

Instrumental Variables Estimation of Exposure Effects on a Time-to-Event Endpoint Using Structural Cumulative Survival Models

Torben Martinussen^{1,*}, Stijn Vansteelandt^{2,**}, Eric J. Tchetgen Tchetgen^{3,***} and David M. Zucker^{4,****}

¹Department of Biostatistics, University of Copenhagen Øster Farimagsgade 5B, 1014 Copenhagen K, Denmark

²Department of Applied Mathematics and Computer Sciences, Ghent University, Krijgslaan 281 S9, 9000 Ghent, Belgium

³Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Massachusetts 02115 Boston, U.S.A.

⁴Department of Statistics, Hebrew University Mount Scopus, Jerusalem 91905, Israel

*email: tma@sund.ku.dk

**email: stijn.vansteelandt@ugent.be

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

****email: david.zucker@mail.huji.ac.il

SUMMARY. The use of instrumental variables for estimating the effect of an exposure on an outcome is popular in econometrics, and increasingly so in epidemiology. This increasing popularity may be attributed to the natural occurrence of instrumental variables in observational studies that incorporate elements of randomization, either by design or by nature (e.g., random inheritance of genes). Instrumental variables estimation of exposure effects is well established for continuous outcomes and to some extent for binary outcomes. It is, however, largely lacking for time-to-event outcomes because of complications due to censoring and survivorship bias. In this article, we make a novel proposal under a class of structural cumulative survival models which parameterize time-varying effects of a point exposure directly on the scale of the survival function; these models are essentially equivalent with a semi-parametric variant of the instrumental variables additive hazards model. We propose a class of recursive instrumental variable estimators for these exposure effects, and derive their large sample properties along with inferential tools. We examine the performance of the proposed method in simulation studies and illustrate it in a Mendelian randomization study to evaluate the effect of diabetes on mortality using data from the Health and Retirement Study. We further use the proposed method to investigate potential benefit from breast cancer screening on subsequent breast cancer mortality based on the HIP-study.

KEY WORDS: Causal effect; Confounding; Current treatment interaction; G-estimation; Instrumental variable; Mendelian randomization.

1. Introduction

A key concern in most analyzes of observational studies is whether sufficient and appropriate adjustment has been made for confounding of the association between the exposure of interest and outcome. This concern can be mitigated to some extent when data are available on an instrumental variable. This 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., Hernán and Robins, 2006). Condition (a) is empirically verifiable, but conditions (b) and (c) are not. However, condition (c) can sometimes be justified in observational studies that incorporate elements of randomization, either by design or by nature (Didelez and Sheehan, 2007). The plausibility of condition (b) can sometimes be argued on the basis of design elements (e.g., blinding) or a priori contextual knowledge.

Instrumental variables have a long tradition in econometrics (e.g., Angrist and Krueger, 2001). They are increasingly popular in epidemiology due to a revival of Mendelian randomization studies (Katan, 1986; Davey-Smith and Ebrahim, 2003). Such studies focus on modifiable exposures known to be affected by certain genetic variants. They then adopt the notion that an association between these genetic variants and the outcome of interest (e.g., all-cause mortality) can only be explained by an effect of the exposure on the outcome. This reasoning presupposes that the genetic variants studied satisfy the aforementioned instrumental variable conditions (Didelez and Sheehan, 2007). That is, they should have no effect on the outcome (e.g., all-cause mortality) other than by modifying the exposure, which can sometimes be justified based on a biological understanding of the functional genetic mechanism. Moreover, their association with the outcome should be unconfounded, which is sometimes realistic

because of Mendelian randomization: the fact that genes are transferred randomly from parents to their offspring.

Instrumental variables estimation of exposure effects is well established for continuous outcomes that obey linear models. Two-stage least squares (2SLS) estimation proceeds via two ordinary least squares regressions: regressing the exposure variable on the instrument in the first stage, and next regressing the outcome variable on the predicted exposure value in the second stage. This approach presumes that the additive exposure effect is the same at all levels of the unmeasured confounders (Hernán and Robins, 2006), which is rarely plausible in the analysis of event times. The IV-analysis of event times is further complicated because of censoring and the fact that the instrumental variables assumptions, even when valid for the initial study population, are typically violated within the risk sets composed of subjects who survive up to a given time. Progress is often made via heuristic adaptations of 2SLS estimation, whereby the second stage regression is substituted by a Cox regression (see Tchetgen Tchetgen et al. (2015), and Rassen et al. (2008), Cai, Small, and Ten Have (2011) for related approaches for dichotomous outcomes), but these have no formal justification outside the limited context of rare events (Tchetgen Tchetgen et al., 2015).

To the best of our knowledge, the first formal IV-approach for the analysis of event times was described in Robins and Tsiatis (1991), who parameterized the exposure effect under a structural accelerated failure time model and developed G-estimation methods for it. Their development is very general, and, in particular, can handle continuous exposures. However, recurring problems in applications have been the difficulty in finding solutions to the estimating equations and obtaining estimators with good precision. This is related to the use of an artificial censoring procedure, where some subjects with observed event times are censored in the analysis in order to maintain unbiased estimating equations. This procedure may lead to an enormous information loss. Moreover, it leads to non-smooth estimating equations (Joffe, Yang, and Feldman, 2012), so that even simple models are difficult to fit. Loeys, Goetghebeur, and Vandebosch (2005) proposed an alternative approach based on structural proportional hazards models. Their development does not require the use of recensoring, but is more parametric than that of Robins and Tsiatis (1991) as it requires modeling the exposure distribution. It is moreover limited to settings with a binary instrument and constant exposure at one level of the instrument, which is characteristic of placebo-controlled randomized experiments without contamination. Cuzick et al. (2007) relax this limitation by adopting a principal stratification approach but, like other such approaches (see e.g., Abadie, 2003; Nie, Cheng, and Small, 2011), restrict their development to binary exposure and instrumental variables. More recently, Tchetgen Tchetgen et al. (2015) independently demonstrated the validity of two-stage estimation approaches in additive hazard models for event times when the exposure obeys a particular location shift model (see Li, Fine, and Brookhart (2015) for a related approach under a more restrictive model; other related approaches are discussed in Tchetgen Tchetgen et al., 2015). In this article, we avoid restrictions on the exposure distribution and develop IV-estimators under a semi-parametric structural cumulative survival model that

is closely related to, but less restrictive than, the additive hazard model in Tchetgen Tchetgen et al. (2015) and Li, Fine, and Brookhart (2014). The proposed approach is general in that it can handle arbitrary exposures and instrumental variables, and can accommodate adjustment for baseline covariates. It neither requires modeling the exposure distribution nor the association between covariates and outcome, and it naturally deals with administrative censoring and certain forms of dependent censoring. Picciotto et al. (2012) studied the different problem of adjusting for time-varying confounding when estimating the effect of a time-varying exposure on a survival outcome. While we also make use of the structural cumulative failure time model, we do this for handling the different problem of estimating the effect of an exposure on a survival outcome in the presence of unobserved confounding using an instrumental variable. Because of this and the fact that we make use of semi-parametric continuous-time models, in contrast to Picciotto et al. (2012) who focus on parametric discrete-time models, the recursive estimators that we propose cannot be immediately compared with those in Picciotto et al. (2012). A further strength of our article is that it develops asymptotic inference for the proposed recursive estimators, which is currently lacking for G-estimators in structural cumulative failure time models. The semi-parametric estimator that we propose requires only a correct model for the conditional mean of the instrumental variable, given covariates, for consistency. Besides deriving its large sample properties we also develop inferential tools allowing us for instance to investigate for time-changing exposure effect. We examine the performance of the proposed method in simulation studies and two empirical studies.

2. Model Specification and Estimation

2.1. Basics

Our goal is to estimate the effect of an arbitrary exposure X on an event time \tilde{T} under the assumption that G is an instrumental variable, conditional on a covariate set L . A data-generating mechanism that satisfies this assumption is depicted in the causal diagram (Pearl, 2009) of Figure 1. Here, the instrumental variables assumptions are guaranteed by the absence of a direct effect of G on \tilde{T} , and by the absence of effects of the unmeasured confounder U on G , and of G on U .

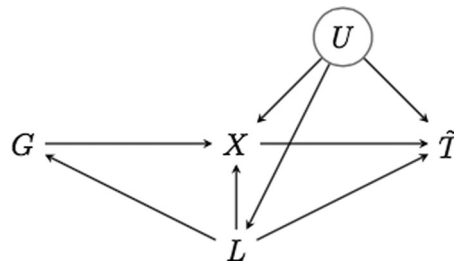


Figure 1. Causal Directed Acyclic Graph. G is the instrument, X the exposure variable, and \tilde{T} the time-to-event outcome. The potential unmeasured confounders are denoted by U , and the observed confounders of the G - \tilde{T} association by L .

To provide insight, we will start by considering uncensored survival data under the following semi-parametric variant of the additive hazards model (Aalen, 1980):

$$E \left\{ d\tilde{N}(t) | \mathcal{F}_t^{\tilde{N}}, G, X, L, U \right\} = \left\{ d\Omega(t, L, U) + dB_X(t)X \right\} \tilde{R}(t), \quad (1)$$

where $\tilde{N}(t) = I(\tilde{T} \leq t)$ denotes the counting process, $\mathcal{F}_t^{\tilde{N}}$ the history spanned by $\tilde{N}(t)$, $\tilde{R}(t) = I(t \leq \tilde{T})$ is the at risk indicator, $\Omega(t, L, U)$ is an unknown, non-negative function of time, L and U , and $B_X(t)$ is an unknown scalar at each time $t > 0$. Note, that the righthand side of this model does not involve G because of the instrumental variables assumptions, which imply that \tilde{T} and G are conditionally independent, given X, L , and U . Note, furthermore that we explicitly choose to leave $\Omega(t, L, U)$ unspecified because U is unmeasured, thus, making assumptions about the hazard's dependence on U rather delicate.

Under model (1),

$$\exp \left\{ -B_X(t)x \right\} = \frac{P(\tilde{T} > t | X = x, G, L, U)}{P(\tilde{T} > t | X = 0, G, L, U)}, \quad (2)$$

which captures the exposure effect of interest by virtue of conditioning on the unmeasured confounder U . By the collapsibility of the relative risk (or the related collapsibility of the hazard difference (Martinussen and Vansteelandt, 2013)), this is also equal to the directly standardized relative survival risk:

$$\exp \left\{ -B_X(t)x \right\} = \frac{E \left\{ P(\tilde{T} > t | X = x, G, L, U) | G, L \right\}}{E \left\{ P(\tilde{T} > t | X = 0, G, L, U) | G, L \right\}},$$

where the averaging is over the conditional distribution of U given (G, L) . Letting \tilde{T}^x , for each fixed x , denote the potential outcome that would have been observed if the exposure were set to x by some intervention, this can also be written as

$$\exp \left\{ -B_X(t)x \right\} = \frac{P(\tilde{T}^x > t | G, L)}{P(\tilde{T}^0 > t | G, L)}. \quad (3)$$

This can be seen because, by virtue of U being sufficient to adjust for confounding of the effect of X on \tilde{T} , we have that \tilde{T}^x is conditionally independent of X , given U, G, L .

That the effect $\exp \left\{ -B_X(t)x \right\}$ can also be defined without making reference to the unmeasured confounder U (i.e., without conditioning on U) is important. Indeed, the lack of data on U as well as the lack of a precise understanding of the variables contained inside U , would otherwise make interpretation difficult (Vansteelandt et al., 2011).

Model (1) is closely related to the structural cumulative survival model:

$$\exp \left\{ -B_X(t)x \right\} = \frac{P(\tilde{T} > t | X = x, G, L)}{P(\tilde{T}^0 > t | X = x, G, L)}. \quad (4)$$

This model is slightly less restrictive than model (1). It makes no assumptions as to how the unmeasured confounders are associated with the event time. It moreover models the

effect of setting the exposure to zero, within exposure subgroups rather than the entire population. By evaluating effects within exposure subgroups, the parameter $B_X(t)$ in model (4) thus encodes a type of treatment effect in the treated. Under the additional assumption that there is no current treatment interaction (Hernán and Robins, 2006), a population-averaged interpretation can be made. In particular, suppose that within levels of G and L , the effect of exposure level x versus 0 on the survival function is the same for subjects with observed exposure $X = x$ as for subjects with a different exposure level in the following sense

$$\frac{P(\tilde{T}^x > t | X = x, G, L)}{P(\tilde{T}^0 > t | X = x, G, L)} = \frac{P(\tilde{T}^x > t | X \neq x, G, L)}{P(\tilde{T}^0 > t | X \neq x, G, L)}. \quad (5)$$

Then model (4) along with the assumption of no current treatment interaction implies (3), so that $B_X(t)$ captures a population-averaged effect (as is also the case under model (1)).

Under the instrumental variables assumptions that X and G are associated, conditional on L , and that \tilde{T}^0 is conditionally independent of G , given L , the estimators of $B_X(t)$ that we will propose in the next section will be consistent estimators of $B_X(t)$ in both models (1) and (4). Condition (5) is not required for the estimation methods that we develop later on; it is only needed to provide the population-level interpretation given in (3).

2.2. Estimation

We will allow for the event time \tilde{T} to be subject to right-censoring. In that case, we only observe whether or not \tilde{T} exceeds a random censoring time C , that is, we observe $\delta = I(\tilde{T} \leq C)$, along with the first time either failure or censoring occurs, that is, we also observe $T = \min(\tilde{T}, C)$. It is assumed that \tilde{T} and C are independent given L, G, X, U , and that $P(C > t | X, G, L, U) = P(C > t | L, U)$. The censoring distribution may thus depend on unobserved characteristics, but we do not have to model such dependencies, in fact no modeling of the censoring distribution is needed under this condition. In the Web Appendix, we give an alternative condition concerning the censoring that also leads to the same estimating equation, see display (6). Let $(T_i, \delta_i, L_i, G_i, X_i)$, $i = 1, \dots, n$, denote n independent identically distributed replicates under the structural cumulative failure time model (1) together with the instrumental variables assumptions, and the independent censoring assumption. The counting processes $N_i(t) = I(T_i \leq t, \delta_i = 1)$, $i = 1, \dots, n$, are observed in the time interval $[0, \tau]$, where τ is some finite time point. Further, we define the at risk indicator $R_i(t) = I(t \leq T_i)$, $i = 1, \dots, n$.

The crux of our estimation method for $B_X(t), t > 0$ is that once the exposure effect has been eliminated from the event time, it only retains a dependence on L and U . It thus becomes conditionally independent of the instrumental variable, given L . In particular, using arguments similar to those of Martinussen et al. (2011), we eliminate the exposure effect from the increments $dN(t)$ by calculating $dN(t) - dB_X(t)X\tilde{R}(t)$, as suggested by (1), and we will eliminate the exposure effect from the at risk indicators $R(t)$ by calculating $R(t) \exp \left\{ B_X(t)X \right\}$, as suggested by (2). It follows

that

$$E \left[\{G - E(G|L)\} e^{B_X(t)X} R(t) \{dN(t) - dB_X(t)X\} \right] = 0, \quad (6)$$

for each $t > 0$. To formally show this we note that

$$E\{dN(t)|\mathcal{F}_{t-}^{L,G,X,U}\} = \tilde{R}(t)\{d\Omega(t, L, U) + dB_X(t)X\} \\ + E\{dM(t)|\mathcal{F}_{t-}^{L,G,X,U}\},$$

where $\mathcal{F}_{t-}^{L,G,X,U}$ is the filtration generated by the counting process and L, G, X, U , and $M(t)$ is the corresponding counting process martingale. Thus

$$E \left[\{G - E(G|L)\} e^{B_X(t)X} R(t) \{dN(t) - dB_X(t)X\} \right] \\ = E \left[\{G - E(G|L)\} e^{B_X(t)X} I(C > t) \tilde{R}(t) d\Omega(t, L, U) \right] \\ = E \left[P(C > t|L) \{G - E(G|L)\} e^{-\Omega(t,L,U)} d\Omega(t, L, U) \right] = 0;$$

here, the last equality follows because $G \perp\!\!\!\perp U|L$. In (6), the $G - E(G|L)$ can be replaced by $H(G, L) - E\{H(G, L)|L\}$ for a general H , but for simplicity we keep (6) as it is.

The unbiasedness of equation (6) suggests a way of estimating the increments $dB_X(t)$ by solving equation (6) for each t with population expectations replaced by sample analogs. This delivers the recursive estimator $\hat{B}_X(t)$ defined by

$$\hat{B}_X(t, \hat{\theta}) = \int_0^t \frac{\sum_i G_i^c(\hat{\theta}) e^{\hat{B}_X(s-)X_i} dN_i(s)}{\sum_i G_i^c(\hat{\theta}) R_i(s) e^{\hat{B}_X(s-)X_i} X_i}, \quad (7)$$

where $G_i^c(\theta) = G_i - E(G_i|L_i; \theta)$, with $E(G_i|L_i; \theta)$ a parametric model for $E(G_i|L_i)$ and $\hat{\theta}$ a consistent estimator of θ (e.g., a maximum likelihood estimator).

The estimator (7) is given by a counting process integral, thus only changing values at observed death times. Because of its recursive structure, we calculate it forward in time, starting from $\hat{B}_X(0) = 0$. In the special case where the exposure X is binary, it can be calculated analytically as shown below. With $A(t) = e^{B_X(t)}$, the equation (6) (with population expectations substituted by sample averages) leads to

$$dA(t) = \frac{\sum_i G_i^c(\theta)(1 - X_i)dN_i(t)}{\sum_i R_i(t)G_i^c(\theta)X_i} + A(t) \frac{\sum_i G_i^c(\theta)X_i dN_i(t)}{\sum_i R_i(t)G_i^c(\theta)X_i}$$

When replacing $A(t)$ with $A(t-)$ on the right side of this expression and integrating, we get the Volterra equation (see Andersen et al. (1993), p. 91)

$$A(t) = W(t) + \int_0^t A(s-)dU(s),$$

where

$$W(t) = \int_0^t \frac{\sum_i G_i^c(\theta)(1 - X_i)dN_i(s)}{\sum_i R_i(s)G_i^c(\theta)X_i}, \\ U(t) = \int_0^t \frac{\sum_i G_i^c(\theta)X_i dN_i(s)}{\sum_i R_i(s)G_i^c(\theta)X_i},$$

and the solution is given by

$$A(t) = W(t) + \int_0^t \left[\prod_{(s,t]} \{1 + dU(v)\} \right] dW(s)$$

If we assume that $B_X(t) = \int_0^t \beta_X(s) ds$, and further that $\beta_X(t) = \beta_X$, that is assuming a time-constant effect, an estimator of β_X may be obtained as

$$\hat{\beta}_X = \int_0^\tau w(t) d\hat{B}_X(t) \quad (8)$$

with $w(t) = \tilde{w}(t) / \int_0^\tau \tilde{w}(s) ds$, $\tilde{w}(t) = R \cdot(t) = \sum_i R_i(t)$.

3. Large Sample Results

The following proposition, whose proof is given in the Appendix, shows that $\hat{B}_X(t)$ is a uniformly consistent estimator of $B_X(t)$, and moreover gives its asymptotic distribution.

PROPOSITION 1. *Under model (4) with the assumption that G is an instrumental variable, conditional on L , and given the technical conditions listed in the Appendix, the IV estimator $\hat{B}_X(t)$ is a uniformly consistent estimator of $B_X(t)$. Furthermore, $W_n(t) = n^{1/2}\{\hat{B}_X(t, \hat{\theta}) - B_X(t)\}$ converges in distribution to a zero-mean Gaussian process with variance $\Sigma(t)$. A uniformly consistent estimator $\hat{\Sigma}(t)$ of $\Sigma(t)$ is given below.*

Let $\epsilon_i^B(t, \theta)$, $i = 1, \dots, n$ be the iid zero-mean processes given by expression (13) in the Appendix. From the proof in the Appendix, it then follows that $W_n(t)$ is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^n \epsilon_i^B(t, \theta)$. The variance $\Sigma(t)$ of the limit distribution can thus be consistently estimated by

$$\hat{\Sigma}(t) = n^{-1} \sum_{i=1}^n \{\hat{\epsilon}_i^B(t, \hat{\theta})\}^2,$$

where $\hat{\epsilon}_i^B(t, \hat{\theta})$ is obtained from $\epsilon_i^B(t, \theta)$ by replacing unknown quantities with their empirical counterparts. These results can be used to construct a pointwise confidence band. The asymptotic behavior of the estimator (8) is easily obtained since:

$$n^{1/2}(\hat{\beta}_X - \beta_X) = n^{-1/2} \sum_{i=1}^n \int_0^\tau w(t) d\epsilon_i^B(t, \theta).$$

To study temporal changes, it is more useful to consider a uniform confidence band. This and tests of the hypothesis of a linear (cumulative) causal effect

$$H_0 : \beta_X(t) = \beta_X \Leftrightarrow H_0 : B_X(t) = \beta_X t$$

can easily be derived based on the above iid representation as also outlined in Martinussen (2010). The above hypothesis can be tested using the following test statistic

$$\sup_{t \leq \tau} |n^{1/2}(\hat{B}_X(t) - \hat{\beta}_X t)|, \quad (9)$$

and since, under the null, $n^{1/2}(\hat{B}_X(t) - \hat{\beta}_X t) = n^{1/2}\{\hat{B}_X(t) - B_X(t) - (\hat{\beta}_X - \beta_X)t\}$ it is easy to get the iid-representation of the test process. This development is based on the following approach of Lin et al. (1993). Let Q_1^m, \dots, Q_n^m be independent standard normal variates. Then, given the data,

$$\hat{W}_m(t) = n^{-1/2} \sum_{i=1}^n \hat{\epsilon}_i(t, \hat{\theta}) Q_i^m$$

also converges in distribution to a zero-mean Gaussian process with variance $\Sigma(t)$. The limit distribution can thus be evaluated by generating a large number, M , of replicates $\hat{W}_m(t)$, $m = 1, \dots, M$. The causal null hypothesis that $B_X(t) = 0$ for all t can thus for example be tested by investigating how extreme the test statistic $\sup_{t \leq \tau} |n^{1/2} \hat{B}_X(t)|$, is in the distribution of $\sup_{t \leq \tau} |\hat{W}_m(t)|$, $m = 1, \dots, M$.

4. Numerical Results

4.1. Simulation Study

To investigate the properties of our proposed methods with practical sample sizes, we conducted a simulation study. We generated data according to the data-generating mechanism of Figure 1 with the following specific models where we leave out the covariate L for simplicity. We considered two different settings where the exposure variable was continuous and binary, respectively. In the first setting, we took G to be binary with $P(G = 1) = 0.5$, and generated X and U , given G , from a normal distribution with $E(X|G = g) = 0.5 + \gamma_G g$, $E(U|G = g) = 1.5$ and with variance-covariance matrix so that $\text{Var}(X|G) = \text{Var}(U|G) = 0.25$, and $\text{Cov}(X, U|G) = -1/6$. The parameter γ_G determines the size of the correlation between exposure and the instrumental variable. Specifically, we looked at correlation ρ equal to 0.3 and 0.5. We generated \tilde{T} according to the hazard model $E\{d\tilde{N}(t)|T \geq t, X, G, U\} = \beta_0(t) + \beta_X(t)X + \beta_U(t)U$, with $\beta_0(t) = 0.25$, $\beta_X(t) = 0.1$ and $\beta_U(t) = 0.15$. Twenty percent were potentially censored

according to a uniform distribution on $(0, 3.5)$, and the rest were censored at $t = 3.5$, corresponding to the study being closed at this time point, leading to a cumulative censoring rate of around 20%. Under this model, as seen in the Section 2.2, (4) holds with $B_X(t) = \int_0^t \beta_X(s) ds = 0.1t$. Under this model it further holds that $E\{d\tilde{N}(t)|T \geq t, X, G\} = \tilde{\beta}_0(t) + \tilde{\beta}_G(t)G + \tilde{\beta}_X(t)X$, with $\tilde{\beta}_X(t) = \beta_X(t) - 2\beta_U(t)/3 = 0$, so the naive Aalen estimator (using X and G as covariates) is biased. In this specific setting $\theta = E(G)$, and we therefore calculated the estimator given in (7) with $\hat{\theta} = \bar{G}$, along with the estimator $\hat{\beta}_X$ given in (8) where we took $\tau = 3$. For this scenario, we considered sample sizes 1600 and 3200 when $\rho = 0.3$, and sample sizes 800 and 1600 when $\rho = 0.5$. Simulation results concerning $\hat{B}_X(t)$, based on 2000 runs for each configuration, are given in Table 1, where (average) biases are reported at time points $t = 1, 2, 3$ for $\hat{B}_X(t)$ along with coverage probability of 95% pointwise confidence intervals $\text{CP}(\hat{B}_X(t))$. Results concerning β_X are given in Web Table 1, first half.

In all scenarios considered the naive Aalen estimator is, as expected, biased; see Table 1. From Table 1, it is also seen that the proposed estimator $\hat{B}_X(t)$ is unbiased. In the case with sample size 1600 and correlation equal to 0.3 the estimated standard error at time point $t = 3$ is a bit too large resulting in a too high coverage probability. However, it is also seen that the estimated standard error approaches the empirical standard deviation as sample size goes up, and overall the 95%-coverage probabilities have the correct size. We also calculated the size of the sup-test (9) that investigates whether the constant exposure effects model is acceptable. For the four considered scenarios of (n, ρ) : (1600, 0.3), (3200, 0.3), (800, 0.5), (1600, 0.5), it was 0.03, 0.04, 0.03, and 0.05, respectively. Hence, when sample size and correlation goes up, the test has the correct size. The results concerning the constant effect estimator, $\hat{\beta}_X$, are reported in the first half of Web Table 1, and from there it is seen that the estimator is unbiased and that the variability is well estimated leading to satisfactory coverage probabilities at least when sample size goes up.

Table 1

Continuous exposure case. Time-constant exposure effect. Bias of $\hat{B}_X(t)$, average estimated standard error, $sd(\hat{B}_X(t))$, empirical standard error, $see(\hat{B}_X(t))$, and coverage probability of 95% pointwise confidence intervals $\text{CP}(\hat{B}_X(t))$ based on the instrumental variables estimator, in function of sample size n and at different strengths ρ (correlation) of the instrumental variable. Bias of $\tilde{B}_X(t)$ is the bias of the naive Aalen estimator.

	n	$\rho = 0.3$			n	$\rho = 0.5$		
		$t = 1$	$t = 2$	$t = 3$		$t = 1$	$t = 2$	$t = 3$
Bias $\hat{B}_X(t)$	1600	-0.003	-0.001	-0.007	800	-0.002	-0.004	-0.015
sd $(\hat{B}_X(t))$		0.139	0.242	0.404		0.109	0.187	0.303
see $(\hat{B}_X(t))$		0.139	0.245	0.439		0.107	0.187	0.314
95% CP $(\hat{B}_X(t))$		95.4	96.5	98.1		95.2	96.1	97.5
Bias $\tilde{B}_X(t)$		-0.101	-0.201	-0.300		-0.099	-0.197	-0.297
Bias $\hat{B}_X(t)$	3200	-0.003	-0.005	-0.014	1600	0.004	0.004	-0.002
sd $(\hat{B}_X(t))$		0.094	0.170	0.267		0.075	0.131	0.209
see $(\hat{B}_X(t))$		0.096	0.166	0.262		0.075	0.130	0.206
95% CP $(\hat{B}_X(t))$		95.6	95.1	96.2		95.0	95.5	95.7
Bias $\tilde{B}_X(t)$		-0.099	-0.200	-0.296		-0.099	-0.200	-0.301

When the exposure is continuous one may also calculate the 2SLS estimator $\hat{\beta}_X$ of Tchetgen et al. (2015); see Web Table 1. From there it is seen that this estimator is also unbiased, and that it is slightly more efficient than the constant effects estimator given in this article. This is not surprising as the 2SLS estimator is targeted at this specific situation while the estimator $\hat{\beta}_X$ is derived from an estimator that can handle much more general situations. We also considered a setup where there was a time-varying exposure effect. Data was generated as described above except that $\beta_X(t)$ was now taken as $\beta_X(t) = 0.1I(t < 1.5) - 0.1I(1.5 \leq t < 3)$. Inducing censoring as above resulted in a cumulative censoring rate of around 25. Results from this study are given in Web Table 2, where we have dropped results for the naive Aalen estimator. From Web Table 2, we see again that the proposed estimator is unbiased and that the variability is well estimated resulting in appropriate coverage. We also calculated the size of the sup-test (9). For the four considered scenarios of (n, ρ) : (1600, 0.3), (3200, 0.3), (800, 0.5), (1600, 0.5), it was 0.07, 0.18, 0.13, and 0.31, respectively. We also ran the situation where $n = 3200$ and $\rho = 0.5$, and obtained the size of the test to be 0.61. We thus see, as expected, that when correlation and sample size goes up the power of the test increases. We also calculated the constant effects estimators $\hat{\beta}_X$ and $\tilde{\beta}_X$, and the mean of them in all four combinations of (n, ρ) was 0.04 thus showing that the constant effects estimators are not appropriate under this scenario.

We next considered settings where the exposure variable X was binary. In the first such setting, we generated data as under the first scenario with $\beta_X(t) = 0.1$, but instead of using the continuous version of X , call it now \tilde{X} , we used $X = I(\tilde{X} > 0.5)$.

We used the same censoring mechanism and also the same hazards model as under the first setting. For this scenario, we considered sample sizes 3200 and 6400 when $\rho = 0.3$, and sample sizes 1600 and 3200 when $\rho = 0.5$. Results, again based on 2000 runs for each configuration, are shown in Table 2. For the case $(n = 3200, \rho = 0.3)$ the coverage probability is a bit too high at $t = 3$. In the other settings, the estimator is

unbiased and coverage is satisfactory. The results concerning the constant effect estimator, $\hat{\beta}_X$, are reported in the second half of Web Table 1, and from there it is seen that the estimator is unbiased and that the variability is well estimated leading to satisfactory coverage probabilities. We also see that 2SLS estimator of Tchetgen et al. (2015) seems to be unbiased in this setting although there is no theoretical underpinning of this. To look further into this and to stress that the 2SLS estimation relies on a correct specification of a model for the exposure X given the instrument G we ran a final study as follows. The instrument G was taken to be normally distributed with mean 2 and variance 1.5^2 . The unobserved U was taken to be $1.5Z^2$ with Z generated as normal with mean 1 and variance 0.25^2 . The exposure X was binary with $P(X = 1|G, U) = \text{expit}\{-1 + 0.2G + 0.5G^2 + U - E(U)\}$. In this way, the correlation between X and G was approximately 0.56. We generated \tilde{T} according to the hazard model $E\{d\tilde{N}(t)|T \geq t, X, G, U\} = 0.05 + 0.4X + 0.3U$, and censored all at $t = 2$ resulting in approximately 25% censorings. We used sample size 1000 and 2000 with 1000 runs for each configuration. We calculated the 2SLS estimator in two ways using different first stage models; we denote the 2SLS estimator based on regressing X on G (despite that X is binary) in the first stage by $\hat{\beta}_{1X}$ and the 2SLS estimator based on a first stage logistic regression model using G as explanatory variable by $\hat{\beta}_{2X}$. We stress that the estimator suggested in this article, $\hat{\beta}_X$, is not based on any modeling of X given G in contrast to the 2SLS estimator. Results are given in Web Table 3 where it is seen that the estimator $\hat{\beta}_X$ is unbiased while the two versions of the 2SLS estimator are both biased.

4.2. Application to the HRS on Causal Association between Diabetes and Mortality

We illustrate the proposed method using data from the Health and Retirement Study (HRS), a cohort initiated in 1992. The same data was used by Tchetgen Tchetgen et al. (2015) (TT) to investigate the causal association between diabetes and mortality. The HRS consists of persons aged 50 years or older and their spouses. There are genotype data for 12,123

Table 2

Binary exposure case. Bias of $\hat{B}_X(t)$, average estimated standard error, $sd(\hat{B}_X(t))$, empirical standard error, $see(\hat{B}_X(t))$, and coverage probability of 95% pointwise confidence intervals $CP(\hat{B}_X(t))$ based on the instrumental variables estimator, in function of sample size n and at different strengths ρ (correlation) of the instrumental variable. Bias of $\tilde{B}_X(t)$ is the bias of the naive Aalen estimator.

	n	$\rho = 0.3$			n	$\rho = 0.5$		
		$t = 1$	$t = 2$	$t = 3$		$t = 1$	$t = 2$	$t = 3$
Bias $\hat{B}_X(t)$	3200	0.000	0.001	-0.017	1600	-0.000	-0.005	-0.022
$sd(\hat{B}_X(t))$		0.109	0.194	0.316		0.102	0.183	0.306
$see(\hat{B}_X(t))$		0.109	0.194	0.331		0.102	0.183	0.302
95% $CP(\hat{B}_X(t))$		95.3	95.4	96.6		95.7	95.6	96.1
Bias $\tilde{B}_X(t)$		-0.082	-0.164	-0.248		-0.085	-0.167	-0.249
Bias $\hat{B}_X(t)$	6400	-0.000	-0.006	-0.015	3200	0.001	0.001	-0.005
$sd(\hat{B}_X(t))$		0.077	0.137	0.221		0.071	0.128	0.202
$see(\hat{B}_X(t))$		0.077	0.135	0.216		0.072	0.128	0.207
95% $CP(\hat{B}_X(t))$		95.1	94.6	95.2		95.1	95.2	95.9
Bias $\tilde{B}_X(t)$		-0.082	-0.167	-0.250		-0.083	-0.168	-0.253

participants, but, like TT, we restrict our analyzes to the 8446 non-Hispanic white persons with valid self-reported diabetes status at baseline. The average follow-up time was 4.10 years with a total of 644 deaths over 34,055 person-years. We used an externally validated genetic risk score predictor of type 2 diabetes as IV. The risk score is based on 39 SNPs that were strongly associated with the diabetes status (Likelihood Ratio Test chi-square statistic equal to 176.75 with 39 degrees of freedom, $P < 10^{-6}$). Like TT we used as observed confounders (L) age, sex, and the top four genomewide principal components to account for possible population stratification. The 2SLS control function approach used in TT is only valid if the instrument is binary unless one makes a further linearity assumption concerning a conditional mean of the un-observed confounder(s), specifically they assume that $E\{\Omega(t, U)|G, X\}$ is linear in G . This assumption is untestable. The method we propose in this article is not restricted to binary instruments; it places no restrictions on the exposure, nor the instrument. The proposed approach furthermore allows the exposure effect to vary with time, unlike the proposal by TT. The analysis used here thus generalizes that of TT in several aspects. Figure 2 shows the estimated causal effect of diabetes status on mortality, $\hat{B}_X(t)$, along with 95% pointwise confidence bands. The straight line corresponds to the constant effects estimator (8). From Figure 2, it seems reasonable to assume a time-constant exposure effect, which we can formally test using the statistic (9). It gives a P-value of 0.61 thus giving no evidence against the time-constant exposure effect model. The estimate of the time-constant exposure effect is $\hat{\beta}_X = 0.036$ with estimated standard error 0.0142 corresponding to the 95% confidence interval (0.008, 0.064). It suggests a causal effect of diabetes on all cause mortality corresponding to an average of 3.6 additional deaths occurring for each year of follow-up in each 100 persons with diabetes alive at the start of the year, compared with each 100 diabetes-free persons alive at the start of the year, conditional on age and sex. This estimated effect is less than half of that obtained by TT suggesting that the linearity assumption used in TT may not hold.

4.3. Application to the HIP Trial on Effectiveness of Screening on Breast Cancer Mortality

The Health Insurance Plan (HIP) of Greater New York was a randomized trial of breast cancer screening that began in

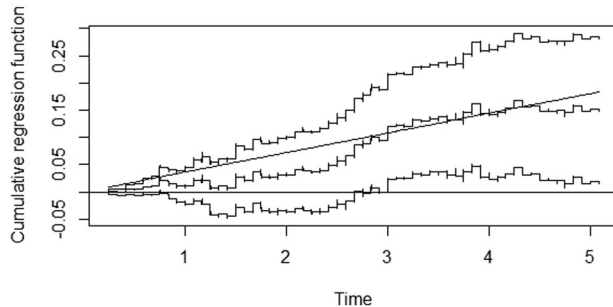


Figure 2. HRS-study. Estimated causal effect of diabetes, $\hat{B}_X(t)$, along with 95% pointwise confidence bands. The straight line corresponds to the constant effects estimator (8).

1963. The purpose was to see whether screening has any effect on breast cancer mortality. About 60,000 women aged 40–60 were randomized into two approximately equally sized groups. Study women were offered the screening examinations consisting of clinical examination, usually by a surgeon, and a mammography. Further, three annual examinations were offered in this group. Control women continued to receive their usual medical care. About 35% of the women that were offered screening refused to participate (non-compliers); see Web Table 4. There were large differences between the study women who participated and those who refused (Shapiro, 1977) and therefore the results from the “as treated” analysis may be doubtful due to unobserved confounding.

The same data were analyzed by Joffe (2001) and as he did, we will also focus on the first 10 years of follow-up. Since screening ended after three years, Joffe argued that focussing on the first 10 years of follow-up will reduce attenuation of the effects of screening in the later periods in which treatment was the same both groups. We will later look into the possibility of a time-varying effect in a more formal way as our estimator $\hat{B}_X(t)$ captures this directly.

A Cox-regression intention-to-treat analysis showed that there is reduced mortality from breast cancer in the screening group (P 0.01). Figure 3 shows the estimated cumulative regression coefficient from a corresponding Aalen additive hazards intention-to-treat analysis (broken curve). The supremum test shows evidence of an overall effect of screening (P 0.005).

We will now apply our suggested method to estimate the causal effect of screening on breast cancer mortality, noting that there is an additional complication due to competing risks: in the first 10 years of follow-up there are 4221 deaths but only 340 were deemed due to breast cancer. We will use the randomization variable G as instrument and denote screening by X . The i th counting process in our estimator

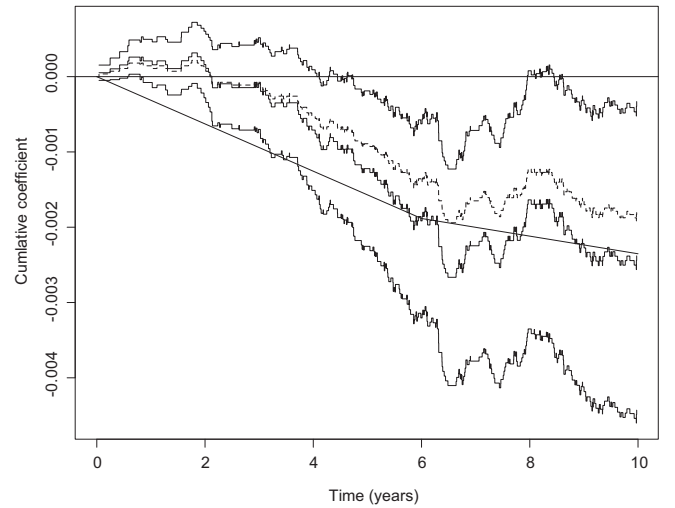


Figure 3. HIP-study. Estimated causal effect of screening, $\hat{B}_X(t)$, along with 95% pointwise confidence bands (solid curves) and the intention to treat estimate (broken curve). A two-parameter piecewise constant estimator $B_X^\dagger(t)$, see (11), is also shown.

(7) is now the counting process that jumps at time point t if the i th women at that point in time dies from breast cancer. We show in a separate report to be communicated elsewhere that $\hat{B}_X(t)$ contrasts the cumulative breast cancer death specific hazards among the treated between scenarios with versus without screening under the assumption that the cause specific hazard of death due to other causes than breast cancer for the screened women would have been the same at all times had they not been screened. To test this assumption one may use the test process

$$H_n(t) = n^{-1/2} \sum_i \int_0^t (G_i - \bar{G}) e^{\hat{B}_X(s^-) X_i} dN_{2i}(s),$$

where $N_{2i}(t)$ is the i th counting process counting non-breast cancer death. Under the null of no causal effect of screening on the non-breast cancer death hazards, this process is a zero-mean process. One may further show that $H_n(t) = n^{-1/2} \sum_i \epsilon_i^H(t) + o_p(1)$, where $\epsilon_i^H(t)$ are independent identically distributed zero-mean processes. Specifically,

$$\begin{aligned} \epsilon_i^H(t) = \{G_i - E(G_i)\} & \left\{ \int_0^t e^{B_X(s) X_i} dN_{2i}(s) - \zeta_1(t) \right\} \\ & + \epsilon_i^B(t, \theta) \zeta_2(t) - \int_0^t \zeta_2(s) d\epsilon_i^B(s, \theta) \end{aligned}$$

considering here the case without covariates so that $\theta = E(G_i)$. In the previous display, $\zeta_1(t)$ and $\zeta_2(t)$ are the limits in probability of

$$\begin{aligned} n^{-1} \sum_i \int_0^t e^{B_X(s) X_i} dN_{2i}(s) \quad \text{and} \\ n^{-1} \sum_i \int_0^t \{G_i - E(G_i)\} e^{B_X(s) X_i} X_i dN_{2i}(s), \end{aligned}$$

respectively. This representation can be used to resample from the limit distribution of $H_n(t)$ under the null. Further, a formal test based on for instance $\sup_{t \leq 10} |H_n(t)|$ may be performed and whether or not it is significant can also be based on resampling from the limit distribution under the null. Figure 4 shows the test process $H_n(t)$ along with 20 resampled processes from its limit distribution under the null, and it is seen that the test process does not seem to deviate in any respect. The supremum test based on 1000 resamples results in a P-value of 0.63. Based on this, we proceed to calculate the estimator $\hat{B}_X(t)$. This estimate along with 95% confidence bands (pointwise) are given in Figure 3 that also shows the intention-to-treat estimate (broken curve). The causal effect of the screening appears to be slightly more pronounced than what is seen from the intention-to-treat estimator and again it is seen that there seems to be a time-varying effect with screening being beneficial in a period of approximately 6 years. The supremum test $\sup_{t \leq 10} |\hat{B}_X(t)|$ is significant (P 0.02).

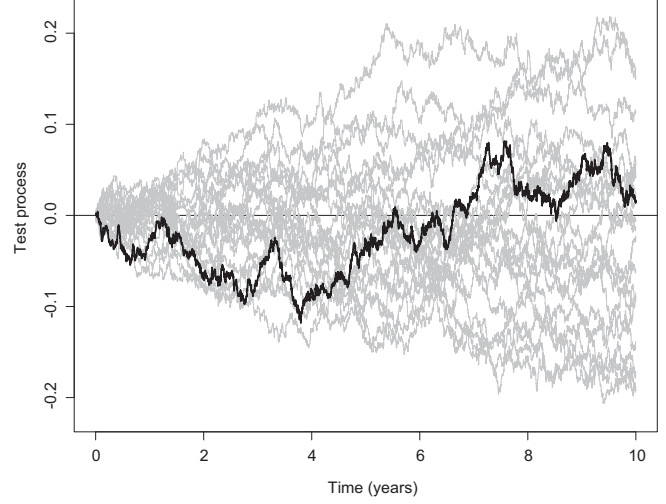


Figure 4. HIP-study. Investigation of whether the cause specific hazard of death due to other causes than breast cancer for the screened patients would have been the same at all times had they not been screened. Test process $H_n(t)$ along with 20 resampled processes from its limit distribution under the null.

Using the proposed approach, it is now possible to study the time-dynamics even further. For illustrative purposes, assume one hypothesized that an effect of screening would, if any, only last for a few years (as screening stopped after 3 years), and let us say it corresponds to roughly 6 years of follow-up. We could then attempt the simpler model

$$\beta_X(t) = \beta_0 I(t < \xi) + \beta_1 I(t \geq \xi). \quad (10)$$

with $\xi = 6$ years. The two parameters β_0 and β_1 are estimated by

$$\hat{\beta}_0 = \int_0^\xi w_0(t) d\hat{B}_X(t) \quad \hat{\beta}_1 = \int_\xi^\tau w_1(t) d\hat{B}_X(t)$$

with $w_0(t) = \tilde{w}(t) / \int_0^\xi \tilde{w}(s) ds$, $w_1(t) = \tilde{w}(t) / \int_\xi^\tau \tilde{w}(s) ds$, $\tilde{w}(t) = R \cdot(t) = \sum_i R_i(t)$. The estimate of $B_X(t)$ under this simplified model is then given by

$$B_X^\dagger(t) = \hat{\beta}_0 I(t < \xi) + \hat{\beta}_0 \xi I(t \geq \xi) + \hat{\beta}_1 (t - \xi) I(t \geq \xi), \quad (11)$$

The constant effects parameters are estimated to $\hat{\beta}_0 = -0.00031$ (SE 0.00011) and $\hat{\beta}_1 = -0.00012$ (SE 0.00020), indicating a significant effect of the screening only in the first 6 years. The estimator $B_X^\dagger(t)$ is also shown in Figure 3. To test whether the simplified model, that is, assuming a constant effect of treatment with a change in the effect after 6 years, gives a reasonable description of the data we consider the test process $TST(t) = n^{1/2} \{\hat{B}_X(t) - B_X^\dagger(t)\}$ which, under the null, can be written as

$$\begin{aligned} n^{1/2} \{\hat{B}_X(t) - B_X(t)\} - n^{1/2} (\hat{\beta}_0 - \beta_0) t I(t < \xi) \\ - n^{1/2} (\hat{\beta}_0 - \beta_0) \xi I(t \geq \xi) - n^{1/2} (\hat{\beta}_1 - \beta_1) (t - \xi) I(t \geq \xi). \end{aligned}$$

Using the iid representation of $W_n(t)$ we can resample from the limit distribution, under the null, of $TST(t)$; such 20 randomly picked processes are shown in Web Figure 1 along with the observed test process $TST(t)$. We may use the supremum test statistic $TST = \sup_{t \leq 10} |TST(t)|$ to investigate whether the test process is deviating. To see whether the observed TST is extreme we sampled 1000 draws from the limit distribution as outlined in Section 3; this gave a P-value of 0.56 suggesting that the constant effects model with a change in the effect after 6 years gives a reasonable fit to the data.

However, it also seen from Figure 3 and Web Figure 1 that the two parameter constant effects model is perhaps not giving a fully satisfactory fit in the first period of follow-up (two years or so). Actually, if instead one uses the test statistic $TST = \sup_{t \leq 6} |TST(t)|$ then one gets a P-value of 0.06 giving some indication of a non-satisfactory fit in the initial phase of the follow up period. One could consider extending the two parameter constant effects model with an additional parameter allowing for a separate effect in the initial phase of two years or so. The cutpoints chosen here were used for illustrative purposes only, in practice they should have been specified ahead of performing the analysis.

5. Concluding Remarks

In this article, we proposed an instrumental variables estimator for the effect of an arbitrary exposure on an event time. In comparison with other instrumental variables estimators for event times, our proposed approach has the advantage that it can handle arbitrary (e.g., continuous) exposures, without the need for modeling the exposure distribution, and that it naturally adjusts for censoring whenever censoring is independent of the event time, exposure, and instrument, conditional on measured and unmeasured confounders. The independent censoring assumption is relatively weak as it allows for a dependence on unmeasured factors. This assumption can be relaxed via inverse probability of censoring weighting under a model for the dependence of censoring on the exposure and/or instrumental variable.

Under the usual instrumental variable assumptions, listed in Section 1, the IV-estimator (8) provides a consistent estimator of the causal exposure effect as opposed to the naive estimator when there is unmeasured confounding. However, in the case of a weak instrument, the IV-estimator may have a large variance. It is therefore of interest to develop semiparametric efficient estimators (Tsiatis, 2006). Along the same lines, it is also of interest to consider estimators that are robust to some model deviations. For instance, consider the following two models

$$\begin{aligned} \lambda_{\tilde{T}0}(t|L) - \lambda_{\tilde{T}0}(t|L=0) &= \psi^T(t)L \text{ and} \\ E\{h(t, G, L)|L\} &= E\{h(t, G, L)|L; \theta\}, \end{aligned} \quad (12)$$

where h is a user defined function such as $h(t, G, L) = G$; and $\psi(t)$ and θ are parameters indexing the two models. Consider then the estimating function

$$\begin{aligned} d(t, L) \{h^*(t, G, L) - \bar{h}^*(t)\} e^{B_X(t)X} R(t) \\ \times \{dN(t) - dB_X(t)X - \psi^T(t)Ldt\}, \end{aligned} \quad (13)$$

where

$$\begin{aligned} h^*(t, G, L) &= h(t, G, L) - E(h(t, G, L)|L; \theta), \\ \bar{h}^*(t) &= \frac{E\{h^*(t, G, L)R(t)e^{B_X(t)X}\}}{E\{R(t)e^{B_X(t)X}\}}, \end{aligned}$$

and $d(t, L)$ is an arbitrary index function. One may then show that (13) has zero mean if either of the two models in (12) hold; the solution to an estimating equation based on estimating function (13) therefore yields a double robust estimator. This estimator has the further advantage of being invariant to linear transformations of the exposure. A detailed study of efficient and double robust estimators will be communicated in a separate report.

Our focus in this article has been on additive hazards models, whose parameters translate directly into relative survival chances. While 2SLS approaches for Cox models enjoy greater popularity at this point, these have no justification (except when few subjects experience an event over the length of follow-up (Tchetgen Tchetgen et al., 2015)). Since it is in particular unclear what causal parameter 2SLS estimators under Cox models are targeting, one cannot use the results obtained from such analyses to evaluate the impact of exposure on the survival time distribution.

6. Supplementary Materials

The Web Appendix referenced in Sections 3 and 4 is available with this article at the *Biometrics* website on Wiley Online Library. The Web Appendix also contains R code.

ACKNOWLEDGEMENTS

Torben Martinussen's work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. Stijn Vansteelandt was supported by IAP research network grant nr. P07/05 from the Belgian government (Belgian Science Policy). Eric Tchetgen Tchetgen is supported by NIH grant R01A1104459.

REFERENCES

- Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. *Lecture Notes in Statistics* **2**, 1–25.
- Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models *Journal of Econometrics* **113**, 231–263.
- Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Berlin: Springer-Verlag.
- Angrist, J. and Krueger, A. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* **15**, 69–85.
- Cai, B., Small, D. S., and Ten Have, T. R. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Statistics in Medicine* **30**, 1809–1824.
- Cuzick, J., Sasieni, P., Myles, J., and Tyrer, J. (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society, Series B* **69**, 565–588.
- Davey-Smith, G. and Ebrahim, S. (2003). Mendelian randomization: can genetic epidemiology contribute to understand-

- ing environmental determinants of disease? *International Journal of Epidemiology* **32**, 1–22.
- Didelez, V. and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* **16**, 309–330.
- Hernán, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology* **17**, 360–372.
- Joffe, M. M. (2001). Administrative and artificial censoring in censored regression models. *Statistics in Medicine* **20**, 2287–2304.
- Joffe, M. M., Yang, W. P., and Feldman, H. (2012). G-Estimation and Artificial Censoring: Problems, Challenges, and Applications. *Biometrics* **68**, 275–286.
- Katan, M. B. (1986) Apolipoprotein E isoforms, serum cholesterol, and cancer. *Lancet* **1**, 507–8.
- Li, J., Fine, J., and Brookhart, A. (2015). Instrumental variable additive hazards models. *Biometrics* **71**, 122–130.
- Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 577–572.
- Loeys, T., Goetghebeur, E., and Vandebosch, A. (2005). Causal proportional hazards models and time-constant exposure in randomized clinical trials. *Lifetime Data Analysis* **11**, 435–449.
- Martinussen, T. (2010). Dynamic path analysis for event time data: large sample properties and inference. *Lifetime Data Analysis* **16**, 85–101.
- 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* **73**, 773–788.
- Martinussen, T. and Vansteelandt, S. (2013). On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis* **19**, 279–296.
- Nie, H., Cheng, J., and Small, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics* **67**, 1397–1405.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Picciotto, S., Hernán, M. A., Page, J., Young, J. G., and Robins, J. M. (2012). Structural nested cumulative failure time models to estimate the effects of hypothetical interventions. *Journal of the American Statistical Association* **107**, 886–900.
- Robins, J. M. and Tsiatis, A. (1991). Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. *Communications in Statistics* **20**, 2609–2631.
- Shapiro, S. (1977). Evidence of screening for breast cancer from a randomised trial. *Cancer* **39**, 2772–2782.
- Tchetgen Tchetgen, E. J., Walter, S., Vansteelandt, S., Martinussen, T., Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26**, 402–410.
- Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.
- Vansteelandt, S., Bowden, J., Babanezhad, M., and Goetghebeur, E. (2011). On instrumental variable estimation of the causal odds ratio. *Statistical Science* **26**, 403–422.

Received November 2016. Revised March 2017.

Accepted March 2017.