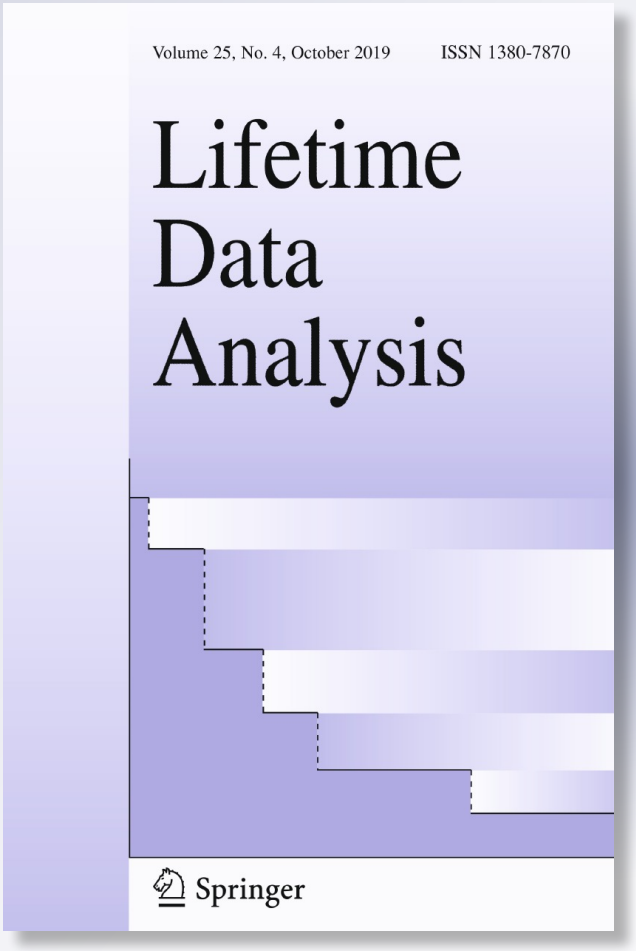
A causal proportional hazards estimator under homogeneous or heterogeneous selection in an IV setting

Ditte Nørbo Sørensen, Torben Martinussen & Eric Tchetgen Tchetgen

Lifetime Data Analysis
An International Journal Devoted to Statistical Methods and Applications for Time-to-Event Data

ISSN 1380-7870
Volume 25
Number 4

Lifetime Data Anal (2019) 25:639-659
DOI 10.1007/s10985-019-09476-y



Springer

Lifetime Data Analysis (2019) 25:639–659
https://doi.org/10.1007/s10985-019-09476-y

A causal proportional hazards estimator under homogeneous or heterogeneous selection in an IV setting

Ditte Nørbo Sørensen[1] · **Torben Martinussen**[1] · **Eric Tchetgen Tchetgen**[2]

Received: 28 March 2018 / Accepted: 26 April 2019 / Published online: 7 May 2019

Abstract

In this paper we present a framework to do estimation in a structural Cox model when there may be unobserved confounding. The model is phrased in terms of a selection bias function and a baseline model that describes how covariates affect the survival time in a scenario without exposure. In this way model congeniality is ensured. The method uses an instrumental variable. Interestingly, the formulated model turns out to have similarities to the so-called Cox–Aalen survival model for the observed data. We exploit this to enhance estimation of the unknown parameters. This also allows us to derive large sample properties of the proposed estimator.

Keywords Causal effect · Structural Cox model · Instrumental variable · Treatment effect on the treated · Selection bias function

1 Introduction

Unobserved confounding is often a shortcoming in analyses of observational data and may lead to less reliable results than analyses of randomized trial data. Theoretically, the existence of an instrumental variable (IV) enables causal inference from observational studies. The simplicity of the IV methods used for linear models is not transferred to nonlinear models and this has led to different proposals on how to use

✉ Torben Martinussen
tma@sund.ku.dk

Ditte Nørbo Sørensen
diso@sund.ku.dk

Eric Tchetgen Tchetgen
ett@wharton.upenn.edu

[1] Section of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B, 1014 Copenhagen K, Denmark

[2] Statistics Department, Wharton, University of Pennsylvania, 467 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, USA

IV's when the standard methods such as two-stage regression lead to bias. For binary outcome, several causal estimands have been considered in the literature of dichotomous instrument and treatment variables, see for example Clarke and Windmeijer (2010, 2012). One of the mentioned estimands is the causal effect among compliers, where a complier is defined in the context of a randomized experiment as a subject who receives the treatment she is assigned to. A complier is thus characterized by the fact that she complies with the observed randomization variable, but also that she would comply had she belonged to the other randomization group. Hence, the classification into compliers and non-compliers is impossible based on the observed data only. Alternatively, one may focus on the treated subjects and estimate the effect of treatment among the treated (ETT). Notice that some of the treated subjects are compliers while others are not, and we can not tell who, since the compliers are defined by a partly unobserved compliance pattern. Based on the IV assumptions alone it is impossible to find the complier effect or the ETT from the observed random variables. Such identification require extra assumptions such as monotonicity or no-effect-modification (NEM). For linear models, the first identifies the effect among compliers and the latter the ETT, see Angrist and Imbens (1994), Robins and Rotnitzky (2004) and Hernan and Robins (2006) for further description of models and assumptions for identification. The effect among compliers and the ETT along with their identifying assumptions are substantially different unless the no contamination assumption holds. No contamination rules out individuals receiving treatment if they are randomized to no treatment, which is often too restrictive for many real-world applications. For survival outcome, the complier effect in a proportional hazards model has been studied by Loeys and Goetghebeur (2003) under no contamination, and by Cuzick et al. (2007) under monotonicity. Situations beyond binary instrument and exposure variable as often met in epidemiological data does not fit into the framework of these methods. One way out is to dichotomize variables by an artificial threshold value to fit into the principal stratification framework, but this loss of information can cause violation of underlying model assumptions. MacKenzie et al. (2014) proposed a proportional hazards model for generic instruments and exposures estimating the log of a population average causal hazard ratio. However, the estimation relies on approximations and assumptions on the nature of the unobserved confounding that are unappealing (Tchetgen Tchetgen et al. 2015). Martínez-Camblor et al. (2017) described a two-stage residual inclusion approach for the Cox proportional hazards model, where they use an estimation procedure based on a frailty model. Under certain assumptions, which imply that the frailty distribution is known, they show that the proposed estimator is consistent. In practice, however, the frailty distribution will not be known. MacKenzie et al. (2016) propose a method based on principal stratification, but that implies a different estimand than the one pursued in this paper. Martinussen et al. (2017a) proposed an estimation procedure for a proportional hazards estimator among the treated that show resemblence with the approaches of Vansteelandt et al. (2011) and Vansteelandt and Goetghebeur (2003). In these papers, the authors introduce working models for the observed data to build estimation equations for the causal parameter. Unfortunately, these association models might be incompatible with the structural model (Robins and Rotnitzky 2004), which is also referred to as incongeniality (Vansteelandt et al. 2011). This means that there is no data generating mechanism such that both the association model and struc-

tural model hold at the same time. Motivated by this problem, we propose in this paper a novel estimator for a causal proportional hazards that relies on a congenial parametrization. A further advantage of this approach is that it does not assume any specific form of the exposure and/or instrumental variables, and that covariates can be included in the model. Direct modelling of the unobserved confounders is not feasible in non-collapsable models such as the Cox model. Instead one can describe the unobserved confounding using a selection bias function. The selection bias function relate the (possibly counterfactual) unexposed survival times for individuals observed to have different characteristics i.e. different levels of exposure, instrument, covariates and survival time. For binary outcome, the selection bias function has been considered by Tchetgen Tchetgen and Vansteelandt (2013) where they relax the NEM assumption and instead restrict the selection bias function in order to achieve theoretical identification of two causal parameters instead of one. They reduce the selection bias to one parameter and use this extra degree of freedom to augment the causal part with a second parameter. In this paper we present a theoretical framework for estimating a causal proportional hazards for censored survival data. The proportional hazards estimator expresses the treatment effect among the treated conditional on the instrument. The structural Cox model is given in terms of a selection bias function and a counterfactual treatment free survival model so that model congeniality is ensured. It turns out that the observed hazard function has similarities with the Cox–Aalen model, see Martinussen and Scheike (2006) for a presentation of this model. We exploit this similarity to develop an estimation procedure, and to explore large sample properties of the proposed estimator. The outline of the paper is as follows. In Sect. 2 the structural Cox model is presented, and in Sect. 2.1 the parametrization using the selection bias function is introduced, and in Sect. 2.2 the link to the conditional exposure model is presented. In Sects. 2.3, 2.4 and 2.5 we outline the estimation procedure under different assumptions concerning the selection bias and structural model. Large sample properties are presented in Sect. 3 with details given in an "Appendix". The estimator is implemented for simulated data in Sect. 4, and applied to real data concerning a potential causal relationship between Body Mass Index (BMI) and Ischaemic Heart Disease (IHD) in Sect. 5. Finally, a discussion follows in Sect. 6.

2 Causal model

Let X and T be the exposure and survival outcome of interest, and let U be some unmeasured variable confounding this relationship. Using a survival regression model for T to naively model the effect of X and possible observed covariates C will be prone to result in a biased estimate of the causal effect of X. Therefore, if we look for a valid instrument G as illustrated in the Directed Acyclic Graph (DAG) in Fig. 1, we can hope to recover the causal effect of the exposure. An instrument is a variable which is (a) associated with the exposure, (b) has no direct effect on the outcome other than through the exposure, and (c) whose association with the outcome is not confounded by unmeasured variables, see e.g. Hernan and Robins (2006). Condition (a) is empirically verifiable, but conditions (b) and (c) are not. Let T^x be the potential survival time had the exposure been set to x tacitly assuming that the intervention

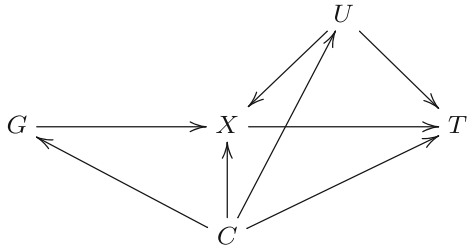


Fig. 1 DAG showing the instrument G, exposure X, survival time T, covariates C and the unobserved confounder U

is meaningful. We assume consistency so that $T^x = T$ for an individual with $X = x$. Both X and G are assumed to be one-dimensional variables, while the observed covariate C may be k-dimensional. For a subject with exposure X, instrument G and covariates C, the survival function is denoted $S_T(t|X, G, C)$, and the hazard function is denoted by $\lambda_T(t|X, G, C)$. The cumulative hazard function is $\Lambda_T(t|X, G, C) = \int_0^t \lambda_T(s|X, G, C)ds$. Likewise, we define the survival function, hazard function and cumulative hazard function for the counterfactual survival time T^x and denote these functions with the subscript T^x instead of T. We define the conditional structural Cox model of interest as

$$\lambda_{T^x}(t|X = x, G, C) = \lambda_{T^0}(t|X = x, G, C)e^{\psi x}, \tag{1}$$

where ψ quantifies the causal effect of the exposure X. Conditional on G and C, the model states that the hazard function of the counterfactual survival times are proportional among the treated individuals within the observed treatment strata. In this sense, the causal effect is an effect of treatment among the treated. In model (1) there is no effect modification by G, but this may be allowed for using the more general model

$$\lambda_{T^x}(t|X = x, G, C) = \lambda_{T^0}(t|X = x, G, C)e^{\gamma(t,x,G,C)},$$

where γ is a known function. The simplest model with effect modification is $\gamma(t, x, G, C) = \psi_1 xG + \psi_0 x(1 - G)$ if the instrument G is binary. Let the observed data be the i.i.d. replicates $(\tilde{T}_i, \delta_i, X_i, G_i, C_i)$ for $i = 1, \ldots, n$ individuals, where $\delta_i = 1$ if the event of interest is observed and then $\tilde{T}_i = T_i$, and $\delta_i = 0$ if the waiting time is independently right-censored with $\tilde{T}_i$ then being the censoring time, U_i. We assume that T_i and U_i are conditional independent given X_i, G_i and C_i. Define the counting process $N_i(t) = I(\tilde{T}_i \leq t, \delta_i = 1)$ that keeps track of whether an event has been observed at a given time. Also define the at risk indicator $R_i(t) = I(\tilde{T}_i \geq t)$. Let the data be observed in the finite time interval $[0, \tau]$ where $\tau < \infty$.

2.1 Selection bias

Martinussen et al. (2017a) proposed an estimation procedure for model (1) by specifying an association model, i.e. a model for the observed data. Unfortunately the association model and the structural model (1) may be incongenial meaning that there is no data generating mechanism so that both models are correctly specified. We

therefore seek an alternative procedure ensuring model congeniality. In this vein we introduce the selection bias function

$$q(t, X, G, C) = -\log\left\{\frac{S_{T^0}(t|X, G, C)}{S_{T^0}(t|X = 0, G, C)}\right\}$$

with $q(t, 0, G, C) = 0$. The selection function q can take any real value; a negative value expresses that $S_{T^0}(t|X, G, C)$ is larger than the reference survival had the individuals in this strata been unexposed. This means that due to some selection mechanism, the exposed individuals have a better survival than the unexposed, had we been able to remove the exposure. For all values of x in the domain of the exposure variable we have $T^x \perp\!\!\!\perp G \mid C$, specifically

$$T^0 \perp\!\!\!\perp G \mid C, \tag{2}$$

see for instance Richardson and Robins (2013). This means that if we intervene by setting the exposure to some fixed level x then the instrument and the outcome are independent given C. By the IV property (2), we can write the counterfactual survival as

$$\begin{aligned} S_{T^0}(t|X = x, G, C) &= \frac{S_{T^0}(t|X = x, G, C)}{S_{T^0}(t|X = 0, G, C)}\left\{\frac{S_{T^0}(t|G, C)}{S_{T^0}(t|X = 0, G, C)}\right\}^{-1} S_{T^0}(t|C) \\ &= \exp\left[-q(t, x, G, C) - \log E\left\{e^{-q(t,X,G,C)}|G, C\right\} - \Omega(t, C)\right], \end{aligned}$$

where

$$\Omega(t, C) = -\log\{S_{T^0}(t|C)\},$$

which we assume to be absolutely continuous, $\Omega(t, C) = \int_0^t \omega(s, C)ds$. This leads to the cumulative baseline hazard function

$$\Lambda_{T^0}(t|X = x, G, C) = q(t, x, G, C) + \log E\left\{e^{-q(t,X,G,C)}|G, C\right\} + \Omega(t, C),$$

and the baseline hazard function in the conditional structural Cox model (1) can therefore be expressed as

$$\lambda_{T^0}(t|X = x, G, C) = q'(t, x, G, C) - \frac{E\left\{e^{-q(t,X,G,C)}q'(t, X, G, C)|G, C\right\}}{E\left\{e^{-q(t,X,G,C)}|G, C\right\}} + \omega(t, C), \tag{3}$$

where q' denotes the derivative of q with respect to its first argument. Note that the form of $\lambda_{T^0}(t|X = x, G, C)$ was obtained without making any assumptions concerning the observed data law, besides that $P(T^0 > t|G, C) = P(T^0 > t|C)$. To proceed, we

need to restrict q and Ω in order to estimate the parameter of interest ψ, but this can be done without such models leading to model incongeniality. Homogeneous selection bias results if we assume

$$q(t, X, G, C) = B_X(t)X, \tag{4}$$

where B_X is absolute continuous with derivative b_X. Heterogeneous selection bias is allowed for if instead

$$q(t, X, G, C) = B_0(t)X(1 - G) + B_1(t)XG \tag{5}$$

for binary G. We further assume an additive hazards model for T^0 given the covariates C,

$$\omega(t, C) = \omega_0(t) + \omega_C(t)C, \tag{6}$$

where ω_C is a $k \times 1$ matrix, and $\Omega_0(t) = \int_0^t \omega_0(s)ds$ and $\Omega_C(t) = \int_0^t \omega_C(s)ds$. Note that for some survival models for T^0 given X, G and C one can motivate a homogeneous selection bias even though this remains untestable. Despite the model choices for the selection bias function and the baseline model for T^0 are hard to justify from the data at hand, they offer an appealing alternative to the association model approach. A specific association model also restricts these functions but implicitly, and in a way that might lead to model incongeniality. The other way around, specific models for q and Ω determine the association model but in a way that might not correspond to a standard choice. In any case, this parametrization will lead to a model that respect the instrumental variable assumption, which is the back bone in the estimation of ψ, rather than a standard association model.

2.2 Exposure model

The baseline hazard function (3) in the structural Cox model depends on the conditional distribution of the exposure X given instrument G and covariates C through the ratio of conditional means

$$\frac{E\left\{e^{-q(t,X,G,C)}q'(t, X, G, C)|G, C\right\}}{E\left\{e^{-q(t,X,G,C)}|G, C\right\}}. \tag{7}$$

To make progress we need a model for X given the instrument G and the covariates C. This model is termed the exposure model. Under homogeneous selection as in (4) the expression (7) reduces to

$$b_X(t)f\{G, C, B_X(t)\},$$

where

$$f(G, C, b) = \frac{E\left\{Xe^{-bX}|G, C\right\}}{E\left\{e^{-bX}|G, C\right\}}. \tag{8}$$

Under heterogeneous selection as in (5) with G being binary, the ratio in (7) simplifies to

$$\{b_0(t)(1-G) + b_1(t)G\} f\{G, C, B_0(t)(1-G) + B_1(t)G\}, \tag{9}$$

where $b_j(t)$ denotes the derivative of $B_j(t)$, $j = 0, 1$. We can further simplify (9) to

$$b_0(t)(1-G) f\{G, C, B_0(t)\} + b_1(t)G f\{G, C, B_1(t)\}$$

using that G is binary. If the exposure X is continuous and its conditional distribution given G and C is normal with mean $\mu(G, C)$ and variance σ^2 then

$$f(G, C, b) = \mu(G, C) - b\sigma^2.$$

Note that we may have a different representation of the exposure variable in the structural model (1) and in the exposure model. We may for instance look for a transformation h_e so that the distribution of $h_e(X)$ given G and C is normal. Without loss of generality we just denote this transformed variable by X. We may use another specification of X in the structural model using another transformation h from $\mathbb{R}$ into some space $\mathbb{E}$. For instance, given some cutpoints $c_1 < c_2 < \ldots < c_r$, $h(x) = \{I(c_1 < x \leq c_2), \ldots, I(c_{r-1} < x \leq c_r), I(x > c_r)\}$ so that h transforms X into a categorized version of X defined by the cutpoints (c_j) using $X < c_1$ as the reference group. The parameter of interest, ψ, is then r-dimensional. We give an illustration of this in Sect. 5 but, in the rest of the paper, we will stick to simple notation where ψ is one-dimensional. If the exposure variable X is binary then

$$f(G, C, b) = \frac{e^{-b}\pi_G(C)}{e^{-b}\pi_G(C) + 1 - \pi_G(C)}$$

with $\pi_g(C) = P(X = 1|G = g, C)$. If this probability fits in a logistic regression for example, then the conditional mean ratio can be calculated using the estimated regression parameters. In general we assume that the conditional distribution of X given G and C is specified via an s-dimensional parameter θ as described above in two special cases.We let θ_0 denote the true parameter value. This exposure model can be estimated from the observed data to get $\hat{\theta}$ by standard methods such as the MLE, and we assume that the obtained estimator $\hat{\theta}$ is regular and asymptotically linear (RAL) (Tsiatis 2006), i.e.

$$n^{1/2}(\hat{\theta} - \theta_0) = n^{-1/2} \sum_{i=1}^{n} \epsilon_i^{\theta},$$

where $(\epsilon_i^\theta)_{i=1,\ldots,n}$ are zero-mean i.i.d. variables. To stress the dependency on θ, we write (8) as $f(G, C, \theta, b)$.

2.3 Homogeneous selection bias and NEM

Under models (4) and (6) the conditional hazard of T^0 is

$$\lambda_{T^0}(t|X = x, G, C) = \omega_0(t) + b_X(t)\,[x - f\{G, C, \theta, B_X(t)\}] + \omega_C(t)C.$$

Under NEM, the parameter of interest ψ is one-dimensional and it follows from the structural Cox model (1) that the observed hazard function

$$\lambda_T(t|X = x, G, C) = e^{\psi x}\Big[\omega_0(t) + b_X(t)\,[x - f\{G, C, \theta, B_X(t)\}] + \omega_C(t)C\Big].$$

is on a Cox–Aalen form, see Ch. 7 in Martinussen and Scheike (2006). Hence,

$$dM_i(t) = dN_i(t) - e^{\psi X_i} R_i(t)\,[X_i - f\{G_i, C_i, \theta, B_X(t-)\}, C_i, 1]\,dA(t)$$

is the increment of a zero-mean martingale, where $A(t) = \{B_X(t), \Omega_C(t), \Omega_0(t)\}^T$. This is the key to estimate the parameter of interest ψ. Let $N = \{N_1(t), \ldots, N_n(t)\}^T$ and $M = \{M_1(t), \ldots, M_n(t)\}^T$ be the n-dimensional counting process and martingale created by stacking the processes for all observations in the data. Let $\phi = (\psi, \theta)$, and write

$$dM(t) = dN(t) - Y\{t, \phi, B_X(t-)\}dA(t)$$

with Y being the $n \times (k + 2)$ matrix with ith row

$$Y_i\{t, \phi, B_X(t-)\} = e^{\psi X_i} R_i(t)\,[X_i - f\{G_i, C_i, \theta, B_X(t-)\}, C_i, 1]\,.$$

Note that we have ensured the predictability of Y seen as a process in time by letting Y, at time t, depend on B_X just before time t. The design Y further depends on the conditional distribution of X given G and C and the parameter of interest ψ. Based on martingale theory, we propose the following estimating equations for ψ and dA

$$\int_0^\tau X^T\,[dN(t) - Y\{t, \phi, B_X(t-)\}dA(t)] = 0 \tag{10}$$

$$Y\{t, \phi, B_X(t-)\}^T W(t)\,[dN(t) - Y\{t, \phi, B_X(t-)\}dA(t)] = 0 \tag{11}$$

where X is the vector of exposures, and $W(t)$ is a an $n \times n$ diagonal weight matrix with entries $w_1, \ldots, w_n$. Following Martinussen and Scheike (2006, Chapter 7) we suggest to use the weights $w_i(t) = R_i(t)e^{-\psi X_i}$.

The specific estimation routine is outlined in the following. For fixed ϕ, the jumps of A are estimated using (11), that is we let

$$\hat{A}(t,\phi) = \int_0^t \left[Y\{s,\phi,\hat{B}_X(s-)\}^T W(s) Y\{s,\phi,\hat{B}_X(s-)\} \right]^{-1} Y\{s,\phi,\hat{B}_X(s-)\}^T W(s) dN(s).$$

This suggests a stepwise procedure. Specifically, let $\tau_1, \tau_2, \ldots$ denote the ordered jump times of N. Initially, $Y(\tau_1,\phi,\hat{B}_X(0))$ is calculated using $\hat{B}_X(0) = 0$, and $d\hat{A}(\tau_1,\phi)$ is found by plugging $t = \tau_1$ into

$$\left[Y\{t,\phi,\hat{B}_X(t-)\}^T W(t) Y\{t,\phi,\hat{B}_X(t-)\} \right]^{-1} Y\{t,\phi,\hat{B}_X(t-)\}^T W(t) dN(t). \quad (12)$$

Next, $Y(\tau_2,\phi,\hat{B}_X(\tau_1))$ is calculated from $\hat{A}(\tau_1,\phi) = \{\hat{B}_X(\tau_1), \hat{\Omega}_C(\tau_1), \hat{\Omega}_0(\tau_1)\}$, with $d\hat{A}(\tau_2,\phi)$ given by (12) for $t = \tau_2$. At every step in the estimation of $d\hat{A}(t,\phi)$ we assume that the matrix

$$Y\{t,\phi,\hat{B}_X(t-)\}^T W(t) Y\{t,\phi,\hat{B}_X(t-)\}$$

is invertible. In practice, we define a matrix to keep track of when the inverse exists as done by Martinussen and Scheike (2006, Chapter 5). Finally, the estimator $\hat{\psi}$ denotes the solution satisfying $U(\hat{\psi}) = 0$ where

$$U(\psi) = \int_0^\tau X^T \left[dN(t) - Y\{t,\tilde{\phi},\hat{B}_X(t-)\} d\hat{A}(t,\tilde{\phi}) \right] \quad (13)$$

is defined as in (10) with A and θ replaced by their estimators and $\tilde{\phi} = (\psi,\hat{\theta})$. Large sample properties of $\hat{\psi}$ are developed in Sect. 3.

2.4 Heterogenous selection bias and NEM

When the selection function is assumed heterogenous as in (5) for binary G then the baseline hazard function in model (1) is given by

$$\begin{aligned}\lambda_{T^0}(t|X=x,G,C) = {} & b_1(t)G\left[X - f\{G,C,\theta,B_1(t)\}\right] \\ & + b_0(t)(1-G)\left[X - f\{G,C,\theta,B_0(t)\}\right] + \omega(t,C).\end{aligned}$$

To do the estimation, expand A to $A(t) = \{B_0(t), B_1(t), \Omega_C(t), \Omega_0(t)\}^T$ and let Y be an $n \times (k+3)$ matrix with ith row $R_i(t)e^{\psi X_i}(Y_i^q, C_i, 1)$, where Y^q is the part of Y dealing with the selection bias function q, that is the ith row of Y^q is

$$\Big((1-G_i)\left[X_i - f\{G_i,C_i,\theta,B_0(t-)\}\right], G_i\left[X_i - f\{G_i,C_i,\theta,B_1(t-)\}\right] \Big).$$

Estimation of ψ and the nuisance parameter A can now be done similarly as for the homogeneous case outlined in Sect. 2.3.

2.5 Effect modification by *G*

One may also extend the estimation procedure to allow for effect modification by G, where

$$\gamma(t, X, G, C) = \psi^1 XG + \psi^0 X(1 - G)$$

with causal parameter $\psi = (\psi^0, \psi^1)$. Under homogeneous selection bias (4) the ith row of Y is then

$$Y_i\{t, \phi, B_X(t-)\} = R_i(t)e^{\psi^1 X_i G_i + \psi^0 X_i(1-G_i)}\left[X_i - f\{G_i, C_i, \theta, B_X(t-)\}, C_i, 1\right].$$

Now A can be estimated using (11) for fixed ψ and $W(t)$ an $n \times n$ weight matrix $W(t) = \text{diag}\{R_i(t)e^{-\psi^1 X_i G_i + \psi^0 X_i(1-G_i)}\}$. Finally, $\hat{\psi}$ can be found by solving $U(\psi) = 0$ for

$$U(\psi) = \int_0^\tau Z^T[dN(t) - Y\{t, \tilde{\phi}, \hat{B}_X(t-)\}d\hat{A}(t, \tilde{\phi})]$$

where Z is an $n \times 2$ matrix with ith row

$$\{X_i(1 - G_i), X_i G_i\}.$$

3 Large sample results

In this section we present large sample properties of the estimators, further details are given in the "Appendix". First we have that

$$n^{1/2}\{\hat{A}(t, \phi_0) - A_0(t)\} = n^{-1/2}\sum_{i=1}^n \epsilon_i^A(t) + o_P(1)$$

with $\epsilon_i^A(t)$ are zero-mean i.i.d. variables as defined in the "Appendix". We further show that

$$n^{1/2}(\hat{\psi} - \psi_0) = n^{-1/2}\sum_{i=1}^n \epsilon_i^\psi + o_p(1)$$

where the ϵ_i^ψ's are zero-mean iid variables given in display (16) in the "Appendix". Thus $n^{1/2}(\hat{\psi} - \psi_0)$ converges in distribution to a zero-mean normal variate with a variance that can be estimated consistently using the above iid-representation. Finally, one can show that

$$n^{1/2}\{\hat{A}(t, \hat{\phi}) - A_0(t)\} = n^{-1/2}\sum_{i=1}^n \tilde{\epsilon}_i^A(t) + o_p(1),$$

Table 1 Performance of the estimator $\hat{\psi}$

n	Mean	SD	MSE
2000	0.495	0.234	0.0545
4000	0.505	0.160	0.0255

where the $\tilde{\epsilon}_i^A(t)$'s are zero-mean iid processes given in the "Appendix".

4 Simulations

In this section we present results based on simulated datasets to investigate the performance of the proposed estimator under NEM, but allowing for heterogen selection. The covariate C was uniformly distributed on (0, 1). The instrument and exposure variables were binaries generated from $\text{logit}P(G = 1|C) = 0.3C$ and $\text{logit}P(X = 1|G, C) = G + 0.3C - 0.5$, where $\text{logit}(p) = \log\{p/(1-p)\}$. The selection function was chosen as

$$q(t; X, G, C) = B_0(t)X(1-G) + B_1(t)XG,$$

where $B_0(t) = 0.2(t - 0.5)I(t > 0.5)$ and $B_1(t) = 0.2t^2$, in this way allowing for heterogen selection. From

$$\frac{S_{T^0}(t|X=1,G,C)}{S_{T^0}(t|X=0,G,C)} = \begin{cases} I(t \le 0.5) + \exp\{-0.2(t-0.5)\}I(t>0.5) & \text{if } G=0 \\ \exp(-0.2t^2) & \text{if } G=1 \end{cases}$$

it is seen that, on average, those being exposed and having $G = 1$, had they been unexposed, they would have had worse survival than those being unexposed with $G = 1$. Similarly for those with instrument $G = 0$, but only when $t > 0.5$, while for $t \le 0.5$ there is no difference. Censoring times were exponentially distributed with mean 5 and approximately 30% of the survival times were censored. The conditional probabilities of X were estimated using logistic regression. We generated 2000 datasets based on the specification of the above structural models with sample sizes equal to 2000 and 4000. Table 1 shows the mean, standard error and mean squared error of the estimates. The proposed estimator for ψ performs well with the mean of the estimates being close to the true parameter value 0.5 in both cases. The estimated asymptotic SD, using the true θ_0, was 0.227 and 0.16, respectively, leading to 95% coverage probabilities of 94.8% and 95.4%.

To compare with the aforementioned approach by Martinussen et al. (2017a), we also fitted different Aalen additive hazard models to the observed data, predicted the survival probabilities $\hat{S}_\beta(t^*|X, G, C)$ from these models, and estimated ψ as the solution to

$$\sum_{i=1}^{n}(G_i - \bar{G})\hat{S}_\beta(t^*|X_i, G_i, C_i) = 0$$

Table 2 Estimates of ψ using Aalen association models; a simple model and a model with additional interaction terms

	n	Mean	SD	MSE
Simple, fixed time	2000	0.762	0.320	0.171
Simple, optimal time (mean 2.459)	2000	0.689	0.385	0.183
Interaction, fixed time	2000	0.537	0.284	0.082
Interaction, optimal time (mean 1.634)	2000	0.678	0.290	0.116
Simple, fixed time	4000	0.764	0.224	0.120
Simple, optimal time (mean 2.508)	4000	0.730	0.276	0.129
Interaction, fixed time	4000	0.549	0.196	0.041
Interaction, optimal time (mean 1.601)	4000	0.625	0.178	0.047

The estimation was either carried out at a fixed time $t^* = 2$, or the optimal time was estimated, in which case the mean optimal time is presented in parenthesis

where $\bar{G}$ is the empirical mean of G, and t^* is a fixed time point. Instead of using t^* in the above estimating equation, we can use the time that minimizes the empirical variance for the estimator for ψ as proposed by Martinussen et al. (2017a). This time point is referred to as an optimal time. The specified structural models lead to a non-standard survival model for the observed data. Nevertheless, we tried to fit additive hazards models to the data in order to see how they performed. A simple model with additive terms for for X, G and C, and a more elaborated model where interaction terms between X and G, and X and C were added. The results, shown in Table 2, using the two different Aalen models are referred to as "simple" (without any interaction terms) and "interaction" (model included interaction terms as previously specified).

In all cases it is seen that the association model approach results in biased estimators. The use of a simple association model resulted in a larger bias and mean squared error than the use of a more complex model. It should come as no surprise that the association model approach results in bias as the association model is misspecified while the data generating model satisfies the assumptions of our newly developed method. The estimation of an optimal time to get the smallest asymptotical variance resulted in larger empirical variance and larger bias. This can be explained by the fact that the association models did not fit the data very well, and thus relying on the models to find the smallest variance was a poor approximation of the true performance of the estimator. We also computed the naive measure of association in a standard Cox model using explanatory variables X and C, and found a proportional hazards mean value of 1.22 (SD 0.046) over all datasets of sample size 4000. Additionally, we also computed a two-stage predictor inclusion estimator using estimates from a correctly specified logistic regression model for the first stage, and a Cox model including C and the predicted probabilities in the second stage. The mean of this two-stage estimator was 0.165 (SD 0.155) over all datasets of sample size 4000. Interestingly, these two standard methods showed significantly larger bias than the association model estimators. Finally, we also estimated an two-stage residual inclusion estimator using a correctly specified first stage model and included the residual as an additional covariate in the second stage Cox model. The residual was calculated as the exposure minus the

predicted probability from the first stage model. The mean of this estimator was 0.743 (SD 0.157), which is less biased than the other naive estimators and comparable to the estimates from the association model approach. To conclude, the naive estimators show bias, they are moreover not theoretically founded and conclusions based on these estimators should be avoided.

5 Causal relationship between BMI and IHD

In this section we analyse data from a Mendelian randomization study to investigate whether adiposity causally affects the onset of cardiovascular disease. The prevalence of obesity is increasing worldwide, and is associated with increased risk of cardiovascular diseases, diabetes, musculoskeletal disorders and certain cancers. An individuals susceptibility to overweight due to biologic inheritance can to some extent be summarized by her profile of associated SNPs, even though the genetic variants only account for a minor proportion of the natural variation in body weight. However, the loci of these SNPs are valuable in the understanding of the mechanisms regulating the human body weight. The Copenhagen General Population Study (CGPS) consists of individuals from the city of Copenhagen. The study was initiated in 2003 with continued enrolment of participants reflecting the adult Danish population. At inclusion the individuals underwent a health examination at Herlev Hospital, answered a questionnaire and had blood samples taken. Moreover, microarrays were examined for SNP detection. The participants were followed prospectively in time after the inclusion. In this application we consider subjects that were between age 40 and 55 at the time of examination. In total, this gives 31,684 subjects in the analysis. We included the following baseline covariates in the analysis: age (in years, and then standardized), sex (0 female; 1 male), income dichotomized (1: high income) and alcohol intake with categories < 7, 7–14, 15–21, > 21. Due to a few missing values in the income and alcohol variables, sample size was reduced to 31,383. To examine the causal effect of BMI on the risk of IHD using Mendelian Randomization (MR) we need an instrumental variable. Besides FTO-rs9939609, the SNPs MC4R-rs17782313, TMEM18-rs6548238, BDNF-rs10767664 and GNPDA2-rs10938397 are also potential instruments that correlate with BMI, see Nordestgaard et al. (2012) for more details. For a given SNP, each individual has either 0, 1 or 2 of the risk alleles. Following the tradition of MR analysis we use an allele score based on the SNPs as given by the total sum of risk alleles thus ranging from 0 to 10. This allele score was also used in Nordestgaard et al. (2012). Comparing the two linear models with BMI as response and the baseline covariates as explanatory variables, and with and without the instrument resulted in a F-statistic of 181. The usual rule of thumb says that an F-statistic below 12 indicates a weak instrument. The median follow-up time was around 5 years, and some were followed for almost 9 years. In total, 518 IHD events were observed, the cumulative incidence function is displayed in Fig. 2.

We first used the intention to treat analysis applying a standard Cox model with the considered baseline covariates and the instrument. The hazard ratio associated with the instrument was estimated to 1.05 (95% CI: 0.98; 1.11). From this there is no convincing evidence of a causal effect of BMI on the risk of experiencing an IHD event

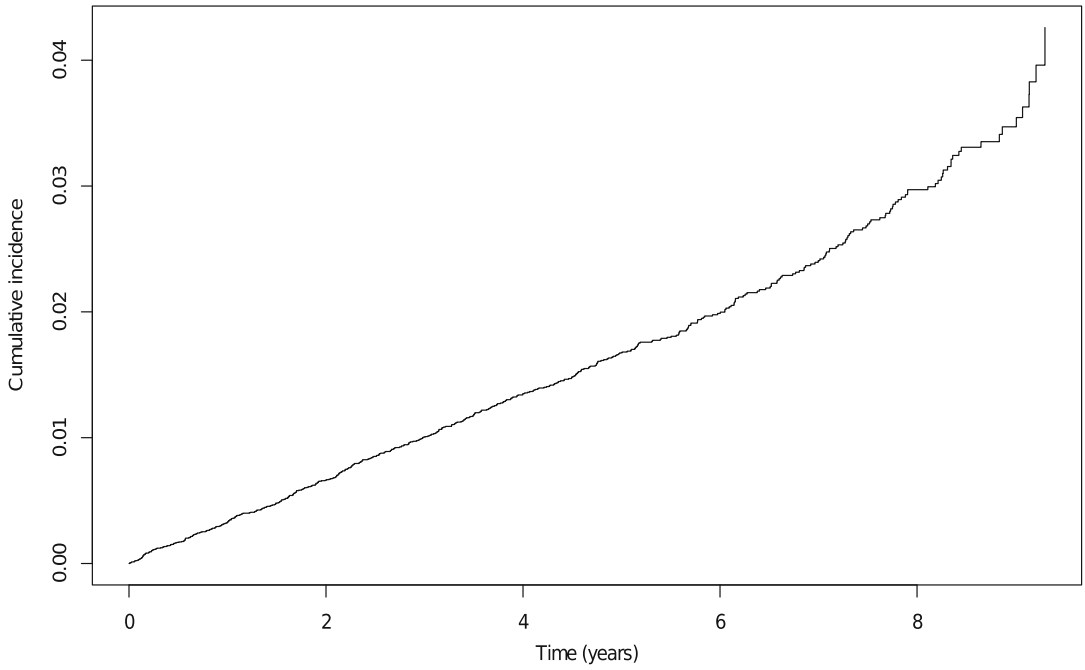


Fig. 2 GCPS-study. Cumulative incidence function

Table 3 CGPS-data. Hazard ratios comparing different BMI groups with 20–25 as reference group

BMI	–20	20–25	25–30	30–
$\exp(\hat{\psi})$	1.08 (0.58; 2.03)	1	1.25 (0.96; 1.63)	1.55 (1.02; 2.33)

Numbers in parentheses are 95% bootstrapped confidence intervals

in the considered age span. We applied next the proposed analysis with a categorized version of BMI in the structural model (1) using cut points 20, 25 and 30 with 20–25 as reference category. In the exposure model we used $X = \sqrt{\text{BMI}} - \sqrt{23}$. We assumed homogeneous selection bias, and used the four mentioned baseline covariates as our C in model (6). To avoid instability in the tail we took $\tau = 7.5$ years. The results are summarised in Table 3.

From this analysis we see again no convincing evidence that there is causal effect of BMI on the risk of experiencing an IHD event in the considered age span. The risk appears, however, to be significantly increased when comparing those with BMI above 30 to the reference group, the causal HR is estimated to 1.55 (1.02; 2.33).

6 Discussion

In this paper we have proposed a novel estimator for a causal proportional hazards. It is a common issue for non-linear models that congeniality issues arise when restrictions on the observed data law and the structural model are not in accordance with the data-generating mechanism. In the presented work, the IV property is built into the model through a parametrization that guarantees congeniality. The method is flexible

in handling homogeneous or heterogeneous selection bias and the NEM assumption can even be relaxed, and covariates can easily be taken into account in the baseline survival model. A drawback of the proposed method is that the conditional distribution of the exposure given instrument and covariates needs to be specified. However this is also a feature of several other IV estimators, for example the naive two-stage IV estimates where the role of the first stage model is to predict the exposure given the instrument. If G and C are categorical variables then the conditional mean ratio can be calculated non-parametrically as done by Chen et al. (2010). To compare with the estimation of a causal proportional hazard using an association model, we found in Sect. 4 that the association model had to include complex interactions terms to correctly specify the hazards and fit the data. However, the association model might be a good approximation in some settings and it is easy to implement. Nevertheless, our proposed reparametrization of the baseline hazard leads to flexible models where the underlying assumptions can be more precisely stated. A drawback is that it is hard to check these assumptions. A possible way forward is to consider the estimating function $U(\psi) = U(\psi, \tau)$, with

$$U(\psi, t) = \int_0^t X^T \left[dN(s) - Y\{s, \tilde{\phi}, \hat{B}_X(s-)\} d\hat{A}(s, \tilde{\phi}) \right],$$

as a process in time. Similar to the derivations in the "Appendix" we may show that

$$n^{-1/2} U(\psi, t) = n^{-1/2} \sum_{i=1}^{n} \delta_i^U(t),$$

where the $\delta_i^U(t)$'s are zero-mean iid processes. This representation allows us to re-sample from the limit distribution (Lin et al. 1993) obtained under the assumption that all the structural models are correctly specified. We may then compare the observed process $n^{-1/2} U(\hat{\psi}, t)$ to these. A deviating observed process is an indication of something being wrong with the model assumptions. We intend to investigate this further in future work. The proposed estimation is enhanced by noting the implied Cox–Aalen type model for the observed data. This allows also that the method may be extended to handle left-truncated data as well. The method we have proposed implies that

$$\begin{aligned} \Lambda_{T^0}(t|x, g, c) = q(t, x, g, c) + \log E \left\{ e^{-q(t, X, G, C)} | G = g, C = c \right\} \\ + \Omega_0(t) + \Omega_C(t) c, \end{aligned}$$

which needs to be an increasing function in t for all (x, g, c), which imposes a non-trivial restriction on the functions $q(t, x, g, c)$, $\omega_0(t)$ and $\omega_C(t)$. The standard Aalen additive hazards model (Aalen 1989) faces a similar problem, but is not found to constitute any serious problem in practice.

IV identification requires in addition to having a valid IV that one makes an additional assumption. In this paper we consider the effect of treatment on the treated (ETT) which can be identified with a valid IV by making one of two possible no interaction assumptions; either by assuming that the magnitude of the ETT does not

vary with the IV, or that the magnitude of the selection bias does not vary with the IV. The former directly restricts the causal effect we are making inferences about which is not ideal, while the latter in principle allows the causal effect to remain unrestricted. Note that under the null of no treatment causal effect in the treated, the first restriction naturally holds and therefore the approach is robust to this assumption under the null which is not true for the second identification strategy. There seems to be no clear simplification or computational advantage of one assumption over the other. The issue of possible effect modification by observed covariates is also of interest. In principle, our approach allows for all such interactions to the extent they can be estimated accurately in finite sample. In practice, particularly when C is high dimensional, the curse of dimensionality will seriously limit our ability to incorporate several two-way or higher interactions, we therefore focus primarily on parsimonious parametric models to incorporate covariates. An important topic for future research is to develop semiparametric doubly robust methods similar to Tchetgen Tchetgen and Vansteelandt (2013) to allow some but not necessarily all models for the effects of C to remain unrestricted.

Acknowledgements We are grateful to Shoaib Afzal and Børge Nordestgaard for giving us access to the CGPS-data.

Appendix: Large sample properties

Large sample properties of the estimator of A

The consistency of $\hat{A}(t, \phi_0)$ may be shown similar to what is done in Martinussen et al. (2017b). We will here focus on the asymptotic distribution of $\hat{A}(t, \phi_0)$. To this end, define the $p \times n$ matrix H as

$$H(s, \phi, b) = \{Y(s, \phi, b)^T W(s) Y(s, \phi, b)\}^{-1} Y(s, \phi, b)^T W(s)$$

for $p = k + 2$ and k the dimension of C. Then we can write

$$\hat{A}(t, \phi_0) = \int_0^t H\{s, \phi_0, \hat{B}_X(s-)\} dN(s).$$

Notice that we can write the true nuisance parameter A_0 as

$$\begin{aligned} A_0(t) &= \int_0^t H\{s, \phi_0, B_X^0(s)\} Y\{s, \phi_0, B_X^0(s)\} dA_0(s) \\ &= \int_0^t H\{s, \phi_0, B_X^0(s)\} dN(s) - \int_0^t H\{s, \phi_0, B_X^0(s)\} dM(s). \end{aligned}$$

Now $n^{1/2}\{\hat{A}(t,\phi_0) - A_0(t)\}$ can be expressed as

$$n^{1/2}\int_0^t H\{s,\phi_0,B_X^0(s)\}dM(s) + n^{1/2}\int_0^t [H\{s,\phi_0,\hat{B}_X(s-)\} - H\{s,\phi_0,B_X^0(s-)\}]dN(s).$$

First we take a closer look at the last integral in this expression. By a Taylor approximation we have

$$n^{1/2}\int_0^t [H\{s,\phi_0,\hat{B}_X(s-)\} - H\{s,\phi_0,B_X^0(s-)\}]dN(s) = n^{1/2}\int_0^t V(ds)\{\hat{A}(s-,\phi_0) - A_0(s-)\} + o_P(1)$$

where V is the $p \times p$ matrix

$$V(s) = \frac{\partial H\{s,\phi_0,B_X^0(s-)\}N(s)}{\partial A_0(s)^T}.$$

As H only depends on the first element of A_0 namely B_X^0 then the first column of V is non-zero and the rest consist of zeros. Define $Z(t,\phi_0) = n^{1/2}\{\hat{A}(t,\phi_0) - A_0(t)\}^T$ and from the above calculations we see that it satisfies the following Volterra equation (Andersen et al. 1993, p. 91)

$$Z(t,\phi_0) = n^{1/2}\int_0^t M(ds)^T H\{s,\phi_0,B_X^0(s-)\}^T + \int_0^t Z(s-,\phi_0)V(ds)^T,$$

and the solution is

$$n^{1/2}\int_0^t M(ds)^T H\{s,\phi_0,B_X^0(s-)\}^T \mathcal{Q}(s,t)$$

where $\mathcal{Q}$ is the product integral

$$\mathcal{Q}(s,t) = \prod_{u\in(s,t]} \{I + V(du)^T\}$$

as defined in Andersen et al. (1993). Finally, we have that

$$n^{1/2}\{\hat{A}(t,\phi_0) - A_0(t)\} = n^{1/2}\int_0^t \mathcal{Q}(s,t)^T H\{s,\phi_0,B_X^0(s-)\}dM(s).$$

In this expression $[n^{-1}Y\{s,\phi_0,B_X^0(s-)\}^T W(s)Y\{s,\phi_0,B_X^0(s-)\}]^{-1}$ converges in probability to some $p \times p$ matrix that we denote $l_1(s)$. Also $\mathcal{Q}(s,t)^T$ converges to some limit in probability that is denoted $l(s,t)$. Further, one can show the convergence

in distribution of $n^{-1/2}Y\{s, \phi_0, B_X^0(s-)\}^T W(s)dM(s)$ to a mean zero process. Then we have the i.i.d. representation of $\hat{A}(t, \phi_0)$

$$n^{1/2}\{\hat{A}(t, \phi_0) - A_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\epsilon_i^A(t) + o_P(1)$$

with

$$\epsilon_i^A(t) = \int_0^t l(s, t)l_1(s)Y_i\{s, \phi_0, B_X^0(s-)\}^T W_i(s)dM_i(s),$$

where the elements of ϵ_i^A are denoted $(\epsilon_i^{B_X}, \epsilon_i^{\Omega_C}, \epsilon_i^{\Omega_0})^T$. This representation ensures convergence of the finite dimensional distribution. Convergence in distribution as a process can be shown similarly to what is done in Martinussen et al. (2017b).

Large sample properties of $\hat{\psi}$

We first note that

$$\begin{aligned} n^{-1/2}U(\hat{\psi}) = n^{-1/2}U(\hat{\psi}, \hat{\theta}) = n^{-1/2}U(\phi_0) + \{n^{-1}D_\psi U\}n^{-1/2}(\hat{\psi} \\ - \psi_0) + \{n^{-1}D_\theta U\}n^{-1/2}(\hat{\theta} - \theta_0) + o_p(1) \end{aligned}$$

so

$$n^{-1/2}(\hat{\psi} - \psi_0) = -\{n^{-1}D_\psi U\}^{-1}[n^{-1/2}U(\phi_0) + \{n^{-1}D_\theta U\}n^{-1/2}(\hat{\theta} - \theta_0)] + o_p(1)$$

and since we have assumed $\hat{\theta}$ to be RAL we just need to find the asymptotic distribution of $n^{-1/2}U(\phi_0)$. We can write this function as

$$\begin{aligned} n^{-1/2}U(\phi_0) = n^{-1/2}\int_0^\tau X^T[dN(t) - Y\{t, \phi_0, B_X^0(t)\}dA_0(t)] \\ - n^{-1/2}\int_0^\tau X^T Y\{t, \phi_0, \hat{B}_X(t-)\}\{d\hat{A}(t, \phi_0) - dA_0(t)\} \quad (14) \\ - n^{-1/2}\int_0^\tau X^T[Y\{t, \phi_0, \hat{B}_X(t-)\} - Y\{t, \phi_0, B_X^0(t)\}]dA_0(t). \quad (15) \end{aligned}$$

The first term on the right hand side of the latter display is the martingale

$$n^{-1/2}\int_0^\tau X^T dM(t) = n^{-1/2}\sum_{i=1}^{n}\int_0^\tau X_i dM_i(t).$$

Note that $Y\{t, \phi_0, \hat{B}_X(t-)\}$ and $Y\{t, \phi_0, B_X^0(t)\}$ share all entries except for the first column. If we let Y_{*1} denote the first column of Y then (15) can be written as

$$-n^{-1/2}\int_0^\tau X^T[Y_{*1}\{t, \phi_0, \hat{B}_X(t-)\} - Y_{*1}\{t, \phi_0, B_X^0(t)\}]dB_X^0(t).$$

Using a Taylor expansion this is asymptotically equivalent to

$$-n^{-1/2}\int_0^\tau \frac{\partial X^T Y_{*1}\{t, \phi_0, B_X^0(t)\}}{\partial B_X^0(t)}\{\hat{B}_X(t-) - B_{X0}(t)\}dB_X^0(t).$$

Let $d_{B_X}(t)$ denote the limit in probability of the derivative $n^{-1}\frac{\partial X^T Y_{*1}\{t,\phi_0,B_X^0(t)\}}{\partial B_X^0(t)}$. By the i.i.d. representation of $n^{1/2}\{\hat{B}_X(t-) - B_X^0(t)\}$ derived earlier in this Appendix, (15) is seen to be asymptotically equivalent to the following sum of zero-mean i.i.d. terms

$$-n^{-1/2}\sum_{i=1}^n \int_0^\tau d_{B_X}(t)\epsilon_i^{B_X}(t-)dB_X^0(t).$$

For the latter of these integrals, (14), we have convergence in distribution of its integrand $n^{1/2}\{\hat{A}(t, \phi_0) - A_0(t)\}$ and the integral can be written as (Kosorok 2008, Lemma 4.2)

$$-n^{-1/2}\int_0^\tau X^T Y\{t, \phi_0, B_X^0(t)\}\{d\hat{A}(t, \phi_0) - dA_0(t)\}$$

since $n^{-1}[X^T Y\{t, \phi_0, \hat{B}_X(t-)\} - X^T Y\{t, \phi_0, B_X^0(t)\}]$ converges in probability to 0. Denote the limit in probability of $n^{-1}X^T Y\{t, \phi_0, B_X^0(t)\}$ as $l_{XY}(t)$. Thus it is asymptotically equivalent to

$$-n^{-1/2}\sum_{i=1}^n \int_0^\tau l_{XY}(t)d\epsilon_i^A(t)$$

Finally, we have that $n^{-1/2}U(\psi_0) = n^{-1/2}\sum_{i=1}^n \epsilon_i^U + o_p(1)$, where

$$\epsilon_i^U = \int_0^\tau X_i dM_i(t) - \int_0^\tau d_{B_X}(t)\epsilon_i^{B_X}(t)dB_X^0(t) - \int_0^\tau l_{XY}(t)d\epsilon_i^A(t).$$

Finally,

$$n^{1/2}(\hat{\psi} - \psi_0) = n^{-1/2}\sum_{i=1}^n \epsilon_i^\psi + o_p(1)$$

where

$$\epsilon_i^{\psi} = -\mathcal{J}_{\psi}^{-1}\{\epsilon_i^{U} + \mathcal{J}_{\theta}\epsilon_i^{\theta}\}, \tag{16}$$

with $\mathcal{J}_{\psi}$ denotes the limit in probability of $n^{-1}D_{\psi}U$ and similarly with $\mathcal{J}_{\theta}$. Based on the above derivations, the i.i.d. decomposition of $n^{1/2}\{\hat{A}(t,\hat{\phi}) - A_0(t)\}$ can easily be obtained since

$$\begin{aligned} n^{1/2}\{\hat{A}(t,\hat{\phi}) - A_0(t)\} &= n^{1/2}\{\hat{A}(t,\hat{\phi}) - \hat{A}(t,\phi_0)\} + n^{1/2}\{\hat{A}(t,\phi_0) - A_0(t)\} \\ &= n^{1/2}\{D_{\psi}\hat{A}(t,\phi_0)\}\{\hat{\psi} - \psi_0\} + n^{1/2}\{D_{\theta}\hat{A}(t,\phi_0)\}\{\hat{\theta} - \theta_0\} \\ &\quad + n^{1/2}\{\hat{A}(t,\psi_0) - A_0(t)\} \end{aligned}$$

where $D_{\psi}\hat{A}(t,\psi_0)$ converges in probability to some limit as $n \to \infty$. Hence also,

$$n^{1/2}\{\hat{A}(t,\hat{\phi}) - A_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\tilde{\epsilon}_i^{A}(t) + o_p(1),$$

where

$$\tilde{\epsilon}_i^{A}(t) = \epsilon_i^{A}(t) + \mathcal{A}_{\psi}\epsilon_i^{\psi} + \mathcal{A}_{\theta}\epsilon_i^{\theta},$$

with $\mathcal{A}_{\psi}$ denoting the limit in probability of $D_{\psi}\hat{A}(t,\phi_0)$, and similarly with $\mathcal{A}_{\theta}$.

References

Aalen OO (1989) A linear regression model for the analysis of life times. Stat Med 8:907–925

Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer, New York

Angrist JD, Imbens GW (1994) Identification and estimation of local average treatment effects. Econometrica 62:467–475

Chen L, Lin DY, Zeng D (2010) Attributable fraction functions for censored event times. Biometrika 97:713–726

Clarke PS, Windmeijer F (2010) Identification of causal effects on binary outcomes using structural mean models. Biostatistics 11:756–70

Clarke PS, Windmeijer F (2012) Instrumental variable estimators for binary outcomes. J Am Stat Assoc 107:1638–1652

Cuzick J, Sasieni P, Myles J, Tyrer J (2007) Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. J R Stat Soc Ser B 69:565–588

Hernan MA, Robins JM (2006) Instruments for causal inference. Epidemiology 17:360–372

Kosorok MR (2008) Introduction to empirical processes and semiparametric inference. Springer, New York

Lin DY, Wei LJ, Ying Z (1993) Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika 80:557–572

Loeys T, Goetghebeur E (2003) A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. Biometrics 59:100–105

MacKenzie TA, Tosteson TD, Morden NE, Stukel TA, O'Malley AJ (2014) Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. Health Serv Outcomes Res Methodol 14:54–68

MacKenzie TA, Løberg M, O'Malley AJ (2016) Patient centered hazard ratio estimation using principal stratification weights: application to the NORCCAP randomized trial of colorectal cancer screening. Obs Stud 2:29–50

Martínez-Camblor P, Mackenzie T, Staiger DO, Goodney PP, O'Malley AJ (2017) Adjusting for bias introduced by instrumental variable estimation in the Cox proportional hazards model. Biostatistics 20:80–96

Martinussen T, Scheike TH (2006) Dynamic regression models for survival data, vol 102. Springer, New York

Martinussen T, Sørensen D, Vansteelandt S (2017a) Instrumental variables estimation under a structural Cox model. Biostatistics 20:65–79

Martinussen T, Vansteelandt S, Tchetgen Tchetgen EJ, Zucker DM (2017b) Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. Biometrics 73(4):1140–1149

Nordestgaard BG, Palmer TM, Benn M, Zacho J, Tybjaerg-Hansen A, Davey Smith G, Timpson NJ (2012) The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a mendelian randomisation approach. PLoS Med 9(5):e1001212

Richardson TS, Robins, JM (2013) Single World Intervention Graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Technical Report 128, Center for Statistics and the Social Sciences, University of Washington

Robins J, Rotnitzky A (2004) Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. Biometrika 91:763–783

Tchetgen Tchetgen EJ, Vansteelandt S (2013) Alternative identification and inference for the effect of treatment on the treated with an instrumental variable. Harvard University Biostatistics Working Paper Series

Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, Martinussen T, Glymour M (2015) Instrumental variable estimation in a survival context. Epidemiology 26:402–10

Tsiatis A (2006) Semiparametric theory and missing data. Springer, New York

Vansteelandt S, Goetghebeur E (2003) Causal inference with generalized structural mean models. J R Stat Soc Ser B 65:817–835

Vansteelandt S, Bowden J, Babanezhad M, Goetghebeur E (2011) On instrumental variables estimation of causal odds ratios. Stat Sci 26:403–422