A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

Wencong Chen [1], Xiang Zhang [2], Douglas E. Faries [2], Wei Shen [2], John W. Seaman, Jr. [1], James D. Stamey [1]

(1) Department of Statistical Science, Baylor University, Waco, TX 76798
(2) Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285

CORRESPONDING AUTHOR: James Stamey, PhD - Baylor University - Department of Statistical Science - Waco, TX 76798 - 254.710.7405 (Ph.) - james_stamey@baylor.edu

ABSTRACT

Purpose: The existence of unmeasured confounding can clearly undermine the validity of an observational study. Methods of conducting sensitivity analyses to evaluate the impact of unmeasured confounding are well established. However, application of such methods to survival data ("time-to-event" outcomes) have received little attention in the literature. The purpose of this study is to propose a novel Bayesian method to account for unmeasured confounding for survival data.

Methods: The Bayesian method is proposed under an assumption that the supplementary information on unmeasured confounding in the form of internal validation data, external validation data or expert elicited prior distributions is available. The method for incorporating such information to Cox proportional hazard model is described. Simulation studies are performed based on the recently published instrumental variable method to assess the impact of unmeasured confounding and to illustrate the improvement of the proposed method over the naïve model which ignores unmeasured confounding.

Results: Simulation studies illustrate the impact of ignoring the unmeasured confounding and the effectiveness of our Bayesian approach. The corrected model had significantly less bias and coverage of 95% intervals much closer to nominal.

Conclusion: The proposed Bayesian method provides a useful and flexible tool in incorporating different types of supplemental information on unmeasured confounding to adjust the treatment estimates when the outcome is survival data. It out-performed the naïve model in simulation studies based on a real world study.

Key words: Bayesian, unmeasured confounding, survival analysis, sensitivity analyses, comparative effectiveness

INTRODUCTION

With the evolvement of electronic healthcare records, observational databases represent a rich source of routine clinical data which can potentially address a variety of medical questions. Due to lack of randomization, the use of such data for comparative effectiveness research is always challenged by selection bias [1-3]. In observational studies, it is critical to improve the study validity by using an appropriate comparator group to avoid potential bias or confounding in the study designs. Furthermore, some statistical methods, e.g. propensity score will help adjust for measured confounding factors, under an assumption of "no unmeasured confounding". If the critical assumption is violated, then under the same study population and design, conclusions may vary significantly depending on the different analytical approaches. In a recent comparative cardiovascular study [4], both the inverse propensity score weighting (IPW) method and the instrumental variable (IV) method were used. As is known to all, IPW requires the "no unmeasured confounding assumption", while IV method does not. Interestingly, the two well-established statistical methods provide inconsistent results. A possible explanation of these inconsistency is the existence of unmeasured confounding, as the authors pointed out "…. In the TRANSLATE-ACS study, 2 confounding factors that are potentially incompletely measured in the data are frailty and coronary disease severity…" In our opinion, this study exemplifies the significant impact unmeasured confounding can have on survival data and illustrates the need for conducting sensitivity analysis to address this problem.

Literature has reported methods for adjusting unmeasured confounding in parametric regression models. Frequentist approaches [2, 3, 5] adjust for unmeasured confounding by using fixed values for the bias parameters (the parameters associated with the unmeasured confounding). By contrast, the Bayesian approach we describe uses of validation data, expert opinion, or a combination of both to build prior probability distributions for these parameters. Monte Carlo sensitivity analysis [6] is possible with frequentist approaches, but the Bayesian paradigm provides an operational approach to combine the different sources of information. For instance, a Bayesian approach to adjust logistic regression for unmeasured confounding with mildly informative priors has been considered [7]. At the cost of inflated interval estimate widths, their use of relatively non-informative priors leads to only slight improvement in the bias of parameter estimates. This work was extended [8] by calibrating the priors for the regression parameters with internal validation data. That is, for a sub-sample of the data, the unmeasured confounder is ascertained. This allows for considerably better bias correction than the much less informative priors used in [7]. These publications focused on Bayesian methods utilizing extra information for parametric regression models when the outcome is continuous, binary or skewed (e.g., cost data).

Unlike parametric regression models, semi-parametric/non-parametric models with unmeasured confounding in comparative observational research have received little attention. In the frequentist paradigm, a simple adjustment using fixed values for the unmeasured confounding parameters in a Cox regression model has been considered [9]. The propensity score calibration approach [3] can also be applied to the Cox model. Here, we extend related work in Bayesian hierarchical modeling [8, 10, 11] by using informative priors and/or validation data to adjust for unmeasured confounding in Cox proportional hazard model.

Our paper is organized as follows. In Section 2 we overview the model and prior distributions for analyzing survival data with an unmeasured confounder. We discuss three approaches to correct for unmeasured confounding: informative priors, internal validation data and external validation data. In Section 3 we report the results of simulation studies showing the effectiveness of the validation data approaches. A discussion along with the concluding remarks is given in Section 4.

METHODS

In observational studies, propensity score [12] can be used as a summary scale to represent multiple confounders and directly adjusted as a regressor in the regression modeling. Theoretically, the Bayesian hierarchical model is able to adjust multiple unmeasured confounders. However, in our simulation studies, we found it would take extreme long to obtain the adjusted estimates if multiple unmeasured confounders existed. Therefore, in our development, we focus on the case of a continuous unmeasured confounder range from 0 to 1, as such single continuous unmeasured confounder could be also viewed as propensity score which summarized multiple unmeasured confounders. We denote the time to the event by T_i, the binary treatment indicator by x_i, the p-dimensional vector of measured confounders by z_i, and the unmeasured confounder by μ_i. Without loss of generality we assume $\mu_i \sim N(\mu_i, \sigma)$. The mean of the continuous unmeasured confounder can be modeled as a function of both the treatment and the measured covariates, specifically,

$\mu_i = \gamma_0 + \gamma_1 x_i + \sum_{(j=1)}^{p} \gamma_{j+1} z_{ji}$. This model can be extended to incorporate multiple unmeasured confounders and multiple treatments in a fairly straightforward way.

Likelihood

We will focus on the simplified model $\mu_i = \gamma_0 + \gamma_1 x_i$ in the discussion below. It has been shown that the simplified model is actually the worst case scenario. This is because

additional covariates may be related to unmeasured confounders, thereby reducing the latter's impact [13].

For the Cox proportional hazards model we follow the development in [14]. Packages such as OpenBUGS and JAGS can be used for posterior computation. We assume the survival times are divided into J intervals with end points $a_1 < a_2 < \ldots < a_J < a_{(J+1)}$. The Cox proportional hazards mode is

$$h(t) = h_0(t) \exp(\beta x + \theta z' + \eta \mu)$$

for scalar β and η, and a px1 vector of coefficients θ. The baseline hazard function, $h_0(t)$, is defined to be the step function

$$h_0(t) = \sum_{j=1} \lambda_j I_{[a_j, a_{j+1}]}(t)$$

Where the λ_j' S are positive scalars and I_A is the indicator function defined on the set A. If J is large, this parametric function nicely approximates an arbitrary baseline hazard function [14]. Simulation studies indicate J rarely needs to be larger than 10 for good performance. The likelihood for each subject is the joint distribution of the survival outcome and the distribution of the unmeasured confounder. The ith likelihood contribution is

$$L_i\left(\beta, \theta, \lambda, \eta, \gamma_0, \gamma_1, \sigma_u\right) = \prod_{j=1}\left[\varphi(i,j)\right]^{d_{ij}} \exp\left[-\varphi(i,j)\right]$$
$$\times \sigma_u^{-1/2} \exp\left[-\frac{(u_i - \gamma_0 - \gamma_1 x_i)^2}{2\sigma_u^2}\right]$$
$$(2)$$

where for the survival portion

$$\varphi(i,j) = \lambda_j \exp(\beta x_i + \theta z_i' + \eta \mu_i)\omega(i,j)$$
$$\omega(i,j) = \begin{cases} a_{j+1} - a_j & \text{if } t_i \geq a_{j+1} \\ t_i - a_j & \text{if } t_i \in [a_j, a_{j+1}), \\ 0 & \text{if } t_i \in [a_{j-1}, a_j) \end{cases}$$

and $d(i,j)$ is Poisson $\phi(i,j)$ with a value of either 0 or 1 depending on whether an event occurs or not. This is related to the so called "zeros trick" [15] to get non-standard distributions. Note that the u_i are being modeled but none are observed in the main study data of size n. The full likelihood is the product of the contributions in (2) across the sample size. Further discussion of the likelihood is provided in the appendix.

The coefficient β, which relates the treatment effect to the survival time, is the primary parameter of interest. Without additional information on γ_0, γ_1 and η the model is not identified. This extra information may come in the form of informative priors or validation data. For the latter, the unmeasured confounder is observed for a subset of the main study data or from a data source external to the main study. We will discuss the various approaches in the following subsections.

Prior distributions

We assume all regression parameters have normal distributions:

$$\beta \sim N(\mu_\beta, \sigma_\beta^2)$$
$$\theta_k \sim N(\mu_{\theta_k}, \sigma_{\theta_k}^2), \quad k = 1, 2, \cdots, p$$
$$\gamma_0 \sim N(\mu_{\gamma_0}, \sigma_{\gamma_0}^2), \gamma_1 \sim N(\mu_{\gamma_1}, \sigma_{\gamma_1}^2) \text{ and}$$
$$\eta \sim N(\mu_\eta, \sigma_\eta^2).$$

The steps, λ_j, in the baseline hazard function (1) are positive. Thus, we have selected gamma priors for them:
$$\lambda_j \sim \text{gamma}(\alpha_j, \kappa_j).$$

Finally, a prior is needed for the standard deviation of the unmeasured confounder, σ_μ. Here we assume a conjugate prior, $\sigma_u^2 \sim$ inverse-gamma(α_0, b_0). If the parameters α_0 and b_0 are chosen to be too small then convergence problems may arise. In this situation, it has been suggested a uniform, half-normal, or half Cauchy prior for the standard deviation be used instead of the inverse-gamma on the variance [16]. For our model, if validation data is not available, informative priors for γ_0, γ_1, and σ_u are required while diffuse priors can be used for β, θ_k and λ_j. An informative prior for η can be helpful but does not appear to be required for model identifiability. Eliciting priors for regression parameters is a non-trivial task. It is known that it is much easier to elicit on observable quantities as opposed to unobservable parameters [12]. Combining the priors with the likelihood given in (1) yields the joint posterior. RJAGS code for this model is available in the online appendix.

Internal Validation

If the study researchers are able to ascertain the unmeasured confounder for a random subsample of the subjects, the resulting informative priors can help mitigate the problems of the unidentified model that results from such confounding. This type of data is called internal validation and has been used in numerous cases to correct for bias due to misclassification and measurement error [17] and for unmeasured confounding [8]. For our case, suppose the total sample size is $n+m$ where n observations have the unmeasured confounder and m have it observed. In general n is much larger than m and m/n is referred to as the validation sample fraction. The resulting likelihood is

$$L(\beta, \theta, \lambda, \eta, \gamma_0, \gamma_1) \propto \prod_{i=1}^{n}\left\{\prod_{j=1}^{J}[\varphi(i,j)]^{d(i,j)} \exp[-\varphi(i,j)] \times \sigma_u^{-1/2} \exp\left[-\frac{(u_i - \gamma_0 - \gamma_1 x_i)^2}{2\sigma_u^2}\right]\right\}$$
$$\times \prod_{i=n+1}^{n+m}\left\{\prod_{j=1}^{J}[\varphi(i,j)]^{d(i,j)} \exp[-\varphi(i,j)] \times \sigma_u^{-1/2} \exp\left[-\frac{(u_i - \gamma_0 - \gamma_1 x_i)^2}{2\sigma_u^2}\right]\right\}$$
$$(3)$$

where, as previously defined, we have
$$\phi(i,j) = \lambda_j \exp(\beta x_i + \theta z_i' + \eta(\tilde{u}_i))\omega(i,j).$$
For the n observations in the main study, the $\tilde{u}_i$ are completely missing. Hence the "~" distinguishes them from

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

e12634-3

the values of u that are observed in the validation data. Since they are observed in the validation sample of size m, inference can proceed with either informative or diffuse priors. Combining the above likelihood with the priors of Section 2.1 yields the joint posterior. Sampling from this posterior can be accomplished with the RJAGS package in R and the code is available in the online appendix.

The advantage of internal validation data, as can be seen in the likelihood, is the data provides direct information on all parameters of interest using data from the target population. The main limitation is the feasibility of collecting additional data from study participants. The cost to obtain internal validation data may be too high. Contacting individuals whose data are routinely collected for administrative purposes may not be possible because of privacy concerns or laws.

External Validation

Since internal validation may be impossible or too expensive to obtain, another approach is to use external validation data. Examples of external validation data include previous studies conducted by the same researchers where subject-level data is likely to be available. Alternatively, published studies might be used where, for instance, information may be available in the form of summary statistics. If subject-level data is available, then the approach for external validation data is identical to that of internal validation and the likelihood given in (3) is used. If summary statistics from previous literature are available, an informative prior can be constructed in the following way. Assume the summary statistics are as in Table 1. Note this is the sufficient statistics for a linear regression with normal errors and a binary covariate. Our informative prior is constructed by first assuming flat priors on all parameters and constructing posteriors with the data in Table 1. These posteriors are then used as priors in the subsequent analyses. In this case, the prior for σ_u is

$$\sigma_u^2 \sim inverse-gamma\left[\frac{n_{x=0}+n_{x=1}-2}{2}, \frac{(n_{x=0}-1)S_{x=0}^2+(n_{x=1}-1)S_{x=1}^2}{2}\right]$$

Conditional on σ_u^2, we have $\gamma_k \mid \sigma_u^2 \sim N(\mu_{\gamma_k}), \sigma_{\gamma_k}^2)$ where $\mu_{\gamma_0} = \overline{\mu_{x=0}}, \sigma_{\gamma_0}^2 = (\frac{1}{n_{x=0}+n_{x=1}} + \frac{\bar{x}}{SSx})\mu_{\gamma_1} = \overline{\mu_{x=1}} - \overline{\mu_{x=0}}, \sigma_{\gamma_1}^2 = \frac{\sigma_u^2}{SSx}, \bar{x} = \sum x_i/n_v$ and $SSx = \sum(x_i - \bar{x})^2$. These data-based priors are equivalent to the posteriors of the linear regression model with Jeffreys priors as derived in [18]. Using a conjugate, non-informative inverse-gamma prior leads to the inverse gamma prior above, but as discussed earlier, if the validation sample size is small, a half-Cauchy prior is preferred. This loses some computational convenience, but is easily accommodated in standard software packages.

Finally, we note that in some cases, the external validation data may "swamp" the main study data, in which case it would be appropriate to down weight the data via a process such as found in [19].

TABLE 1. Summary statistics for unmeasured confounder across binary exposure variable

	x = 0	x = 1
Mean	$\overline{\mu_{x=0}}$	$\overline{\mu_{x=0}}$
Standard Deviation	$S_{x=0}$	$S_{x=0}$
Sample Size	$n_{x=0}$	$n_{x=0}$

RESULTS

We used simulated datasets to illustrate the performance of the proposed Bayesian model. The parameters of the simulation are based on the empirical data from [4]. The publication is a secondary analysis of data from the Treatment With Adenosine Diphosphate Receptor Inhibitors–Longitudinal Assessment of Treatment Patterns and Events After Acute Coronary Syndrome (TRANSLATE-ACS) study. Included in the study were patients undergoing percutaneous coronary intervention for myocardial infarction and the study dates were April 4, 2010, to October 31, 2012. We used information on the key demographic variables age, gender and race, and the proportion of patients receiving different ADP inhibitors (Table 1, [4]) to inform our simulation setting. To be more specific, x_i represents patients receiving different treatments, and z_1, z_2, z_3 correspond to age, gender and race, respectively. The unmeasured confounder was scaled to range from 0 to 1 to represent either a single unmeasured confounder or a propensity score summarizing multiple confounders.

For every simulation configuration we generated 100 data sets. We examined scenarios with internal and external validation data along with a situation that mimics informative priors.

The model used for the unmeasured confounder for the simulations was

$u_i \sim N(0.36+0.1x_{1i}, 0.2)$

Note that only the treatment is included in the unmeasured confounding model, which as previously mentioned, is the worst case scenario in terms of bias. For the Cox model we used

$exp(-0.3x_i+0.01z_{1i}+0.013z_{2i}+0.03z_{3i}+1.06u_i)$

Such simulation parameters were chosen so that the unmeasured confounder had at least moderate impact on the outcome. If the unmeasured confounder only had mild impact on the outcome, then ignoring it may not lead to significant bias in the treatment effect estimate. We assumed a main study sample size of 1000 and sample sizes for validation data of 50, 100, and 200. We generated an internal validation data set with all variables including the unmeasured confounder and an external validation data set where only summary statistics for the unmeasured confounder and exposure variable were obtained. Thus, no information for is provided for the external validation data. The informative priors were obtained by generating

TABLE 2. Simulation results across 100 data sets with censoring rate of 60%. Empirical coverage is for 95% nominal intervals.

Validation Sample Size	Validation Type	Bias	Coverage Probability	Interval Width
50	Internal	0.015	0.93	0.75
	External	0.076	0.92	0.95
	Informative	0.098	0.90	0.78
	Naïve	0.124	0.80	0.46
100	Internal	0.039	0.96	0.54
	External	0.069	0.94	0.83
	Informative	0.095	0.89	0.78
	Naïve	0.124	0.80	0.46
200	Internal	0.045	0.98	0.48
	External	0.079	0.85	0.69
	Informative	0.082	0.95	0.72
	Naïve	0.124	0.80	0.46

Footnote: Naïve model ignores unmeasured confounders in the analyses.

TABLE 3. Simulation results across 100 data sets with censoring rate of 90%. Empirical coverage is for 95% nominal intervals.

Validation Sample Size	Validation Type	Bias	Coverage Probability	Interval Width
50	Internal	0.032	0.98	1.41
	External	0.032	0.89	1.84
	Informative	0.100	0.97	1.95
	Naïve	0.080	0.93	1.00
100	Internal	0.013	0.98	1.18
	External	0.090	098	1.71
	Informative	0.034	0.97	1.77
	Naïve	0.080	0.93	1.00
200	Internal	0.034	0.93	1.00
	External	0.052	0.96	1.58
	Informative	0.034	0.97	1.71
	Naïve	0.080	0.93	1.00

Footnote: Naïve model ignores unmeasured confounders in the analyses.

a single data set for each of the validation sample sizes and using this data to inform independent normal priors for each of the regression parameters. These priors were then used for every data set in the simulation as opposed to the validation data methods where a new validation data set was used each time. This was done to allow comparison of the methods with basically the same sample size equivalence but the informative prior is the same every time. We considered censoring rates of 60%, and 90%. In the real data, the censoring was close to 90%. For the validation data scenarios we used fairly diffuse priors for all regression parameters, specifically, normal with mean 0 and standard deviation of 10. Relatively diffuse priors were used for σ_u^2 and the λ_l. For all analyses, we used the RJAGS package in R. Two chains were run with a burn-in of 7000 iterations which was followed by 20000 iterations keeping every 10th for inference. Convergence was checked with Gelman-Rubin statistics and monitoring history plots. Sample plots are provided in the online appendix.

We provide results for the main parameter of interest, β, for 60% censoring in Table 2 and 90% censoring in Table 3. The results are the average bias along with empirical coverage and widths for 95% nominal intervals across the 100 simulated data sets for each different scenario. The bias was calculated as the absolute difference between the posterior mean and the true parameter value. Several interesting points are worth mentioning. The bias for each of the correction methods is smaller than the naïve model for all scenarios considered except one case with 90% censoring where the informative prior case is slightly larger. Coverage is close to nominal in most cases for the corrected models while the naïve model has coverage considerably below nominal for the 60% censoring case. All confidence intervals are fairly wide due to the large censoring rate (60% and 90%). Internal validation has consistently narrower widths than the external and informative cases. In fact, when the internal validation sample size is 200 (which is 20% of the total sample) the widths of the 95% intervals are very close to those in the naïve case. While validation data of size 50, or informative priors based on an equivalent sample size of 50 do seem to correct the bias and provide intervals with nominal coverage, we found that using validation data samples of less than 50 resulted in considerable convergence problems, indicating that perhaps validation

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

e12634-5

fractions (ratio of validation sample size to main study sample size) much below 5% are not particularly useful.

Another interesting point about the 90% censoring is that even though the naïve model is the most biased, it has coverage that is fairly close to nominal. This is because with high censoring the uncertainty in estimation is fairly high for all the models thus the 95% intervals are all wide and cover the truth most of the time even with the bias. We considered other parameter configurations (motivated by another real world example where unmeasured confounder was weakly associated with treatment choice but strongly impact the outcome), along with the case of a binomial unmeasured confounder and these simulations are described in the online appendix. In general, the results of the simulations with different parameter setting yields very consistent results as the results we presented here.

DISCUSSION

Unmeasured confounding continues to be a major problem in observational studies. New methods that correct for bias due to unmeasured confounding for more complicated models than linear and logistic regression are important areas of research. Several advanced statistical methods – both from frequentist and Bayesian perspectives – have been developed during the past decades in addressing this challenge. However, only a few frequentist methods have been proposed for survival models, and none of them focus on time-to-event data specifically. Here, we have provided an overview of a Bayesian model-based approach to account for unmeasured confounding when supplementary information is available in the form of internal or external validation data. Subjectively elicited informative prior distributions are another alternative. A flexible parametric approximation of the Cox proportional hazard model from [14] was implemented to construct the posterior likelihood estimation incorporating both main study and supplemental information. Also, we have conducted simulation studies under different scenarios based on empirical data from a published comparative cardiovascular study. We found that even when the unmeasured confounder is only mildly related to the exposure, as long as the unmeasured confounding at least moderately impact the outcome, non-ignorable bias still results in the naïve model. Further, we demonstrated that our Bayesian method can correct for such bias using rather modest amounts of validation data.

There are several limitations of our methods that should be discussed to perhaps motivate further research. First we are assuming the unmeasured confounder is, in fact, measurable in either the current or a similar data set. If it is an unmeasurable latent trait, our methods will not apply. The use of external validation data requires several assumptions. First, the population that the validation data comes from is similar enough to the current data so

that the parameters being estimated are approximately the same. Secondly, we require that the same variables be measured. Finally, we focus on relatively simple relationships with the unmeasured confounder. Indeed, we are assuming the information available (either validation data or expert opinion) will rarely be rich enough to model complex relationships.

In this paper we have focused on simulated data where the true value of the parameters is known and the extent of the bias due to the unmeasured confounding can be ascertained precisely. In future research we will provide a thorough analysis of real world data and provide further guidelines on how to address unmeasured confounding in time-to-event data.

References

1. Groenwold, R H, de Groot, M C, Ramamoorthy, D, Souverein, P C, & Klungel, O H. Unmeasured confounding in pharmacoepidemiology. Annals of Epidemiology. 2016; 26, 85.

2. Schneeweiss, Sebastian. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiology and Drug Safety. 2006; 15, 291-303.

3. Stürmer, Til, Schneeweiss, S., Avorn, J, and Glynn, RJ, Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. American Journal of Epidemiology. 2005; 162, 279-289.

4. Federspiel, JJ, Anstrom, KJ, Xian, Y, McCoy, LA, Effron, MB, Faries, DE, Wang, T Y. Comparing Inverse Probability of Treatment Weighting and Instrumental Variable Methods for the Evaluation of Adenosine Diphosphate Receptor Inhibitors After Percutaneous Coronary Intervention. Jama Cardiology. 2016; 1, 655-665.

5. Rosenbaum, Paul R., and Donald B. Rubin. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society. Series B. 1983; 212-218.

6. Greenland, Sander. Multiple bias modelling for analysis of observational data. Journal of the Royal Statistical Society: Series A. 2005; 168, 267-306.

7. McCandless, L. C., Gustafson, P., & Levy, A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. Statistics in medicine. 2007; 26, 2331-2347.

8. Faries, D., Peng, X., Pawaskar, M., Price, K., Stamey, J. D., & Seaman, J. W. Evaluating the impact of unmeasured confounding with internal validation data: an example cost evaluation in type 2 diabetes. Value in Health. 2013; 16, 259-266.

9. Lin, D., Psaty, B., and Kronmal, R. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998; 1, 948-963

10. Stamey, J. D., Beavers, D. P., Faries, D., Price, K. L., & Seaman, J. W. Bayesian modeling of cost effectiveness studies with unmeasured confounding: a simulation study. Pharmaceutical statistics, 2014; 13, 94-100.

11. Zhang, X., Faries, D. E., Boytsov, N., Stamey, J. D., & Seaman, J. W. A Bayesian sensitivity analysis to evaluate the impact of unmeasured
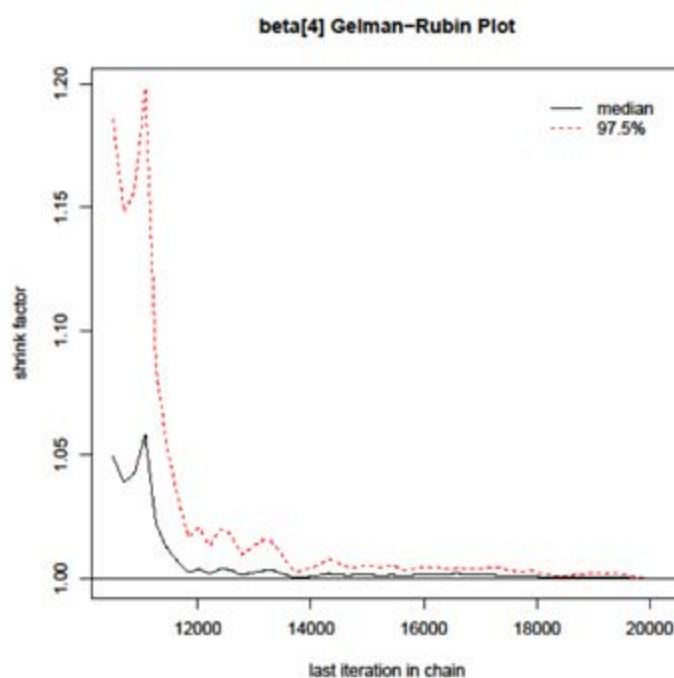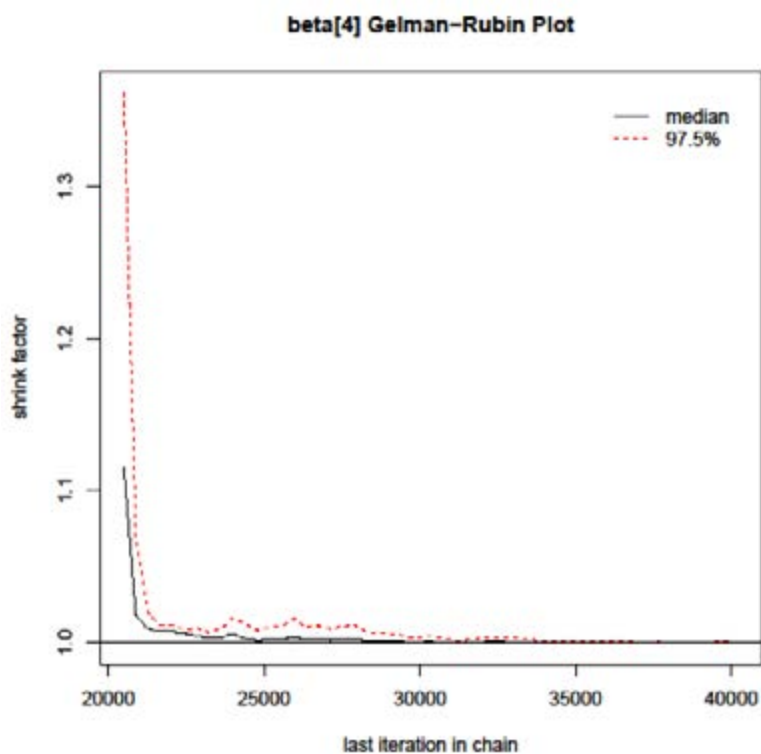
e12634-6

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

confounding with external data: a real world comparative effectiveness study in osteoporosis. Pharmacoepidemiology and Drug Safety. 2016; 25, 982-992.

12. Rosenbaum, P. R., & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. Biometrika, 1983; 1, 41-55.

13. Fewell, Z., Smith, G. D., & Sterne, J. A. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. American Journal of Epidemiology. 2007; 166, 646-655.

14. Christensen, Ronald, Wesley Johnson, Adam Branscum, and Timothy E. Hanson. Bayesian ideas and data analysis: an introduction for scientists and statisticians. CRC Press, 2011.

15. Spiegelhalter, D. J., Thomas, A., Best, N. and Lunn, D. (2007) WinBugs Version 1.4.3: User Manual. Cambridge: Medical Research Council Biostatistics Unit.

16. Polson, N. G., & Scott, J. G. On the half-Cauchy prior for a global scale parameter. Bayesian Analysis. 2012; 7, 887-902.

17. Stamey, J. D., Young, D. M., & Seaman, J. W. A Bayesian approach to adjust for diagnostic misclassification between two mortality causes in Poisson regression. Statistics in medicine, 2009; 27, 2440-2452.

18. Box, G. E. P.; Tiao, G. C. (1973). Bayesian Inference in Statistical Analysis. Wiley.

19. Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science, 2017; 32, 1-28

*

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets
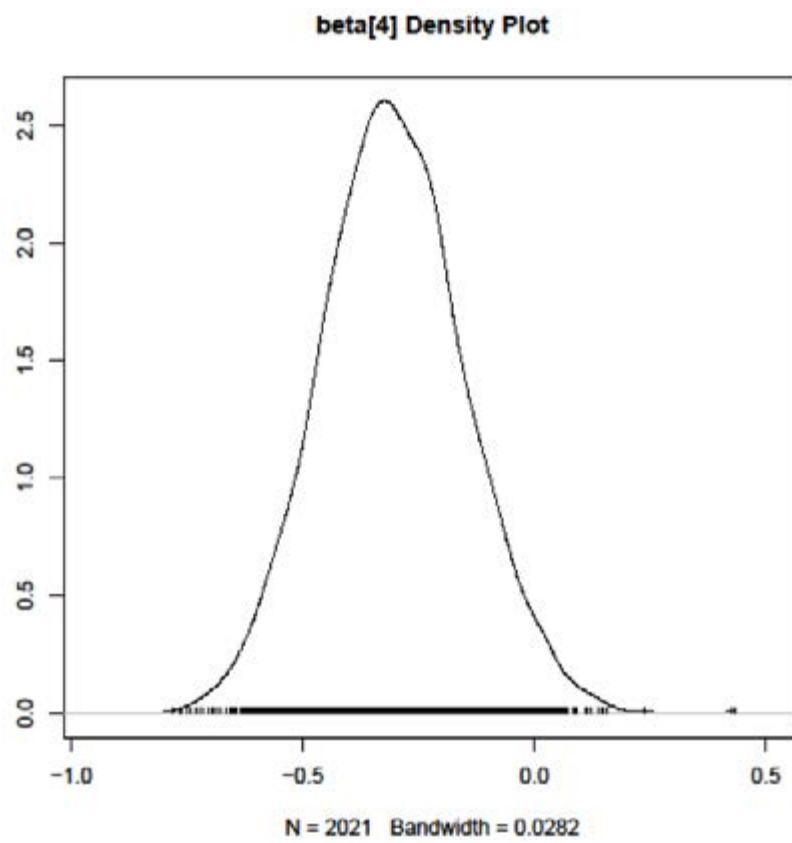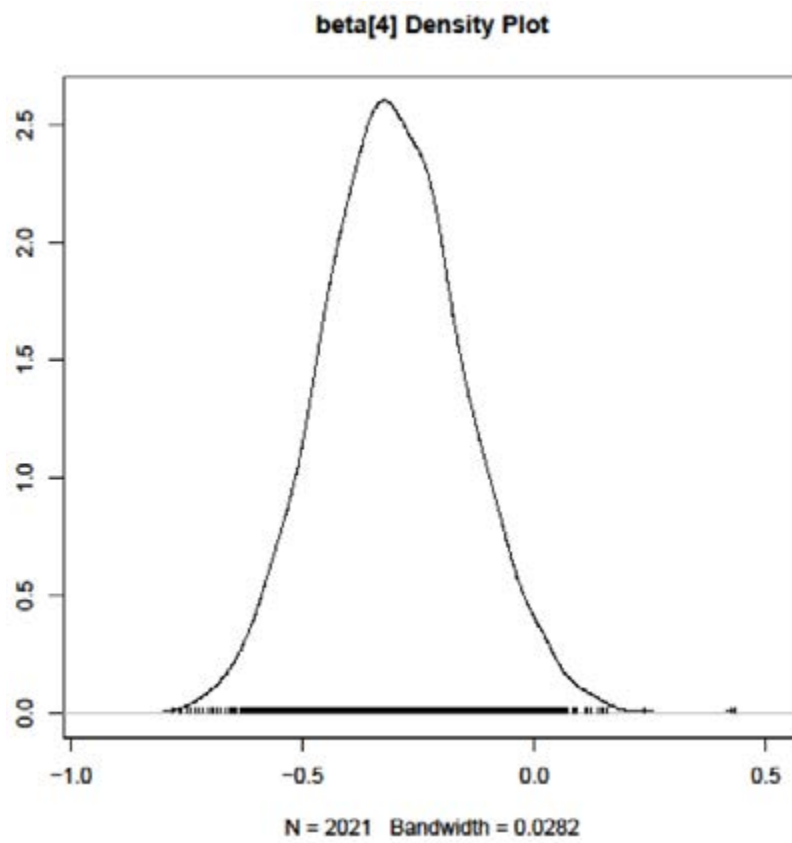
e12634-7

Supplementary material
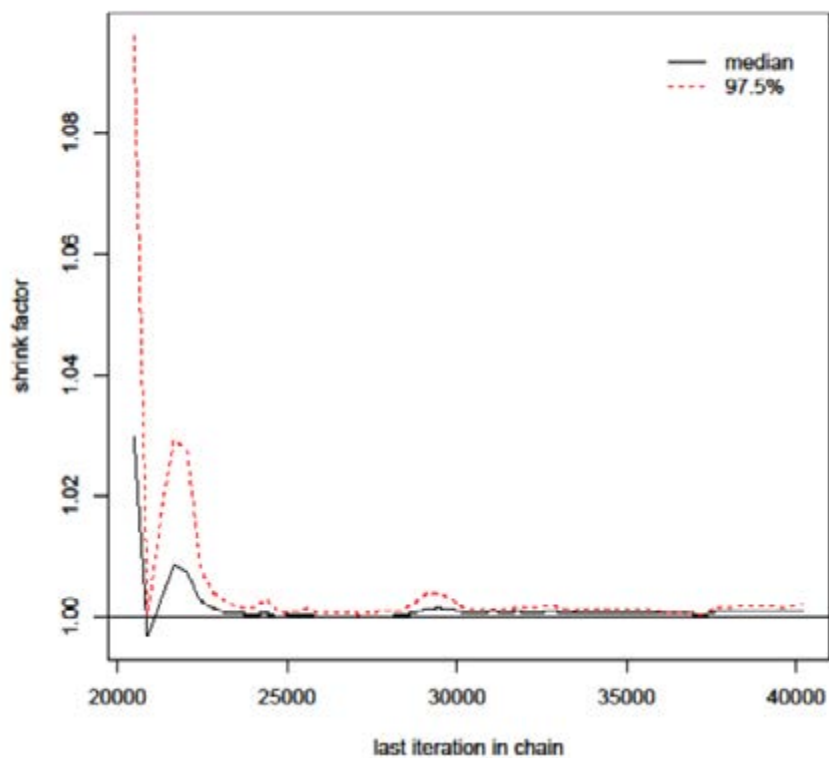
1. Convergence diagnostics

Below are representative convergence plots for the simulation studies illustrating convergence. These are for 60% censoring for the informative prior case (no validation data).

beta[4] Gelman−Rubin Plot

beta[4] Gelman−Rubin Plot

beta[4] Density Plot

N = 2021   Bandwidth = 0.0282

beta[4] Density Plot

N = 2021   Bandwidth = 0.0282

eta[2] Gelman−Rubin Plot

eta[2] Density Plot

N = 2021   Bandwidth = 0.002582

e12634-10

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets
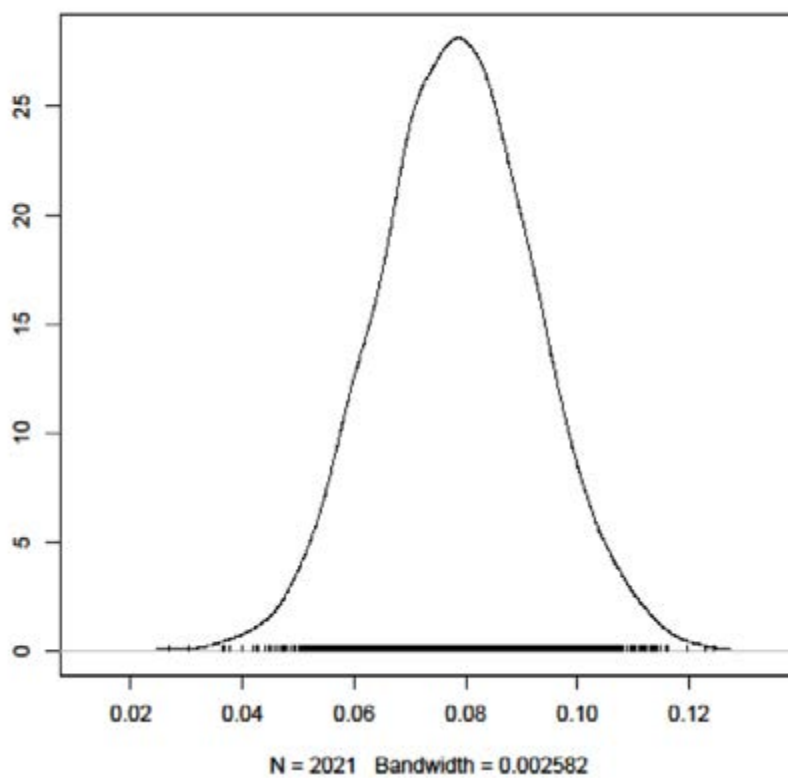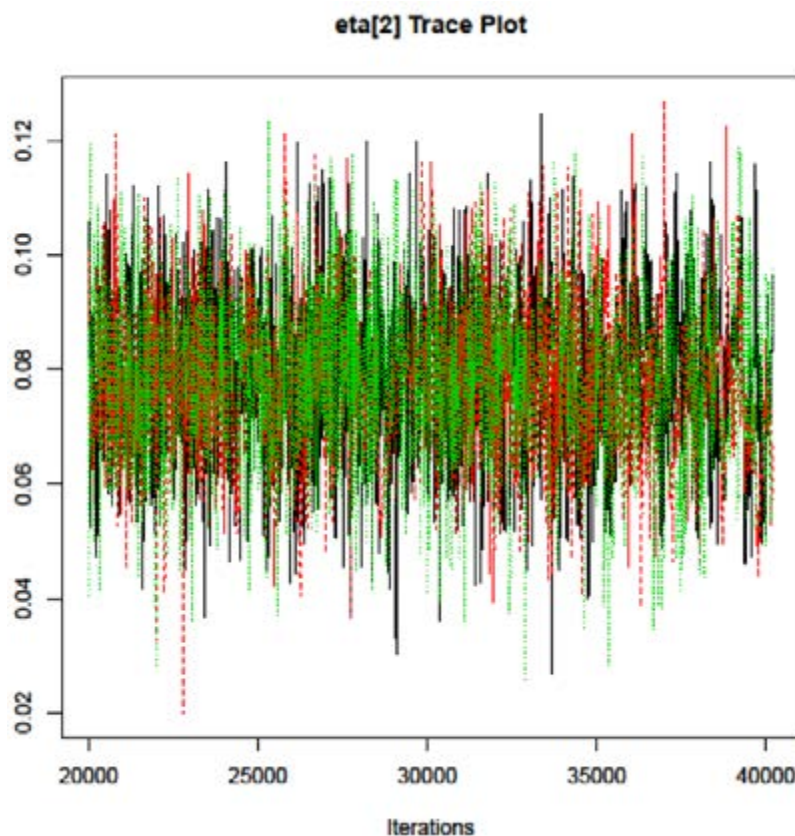
eta[2] Trace Plot

2. Additional Simulation Results

We now describe some results from additional simulations. The data parameters chosen were motivated by the results in Connors (1996). Note that there is only a very small relationship between the treatment and the unmeasured confounder. The exposure variable, x, and three measured covariates were all generated from binary distributions with probabilities 0.4, 0.08, 0.24 and 0.11 respectively. The model used for the unmeasured confounder for the simulations was

$u_i \sim N(0.6-04x_{1i}, 0.2)$.

For the Cox model we used

$exp(0.3x_i - 1.2z_{1i} + 0.4z_2 + z3 - 3u)$

We assumed a main study sample size of 1000 and validation sample size of 100. We generated an internal validation data set with all variables including the unmeasured confounder and an external validation data set where only summary statistics for the unmeasured confounder and exposure variable were obtained. We considered censoring rates of 30%, 60%, and 90%. We used fairly diffuse priors for all regression parameters, specifically, normal with mean 0 and standard deviation of 10. We used a diffuse inverse gamma prior for   and diffuse gamma priors for the . The results in Supplemental Table 1 yield similar relationships as the results discussed in the main manuscript. Using validation data improves bias and coverage for lower censoring rates but only improves bias in higher censoring rate cases due to wide intervals for all methods.

TABLE 1. Simulation results for β.

Censoring	Validation Type	Bias	Coverage Probability	Interval Width
	Internal	0.00	0.975	0.34
30%	External	0.01	0.95	0.34
	Naive	0.17	0.825	0.31
	Internal	0.07	0.98	1.11
90%	External	0.13	0.93	1.05
	Naive	0.26	0.95	0.83

We now also consider the case of a binary unmeasured confounder. Specifically, we generate the data via the following:

$u_i \sim$ Ber (p_i)
$logit (p_i) = -1.3 - 1.6x$

For the Cox model, we used

$$exp(0.3x_{li} - 1z_{li} + 0.3z_{2i} + 0.005z_{2i} - 1.4u_i)$$

Again, the main study size was 1000 with 100 subjects for the validation samples. The results for 30% and 90% censoring are provided in Supplemental Table 2. For the binary unmeasured confounder case the naïve model does worse than it did with a continuous unmeasured confounder with more bias and lower coverage. The internal and external validation models again reduce the bias and provide coverage at close to nominal but this is with the price of wider intervals. As in previous simulation examples, the case of 90% censoring has extremely wide intervals.

TABLE 2. Simulation results for β with binary unmeasured confounder.

Censoring	Validation Type	Bias	Coverage Probability	Interval Width
30%	Internal	0.00	0.98	0.35
	External	0.06	0.95	0.43
	Naive	0.12	0.78	0.31
90%	Internal	0.01	0.93	1.37
	External	0.13	0.97	2.09
	Naive	0.35	0.65	0.95

4. JAGS Code

The JAGS code for the model with a continuous unmeasured confounder and subject level validation data (either internal or external) is provided below.

```
model
{

#loop for main study data

for(i in 1:n){
    for (k in 1:J){
    d[i,k]~dpois(mu[i,k])
    mu[i,k]<-(min(dur[i],a[k+1])-a[k])*step(dur[i]-a[k])*
    lam[k]*exp(beta*x1[i]+theta[1]*z1[i]+theta[2]*z2[i]+theta[3]*z3[i]+eta*u[i])
    }
     u[i]~dnorm(gamma[1]+gamma[2]*x1[i], sigu)

}

## loop for validation data

for(j in 1:n.tilde){
    for (m in 1:J){
    d.tilde[j,m]~dpois(mu.tilde[j,m])
    mu.tilde[j,m]<-(min(dur.tilde[j],a[m+1])-a[m])*step(dur.tilde[j]-a[m])*
    lam[m]*exp(beta*x1.tilde[j]+theta[2]*z1.tilde[j]+theta[3]*z2.tilde[j]+
```

e12634-12

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

```
theta[4]*z3.tilde[j]+eta*u.tilde[j])
}
  u.tilde[j]~dnorm(gamma[1]+gamma[2]*x1.tilde[j], sigu)
}

# prior information
beta~dnorm(0, 0.1)
for(k in 1:J){lam[k]~dgamma(0.01,0.01)}
for(l in 1:3){theta[l]~dnorm(0,0.1)}
for(l in 1:2){gamma[l]~dnorm(0,0.1)}
sigu~dgamma(0.01, 0.01)
eta~dnorm(0,0.1) ## non-informative prior
}
```

JAGS code for continuous unmeasured confounder using informative priors is given below. Note normal priors are in terms of means and precisions, so standard deviations have to be transformed, precision = 1/variance.

```
model
{
for(i in 1:n){
    for (k in 1:J){
    d[i,k]~dpois(mu[i,k])
    mu[i,k]<-(min(dur[i],a[k+1])-a[k])*step(dur[i]-a[k])*
    lam[k]*exp(beta*x1[i]+theta[1]*z1[i]+theta[2]*z2[i]+theta[3]*z3[i]+eta*u[i])
    }
    u[i]~dnorm(gamma0+gamma1*x1[i], taum)
}

    # prior information
  for(k in 1:J){lam[k]~dgamma(0.01,0.01)}
  for(l in 1:3){theta[l]~dnorm(theta.mean[l], theta.prec[l])}
  gamma0~dnorm(gamma0.mean, gamma0.prec)
  gamma1~dnorm(gamma0.mean, gamma0. prec)
  taum~dgamma(tau.a, tau.b)
  beta~dnorm(beta.mean, beta. prec) ## non-informative prior
  eta~dnorm(eta.mean, eta prec) ## non-informative prior
}
```

*

A Bayesian approach to correct for unmeasured or semi-unmeasured confounding in survival data using multiple validation data sets

e12634-13