Received: 4 August 2019 | Revised: 30 May 2020 | Accepted: 3 June 2020

DOI: 10.1002/sim.8672

RESEARCH ARTICLE

Statistics in Medicine WILEY

Sensitivity analysis of treatment effect to unmeasured confounding in observational studies with survival and competing risks outcomes

Rong Huang[1] | **Ronghui Xu**[1,2] | **Parambir S. Dulai**[3]

[1]Department of Mathematics, University of California San Diego, La Jolla, California

[2]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California

[3]Department of Medicine, University of California San Diego, La Jolla, California

Correspondence
Ronghui Xu, Department of Mathematics, University of California San Diego, 9500 Gilman Drive, Mail code 0112, La Jolla, CA 92093, USA.
Email: rxu@health.ucsd.edu

Funding information
National Institutes of Health, Grant/Award Number: UL1TR001442 of CTSA

No unmeasured confounding is often assumed in estimating treatment effects in observational data, whether using classical regression models or approaches such as propensity scores and inverse probability weighting. However, in many such studies collection of confounders cannot possibly be exhaustive in practice, and it is crucial to examine the extent to which the resulting estimate is sensitive to the unmeasured confounders. We consider this problem for survival and competing risks data. Due to the complexity of models for such data, we adapt the simulated potential confounder approach of Carnegie et al (2016), which provides a general tool for sensitivity analysis due to unmeasured confounding. More specifically, we specify one sensitivity parameter to quantify the association between an unmeasured confounder and the exposure or treatment received, and another set of parameters to quantify the association between the confounder and the time-to-event outcomes. By varying the magnitudes of the sensitivity parameters, we estimate the treatment effect of interest using the stochastic expectation-maximization (EM) and the EM algorithms. We demonstrate the performance of our methods on simulated data, and apply them to a comparative effectiveness study in inflammatory bowel disease. An R package "survSens" is available on CRAN that implements the proposed methodology.

KEYWORDS

causal inference, Cox model, expectation-maximization algorithm, inverse probability weighting, proportional hazards regression, regression adjustment, simulated confounder, stochastic EM

1 | INTRODUCTION

One widely used yet untestable assumption when analyzing data from observational studies is that there is no unobserved confounding, which means that the treatment received and the potential outcomes are independent conditional on the observed pretreatment covariates. Sensitivity analysis offers an approach to assess the extent to which the inference is robust to violation of this assumption. Rosenbaum[1] contains a nice introduction describing the idea based on association between the unobserved confounder and the treatment, and between the unobserved confounder and the outcome. Analytical approaches have been developed for simpler outcomes such as binary,[2] as well as for survival outcomes under the assumption that the event is rare or the effect of the unmeasured confounder on the survival time is small.[3] Li et al[4] and

Shen et al[5] considered sensitivity analysis methods for inverse probability weighted (IPW) estimators using propensity scores, an approach that was gaining popularity in practice.[6]

Our motivation came from studies in inflammatory bowel disease (IBD). IBD is an umbrella term for two conditions, ulcerative colitis (UC) and Crohn's disease (CD), that are characterized by chronic inflammation of the gastrointestinal tract.[7] With rapid growth in treatment options, head-to-head comparisons are entirely lacking due to difficulty in performing randomized clinical trials. In order to compare the effectiveness between Vedolizumab and tumor necrosis factor (TNF)-antagonist therapies for UC and CD patients, data were collected between May 2014 and December 2017 from a North American based consortium registry,[8,9] which is a multicenter collaborative research group where data are pooled for consecutive UC and CD patients treated with biologics. Our primary endpoint is time from treatment initiation to clinical remission. Although data collection was rather extensive and accounted for most known measurable confounders, treatment selection for IBD is known to be preference sensitive and influenced by patient and provider perceptions, experiences, and understandings of potential benefit and risk based on the data available to them, all of which are unmeasurable. We aim to assess to what extent our inference from the data is affected by potentially unmeasured confounding.

Time from treatment initiation to clinical remission may be observed, or may be censored by the patient switching to surgical management. Wide variability exists across centers, patients, and providers, for their preference to proceed with surgery while awaiting response to therapy. Therefore, surgery presents a competing risk to clinical remission, in that surgery prevents the event of achieving clinical remission. In Lukin et al[10] and Bohm et al[11] the authors considered propensity score methods with IPW as the primary approach to account for the observed covariates. However, it is possible that there might be confounders not captured by the observed covariates. To carry out sensitivity analysis for this type of complex outcomes, we found the simulated unobserved confounder approach[12] to be useful, in particular since the analytical approaches seem difficult to derive for competing risks.

The article is organized as follows. We describe our models in Section 2, including both the survival models and the competing risks models. We consider estimation in Section 3, using both the expectation-maximization (EM) algorithm and a stochastic EM algorithm. In Section 4, we demonstrate the performance of our algorithms via simulations. We apply our methods to the IBD data in Section 5. Finally, we conclude with discussion in Section 6.

2 | MODELS

2.1 | Survival outcomes

Denote T^0 a time-to-event outcome, Z a binary treatment indicator, and $\boldsymbol{X}$ a vector of observed covariates. Due to possible right censoring, we observe $T = \min(T^0, C)$ and $\delta = I(T^0 \leq C)$, where C is the censoring time, and $I(\cdot)$ the indicator function. We consider U which represents the portion of unmeasured confounder(s) that is independent of $\boldsymbol{X}$, and will simply refer to U as the unmeasured confounder for the rest of the article. We assume U to be binary for ease of implementation, although other distributions are possible and will be discussed later. Given Z, $\boldsymbol{X}$ and U, the hazard rate of T^0 is modeled using the Cox proportional hazards (PH) regression:[13]

$$\lambda(t|Z, \boldsymbol{X}, U) = \lambda_0(t)\exp(\tau Z + \boldsymbol{X}'\boldsymbol{\beta} + \zeta U), \tag{1}$$

where $\lambda_0(\cdot)$ is the baseline hazard function, and τ, $\boldsymbol{\beta}$, and ζ are the regression coefficients. In addition, we assume that given $\boldsymbol{X}$ and U, Z follows a generalized linear model; for illustration purposes we assume a probit link below, although logistic would be an obvious alternative:

$$\mathbb{P}(Z = 1|\boldsymbol{X}, U) = \Phi(\boldsymbol{X}'\boldsymbol{\beta}_{\boldsymbol{z}} + \zeta_z U), \tag{2}$$

where Φ is the standard normal cumulative distribution function, and $\boldsymbol{\beta}_{\boldsymbol{z}}$ and ζ_z are the regression coefficients. In the above ζ_z and ζ are sensitivity parameters, which quantify the relationships between the unobserved confounder and the treatment received and the outcome, respectively. Finally, we assume that $U \sim$ Bernoulli(π), and we set $\pi = 0.5$.

Our goal is to simulate U given the observed T, δ, Z, and $\boldsymbol{X}$. We note that if the parameters in the above models are known, then

$$U|T,\delta,Z,\boldsymbol{X} \sim \text{Bernoulli}\left(\frac{\mathbb{P}(T,\delta,Z,U=1|\boldsymbol{X})}{\mathbb{P}(T,\delta,Z|\boldsymbol{X})}\right), \tag{3}$$

where $\mathbb{P}(T,\delta,Z,U=u|\boldsymbol{X})$ is the joint probability of $(T,\delta,Z,U=u)$ given $\boldsymbol{X}$ for $u=0,1$, and $\mathbb{P}(T,\delta,Z|\boldsymbol{X}) = \mathbb{P}(T,\delta,Z,U=1|\boldsymbol{X}) + \mathbb{P}(T,\delta,Z,U=0|\boldsymbol{X})$. In particular,

$$\begin{aligned}\mathbb{P}(T,\delta,Z,U|\boldsymbol{X}) &= \mathbb{P}(U|\boldsymbol{X})\cdot\mathbb{P}(Z|\boldsymbol{X},U)\cdot\mathbb{P}(T,\delta|Z,\boldsymbol{X},U)\\ &= \pi^U(1-\pi)^{1-U}\{\Phi(\boldsymbol{X}'\boldsymbol{\beta_z}+\zeta_z U)\}^Z\{1-\Phi(\boldsymbol{X}'\boldsymbol{\beta_z}+\zeta_z U)\}^{1-Z}\\ &\quad\cdot\{\lambda_0(T)e^{\tau Z+\boldsymbol{X}'\boldsymbol{\beta}+\zeta U}\}^{\delta}\exp\{-\Lambda_0(T)\cdot e^{\tau Z+\boldsymbol{X}'\boldsymbol{\beta}+\zeta U}\}.\end{aligned} \tag{4}$$

Expression (3) will be used to simulate U_i given the observed T_i, δ_i, Z_i, and $\boldsymbol{X}_i$, where $(T_i,\delta_i,Z_i,\boldsymbol{X}_i)$ for $i=1,\ldots,n$ are independent and identically distributed (i.i.d.) from the distribution of $(T,\delta,Z,\boldsymbol{X})$.

2.2 | Competing risks

In the presence of competing risks, when an event occurs it may be one of m distinct types of failures indexed by $j=1,2,\ldots,m$. Again denote T^0 the time-to-event, Z a binary treatment indicator, and $\boldsymbol{X}$ a vector of observed covariates. We observe $T=\min(T^0,C)$, $\delta=I(T^0\le C)$, J the type of failure if $\delta=1$ and $J=0$ otherwise for notational purpose. We again consider an unmeasured binary confounder U that is independent of $\boldsymbol{X}$. The cause-specific hazard function[14] for the jth failure type is $\lambda_j(t|Z,\boldsymbol{X},U)=\lim_{\Delta t\to 0}\mathbb{P}(t\le T^0<t+\Delta t, J=j|T^0\ge t,Z,\boldsymbol{X},U)/\Delta t$. Using the PH modeling of the cause-specific hazard function, we have:

$$\lambda_j(t|Z,\boldsymbol{X},U)=\lambda_{j0}(t)\exp(\tau_j Z+\boldsymbol{X}'\boldsymbol{\beta_j}+\zeta_j U),\quad j=1,2,\ldots,m. \tag{5}$$

where $\lambda_{j0}(\cdot)$ is the baseline hazard function for type j, and $(\tau_j,\boldsymbol{\beta_j},\zeta_j)$ are the regression coefficients. The interpretation of treatment effect in the presence of competing risks needs caution, and is further discussed in the data analysis section. As before we also assume that given $\boldsymbol{X}$ and U, Z follows a generalized linear model (2) with a probit link. Then parallel to (4) we have the joint probability of (T,δ,J,Z,U) given $\boldsymbol{X}$ as Reference 14

$$\begin{aligned}\mathbb{P}(T,\delta,J,Z,U|\boldsymbol{X}) &= \pi^U(1-\pi)^{1-U}\{\Phi(\boldsymbol{X}'\boldsymbol{\beta_z}+\zeta_z U)\}^Z\{1-\Phi(\boldsymbol{X}'\boldsymbol{\beta_z}+\zeta_z U)\}^{1-Z}\\ &\quad\cdot\prod_{j=1}^{m}\{\lambda_{j0}(T)e^{\tau_j Z+\boldsymbol{X}'\boldsymbol{\beta}_j+\zeta_j U}\}^{I(\delta=1,J=j)}\exp\{-\Lambda_{j0}(T)\cdot e^{\tau_j Z+\boldsymbol{X}'\boldsymbol{\beta}_j+\zeta_j U}\},\end{aligned} \tag{6}$$

where $I(\delta=1,J=j)$ indicates whether subject had the event j. The posterior probability of U is then obtained similar to (3). In general, if there are m distinct types of failures, there would be $m+1$ sensitivity parameters, $\zeta_z,\zeta_1,\ldots,\zeta_m$.

3 | ESTIMATION

In order to simulate U given the observed data, we first need to estimate the unknown parameters. As before denote n the number of subjects. Conditional on the unobserved U as well as Z and $\boldsymbol{X}$, the likelihood function of the survival outcome without competing risks is

$$\begin{aligned}L_1(\tau,\boldsymbol{\beta},\zeta;T,\delta|Z,\boldsymbol{X},U) &= \prod_{i=1}^{n}\lambda(t_i|z_i,\boldsymbol{x_i},u_i)^{\delta_i}\exp\{-\Lambda(t_i|z_i,\boldsymbol{x_i},u_i)\}\\ &= \prod_{i=1}^{n}\{\lambda_0(t_i)e^{\tau z_i+\boldsymbol{x_i}'\boldsymbol{\beta}+\zeta u_i}\}^{\delta_i}\exp\{-\Lambda_0(t_i)e^{\tau z_i+\boldsymbol{x_i}'\boldsymbol{\beta}+\zeta u_i}\}.\end{aligned} \tag{7}$$

Similarly, the likelihood function of the competing risks outcome is

$$L_1(\tau, \boldsymbol{\beta_j}, \zeta_j; T, \delta, J|Z, \boldsymbol{X}, U) = \prod_{i=1}^{n}\prod_{j=1}^{m} \lambda_j(t_i|z_i, \boldsymbol{x_i}, u_i)^{\delta_{ij}} \exp\{-\Lambda_j(t_i|z_i, \boldsymbol{x_i}, u_i)\}$$
$$= \prod_{i=1}^{n}\prod_{j=1}^{m} \{\lambda_{j0}(t_i) e^{\tau_j z_i + \boldsymbol{x_i}'\boldsymbol{\beta}_j + \zeta_j u_i}\}^{\delta_{ij}} \exp\{-\Lambda_{j0}(t_i) e^{\tau_j z_i + \boldsymbol{x_i}'\boldsymbol{\beta}_j + \zeta_j u_i}\}, \quad (8)$$

where $\delta_{ij} := I(\delta_i = 1, J_i = j)$ indicates whether subject i had event j.

3.1 | The EM algorithm

The EM algorithm[15] is a commonly used approach to handle missing data, in this case U, in the likelihood function. Let $\boldsymbol{\theta}$ denote the unknown parameters, and y_i the survival outcome for subject i. The EM algorithm iterates between the E-steps and the M-steps that are described below, where in the notation the covariate $\boldsymbol{x_i}$ is suppressed which is always being conditioned upon. The initial values can be set using the parameter estimates from the regression models ignoring U. We note that the sensitivity parameters, as well as $\pi = 0.5$, are known.

E-step

In the E-step we compute the conditional expectation of the log-likelihood of the complete data (y_i, z_i, u_i) given the observed data and the current parameter value $\tilde{\boldsymbol{\theta}}$. For the survival outcome without competing risks, let

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \mathbb{E}[l(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}, \mathbf{u})|\mathbf{y}, \mathbf{z}, \tilde{\boldsymbol{\theta}}] \\ &= \mathbb{E}[l_1(\boldsymbol{\beta}, \tau, \lambda_0; \mathbf{y}|\mathbf{z}, \mathbf{u})|\mathbf{y}, \mathbf{z}, \tilde{\boldsymbol{\theta}}] + \mathbb{E}[l_2(\boldsymbol{\beta_z}; \mathbf{z}|\mathbf{u})|\mathbf{y}, \mathbf{z}, \tilde{\boldsymbol{\theta}}] + \mathbb{E}[l_3(\mathbf{u})|\mathbf{y}, \mathbf{z}, \tilde{\boldsymbol{\theta}}] \\ &:= Q_1(\boldsymbol{\beta}, \tau, \lambda_0) + Q_2(\boldsymbol{\beta_z}) + Q_3, \end{aligned} \quad (9)$$

where

$$Q_1(\boldsymbol{\beta}, \tau, \lambda_0) = \sum_{i=1}^{n} \big[\delta_i \{\log \lambda_0(t_i) + \boldsymbol{x_i}'\boldsymbol{\beta} + \zeta \mathbb{E}[u_i|y_i, z_i, \tilde{\boldsymbol{\theta}}] + \tau z_i\}$$
$$- \Lambda_0(t_i) \exp\{\boldsymbol{x_i}'\boldsymbol{\beta} + \log \mathbb{E}[e^{\zeta u_i}|y_i, z_i, \tilde{\boldsymbol{\theta}}] + \tau z_i\}\big], \quad (10)$$

$$Q_2(\boldsymbol{\beta_z}) = \sum_{i=1}^{n} \{z_i \mathbb{E}[\log(\Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z} + \zeta_z u_i))|y_i, z_i, \tilde{\boldsymbol{\theta}}] + (1 - z_i)\mathbb{E}[\log(1 - \Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z} + \zeta_z u_i))|y_i, z_i, \tilde{\boldsymbol{\theta}}]\}, \quad (11)$$

$$Q_3 = \sum_{i=1}^{n} \{\log \pi \mathbb{E}[u_i|y_i, z_i, \tilde{\boldsymbol{\theta}}] + \log(1 - \pi)\mathbb{E}[1 - u_i|y_i, z_i, \tilde{\boldsymbol{\theta}}]\}. \quad (12)$$

We note that Q_3 is in fact not used in the M-step since it does not involve unknown parameters. As described earlier, given the observed data, U follows Bernoulli$(\tilde{\pi}_i)$ as in (3) where $\tilde{\pi}_i$ is calculated based on the current parameter value $\tilde{\boldsymbol{\theta}}$. So for any function $h(u_i)$ in (10) and (11), we have $\mathbb{E}[h(u_i)|y_i, z_i, \tilde{\boldsymbol{\theta}}] = h(1)\tilde{\pi}_i + h(0)(1 - \tilde{\pi}_i)$.

For competing risks outcome, from (8) we see that the likelihood function is a product of m likelihoods, one for each type of event with its own type specific parameters. The corresponding Q_1 function is then a sum of $Q_{1j}(\boldsymbol{\beta_j}, \tau_j, \lambda_{j0})$'s, each having the same form as $Q_1(\boldsymbol{\beta}, \tau, \lambda_0)$ above but with parameters $\boldsymbol{\beta_j}$, τ_j, λ_{j0} and data for the event type j instead.

M-step

From (9) it is clear that in the M-step we can update $(\boldsymbol{\beta}, \tau, \lambda_0)$ and $\boldsymbol{\beta_z}$ separately. In order to maximize Q_1, we note that it has the same form as the log-likelihood in a Cox regression model with known offset $\log \mathbb{E}[e^{\zeta u_i}|y_i, z_i, \tilde{\boldsymbol{\theta}}]$, just like the Cox model with random effects.[16] For competing risks again because Q_1 is a sum of $Q_{1j}(\boldsymbol{\beta_j}, \tau_j, \lambda_{j0})$'s for $j = 1, \ldots, m$, each set of

parameters $\boldsymbol{\beta_j}, \tau_j, \lambda_{j0}$ is updated separately using the Cox model software with offsets, the same way as a single survival outcome.

To maximize Q_2, we have

$$Q_2(\boldsymbol{\beta_z}) = \sum_{i=1}^{n} \big(z_i[\log\{\Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z} + \zeta_z)\}\tilde{\pi}_i + \log\{\Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z})\}(1-\tilde{\pi}_i)] + (1-z_i)[\log\{1-\Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z} + \zeta_z)\}\tilde{\pi}_i + \log\{1-\Phi(\boldsymbol{x_i}'\boldsymbol{\beta_z})\}(1-\tilde{\pi}_i)]\big). \quad (13)$$

This function can be maximized using the R function "optim."

Variance estimation

As in typical nonparametric maximum likelihood inference under semiparametric models, the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is estimated by the inverse of a discrete observed information matrix $I(\hat{\boldsymbol{\theta}})$ following the EM algorithm, which is given by Louis' formula[17] based on missing information principle:

$$I(\boldsymbol{\theta}) = \mathbb{E}[-\ddot{l}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}, \mathbf{u})|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}] - \mathbb{E}[s(\theta; \mathbf{y}, \mathbf{z}, \mathbf{u})s(\theta; \mathbf{y}, \mathbf{z}, \mathbf{u})'|\mathbf{y}, \mathbf{z}, \boldsymbol{\theta}], \quad (14)$$

where $\ddot{l}$ and s denote the second and first derivatives of l with respect to $\boldsymbol{\theta}$. The components of $\ddot{l}$ and s are given in the Appendix.

3.2 | The stochastic EM algorithm

Instead of the EM algorithm described above, the stochastic EM algorithm was used in Carnegie et al,[12] we think primarily due to its ease of implementation for practitioners as well as intuitive appeal. It is similar to a Monte Carlo EM (MCEM) but in the E-steps only a single U is drawn from the conditional distribution of U given the observed data, so that in the M-steps the parameters are updated using that single sample of U as if it were observed. A typical MCEM would otherwise draw many samples of U in order to approximate the conditional expectations in the E-steps. The E- and M-steps are as described above for the models that we consider in this article, for both survival and competing risks outcomes.

In order to obtain a more accurate estimate, the whole procedure is repeated K times, and the final estimate of the treatment effect on the survival outcome is $\hat{\tau} = \sum_{k=1}^{K} \hat{\tau}_k/K$, with the corresponding standard error

$$\hat{\sigma}_{\hat{\tau}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\hat{\sigma}^2_{\hat{\tau}_k} + \left(1+\frac{1}{K}\right)\frac{1}{K-1}\sum_{k=1}^{K}(\hat{\tau}_k - \hat{\tau})^2}, \quad (15)$$

where $\hat{\sigma}^2_{\hat{\tau}_k}$ is estimated variance of $\hat{\tau}_k$ pretending that the singly sampled U_k is observed. For competing risks we have similarly for type j event $\hat{\tau}_j = \sum_{k=1}^{K} \hat{\tau}_{jk}/K$, and the corresponding standard error is obtained using (15) with $\hat{\tau}_k$ replaced by $\hat{\tau}_{jk}$ and $\hat{\tau}$ replaced by $\hat{\tau}_j$.

Nielsen et al[18] studied the asymptotic behavior of the stochastic EM algorithm, and showed that under certain assumptions it is root-n consistent but not fully efficient. We show in our data analysis that it can be naturally adapted to the IPW approach and obtain inferential results in sensitivity analysis.

The implementation of all of the above methods is available in the R package "survSens" on CRAN.

4 | SIMULATIONS

We conducted simulation studies to investigate the performance of the EM as well as the stochastic EM algorithms, as compared with the estimation of the treatment effect using the true confounder U with the given sensitivity parameters. For both survival and competing risks outcomes, we set sample size $n = 1000$, $U \sim$ Bernoulli(0.5), and two independent

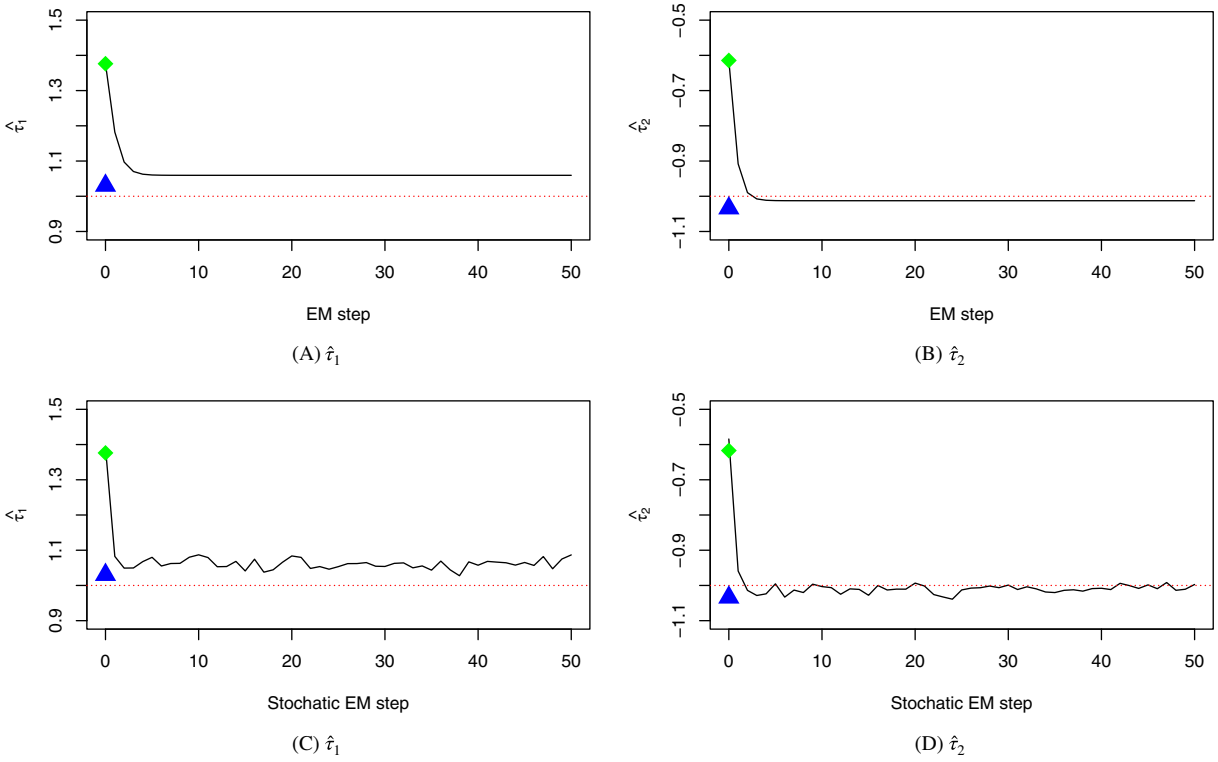


FIGURE 1 Convergence of the EM (top) and stochastic EM (bottom) algorithms in a single simulation run. The red horizontal lines indicate the true values of τ_j's, and $\zeta_1 = \zeta_2 = \zeta_z = 1$. The blue triangles correspond to the estimated treatment effects with true U, the green diamonds correspond to the estimates without U, and the black lines show the values of $\hat{\tau}_j$ during the first 50 steps. All sequences met the convergence criterion in 50 steps. EM, expectation-maximization [Colour figure can be viewed at wileyonlinelibrary.com]

covariates $X_1 \sim N(0,1)$, $X_2 \sim N(1,1)$ with $\boldsymbol{\beta_z} = (0.25, -0.25)'$ in (2). The number of EM or stochastic EM steps was set to 20 (see Figure 1 and related discussion below), and true sensitivity parameter values were used in fitting the models. The final estimates from the stochastic EM were obtained by averaging over $K = 40$ estimates to reduce the variability. For each case we show the results of 200 simulation runs.

4.1 | Survival outcomes

To simulate survival outcomes under model (1), we set $\lambda_0(t) = 1$, $\boldsymbol{\beta} = (0.5, -1)'$, and $\tau = 1$. In addition, we set censoring times $C\sim$ Uniform(1, 2) which led to between 25% and 60% censoring, depending on the combinations of the parameter values.

We run simulations over each combination of $\zeta_z \in \{0, 1, 2\}$ and $\zeta \in \{-2, -1, 0, 1, 2\}$. The results of the simulation are reported in Table 1 and Supplement Figure 1. From the table and figure it is clear that ignoring U led to bias in the estimated treatment effect as long as $\zeta \neq 0$; this bias also increases with the magnitude of ζ as well as the magnitude of ζ_z. On the other hand, both the stochastic EM and the EM algorithms gave good estimates of the treatment effect compared with the estimates using the true U's. Closer comparison of the results in Table 1 shows that the EM algorithm gave more accurate estimates than the stochastic EM algorithm, both in terms of generally less bias and smaller variances.

4.2 | Competing risks outcomes

To simulate competing risks outcomes, we followed the approach designed in Beyersmann et al.[19] We assumed that $m = 2$, the baseline hazard functions for type 1 and type 2 failures to be $\lambda_{10}(t) = \lambda_{20}(t) = 1$, and $\boldsymbol{\beta_1} = (0.5, -1)'$, $\tau_1 = 1$, $\boldsymbol{\beta_2} = (-0.5, 0.2)'$, $\tau_2 = -1$ in model (5). We then simulated the survival times with all-causes hazard $\lambda = \lambda_1 + \lambda_2$, and

TABLE 1 Estimated treatment effect (standard deviation) for the simulated survival data with $\tau = 1$

	Method	$\zeta_z = 0$	$\zeta_z = 1$	$\zeta_z = 2$
$\zeta = -2$	True U	1.0171 (0.1244)	1.0108 (0.1171)	1.0033 (0.1166)
	EM	1.0216 (0.1496)	1.0101 (0.1469)	1.0015 (0.1297)
	Sto EM	1.0257 (0.1502)	1.0132 (0.1468)	1.0066 (0.1296)
	No U	0.7873 (0.1219)	0.1512 (0.1257)	−0.2052 (0.1141)
$\zeta = -1$	True U	1.0206 (0.1015)	1.0144 (0.1067)	1.0129 (0.1071)
	EM	1.0203 (0.1153)	1.0121 (0.1173)	1.0104 (0.1103)
	Sto EM	1.0220 (0.1157)	1.0125 (0.1172)	1.0118 (0.1105)
	No U	0.9310 (0.1068)	0.5664 (0.1109)	0.3524 (0.1068)
$\zeta = 0$	True U	1.0159 (0.0868)	1.0124 (0.0996)	1.0095 (0.1035)
	EM	1.0159 (0.0868)	1.0124 (0.0996)	1.0095 (0.1035)
	Sto EM	1.0159 (0.0868)	1.0124 (0.0996)	1.0095 (0.1035)
	No U	1.0159 (0.0868)	1.0124 (0.0996)	1.0095 (0.1035)
$\zeta = 1$	True U	1.0148 (0.0797)	1.0134 (0.0896)	1.0110 (0.1004)
	EM	1.0188 (0.0891)	1.0167 (0.0977)	1.0139 (0.1072)
	Sto EM	1.0195 (0.0894)	1.0183 (0.0977)	1.0164 (0.1068)
	No U	0.9059 (0.0802)	1.2601 (0.0878)	1.4993 (0.0971)
$\zeta = 2$	True U	1.0133 (0.0768)	1.0164 (0.0875)	1.0154 (0.1031)
	EM	1.0226 (0.1047)	1.0260 (0.1122)	1.0263 (0.1218)
	Sto EM	1.0225 (0.1052)	1.0271 (0.1127)	1.0303 (0.1228)
	No U	0.6946 (0.0783)	1.2618 (0.0835)	1.6734 (0.0942)

Abbreviation: EM, expectation-maximization.

the cause J was generated from Bernoulli trials with $\mathbb{P}(J = 1|Z, \boldsymbol{X}, U) = \lambda_1/(\lambda_1 + \lambda_2)$. We also set censoring times $C\sim$ Uniform(0.3, 0.7).

Similarly as the survival model, we first ran simulations over each combination of $\zeta_z \in \{0, 1, 2\}$ and $\zeta_1 = \zeta_2 \in \{-2, -1, 0, 1, 2\}$. This gave about 20% to 60% censoring, depending on the combinations of the parameter values, and about equal numbers of type 1 and type 2 events. In a second scenario, we fixed $\zeta_1 = 1$ and $\zeta_2 \in \{-2, -1, 0, 1, 2\}$ as before, which gave about 20% to 40% censoring, and type 1/2 event rates between 40%/20% and 30%/50%, again depending on the combinations of the parameter values. The results of experiments are reported in Tables 2 to 4, Supplement Table 1 and Figures 2 to 5. All results show that for each type of failure, the estimated treatment effect by either the stochastic EM or the EM recovered the true treatment effect quite well, while ignoring U induced a substantial bias. In particular, Table 3 and Supplement Figure 4 show that varying ζ_2 had a noticeable impact on the estimation of τ_1, that is, unobserved confounding for type 2 failure had a noticeable impact on the estimation of the treatment effect on type 1 failure.

Finally, we take a closer look at the EM and the stochastic EM algorithms in a single run. Figure 1 plots the values of the corresponding $\hat{\tau}_j$'s during the first 50 EM or stochastic EM steps. Such plots are often used to examine the behavior and convergence of EM type algorithms for a given dataset. It is seen that the EM sequence displays a much smoother line than the stochastic EM sequence; and even at convergence, the stochastic EM sequence has quite some fluctuation compared with the EM sequence.

5 | SENSITIVITY ANALYSIS OF THE IBD DATA

5.1 | UC data

UC is one type of IBD that occurs in the large intestine (colon) and the rectum, which is characterized clinically by bloody diarrhea and urgency. We are interested in comparing the effectiveness between Vedolizumab and TNF-antagonist

TABLE 2 Treatment effect estimate (standard deviation) on type 1 failures for the simulated competing risks data with $\tau_1 = 1$

	method	$\zeta_z = 0$	$\zeta_z = 1$	$\zeta_z = 2$
$\zeta_1 = \zeta_2 = -2$	True U	1.0150 (0.1428)	1.0293 (0.1603)	1.0305 (0.1644)
	EM	1.0154 (0.1767)	1.0357 (0.1801)	1.0354 (0.1831)
	StoEM	1.0180 (0.1788)	1.0390 (0.1825)	1.0428 (0.1844)
	No U	0.9312 (0.1583)	0.3192 (0.1670)	−0.0215 (0.1678)
$\zeta_1 = \zeta_2 = -1$	True U	1.0141 (0.1327)	1.0185 (0.1542)	1.0269 (0.1613)
	EM	1.0141 (0.1390)	1.0186 (0.1585)	1.0280 (0.1677)
	StoEM	1.0150 (0.1388)	1.0191 (0.1584)	1.0304 (0.1675)
	No U	0.9817 (0.1339)	0.6243 (0.1536)	0.4260 (0.1635)
$\zeta_1 = \zeta_2 = 0$	True U	1.0153 (0.1212)	1.0258 (0.1329)	1.0317 (0.1593)
	EM	1.0153 (0.1212)	1.0258 (0.1329)	1.0317 (0.1593)
	StoEM	1.0153 (0.1212)	1.0258 (0.1329)	1.0317 (0.1593)
	No U	1.0153 (0.1212)	1.0258 (0.1329)	1.0317 (0.1593)
$\zeta_1 = \zeta_2 = 1$	True U	1.0078 (0.1115)	1.0276 (0.1259)	1.0348 (0.1524)
	EM	1.0088 (0.1226)	1.0251 (0.1415)	1.0320 (0.1549)
	StoEM	1.0095 (0.1230)	1.0272 (0.1422)	1.0346 (0.1549)
	No U	0.9745 (0.1130)	1.3518 (0.1319)	1.5750 (0.1470)
$\zeta_1 = \zeta_2 = 2$	True U	1.0064 (0.1094)	1.0263 (0.1238)	1.0345 (0.1570)
	EM	1.0022 (0.1370)	1.0174 (0.1600)	1.0266 (0.1645)
	StoEM	1.0031 (0.1369)	1.0230 (0.1615)	1.0345 (0.1652)
	No U	0.9238 (0.1141)	1.5064 (0.1300)	1.8424 (0.1411)

Abbreviation: EM, expectation-maximization.

therapy for UC patients. The data were collected between May 2014 and December 2017 from the North American based VICTORY consortium registry.[8] In brief, a total of 719 (453 treated with Vedolizumab, 266 with TNF-antagonist) UC patients with a median follow-up of 12 months were included. We focus on the treatment effect of Vedolizumab ($Z = 1$) vs TNF-antagonist ($Z = 0$) on clinical remission, which is defined as resolution of diarrhea, rectal bleeding and urgency. In the Vedolizumab group, 187 patients had clinical remission and no one had surgery, while in the TNF-antagonist group, 100 patients had clinical remission and three patients had surgery. Since there were only three competing events of surgery, too few to fit any model, we had to simply treat surgery as independent censoring and applied our approach under the survival models (ie, without competing risks) to approximate the treatment effect of Vedolizumab.

In Lukin et al,[10] the propensity score for each subject i, denoted PS_i, was calculated using the R package "twang"[20] based on pretreatment variables, including age, disease extent, clinical disease severity, UC related hospitalization within the preceding 1-year, prior TNF-antagonist exposure, baseline steroid dependency or refractoriness, concomitant steroid use, and concomitant immunomodulator use.

To be consistent with Lukin et al,[10] here we consider a single covariate $X_i = \Phi^{-1}(\text{PS}_i)$ in our models, as this quantity is more likely to be normally distributed than PS_i. In the models without unmeasured confounding ($\zeta_z = \zeta = 0$), the estimates were $\hat{\beta}_z$ (SE) = 1.1002 (0.0926), $\hat{\beta}$ (SE) = −0.3250 (0.0994), and $\hat{\tau}$ (SE) = 0.5756 (0.1423), where "SE" stands for standard error. We note that $\hat{\beta}_z$ would have been exactly one if, instead of "twang," probit regression had been used to fit the propensity score model. In addition, the estimated treatment effect $\hat{\tau}$ here was obtained by regression adjustment, compared with the IPW estimate of Lukin et al[10] (see sensitivity analysis for IPW below also).

We then assume that there is an unmeasured confounder $U \sim$ Bernoulli(0.5). To determine the range for the sensitivity parameters, we take into consideration the observed association between a measured confounder and the treatment received or the outcome, in this case all less than one in absolute value in terms of log odds ratio (OR) or log hazard ratio (HR). In addition, a probit coefficient on a binary variable (U) is likely to lie in [− 2,2] in practice as suggested by Carnegie

TABLE 3 Treatment effect estimate (standard deviation) on type 1 failures for the simulated competing risks data with $\tau_1 = 1$ and $\zeta_1 = 1$ fixed

	method	$\zeta_z = 0$	$\zeta_z = 1$	$\zeta_z = 2$
$\zeta_2 = -2$	true U	0.9971 (0.1011)	1.0187 (0.1163)	1.0271 (0.1505)
	EM	0.9963 (0.1093)	1.0174 (0.1320)	1.0263 (0.1545)
	stoEM	0.9971 (0.1101)	1.0189 (0.1320)	1.0282 (0.1549)
	no U	0.8893 (0.0998)	1.2872 (0.1214)	1.5816 (0.1468)
$\zeta_2 = -1$	true U	1.0020 (0.1008)	1.0212 (0.1227)	1.0292 (0.1508)
	EM	1.0028 (0.1080)	1.0203 (0.1369)	1.0280 (0.1544)
	stoEM	1.0034 (0.1091)	1.0214 (0.1368)	1.0299 (0.1548)
	no U	0.9023 (0.0992)	1.2980 (0.1264)	1.5844 (0.1465)
$\zeta_2 = 0$	true U	1.0033 (0.0966)	1.0271 (0.1183)	1.0326 (0.1481)
	EM	1.0040 (0.1045)	1.0258 (0.1338)	1.0305 (0.1507)
	stoEM	1.0050 (0.1045)	1.0275 (0.1338)	1.0325 (0.1510)
	no U	0.9243 (0.0965)	1.3220 (0.1244)	1.5863 (0.1431)
$\zeta_2 = 1$	true U	1.0078 (0.1115)	1.0276 (0.1259)	1.0348 (0.1524)
	EM	1.0088 (0.1226)	1.0251 (0.1415)	1.0320 (0.1549)
	stoEM	1.0095 (0.1230)	1.0272 (0.1422)	1.0346 (0.1549)
	no U	0.9745 (0.1130)	1.3518 (0.1319)	1.5750 (0.1470)
$\zeta_2 = 2$	true U	1.0148 (0.1231)	1.0245 (0.1402)	1.0301 (0.1592)
	EM	1.0135 (0.1353)	1.0195 (0.1506)	1.0255 (0.1609)
	stoEM	1.0146 (0.1356)	1.0223 (0.1511)	1.0295 (0.1615)
	no U	1.0452 (0.1271)	1.3591 (0.1415)	1.5158 (0.1520)

Abbreviation: EM, expectation-maximization.

et al.[12] Similarly under the Cox PH model, the log HR of ± 2 is very substantial for a binary variable. Therefore, we focused on $\zeta_z \in [-2, 2]$ and $\zeta \in [-2, 2]$.

The EM and stochastic EM algorithms were then applied as described in Section 3. The estimates from the stochastic EM were obtained by averaging over $K = 100$ estimates. The sensitivity analysis results are reported in Figure 2A,B and Supplement Table 2. Figure 2A,B show that over a wide range of sensitivity parameters, the EM and the stochastic EM gave very similar results. Note that except for very small random fluctuation in the stochastic EM results, the contours and curves are symmetric about the origin $(\zeta_z, \zeta) = (0, 0)$, where the estimated $\hat{\tau} = 0.5756$ is marked.

A main usage of these sensitivity plots is to identify the magnitude of the unmeasured confounding, that is, the sensitivity parameters (ζ_z, ζ), needed to alter a conclusion on the treatment effect. This can be reflected in two ways: (1) to drive the estimated treatment effect to zero, or (2) to lead to a nonsignificant estimated treatment effect in this case. From the plots we see that (ζ_z, ζ) will need to be close to (1.5, 1) or (1, 1.5), for example, in order to drive the estimated treatment effect to zero. To understand whether such a magnitude is likely in practice, we may again compare them to the observed association between a measured confounder and the treatment or the outcome, which were all less than one in absolute value in terms of log OR or log HR as we noted earlier (the largest log HR being just under 0.6 in absolute value). We may also compare them to the fitted values of $\hat{\beta}_z = 1.1002$ and $\hat{\beta} = -0.3250$ above. We see that such a very strong association between U and the survival outcome, in particular, seems unlikely.

In Figure 2A,B any combination of (ζ_z, ζ) in the region between two red curves in the upper right or lower left quadrant leads to a nonsignificant estimated treatment effect at 0.05 level two-sided. For example, (ζ_z, ζ) will need to be close to (1, 0.8), in order to drive the estimate to be nonsignificant. Similar to the discussion above, such a magnitude of unmeasured confounding seems unlikely in practice.

Finally, as IPW with PS_i was the main statistical approach used in Lukin et al[10] to estimate the treatment effect, we also carried out sensitivity analysis for this approach. We implemented this by combining the stochastic EM with IPW as

TABLE 4 Treatment effect estimate (standard deviation) on type 2 failures for the simulated competing risks data with $\tau_2 = -1$ and $\zeta_1 = 1$ fixed

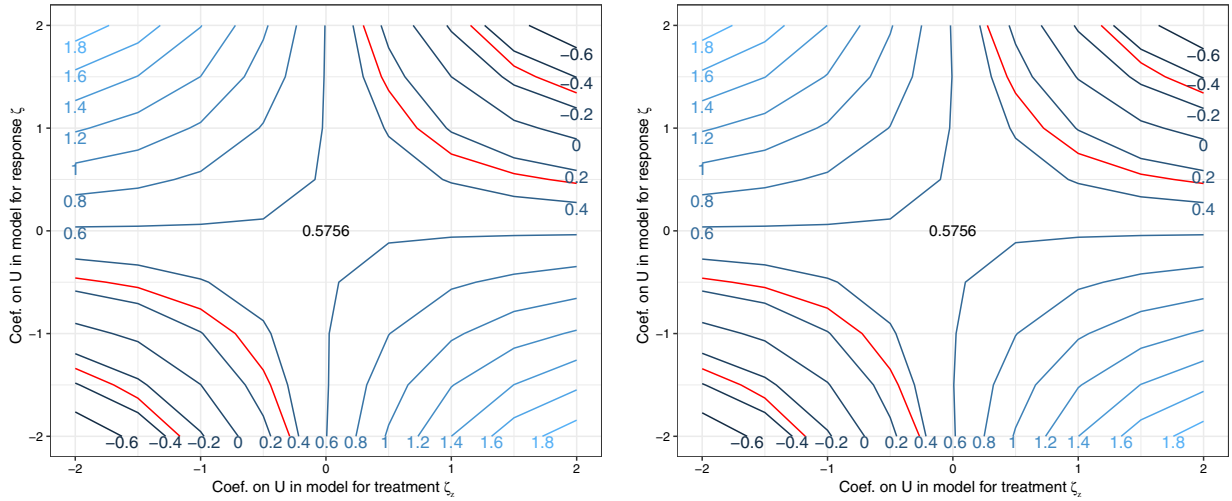
	method	$\zeta_z = 0$	$\zeta_z = 1$	$\zeta_z = 2$
$\zeta_2 = -2$	true U	−1.0113 (0.1926)	−1.0100 (0.1818)	−1.0107 (0.1747)
	EM	−1.0118 (0.2107)	−0.9985 (0.1886)	−1.0031 (0.1832)
	stoEM	−1.0142 (0.2111)	−1.0001 (0.1898)	−1.0002 (0.1831)
	no U	−0.8715 (0.1996)	−1.4541 (0.1786)	−1.8513 (0.1766)
$\zeta_2 = -1$	true U	−1.0224 (0.1803)	−1.0066 (0.1556)	−1.0082 (0.1537)
	EM	−1.0222 (0.1887)	−0.9995 (0.1546)	−1.0038 (0.1567)
	stoEM	−1.0227 (0.1888)	−1.0001 (0.1549)	−1.0035 (0.1568)
	no U	−0.9510 (0.1841)	−1.2829 (0.1511)	−1.5228 (0.1547)
$\zeta_2 = 0$	true U	−1.0170 (0.1693)	−1.0010 (0.1383)	−0.9957 (0.1329)
	EM	−1.0170 (0.1693)	−1.0010 (0.1383)	−0.9957 (0.1329)
	stoEM	−1.0170 (0.1693)	−1.0010 (0.1383)	−0.9957 (0.1329)
	no U	−1.0170 (0.1693)	−1.0010 (0.1383)	−0.9957 (0.1329)
$\zeta_2 = 1$	true U	−1.0088 (0.1316)	−0.9886 (0.1019)	−0.9836 (0.1097)
	EM	−1.0108 (0.1423)	−0.9950 (0.1109)	−0.9887 (0.1115)
	stoEM	−1.0110 (0.1434)	−0.9953 (0.1117)	−0.9883 (0.1114)
	no U	−0.9776 (0.1328)	−0.5956 (0.1039)	−0.3335 (0.1074)
$\zeta_2 = 2$	true U	−1.0054 (0.1153)	−0.9918 (0.0933)	−0.9876 (0.0969)
	EM	−1.0143 (0.1513)	−1.0037 (0.1272)	−0.9976 (0.1175)
	stoEM	−1.0139 (0.1522)	−1.0026 (0.1276)	−0.9927 (0.1174)
	no U	−0.7921 (0.1136)	−0.1505 (0.0981)	0.3065 (0.0977)

Abbreviation: EM, expectation-maximization.

follows. At convergence of the algorithm we simulated U_i and estimated the propensity score $\mathbb{P}(Z = 1|X, U)$ by regressing Z_i on $X_i = \Phi^{-1}(\text{PS}_i)$ and the simulated U_i, $i = 1, \ldots n$. Stabilized weights were obtained and further trimmed to be within (0.1, 10) if necessary. The IPW approach was then applied. The final estimates were also obtained by averaging over $K = 100$ estimates, with the corresponding standard errors obtained using (15) where $\hat{\sigma}^2_{\hat{\tau}_k}$ was the sandwich variance estimator following the IPW. The results are reported in Figure 2C and Supplement Table 2. It is seen that unlike the regression adjustment results above, where the estimated treatment effect remained the same as long as $\zeta = 0$, here instead the estimated treatment effect remained the same as long as $\zeta_z = 0$. We also note much larger standard errors when $|\zeta_z|$ is large, perhaps understandable as the treatment groups become more imbalanced. However, similar to the regression adjustment results above, in order to drive the estimated treatment effect to zero, (ζ_z, ζ) will need to be close to (1.5, 1) or (1, 1.5). On the other hand, the estimated treatment effect may become nonsignificant at 0.05 level if $(\zeta_z, \zeta) = (0.5, 1)$ or $(\zeta_z, \zeta) = (0.8, 0.5)$.

5.2 | CD data

CD is another type of IBD that can cause inflammation along anywhere of the digestive tract. We are again interested in comparing the effectiveness between Vedolizumab and TNF-antagonist therapy for CD patients. The data were collected between May 2014 and December 2017 from the North American based consortium registry.[9] A total of 1242 patients were included (655 treated with Vedolizumab, 587 with TNF-antagonist therapy). The primary interest is the treatment effect of Vedolizumab ($Z = 1$) vs TNF-antagonist ($Z = 0$) on clinical remission, which is defined as complete resolution of CD-related symptoms. In the Vedolizumab group, 196 patients had clinical remission and nine had surgery, while in the TNF-antagonist group, 255 patients had clinical remission and 18 patients had surgery. Supplement Figure 6 shows the cumulative incidence curves for time to clinical remission and time to surgery in these patients. Due to the



(A) stochastic EM

(B) EM

(C) IPW

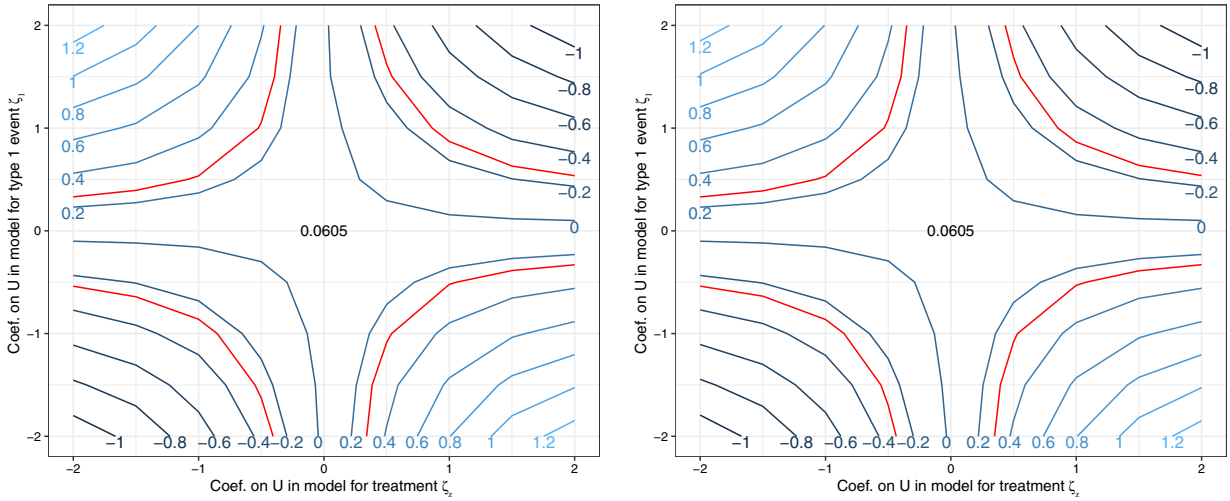
FIGURE 2 Sensitivity analysis results for UC patients data for outcome clinical remission. In all plots, the blue contours show the sensitivity parameter values corresponding to the estimated treatment effect $\hat{\tau}$, and the red curves correspond to where the absolute value of the t-statistic $|t| = |\hat{\tau}/\hat{\sigma}_{\hat{\tau}}| = 1.96$ [Colour figure can be viewed at wileyonlinelibrary.com]

presence of competing events, we applied our approach under the competing risks models to estimate the treatment effect of Vedolizumab.

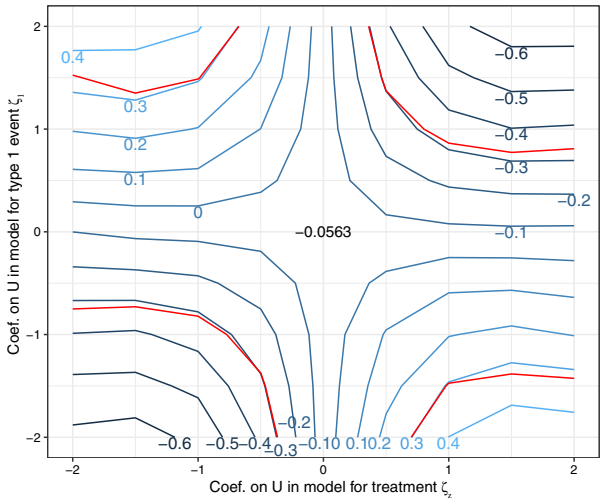
In Bohm et al,[11] the propensity score for each subject i, denoted PS_i, was calculated using the R package "twang"[20] based on pretreatment variables, including prior TNF-antagonist exposure and number of prior TNF-antagonists exposed, disease extent, history of fistulizing disease, prior bowel surgery, disease phenotype, clinical disease severity, CD related hospitalization within the preceding 1-year, baseline steroid dependency or refractoriness, concomitant steroid use, or concomitant immunomodulator use.

To be consistent with Bohm et al,[11] we consider a single covariate $X_i = \Phi^{-1}(\text{PS}_i)$ in our models. In the models without unmeasured confounding ($\zeta_z = \zeta_1 = \zeta_2 = 0$), the estimate of β_z as defined in model (2) is $\hat{\beta}_z$ (SE) = 1.0631 (0.0513), the estimates of β_j ($j = 1,2$) as defined in model (5) are $\hat{\beta}_1$ (SE) = −0.1664 (0.0562) and $\hat{\beta}_2$ (SE) = −0.2601 (0.2401), and the estimates of τ_j ($j = 1,2$) are $\hat{\tau}_1$ (SE) = 0.0605 (0.1318) and $\hat{\tau}_2$ (SE) = −0.0537 (0.5705).

We then assume an unmeasured confounder $U \sim$ Bernoulli(0.5). The range for the sensitivity parameters is determined similarly as the UC data. We focus on $\zeta_z \in [-2, 2]$, $\zeta_1 \in [-2, 2]$, and $\zeta_2 \in \{-2, 0, 2\}$. The EM and stochastic EM algorithms were then applied as described in Section 3. The estimates from the stochastic EM were obtained by averaging over $K = 100$ estimates. The sensitivity analysis results are reported in Figure 3A,B and Supplement Tables 3 to 5.



(A) stochastic EM, $\zeta_2 = 0$ (B) EM, $\zeta_2 = 0$



(C) IPW, $\zeta_2 = 0$

FIGURE 3 Sensitivity analysis results for CD patients data for outcome clinical remission. In all plots, the blue contours show the values of (ζ_z, ζ_1) corresponding to the estimated treatment effect $\hat{\tau}_1$, and the red curves correspond to where the absolute value of the t-statistic $|t| = |\hat{\tau}_1/\hat{\sigma}_{\hat{\tau}_1}| = 1.96$ [Colour figure can be viewed at wileyonlinelibrary.com]

Note that by our algorithms, ζ_2 affects $\hat{\tau}_1$ only through the conditional probability of U as shown in (3). In these data, as the number of surgery is relatively small compared with the number of clinical remission, the effect of ζ_2 on $\hat{\tau}_1$ is subtle (Supplement Tables 3 to 5). This is, of course, not necessarily true if the number of the competing risk events is comparable to the number of events of interest. We further discuss the impact of the competing risk toward the end of this analysis.

Figure 3A,B show that when $\zeta_2 = 0$, over a wide range of (ζ_z, ζ_1), the EM and the stochastic EM gave similar results. In the plots, the blue contours show the values of (ζ_z, ζ_1) corresponding to the estimated treatment effect $\hat{\tau}_1$, and the red curves correspond to where the absolute value of the $t-$statistic $|t| = |\hat{\tau}_1/\hat{\sigma}_{\hat{\tau}_1}| = 1.96$. Hence, any combination of (ζ_z, ζ_1) in the region surrounded by four red curves leads to a nonsignificant estimated treatment effect at level 0.05 two-sided. Except for very small random fluctuation in the stochastic EM results, the contours and curves are symmetric about the origin $(\zeta_z, \zeta_1) = (0, 0)$, where the estimated $\hat{\tau}_1 = 0.0605$ is marked. We see that in order to drive the estimated treatment effect to being significant, given $\zeta_2 = 0$, (ζ_z, ζ_1) will need to be close to (1,1) or (−0.8, 0.8), for example. Compared with $\hat{\beta}_z = 1.0631$ and $\hat{\beta}_1 = -0.1664$ above, such a strong association between U and the outcome seems unlikely in practice.

As IPW with PS_i was the main statistical approach used in Bohm et al[11] to estimate the treatment effect, we also carried out sensitivity analysis for this approach by combining the stochastic EM with IPW as under the survival models.

The final estimates were also obtained by averaging over $K = 100$ estimates. The results are reported in Figure 3C and Supplement Tables 3 to 5. Similar to the regression adjustment results, in order to drive the estimated treatment effect to being significant, given $\zeta_2 = 0$, (ζ_z, ζ_1) will need to be close to (1, 1) or (−1, 1.5), which seems unlikely in practice.

We emphasize that the interpretation of treatment effect in the presence of competing risks needs caution in general. Here, the comparison of the two treatment groups as reflected in the effect τ_1 for clinical remission is among those without surgery. In this case, the effect of treatment on the competing risk, that is, time to surgery, is not significant for a broad range of sensitivity parameter values (data not shown). Therefore, we are not in a situation where a treatment appears to increase the risk of one type of events while reducing the risk of another type of events, which could otherwise happen in practice as the probabilities from different types of events must sum up to one as time goes to infinity.

Finally, as suggested by a reviewer, we explore the sensitivity analyses when the distribution of U, instead of being symmetric, has $\pi = 0.7$ or 0.3. The results are in the Supplement Figures 7 to 10. It is seen that for the same values of ζ or ζ_z, the change in the estimated treatment effect is not as large. A possible explanation is that the magnitude of unmeasured confounding, as reflected in the variance of U, is reduced when $\pi = 0.7$ or 0.3. We also note that in this case the contours are no longer symmetric about the origin $(\zeta_z, \zeta_1) = (0, 0)$.

6 | DISCUSSION

In this article, we developed approaches to perform sensitivity analysis of the estimated treatment effect with regard to unobserved confounding in observational studies with survival or competing risks outcomes. The approaches we developed are based on models for survival or competing risks outcomes, which allow simulating the unobserved confounder given the observed data. The sensitivity parameters reflect the association between the unobserved confounder and the outcomes, as well as the association between the unobserved confounder and the treatment received. The interpretation of these sensitivity parameters is straightforward, which leads to relative ease in choosing plausible ranges for them. Simulation studies show that both the EM and the stochastic EM algorithms are able to recover the true treatment effect if the correct sensitivity parameter values are used. The EM algorithm is clearly optimal in theory,[18] although the stochastic EM allows easy incorporation of IPW approaches for estimating treatment effects, which are commonly used in practice and as we have illustrated in our data analysis.

Lin et al[3] developed an analytic approach with closed-form formulas for assessing the sensitivity of regression results to unmeasured confounders in observational studies with either binary or survival outcomes. Under certain conditions, they derived simple algebraic relationships between the true treatment effect and the apparent treatment effect ignoring the unmeasured confounder U. For survival data, they assumed that the event was rare or the effect of U on the survival outcome was small. Their parameterization is different from ours here. Lin et al parametrized the conditional distribution of U given the treatment Z, and assumed that U was independent of the observed covariates X given treatment received Z; this isolates the unmeasured confounder effect from those of the observed covariates. By contrast we have modeled the distribution of Z given X and U, which is perhaps a more general and flexible definition of the unmeasured confounder. The Lin et al approach does not apply to competing risks.

For the distribution of the unobserved confounder we used binary 0, 1 with probability 0.5 each, which were recommended and used throughout the book by Rosenbaum.[1] It is also possible to incorporate normally distributed U, such as in Shen et al[5] and Xu et al,[21] in which case the probit link in model (2) allows closed-form marginal propensity scores given X after integrating out U. The Q_1 part of the EM algorithm would be similar to that under the proportional hazards mixed-effects model and Monte Carlo approximation would be needed in the E-steps.[16]

Carnegie et al[12] discussed the advantages and disadvantages of using parametric vs nonparametric approaches in sensitivity analysis. Parametric approaches are typically needed in order to simulate the unobserved confounder; in survival analysis, however, the outcome models are often semiparametric, allow flexibility in modeling in particular the nuisance parameters. On the other hand, nonparametric bounds might be considered under minimal assumptions in place of sensitivity analysis.[22] However, such bounds can be very difficult to derive for complex outcomes like what we consider here in the presence of right censoring, which is unlike in Shen et al[5] where it is possible to derive these bounds for binary or continuous outcomes without censoring. Also evident in Shen et al[5] is that parametric settings are often needed in order to aid in the interpretation of the sensitivity parameters in the corresponding nonparametric settings, and extensive simulations have to be conducted in order to determine sensible ranges for these sensitivity parameters.[21]

ORCID

Ronghui Xu https://orcid.org/0000-0002-2822-0561

REFERENCES

1. Rosenbaum PR. *Observational Studies*. 2nd ed. New York, NY: Springer-Verlag; 2002.
2. Liu W, Kuramoto SJ, Stuart EA. An introduction to sensitivity analysis for unobserved confounding in nonexperimental prevention research. *Prev Sci*. 2013;14(6):570-580.
3. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*. 1998;54:948-963.
4. Li L, Shen C, Wu AC, Li X. Propensity score-based sensitivity analysis method for uncontrolled confounding. *Am J Epidemiol*. 2011;174(3):345-353.
5. Shen C, Li X, Li L, Were MC. Sensitivity analysis for causal inference using inverse probability weighting. *Biom J*. 2011;53(5):822-837.
6. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34:3661-3679.
7. Katz JA, Melmed G, Sands BE. *The Facts About Inflammatory Bowel Diseases*. New York, NY: Crohn's & Colitis Foundation of America; 2011.
8. Narula N, Peerani F, Meserve J, et al. Vedolizumab for ulcerative colitis: treatment outcomes from the VICTORY Consortium. *Am J Gastroenterol*. 2018;113(9):1345.
9. Dulai PS, Singh S, Jiang X, et al. The real-world effectiveness and safety of vedolizumab for moderate–severe Crohn's disease: results from the US VICTORY consortium. *Am J Gastroenterol*. 2016;111(8):1147.
10. Lukin D, Faleck D, Xu R, et al. Comparative safety profile of vedolizumab and tumor necrosis factor-antagonist therapy for inflammatory bowel disease: a multicenter consortium propensity score-matched analysis. *Gastroenterology*. 2019;154(6):S68.
11. Bohm M, Xu R, Zhang Y, et al. Comparative effectiveness of vedolizumab and tumor necrosis factor-antagonist therapy in Crohn's disease: a multicenter consortium propensity score-matched analysis. *Gastroenterology*. 2019;154(6):S369-S370.
12. Carnegie NB, Harada M, Hill JL. Assessing sensitivity to unmeasured confounding using a simulated potential confounder. *J Res Educ Effect*. 2016;9(3):395-420.
13. Cox DR. Regression models and life-tables. *J R Stat Soc Series B (Methodological)*. 1972;34(2):187-202.
14. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. John Wiley & Sons; 2011.
15. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Series B (Methodological)*. 1977;39(1):1-22.
16. Vaida F, Xu R. Proportional hazards model with random effects. *Stat Med*. 2000;19(24):3309-3324.
17. Louis TA. Finding the observed information matrix when using the EM algorithm. *J R Stat Soc Series B (Methodological)*. 1982;44(2):226-233.
18. Nielsen SF. The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli*. 2000;6(3):457-489.
19. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med*. 2009;28(6):956-971.
20. McCaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychol Methods*. 2004;9(4):403.
21. Xu R, Honerkamp-Smith G, Chambers CD. Statistical sensitivity analysis for the estimation of fetal alcohol spectrum disorders prevalence. *Reprod Toxicol*. 2019;86:62-67.
22. Richardson A, Hudgens MG, Gilbert PB, Fine JP. Nonparametric bounds and sensitivity analysis of treatment effects. *Stat Sci*. 2014;29(4):596.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Huang R, Xu R, Dulai PS. Sensitivity analysis of treatment effect to unmeasured confounding in observational studies with survival and competing risks outcomes. *Statistics in Medicine*. 2020;39:3397–3411. https://doi.org/10.1002/sim.8672

APPENDIX

In the following we write out the components of $\ddot{l}$ and s for competing risks with $j = 1, \ldots, m$. For a single survival outcome without competing risks, we should simply take $m = 1$ and the corresponding parameters are the same as without the subscript j.

The components of s are:

$$\frac{\partial l}{\partial \boldsymbol{\beta_j}} = \sum_{i=1}^{n} \boldsymbol{x_i} \{\delta_{ij} - \Lambda_{j0}(t_i) \exp(\tau_j z_i + \boldsymbol{x_i}' \boldsymbol{\beta_j} + \zeta_j u_i)\} \tag{A1}$$

$$\frac{\partial l}{\partial \tau_j} = \sum_{i=1}^{n} z_i \{\delta_{ij} - \Lambda_{j0}(t_i) \exp(\tau_j z_i + \boldsymbol{x_i}' \boldsymbol{\beta_j} + \zeta_j u_i)\} \tag{A2}$$

$$\frac{\partial l}{\partial \lambda_{j0}(t_i)} = \frac{1}{\lambda_{j0}(t_i)} - \sum_{t_k \geq t_i} \exp(\tau_j z_k + \boldsymbol{x_k}' \boldsymbol{\beta_j} + \zeta_j u_k) \tag{A3}$$

$$\frac{\partial l}{\partial \boldsymbol{\beta_z}} = \sum_{i=1}^{n} \left\{ z_i \frac{\phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)}{\Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)} - (1 - z_i) \frac{\phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)}{1 - \Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)} \right\} \boldsymbol{x_i} \tag{A4}$$

for $j = 1, \ldots, m$. For the second derivatives,

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta_j}^2} = -\sum_{i=1}^{n} \boldsymbol{x_i}^{\otimes 2} \Lambda_{j0}(t_i) \exp(\tau_j z_i + \boldsymbol{x_i}' \boldsymbol{\beta_j} + \zeta_j u_i) \tag{A5}$$

$$\frac{\partial^2 l}{\partial \tau_j^2} = -\sum_{i=1}^{n} z_i \Lambda_{j0}(t_i) \exp(\tau_j z_i + \boldsymbol{x_i}' \boldsymbol{\beta_j} + \zeta_j u_i) \tag{A6}$$

$$\frac{\partial^2 l}{\partial \lambda_{j0}(t_i)^2} = -\frac{1}{\lambda_{j0}(t_i)^2} \tag{A7}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta_j} \partial \tau_j} = -\sum_{i=1}^{n} z_i \boldsymbol{x_i} \Lambda_{j0}(t_i) \exp(\tau_j z_i + \boldsymbol{x_i}' \boldsymbol{\beta_j} + \zeta_j u_i) \tag{A8}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta_j} \partial \lambda_{j0}(t_i)} = -\sum_{t_k \geq t_i} \boldsymbol{x_k} \exp(\tau_j z_k + \boldsymbol{x_k}' \boldsymbol{\beta_j} + \zeta_j u_k) \tag{A9}$$

$$\frac{\partial^2 l}{\partial \tau_j \partial \lambda_{j0}(t_i)} = -\sum_{t_k \geq t_i} z_k \exp(\tau_j z_k + \boldsymbol{x_k}' \boldsymbol{\beta_j} + \zeta_j u_k) \tag{A10}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \boldsymbol{\beta_z}^2} = -\sum_{i=1}^{n} \phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i) \Bigg\{ & z_i \frac{\phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i) + (\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i) \Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)}{\Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)^2} \\ & + (1 - z_i) \frac{\phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i) - (\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i)(1 - \Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i))}{(1 - \Phi(\boldsymbol{x_i}' \boldsymbol{\beta_z} + \zeta_z u_i))^2} \Bigg\} \boldsymbol{x_i}^{\otimes 2} \end{aligned} \tag{A11}$$

where $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}'$ for a vector $\boldsymbol{a}$, ϕ is the probability density function of the standard normal distribution, and all other off-diagonal elements are zeros. The computation of the first term in (14) is similar to the computation in the E-step for different functions $h(u_i)$. To calculate the second term in (14), we sample U from Bernoulli($\tilde{\pi}$) for 1000 times after convergence of the EM, and take the average of $s(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}, \mathbf{u}) s(\boldsymbol{\theta}; \mathbf{y}, \mathbf{z}, \mathbf{u})'$ over the sampled U's.