Instrumental variables estimation with competing risk data

TORBEN MARTINUSSEN*

*Section of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B,
1014 Copenhagen K, Denmark*

tma@sund.ku.dk

STIJN VANSTEELANDT

*Department of Applied Mathematics, Computer Sciences and Statistics, Ghent University, Krijgslaan
281 (S9), 9000 Gent, Belgium and Department of Medical Statistics, Centre for Statistical Methodology,
London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*

Summary

Time-to-event analyses are often plagued by both—possibly unmeasured—confounding and competing
risks. To deal with the former, the use of instrumental variables (IVs) for effect estimation is rapidly gaining
ground. We show how to make use of such variables in competing risk analyses. In particular, we show
how to infer the effect of an arbitrary exposure on cause-specific hazard functions under a semi-parametric
model that imposes relatively weak restrictions on the observed data distribution. The proposed approach
is flexible accommodating exposures and IVs of arbitrary type, and enabling covariate adjustment. It
makes use of closed-form estimators that can be recursively calculated, and is shown to perform well in
simulation studies. We also demonstrate its use in an application on the effect of mammography screening
on the risk of dying from breast cancer.

Keywords: Causal effect; Competing risk; Instrumental variable; Time-to-event; Unobserved confounding.

1. Introduction

In most observational studies unobserved confounding cannot be ruled out. This can make the results on
exposure effects, as obtained via standard regression methods, questionable. Sometimes, however, it may
be possible to estimate an exposure effect without (large sample) bias when an instrumental variable (IV)
is available. This is a variable which is (i) associated with the exposure, (ii) has no direct effect on the
outcome other than through the exposure, and (iii) whose association with the outcome is not confounded
by unmeasured variables (e.g. Hernán and Robins 2006). Condition (i) is empirically verifiable, but
conditions (ii) and (iii) are not.

IVs estimation of exposure effects is well established for continuous outcomes that obey linear models.
One popular technique is two-stage predictor substitution (2SLS) estimation, sometimes also referred to
as the 2SPS method (Cai *and others*, 2011). Here, the exposure variable is regressed on the instrument in

---

*To whom correspondence should be addressed.

the first stage, and then the outcome variable is regressed on the predicted exposure value in the second stage. The regression coefficient of the predicted exposure in the second stage is then interpreted as the exposure effect of interest.

Recently there has been a focus on extending these methods to handle also right censored failure time data. Robins and Tsiatis (1991) initiated this work, but although they developed a general estimating equations-based method under structural accelerated failure time models, their proposal suffers from a lack of smoothness of the estimating equations due to the way how censoring is handled (Joffe *and others*, 2012). Tchetgen Tchetgen *and others* (2015) developed an easy-to-use two-stage estimation approach under additive hazard models for event times, which works when the exposure obeys a location shift model (see Li *and others*, 2015 for a related approach). Martinussen *and others* (2017) generalized these methods by working under a less restrictive semiparametric structural cumulative failure time model, imposing no restrictions on distribution (or type) of instrument or exposure. Their proposal has the further advantage of enabling non-parametric estimation of a possibly time-varying exposure effect. Kjaersgaard and Parner (2016) suggested an alternative approach based on pseudo-observations. Their 2SLS method requires a latent additive model for the target parameter which is not so attractive when focusing on a distribution function.

Motivated by an analysis of the HIP-study (Health Insurance Plan of Greater New York, see Joffe 2001), which was designed to assess the potential effect of breast cancer screening, here, we aim at extending the methods of Martinussen *and others* (2017) to handle competing risk data. The HIP-study comprised approximately 60 000 women, who were randomized into two approximately equally sized groups. About 35% of the women, who were offered screening, refused to participate, resulting in a problem of non-compliance. We planned to correct for this using randomization as an IV. In the first 10 years of follow-up, there were 4221 deaths, but only 340 were deemed due to breast cancer, making competing risks a major issue in these data.

Richardson *and others* (2017) proposed a method that can deal with competing risk data also using an IV approach. Their approach requires the instrument as well as exposure variable to be binary variables. Essentially they generalize the standard IV Wald estimator for survival probabilities to estimate cumulative incidence probabilities. In this way, they estimate the so-called complier average treatment effect. The method we propose puts no restriction on the type of instrument nor on the exposure, and it can also incorporate covariates which is not possible using the Wald type estimator of Richardson *and others* (2017). Our method is thus much more general. Zheng *and others* (2017) suggest a two-stage approach where they directly model the subdistribution hazard. This quantity in itself is hard to interpret as it involves a risk set of those alive and those that have died from the competing cause, but it can be seen just as a tool to model the cumulative incidence function, and it is widely used. For further discussion on this, we refer to Andersen and Keiding (2012) and Andersen *and others* (2012). Zheng *and others* (2017) require a model for the exposure distribution that involves unmeasured variables. This model is assumed to be linear which is geared toward continuous exposure as it may not be natural to use for a binary exposure. They also need that observed and unobserved covariates are independent. Our method avoid such an assumption and puts no restriction on the exposure distribution. In particular, it can handle binary and continuous exposures.

The article is structured as follows. In Section 2, we specify the model and outline the estimation procedure. Section 3 contains large sample results. In Section 4, we study by simulations the practical behavior of the proposed estimator and also analyze the HIP-data. Section 5 contains some closing remarks and technical details are deferred to the Appendix.

2. Model specification and estimation

We let $\tilde{T}$ denotes the time until one of the two competing events happens and let $\delta = 1, 2$ denotes which of the two that takes place. Our aim is to assess the effect of an arbitrary exposure X on the cause-specific

hazard of each of these competing events, by making use of an IV G. This variable is such that, possibly conditional on measured covariates L, G is associated with the exposure X, but is not associated with the event time $\tilde{T}$, nor the event type δ, except because of a possible exposure effect. More formally, let $(\tilde{T}^x, \delta^x)$ denote the counterfactual event time and event type that would be observed for given subject if the exposure of that subject were set to x. We will make the consistency assumption that these coincide with the observed event time $\tilde{T}$ and event type δ for those subjects who happen to have exposure level $X = x$. This notation enables us to be clear about our target of inference, which is the contrast between the counterfactual cause-specific hazard functions

$$\lambda_{T^x}^j(t|X = x, G, L) - \lambda_{T^0}^j(t|X = x, G, L), \tag{2.1}$$

for $j = 1, 2$, where

$$\lambda_{T^x}^j(t|X, G, L) = \frac{\frac{\mathrm{d}}{\mathrm{d}t} P\left(\tilde{T}^x \le t, \delta^x = j|X, G, L\right)}{P\left(\tilde{T}^x > t|X, G, L\right)}, \quad j = 1, 2. \tag{2.2}$$

Because T^0 and δ^0 are unobserved for subjects with non-zero exposure, we will rely on the assumption that G is an IV for the exposure effect (conditional on L), in the sense that $(\tilde{T}_0, \delta^0)$ is conditionally independent of G, given L (Hernán and Robins, 2006). This assumption expresses that, if all subjects received zero exposure, both events would have cause-specific hazards (conditional on L) that would be same at all levels of G. This would be the case when, as in the causal directed acyclic graph of Figure 1, G shares no common causes with the event time and type, and does not influence those in the absence of exposure. As in other IVs problems, these IVs assumptions will not generally suffice to identify the contrast (2.1) at all levels of X, G, and L. For that reason, as well as for reasons of parsimony, we will assume that the following structural model holds

$$\lambda_{T^x}^j(t|X = x, G, L) - \lambda_{T^0}^j(t|X = x, G, L) = \beta_j(t)x, \tag{2.3}$$

for all $t > 0$ and for $j = 1, 2$, with $\beta_j(t)$ an unknown, locally integrable function. Our aim is then to estimate $B_j(t) = \int_0^t \beta_j(s)\,\mathrm{d}s$ for all $t > 0, j = 1, 2$. As in Martinussen *and others* (2017), it can be shown that this model is satisfied when the causal directed acyclic graph of Figure 1 holds and, moreover,

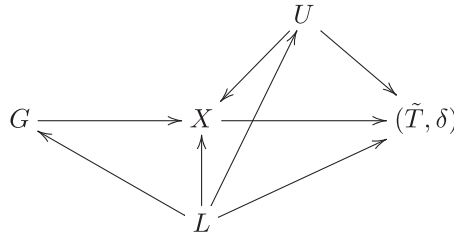$$\lambda^j(t|X, G, L, U) = \lambda_j(t) + \beta_j(t)X + \psi_j(t, U, L),$$

Fig. 1. Causal directed acyclic graph. G is the instrument, X is the exposure variable, and $\tilde{T}$ is the time-to-event and δ indicates which of the two competing events that is taking place. The potential unmeasured confounders are denoted by U, and the observed confounders by L.

with the function $\psi_j(.), j = 1, 2$ left unspecified and the functions $\lambda_j(.)$ and $\beta_j(t)$ unknown. Model (2.3) excludes interactions of the exposure effect with both G and L. Whereas the model can be extended to allow for interactions between X and L (see Section 5), the absence of interactions between X and G is assumed in order to obtain identification of the exposure effect. Others (e.g. Richardson *and others* 2017) avoid this no-interaction assumption by restricting themselves to settings with dichotomous instrument and exposure, and focusing on the treatment effect in compliers; here, compliers refer to individuals who would take treatment if assigned to one level of the instrument (say, Level 1), but not the other (say, Level 0). They instead invoke a monotonicity assumption that increasing the level of the instrument cannot decrease the exposure for any individual. Here, we avoid this framework because it does not easily extend to continuous instruments or exposures, and because in our opinion, there is usually greater scientific interest in knowing the exposure effect for the general study population.

Throughout, 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, i.e. we observe $D = I(\tilde{T} \leq C)$, along with the first time either failure or censoring occurs, i.e. we also observe $T = \min(\tilde{T}, C)$ as well as δ if $D = 1$. Let $\delta = 0$ if $D = 0$. Define also the observed counting processes $N^j(t) = I(T \leq t, D = 1, \delta = j)$, $j = 1, 2$, and the at risk indicator $R(t) = I(t \leq T)$. We assume that the censoring time satisfies the following condition

$$\tilde{T} \perp\!\!\!\perp C | X, G, L \text{ and } P(C > t | X, G, L) = P(C > t | L) \tag{C}$$

The above condition on the censoring distribution can be relaxed to $P(C > t | X, G, L, U) = P(C > t | L, U)$ for some variable $U \perp\!\!\!\perp G | L$.

The following proposition lays the basis of the estimation procedure for $B_j(t)$, which we will describe next.

PROPOSITION 1 Assume the structural model (2.3) with the assumption that G is an instrumental variable, conditional on L, and further that the censoring time satisfies condition (C). Then

$$E\left[\{G - E(G|L)\} e^{B_1(t)X + B_2(t)X} R(t) \left\{dN^j(t) - dB_j(t)X\right\}\right] = 0, \tag{2.4}$$

for each $t, j = 1, 2$.

Proof. By the independent censoring assumption (C) and

$$\frac{P\left(\tilde{T}^0 > t | X, G, L\right)}{P\left(\tilde{T} > t | X, G, L\right)} = e^{B_1(t)X + B_2(t)X}$$

it follows, for $j = 1, 2$, that

$$E\left[\{G - E(G|L)\} e^{B_1(t)X + B_2(t)X} R(t) \left\{dN^j(t) - dB_j(t)X\right\}\right]$$
$$= E\left[\{G - E(G|L)\} e^{B_1(t)X + B_2(t)X} I(C > t) I(\tilde{T} > t) \lambda_{\tilde{T}^0}^j(t|X, G, L) dt\right]$$

$$= E\left[P(C > t|L)\{G - E(G|L)\}P(\tilde{T}^0 > t|X, G, L)\lambda^j_{\tilde{T}^0}(t|X, G, L)\mathrm{d}t\right]$$

$$= E\left[P(C > t|L)\{G - E(G|L)\}\frac{\mathrm{d}}{\mathrm{d}t}P(\tilde{T}^0 \le t, \delta^0 = j|X, G, L)\mathrm{d}t\right]$$

$$= \frac{\mathrm{d}}{\mathrm{d}t}E\left[P(C > t|L)\{G - E(G|L)\}P(\tilde{T}^0 \le t, \delta^0 = j|G, L)\mathrm{d}t\right]$$

$$= 0$$

because, for any function $g_t(L)$, we have

$$E\left[g_t(L)\{G - E(G|L)\}P(\tilde{T}^0 \le t, \delta^0 = j|G, L)\mathrm{d}t\right] = 0$$

since $G \perp\!\!\!\perp (\tilde{T}^0, \delta^0)|L$. This completes the proof.   □

The above proposition gives two unbiased estimating functions for each time t, on the basis of which we can construct a consistent estimator of $B_j(t)$. In particular, let $(T_i, D_i, \delta_i, L_i, G_i, X_i)$, $i = 1, \ldots, n$, denote n independent identically distributed replicates under the structural model (2.3) together with the IVs assumptions. Suppose that the counting processes $N^j_i(t) = I(T_i \le t, D_i = 1, \delta_i = j)$, $i = 1, \ldots, n$, $j = 1, 2$, are observed in the time interval $[0, \tau]$, where τ is some finite time point. Solving equation (2.4), with population expectations substituted by sample analogs leads to the recursive estimator $\hat{B}_j(t)$ defined by

$$\hat{B}_j(t, \hat{\theta}) = \int_0^t \frac{\sum_i G^c_i(\hat{\theta})\mathrm{e}^{\{\hat{B}_1(s-)+\hat{B}_2(s-)\}X_i}\mathrm{d}N^j_i(s)}{\sum_i G^c_i(\hat{\theta})R_i(s)\mathrm{e}^{\{\hat{B}_1(s-)+\hat{B}_2(s-)\}X_i}X_i}, \tag{2.5}$$

where $G^c_i(\theta) = G_i - E(G_i|L_i; \theta)$, 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). Note that $\hat{B}_j(t, \hat{\theta})$ is a step function that is well defined by setting $\hat{B}_j(0, \hat{\theta}) = 0$, $j = 1, 2$. Furthermore, note that—unlike many other IV estimators—the estimator (2.5) can be evaluated for discrete as well as continuous exposures and instruments. It does not require distributional assumptions for the exposure and does not make assumptions as to how measured covariates relate to the event time.

In practice, it may be more appealing to report some function of the cumulative incidence functions. In the following, we focus on the situation where the exposure is binary. One may for instance focus on the relative risk function

$$\mathrm{RR}(t) \equiv \frac{P(T^0 \le t, \delta^0 = 1|X = 1, G, L)}{P(T^1 \le t, \delta^1 = 1|X = 1, G, L)} \tag{2.6}$$

or the absolute difference function

$$\mathrm{AD}(t) \equiv P(T^0 \le t, \delta^0 = 1|X = 1, G, L) - P(T^1 \le t, \delta^1 = 1|X = 1, G, L). \tag{2.7}$$

We have

$$P(T^1 \le t, \delta^1 = 1|X = 1, G, L) = P(T \le t, \delta = 1|X = 1, G, L),$$

which is expressed via the observed data. Also

$$P(T^0 \leq t, \delta^0 = 1 | X = 1, G, L)$$

$$= \int_0^t e^{-\Lambda^1_{T^0}(s|X=1,G,L) - \Lambda^2_{T^0}(s|X=1,G,L)} d\Lambda^1_{T^0}(s|X=1,G,L) ds$$

$$= \int_0^t e^{-\Lambda^1_T(s|X=1,G,L) - \Lambda^2_T(s|X=1,G,L) + B_1(s) + B_2(s)} \{ d\Lambda^1_T(s|X=1,G,L) - dB_1(s) - dB_2(s) \}$$

By formulating a model for the observed data via the cause-specific hazard functions $\lambda^j_T(s|X,G,L)$, $j = 1, 2$, and using the estimator (2.5) we may then estimate $RR(t)$ and $AD(t)$. In general, however, specific models for the observed data may not be congenial with the structural model meaning that there may no data generating mechanism such that both models hold, see Robins and Rotnitzky (2004). In the situation without the covariate L, and where exposure and instrument are both binaries we may estimate the cumulative cause-specific hazard functions for the observed data using non-parametric estimators and hence avoiding the issue of non-congeniality. This is the situation for the HIP-data that we study in Section 4.2. In practice, one can calculate $\hat{\Lambda}^j_T(s|X=1,G), j = 1, 2$, using the `aalen`-function of the R-package `timereg`.

3. LARGE SAMPLE PROPERTIES

The following proposition, whose proof is sketched in the Appendix, shows that the estimators $\hat{B}_j(t)$, $j = 1, 2$, are uniformly consistent. It moreover gives the asymptotic distribution of the two estimators.

PROPOSITION 2   Under model (2.3) with the assumption (C) and the assumption that G is an IV, conditional on L, and given the technical conditions listed in the Appendix, the IV estimators $\hat{B}_j(t)$, $j = 1, 2$, are uniformly consistent. Furthermore, $W^j_n(t) = n^{1/2} \{ \hat{B}_j(t, \hat{\theta}) - B_j(t) \}$ converges in distribution to a zero-mean Gaussian process with variance $\Sigma_j(t)$. A uniformly consistent estimator $\hat{\Sigma}_j(t)$ of $\Sigma_j(t)$ is given below.

Let $\epsilon^{B_j}_i(t, \theta), i = 1, ..., n$ be the iid zero-mean processes given by expression (A.1) in the Appendix. From the proof in the Appendix, it then follows that $W^j_n(t)$ is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^n \epsilon^{B_j}_i(t, \theta)$. The variance $\Sigma_j(t)$ of the limit distribution can thus be consistently estimated by

$$\hat{\Sigma}_j(t) = n^{-1} \sum_{i=1}^n \{ \hat{\epsilon}^{B_j}_i(t, \hat{\theta}) \}^2, \tag{3.1}$$

where $\hat{\epsilon}^{B_j}_i(t, \hat{\theta})$ is obtained from $\epsilon^{B_j}_i(t, \theta)$ by replacing unknown quantities with their empirical counterparts. These results can be used to construct a pointwise confidence band.

In the case without additional covariates L, and with X and G being binaries, $n^{1/2} \{ \hat{\Lambda}^j_T(s|X=1,G) - \Lambda^j_T(s|X=1,G) \}$ also allows an iid-representation (Martinussen and Scheike, 2006), and it can furthermore be extracted from the `aalen`-function in R. Now using the functional delta theorem we may then obtain iid-representations of $n^{1/2} \{ \widehat{RR}(t) - RR(t) \}$, and of $n^{1/2} \{ \widehat{AD}(t) - AD(t) \}$, that may then be used to also calculate confidence bands. We illustrate this using the HIP-data (see Section 4.2).

4. Numerical results

In this section, we investigate the practical behavior of the proposed estimator. In Section 4.1, we study the small sample performance using simulations, and in Section 4.2, we give a worked application using the HIP-data on the potential effect of breast cancer screening on death due to breast cancer.

4.1. *Simulation study*

To investigate the properties of our proposed methods with practical sample sizes, we conducted a simulation experiment, whereby we generated data under the causal Directed Acyclic Graph of Figure 1. 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 $Var(X|G) = Var(U|G) = 0.25$, and $Cov(X, U|G) = -1/6$. The parameter γ_G determines the size of the correlation between exposure and the IV. Specifically we looked at correlation ρ equal to 0.3 and 0.5. The two cause-specific hazards were given as

$$\lambda^1(t|X, G, U) = 0.1 + 0.1U$$
$$\lambda^2(t|X, G, U) = 0.1 + 0.2X + 0.1U$$

so $\tilde{T}$ is generated according to the hazard

$$\lambda_{\tilde{T}}(t|X, G, U) = 0.2 + 0.2X + 0.2U$$

and failure of Type 1 happens with probability

$$\frac{\lambda^1(t|X, G, U)}{\lambda^1(t|X, G, U) + \lambda^2(t|X, G, U)},$$

and likewise with failure of Type 2. It is easily seen that model (2) holds under this model. 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 an overall censoring rate of around 17. For this scenario, we considered sample size 1600 when $\rho = 0.3$, and sample size 1000 when $\rho = 0.5$. Simulation results are given in Table 1 based on 2000 runs for each configuration, where we report average biases at time points $t = 0.5, 1.5, 2.5$ for $\hat{B}_j(t)$. We also report the empirical standard errors as well as estimated standard errors based on formula (3.1) along with coverage probability of 95% pointwise confidence intervals $CP(\hat{B}_j(t))$. Biases from the naive Aalen estimator, denoted as $\tilde{B}_j(t)$ in the table, running Aalen's additive hazards model (Martinussen and Scheike 2006, Ch. 5) for the two cause-specific hazards using X and G as covariates are also given.

It is seen from Table 1 that the suggested estimators are unbiased and also that the estimated standard errors estimate well the variability resulting in satisfactory coverage probabilities. As expected, the naive estimators $\tilde{B}_j(t), j = 1, 2$, are biased.

An additional simulation study is reported in the supplementary material available at *Biostatistics* online. In the supplementary material available at *Biostatistics* online, we took both the exposure variable and the instrument to be continuous variables. To our knowledge, there are no other available methods to handle such a situation.

4.2. *Application to the HIP trial on effectiveness of screening on breast cancer mortality*

The HIP of Greater New York was a randomized trial of breast cancer screening that began in 1963. About 60 000 women aged 40–60 were randomized into two approximately equally sized groups. Study

Table 1. *Continuous exposure and binary instrument*

	n	$\rho = 0.3$ $t = 0.5$	$t = 1.5$	$t = 2.5$	n	$\rho = 0.5$ $t = 0.5$	$t = 1.5$	$t = 2.5$
Bias $\hat{B}_1(t)$	1600	0.001	0.007	−0.007	1000	−0.000	−0.001	−0.004
SD $(\hat{B}_1(t))$		0.061	0.132	0.233		0.042	0.089	0.149
See $(\hat{B}_1(t))$		0.061	0.130	0.231		0.042	0.089	0.149
95% CP$(\hat{B}_1(t))$		95.4	95.3	96.6		95.0	95.5	95.7
Bias $\hat{B}_2(t)$		0.002	−0.001	−0.010		0.000	−0.000	−0.005
SD $(\hat{B}_2(t))$		0.077	0.165	0.300		0.053	0.113	0.198
See $(\hat{B}_2(t))$		0.076	0.165	0.296		0.054	0.117	0.199
95% CP$(\hat{B}_2(t))$		94.8	95.6	96.0		95.2	96.1	96.5
Bias $\tilde{B}_1(t)$		−0.033	−0.099	−0.167		−0.034	−0.101	−0.168
Bias $\tilde{B}_2(t)$		−0.034	−0.102	−0.170		−0.035	−0.102	−0.168

The two causes are given by $j = 1, 2$. Bias of $\hat{B}_j(t)$, average estimated standard error, sd$(\hat{B}_j(t))$, empirical standard error, see$(\hat{B}_j(t))$, and coverage probability of 95% pointwise confidence intervals CP$(\hat{B}_j(t))$ based on the instrumental variables estimator, in function of sample size n and at different strengths ρ (correlation) of the instrumental variable. Bias of $\tilde{B}_j(t)$ is the bias of the naive Aalen estimator.

women were offered the screening examinations consisting of clinical examination, and a mammography. Further three annual examinations were offered in this group. Control women continued to receive their usual medical care. There were 30 565 women in the control group and 30 130 in the screening group of which 9984 (35%) refused to participate (non-compliers). 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 potential unobserved confounding.

In this setting, the exposure is the treatment and the instrument is the original randomization, which is unrelated to any unobserved confounders. Further, there is no reason to believe that the instrument should have any effect in itself on the risk of dying of any cause, only (potentially) via the treatment. Exposure and instrument were correlated, estimated correlation coefficient is 0.71, being highly significant. We applied the estimator given by (2.5) to these data focusing on the first 10 years of follow-up. This estimator is shown in Figure 2 along with 95% pointwise confidence intervals. The left panel gives results for breast cancer and the right panel for other causes. The intention to treat estimator is also shown (dotted curves). It is seen from Figure 2 that breast cancer screening appears to lower the risk of dying from breast cancer while there is no evidence of an effect of screening on the risk of dying from other causes. The impact of the screening on the risk of dying from breast cancer seems to be slightly more pronounced than what is indicated by the intention to treat estimator. As the specific value of $B_1(t)$ may be hard to interpret, we suggest also to report the estimated relative risk function $\widehat{RR}(t)$ and/or the absolute risk function $\widehat{AD}(t)$. Figure 3 displays $\widehat{RR}(t)$ along with 95% pointwise confidence bands. It is seen that for women who received screening on the active arm the risk of dying from breast cancer within 5 (10) years would have been approximately twice (1.5 times) as large had they not received screening.

Figure 4 displays $\widehat{AD}(t)$ along with 95% pointwise confidence bands. Again it is seen that for women who received screening on the active arm the risk of dying from breast cancer within 5 (10) years would have a larger risk had they not received screening. We also see, however, that the risk of dying of due to breast cancer is low for the women in this study giving a small (but significant) difference of the two cumulative incidence functions. Figure 4 also displays the Wald-type estimator of Richardson *and others* (2017), dotted curve. Their estimator estimates the causal effect among compliers and is constructed to
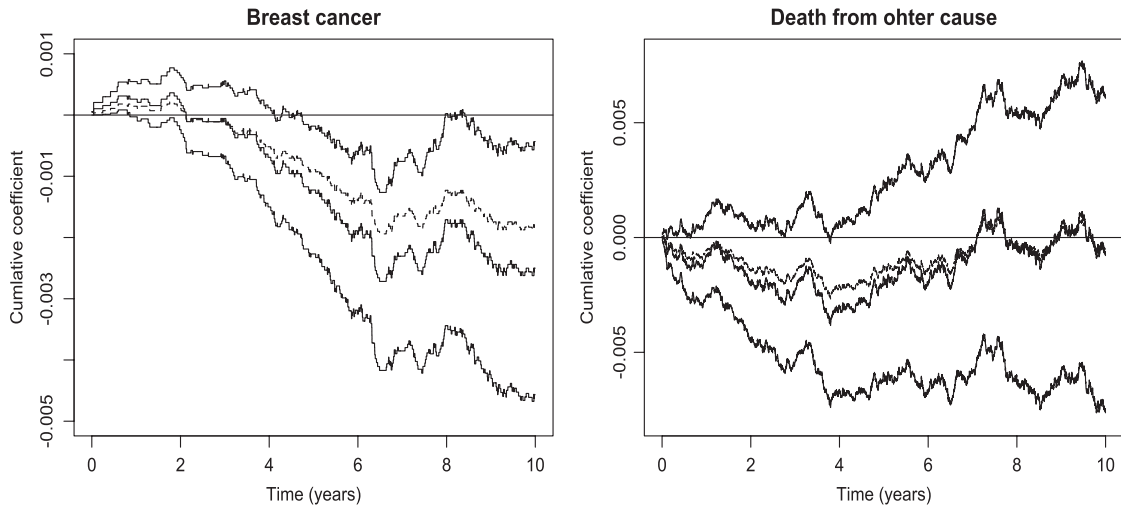
Fig. 2. HIP-study. Estimated causal effect of screening, $\hat{B}(t)$ along with 95% pointwise confidence bands and the intention to treat estimate (broken curve). Curves are given for the two competing events.
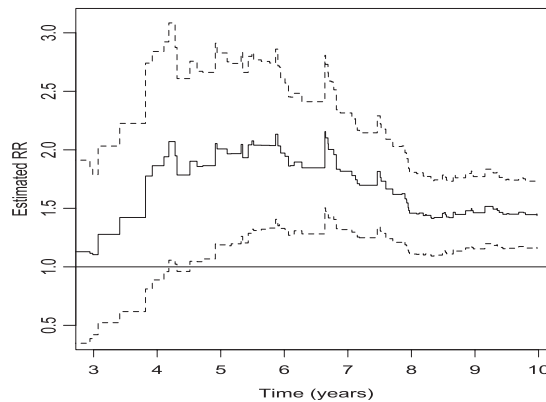
Fig. 3. HIP-study. Estimated relative risk function (breast cancer death), see display (2.6), along with 95% pointwise confidence bands (dotted curves).

the situation where both exposure and instrument are binaries. For the HIP-data, those randomized not to receive screening never got access to screening (so-called no contamination), and therefore the Wald-type estimator, in this special case, estimates the causal effect among the treated as is the case for our estimator (2.5). Hence in the special setting in the HIP-study the two estimators estimate the same quantity, and indeed it is seen from Figure 4 that they give almost identical results.

5. Concluding remarks

In this article, we have proposed an approach to estimate causal effects in a competing risk setting where there may be unobserved confounding. The proposal is based on the availability of an IV. It can accommodate adjustment for baseline covariates, which is sometimes needed to make the IVs assumptions more plausible. Unlike available IV methods, it makes no restriction on the type, nor the distribution of
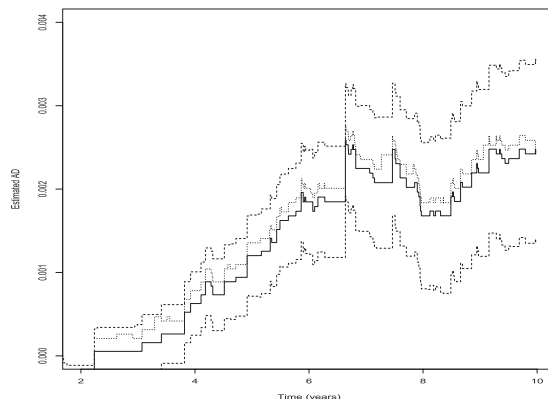
Fig. 4. HIP-study. Estimated absolute risk function concerning breast cancer death (full line), see display (2.7), along with 95% pointwise confidence bands (broken curves). The Wald-type estimator of Richardson *and others* (2017) is also shown (dotted curve).

exposure or instrument. We can in particular deal with a situation where the exposure is continuous and the instrument is categorical, or where both are continuous. Dichotomization of the exposure, which is sometimes considered by simpler proposals, is no valid remedy in such cases as it may entail a violation of the exclusion restriction.

A further strength of the approach is that it naturally adjusts for censoring whenever censoring is independent of the event time, exposure and instrument conditional on the confounders (even when unmeasured). Although this assumption could fail, it can be remedied by applying inverse probability of censoring weighting; that is, by redefining $e^{B_1(t)X + B_2(t)X} R(t)$ in (2.4) to

$$
e^{B_1(t)X + B_2(t)X} R(t) \frac{P(C \geq t | \mathcal{F}_t^N \vee L)}{P(C \geq t | \mathcal{F}_t^N \vee X \vee G \vee L)},
$$

where $\mathcal{F}_t^N$ denotes the history spanned by the counting processes. Using this modification requires postulating two models for the cause-specific hazard of censoring: one conditional on X, G, and L, and one conditional on L only. However, only the former model must be correctly specified to maintain a consistent estimator.

The results in this article may be extended to more general models than (2.3), to also handle interactions with the confounder L. For instance, suppose that instead of (2.3) the following model holds, for $j = 1, 2$,

$$
\lambda_{T^x}^j(t | X = s, G, L) - \lambda_{T^0}^j(t | X = x, G, L) = \beta_j(t)x + \beta_{jXL}^T(t)xl \tag{5.1}
$$

where $\beta_{jXL}(t)$ is of dimension corresponding to L, say p. Let $B_*(t)$ denote the integral from 0 to t of the corresponding $\beta_*(t)$, and define $B(t) = \{B_1(t), B_2(t), B_{1XL}^T(t), B_{2XL}^T(t)\}$. Let $dN(t) = \{dN_1^1(t), dN_1^2(t), \ldots, dN_n^1(t), dN_n^2(t)\}^T$. We can the write the estimator of $B(t)$ as

$$
\hat{B}(t, \theta) = \int_0^t \left[\{J_1(s, \theta)J_2(s, \theta)\}^T \{J_1(s, \theta)J_2(s, \theta)\} \right]^{-1} \left[\{J_1(s, \theta)J_2(s, \theta)\}^T J_1(s, \theta) dN(s), \right.
$$

where $J_1(t, \theta)$ is $1 \times 2n$-matrix with ith row

$$\{G_i - E(G_i|L_i)\} \exp\left\{\sum_{j=1}^{2} \hat{B}_j(t-, \theta)X_i + \hat{B}_{jXL}^T(t-, \theta)X_iL_i\right\}R_i(t),$$

and $J_2(t)$ is $2n \times (2 + 2p)$-matrix consisting of n blocks of size $2 \times (2 + 2p)$, where the ith block is

$$\begin{bmatrix} X_i & 0 & X_iL_i^T & 0 \\ 0 & X_i & 0 & X_iL_i^T \end{bmatrix}.$$

We have shown how to estimate functions of the cumulative incidence functions, $RR(t)$ and $AD(t)$, in the case without covariates L and with the exposure and instrument both being binaries. In principle, this may be extended to the general situation with covariates L and without restrictions on the exposure and instrument. However, one then need to specify the cause-specific hazard functions, and such models may not be congenial with the assumed structural model. We plan to investigate this further in future work.

6. Software

Simulation script with data is available at https://github.com/TorbenMart/IV_Compet_Risk_2018.

Supplementary material

Supplementary material is available online at http://biostatistics.oxfordjournals.org.

Acknowledgments

Funding

Appendix: large sample properties

Consistency

Let $\mu(L; \theta) = E(G|L; \theta)$ be the conditional mean of the instrument given observed confounders L, which is function of an unknown finite-dimensional parameter θ. In the case of no observed confounders $\mu(\theta) = \theta = E(G)$ and $\hat{\theta} = \bar{G}$. We assume that $n^{1/2}(\hat{\theta} - \theta) = n^{-1/2}\sum_i \epsilon_i^\theta + o_p(1)$, where the ϵ_i^θ's are zero-mean iid variables.

We write $\|g\|_\infty = \sup_{t\in[0,\tau]} |g(t)|$. Let $B_j^\circ(t)$ denote the true value of $B_j(t)$, and let $M_j^\circ = \|B_j^\circ\|_\infty < \infty$. Conditions:

- (i) We assume that X and G are bounded.
- (ii) Define $a(s, h) = E[R(s)XG^ce^{hX}]$. We assume that there exist $M > M_j^\circ, j = 1, 2$, and $\nu > 0$ such that $\inf_{s\in[0,\tau],h\in[-M,M]} a(s, h) \geq 1.01\nu$.

The quantities M_j° and M do not necessarily need to be known. Under these assumptions we may modify the arguments given in Martinussen *and others* (2017) to also cover the competing risk situation described here. Hence, consistency can be inferred similarly.

Asymptotic normality

Let $N(t)$ be the $n \times 2$ matrix with ith row $\{N_i^1(t), N_i^2(t)\}$ and $X = (X_1, \ldots, X_n)^T$. For known θ, we can write

$$\hat{B}(t,\theta) = \{\hat{B}_1(t,\theta), \hat{B}_2(t,\theta)\} = \int_0^t H_\theta\{s, \hat{B}_1(s-,\theta) + \hat{B}_2(s-,\theta)\}dN(s),$$

where the kth element of the n-vector $H_\theta\{t, a\}$ is

$$R_k(t)G_k^c(\theta)e^{aX_k} / \sum_{i=1}^n G_i^c(\theta)R_i(t)e^{aX_i}X_i.$$

with $a = \hat{B}_1(t-,\theta) + \hat{B}_2(t-,\theta)$. Let $W_n(t,\theta) = n^{1/2}\{\hat{B}(t,\theta) - B(t)\}$, $b = (1,1)^T$, and let $\dot{H}$ denote the derivative of H with respect to its second argument. It is then easy to see that

$$W_n(t,\theta) = n^{1/2}\int_0^t H\{s, B_1(s-) + B_2(s-)\}\,[dN(s) - X\,dB(s)]$$

$$+ \int_0^t V(s-,\theta)\{1 + o_p(1)\}b\dot{H}\{s, B_1(s-) + B_2(s-)\}dN(s)$$

which is a Volterra-equation, see Andersen *and others* (1993, p 91). The solution to this equation is given by

$$W_n(t,\theta) = \int_0^t n^{1/2}H\{s, B_1(s-) + B_2(s-)\}\,[dN(s) - X\,dB_X(s)]\,\mathcal{F}(s,t) + o_p(1),$$

where

$$\mathcal{F}(s,t) = \prod_{(s,t]}\left\{I + b\dot{H}(\cdot, B_1(\cdot) + B_2(\cdot))dN(\cdot)\right\}$$

with the latter being a product integral that converges in probability to some limit. This leads to the iid-representation

$$W_n(t,\theta) = n^{-1/2}\sum_{i=1}^n \epsilon_i^B(t)$$

with the $\epsilon_i^B(t)$'s being zero-mean iid terms. Specifically

$$\epsilon_i^B(t) = \int_0^t n^{1/2}\{H\{s, B_1(s-) + B_2(s-)\}\}_i\,[dN(s) - X\,dB_X(s)]_i\,\mathcal{F}(s,t)$$

with a_i being the ith element of the vector a, and $[A]_i$ being the ith row of the matrix A. This together with

$$n^{1/2}\{\hat{B}(t,\hat{\theta}) - B(t)\} = n^{1/2}\{\hat{B}(t,\theta) - B(t)\} + n^{1/2}\{\hat{B}(t,\hat{\theta}) - \hat{B}(t,\theta)\}$$
$$= n^{1/2}\{\hat{B}(t,\theta) - B(t)\} + D_\theta(\hat{B}(t,\theta)^T)_{|\hat{\theta}}n^{1/2}(\hat{\theta} - \theta) + o_p(1),$$

where $D_\theta\{\hat{B}(t,\theta)^T\}$ is the first order derivative of $\hat{B}(t,\theta)^T$ w.r.t. θ gives an iid-decomposition of $n^{1/2}\{\hat{B}(t,\hat{\theta}) - B(t)\}$:

$$n^{1/2}\{\hat{B}(t,\hat{\theta}) - B(t)\} = n^{-1/2}\sum_{i=1}^{n}\epsilon_i^B(t,\theta) + o_p(1),$$

where

$$\epsilon_i^B(t,\theta) = \epsilon_i^B(t) + D_\theta(\hat{B}(t,\theta)^T)_{|\theta}\epsilon_i^\theta. \tag{A.1}$$

It thus follows that

$$n^{1/2}\{\hat{B}(t,\hat{\theta}) - B(t)\}$$

converges to a zero-mean Gaussian process with a variance that is consistently estimated by

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

The derivative $D_\theta(\hat{B}(t,\theta))_{|\hat{\theta}}$ can be calculated recursively as $\hat{B}(t,\hat{\theta})$ is constant between the observed death times. Denote the jump times by $\tau_1, \ldots, \tau_m$. Hence

$$\hat{B}(\tau_j,\theta) = \hat{B}(\tau_{j-1},\theta) + d\hat{B}(\tau_j,\theta)$$

which then also holds for the derivative. Since $\hat{B}(0,\theta) = 0$ and the derivative of the increment in the first jump time, $d\hat{B}(\tau_1,\theta)$, is easily calculated, then we have a recursive way of calculating the derivatives of $\hat{B}(\cdot,\theta)$. We now argue that the process $W_n^j(t,\theta)$, $j = 1,2$, converges in distribution as a process using arguments similar to what is done in Lin *and others* (2000, p. 726). It is seen from (2.3) that $B_j(t)$ can be written as a difference of two monotone functions. Let $\tilde{H}_{ijk}(s)$ be the limit in probability of $\mathcal{F}_{kj}(s,t)H_i(s,B_X(s-))$. Now, split $\tilde{H}_{ijk}(s)$ into its positive and negative parts, $\tilde{H}_{ijk}^+(s)$ and $\tilde{H}_{ijk}^-(s)$, and similarly with X_i, X_i^+ and X_i^-. Then $\int_0^t \tilde{H}_{ijk}(s)[dN_i(s) - X_i dB_X(s)]_{ik}$ can be written as a difference of two monotone functions, and then we follow the arguments of Lin *and others* (2000) (or use example 2.11.16 of van der Vaart and Wellner, 1996). Convergence in distribution for the process $W_n^j(t,\hat{\theta})$ also holds using the above Taylor expansion. It thus follows that

$$n^{1/2}\{\hat{B}(t,\hat{\theta}) - B(t)\}$$

converges to a zero-mean Gaussian process.

References

ANDERSEN, P. K., BORGAN, O., GILL, R. D. AND KEIDING, N. Statistical Models Based on Counting Processes. Berlin: Springer-Verlag.

ANDERSEN, P. K., GESKUS, R. B., DE WITTE, T. AND PUTTER, H. (2012). Competing risks in epidemiology: possibilities and pitfalls. *International Journal Epidemiology*, **41**, 861–70.

ANDERSEN, P. K. AND KEIDING, N. (2012). Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine* **31**, 1074–1088.

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.

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.

KJAERSGAARD, M. I. S. AND PARNER, E. T. (2016). Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics* **72**, 463–472.

LI, J., FINE, J. AND BROOKHART, A. (2014). Instrumental variable additive hazards models. *Biometrics*, **71**, 122–130.

MARTINUSSEN, T. AND SCHEIKE, T. H. (2006). *Dynamic Regression Models for Survival Data.* New York, NY: Springer.

MARTINUSSEN, T., VANSTEELANDT, S., TCHETGEN TCHETGEN, E. J. AND ZUCKER, D. M. (2017). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics*. doi:10.1111/biom.12699.

RICHARDSON, A., HUDGENS, M. G., FINE, J. AND BROOKHART, A. (2017). Nonparametric binary instrumental variable analysis of competing risks data. *Biostatistics* **18**, 48–61.

ROBINS, J. M. AND 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.

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.

ZHENG, C., DAI, R., HARI, P. N. AND ZHANG, M. J. (2017). Instrumental variable with competing risk model. *Statistics in Medicine* doi:10.1002/sim.7205.