Adjusting for bias introduced by instrumental variable estimation in the Cox proportional hazards model

PABLO MARTÍNEZ-CAMBLOR*

The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine, Dartmouth College, 7 Lebanon Street, Suite 309, Hinman Box 7251, Hanover, NH 03755, USA

Pablo.Martinez.Camblor@Dartmouth.edu

TODD MACKENZIE

Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, USA

DOUGLAS O. STAIGER

Department of Economics, Dartmouth College, The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, USA

PHILIP P. GOODNEY

Section of Vascular Surgery, Dartmouth-Hitchcock Medical Center, The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, USA

A. JAMES O'MALLEY

Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, USA

SUMMARY

Instrumental variable (IV) methods are widely used for estimating average treatment effects in the presence of unmeasured confounders. However, the capability of existing IV procedures, and most notably the two-stage residual inclusion (2SRI) algorithm recommended for use in non-linear contexts, to account for unmeasured confounders in the Cox proportional hazard model is unclear. We show that instrumenting an endogenous treatment induces an unmeasured covariate, referred to as an individual frailty in survival analysis parlance, which if not accounted for leads to bias. We propose a new procedure that augments 2SRI with an individual frailty and prove that it is consistent under certain conditions. The finite sample-size behavior is studied across a broad set of conditions via Monte Carlo simulations. Finally, the proposed methodology is used to estimate the average effect of carotid endarterectomy versus carotid stenting on the mortality of patients suffering from carotid artery disease. Results suggest that the 2SRI-frailty estimator generally reduces the bias of both point and interval estimators compared to traditional 2SRI.

Keywords: Individual frailty effect; Instrumental variable; Two-stage residual inclusion; Unmeasured confounding.

---

*To whom correspondence should be addressed.

1. INTRODUCTION

Ever since Cox's seminal paper in 1972 (Cox, 1972), the Cox proportional hazards models has become one of the most widely used statistical models due to the ubiquity in medicine of time-to-event outcomes subject to censoring. Although the Cox model was traditionally applied to analyze small data sets from randomized clinical trials (RCT), the increasing cost of RCTs and increasing availability of observational data have led to increased utilization of the Cox model in non-randomized settings. A clever observational study, perhaps in conjunction with a small RCT, can overcome the necessity for a large and highly expensive RCT. Registries containing the procedures and outcomes of all patients with a particular diagnosis (e.g., the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, https://seer.cancer.gov/about/), or that have undergone a certain procedure (e.g., the Vascular Quality Initiative, VQI, registry, http://www.vascularqualityinitiative.org/), are typically designed to measure all known risk factors and so yield high quality observational data. Nonetheless, unmeasured confounding is a concern whenever randomization is lacking. Provided its assumptions holds, the instrumental variable (IV) method (Angrist *and others*, 1996) allows causal interpretations and estimation from an analysis of observation data. They may also be applied to RCTs with imperfect compliance in which the objective is to estimate the effect of treatment received (the average treatment effect of the treated) rather than just intention-to-treat. However, even if a valid IV is available (e.g., treatment assignment in a RCT), an unresolved question in statistics and econometrics is the best way of using an IV with the Cox model. In this article, we seek to answer this question.

There are many real-world applications in which hazard ratio (HR) estimates from RCT comparisons and Cox models estimated on observational data differ even after controlling for detailed clinical information on patients, suggesting the need to develop IV procedures for Cox models. For example, large differences were observed in the HR of death for carotid endarterectomy (CEA) versus carotid stenting (CAS) estimated using the Cox model adjusted for covariates in the VQI and the results of recent RCT (Rosenfield *and others*, 2016; Brott *and others*, 2016). The VQI analysis finds an estimated HR of 0.693 [95% confidence interval (95% CI): (0.633; 0.760)], implying longer survival times under endarterectomy, whereas both clinical trials found almost null effects [approximate HR of 0.97, 95% CI (0.80; 1.19) and 0.91 (0.69; 1.20), respectively]. The profound difference in these results may be due to unmeasured confounders that have a strong effect on the procedure a VQI patient receives and their subsequent survival time. Therefore, the development of an IV procedure for survival time data that yields RCT-like estimates for a more general VQI population is highly desired.

In the context of structural linear equations, the two-stage least squares estimator, 2SLS, was introduced in econometrics in the 1950s by various authors (see Anderson, 2005 and references therein). IV procedures have subsequently been used to estimate causal effects from observational studies (Martens *and others*, 2006) in statistics, biomedicine, and many other applied research fields.

Conventional linear IV methods can produce substantial estimation bias when the true underlying models are non-linear (Terza *and others*, 2008). The direct non-linear generalization to 2SLS is the two-stage predictor substitution, 2SPS, procedure (Greene and Zhang, 2003). In the first stage, the relationship between the IV, W, and the exposure, X, is estimated by any consistent estimation technique. Then, the resulting fitted exposure status replaces the real observed exposure in the outcome model.

Alternatively, Hausman (1978) proposed the two-stage residual inclusion, 2SRI, or *control function estimator*. The 2SRI computes the expected exposure as for 2SPS but for the second-stage augments the target model with the residuals from the first stage. The first- and second-stage models can be linear or non-linear. Although there exists some debate about the relative performance of the 2SPS and 2SRI procedures, 2SRI is generally considered to have theoretical and practical advantages over 2SPS (Terza *and others*, 2008; O'Malley *and others*, 2011). Cai *and others* (2011) compared the bias of the 2SPS and 2SRI procedures at estimating the odds ratio among compliers under the principal stratification framework.
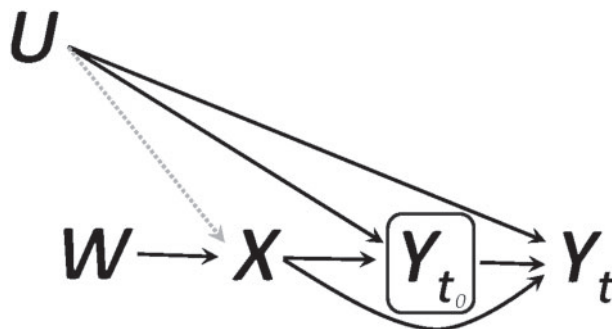
Fig. 1. DAG showing the unmeasured confounder U, treatment X, and the time-to-event outcome Y at t_0 and $t = t_0 + \epsilon$, where ϵ represents an arbitrarily small amount of time. The IV W is related with Y_{t_0} and Y_t only through X is an instrumental variable.

They found that 2SRI is asymptotically unbiased in the absence of unmeasured confounding but bias increasingly occurs with the magnitude of unmeasured confounding while 2SPS was always biased. (See Klungel *and others* (2015) for an extensive review of different IV procedures for both linear and non-linear contexts.)

The estimation of treatment effects by IV for time-to-event outcomes has received attention recently. The challenge is the presence of right censoring and non-linearity. Due to the presence of additive effects and an explicit error term, accelerated failure time models (Robins and Tsiatis, 1991) and additive hazard models (see, for instance, Li *and others*, 2015; Tchetgen Tchetgen *and others*, 2015, and references therein) are the most common frameworks. However, in biomedical and epidemiological research proportional hazard models are overwhelmingly used to analyze time-to-event outcomes.

Most practitioners are familiar with Cox's proportional hazard model (Cox, 1972) and interpreting research results in terms of HRs. An IV estimator of the HR has been proposed that assumes that the omitted covariates have an additive effect on the hazard (MacKenzie *and others*, 2014). A consistent estimator of the HR has been derived for the setting of a binary instrument (e.g., randomization), (MacKenzie *and others*, 2016). However, from a causal standpoint estimating the HR is problematic (Hernán, 2010). Because the hazard function is the instantaneous change in the survival probability, survival status Y_{t_0} just prior the instantaneous time period is conditioned on thereby inducing an association between treatment and any unmeasured predictor of survival, a phenomena known as collider bias. Intuitively, the problem is that even if X is independent of U at time 0, it is not independent at any time $t > 0$ because selection out of the sample depends on U. The Directed Acyclic Graph (DAG) (Pearl, 1995) in Figure 1 represents this scenario.

A consequence of collider bias is that if the path $U \rightarrow X$ exists then blocking it using an IV is not sufficient to identifying the causal estimate for the HR of X on Y_t, bringing the performance of IV methodology at estimating HRs into question. This is seen from evaluations of the bias of both the 2SPS and the 2SRI procedures on causal HR estimation under Weibull models (Cai *and others*, 2011) and in other contexts (Wan *and others*, 2015; Hernán, 2010), which found that no procedure consistently performed the best and that the bias of the IV estimator of the causal HR depended on the form of the unmeasured confounder, the magnitude of its effect, and the data structure.

Even if U does not have a causal effect on X, it is known that under the proportional hazard assumption the true effect of measured covariates on the HR is underestimated if U is ignored (Chamberlain, 1985; Gail *and others*, 1984). Therefore, in survival time models, omitted predictors unrelated with treatment assignment should be accounted for to avoid misspecification (Schmoor and Schumacher, 1997), a contrast

to a linear regression model where an explicit error term would absorb $\boldsymbol{U}$. Because omitted predictors of survival that are unrelated with any of the measured predictors manifest as frailties, the addition of an individual frailty (Wienke, 2010) appears to be a possible remedy.

The crux of the research in this article rests on the observation that if (i) a continuously valued exposure, X, is related to an IV W and a covariate U by the linear model, $X = \alpha_W \cdot W + \alpha_U \cdot U + \epsilon$ and (ii) the effects of X and U on a time-to-event, Y, satisfies the Cox model, $\mathcal{P}\{Y \geq t | X = x, U = u\} = \exp\{-\Lambda_0(t)\exp\{\beta_X \cdot x + \beta_U \cdot u\}\}$, then the conditional distribution of the time-to-event given the exposure, X, and $R = X - \alpha_W \cdot W$ satisfies the Cox model with frailty term: $\mathcal{P}\{Y \geq t | X = x, R = r\} = \mathcal{P}\{Y \geq t | X = x, \alpha_U \cdot U + \epsilon = r\} = \mathcal{P}\{Y \geq t | X = x, U = \alpha_U^{-1} \cdot (r - \epsilon)\} = \mathbb{E}_\epsilon[\mathcal{P}\{Y \geq t | X = x, U = \alpha_U^{-1} \cdot (r - \epsilon)\}] = \mathbb{E}_\epsilon[\exp\{-\Lambda(t) \cdot z \cdot \exp\{\beta_X \cdot x + \beta_U \cdot \alpha_U^{-1} \cdot r\}\}]$ where $z = \exp\{-\beta_U \cdot \alpha_U^{-1} \cdot \epsilon\}$ is the frailty term. The above expectation has the form of a Cox model with a frailty, suggesting that 2SRI will work in the context of Cox's model if the second stage is implemented with a frailty. Specifically, the procedure requires that the distribution specified for the frailty matches the distribution of the noise term in the linear model of X given W and U. This strategy assumes that $\alpha_U \neq 0$, the omitted covariates are related with treatment assignment, and that α_W is known thereby motivating evaluation of what happens when α_W is estimated (and hence fitted values of $\boldsymbol{X}$ are used in the second stage), when $\boldsymbol{X}$ is binary, and when the wrong frailty distribution is assumed.

The novel contributions of this article are to: (i) confirm the above reasoning that 2SRI applied to the Cox model induces a frailty, (ii) derive a 2SRI algorithm with an individual frailty term in the second-stage estimation, (iii) explore the performance of the procedure both when $\boldsymbol{X}$ is continuous and binary, and (iv) evaluate the robustness of the procedure to misspecification of the frailty distribution. The remainder of the article is structured as follows. The notation and assumed underlying model are defined in Section 2. Theoretical justification of the 2SRI-frailty procedure is provided in Section 3. Section 4 evaluates the operating characteristics of the new 2SRI-frailty procedure using Monte Carlo simulations. We consider the common situation in which we have a binary treatment with continuous IV and continuous measured and unmeasured covariates. We first consider the case where the exact same unmeasured predictors are present in both the treatment selection and the survival time models. We then consider the more realistic scenario where there are related but different unmeasured confounders in the survival and the treatment selection models, in which the first scenario is a special case. In Section 5, the proposed methodology is applied to an observational study where the goal is to estimate the average effect on mortality of the treatment [CEA versus carotid stenting (CAS)] received by patients suffering from carotid artery disease. The article concludes in Section 6. Additional simulations, tables, and figures are provided in supplementary material available at *Biostatistics* online. Besides, simulated data with the same structure to the considered real-world problem and the used code are available at https://github.com/pmcamblor/Inducement-frailty.

2. NOTATION AND MODELS

Conventionally, the right-censored framework assumes the sample $\{(t_i, \delta_i)\}_{i=1}^N$, where $t_i = \min\{y_i, c_i\}$ and $\delta_i = I_{(-\infty, c_i]}(y_i)$ ($I_A(x)$ stands for the usual indicator function, taking value 1 if $x \in A$ and 0 otherwise), and c_i and y_i are the censoring and the survival times for the ith subject ($1 \leq i \leq N$), respectively. Let $S_Y(\cdot) = \mathcal{P}\{Y > \cdot\}$ denote the survival function of Y. The Cox proportional hazard model is given by

$$d[\log S_Y(t | X, \boldsymbol{Z}, \boldsymbol{U})] = -\lambda_0(t) \cdot \exp\{\beta_X \cdot X + \boldsymbol{\beta}_Z^t \cdot \boldsymbol{Z} + \boldsymbol{\beta}_U^t \cdot \boldsymbol{U}\}, \tag{2.1}$$

where $\lambda_0(\cdot)$ is the baseline hazard function, X is a random variable representing the study treatment, $\boldsymbol{Z}$ is a random vector of exogenous measured predictors, and $\boldsymbol{U}$ is a random vector of unmeasured predictors.

The goal is to estimate the value of β_X; that is, the average change in the risk of an individual caused by a change in the received treatment.

We assume that an individual's received treatment is the result of the selective process depicted by the equation,

$$X = \alpha_0 + \boldsymbol{\alpha}_W^t \cdot \boldsymbol{W} + \boldsymbol{\alpha}_Z^t \cdot \boldsymbol{Z} + \boldsymbol{\alpha}_V^t \cdot \boldsymbol{V} + \epsilon, \tag{2.2}$$

with $\boldsymbol{W}$ a measured random vector and $\boldsymbol{V}$ a random vector of unmeasured variables that may be correlated with unmeasured variables in $\boldsymbol{U}$. The term ϵ is an independent random component representing the difference between the variable X and the predicted values obtained in the linear model. Note that the uncontrolled relationship between the unmeasured covariates affecting survival and treatment selection operate through the relationship between $\boldsymbol{U}$ and $\boldsymbol{V}$ and/or $\boldsymbol{Z}$. If $\boldsymbol{V} \perp\!\!\!\perp \boldsymbol{U}$, the survival model just contains unmeasured covariates, $\boldsymbol{U}$, that are not unmeasured confounders.

If the random vector $\boldsymbol{W}$ satisfies:

C_1. $\boldsymbol{W} \not\perp\!\!\!\perp (X|\boldsymbol{Z}, \boldsymbol{U}, \boldsymbol{V})$,
C_2. $\boldsymbol{W} \perp\!\!\!\perp (Y|X, \boldsymbol{Z}, \boldsymbol{U})$ (exclusion restriction assumption),
C_3. $\boldsymbol{W} \perp\!\!\!\perp (\boldsymbol{V}, \boldsymbol{U}|\boldsymbol{Z})$ (randomization assumption),

then $\boldsymbol{W}$ can be considered an *instrumental variable*. The strength of the instrument, that is, the ability of the instrument variable to capture the unmeasured confounding, is reflected through the strength of the relationship between $\boldsymbol{W}$ and X. In a clinical trial with perfect compliance and no loss to follow-up, random assignment is a perfect instrument. Assumptions C_1, C_2, and C_3 can be reformulated and combined with the stable unit treatment value assumption, SUTVA, and the monotonicity assumption (Hernán and Robins, 2006) between the instrument and the treatment to complete the conditions under which the IV $\boldsymbol{W}$ identifies the causal effect of X non-parametrically (i.e., without relying on (1) and (2)). Common instruments include prior institutional affinity for using a particular procedure, geographic region of residence, an individual's differential access to certain treatments, and an individual' s genes (aka Mendelian randomization, Thanasassoulis and O'Donnell, 2009).

3. Proposed methodology

The proposed methodology considers a standard first stage in which the relationship among the treatment, X, the instrument variable, $\boldsymbol{W}$, and the measured confounding, $\boldsymbol{Z}$, is estimated by any consistent method. We use simple linear regression models with standard least squares estimation but more flexible procedures could also be implemented. The first-stage procedure is:

Step 1. From the data, estimate the parameters to obtain the predicted values

$$\hat{X} = \hat{\alpha}_0 + \hat{\boldsymbol{\alpha}}_W^t \cdot \boldsymbol{W} + \hat{\boldsymbol{\alpha}}_Z^t \cdot \boldsymbol{Z}.$$

Then, compute the residuals, $\hat{R} = X - \hat{X}$.

It is worth noting that, due to C_3, under model (2) and assuming $\boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{V}$, $\hat{R}$ is a consistent estimator of $[\boldsymbol{\alpha}_V^t \cdot \boldsymbol{V} + \epsilon]$. If $\boldsymbol{Z} \not\perp\!\!\!\perp \boldsymbol{V}$, $\hat{R}$ estimates $[(\boldsymbol{\alpha}_Z - \hat{\boldsymbol{\alpha}}_Z^t) \cdot \boldsymbol{Z} + \boldsymbol{\alpha}_V^t \cdot \boldsymbol{V} + \epsilon]$. Therefore, in the last case, almost sure convergence of $\hat{\boldsymbol{\alpha}}_Z$ to $\boldsymbol{\alpha}_Z$ is not guaranteed. The residual $\hat{R}$ contains all the information about the unmeasured vector $\boldsymbol{U}$ related with the treatment assignment and unrelated with $\boldsymbol{Z}$; that is, all the

available information about unmeasured confounding is contained in the residuals provided by the model (1). However, $\hat{R}$ also contains *white noise* pertaining to idiosyncratic or purely random factors affecting an individual's treatment selection, which corresponds to the difference between the unmeasured covariates in the two models, V and U, and to the independent random term ϵ in model (2). We conjecture that the component of $\hat{R}$ due to white noise can be handled by specifying an individual *frailty* in the outcome model in order to allow $\hat{R}$ to more fully be able to perform its intended task of controlling for U. The proposed second stage is:

Step 2. Estimate the Cox proportional hazard regression with individual frailty:

$$d[\log \hat{S}_Y(t|X, \mathbf{Z}, \hat{R}, F)] = -\phi \cdot \hat{\lambda}_0^*(t) \cdot \exp\{\hat{\beta}_X^{\mathrm{IV}} \cdot X + \hat{\boldsymbol{\beta}}_{\mathbf{Z}}^{*t} \cdot \mathbf{Z} + \hat{\beta}_{\hat{R}} \cdot \hat{R}\},$$

where $\phi = \exp\{F\}$ is the individual frailty term. A distribution should be specified for F (e.g., log-Gaussian, Gamma, ...). The parameter estimate of β_X that results from this procedure is denoted $\hat{\beta}_X^{\mathrm{IV}}$.

Standard algorithms for estimating Cox models with frailties may be used to implement the procedure. For example, Therneau *and others* (2003) proved that maximum likelihood estimation for the Cox model with a Gamma frailty can be accomplished using a general penalized routine, and Ripatti and Palmgren (2000) derived a similar argument for the Cox model with a Gaussian frailty.

3.1. *Asymptotic properties of $\hat{\beta}_X^{IV}$*

By adapting the convergence results for Cox's partial likelihood (see, for instance, Theorem 5.3.1 in Kalbfleisch and Prentice, 2002) we obtain the following convergence results.

THEOREM Assume the causal models

$$d[\log S_Y(t|X, Z, U)] = -\lambda_0(t) \cdot \exp\{\beta_X \cdot X + \beta_Z \cdot Z + \beta_U \cdot U\}, \tag{3.1}$$

$$X = \alpha_0 + \alpha_W \cdot W + \alpha_Z \cdot Z + \alpha_V \cdot V + \epsilon, \tag{3.2}$$

with the random variables X (the treatment), Z (measured covariate), and U (unmeasured covariate). In addition, assume that the pair (U, V) is normally distributed, that ϵ is independently normally distributed random noise, and that W (the instrument) is a random variable satisfying C_1–C_3. Then, if the censoring time, C, satisfies $C \perp\!\!\!\perp (Y|X, Z, U, W)$, we have the weak convergence,

$$\sqrt{n} \cdot \left\{\hat{\beta}_X^{\mathrm{IV}} - \beta_X\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_X^{\mathrm{IV}}\right), \tag{3.3}$$

where $\left(\sigma_X^{\mathrm{IV}}\right)^2 = n \cdot \mathbb{V}[\hat{\beta}_X^{\mathrm{IV}}]$ ($\mathbb{V}$ stands for the variance operator) can be consistently estimated from the survival and at-risk counting processes.

Proof. In absence of the frailty term, it is well known that the estimator of $\boldsymbol{\beta}$ that maximizes the Cox-model partial likelihood function obeys the asymptotic law

$$\sqrt{n} \cdot \{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\} \xrightarrow{\mathcal{L}} \mathcal{N}_p\left(0, \boldsymbol{I}^{-1}(\boldsymbol{\beta})\right), \tag{3.4}$$

where $\boldsymbol{\beta} = \{\beta_0^*, \beta_X, \beta_Z^*\}$, and $\boldsymbol{I}(\boldsymbol{\beta})$ is the $p \times p$ (p is the dimension of the vector $\boldsymbol{\beta}$) information matrix (Kalbfleisch and Prentice, 2002), which can be consistently estimated by $\boldsymbol{I}(\hat{\boldsymbol{\beta}})$.

In the presence of an individual frailty, different estimation methods have been proposed. Particularly, Ripatti and Palmgren (2000) and Vaida and Xu (2000) studied the case of multiplicative log-normal distributed frailties. The proposed methodology obtains maximum-likelihood estimates of the regression parameters, the variance components and the baseline hazard, as well as empirical Bayes estimates of the random effects (frailties). Therefore, it suffices to prove that the second stage of our proposed 2SRI-frailty procedure is a Cox proportional hazard model with a Gaussian frailty term in which the treatment variable coefficient is unchanged from the original model (i.e., β_X).

We can assume, without loss of generality, that both U and V have standard normal distributions with $\mathbb{E}[U \cdot V] = \rho$. The case $\rho = 0$ implies that there is no unmeasured confounding and it is not considered here. Therefore, if $S = V - \rho \cdot U$, (3.2) can be written as

$$
\begin{aligned}
X &= \alpha_0 + \alpha_W \cdot W + \alpha_Z \cdot Z + \alpha_V \cdot (\rho \cdot U + S) + \epsilon \\
&= \alpha_0 + \alpha_W \cdot W + \alpha_Z \cdot Z + \alpha_U \cdot U + \epsilon^*,
\end{aligned}
\tag{3.5}
$$

where $\epsilon^* \; (= \alpha_V \cdot S + \epsilon)$ is a normal random variable independent of U and $\alpha_U = \alpha_V / \rho$.

Condition C_1 implies that if $(\alpha_W - \hat{\alpha}_W) = \mathcal{O}(n^{-1/2})$, then given the causal equation in (3.5) and conditions $C_1 - C_3$, Step 1 yields

$$
\hat{R} = (\alpha_0 - \hat{\alpha}_0) + (\alpha_Z - \hat{\alpha}_Z) \cdot Z + \alpha_U \cdot U + \epsilon + \mathcal{O}(n^{-1/2}).
$$

Hence, $\alpha_U \cdot U = \hat{R} - \{(\alpha_0 - \hat{\alpha}_0) + (\alpha_Z - \hat{\alpha}_Z) \cdot Z + \epsilon + \mathcal{O}(n^{-1/2})\}$. Assuming $\alpha_U \neq 0$ ($\alpha_U = 0$ also implies there is no unmeasured confounding), the linear part of the model in (3.1) can be rewritten as,

$$
\mathfrak{L} = \beta_X \cdot X + \beta_Z \cdot Z + \beta_U \cdot U = \beta_0^* + \beta_X \cdot X + \beta_Z^* \cdot Z + \beta_{\hat{R}} \cdot \hat{R} + F,
$$

with $\beta_{\hat{R}} = \beta_U / \alpha_U$, and where, in this case, $\beta_0^* = -\beta_{\hat{R}} \cdot (\alpha_0 - \hat{\alpha}_0)$, $\beta_Z^* = \beta_Z - \beta_{\hat{R}} \cdot (\alpha_Z - \hat{\alpha}_Z)$ and $F = \beta_{\hat{R}} \cdot \epsilon^* + \mathcal{O}(n^{-1/2})$. Therefore, F is asymptotically independent and normally distributed due to ϵ^* being independent and normally distributed. Hence, if $\lambda_0^*(t) = \lambda_0(t) \cdot \exp(\beta_0^*)$, the survival model is given by

$$
d[\log S_Y(t|X, Z, U)] = -\phi \cdot \lambda_0^*(t) \cdot \exp\{\beta_X \cdot X + \beta_Z^* \cdot Z + \beta_{\hat{R}} \cdot \hat{R}\},
$$

which has the form of a Cox proportional hazards model with frailty $\phi = \exp\{F\}$. Therefore, if $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_0^*, \hat{\beta}_X^{\mathrm{IV}}, \hat{\beta}_Z^*\}$ is the estimator resulting from step S_2, invoking the censoring time assumptions and using the convergence of the partial maximum-likelihood method given a consistent method of estimating the product of the baseline risk and the frailty (Ripatti and Palmgren, 2000; Vaida and Xu, 2000), it follows that

$$
\sqrt{n} \cdot \left\{ \hat{\beta}_X^{\mathrm{IV}} - \beta_X \right\} \xrightarrow{\mathcal{L}}_n \mathcal{N}\left(0, \sigma_X^{\mathrm{IV}}\right),
\tag{3.6}
$$

with $(\sigma_X^{\mathrm{IV}})^2$ the component in the matrix $\boldsymbol{I}^{-1}(\boldsymbol{\beta})$ corresponding to β_X. □

REMARK 1 Normality of the pair (U, V) is sufficient but not necessary condition to guarantee normality of ϵ^*.

REMARK 2 The estimator $\hat{\beta}_Z^*$ is unbiased for the parameter $\beta_Z^* = \beta_Z - \beta_{\hat{R}} \cdot (\alpha_Z - \hat{\alpha}_Z)$. Notice that, under the current assumption, we just can guarantee the convergence $|\beta_Z^* - \beta_Z| \to_n 0$ if $Z \perp\!\!\!\perp V$. Therefore, in

general, $\hat{\beta}_Z^*$ should not be interpreted as an estimator of the measured covariates effects. Empirical results and additional discussion are reported in Table S5 of supplementary material available at *Biostatistics* online.

4. MONTE CARLO SIMULATION STUDY

To evaluate the behavior of the proposed methodology in finite samples, we conducted a range of Monte Carlo simulations. We found that, beyond the expected effect on precision of estimation, neither the baseline shape of the survival times nor the censorship distribution have any meaningful effect on the observed results. Therefore, we only show results when the baseline survival times follow a Weibull distribution with shape parameter two and scale parameters coherent with the proportional hazard assumption; that is, the scale parameter equates to $\exp\{\beta_X \cdot (X - \bar{X}) + Z + \beta_U \cdot U\}$. Here β_X is the target. We subtract $\bar{X}$ from X, as opposed to drawing X from a distribution with mean 0, to ensure that we obtain realistic survival times with binary X without needing to alter the intercept. Z and U denote the measured and unmeasured confounders, respectively. The measured covariate (Z) follows a standard normal distribution, $\mathcal{N}_{0,1}$, while an independent standard normal and a centered Gamma with parameter one distributions are considered for the unmeasured covariate (U). Censoring was independently drawn from a Weibull distribution such that the expected censorship was 20%. Treatment assignment is based on the linear equation $X^* = \alpha_W \cdot W + Z + V + \epsilon$, where W is the instrument, V is the unmeasured covariate, and ϵ is random noise. We set $X = X^*$ for a continuous exposure and $X = I(X^* \geq 0)$ for a binary exposure. All of W and ϵ are drawn from independent standard-normal distributions. For V, we have considered the same distributions as for U. Notice that, after fixing the rest of the parameters, decreasing α_W yields an IV of lesser quality. Sample size was fixed at $N = 500$.

4.1. *All omitted covariation is unmeasured confounding*

We first suppose $X = X^*$, $U = V$ and U follows a standard normal or a centered Gamma distribution, and ϵ follows an independent standard normal distribution. That is, the endogenous variable is continuous and there are no unmeasured predictors of survival time unrelated with treatment selection (i.e., while the true Cox model of the survival times may include a shared covariate with the treatment selection process, it does not include a frailty).

Figure 2 (top) shows the median of the bias observed in 2000 Monte Carlo iterations for a weaker, $\alpha_W = 1$, instrument (results for a stronger instrument, $\alpha_W = 2$, are provided in Figure S1 of supplementary material available at *Biostatistics* online). The Naive Cox model, which ignores the presence of omitted covariates always performed poorly. The proposed algorithm, 2SRI-frailty (2SRI-F), reduced bias the most when U had strong effects and is normally distributed. When the presence of unmeasured confounding was weaker (smaller β_U-values), and when there was no effect of the treatment, $\beta_X = 0$, both 2SRI and 2SRI-F obtained similar results. When the true and assumed frailty distribution differs, the 2SRI-F bias reduction is smaller but, crucially, observed biases were always smaller than under the 2SRI.

As is well known, two-stage IV methods lead to estimators ($\hat{\beta}_X^{IV}$) with the degree of variance inflation depending on the strength of the IV. The standard asymptotic variance obtained from the second-stage Cox regression models of the respective procedures, ignoring the first-stage variability, under-estimates the true value (computed from the Monte Carlo simulations). This difference disappears when we compute the trimmed variability (95% level). It appears that the 2SRI-F's assumption of the distribution of the frailty is helpful in addressing bias and does not suffer from the excessive and inappropriate gains in variability that accompany many procedures with parametric components. Complete information is reported in Table S1 of supplementary material available at *Biostatistics* online.
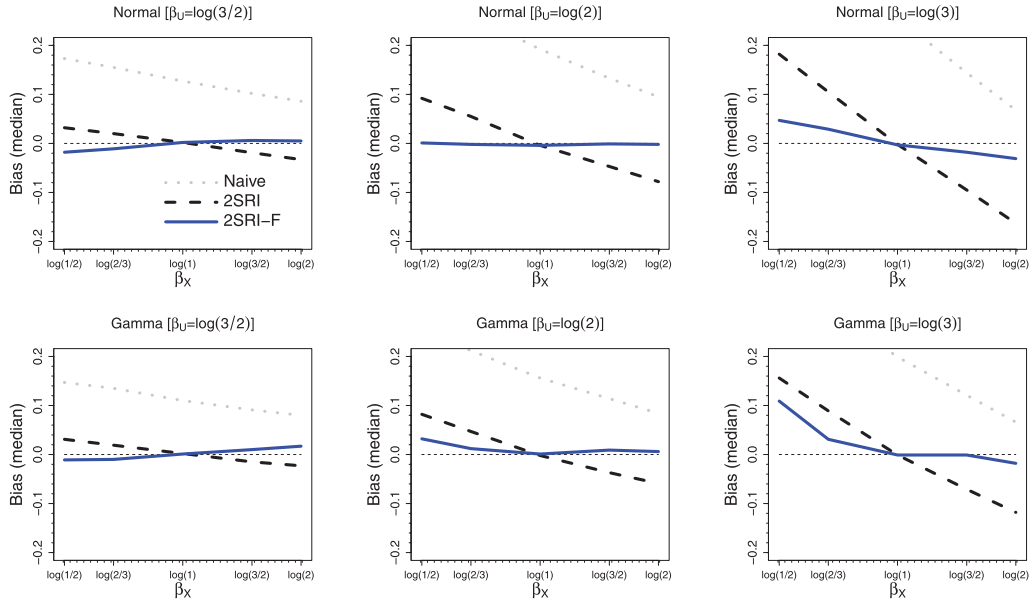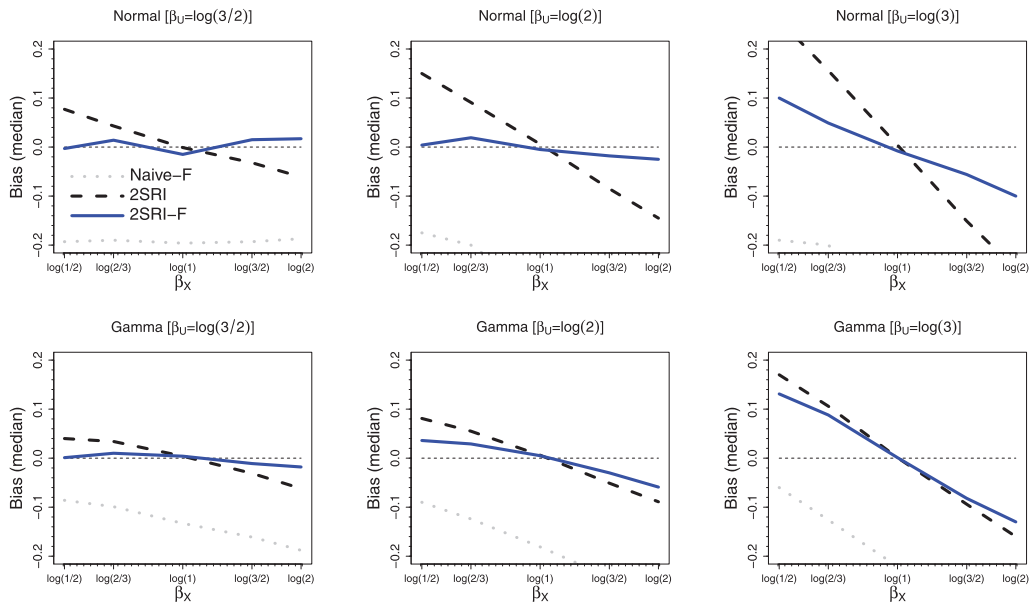
Fig. 2. Median bias for Weibull($\exp\{\beta_X \cdot X + Z + \beta_U \cdot U\}, 2$) survival times and $X^* = Z + W + V + \epsilon$, where Z, W, and ϵ follow independent standard normal distributions and centered Gamma(1,1) and standard normal distributions are considered for U and V. Top-segment, $V = U$ and $X = X^*$. Bottom-segment, $X = I\{X^* > 0\}$ and $\mathbb{C}(U, V) = 1/2$. Gray-dots, naive Cox model (omitted covariate is ignored); black-dashed 2SRI, procedure; continuous, 2SRI-F (2SRI plus Gaussian frailty). A continuous black thin dashed line stands for the zero-bias situation.

4.2. *Omitted covariation is a mixture of unmeasured confounding and a pure individual frailty*

We consider the case in which U and V follow standard normal distributions ($\alpha_W = 1$) and the correlation between them, $\mathbb{C}(U, V)$, is ρ. Note that $\rho = 0$ implies that the survival model does not contain unmeasured confounders, only unmeasured covariates, while $\rho = \pm 1$ implies that all omitted covariation manifests as unmeasured confounding (i.e., is also related to treatment assignment). In the $\rho = 0$ case, it is reassuring that the 2SRI and the 2SRI-F procedures perform nearly as well as the naive Cox regression model, the true model in this scenario. The advantage of using 2SRI-F versus 2SRI was larger for ρ close to zero, which makes sense as the pure frailty variation is at its maximum, whereas at values close to ± 1 the frailty has all but disappeared. Full results are reported in Figure S2 of supplementary material available at *Biostatistics* online.

4.3. *Omitted covariation is mixture distributed*

In order to check the robustness of the approach that assumes a Gaussian frailty with respect to deviations from normality of the unmeasured covariates distribution, we study the case where the unmeasured covariates, U and V, are not normally distributed. In particular, we considered the following scenarios:

M-I. $U = \chi_a \cdot \gamma_1 + (1 - \chi_a) \cdot \gamma_2$; $V = \chi_a \cdot \gamma_1 + (1 - \chi_a) \cdot \gamma_3$.
M-II. $U = \chi_a \cdot \gamma_1 + (1 - \chi_a) \cdot \eta_2$; $V = \chi_a \cdot \gamma_1 + (1 - \chi_a) \cdot \eta_3$.
M-III. $U = \chi_a \cdot \tau_1 + (1 - \chi_a) \cdot \eta_2$; $V = \chi_a \cdot \tau_1 + (1 - \chi_a) \cdot \eta_3$.
M-IV. $U = \chi_a \cdot \eta_1 + (1 - \chi_a) \cdot \gamma_2$; $V = \chi_a \cdot \eta_1 + (1 - \chi_a) \cdot \gamma_3$.

where χ_a is an independent random variable which takes the value 1 with probability a and 0 otherwise; γ_1, γ_2, and γ_3 follow independent centered Gamma(1,1); η_1, η_2, and η_3 follow independent standard normal distribution; and τ_1 follows a Student T distribution with four degrees of freedom. Notice that a determines both the distribution and the relationship between U and V with $\mathbb{C}(U, V) = a$. In the extremes, $a = 0$ when there is no unmeasured confounding (just unmeasured covariates) and $a = 1$ when $U = V$. Figure 3 shows the median of the bias of 2SRI-F algorithm observed in 2000 Monte Carlo iterations under the previous models with $a \in [0, 1]$. Results suggest minimal impact of the unmeasured covariates distribution. The proposed methodology fails to completely remove the bias when there are strong unmeasured confounding effects but always produces better results than the standard 2SRI method.

4.4. *Non-linear treatment selection model: binary exposure*

The second scenario supposes $X = I_{[0,\infty)}(X^*)$, a binary exposure. Because other values of ρ produced similar results, we only report results for which the correlation between U and V was fixed at $\rho = 1/2$.

Figure 2 (Binary treatment) depicts the median bias over 2000 Monte Carlo iterations when β_X was directly estimated from a Cox regression with Gaussian frailty, (due to the presence of an unmeasured covariate unrelated with treatment assignment, this model is the correct model in the absence of unmeasured confounding), the 2SRI procedure, and 2SRI-F. Not surprisingly, ignoring the presence of the frailty and estimating a standard Cox regression model results in larger bias. The 2SRI algorithm helps us to control just the part of the bias related with the treatment assignment but also fails to handle the frailty and its performance suffers as a result. However, the proposed 2SRI-F methodology produces some bias when the presence of unmeasured confounding is strong although, crucially, the results are always better than those obtained by 2RSI. These results reveal that there is clear benefit to be gained from using 2SRI-F as an IV procedure for time-to-event data.

The estimated standard deviation of the Monte Carlo simulations and the mean of standard deviation reported by the Cox regression model in the second stage, ignoring the first-stage variability, were similar
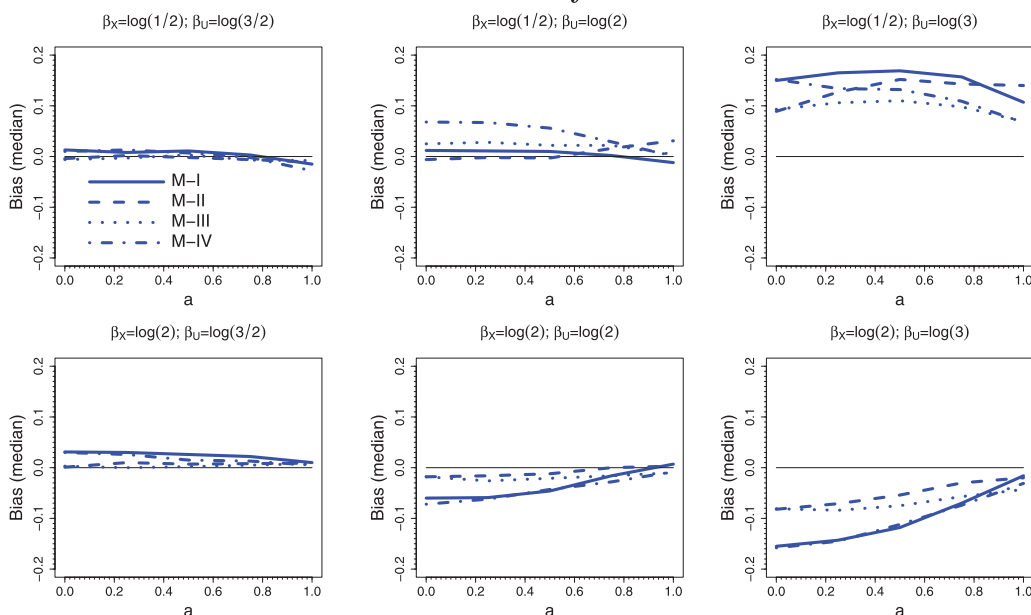
Non-normal frailty distributions

Fig. 3. Median bias for Weibull($\exp\{\beta_X \cdot X + Z + \beta_U \cdot U\}$, 2) survival times and $X = Z + W + U + \epsilon$, where Z and ϵ follow standard normal distributions, U and V are as previously specified, and estimation is performed by the 2SRI-F algorithm against parameter determining the mixture model, a. A continuous black thin line stands for the zero-bias situation.

(complete results reported in the Table S2 of supplementary material available at *Biostatistics* online). These results are consistent with the results for continuous exposures; the variance inflation under the 2SRI-F procedure exceeds that under 2SRI which in-turn exceeds that under Naive-F.

5. Real-world application: the VQI data set

We apply 2SRI-F to nationwide data from the VQI (http://www.vascularqualityinitiative.org) on patients diagnosed with carotid artery disease (carotid stenosis). These data contain comprehensive information on all patients suffering from carotid stenosis and is continually updated over time to facilitate determination of the best procedure or treatment approach to use on average and to determine which types of patients benefit the most from each procedure. However, the data are exposed to a plethora of selection biases raising concerns that naive analyses will yield biased results. Because the outcomes of most interest are events such as stroke or death that can occur at any point during follow-up, these data are ideal for application of the 2SRI-F procedure.

We employed 2SRI-F to estimate the comparative effectiveness of CEA versus carotid stenting (CAS), the two surgical procedures used to intervene on patients with carotid stenosis. The data consist of 28 712 patients who received CEA and 8117 who received CAS, between 15 and 89 years of age, over 2003–2015. During follow-up, there were 3955 and 807 deaths in the CEA and CAS groups, respectively.

Figure 4 (top) depicts two Kaplan–Meier survival curves: crude (left) and adjusted by patient age, gender, ethnicity, race, type of surgery (elective/not elective), symptoms (yes/no), hypertension diabetes, smoking history (yes/no), positive stress test, coronary disease, heart failure, diabetes, chronic obstructive
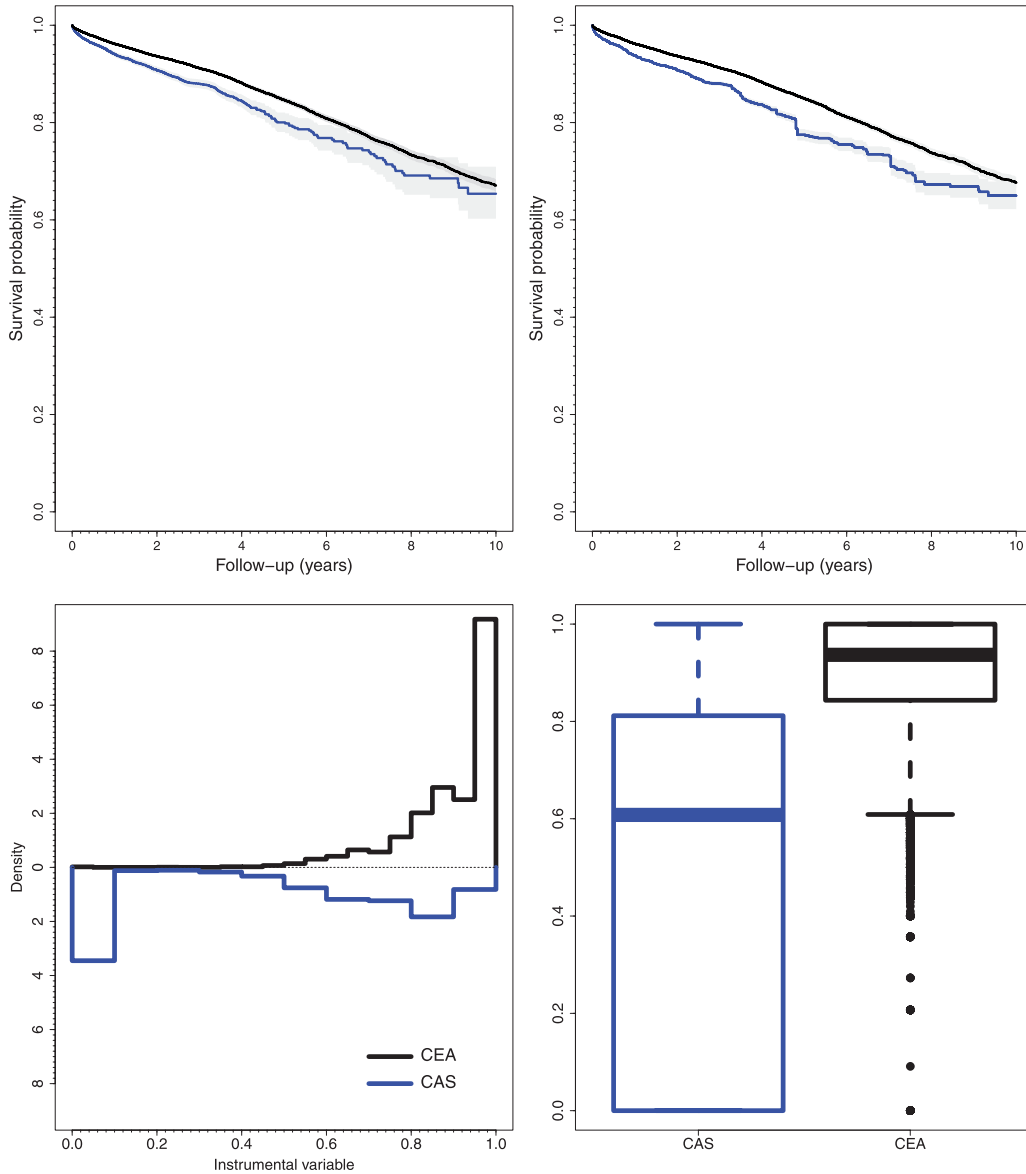
Fig. 4. Top-segment, estimated Kaplan–Meier curves for both the CEA and CAS groups with 95% confidence bands: crude (left) and adjusting by the covariates described in Table S3 of supplementary material available at *Biostatistics* online and using the weighted inverse propensity procedure (right). Bottom-segment, (left) histograms for the instrument variable, W, and (right) boxplot by received treatment.

pulmonary disease (COPD), renal insufficiency, dialysis status (haemodialysis, HD, yes/no), prior ipsilateral CEA, and the use of antiplatelet therapy, beta-blocker, and statin use by using the weighted inverse propensity procedure (MacKenzie *and others*, 2012). The crude HR comparing CEA to CAS was 0.719 [95% CI of (0.666; 0.777)]. Adjusted HR was 0.693 (0.633; 0.760). The last HR is slightly modified when a frailty term is included: 0.685 (0.624; 0.753) and 0.676 (0.613; 0.745) for the Gaussian and Gamma cases, respectively.

Table 1. *HRs and 95% CIs by estimation method*

	HR (95% CI)
Crude	0.719 (0.666;0.777)
Cox model (Naive)	0.693 (0.633;0.760)
Naive + frailty (Gaussian)	0.685 (0.624;0.753)
Naive + frailty (Gamma)	0.676 (0.613;0.745)
2SRI	0.901 (0.737;1.100)
2SRI-F (Gaussian)	0.887 (0.724;1.087)
2SRI-F (Gamma)	0.882 (0.716;1.086)

The IV we use is the center level frequency of CEA versus CAS procedures over the 12 months prior to the current patient; that is, the total number of CEA divided by total of CEA and CAS procedures in the 12 months prior to the current patient. This variable is justified as an instrument due to: (i) hospitals that performed a high relative amount of a certain procedure in the past are likely to keep doing so; (ii) there should be no effect of the relative frequency of CEA versus CAS on a patient's outcome except through its effect on treatment choice for that patient; and (iii) we know of no factors that would influence both this frequency and a patient's outcome. Reasons (ii) and (iii) are contingent on adjusting for the total number of CEA and CAS procedures performed at the center over the past 12 months.

On the VQI data the IV is highly associated with treatment choice. The probability that a randomly selected subject undergoing CEA has a larger value of the instrument than a randomly selected subject undergoing CAS was 0.881 [95% CI of (0.876; 0.885)]. This IV was unrelated with all of the measured confounders suggesting anecdotally that it may also be uncorrelated with any unmeasured confounders. Hence, it is reasonable to assume that the relationship of the instrument with mortality is solely due to its relationship with the treatment. Figure 4 (bottom; left side) shows the histogram of W in both the CEA and CAS groups and (at right) shows the boxplot for the IV by surgical procedure.

The treatment effect almost disappears when 2SRI is applied on the data set: HR of 0.901 with a 95% CI (0.737; 1.100). When the proposed 2SRI-F algorithm is used we obtain, a HR of 0.887 with associated 95% CI (0.724; 1.087) and in the first-stage equation $\hat{\beta}_X^{IV} = -0.120$. Similar results were also obtained under a Gamma-distributed frailty instead of the Gaussian frailty. Table 1 shows the HRs and 95% CIs.

6. DISCUSSION

IV methods are often used to account for unmeasured confounders. Although these methods have been widely studied in a variety of situations, their suitability for estimating Cox proportional hazard models is unclear. It is well known that, in this case, model misspecification can produce bias even when the omitted variables are unrelated with the treatment assignment (Aalen *and others*, 2015); that is, when they only affect the survival time. As suggested by our structural argument in the Introduction, an individual frailty appears able to solve this problem. We showed that the presence of idiosyncratic variation affecting treatment selection may induce a frailty in the instrumented survival time model even if there is no frailty in the original survival model. In practice, the most likely scenario is that both a true frailty and unmeasured confounding factors affect survival. For these reasons, we were motivated to develop and evaluate an IV procedure that, in the second stage, incorporates a frailty term.

Because the Cox model is non-linear, our base strategy for dealing with unmeasured confounders was to use the 2SRI algorithm adapted to the Cox model. As noted above, even when the true survival model does not contain omitted covariates, the 2SRI procedure induces a frailty in the second-stage Cox regression

model from the inclusion of the residuals computed in the first stage. To account for this phenomenon, we added an individual frailty in the second-stage (instrumented) statistical model. Under standard reliability conditions, we proved the asymptotic consistency of the estimator defined under our 2SRI-F procedure for the case when the univariate frailty distribution is correctly assumed to be Gaussian.

Monte Carlo simulations suggested that the proposed methodology (2SRI-F) produces an important bias reduction and is superior to the 2SRI, particularly in the presence of an individual frailty due to unmeasured covariates unrelated with treatment assignment. A very important finding is that the bias of the 2SRI-F method was always close to zero even when the residuals from the treatment selection equation were not normally distributed. The Gaussian distribution can be directly justified when each individual frailty is the sum of different independent sources of heterogeneity. Furthermore, because the procedure with the Gaussian frailty was surprisingly robust to erroneously assumed frailty distributions, we recommend using a Gaussian frailty.

A controversial feature of our procedure is the inclusion of the individual frailty term. Although there exists a vast literature for the case where the frailty is common to a group of individuals (shared frailty), the number of references dealing with individual frailties is minimal. Consistency properties of the common estimation algorithm for Cox models with frailties were proved previously (Nielsen *and others*, 1992; Murphy, 1995). We adapted these theoretical results in deriving the consistency results presented herein. By specifying a distribution for its values, the individual frailty accounts for the omitted covariates unrelated with the treatment assignment—the extra variability introduced in the survival model from the first stage of the algorithm (ϵ) and the *portion* of V independent of U, freeing the augmented first-stage residual (the control function) to account for unmeasured confounding. The resulting procedure estimates the average treatment effect conditional on the unmeasured confounder and the frailty (Yashin *and others*, 2001). Because in practice specification of the distribution of the frailty can be arbitrary, the observed results should be handled with caution (Hougaard, 1995) and they should be supported by sensitivity analyses considering different frailty distributions. In the VQI application, the empirical results were found to only have a slight dependence on the distribution of the frailty. This is a key finding that justifies the use of the 2SRI-F procedure and represents a major advance in the IV literature for the analysis of time-to-event outcomes in observational settings.

In the real-world application, a small but significant (at the 0.05 level) effect of the treatment is detected when the presence of omitted covariates on the Cox model is ignored. This effect almost disappears under 2SRI. When the 2SRI-F method is used, the estimated effect of CEA over CAS is slightly larger suggesting that the effect of the procedure a patient receives is underestimated when unmeasured confounding is ignored (Chamberlain, 1985; Gail *and others*, 1984). It is worth noting that different patient enrollment rates were observed by treatment, while CAS patient censorship is constant across the follow-up time-period, most of the CEA patients have follow-up above 2 years with an important number censored between the second and the fourth years. To the extent these differences are caused by an unmeasured confounder, this can introduce additional bias in the standard naive estimates and strongly motivates the use of an adequate IV procedure.

The method we developed conditions on all omitted covariates and assumes they have multiplicative effects on the hazard function under the Cox model, unlike recently developed methods that make unusual additive hazard assumptions in order to more simply account for unmeasured confounding (MacKenzie *and others*, 2014; Tchetgen Tchetgen *and others*, 2015). Therefore, we anticipate that our proposed and proven procedure will hold extensive appeal and be widely used in practice.

While it is encouraging that our null results cohere with those of recent RCTs (Rosenfield *and others*, 2016; Brott *and others*, 2016), thereby overcoming the unfavorable CAS results of the non-IV observational study analyses, an effect close to 0 makes it difficult to distinguish our proposed IV procedure from the incumbent two-stage residual inclusion method. However, when the true effect is 0 (HR of 1),

the bias from ignoring the frailty is 0 due to the fact that omitting a frailty shrinks the true coefficient towards 0. Therefore, the differences between the 2SRI-F and the standard 2SRI procedure for the Cox model are expected to converge to 0 as the true treatment effect approaches 0. In this sense, the lack of extensive differences between the various 2SRI (frailty and standard) procedures is a real-data endorsement that our proposed 2SRI-F procedure for the Cox model performs as it should by not rejecting the null hypothesis when the RCT results and the 2SRI results suggest that the true effect is close to 0.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

FUNDING

REFERENCES

AALEN, O. O., COOK, R. J. AND RØYSLAND, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis* **21**, 579–593.

ANDERSON, T. W. (2005). Origins of the limited information maximum likelihood and two-stage least squares estimators. *Journal of Econometrics* **127**, 1–16.

ANGRIST, J. D., IMBENS, G. W. AND RUBIN, D. G. (1996). Identification of causal effects using instrumental variables identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.

BROTT, T. G., HOWARD, G., ROUBIN, G. S., MESCHIA, J. F., MACKEY, A., BROOKS, W. *and others*. (2016). Long-term results of stenting versus endarterectomy for carotid-artery stenosis. *New England Journal of Medicine* **374**, 1021–1031.

CAI, B., SMALL, D. S. AND HAVE, T. R. T. (2011). Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Statistics in Medicine* **30**, 1809–1824.

CHAMBERLAIN, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J. J. and Singer, B. S. (editors), *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

COX, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistics Society (B)* **34**, 187–220.

GAIL, M. H., WIEAND, S. AND PIANTADOSI, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431–444.

GREENE, W. H. AND ZHANG, G. (2003). *Econometric Analysis*. New Jersey, USA: Prentice Hall.

HAUSMAN, J. A. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251–1271.

HERNÁN, M. A. (2010). The hazards of hazard ratios. *Epidemiology* **21**, 13–15.

HERNÁN, M. A. AND ROBINS, J. M. (2006). Instruments for causal inference: an epidemiologist's dream? *Epidemiology* **17**, 360–372.

HOUGAARD, P. (1995). Frailty models for survival data. *Lifetime Data Analysis* **1**, 255–273.

KALBFLEISCH, J. D. AND PRENTICE, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

KLUNGEL, U., DE BOER A., BELITSER, S. V., GROENWOLD, R. H., AND ROES, K. C. (2015). Instrumental variable analysis in epidemiologic studies: an overview of the estimation methods. *Pharmaceutica Analytica Acta* **6**, 1–9.

LI, J., FINE, J. AND BROOKHART, A. (2015). Instrumental variable additive hazards models. *Biometrics* **71**, 122–130.

MACKENZIE, T. A., BROWN, J. R., LIKOSKY, D. S., WU, Y. X. AND GRUNKEMEIER, G. L. (2012). Review of case-mix corrected survival curves. *The Annals of Thoracic Surgery* **93**, 1416–1425.

MACKENZIE, T. A., LØBERG, M. AND O'MALLEY, A. J. (2016). Patient centered hazard ratio estimation using principal stratification weights: application to the norccap randomized trial of colorectal cancer screening. *Observational Studies* **2**, 29–50.

MACKENZIE, T. A., TOSTESON, T. D., MORDEN, N. E., STUKEL, T. A. AND O'MALLEY, A. J. (2014). Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding. *Health Services and Outcomes Research Methodology* **14**, 54–68.

MARTENS, E. P., PESTMAN, W. R., DE BOER, A., BELITSER, S. V. AND KLUNGEL, O. H. (2006). Instrumental variables: application and limitations. *Epidemiology* **17**, 261–267.

MURPHY, S. A. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics* **23**, 182–198.

NIELSEN, G. G., GILL, R. D., ANDERSEN, P. K. AND SØRENSEN, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.

O'MALLEY, A. J., FRANK, R. G. AND NORMAND, S.-L. T. (2011). Estimating cost-offsets of new medications: use of new antipsychotics and mental health costs for schizophrenia. *Statistics in Medicine* **30**, 1971–1988.

PEARL, J.. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669.

RIPATTI, S. AND PALMGREN, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022.

ROBINS, J. M. AND TSIATIS, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods* **20**, 2609–2631.

ROSENFIELD, K., MATSUMURA, J. S., CHATURVEDI, S., RILES, T., ANSEL, G. M., METZGER, D. C. *and others*. (2016). Randomized trial of stent versus surgery for asymptomatic carotid stenosis. *New England Journal of Medicine* **374**, 1011–1020.

SCHMOOR, C. AND SCHUMACHER, M. (1997). Effects of covariate omission and categorization when analysing randomized trials with the Cox model. *Statistics in Medicine* **16**, 225–237.

TCHETGEN TCHETGEN, E. J., WALTER, S., VANSTEELANDT, S., MARTINUSSEN, T. AND GLYMOUR, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology* **26**, 402–410.

TERZA, J. V., BASU, A. AND RATHOUZ, P. J. (2008a). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics* **27**, 531–543.

TERZA, J. V., BRADFORD, W. D. AND DISMUKE, C. E. (2008b). The use of linear instrumental variables methods in health services research and health economics: a cautionary note. *Health Research and Educational Trust* **43**, 1102–1120.

Thanasassoulis, P. and O'Donnell, T. J. (2009). Mendelian randomization. *Journal of American Medical Association* **301**, 2386–2388.

Therneau, T. M., Grambsch, P. M. and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12**, 156–175.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine* **19**, 3309–3324.

Wan, F., Small, D., Bekelman, J. E. and Mitra, N. (2015). Bias in estimating the causal hazard ratio when using two-stage instrumental variable methods. *Statistics in Medicine* **34**, 2235–2265.

Wienke, A. (2010). *Frailty Models in Survival Analysis.* CRC Biostatistics Series. Florida: Chapman & Hall.

Yashin, A. I., Machine, I. A., Begun, A. Z. and Vaupel, J. W. (2001). Hidden frailty: myths and reality. *Research Report*. Department of Statistics and Demography, Odense University.