Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2017 January ; 66(1): 159–185. doi:10.1111/rssc.12158.

Estimating the Causal Effect of Treatment in Observational Studies with Survival Time Endpoints and Unmeasured Confounding

Jaeun Choi[†] and

Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

A. James O'Malley

Department of Biomedical Data Science and The Dartmouth Institute of Health Policy & Clinical Practice, Geisel School of Medicine at Dartmouth College, Lebanon, NH 03756, USA

Summary

Estimation of the effect of a treatment in the presence of unmeasured confounding is a common objective in observational studies. The Two Stage Least Squares (2SLS) Instrumental Variables (IV) procedure is frequently used but is not applicable to time-to-event data if some observations are censored. We develop a simultaneous equations model (SEM) to account for unmeasured confounding of the effect of treatment on survival time subject to censoring. The identification of the treatment effect is assisted by IVs (variables related to treatment but conditional on treatment not to the outcome) and the assumed bivariate distribution underlying the data generating process. The methodology is illustrated on data from an observational study of time to death following endovascular or open repair of ruptured abdominal aortic aneurysm. As the IV and the distributional assumptions cannot be jointly assessed from the observed data, we evaluate the sensitivity of the results to these assumptions.

Keywords

Comparative effectiveness research; Instrumental variable; Observational study; Simultaneous equations model; Survival analysis

1. Introduction

Instrumental Variable (IV) procedures such as Two Stage Least Squares (2SLS) are frequently used to account for unmeasured confounding, whose presence is a common concern in observational studies, with regard to estimation of the causal effect of treatment. However, IV procedures are not applicable to time-to-event data if some observations are censored. Thus, there is a clear need for an IV procedure that can be used with censored

[†]Address for correspondence: Jaeun Choi, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA. jaeun.choi@einstein.yu.edu.

Supplementary information: Additional 'supporting information' may be found with the online version of this article: 'Supplementary material'.

survival data as the exclusion of censored observations is subject to selection bias whenever there is differential loss to follow-up (Hernan, Hernandez-Diaz, and Robins 2004). On the other hand, survival time models that ignore unmeasured confounding will yield biased treatment effects.

Although there recently has been some application of IV methods to survival analysis, the methods used have not accounted for censoring, were not fully validated, were limited to reduced forms of the outcome (e.g. quintiles), were specific to use of random assignment as the IV, or aimed to resolve a different issue than unmeasured confounding. For example, Terza, Basu, and Rathouz (2008) showed that for nonlinear situations (for which 2SLS is less justified) Two Stage Residual Inclusion (2SRI) estimation is consistent and performed well compared to 2SLS and generalizations of it across Weibull models of complete data and other nonlinear models. However, 2SRI does not account for censoring. In other work, O'Malley *et al.* (2011a) used a bivariate probit model (Heckman 1978; Goldman *et al.* 2001; Zeng *et al.* 2006; Bhattacharya, Goldman, and McCaffrey 2006) to jointly model both binary treatment and outcome by a simultaneous equation system; the equation for treatment is linked to an equation for the outcome through both treatment and a selection parameter that reflects the sign and magnitude of the correlation between the error terms of the two equations. Further, O'Malley, Frank, and Normand (2011b) modeled the causal effect of a treatment on a continuous outcome using a simultaneous equations model with IVs, but again censoring was not considered.

In the context of survival data subject to censoring, some studies applied the IV procedures but only illustrated their approaches through examples without validating whether they worked in the setting with censored observations: Abbring and van den Berg (2005) examined the empirical analysis of treatment effects on a duration outcome using random treatment assignment as an IV for treatment received in social experiments; Bosco *et al.* (2010) considered an IV-like estimation, similar to Brookhart and Schneeweiss (2007)'s preference-based IV method, by using logistic regression in the first step and Cox proportional hazards regression in the second step; Recently, Gore *et al.* (2010) and Palmer (2013) used the control function approach of which 2SRI is a special case – Gore *et al.* (2010) applied 2SRI by considering Weibull survival models in the second stage, and Palmer (2013) employed a nonlinear IV approach in a proportional hazards model with an unmeasured confounder having a normal distribution. However, these methods heuristically adapt approaches for linear models as opposed to satisfying theoretical criteria to justify them.

Several papers using IVs for causal inference on survival outcomes have proposed methods for analyzing randomized trials with non-compliance. These use treatment assignment as an IV. Robins and Tsiatis (1991) introduced causal (or structural) AFT models to adjust for non-compliance and called them the rank-preserving structural failure time models (RPSFTM). Joffe (2001) provided a rationale for the use of artificial censoring in the models by Robins and Tsiatis (1991) with comparisons to other methods of survival analysis. The approach by Robins and Tsiatis (1991) was extended to causal proportional hazards models with observed censoring times by Loeys and Goetghebeur (2003). Baker (1998) extended likelihood-based methodology for all-or-none compliance in a randomized trial to the

analysis of discrete-time survival data. Baker's estimators are equivalent to the standard IV estimators in the setting with a survival outcome at a specific time but can provide negative estimates of hazards and be inefficient in some situations. Based on Baker (1998), Nie *et al.* (2011) developed an efficient plug-in nonparametric empirical maximum likelihood estimation (PNEMLE) approach and compared it to standard IV methods and parametric maximum likelihood methods. Their contribution was to gain efficiency by making use of the mixture distribution structure of outcomes due to the presence of compliers and never-takers in both control and treatment arms without fully specifying the outcome distribution. However, these contributions only apply to the binary IV, random assignment, and are not applicable when some observations are censored.

As alternatives to the direct application of IVs in survival analysis, Blundell and Powell (2007) and Chernozhukov, Fernandez-Val, and Kowalski (2011) used IVs to account for unmeasured confounding in duration models, but their methods were based on quantile estimation of a censored regression model and thus the treatment effect is a conditional quantile of the outcome variable. Bijwaard (2009) proposed the Instrumental Variable Linear Rank Estimation (IVLRE) method for duration models that adjusts for the possible endogeneity of the intervention. The method used the inverse of the rank test for the significance of a covariate on the hazard and considered a Generalized Accelerated Failure Time (GAFT) model, but the causal effect in the GAFT model is a difference of quantiles of the outcome distribution. Furthermore, the IVLRE method assumes that the censoring time is known in advance even for uncensored observations. Yu *et al.* (2012) evaluated a prior event rate ratio (PERR) adjustment method, and an alternative (PERR-ALT), in a Cox proportional hazards model. However, PERR and PERR-ALT rely on the availability of survival data from a prior time-period in order to account for baseline differences between the study groups in the absence of the intervention treatment and do not involve formal use of IVs. That is, the study design used by the PERR and PERR-ALT was different from the general setting of survival analysis in observational studies, the focus of this paper.

More recently, two papers utilized IV techniques incorporating censored data in additive hazard formulations. Tchetgen Tchetgen *et al.* (2015) developed two IV approaches: a 2-stage regression approach for a continuous endogenous variable, analogous to 2SLS IV procedures in linear regression; a control function approach for a binary or discrete exposure which is an extension of the 2SRI approach of Terza *et al.* (2008). Li, Fine, and Brookhart (2015) assumed linear structural equation models for the hazard function and developed a closed-form, 2-stage estimator which can be applied for the causal effect of both continuous and discrete exposures. In this paper we are not limited to additive hazards.

In addition, as a different type of application of IVs in causal inference of survival data, Song and Wang (2014) adopted IVs to deal with covariates with measurement error. They developed a simple nonparametric correction approach for estimation in the proportional hazards model using the subset of the sample where IVs are observed and further proposed a generalized methods of moments nonparametric correction estimator to improve the efficiency over the simple correction approach. However, measurement error is a different problem to unmeasured confounding in that bias occurs only in the form of attenuation of estimates towards 0 (a trait of noisy measurements) whereas confounding bias is directional

(a feature of the observed predictor depending on unmeasured factors that are also related to the outcome).

The lack of an established IV methodology for survival data motivates development of a new method to account for unmeasured confounding of the causal effect of treatment on survival time subject to censoring. For this purpose, we propose a structural equations model (SEM) by linking the survival time and treatment selection equations. We consider the case of a log-normal survival model and assume the log-survival time and propensity to be “treated” have an underlying bivariate normal distribution; the method can thus be viewed as an extension of the bivariate probit model to survival time data. From the perspective of missing data, the method is a missing-not-at-random (MNAR) procedure as missingness depends on unobservables. The identification of the treatment effect is assisted by IVs that are related to treatment but conditional on treatment do not directly affect the outcome. Bayesian Markov Chain Monte Carlo (MCMC) methods are used for estimation. We use a novel imputation based procedure to estimate censored survival times using MCMC.

The methodology is illustrated by assessing the comparative effectiveness of endovascular (endo) repair vs open surgical (open) repair in a nationally-representative sample of United States Medicare patients with ruptured abdominal aortic aneurysm (rAAA). For patients with rAAA, if the underlying health status of endo recipients is worse than those undergoing open repair, a naïve comparison of observed survival times will be biased against finding that endo is more effective. In the meantime, if there are more patients having received endo alive at the end of study, disenrolled from their health plans or lost to follow-up, traditional IV methods that exclude those patients will also yield biased estimates of the endo effect. Our goal is to advance methods for comparative effectiveness research involving observational survival data using the method we develop in this paper and showing that it compares favorably to existing methods, including those that only analyze non-censored samples or ignore unmeasured confounding.

Finite sample properties of the proposed method are investigated through simulation studies under various settings including strong and weak IVs and different censoring rates. In addition to when the model holds, we also conduct simulations to examine robustness under model misspecification.

The remainder of the paper is organized as follows. In Section 2, we present our proposed SEM approach and describe the estimation procedure with non-crucial technical details consigned to the appendix and online supplementary material. The proposed method is illustrated with the rAAA data in Section 3, numerical results from simulation studies are provided in Section 4, and the paper concludes in Section 5. The online supplementary material can be obtained from <http://www.blackwellpublishing.com/rss>.

2. Methods

2.1. Model Formulation and Notation

We denote the survival time, the right-censoring time, treatment, a vector of exogenous covariates, and an unmeasured confounding variable for the i -th of n subjects by Y_i , C_i , W_i ,

X_i and U_i respectively. While (Y_i, C_i) may be considered a bivariate random variable, we only observe $Y_i^{obs} = \min(Y_i, C_i)$, the followup time, and $\delta_i = I(Y_i < C_i)$, the status of Y_i^{obs} . We assume that Y_i and C_i are independent conditional on X_i . In this paper we assume treatment is binary (e.g. endo versus open surgery) and Y_i^{obs} (i.e., Y_i and C_i) to have undergone a log transformation which is a natural approach to provide structural flexibility to models of time-to-event data that are positively-valued and generally highly skewed.

To present our assumed causal model we let $Y_i(W)$ denote the potential survival time for subject i under treatment $W_i \in \{0, 1\}$. The survival time that would be observed in the absence of censoring satisfies the consistency equation, $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$. The causal model for the survival times has the form

$$E[Y_i(W)|X_i, U_i] = \psi W_i + \beta^T X_i + \beta_u U_i. \quad (1)$$

The parameter, ψ , denoting the effect of treatment on survival time is the target of inference. A defining feature of the model assumed here is that W_i , the observed treatment for subject i , and U_i are *not* assumed to be independent.

Next we formalize the dependence of W_i on other variables, often referred to as selection variables, that influence the treatment an individual is assigned. We now introduce a vector of variables, denoted Z_i , that are known as “instruments” as they manipulate treatment without fully controlling it and affect the outcome only indirectly through their manipulation of the treatment (Imbens and Angrist 1994). The potential treatment an individual receives can be represented by its own causal model. We specifically assume $W_i(Z) = I(W_i^*(Z) > 0)$, where

$$E[W_i^*(Z)|X_i, U_i] = \lambda^T Z + \theta^T X_i + \theta_u U_i, \quad (2)$$

for $Z \in \mathcal{R}(Z)$, the set of all possible values of Z . The quantity $W_i^*(Z)$ is an unbounded quantity representing the underlying propensity of individual i to be treated ($W_i = 1$) under the counterfactual $Z_i = Z$. For the IV Z_i we make the following assumptions:

(A1) Z_i is associated with W_i conditional on X_i and U_i .

(A2) Exclusion Restriction: Z_i is uncorrelated with Y_i conditional on W_i , X_i , and U_i . That is, $Y_i(Z_i|W_i, X_i, U_i) = Y_i(Z_i'|W_i, X_i, U_i)$ for all Z_i, Z_i' , W_i , X_i , U_i and all subjects i .

(A3) Z_i is unrelated to U_i conditional on X_i .

(A2) says that there is no direct effect of Z_i on Y_i (Angrist *et al.* 1996) while (A3) says that the mechanism of generating Z_i is ignorable (Rubin 1978) and thus Z_i shares no common

causes with Y_i . When Y_i is modeled with an explicit error term $\varepsilon_{1,i}$ (A2) and (A3) reduce to the assumption that $\mathbf{Z}_i$ and $\varepsilon_{1,i}$ (which includes U_i) are uncorrelated conditional on W_i and $\mathbf{X}_i$. Besides (A1)–(A3), Equations 1 and 2 imply the following assumptions for causal identification:

(A4) Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1978)

- i. Non-interference between units: Treatment status of any unit does not affect the potential outcomes of the other units. That is, $Y_i(W_1, W_2, \dots, W_n) = Y_i(W_i)$
- ii. No variation in treatment: There is only a single version of each treatment level.

(A5) Monotonicity (Imbens and Angrist 1994): $W_i(\mathbf{Z}) \geq W_i(\mathbf{Z}')$ for any $\mathbf{Z} \geq \mathbf{Z}'$ and all subjects i

The causal model characterized by Equations 1 and 2 describes the data generating mechanism under which (i) individuals are assigned (or select) a treatment and (ii) observed and unobserved selection factors including treatment affect outcomes. In order to estimate ψ and the other parameters in the presence of censoring, we embed these causal equations in the following SEM:

$$Y_i = \psi W_i + \beta^T \mathbf{X}_i + \varepsilon_{1,i} \text{ and } W_i^* = \lambda^T \mathbf{Z}_i + \theta^T \mathbf{X}_i + \varepsilon_{2,i}, \quad (3)$$

where $W_i = I(W_i^* > 0)$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{1,i}, \varepsilon_{2,i}) = (\beta_u U_i + \delta_{1,i}, \theta_u U_i + \delta_{2,i})$, and $(\delta_{1,i}, \delta_{2,i})$ is a bivariate vector of random variables that are independent of all other variables. Because $\boldsymbol{\varepsilon}_i$ depends on U_i it follows that $\text{Cov}(W_i, \varepsilon_{1,i}) \neq 0$, which violates the conditions under which least squares or other methods for estimating linear regression models are valid.

To complete the specification of the model we assume U_i , $\delta_{1,i}$ and $\delta_{2,i}$ are independent random normal variables with means 0 and variances, σ_u^2 , τ_1^2 , and τ_2^2 , respectively. Therefore, $\boldsymbol{\varepsilon}_i$ has a bivariate normal distribution with mean $(0, 0)^T$ and variance

$$\text{cov}(\boldsymbol{\varepsilon}_i) = \begin{pmatrix} \beta_u^2 \sigma_u^2 + \tau_1^2 & \beta_u \theta_u \sigma_u^2 \\ \beta_u \theta_u \sigma_u^2 & \theta_u^2 \sigma_u^2 + \tau_2^2 \end{pmatrix}. \quad (4)$$

Because we only observe the binary treatment W_i , not the underlying propensity to be treated W_i^* , the restriction $\theta_u^2 \sigma_u^2 + \tau_2^2 = 1$ is imposed in order to identify the model parameters (see the online supplementary material A.1 for justification). It follows that we may complete (3) by specifying

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix} \right), \quad (5)$$

where σ_1^2 is interpreted as the residual variance of the survival time distribution. The quantity ρ represents the net effect of unmeasured confounding by characterizing the extent to which unobserved factors (U_i) affecting W_i^* are correlated with those affecting Y_i . A positive correlation ($\rho > 0$) indicates (in our case) endo-favorable selection because unobserved factors that make subjects more likely to receive endo lengthen survival times.

This model formulation resembles the models in Chib and Hamilton (2000, 2002) in which a bivariate normal distribution underlies the assignment of a binary treatment and realization of a continuous outcome. Together with the inclusion of valid IVs in the SEM, the assumption of a specific family of distributions allows censoring and unmeasured confounding to be simultaneously accounted for and distinguished in order to estimate the causal effect of treatment on survival time.

2.2. Likelihood

Bayesian methods are used for estimation, so we must specify both the likelihood function for the observed data and a prior distribution for the unknown model parameters to complete the specification of the model. In this subsection we develop the likelihood function while in the next we specify the prior.

For each subject we observe Y_i^{obs} , Δ_i and W_i conditional on X_i and Z_i . Therefore, the likelihood function of the model parameters $(\psi, \beta, \lambda, \theta, \sigma_1^2, \rho)$ given these observations is the product over $i = 1, \dots, n$ of terms of the form:

$$\begin{aligned} & f(Y_i^{obs}, \Delta_i, W_i | X_i, Z_i) \\ &= \left[\phi(Y_i^{obs}; \mu_{y,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|y,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|y,i} | X_i, Z_i))^{1-W_i} \right]^{\Delta_i} \\ & \times \left[S(Y_i^{obs} | W_i, X_i, Z_i) f(W_i | X_i, Z_i) \right]^{1-\Delta_i}, \end{aligned} \quad (6)$$

which are obtained by integrating the density of models (3) and (5) over W_i^* (i.e. integrating over $\varepsilon_{2,i}$). The components of (6) are given by:

$$\mu_{y,i} = \psi W_i + \beta^T X_i, \mu_{w|y,i} = \frac{\lambda^T Z_i + \theta^T X_i + \rho(Y_i^{obs} - \mu_{y,i})/\sigma_1}{\sqrt{1 - \rho^2}},$$

$$S(Y_i^{obs} | W_i, X_i, Z_i) = \Pr(Y_i \geq Y_i^{obs} | W_i, X_i, Z_i) = \int_{Y_i^{obs}}^{\infty} \phi(Y_i; \mu_{y,i}, \sigma_1^2 | W_i, X_i, Z_i) dY_i, \text{ and} \quad (7)$$

$$f(W_i | X_i, Z_i) = \Phi(\mu_{w^*,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w^*,i} | X_i, Z_i))^{1 - W_i} \text{ with } \mu_{w^*,i} = \lambda^T Z_i + \theta^T X_i,$$

where $\phi(v; \mu, \sigma^2)$ and $\Phi(v)$ denote the probability density function of $N(\mu, \sigma^2)$ and the cumulative density function of $N(0, 1)$, respectively, evaluated at v , for $\rho = 1$. The derivation of (6) and (7) is given in Appendix A.

2.2.1. Parameter expansion to censored survival times

The estimation procedure based on the likelihood (6) is complicated due to the joint problem of censoring and unmeasured confounding: in survival analysis, when the i -th subject is censored (i.e. $Y_i^{obs} = C_i$ and $\Delta_i = 0$), the observation contributes to the likelihood through the survival probability for the observed censoring time. The survival function $S(Y_i^{obs} | W_i, X_i, Z_i)$ in (6) is obtained by integrating its conditional survival function over the distribution of the unmeasured confounder; when the resulting function does not have a closed form, this integration becomes too complex and computationally intensive. To simplify computations, instead of directly evaluating survival probabilities we propose to impute censored survival times bounded below by the observed censoring time (see Section 2.4.1). Therefore, for this imputation, we expand the parameter space by treating the censored survival time as a range restricted unknown parameter denoted Y_i^* . Then, the likelihood function for

$(\psi, \beta, \lambda, \theta, \sigma_1^2, \rho, Y_i^* | Y_i = Y_i^* \geq C_i)$ is given by:

$$\begin{aligned} & [\phi(Y_i; \mu_{y,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|y,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|y,i} | X_i, Z_i))^{1 - W_i}]^{\Delta_i} \\ & \times [\phi(Y_i^*; \mu_{Y^*,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|Y^*,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|Y^*,i} | X_i, Z_i))^{1 - W_i}]^{1 - \Delta_i} = \phi(Y_i^a; \\ & \mu_{Y^a,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|Y^a,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|Y^a,i} | X_i, Z_i))^{1 - W_i}, \end{aligned} \quad (8)$$

where $Y_i^a = Y_i^{\Delta_i} (Y_i^*)^{(1 - \Delta_i)}$ and $\mu_{Y^*,i}$ and $\mu_{w|Y^*,i}$ have Y_i^* in place of Y_i in (7), and $\mu_{Y^a,i}$ and $\mu_{w|Y^a,i}$ have Y_i^a in place of Y_i in (7). A comparison to the integrated (Y_i^*) marginalized

likelihood in Appendix A reveals that conditioning the value of censored observations on Y_i^a yields likelihood-function contributions that have a closed and analogous form across all observations, which is highly amenable to MCMC updating.

2.3. Prior and Posterior Distribution

The posterior distribution of the parameters $\Theta = (\psi, \beta, \lambda, \theta, \sigma_1^2, \rho, Y^*)$ in (8) given observed data has the generic form

$$P(\Theta | \text{Data}) = \frac{P(\text{Data} | \Theta)P(\Theta)}{\int P(\text{Data} | \Theta)P(\Theta)d\Theta} = \frac{P(\text{Data} | \Theta)P(\Theta)}{P(\text{Data})} \propto P(\text{Data} | \Theta)P(\Theta).$$

In the following we re-parameterize λ , θ , and ρ in (7) to parameters that are less structurally dependent (i.e., closer to orthogonal) and for which closed-form posterior distributions are more accessible:

$$\tilde{\lambda} = \frac{\lambda}{\sqrt{1 - \rho^2}}, \tilde{\theta} = \frac{\theta - \beta\rho/\sigma_1}{\sqrt{1 - \rho^2}}, \text{ and } \tilde{\rho} = \frac{\rho}{\sigma_1\sqrt{1 - \rho^2}}. \quad (9)$$

Re-parameterization is further discussed in the online supplementary material A.2. Then, to complete the model specification, we specify priors for three original parameters $(\psi, \beta, \sigma_1^2)$ and the three transformed parameters $(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho})$. A prior distribution does not need to be specified for the censored survival outcome (Y_i^*) as we assume observations are exchangeable and, therefore, the survival time density function and observed censoring time are the basis for informing Y_i^* .

We assume diffuse priors for the model parameters other than ρ . Because we are interested in sensitivity to unmeasured confounding, prior distributions for ρ of varying precision are used. Non-informative flat priors are assumed for ψ , β , $\tilde{\lambda}$ and $\tilde{\theta}$. The Inverse Gamma prior, the conjugate prior for the variance of a normal distribution, is assumed for σ_1^2 . Specifically, $p(\psi) \propto 1$, $p(\beta) \propto 1$, $p(\tilde{\lambda}) \propto 1$, $p(\tilde{\theta}) \propto 1$, and $\sigma_1^2 \sim IG(\omega_1, \omega_2)$, where $\omega_1 (=0.001)$ and $\omega_2 (=0.001)$ are chosen such that the variance far exceeds the mean. We assume a Beta prior for $\rho^* = (\rho + 1)/2$, (i.e. $\rho^* \sim \text{Beta}(\nu_1, \nu_2)$), implying a Beta-type prior for ρ over $(-1, 1)$. Thus, ρ has the prior density:

$$p(\rho; \nu_1, \nu_2) = \left(\frac{1}{2}\right)^{\nu_1 + \nu_2 - 1} \frac{1}{B(\nu_1, \nu_2)} (1 + \rho)^{\nu_1 - 1} (1 - \rho)^{\nu_2 - 1},$$

where $B(\nu_1, \nu_2) = \Gamma(\nu_1)\Gamma(\nu_2)/\Gamma(\nu_1 + \nu_2)$. Therefore, $\tilde{\rho}$ has the prior,

$$p(\tilde{\rho}, \nu_1, \nu_2) = \left(\frac{1}{2}\right)^{\nu_1 + \nu_2 - 1} \frac{1}{B(\nu_1, \nu_2)} \frac{\sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}} \left(1 + \frac{\tilde{\rho} \sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}}\right)^{\nu_1 - 1} \left(1 - \frac{\tilde{\rho} \sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}}\right)^{\nu_2 - 1}. \quad (10)$$

In the special case where $\nu_1 = \nu_2 = 1$, ρ follows a uniform $(-1, 1)$ distribution and $\tilde{\rho}$ has a t -distribution with 2 degrees of freedom. The model assumes that censored and non-censored observations are exchangeable and that, in general, censoring is non-informative. Therefore, the conditional posterior for Y_i^* equals the marginal density for Y_i^* (obtained by integrating over the density function of the latent propensity W_i^* , or equivalently $\varepsilon_{2,i}$, as in Appendix A) truncated on the left at C_i . All conditional posterior distributions needed for the computation described in Section 2.4 are given in Appendix B. We refer to the model given in (3) and (5) and the above prior distributions as the Bayesian structural equations model (BSEM).

2.4. Bayesian Computation

We briefly outline the Markov Chain Monte Carlo (MCMC) procedure used to fit the BSEM. For the n_0 subjects with right-censored survival times, let $\mathbf{Y}_0^*$ denote a $n_0 \times 1$ vector of censored survival outcomes with the j -th element, $Y_{0,j}^*$, $j = 1, \dots, n_0$ (Recall that $Y_i^* \geq C_i$ denotes the parameter for the censored survival outcome for subject i if $i = 0$.)

The MCMC procedure first selects initial values

$(\mathbf{Y}_0^{*(0)} = (Y_{0,1}^{*(0)}, \dots, Y_{0,n_0}^{*(0)}), \psi^{(0)}, \boldsymbol{\beta}^{(0)}, \lambda^{(0)}, \boldsymbol{\theta}^{(0)}, \sigma_1^{2(0)}, \rho^{(0)})$. They are updated in order to obtain

$Y_{0,1}^{*(1)}, \dots, Y_{0,n_0}^{*(1)}, \boldsymbol{\beta}^{(1)}, \sigma_1^{2(1)}, (\lambda^{(1)}, \boldsymbol{\theta}^{(1)}, \rho^{(1)})$, and $\psi^{(1)}$. For $(\lambda^{(1)}, \boldsymbol{\theta}^{(1)}, \rho^{(1)})$, their transformed counterparts $(\tilde{\lambda}^{(1)}, \tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\rho}^{(1)})$ are jointly updated and then transformed back to update the original parameters. $\boldsymbol{\beta}^{(1)}$ and $\sigma_1^{2(1)}$ are directly generated from closed-form conditional posteriors, a normal and an Inverse-Gamma distribution, respectively, using Gibbs sampling steps. Candidate values for $Y_{0,1}^{*(1)}, \dots, Y_{0,n_0}^{*(1)}, (\tilde{\lambda}^{(1)}, \tilde{\boldsymbol{\theta}}^{(1)}, \tilde{\rho}^{(1)})$ and $\psi^{(1)}$ are generated from

their conditional posteriors by the Metropolis-Hastings (M-H) algorithm; each current value is updated with the newly generated value if the latter is accepted based on the M-H acceptance probability (see the online supplementary material A.3). Otherwise, the current value is retained. The procedure is then iterated.

We found that running the MCMC procedure for 40,000 samples following burn-in of 10,000 samples was sufficient by monitoring trace plots, the Gelman-Rubin diagnostic, and the Geweke criterion for assessing convergence of the MCMC procedure. The posterior estimates of the parameters were evaluated as Monte Carlo averages using the $M (= 40,000)$ sample draws from the joint posterior distribution. The detailed MCMC procedure is given

in the online supplementary material A.4, and the convergence diagnostics are further discussed in the online supplementary material A.5.

2.4.1. Imputing censored survival times—In Section 2.2.1 we represented the censored survival times as range restricted unknown parameters to avoid direct computation of $S(Y_i^{obs} | W_i, X_i, Z_i)$, which is more computationally demanding. In the course of our MCMC algorithm we use a novel procedure to update the parameter Y_i^* , representing the true value of Y_i . We now justify this procedure.

When the survival time for the i -th subject is observed ($Y_i^{obs} = Y_i$), the joint marginal density for (Y_i, W_i) evaluated at (Y_i^{obs}, W_i) given in (6) is

$$f(Y_i^{obs}, W_i | X_i, Z_i) = \phi(Y_i^{obs}; \mu_{y,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|y,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|y,i} | X_i, Z_i))^{1-W_i}. \quad (11)$$

When the i -th subject is censored ($Y_i^{obs} = C_i$), Y_i^* the realization of the survival time Y_i that would have been observed in the absence of censoring, is left-truncated by C_i .

To obtain draws of Y_i^* from its conditional posterior, we use the truncated normal distribution, $f(Y_i^* | W_i, X_i, Z_i) \propto \phi(Y_i^*; \mu_{Y^*,i}, \sigma_1^2 | W_i, X_i, Z_i) I(Y_i^* \geq C_i)$, as a candidate generating density and apply the M-H algorithm (see Appendix B). Since $\mu_{Y^*,i} = \psi W_i + \beta^T X_i$ does not depend on Y_i^* , at each iteration the candidates are drawn independently of the current value of Y_i^* (an independence chain, Tierney 1994), in contrast to a random walk chain. Under this strategy, the probability of accepting a new candidate value $Y_i^{*,new}$ is given by

$$\min \left\{ 1, \frac{\Phi(\mu_{w|y^*,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|y^*,i} | X_i, Z_i))^{1-W_i}}{\Phi(\mu_{w|y^*,i} | X_i, Z_i)^{W_i} (1 - \Phi(\mu_{w|y^*,i} | X_i, Z_i))^{1-W_i}} \right\}.$$

We found the acceptance rate to be very high (around 0.9), which implies the shape of the truncated normal distribution used to generate posterior samples is close to the shape of the true conditional posterior.

To examine the predictability of Y_i^* under the fitted model, we generated data emulating our observed data and for a specified treatment effect plotted the densities of log(estimated survival time – censoring time) for all censored observations for each treatment group (Figure 1). The vertical lines in Figure 1 indicate the average of log(true survival time –

censoring time) over the censored observations in the corresponding treatment group. Figure 1 shows treatment is predictive of the true survival time beyond its association with the observed censoring time and, within each treatment group, the estimated $\log(\text{survival time} - \text{censoring time})$ are centered around the average of their true values suggesting the imputations are calibrated appropriately and have minimal bias.

3. Application to Vascular Surgery Data

In this section, the analysis of the BSEM is illustrated using vascular surgery data and the results are compared to those for standard survival analysis, which ignores confounding, and the 2SLS procedure, which can only be applied to the non-censored subset of observations. Standard survival analysis is implemented using a log-normal accelerated failure time (AFT) model.

Endovascular (endo) repair was introduced in 1999 as a less invasive alternative to elective open surgical (open) repair of AAA, which was traditionally performed to prevent ruptures. The goal of this analysis is to evaluate the effectiveness of endo repair on survival of patients with rAAA, one of the most fatal surgical emergencies, compared to that of open repair.

We used Medicare claims data to identify all open and endo repairs of rAAA that occurred during 2001–2008. To be eligible for analysis, patients were required to have at least 2 years of prior Medicare enrollment. This restriction ensured that comorbidities that might influence the choice of approach and outcomes of rAAA repair could be measured equitably for all patients. We only included each patient in the data set once by only using the first case for the few individuals who appeared multiple times. A total of 2,853 ruptured cases met the criteria for analysis, yielding 2,201 (77.15%) and 652 (22.85%) patients who underwent endo and open repair, respectively. The proportion of survival times that were censored is high (66.28%). Patients in the endo group had an average of 1579 (SD=780) days of maximal follow-up while those in open repair group had an average of 1854 (SD=887) days of maximal follow-up.

The survival time is the number of days from procedure date to death. Treatment is coded 1 for endo and 0 for open surgery. Based on a prior study (O'Malley *et al.* 2011a), we adjusted for the following observed confounders on which both treatment selection and survival may depend: gender, race, age at procedure, seven indicators of specific comorbidities (previous renal failure, congestive heart failure, chronic pulmonary disease, peripheral vascular disorders, vascular disease, neurovascular disease, and prior AAA diagnosis), two procedural factors (year of procedure and urgent case – defined as emergency department charges of \$50 or more), and total number of AAA procedures over the prior 365 days at the hospital where the procedure was performed. There are suspected to be unmeasured confounders that influence choice of procedure (some unknown to even the referring physician). Possible unmeasured confounders in the rAAA example may be patients' underlying health status and physicians' preference for new treatment.

The proportion of endo cases over the prior 365 days of each procedure at the hospital the patient attended is used as an IV. The rationale behind this choice is that the likelihood a

rAAA patient receives endo is likely to increase with the proportion of endo cases performed at the hospital over the prior 365 days. To be a valid IV, the proportion of endo procedures performed over the past 365 days from a patient's procedure must be independent of their survival time conditional on their treatment and other covariates, including the total number of AAA cases and procedure date. Among the rAAA patients in this analysis, the value of our IV is greater on average for patients who received endo (59%) than those who received open (32%) (Table 2).

The characteristics of the rAAA patients who received AAA procedures are described in Tables 1 and 2. The majority of patients were white and male, did not have urgent admission, and had a prior AAA diagnosis without rupture. The volume of open repair decreased over time while that of endo gradually increased, except during the later years. A higher proportion of patients were urgently admitted for open repair while significantly higher proportions with congestive heart failure, vascular disease or prior AAA diagnosis without rupture underwent endo. The total volume of AAA procedures over the prior 365 days at the facility where the procedure was performed was higher for endo than open repair. The observed confounders among censored and uncensored observations are similarly distributed, except for year of procedure. Because their procedures were performed closer to the end of the follow-up period, more patients have censored survival times in the latter study years.

To enhance interpretation and lessen the vulnerability of the model to misspecification, we use the log as opposed to an optimal Box-Cox transformation to normality. To evaluate whether the assumed model is reasonable we plot the distribution of the residuals for observed survival times, which is appropriate for checking the survival time distribution if censoring is completely at random, and also evaluate a posterior predictive check, which provides a more rigorous assessment of the appropriateness of the full model. The former is a descriptive comparison reliant on only the model for the outcome while the latter is based on the full BSEM. Both approaches are described in the following four paragraphs.

Figure 1 of the online supplementary material B shows a histogram of the standardized residuals of the non-censored actual observations and a super-imposed smooth density obtained when a log-normal AFT model is fitted to the survival times. Although the presence of censored observations inhibits direct interpretation of the plot as an assessment of normality, the symmetric appearance of the estimated density offers some support for log-normality.

If the overall level of skewness in the data is not accounted for by the model, the results have the potential to be biased. Therefore, we use a posterior predictive check (PPC) with the skewness coefficient as the test statistic to check whether the BSEM model recaptures the overall level of skewness in the data, which, in turn evaluates whether the assumption of bivariate normality (of the log survival time and the underlying propensity to undergo endo) is reasonable. The procedure we use is described below.

To emulate the rAAA study we constructed an empirical distribution of censoring times by computing the difference in time from when each patient underwent their procedure until the

day when they were censored or (if they weren't censored) to December 31, 2009 (the maximum possible time of follow-up). We randomly assigned a censoring time to each subject and designated them as censored if their predictive survival time was greater.

To compute the PPC, we first sampled 100 sets of values from the posterior predictive distribution corresponding to each MCMC draw of parameter values from the joint posterior distribution of the model parameters. For each vector of parameters, we generated a random draw from the predictive survival time distribution for each patient in the data set. Then for each set of predictive survival times we impose the censoring procedure described above to each observation and compute the skewness coefficient, denoted S , on the resulting set of log-scaled $\min(\text{predictive survival time, newly assigned censoring time})$ values. Figure 2 shows the predictive and actual distributions of these values for subjects with procedures in 2005 and 2008. Finally, we calculate skewness for the observed data, $S(\mathbf{y})$, which equaled -1.86, and for each set of drawn log predictive values, $S(\mathbf{y}_{rep}|\mathbf{y})$, $rep = 1, 2, \dots, 100$. The posterior predictive p-value equals the fraction of times that $S(\mathbf{y}_{rep}|\mathbf{y}) > S(\mathbf{y})$. A posterior predictive p-value close to 0.5 suggests that the model fits the data well while extreme values (close to 0 and close to 1) are suggestive of lack-of-fit. We obtain a p-value = 0.39 suggesting that the model is successfully reproducing the overall amount of skewness in the data.

We also graphically checked two structural assumptions of predictors: the linearity of their relationship to the log survival time in (1); the linearity of their relationship with the propensity to undergo the endovascular procedure in (2) (see Figures 2 and 3 of the online supplementary material B). We report results for the variables 'facility total volume of AAA procedures over prior 365 days' and 'facility proportion of endo cases over the prior 365 days' (the variable used as an IV). For both of these variables we use a smoothing function, lowess (locally weighted scatterplot smoothing), to obtain a data-driven visual test of the appropriateness of the assumed linear relationship with the left-hand-side of (1) and (2). The lowess fits of 'facility total volume of AAA procedures over prior 365 days' appear linear for both the log survival time and latent propensity of endo although their slopes are close to 0. The IV 'facility proportion of endo cases over prior 365 days,' which is only included in and thus assessed for the treatment equation, shows obvious linearity with the fitted latent propensity to receive endo. All the other predictor variables are categorical so assessing linearity is not a concern.

For the BSEM, we assume the prior mean of ρ to be 0 because (i) a priori we did not know whether any unmeasured confounders of treatment and survival time for rAAA cases would be positively or negatively correlated; (ii) given that many observed confounders are already adjusted for in this analysis, we felt it unlikely that ρ would be far from 0. The prior of σ_1^2 is assumed to be $IG(0.001, 0.001)$.

Table 3 compares the estimates of treatment effect (ψ) between estimation methods both when censored observations are included and excluded. Estimates are obtained using the BSEM with and without an IV, the log-normal AFT model (which assumes no unmeasured confounding), and the 2SLS procedure (excludes censored cases). For comparison, the BSEM and the AFT model are also evaluated when censored observations are excluded. For

the BSEM in this table, we assume a vague extended Beta prior with parameters=(1,1) for ρ , which corresponds to a uniform $(-1, 1)$ distribution (mean 0 and variance 0.33).

There are several interesting findings in Table 3. First, the BSEM and the AFT model produce estimates of ψ that are smaller in magnitude when censoring is accounted for than when only the non-censored data are analyzed. Thus, removing censored observations likely results in overestimation of ψ . Second, ignoring confounding leads to estimates of smaller magnitude of ψ in the standard AFT model than when accounting for unmeasured confounding under the BSEM, irrespective of whether censoring is accounted. However, 2SLS yields the smallest estimated ψ among the methods for non-censored data. This may reflect a selection bias incurred from excluding the large number of censored observations in the rAAA data. Third, the BSEM accounting for both censoring and unmeasured confounding yields negative point estimates of ψ implying shorter survival under endo while all the other point estimates of ψ are positive, suggesting longer survival under endo. However, when accounting for censoring, the 95% credible intervals from the standard AFT model and the BSEM reassuringly include 0, implying endo neither significantly increased nor decreased survival compared to open. Fourth, the BSEM without an IV appears to magnify the point estimate of ψ and produces wider interval estimates than the BSEM with an IV, which confirms that including an IV in the BSEM helps the identification and precision of estimation of ψ .

To assess the impact of a realistic unmeasured confounder in the rAAA data, we created an unmeasured confounder with known effect sizes by omitting the observed predictor with the largest product of effects across the model equations and compared the results in Table 3 for the full model to those under the reduced model. We omitted 'urgent admission', an indicator of whether a patient was urgently admitted, as it had the greatest overall impact on treatment and survival time among the observed predictors (the estimates of coefficients of the observed confounders are presented in and 2 of the online supplementary material B).

The results in Table 4 reveal that omitting the urgent admission indicator led to an increase of 0.026 in the estimate of ρ and a decrease of 0.023 in the estimate of ψ . The fact that these estimates changed by relatively small amounts compared to when urgent admission was included in the model suggests that a simulation study that considered ρ in the range $[-0.2, 0.2]$ is likely to encompass the range of realistic scenarios for the strength of an unmeasured confounder. In the simulations described in Section 4 we also consider the cases with $\rho = \pm 0.5$, thereby testing our method under more extreme conditions would be considered likely to occur in practice.

We also assessed the sensitivity of the BSEM to different levels of precision of the prior for ρ (Table 5). Beta-type distributions with parameters=(1,1), (5,5) and (10,10) form the vague (uniform or flat), medium, and precise priors of ρ , respectively; prior mean=0 and variance=0.33, 0.09, and 0.05. The BSEM appears insensitive to the different priors for ρ .

The reversed sign of the point estimate of ψ between the AFT model and the BSEM on the full data set (Table 3) indicates that accounting for unmeasured confounders has an important impact. But can the BSEM results be trusted? Answering this question involves

the evaluation of ρ . As mentioned earlier in this section, it is unlikely that unmeasured variables could exist such that ρ is far from 0 because we have a very detailed data set and AAA has been studied extensively. Furthermore, in Copas and Li (1997), it is suggested that ρ estimated to be near ± 1 indicates model lack-of-fit. Therefore, a test of the suitability of the BSEM is whether ρ is moderately close to 0. It is reassuring that the estimates of ρ under all specifications of the BSEM are within the range $(-0.2, 0.2)$ (Table 5).

It has been noted that IG priors with large variances can be problematic when used for variance parameters at levels of the model above the observation level (Gelman 2006). Although σ_1^2 is not a hierarchical variance parameter we nonetheless compared results between the $IG(0.001, 0.001)$ and $IG(0.1, 0.1)$ priors (see Table 3 of the online supplementary material B). The posterior point and interval estimates of σ_1^2 under $IG(0.1, 0.1)$ are almost identical to those under $IG(0.001, 0.001)$, consistent with Gelman (2006)'s comment that typically any reasonable non-informative prior distribution can be used for variance parameters at the observation level.

4. Simulation Studies

In this section, we conduct simulation studies to evaluate the sensitivity of $\hat{\psi} = E[\psi|\text{Data}]$, the posterior mean estimator of ψ , to: (1) the prior of ρ , (2) strength of the IV, (3) different censoring rates, and (4) misspecification of the distribution of ϵ_i .

We assume X_i and Z_i are univariate and generated from a uniform (0, 1) distribution. Further, $\epsilon_i = (\epsilon_{1,i}, \epsilon_{2,i})$ has a bivariate normal distribution except in those simulations evaluating the impact of a wrongly assumed error distribution. W_i^* is generated by the linear model in (3) (in Section 2.1) with $W_i = 1$ for $W_i^* > 0$ and 0 otherwise. The transformed outcome Y_i is generated by the linear model in (3); the exponential of Y_i can be thought of as the original positively-valued survival time under a log-normal model. The exponential of C_i is generated from a uniform (0, a) distribution and log-transformed to C_i . The censoring proportion is 30% except when evaluating the performance of the estimator at different censoring rates. Performance is evaluated by computing bias, mean squared error (MSE), and coverage probabilities over 100 simulated data sets.

4.1. Finite Sample Properties and Sensitivity to ρ

The values of the survival-time parameters (Equation 1) for data generation were $\psi = -0.5$, $\beta_0 = 0.5$, and $\beta_1 = 0.5$ and the treatment-selection parameters (Equation 2) were $\lambda = 0.5$, $\theta_0 = -0.5$ and $\theta_1 = 0.5$. The variance and correlation parameters in the bivariate normal error distribution (5) were set to $\sigma_1^2 = 0.5$ and $\rho_{GEN} = 0/\pm 0.2/\pm 0.5$, where the additional notation ρ_{GEN} is used to denote the data generated value of ρ when ρ is also being referred to as an unknown parameter.

As in Section 3, Beta-type distributions with parameters (1,1), (5,5) and (10,10) represent vague, medium, and precise priors of ρ , respectively; prior mean = 0 in all cases and variance = 0.33, 0.09, and 0.05. As mentioned in Section 3, the choice of 0 as the prior mean

of ρ (corresponding to no unmeasured confounding) is reasonable when a priori the direction of unmeasured confounding is unknown and we want to allow the data to be solely responsible for moving one way or the other. However, when investigators know of a specific unmeasured confounder between treatment and outcome, the prior of ρ should be centered according to this knowledge. Hence, we also conduct simulations with the prior mean of ρ , $E[\rho]$, equal to 0.2. To evaluate sensitivity of the BSEM analysis with respect to the prior for σ_1^2 , simulations were conducted with both $IG(0.001, 0.001)$ and $IG(0.1, 0.1)$ as its prior distribution. To investigate the asymptotic properties of $\hat{\psi}$, the Bayesian posterior mean estimator of the treatment effect, simulations were performed with $n = 1000$ and $n = 5000$.

The results in Table 6 and Figure 3(a) reveal that the bias of $\hat{\psi}$ increases as ρ_{GEN} diverges from the prior mean, $E[\rho] = 0$, and as the prior of ρ becomes more informative. Conversely, bias is very close to 0 when $\rho_{GEN} = E[\rho]$ regardless of informativeness of the prior. Likewise, RMSE and coverage improve with $\text{var}(\rho)^{-1}$, the prior-precision of ρ , when $\rho_{GEN} = 0$ (i.e. $\rho_{GEN} = E[\rho]$) whereas they deteriorate with increased prior precision when $\rho_{GEN} = \pm 0.2, \pm 0.5$ (i.e. $\rho_{GEN} \neq E[\rho]$). As n increases from 1000 to 5000, the informativity of the priors for ρ has less influence with bias and RMSE becoming smaller and the coverage probability converging upon the 95% nominal level. However, there is one case where $\rho_{GEN} = 0$ in which the coverage probability is further from 95% with $n = 5000$ as compared to $n = 1000$, especially when the prior for ρ is vague. This apparent lack of consistency was found to be a consequence of the MCMC chain taking longer to converge when ρ is weakly identified, which can occur under a flat prior. Under 100,000 iterations of the MCMC chain, the coverage with $n = 5000$ approached the 95% nominal level (see Table 4 of the online supplementary material B). The fact that our analyses of real data appeared to have converged by 40,000 iterations is consistent with the true value of ρ not being exactly equal to 0 and perhaps that the real data were more informative than the simulated data.

Simulations comparing the $IG(0.001, 0.001)$ and $IG(0.1, 0.1)$ priors for σ_1^2 (Table 5 of the online supplementary material B) showed very similar results under all values of ρ_{GEN} , confirming that estimates of σ_1^2 under the BSEM are reasonably robust to the parameters of the IG prior. In addition, there is minimal change in the bias of the estimates of ψ and ρ between these two IG priors for σ_1^2 as ρ_{GEN} increases.

As shown in Figure 4 by the posterior distributions of ρ when $\rho_{GEN} = 0, 0.2$ and 0.5 , the posterior means of ρ ($\hat{\rho} = 0.008, 0.182$ and 0.472 indicated by vertical lines) are close to ρ_{GEN} . Even when the prior mean $E[\rho]$ is far from ρ_{GEN} the bias of $\hat{\rho}$ is small but does increase as the prior for ρ becomes more informative. A similar trend is observed for estimates of ψ .

The results from additional simulations by the BSEM when the prior mean for $\rho = 0.2$ (provided in Table 6 of the online supplementary material B) confirms that the further the prior mean for ρ is from ρ_{GEN} the less accurate the point and interval estimates of the effect

of treatment on the outcome, ψ . The estimates of all the other parameters were fairly robust to ρ_{GEN} and the informativeness of the prior for ρ .

4.2. Sensitivity to the Strength of IVs

To illustrate the importance of a strong IV to the BSEM, we evaluate the operating characteristics of $\hat{\psi}$ at $\lambda = 0, 0.5, 1$, and 2 (the effect of the IV). A vague prior for ρ is assumed so that the IV is responsible for identifying ψ separate from ρ . The other model specifications and simulation settings are as in Section 4.1.

Table 7 and Figure 3(b) reveal that the estimate of ψ is less biased when λ is bigger (stronger IV). Also, ψ is estimated with more precision as the strength of the IV increases (Table 7). It is clear that the benefit of a strong IV is relatively more evident for larger ρ_{GEN} (Table 7 and Figure 3(b)) due to the fact that the higher correlation between the error terms of the survival time and the treatment selection equations increases the information available to estimate ψ provided unmeasured confounding is accounted.

4.3. Sensitivity to Censoring Rate

To assess sensitivity to different censoring proportions we set the censoring rate to be 0%, 30%, and 60% while fixing the other simulation parameters to the same values as previously and assuming the uniform $(-1, 1)$ prior for ρ .

Because bias is close to 0 for all censoring rates, detailed results are consigned to the online supplemental material B (Table 7 and Figure 4). Although lower censoring rates lead to better precision, the differences are small. This indicates that the information in the censoring times makes an important contribution to estimation; even in the presence of substantial censoring, precision is only modestly compromised.

4.4. Sensitivity to Model Misspecification

We finally examine the impact of departures from the bivariate normal assumptions. Data were generated assuming the true distribution of the random errors of the survival time and treatment selection equations to be (i) a bivariate t-distribution and (ii) a mixture of a bivariate normal distribution and a bivariate Gamma distribution. These allow sensitivity to outliers and skewness to be assessed, respectively.

4.4.1. Bivariate t-distribution—Under a bivariate t-distribution, the random errors in the survival and treatment selection equations have the probability density function (PDF) given by

$$\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1,i} \\ \varepsilon_{2,i} \end{pmatrix} \sim t_{df}(\boldsymbol{\mu}_\varepsilon, \boldsymbol{\Sigma}), \text{ where } \boldsymbol{\mu}_\varepsilon = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix}. \quad (12)$$

The smaller the degree-of-freedom (df) the heavier the tails of the PDF in (12). We consider $df=3, 10$, and 30 and evaluate the operating characteristics (bias, MSE, and coverage) of $\hat{\psi}$

when $\rho_{GEN}=0, 0.2$, and 0.5 assuming a uniform $(-1, 1)$ prior for ρ . Other model specifications and simulation settings are as in Section 4.1.

As df increases and the true distribution approaches normality, we find the BSEM estimator of ψ is less biased (Table 8 and Figure 3(c)), more precise and has better coverage (Table 8). Hence, the combination of unmeasured confounding (i.e. under $\rho_{GEN}=0.2$) and assuming a survival time distribution whose tails are not sufficiently thick leads to substantial bias whereas the bias is small regardless of df under no unmeasured confounding.

4.4.2. Mixture of a bivariate normal distribution and a bivariate Gamma

distribution—To isolate the impact of skewness, we generate data from a bivariate normal-gamma mixture distribution having the same mean and covariance as for the bivariate normal model in Section 2.1. That is,

$$\mathbf{e}_i = (e_{1,i}, e_{2,i}) \sim \pi \times N(\mathbf{e}_i; \boldsymbol{\mu}_e, \boldsymbol{\Sigma}) + (1 - \pi) \times G(\mathbf{e}_i; \boldsymbol{\mu}_e, \boldsymbol{\Sigma}), \quad (13)$$

where $\boldsymbol{\mu}_e$ and $\boldsymbol{\Sigma}$ are as in (12), π is the mixing proportion, and $N(\mathbf{e}_i; \boldsymbol{\mu}_e, \boldsymbol{\Sigma})$ denotes the bivariate normal distribution and $G(\mathbf{e}_i; \boldsymbol{\mu}_e, \boldsymbol{\Sigma})$ denotes the bivariate gamma distribution with the mean $\boldsymbol{\mu}_e$ and variance $\boldsymbol{\Sigma}$. Draws from the mixture distribution are made by evaluating $\mathbf{e}_i = \boldsymbol{\Lambda}^T (\pi \mathbf{v}_1 + (1 - \pi) \mathbf{v}_2)$, where $\boldsymbol{\Lambda}$ is the Choleski decomposition of $\boldsymbol{\Sigma}$ (i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}^T \boldsymbol{\Lambda}$), assigning the elements of $\mathbf{v}_1$ to be standard normal random variables, and assigning the elements of $\mathbf{v}_2$ to be univariate gamma random variables with both mean and variance equal to 1. Smaller π allows greater departure from the bivariate normal distribution, generating more skewed data. We consider the six settings of $\rho_{GEN}=0$ and 0.2 by $\pi=0.2, 0.5$, and 0.8 . The other model specifications and simulation settings are as in Section 4.4.1.

As π increases, the BSEM estimator of ψ is less biased (Table 8 and Figure 3(d)), more precise and has better coverage (Table 8). The changes in bias, precision and coverage due to skewness are much more pronounced in the presence of unmeasured confounding ($\rho_{GEN}=0.2$) and are greater than those under no unmeasured confounding ($\rho_{GEN}=0$). However, it was re-assuring to observe that under modest departures from the parametric assumptions (the scenarios in Table 8 other than the greatest departure) the BSEM estimator performed reasonably well.

5. Discussion

In this paper, we have developed a novel Bayesian structural equations model (BSEM) to estimate the causal effect of treatment on survival by jointly modeling survival time and treatment using a simultaneous equations model to account for censoring and unmeasured confounding. The approach assumes an underlying bivariate normal distribution for the log-survival time and propensity to be “treated.” Bayesian MCMC techniques were used for estimation which, for computational efficiency, included treating the censored survival times as unknown parameters to be estimated. The methodology extends comparative effectiveness

research methodology to account for both unmeasured confounding and censoring, whereas almost all prior work has focused on one problem or the other.

In the rAAA data analysis, the BSEM appeared to yield more justifiable results than the AFT model and the 2SLS procedure. It was also robust to different priors for ρ (the unknown selection parameter), obviating the possibility that a subjective prior assumption impacted the results.

The BSEM performed well in finite samples with the Bayesian estimator appearing to be consistent under the model even when the prior for ρ was not centered on the true mean. Other simulations revealed that a stronger IV led to more robust estimation of the casual effect of treatment and that different censoring rates had little impact, implying the model made efficient use of information in the censored survival times. However, when unmeasured confounding was present, the BSEM was sensitive to substantial departures from the bivariate normal distribution particularly in terms of skewness.

The censoring mechanism assumed in this paper is non-informative in that the censoring time had no causal dependence on the potential survival time. The method can in theory be extended to informative censoring by adding an equation for censored time that depends on both treatment and confounders and by allowing unmeasured confounders among the three equations for survival time, censoring time, and treatment selection to be correlated. However, this topic is beyond the current paper.

The selection parameter ρ is easily interpreted as a correlation coefficient. However, estimation of ρ and the other model parameters relies on the bivariate normality assumption, which cannot be conclusively tested, raising concerns that misspecification will result in substantial bias, as evinced under the simulations for the sensitivity of the BSEM to wrongly assumed distributions in Section 4.4. An alternative to full parametric modeling is to assume a Dirichlet process prior (DPP) for ρ (MacEachern and Muller 1998). Although the DPP itself is fundamentally nonparametric, the base distribution of a DPP is typically a standard normal distribution, which may limit flexibility. A uniform based distribution could be considered or, further afield, the Gaussian copula approach of Song (2000) would be an alternative approach.

Another alternative involves accounting for censoring and confounding in sequence, not simultaneously. For example, first use the BSEM solely to multiply impute the censored survival times. Then apply a traditional IV analysis to each completed data set. This method might allow censoring to be taken into account while preserving the robustness of the IV procedure to misspecification of parametric distributions. Therefore, the methodology derived here is not limited to parametric inference as it can be embedded in a procedure that seeks to overcome the censoring and confounding problems in sequence.

Our simulations reinforced the notation that using an IG prior distribution is not problematic for variance parameters at the level of the data - an approximately 100-fold change in the prior parameters (from 0.001 to 0.1) had minimal impact on results. However, because a half- t prior distribution has been recommended in general for variance parameters (Gelman

2006), studying its use on the BSEM is an interesting topic of investigation for a future study.

In summary, the combination of unmeasured confounding and censored survival times presents a challenging scenario to overcome in an observational data analysis. Despite a large quantity of work involving unmeasured confounding and IV methods and a large body of work involving survival or time-to-event analysis, relatively little has been published at the intersection of these areas. The model and analysis in this paper thus presents a novel contribution which we hope will be a catalyst to more work being conducted and results published at the interface of unmeasured confounding and survival time (censored data) analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study is part of the NIH-funded project (1RC4MH092717-01 NIH/NIMH) to develop novel methods for combining the strengths of RCT and observational data.

Appendices

Appendix A. Likelihood

The joint marginal density $f(Y_i, W_i | X_i, Z_i)$ of the i -th subject is obtained by integrating the density of the models in (3) and (5) over W_i^* (i.e. integrating over $\varepsilon_{2,i}$). The likelihood function (6) of $(\psi, \beta, \lambda, \theta, \sigma_1^2, \rho)$ is thus the product over $i = 1, \dots, n$ of:

$$\begin{aligned} f(Y_i^{obs}, \Delta_i, W_i | X_i, Z_i) &= [f(Y_i^{obs} | W_i, X_i, Z_i)]^{\Delta_i} [S(Y_i^{obs} | W_i, X_i, Z_i)]^{1-\Delta_i} f(W_i | X_i, Z_i) \\ &= \int [f(Y_i^{obs} | \varepsilon_{2,i}, W_i, X_i, Z_i)]^{\Delta_i} [S(Y_i | \varepsilon_{2,i}, W_i, X_i, Z_i)]^{1-\Delta_i} f(\varepsilon_{2,i}, W_i | X_i, Z_i) d\varepsilon_{2,i} \\ &= \left[\int f(Y_i^{obs} | \varepsilon_{2,i}, W_i, X_i, Z_i) f(\varepsilon_{2,i}, W_i | X_i, Z_i) d\varepsilon_{2,i} \right]^{\Delta_i} \times \left[\int S(Y_i^{obs} | \varepsilon_{2,i}, W_i, X_i, Z_i) f(\varepsilon_{2,i}, W_i | X_i, Z_i) d\varepsilon_{2,i} \right]^{1-\Delta_i} \\ &= \left[\phi(Y_i^{obs}; \mu_{y,i}, \sigma_1^2 | W_i, X_i, Z_i) \Phi(\mu_{w|y,i} | X_i, Z_i) W_i (1 - \Phi(\mu_{w|y,i} | X_i, Z_i))^{1-W_i} \right]^{\Delta_i} \times [S(Y_i^{obs} | W_i, X_i, Z_i) f(W_i | X_i, Z_i)]^{1-\Delta_i} \end{aligned}$$

Appendix B. Posterior distributions

The conditional posterior distributions discussed in Section 2.3 are provided below.

1) Posterior distribution of Y_i^* (potential survival time)

Assuming a non-informative prior, we derive the conditional posterior distribution of Y_i^* ,

$$p(Y_i^* | \psi, \beta, \lambda, \theta, \sigma_1^2, \rho, C_i) \propto \phi(Y_i^*; \mu_{y^*,i}, \sigma_1^2) \Phi(\mu_{wly^*,i})^{W_i} (1 - \Phi(\mu_{wly^*,i}))^{1-W_i} I(Y_i^* \geq \log(C_i)),$$

where $\mu_{y^*,i} = \psi W_i + \beta^T X_i$ and $\mu_{wly^*,i} = [\lambda Z_i + \theta^T X_i + \rho(Y_i^* - \mu_{y^*,i})/\sigma_1] / \sqrt{1 - \rho^2}$ for $\rho \neq 1$.

Candidate values of Y_i^* are generated using a Metropolis-Hastings (M-H) independence step (Tierney 1994). Instead of the typically-used random walk step, the candidate generating density is the fixed approximation to the true conditional posterior given by

$$p(Y_i^*; \mu_{y^*,i}, \sigma_1^2) \propto \phi(Y_i^*; \mu_{y^*,i}, \sigma_1^2) I(Y_i^* \geq \log(C_i)).$$

2) Posterior distribution of β

Under the improper uniform prior for β (all values of β considered equally likely), the conditional posterior distribution of β is available in closed form, allowing use of a Gibbs Sampling step to generate values of β . The conditional posterior distribution is given by

$$\beta | \sigma_1^2, \tilde{\lambda}, \tilde{\theta}, \tilde{\rho} \sim N((X^T X)^{-1} X^T \tilde{Y}, \sigma_1^2 (X^T X)^{-1}),$$

where X is the matrix with i -th row X_i^T and $\tilde{Y}$ has the i -th element $\tilde{Y}_i = Y_i^a - \psi W_i$, where

$$Y_i^a = Y_i \Delta_i(Y_i^*)^{(1-\Delta_i)}.$$

3) Posterior distribution of σ_1^2

Under the Inverse Gamma, IG (ω_1, ω_2), conjugate prior for σ_1^2 yields the conditional posterior distribution,

$$\sigma_1^2 | \beta, \tilde{\lambda}, \tilde{\theta}, \tilde{\rho} \sim IG\left(\omega_1 + \frac{n}{2}, \omega_2 + \frac{1}{2} \sum_{i=1}^n (\tilde{Y}_i - \beta^T X_i)^2\right)$$

We choose $\omega_1 (= 0.001)$ and $\omega_2 (= 0.001)$ so that the prior is diffuse with large variance ($= \infty$).

4) Posterior distribution of $(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho})$

Assuming non-informative priors for the two transformed parameters $\tilde{\lambda}$ and $\tilde{\theta}$ and the prior given in (10) for $\tilde{\rho}$, we obtain the joint conditional posterior distribution for $(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho})$ given by

$$p(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho} | \beta, \sigma_1^2, \psi) \propto \prod_{i=1}^n \Phi(\mu_{w|y,i})^{W_i} (1 - \Phi(\mu_{w|y,i}))^{1-W_i} \frac{\sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}} \left(1 + \frac{\tilde{\rho} \sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}}\right)^{\nu_1 - 1} \left(1 - \frac{\tilde{\rho} \sigma_1}{\sqrt{1 + \tilde{\rho}^2 \sigma_1^2}}\right)^{\nu_2 - 1}.$$

The covariance parameter used in the vector-normal distribution for updating $(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho})$ is the covariance of the maximum likelihood estimator of $(\tilde{\lambda}, \tilde{\theta}, \tilde{\rho})$ under a normal PDF based on only the non-censored data. After generating each new candidate value, the values of $\tilde{\lambda}$, $\tilde{\theta}$, and $\tilde{\rho}$ are transformed back to the original parameters λ , θ , and ρ for use in the other steps of the MCMC algorithm.

5) Posterior distribution of ψ

Under exchangeability, the conditional posterior distribution of ψ is given by

$$p(\psi | \beta, \sigma_1^2, \tilde{\lambda}, \tilde{\theta}, \tilde{\rho}) \propto \phi(Y_i^a; \mu_{y,i}^a, \sigma_1^2) \Phi(\mu_{w|y,i}^a)^{W_i} (1 - \Phi(\mu_{w|y,i}^a))^{1-W_i},$$

where $Y_i^a = Y_i^{\Delta_i} (Y_i^*)^{(1-\Delta_i)}$ and $\mu_{y,i}^a$ and $\mu_{w|y,i}^a$ have Y_i^a in place of Y_i in (7). A M-H step with a normal distribution as the random-walk candidate generating function is used for updating ψ .

References

- Abbring JH, van den erg GJ. 2005 Social experiments and instrumental variables with duration outcomes. Working Paper, IFAU - Institute for Labour Market Policy Evaluation, No 2005. :11.
- Angrist JD, Imbens GW, Rubin DR. 1996; Identification of Causal Effects Using Instrumental Variables. *Journal of American Statistical Association*. 91:444–472.
- Baker SG. 1998; Analysis of Survival Data from a Randomized Trial with All-or-None Compliance: Estimating the Cost-Effectiveness of a Cancer Screening Program. *Journal of American Statistical Association*. 93(443):929–934.
- Bijwaard GE. 2009 Instrumental Variable Estimation for Duration Data. Tinbergen Institute Discussion Paper.
- Bhattacharya J, Goldman D, McCaffrey D. 2006; Estimating Probit Models with Self-selected Treatments. *Statistics in Medicine*. 25(3):389–413. [PubMed: 16382420]
- Blundell R, Powell JL. 2007; Censored regression quantiles with endogenous regressors. *Journal of Econometrics*. 141:65–83.
- Bosco JLF, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, Yood MU, Haque R, Wei F, Lash TL. 2010; A Most Stubborn Bias: No Adjustment Method Fully Resolves Confounding by Indication in Observational Studies. *Journal of Clinical Epidemiology*. 63(1):64–74. [PubMed: 19457638]
- Brookhart MA, Schneeweiss S. 2007; Preference-based Instrumental Variable Methods for the Estimation of Treatment Effects: Assessing Validity and Interpreting Results. *The International Journal of Biostatistics*. 3(1):14.

- Chernozhukov V, Fernandez-Val I, Kowalski K. 2011 Quantile Regression with Censoring and Endogeneity. Cowels Foundation Discussion Paper, No 1797.
- Chib S, Hamilton BH. 2000; Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics*. 97:25–50.
- Chib S, Hamilton BH. 2002; Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics*. 110:67–89.
- Copas JB, Li HG. 1997; Inference for Non-random Samples. *Journal Royal Statistical Society B*. 59:55–77.
- Gelman A, Meng XL, Stern H. 1996; Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*. 6:733–760.
- Goldman DP, Bhattacharya J, McCaffrey DF, Duan N, Leibowitz AA, Joyce GF, Morton SC. 2001; Effect of Insurance on Mortality in an HIV-positive Population in Care. *Journal of American Statistical Association*. 96(455):883–894.
- Gore JL, Litwin MS, Lai J, Yano EM, Madison R, Setodji C, Adams JL, Saigal CS. 2010; Use of Radical Cystectomy for Patients With Invasive Bladder Cancer. *Journal of the National Cancer Institute*. 102(11):802–811. [PubMed: 20400716]
- Imbens GW, Angrist JD. 1994; Identification and Estimation of Local Average Treatment Effects. *Econometrica*. 62(2):467–475.
- Joffe MM. 2001; Administrative and artificial censoring in censored regression models. *Statistics in Medicine*. 20:2287–2304. [PubMed: 11468764]
- Heckman JJ. 1978; Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica*. 46:931–960.
- Hernan MA, Hernandez-Diaz S, Robins JM. 2004; A Structural Approach to Selection Bias. *Epidemiology*. 15:615–625. [PubMed: 15308962]
- Li J, Fine J, Brookhart A. 2015; Instrumental Variable Additive Hazards Models. *Biometrics*. 71:122–130. [PubMed: 25298257]
- Loeys T, Goetghebeur E. 2003; A Causal Proportional Hazards Estimator for the Effect of Treatment Actually Received in a Randomized Trial with All-or-Nothing Compliance. *Biometrics*. 59:100–105. [PubMed: 12762446]
- MacEachern SN, Miller P. 1998; Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics*. 7:223–238.
- Nie H, Cheng J, Small DS. 2011; Inference for the Effect of Treatment on Survival Probability in Randomized Trials with Noncompliance and Administrative Censoring. *Biometrics*. 67:1397–1405. [PubMed: 21385167]
- O'Malley AJ, Cotterill P, Schermerhorn ML, Landon BE. 2011a; Improving Observational Study Estimates of Treatment Effects Using Joint Modeling of Selection Effects and Outcomes: the Case of AAA Repair. *Medical Care*. 49(12):1126–1132. [PubMed: 22011709]
- O'Malley AJ, Frank RG, Normand SL. 2011b; Estimating Cost-offsets of New Medications: Use of New Antipsychotics and Mental Health Costs for Schizophrenia. *Statistics in Medicine*. 30(16):1971–1988. [PubMed: 21520217]
- Palmer, C. Why Did So Many Subprime Borrowers Default During the Crisis: Loose Credit or Plummeting Prices?. Job Market Paper. 2013. http://web.mit.edu/cjpalmer/www/CPalmer_JMP.pdf
- Robins JM. 2005; Structural Nested Failure Time Models. *Encyclopedia of Biostatistics*. doi: 10.1002/0470011815.b2a11071
- Robins JR, Tsiatis AA. 1991; Correcting for Non-compliance in Randomized Trials Using Rank Preserving Structural Failure Time Models. *Communications in Statistics: Theory and Methods*. 20:2609–2631.
- Rubin D. 1978; Bayesian Inference for Causal Effects. *The Annals of Statistics*. 6:34–58.
- Song PXX. 2000; Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*. 27:305–320.

- Song X, Wang CY. 2014; Proportional Hazards Model With Covariate Measurement Error and Instrumental Variables. *Journal of American Statistical Association*. 109(508):1636–1646. DOI: 10.1080/01621459.2014.896805
- Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, Martinussen T, Glymour M. 2015; Instrumental Variable Estimation in a Survival Context. *Epidemiology*. 26:402–410. [PubMed: 25692223]
- Terza J, Basu A, Rathouz P. 2008; Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of Health Economics*. 27:531–543. [PubMed: 18192044]
- Tierney L. 1994; Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*. 22:1701–1728.
- Yu M, Xie D, Wang X, Weiner MG, Tannen RL. 2012; Prior Event Rate Ratio Adjustment: Numerical Studies of a Statistical Method to Address Unrecognized Confounding in Observational Studies. *Pharmacoepidemiology and Drug Safety*. 21(S2):60–68. [PubMed: 22552981]
- Zeng F, O'Leary JF, Sloss EM, Lopez MS, Dhanani N, Melnick G. 2006; The Effect of Medicare Health Maintenance Organizations on Hospitalization Rates for Ambulatory Care-sensitive Conditions. *Medical Care*. 44(10):900–907. [PubMed: 17001260]

Fig. 1.

Density plot of $\log(\text{estimated survival time} - \text{censoring time})$. Solid (—) and dotted (- -) lines denote treatment=1 and 0, respectively, and vertical lines indicate the average $\log(\text{true survival time} - \text{censoring time})$ for censored observations within each treatment group. These assume $n = 5000$, censoring rate = 30%, number of MCMC posterior samples = 40000, $\psi = 2$, $\rho = 0.2$, and a vague prior for ρ .

Fig. 2.

Posterior predictive checks (PPC): Density plots from one replicated set of the BSEM posterior predictive values for the Medicare patients who received endo or open repair for the treatment of rAAA in years of 2005 and 2008 with the largest and the smallest group of patients, respectively. Solid (—) and dashed (- -) lines denote observed and predictive values, respectively.

Fig. 3.

Bias plots of treatment effect (ψ) in simulation studies. Solid (—), dashed (---) and dotted (···) lines denote vague, medium and precise prior of ρ in (a) and $\rho_{GEN}=0, 0.2$ and 0.5 in (b), respectively. Solid (—) and dashed (---) lines denote $\rho_{GEN}=0$ and 0.2 , respectively, in (c) and (d).

Fig. 4.

Density of posterior estimates of ρ with vague prior mean=0. Dotted (---), dashed (---) and solid (—) lines denote true value of correlation, $\rho_{GEN}=0, 0.2$ and 0.5 , respectively, and vertical lines indicate the average posterior estimate of ρ for the corresponding value of ρ_{GEN} . These assume $n = 5000$, censoring rate = 30%, number of MCMC posterior samples = 40000.

Table 1
Categorical characteristics of the Medicare patients who received endovascular (endo) or open surgical repair (open) for ruptured abdominal aortic aneurysm (rAAA) during 2001-2008

Variable	Level (%)	AAA repair			Censorship	
		Endo	Open	P*	Uncensored	Censored
AAA procedure	Endo	100.00	0		76.20	77.63
	Open	0	100.00		23.80	22.37
Race (Ref=Hispanic+Others)	White	93.87	93.25	.8219	93.66	93.76
	Black	3.77	4.29		3.85	3.91
Gender (Ref=Male)	Female	19.40	25.77	.0004	22.77	19.88
Age at procedure (Ref=85+)	65-69	9.27	15.80	<.0001	7.80	12.27
	70-74	26.40	30.83		20.79	30.78
	75-79	28.44	31.60		25.88	30.83
	80-84	23.99	17.18		30.56	18.30
Year at procedure (Ref=2008)	2001	10.77	24.39	<.0001	22.56	9.47
	2002	12.22	17.48		16.42	11.90
	2003	12.95	17.18		15.59	13.06
	2004	13.49	12.58		14.24	12.80
	2005	16.27	8.90		12.27	15.76
	2006	13.36	7.36		7.90	14.07
	2007	12.13	6.75		7.48	12.64
Urgent admission	Yes	4.09	8.59	<.0001	6.96	4.18
Prior AAA diagnosis w/o rupture	Yes	74.83	64.57	<.0001	72.45	72.50
Renal failure	Yes	6.45	5.37	.3138	8.32	5.13
Congestive heart failure	Yes	14.31	10.43	.0106	18.92	10.63
Chronic pulmonary disease	Yes	26.12	23.62	.1977	31.29	22.63
Peripheral vascular disorders	Yes	20.22	19.02	.5007	22.97	18.40
Vascular disease	Yes	11.13	7.82	.0149	11.54	9.78
Neuro vascular disease	Yes	12.18	11.96	.8836	13.10	11.63

* : Chi-square tests were used to evaluate p-values.

Note: Column shows distributions of variables.

Table 2
Continuous characteristics of the Medicare patients who received endovascular (endo) or open surgical repair (open) for ruptured abdominal aortic aneurysm (rAAA) during 2001-2008

Variable	AAA repair		P*	Censorship	
	Endo Mean(SD)	Open Mean(SD)		Uncensored Mean(SD)	Censored Mean(SD)
Facility total volume of AAA procedures over prior 365 days	51.02(45.28)	38.58(41.08)	<.0001	45.85(43.11)	49.37(45.39)
Facility proportion of endo cases over prior 365 days (IV)	.59(.21)	.32(.25)	<.0001		

* : T-tests were used to evaluate p-values.
Note: Column shows distributions of variables.

Table 3

Results of estimates of ψ , the effect of endo versus open repair for the treatment of rAAA, in the Medicare population over 2001–2008.

Censored data	Procedure	Treatment effect (ψ)		
		Estimate	Std Err	Interval
Exclude	AFT model	.230	.060	(.112, .347)
	2SLS	.213	.135	(-.052, .477)
	BSEM	.283	.135	(.034, .577)
	BSEM w/o IV	1.168	.189	(.793, 1.529)
Include	AFT model	.076	.063	(-.048, .200)
	BSEM	-.112	.129	(-.353, .145)
	BSEM w/o IV	-.122	.238	(-.529, .333)

Note: The BSEM used a vague extended Beta prior with parameters=(1,1) for ρ (i.e. $E[\rho]=0$ and $\text{Var}[\rho]=.33$) over a range (-1,1) which is same as uniform (-1, 1).

Table 4

Results of the BSEM estimates of treatment effect (ψ) and correlation (ρ) by omitting/including 'urgent admission' among observed confounders for all rAAA data in the Medicare population over 2001–2008.

Confounders	Parameter	Estimate	Std Err	Interval
Include all observed confounders	ψ	-.112	.129	(-.353, .145)
	ρ	.119	.071	(-.014, .259)
Omit an observed confounder (urgent admission)	ψ	-.135	.144	(-.436, .133)
	ρ	.145	.080	(-.009, .309)

Note: The BSEM used a vague extended Beta prior with parameters=(1,1) for ρ (i.e. $E[\rho]=0$ and $\text{Var}[\rho]=.33$) over a range (-1,1) which is same as uniform (-1, 1).

Table 5

Results of the BSEM estimates of treatment effect (ψ) and correlation (ρ) for all rAAA data in the Medicare population over 2001–2008.

Parameter	Prior of ρ	Estimate	Std Err	Interval
ψ	Vague	-.112	.129	(-.353, .145)
	Medium	-.094	.135	(-.367, .175)
	Precise	-.067	.130	(-.327, .176)
ρ	Vague	.119	.071	(-.014, .259)
	Medium	.108	.075	(-.044, .254)
	Precise	.092	.071	(-.042, .234)

Note: Vague, medium, and precise priors for ρ are Beta-type distributions with parameters=(1,1), (5,5) and (10,10), respectively. Thus, $E[\rho] = 0$ in each case and $\text{Var}[\rho] = .33, .09$, and $.05$, respectively, over a range (-1,1).

Operating characteristics of the Bayesian estimator (posterior mean) of the treatment effect (ψ) under different sample sizes when the data are generated based on ρ_{GEN} but the analysis is performed using different priors for ρ (with prior mean=0) that do not depend on ρ_{GEN} .

Table 6

Sample size (n)	Prior of correlation (ρ)	Treatment effect (true ψ = -.5)	True correlation (ρ_{GEN})				
			-.5	-.2	0	.2	.5
1000	Vague	Bias	-.128	-.063	.021	.060	.142
		$\sqrt{MSE}$	.481	.549	.549	.546	.493
		Coverage (%)	97	97	92	98	97
	Medium	Bias	-.263	-.150	-.015	.113	.317
		$\sqrt{MSE}$	.457	.434	.377	.396	.481
		Coverage (%)	87	95	98	100	89
	Precise	Bias	-.382	-.166	.021	.161	.378
		$\sqrt{MSE}$	.474	.326	.286	.323	.470
		Coverage (%)	60	99	97	100	64
5000	Vague	Bias	-.029	.010	-.011	.011	.020
		$\sqrt{MSE}$	.195	.268	.316	.261	.211
		Coverage (%)	95	91	84	94	95
	Medium	Bias	-.069	-.010	.000	.026	.097
		$\sqrt{MSE}$	.200	.235	.256	.235	.223
		Coverage (%)	97	93	91	96	95
	Precise	Bias	-.148	-.076	.000	.074	.148
		$\sqrt{MSE}$	.238	.225	.210	.210	.242
		Coverage (%)	84	91	99	96	85

Note: Vague, medium, and precise priors for ρ are Beta-type distributions with parameters=(1,1), (5,5) and (10,10), respectively. Thus, $E[\rho] = 0$ in each case and $\text{Var}[\rho] = .33$, .09, and .05, respectively.

Operating characteristics of the Bayesian estimator (posterior mean) of the treatment effect (ψ) as a function of the strength of the IV (λ)

Table 7

Sample size (n)	True correlation (ρ_{GEN})	Treatment effect (true $\psi = -.5$)	IV effect (λ)				
			0	.5	1.0	2.0	
1000	0	Bias	.043	.021	-.025	.002	
		$\sqrt{MSE}$	.692	.549	.347	.176	
		Coverage (%)	93	92	92	93	
	.2	Bias	.293	.060	-.005	.015	
		$\sqrt{MSE}$	.750	.546	.304	.161	
		Coverage (%)	91	98	97	97	
	.5	Bias	.426	.142	.012	-.030	
		$\sqrt{MSE}$	.762	.493	.260	.163	
		Coverage (%)	93	97	96	96	
5000	0	Bias	.050	-.011	-.015	.003	
		$\sqrt{MSE}$	.483	.316	.148	.074	
		Coverage (%)	87	84	93	95	
	.2	Bias	.338	.011	.012	-.005	
		$\sqrt{MSE}$	.599	.261	.140	.073	
		Coverage (%)	77	94	95	99	
	.5	Bias	.386	.020	.001	-.014	
		$\sqrt{MSE}$	.609	.211	.125	.077	
		Coverage (%)	78	95	93	94	

Note: The vague prior for ρ with $E[\rho] = 0$ and $\text{Var}[\rho] = .33$ is used.

Table 8
Results of estimates of treatment effect (ψ) from the simulations for sensitivity to model misspecification with two true distributions: (A) bivariate t-distribution and (B) the mixture of bivariate normal and bivariate gamma distributions

(A) True distribution: bivariate t-distribution									
Sample size (n)	Treatment effect (true $\psi = -.5$)	True correlation ($\rho_{GEN} = 0$) df of true biv. t-dist			True correlation ($\rho_{GEN} = .2$) df of true biv. t-dist				
		3	10	30	3	10	30	10	30
1000	Bias	-.070	-.112	.031	-.305	-.083	.028		
	$\sqrt{MSE}$	.759	.550	.542	.687	.550	.527		
	Coverage (%)	55	89	95	49	88	91		
	Bias	.014	-.034	-.015	-.293	-.073	-.008		
5000	$\sqrt{MSE}$	.637	.287	.254	.674	.291	.281		
	Coverage (%)	57	87	95	32	89	91		
(B) True distribution: mixture of bivariate normal and bivariate gamma distributions									
Sample size (n)	Treatment effect (true $\psi = -.5$)	True correlation ($\rho_{GEN} = 0$)			True correlation ($\rho_{GEN} = .2$)				
		Prop. of normal (π)			Prop. of normal (π)				
		.2	.5	.8	.2	.5	.8		
1000	Bias	-.166	-.017	.007	-.438	-.129	.064		
	$\sqrt{MSE}$	.528	.513	.563	.661	.530	.560		
	Coverage (%)	84	93	91	54	86	92		
	Bias	-.111	-.022	-.007	-.573	-.138	-.001		
5000	$\sqrt{MSE}$	.333	.272	.278	.630	.349	.261		
	Coverage (%)	82	89	89	19	78	93		

Note: The vague prior for ρ with $E[\rho] = 0$ and $\text{Var}[\rho] = .33$ is used in both (A) and (B).