sum_{i=1}^n \int_0^\infty \Delta_i(\beta) R_i(t) \exp(\psi A_i t) \{dN_i(t) - d\Omega(t, L_i) - \psi A_i dt\}$$

and ψ can therefore be estimated as a solution to the equations

$$0 = \sum_{i=1}^n \Delta_i(\hat{\beta}) \left\{ 1 - \int_0^{T_i} d\Omega(s, L_i) \exp(\psi A_i s) \right\} \quad (16)$$

It is vital for identification that the arbitrary function $d\Omega(t, L)$ is non-zero over a set of times t with positive Lebesgue measure. Otherwise, as we integrate to ∞ , all information about the parameter ψ is lost. Therefore, $d\Omega(t, L)$ can be seen as weighting term that prevents the integral in (15) from equaling zero at all ψ .

Setting $d\Omega(t, L_i) = 1$ for all $t > 0$, the above equations reduce to $0 = \sum_{i=1}^n U_i(\psi, \hat{\beta})$, where

$$U_i(\psi, \hat{\beta}) = \begin{cases} \Delta_i(\hat{\beta})\{\exp(\psi A_i T_i) - 1\}/\psi A_i & \text{if } \psi A_i \neq 0 \\ \Delta_i(\hat{\beta})T_i & \text{if } \psi A_i = 0 \end{cases}$$

It follows from the theory of M-estimation that under standard regularity conditions, the solution $\hat{\psi}_{BPS}$ to equation (15) delivers an estimator which is consistent and asymptotically normal under model $\mathcal{M} \cap \mathcal{A}_E$.

Doubly robust extensions to the previous proposal can also be made. Returning to equation (15), then rather than setting $d\Omega(t, L_i) = 1$ for all $t > 0$, we now postulate a model \mathcal{B} for $d\Omega(t, L)$, such as $d\Omega(t, L) = d\Omega\{L; \Gamma(t)\}$. After setting $d(t, L_i) = 1$ again for all $t > 0$, it follows that ψ can be estimated as the solution to

$$\begin{aligned} 0 &= \sum_{i=1}^n \Delta_i(\hat{\beta}) \left\{ \left(\int_0^{T_i} J(s) \exp(\psi A_i s) [dN_i(s) \right. \right. \\ &\quad \left. \left. - d\Omega\{L_i; \hat{\Gamma}(s)\}] - \exp(\psi A_i T_i) \right\} \end{aligned} \quad (17)$$

The resulting solution $\hat{\psi}_{BPS-DR}$ is consistent and asymptotically normal under the model $\mathcal{M} \cap (\mathcal{A}_E \cup \mathcal{B})$.

It is straightforward to show that the efficient subclass score (7) is invariant to centering A by its conditional mean; if we are willing to work with a non-efficient estimator, it is also desirable that it has this property. We therefore recommend that A be substituted by $\Delta(\beta)$ in (16) and (17), such that the estimating equations implied by (16) reduce to $U_i(\psi, \hat{\beta}) = [\exp\{\psi \Delta_i(\hat{\beta}) T_i\} - 1]/\psi$ if $\Delta_i(\hat{\beta}) \neq 0$ and $\Delta_i(\hat{\beta}) T_i$ otherwise. Centering will prevent the exponential terms in the estimating equations from becoming large at later time points, which could lead to improved finite-sample performance.

Solving the equation $U(\psi, \hat{\beta})$ for the singly robust estimator could also be a computationally-fast first step towards an estimator that is nearly efficient (in the general class), if one is willing to specify the distribution $f(A|L)$. This is because an initial estimate $\hat{\psi}_{init}$ could be plugged into ratio term (9) in the estimating function (8), making it linear in ψ . Under model \mathcal{M} , the resulting two-step estimator is consistent and asymptotically normal if either or both of the models \mathcal{A}_D or \mathcal{B} hold.

4.3. Censoring

When the survival time is censored by C , all of the approaches described above are valid under the assumption that censoring is independent of T and A , conditional on L , in the sense that $C \perp (T, A) | L$. All approaches are thus consistent when censoring depends only on L . The doubly robust estimators are moreover consistent when censoring depends additionally on A , and the additive hazards model is correctly specified. We can relax these assumptions by using inverse probability of censoring weighting (Scharfstein and Robins, 2002), under a model for the censoring mechanism:

$$E \{ dN_C(t) | T \geq t, C \geq t, L, A, \bar{V}_t \} = a(t, L, A, \bar{V}_t; \pi),$$

Here, $N_C(t)$ is the counting process for the censoring time; $\bar{V}_t = \{V_s : s < t\}$, where V_s is a collection of covariates measured at time s ; and $a(t, L, A, \bar{V}_t; \pi)$ is a known function, smooth in an unknown parameter π . An additive or multiplicative hazards model could be postulated here. An individual's contribution to the estimating function (12) at time t is then weighted by

$$\frac{1}{\exp\{-\int_0^t a(s, L, A, \bar{V}_s; \pi) ds\}},$$

such that ψ can be estimated as the solution to

$$0 = \sum_{i=1}^n \int_0^\infty \left[\frac{1}{\exp\{-\int_0^t a(s, L_i, A_i, \bar{V}_{si}; \hat{\pi}) ds\}} \right] \Delta_i \{ \hat{\theta}(t) \} R_i(t) J(t) \times [dN_i(t) - d\Omega\{L_i; \hat{\Gamma}(t)\} - \psi A_i dt] \quad (18)$$

Weights can also be added to the estimating equations in (15), such that ψ can be estimated as the solution to

$$0 = \sum_{i=1}^n \Delta_i(\hat{\beta}) \left(1 - \int_0^{T_i} \left[\frac{1}{\exp\{-\int_0^s a(u, L_i, A_i, \bar{V}_{ui}; \pi) du\}} \right] \times d\Omega(s, L_i) \exp(\psi A_i s) ds \right),$$

and likewise for the doubly robust estimator given by (17).

When the resulting weights are highly variable, stabilized inverse probability weights can be obtained under an additional model for the censoring mechanism:

$$E \{ dN_C(t) | T \geq t, C \geq t, L \} = a(t, L; \kappa),$$

where $a(t, L; \kappa)$ is a known function, smooth in an unknown parameter κ . Misspecification of the latter model

does not affect the consistency of the estimator of ψ (Robins et al., 2000).

5. Simulation Study

We considered 4 estimators: *i*) the singly robust estimator $\hat{\psi}_{BPS}$ described in Section 4.2 that is consistent under model $\mathcal{M} \cap \mathcal{A}_E$; *ii*) the doubly robust estimator $\hat{\psi}_{BPS-DR}$ based on display (17) that is consistent under model $\mathcal{M} \cap (\mathcal{A}_E \cup \mathcal{B})$; *iii*) the doubly robust estimator $\hat{\psi}_{TVPS-DR}$ given in closed-form in (12) that is consistent under model $\mathcal{M} \cap (\mathcal{A}_{TV} \cup \mathcal{B})$; and *iv*) the Aalen least-squares estimator $\hat{\psi}_{ALS}$ of the time-constant treatment coefficient from a covariate-adjusted additive hazards model (where the effects of the baseline covariates were allowed to vary over time) that is consistent under model $\mathcal{M} \cap \mathcal{B}$. Model-based standard errors were used to construct 95% confidence intervals for $\hat{\psi}_{ALS}$ (the variance estimators used to construct the other 95% confidence intervals are described in Web Appendix B).

In order to evaluate the four different estimators, we considered eight different experiments. For each experiment, we simulated 1000 data sets of 1000 observations. We generated covariates L_1 and L_2 , and exposure A , event time T and censoring time C ; an individual's follow up time was taken as $\min(T, C)$. In experiments 1–4, the exposure A was continuous (e.g., the increase in the dose of a drug), whereas it was binary in experiments 5–8 (see Table 1 for a descriptions of the data generating mechanisms). In experiments 1, 2, 5, and 6, all working models included only terms for L_1 , L_2 , and an intercept. In experiments 1 and 5, all models were correctly specified, whereas in experiments 2 and 6, the models \mathcal{A}_E and \mathcal{A}_{TV} were misspecified, as they excluded an interaction term. In experiments 3 and 7, the models \mathcal{A}_E and \mathcal{A}_{TV} correctly included an interaction term, whereas this term was excluded in experiments 4 and 8 (all models are wrong). In experiments 1, 2, 5, and 6, those for whom $\min(T, C) \geq 1.6$ were censored at this time point. The same was done in experiments 3, 4, 7 and 8 at $t = 1.3$. For all experiments, the chosen censoring mechanisms lead to 25–30% of subjects being censored (with around 10% censored at the end of the study).

The results of the simulations are given in Table 2, and largely corroborate the theory. The doubly robust estimators were empirically unbiased when either of the working models was correctly specified, and when both models were correctly specified, they were more efficient than the estimator $\hat{\psi}_{BPS}$. There was little difference in efficiency between the estimators, and they also behaved similarly in terms of bias when both working models were misspecified. In general, the standard errors performed well. When the treatment was normal however, under misspecification of model \mathcal{A}_E , performance of the standard errors for $\hat{\psi}_{BPS}$ and $\hat{\psi}_{BPS-DR}$ was less than optimal because the distribution of the estimation function was characterized by outlying values.

6. Data Analysis

We applied our methods to data from the five-center SUPPORT study that took place between 1989 and 1994. Previously, Connors et al. (1996) analyzed the dataset in order to evaluate the effect of right heart catheterization (RHC)

Table 1

A description of the data-generating mechanisms behind experiments 1–8. In experiments 1–4, we standardized the exposure to give it mean zero and standard deviation 1. $B(1, p)$: Bernoulli distribution with expectation p ; $\mathcal{N}(\mu, \sigma^2)$: normal distribution with expectation μ and variance σ^2 ; $\text{Exp}(\lambda)$: exponential distribution with rate λ ; $\text{unif}(a, b)$: uniform distribution with minimum and maximum values a and b , respectively. We use $L = (1, L_1, L_2)^T$; in all settings, for model \mathcal{B} , we fitted $E\{dN(t)|\mathcal{F}_t, L, A = 0\} = d\Gamma(t)^T L R(t)$.

Exp.	Data-generating mechanism	Fitted exposure model
1	$L_1 \sim B(1, 0.6)$ $L_2 \sim B[1, \text{expit}(0.5L_1)]$ $A \sim \mathcal{N}\{-1 + 0.25(L_1 - L_2), 0.09\}$ $T \sim \text{Exp}(0.5 + 0.5L_1 + L_2 + 0.1A)$ $C \sim \text{unif}(0, 3.5)$	$\mathcal{A}_E : \beta^T L$ $\mathcal{A}_{TV} : \theta(t)^T L$
2	$L_1 \sim B(1, 0.6)$ $L_2 \sim B[1, \text{expit}(0.5L_1)]$ $A \sim \mathcal{N}\{-1 + 0.25(L_1 - L_2) + 0.5L_1L_2, 0.09\}$ $T \sim \text{Exp}(0.5 + 0.5L_1 + L_2 + 0.1A)$ $C \sim \text{unif}(0, 3.5)$	$\mathcal{A}_E : \beta^T L$ $\mathcal{A}_{TV} : \theta(t)^T L$
3, 4	$L_1 \sim \mathcal{N}(0, 1)$ $L_2 \sim \mathcal{N}(L_1, 1)$ $A \sim \mathcal{N}\{-1 + 0.25(L_1 - L_2) + 0.5L_1L_2, 0.09\}$ $T \sim \text{Exp}\{0.3 + L_1 + \log(1 + L_2) + 0.1A\}$ $C \sim \text{unif}(0, 3)$	3 - $\mathcal{A}_E : \beta_1^T L + \beta_2 L_1 L_2$ $\mathcal{A}_{TV} : \theta_1(t)^T L + \theta_2(t) L_1 L_2$ 4 - $\mathcal{A}_E : \beta^T L$ $\mathcal{A}_{TV} : \theta(t)^T L$
5	$L_1 \sim B(1, 0.6)$ $L_2 \sim B[1, \text{expit}(0.5L_1)]$ $A \sim B[1, \text{expit}\{-1 + 0.25(L_1 - L_2)\}]$ $T \sim \text{Exp}(0.5 + 0.5L_1 + L_2 + 0.1A)$ $C \sim \text{unif}(0, 3.5)$	$\mathcal{A}_E : \text{expit}(\beta^T L)$ $\mathcal{A}_{TV} : \text{expit}\{\theta(t)^T L\}$
6	$L_1 \sim B(1, 0.6)$ $L_2 \sim B[1, \text{expit}(0.5L_1)]$ $A \sim B[1, \text{expit}\{-1 + 0.25(L_1 - L_2) + 0.5L_1L_2\}]$ $T \sim \text{Exp}(0.5 + 0.5L_1 + L_2 + 0.1A)$ $C \sim \text{unif}(0, 3.5)$	$\mathcal{A}_E : \text{expit}(\beta^T L)$ $\mathcal{A}_{TV} : \text{expit}\{\theta(t)^T L\}$
7, 8	$L_1 \sim \mathcal{N}(0, 1)$ $L_2 \sim \mathcal{N}(L_1, 1)$ $A \sim B[1, \text{expit}\{-1 + 0.25(L_1 - L_2) + 0.5L_1L_2\}]$ $T \sim \text{Exp}\{0.3 + L_1 + \log(1 + L_2) + 0.1A\}$ $C \sim \text{unif}(0, 3)$	7 - $\mathcal{A}_E : \text{expit}(\beta_1^T L + \beta_2 L_1 L_2)$ $\mathcal{A}_{TV} : \text{expit}\{\theta_1(t)^T L + \theta_2(t) L_1 L_2\}$ 8 - $\mathcal{A}_E : \text{expit}(\beta^T L)$ $\mathcal{A}_{TV} : \text{expit}\{\theta(t)^T L\}$

Table 2

Simulation results from experiments 1 to 8. Monte Carlo bias multiplied by 10 (bias); Monte Carlo standard deviation multiplied by 10 (SD); coverage of 95% Wald confidence intervals (Cov).

Exp	$\hat{\psi}_{BPS}$			$\hat{\psi}_{BPS-DR}$			$\hat{\psi}_{TVPS-DR}$			$\hat{\psi}_{ALS}$		
	Bias	SD	Cov	Bias	SD	Cov	Bias	SD	Cov	Bias	SD	Cov
1	-0.00	0.6	95.4	0.00	0.5	95.0	-0.00	0.5	94.9	-0.00	0.5	95.3
2	-0.57	0.6	86.4	-0.07	0.5	93.9	0.02	0.6	94.5	0.02	0.6	94.6
3	0.03	2.1	94.7	0.04	1.8	94.9	-0.00	1.8	95.3	7.19	1.1	0.0
4	7.80	1.8	0.0	7.86	1.8	0.0	7.17	1.1	0.0	7.17	1.1	0.0
5	0.01	1.2	94.9	0.05	1.1	94.8	0.04	1.1	94.8	0.04	1.1	94.7
6	-0.10	1.1	94.2	0.03	1.0	95.4	0.02	1.0	95.4	0.02	1.0	95.7
7	-0.01	1.5	95.0	0.03	1.3	95.0	-0.02	1.4	94.5	3.80	1.4	17.3
8	3.91	1.5	25.8	3.68	1.3	19.1	3.70	1.3	19.3	3.75	1.4	18.2

on patient mortality. Many cardiologists believed that RHC was beneficial, but due to ethical concerns, this benefit had not been demonstrated via a randomized controlled trial. Connors et al. (1996) constructed propensity scores using 72 potential confounders identified by clinicians. As part of the original analysis, Connors et al. (1996) evaluated the association between RHC and survival by fitting a Cox model, adjusted for treatment (RHC), the propensity score and a reduced set of outcome adjustment variables. They considered only the first thirty days after entry into the study, such that people who survived beyond thirty days were considered administratively censored. Surprisingly, the investigators found that undergoing RHC (compared with no RHC) led to a decrease in survival.

We attempted to fit an additive hazards model using Aalen least-squares, adjusting for treatment and all 72 covariates. Random noise from a uniform distribution on $(0, 0.001)$ was first added to the survival times in order to break ties. We also forced the effects of all covariates (excluding the intercept) to be constant over time. The estimate of the coefficient for treatment in the final adjusted model was 0.00365 (SE = 0.00077, 95%CI 0.00213–0.00516). To assess the fitted model, we obtained predicted hazards at each time point (day 1, 2, ..., 30); over 99% of the predicted hazards at the 25th, 50th, and 75th percentiles of the distribution of survival times were less than zero. We would be cautious about drawing inferences from such a model, as the misspecification (suggested by the invalid predictions) could lead to biased estimates of the treatment effect and misleading model-based standard errors. In view of this, we will next report estimates that rely on propensity scores instead.

We fitted a logistic regression model for treatment, adjusting for all 72 covariates. We obtained the predicted values from this model in order to estimate the effect of RHC using the estimator $\hat{\psi}_{BPS}$. We also postulated a simplified additive hazards model, which included treatment and the outcome adjustment variables listed above, and allowed their effects to depend on time. The purpose of fitting this model was to

construct the doubly robust estimators $\hat{\psi}_{BPS-DR}$ and $\hat{\psi}_{TVPS-DR}$. For $\hat{\psi}_{TVPS-DR}$, we obtained the predictions after fitting a logistic regression model that included all 72 variables at each event time. R code is available in the Supplementary Materials (see also Web Appendix C).

The estimated hazard difference from the singly-robust estimator $\hat{\psi}_{BPS}$ was 0.00406 (SE = 0.00084, 95%CI 0.00241–0.0057). The doubly robust estimator $\hat{\psi}_{BPS-DR}$ gave a hazard difference of 0.00334 (SE = 0.00074, 95%CI 0.0019–0.00479), and the estimator $\hat{\psi}_{TVPS-DR}$ gave 0.00363 (SE = 0.00078, 95%CI 0.0021–0.00515). We also computed $E\{\text{pr}(T \geq t | A = a, L)\}$, the average survival probability at time t if everyone received treatment a , standardized with respect to the observed distribution of L . This was done by taking sample averages of $R(t) \exp\{\hat{\psi}(A - a)t\}$. The adjusted survival curves are plotted in Figure 1; compared with the standard Kaplan–Meier estimates, the difference between the treatment groups is slightly shrunken towards the null. This is consistent with the results of Connors et al. (1996) and Vermeulen and Vansteelandt (2015), where small differences between unadjusted and adjusted treatment effect estimates (in the same direction) were observed.

7. Discussion

In this article, we have developed a theory of estimation for the adjusted hazard difference/relative chance of survival using semiparametric additive hazards models. We have used this theory to develop several classes of doubly robust estimators, each strategy with its own strengths and limitations. The closed-form estimators described in Section 4.1, which are consistent and asymptotically normal under model $\mathcal{M} \cap (\mathcal{A}_{TV} \cup \mathcal{B})$, have several important advantages over competing strategies: under omitted interactions and/or time effects in model \mathcal{M} , they converge to a convex combination of time/covariate-specific treatment effects (assuming that model \mathcal{A}_{TV} is correct); since the term $\exp(\psi At)$ does not appear in the estimating functions, their behavior

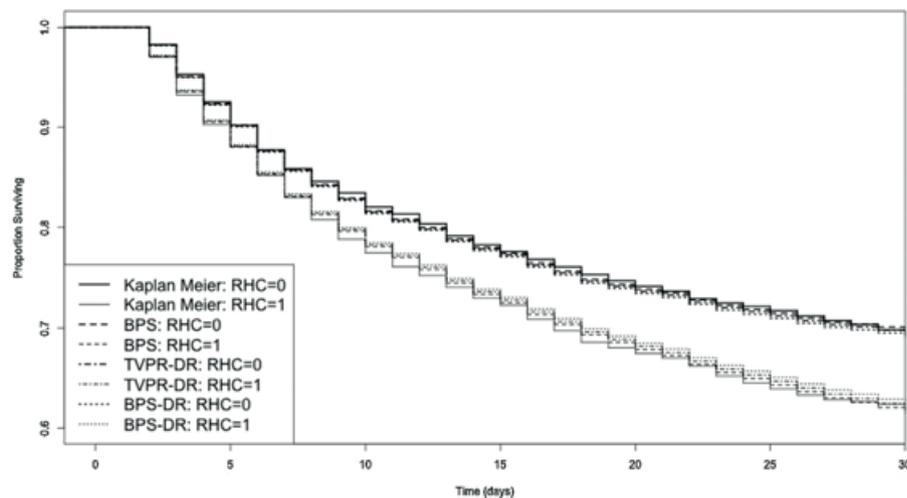


Figure 1. Thirty-day survival curves. $\hat{\psi}_{BPS}$, $\hat{\psi}_{BPS-DR}$ and $\hat{\psi}_{TVPS-DR}$ were used to compute $E\{\text{pr}(T \geq t | A = a, L)\}$, the average survival probability at time t if everyone received treatment a (RHC or no RHC), standardized with respect to the observed distribution of the covariates L .

is likely to be more stable; and simulations suggest that these estimators are reasonably efficient (although information may be lost by ignoring the fact that model \mathcal{A}_{TV} contains information about ψ). In Web Appendix D, we compare these estimators with those of Kang et al. (2018) in additional simulations, and find ours to be more efficient in a low-dimensional setting.

Although for simplicity we have considered a scalar treatment effect, in many settings ψ will be a vector of parameters. For instance, when the effect of A on the hazard is modified by Z , then the semiparametric model is now defined by the restriction

$$E \{dN(t)|\mathcal{F}_t, L, A\} = \{d\Omega(t, L) + (\psi_1 + \psi_2^T Z)Adt\} R(t)$$

where ψ_2 is a vector, and ψ_1 and ψ_2 give the adjusted effect of A within different levels of Z . We note that unlike the estimators of Wang et al. (2017), all of our proposals remain doubly robust outside of the ‘no treatment heterogeneity’ model (and are considerably more efficient under the homogeneous model by avoiding inverse probability weighting). If the vector-valued $\psi = (\psi_1, \psi_2)^T$ is estimated via the approach described in Section 4.1 that requires a model for $E(A|T \geq t, L)$, then to ensure a congenial model specification, the model \mathcal{A}_{TV} must now also include Z , with its regression coefficient(s) allowed to depend on time.

The proposals described in this article are closely related to the method of *G-estimation* (Robins and Tsiatis, 1991; Robins, 1994). Picciotto et al. (2012) recently introduced the class of discrete-time structural cumulative failure time models, along with accompanying G-estimators. They postulate a multiplicative model for the probability of failure, rather than survival. Note that all of the methods described above can be adjusted in order to estimate the relative chance of failure. In Web Appendix A, we further investigate the efficiency of Picciotto et al.’s estimators.

In future work, the new class of semiparametric additive hazards models and the accompanying doubly robust estimators will be extended to estimate controlled direct effects in the presence of mediators and/or time varying confounders. Regarding mediation problems, Martinussen et al. (2011) previously considered the estimation of direct effects using additive hazards models; however, their approach was limited to settings with binary, randomly assigned treatments. By taking a semiparametric approach, the methods described in this article would be able to accommodate different types of treatment and adjustment for baseline covariates. Regarding the problem of time-varying confounding, such extensions would be useful in light of issues facing the two principal approaches used in survival analysis, marginal structural models (Robins et al., 2000) and structural accelerated failure time models (Robins and Tsiatis, 1991). Inference for the former requires inverse probability weighting; estimators can suffer heavily from large finite-sample bias and imprecision due to highly variable weights. Marginal structural models also prohibit investigation into effect modification by time-varying covariates. G-estimation has turned out to be problematic for structural accelerated failure time models, because administrative censoring is dealt with through

an artificial recensoring process which can induce a lack of smoothness in the estimating equations.

8. Supplementary Materials

Web Appendices referenced in Sections 2–7 are available with this article at the *Biometrics* website on Wiley Online Library. R code for the data analysis is also available here.

ACKNOWLEDGEMENTS

Oliver Dukes is supported by a Strategic Basic Research PhD grant from the Research Foundation—Flanders (FWO). Torben Martinussen’s work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. Eric J. Tchetgen Tchetgen is funded by the National Institutes of Health grant A1104459. The authors are grateful to Shaun Seaman for helpful discussions, and to Wenbin Lu and Suhyun Kang for providing R code.

REFERENCES

- Aalen, O. (1980). A model for nonparametric regression analysis of counting processes. In *Lecture Notes in Statistics*, Vol. 2, 1–25. New York, NY: Springer New York.
- Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins series in the mathematical sciences. Baltimore: Johns Hopkins University Press.
- Connors, A. F., Speroff, T., Dawson, N. V., Thomas, C., Harrell, F. E., Wagner, D., et al. (1996). The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA* **276**, 889–897.
- Farrell, M. H. (2015). Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics* **189**, 1–23.
- Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21**, 13–15.
- Kang, S., Lu, W., and Zhang, J. (2018). On estimation of the optimal treatment regime with the additive hazards model. *Statistica Sinica* In press. DOI: 10.5705/ss.202016.0543.
- Martinussen, T., Vansteelandt, S., Gerster, M., and Hjelmberg, J. v. B. (2011). Estimation of direct effects for survival data by using the Aalen additive hazards model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 773–788.
- McKeague, I. W. and Sasieni, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–514.
- Picciotto, S., Hernán, M. A., Page, J. H., Young, J. G., and Robins, J. M. (2012). Structural nested cumulative failure time models to estimate the effects of interventions. *Journal of the American Statistical Association* **107**, 886–900.
- Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics—Theory and Methods* **23**, 2379–2412.
- Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.
- Robins, J. M. and Rotnitzky, A. (2001). Comments. *Statistica Sinica* **11**, 920–936.
- Robins, J. M. and Tsiatis, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving

- structural failure time models. *Communications in Statistics—Theory and Methods* **20**, 2609–2631.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Scharfstein, D. O. and Robins, J. M. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89**, 617–634.
- Tchetgen Tchetgen, E. J., Robins, J. M., and Rotnitzky, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. *Biometrika* **97**, 171–180.
- Vansteelandt, S., Martinussen, T., and Tchetgen Tchetgen, E. J. (2014). On adjustment for auxiliary covariates in additive hazard models for the analysis of randomized experiments. *Biometrika* **101**, 237–244.
- Vermeulen, K. and Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. *Journal of the American Statistical Association* **110**, 1024–1036.
- Wang, Y., Lee, M., Liu, P., Shi, L., Yu, Z., Abu Awad, Y., et al. (2017). Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. *Epidemiology* **28**, 771–779.

Received June 2017. Revised May 2018. Accepted May 2018.