Bias and Sensitivity Analysis When Estimating Treatment Effects from the Cox Model with Omitted Covariates

Nan Xuan Lin,[1,2] Stuart Logan,[1] William Edward Henley[1,2,*]

[1]Institute of Health Research, University of Exeter Medical School, Exeter, U.K.
[2]Centre for Health and Environmental Statistics, University of Plymouth, Plymouth, U.K.
*email: w.e.henley@exeter.ac.uk

SUMMARY. Omission of relevant covariates can lead to bias when estimating treatment or exposure effects from survival data in both randomized controlled trials and observational studies. This paper presents a general approach to assessing bias when covariates are omitted from the Cox model. The proposed method is applicable to both randomized and non-randomized studies. We distinguish between the effects of three possible sources of bias: omission of a balanced covariate, data censoring and unmeasured confounding. Asymptotic formulae for determining the bias are derived from the large sample properties of the maximum likelihood estimator. A simulation study is used to demonstrate the validity of the bias formulae and to characterize the influence of the different sources of bias. It is shown that the bias converges to fixed limits as the effect of the omitted covariate increases, irrespective of the degree of confounding. The bias formulae are used as the basis for developing a new method of sensitivity analysis to assess the impact of omitted covariates on estimates of treatment or exposure effects. In simulation studies, the proposed method gave unbiased treatment estimates and confidence intervals with good coverage when the true sensitivity parameters were known. We describe application of the method to a randomized controlled trial and a non-randomized study.

KEY WORDS: Bias analysis; Cox model; Omitted covariates; Sensitivity analysis; Survival analysis; Unmeasured confounding

1. Introduction

Treatment or exposure effects are commonly estimated from survival or other time-to-event data using the Cox model. The gold standard design for conducting such evaluations is the randomized controlled trial because randomization acts to balance measured and unmeasured confounders. Although it is common for researchers to present unadjusted analyses, it is recommended to adjust proportional hazards models for all measured covariates in randomized studies to maximise power to detect treatment effects (Hernandez, Eijkemans, and Steyerberg, 2006). Gail, Wieand, and Piantadosi (1984) derived asymptotic formulae for the bias in estimates of treatment effects when balanced covariates are omitted from the Cox model. It was shown that when censoring is moderate, the Cox model yielded more biased estimates of treatment effect than analysis with the exponential model.

In practice, randomized experiments may be difficult to conduct for reasons of cost, logistics or ethics (Black, 1996). The increasing availability of electronic medical record databases and population-based studies is creating new opportunities for using observational data to assess the effect of medical treatments and exposures (Ghani et al., 2001; Tannen, Weiner, and Xie, 2009). A major challenge in using clinical databases in this way is addressing the potential bias introduced due to unmeasured differences between the treatment groups (Klungel et al., 2004). Lin, Psaty, and Kronmal (1998) presented approximate formulae for the bias due to omission of a binary or continuous confounder when estimating treatment effects from censored survival time data using the Cox model. The bias formulae were used as the basis for a method of conducting sensitivity analysis to assess how the point and interval estimates of the treatment effect vary under a range of assumptions about the unmeasured confounder. The idea behind this approach is that the plausibility of the estimated treatment effects will increase if the inferences are insensitive over a wide range of relevant scenarios.

In this paper, we develop a general framework for estimating bias and conducting sensitivity analysis when covariates are omitted from the Cox model. Formulating the problem more broadly than previous work, we consider the combined influence of three different sources of bias: (1) bias due to omitting a balanced covariate; (2) bias due to censoring; (3) bias due to the missing covariate being a confounder. The proposed approach is applicable to both randomized trials and observational studies, and provides explicit formulae for arbitrary distributions of measured and unmeasured confounders. We consider the general case in which the censoring distribution can depend on treatment or other covariates. The treatment variable can be either a binary or continuous exposure.

The paper is organized as follows. Asymptotic bias formulae, derived from the large sample properties of the partial maximum likelihood estimators, are presented in Section 2. Simulation studies conducted to investigate the accuracy of the bias formulae and to characterize the impact of the different sources of bias are presented in Section 3. Section 4 discusses how the bias formulae can be used to develop a

new method of sensitivity analysis for treatment effects in proportional hazards models. The method is applied to data from a randomized controlled trial and a non-randomized study in Section 5.

2. Bias Formulae

We denote random variables by upper case letters and their values by lower case letters. Suppose $X = (X_1, \ldots, X_K)$ are K measured covariates with joint distribution $f(x)$, and $C = (C_1, \ldots, C_q)$ are q unmeasured covariates with conditional joint distribution $f(c|x)$. Let T and T^+ represent the true event/failure time and possible censoring time respectively. We assume failure and censoring times are independent conditional on x (i.e., $T^+ \perp T|x$). We observe $(\tilde{T}, \Delta)$, where $\tilde{T} = \min(T, T^+)$, and $\Delta = 1$ if $T \leq T^+$ and 0 otherwise. The true hazard is assumed to be

$$h(t|X, C) = h_0(t) \exp(X\theta + C\beta), \qquad (1)$$

where $h_0(t)$ is the baseline hazard function and $\theta = (\theta_1, \ldots, \theta_K)^{tr}$ and $\beta = (\beta_1, \ldots, \beta_q)^{tr}$ are coefficients for X and C, respectively. But since C is omitted, one is forced to fit the model

$$h(t|X) = h_0^*(t) \exp(X\theta^*), \qquad (2)$$

where $\theta^* = (\theta_1^*, \ldots, \theta_K^*)^{tr}$ are the coefficients when C is missing. Let $(\tilde{T}_i, \Delta_i, X_i)$ be $\tilde{n}$ independent replicates of $(\tilde{T}, \Delta, X)$. Then the average partial log-likelihood based on (2) is

$$l(\theta^*) = \frac{1}{\tilde{n}} \sum_i^{\tilde{n}} \Delta_i \left(X_i \theta^* - \log \sum_j^{\tilde{n}} Y_j(\tilde{T}_i) \exp(X_j \theta^*) \right), \quad (3)$$

where $Y_j(\tilde{T}_i) = 1$ if $\tilde{T}_j \geq \tilde{T}_i$ and 0 otherwise. It is shown in Web Appendix A that as $\tilde{n} \to \infty$, the score function $\partial l(\theta^*)/\partial \theta_k^*$ has the limit

$$0 = U_k = U_k(\theta^*, \theta; \beta)$$

$$= E_{OBS} \left[X_k - \frac{E_{xc} \left\{ e^{X\theta^*} e^{-H_0(T)e^{X\theta+C\beta}} X_k S^+(T|X) \right\}}{E_{xc} \left\{ e^{X\theta^*} e^{-H_0(T)e^{X\theta+C\beta}} S^+(T|X) \right\}} \right] \quad (4)$$

for $k = 1, \ldots, K$, where E_{OBS} (OBS=observed events) is the mean over the uncensored subjects, E_{xc} is under the density $f(c|x)f(x)$ and $S^+(T|X)$ is the survival function of censoring time conditional on X. Inclusion of $S^+(T|X)$ allows the censoring distribution to depend on covariates.

The system of equations (4), $0 = U = (U_1, \ldots, U_K)^{tr}$, relate θ^* and θ, and therefore the asymptotic biases $\theta^* - \theta$ can be evaluated from them. The first-order Taylor series approximation is

$$\theta^* - \theta \approx I^{-1}(\theta)U(\theta, \theta; \beta), \quad \text{with } I(\theta^*) = -\partial U/\partial \theta^{*tr}. \quad (5)$$

2.1. The Distributions of Uncensored Subjects
Let

$$\varphi_{xc} = P(T^+ \geq T|x, c) = \int P(T^+ \geq t|x, c)P(T = t|x, c)\mathrm{d}t$$

$$= \int S^+(t|x, c) f(t|x, c) \,\mathrm{d}t$$

be the uncensoring probability conditional on x and c, where $f(t|x, c)$ is the density of model (1) and $S^+(t|x, c)$ is the survival function of censoring time.

The density of the observed event times is then given by

$$f_{OBS}(t|x, c) = P(T = t|T^+ \geq T, x, c)$$

$$= \frac{P(T^+ \geq t|x, c)P(T = t|x, c)}{P(T^+ \geq T|x, c)}$$

$$= \frac{S^+(t|x, c)}{\varphi_{xc}} f(t|x, c).$$

The mean of X_k for uncensored subjects is

$$E_{OBS}(X_k) = \int x_k P(x_k|T^+ \geq T)\mathrm{d}x_k$$

$$= \frac{\int x_k P(T^+ \geq T|x_k) P(x_k)\mathrm{d}x_k}{P(T^+ \geq T)}$$

$$= \frac{E_{xc}(X_k \varphi_{XC})}{E_{xc}(\varphi_{XC})}. \quad (6)$$

2.2. Extension of the Results of Lin et al. (1998)
Lin et al. (1998) proposed bias formulae for survival analysis with unmeasured confounders based on the assumption of rare events (small $H_0(t)$) or small $|\beta|$. For binary x, the proposed bias approximation is

$$\theta^* - \theta \approx \begin{cases} \log \dfrac{e^\beta \rho_1 + (1 - \rho_1)}{e^\beta \rho_0 + (1 - \rho_0)} & \text{if } C|x \sim B(1, \rho_x) \\ \beta(\mu_1 - \mu_0) & \text{if } C|x \sim N(\mu_x, \sigma^2) \end{cases} \quad (7)$$

The simulation of Lin et al. (1998) showed that (7) are good approximations when t^+ was generated from the uniform $(0, \tau)$ distribution and the censoring percentage is 90%.

Using the assumption of rare events and the simulation settings in Lin et al. (1998), Web Appendix B shows that Equation (4) reduces to a simple equation of θ and θ^*:

$$\frac{E_x(Xe^{\theta^* X})}{E_x(e^{\theta^* X})} \approx \frac{E_{xc}(Xe^{\theta X + \beta C})}{E_{xc}(e^{\theta X + \beta C})}, \quad (8)$$

which leads to the formulae (7) when $X \sim B(1, P)$ and $C|x \sim B(1, \rho_x)$ or $C|x \sim N(\mu_x, \sigma^2)$. The Equation (8) therefore provides a general extension of the results in Lin et al. (1998) to arbitrary distributions of X and C.

3. Bias Analysis

3.1. *Bias Analysis for a Binary Treatment with a Single Omitted Covariate*

We now show the asymptotic formula for the bias in the important special case of a single missing covariate C and a binary exposure variable X taking values 1 or 0 with probabilities p and $1 - p$, respectively.

The Equation (4) leads to (see proof in Web Appendix C)

$$E_{OBS}\left\{V(\theta^*)\right\} = \left\{(1-p)\frac{\varphi_0}{\varphi_1} + p\right\}^{-1} \qquad (9)$$

with

$$V(\theta^*) = \left\{(1-p)\mathrm{e}^{-\theta^*}\frac{S^+(T|x=0)E_{c|x=0}\left(\mathrm{e}^{-H_0(T)\mathrm{e}^{\beta C}}\right)}{S^+(T|x=1)E_{c|x=1}\left(\mathrm{e}^{-H_0(T)\mathrm{e}^{\theta+\beta C}}\right)} + p\right\}^{-1},$$

where the expectations $E_{c|x}$ and E_{OBS} are under $f(c|x)$ and

$$f_{OBS}(t) = \int_x \int_c f_{OBS}(t|x,c)f(c|x)f(x)\mathrm{d}c\,\mathrm{d}x$$

respectively, and $\varphi_0/\varphi_1 = E_{c|x=0}(\varphi_{XC})/E_{c|x=1}(\varphi_{XC})$ is the ratio of uncensoring rates between control and treatment groups.

From (9), it can be seen that the relation between θ^* and θ mainly depends on three factors (corresponding to the three sources of bias): the effect of the missing covariate, β; censoring mechanism, E_{OBS}, $S^+(\cdot)$ and φ_0/φ_1; and the ratio of conditional expectations, $E_{c|x=0}[\cdot]/E_{c|x=1}[\cdot]$. The latter ratio represents how much the density of C varies between $x = 0$ and $x = 1$ and, hence, measures the extent to which C is a confounder.

The bias is also affected by the cumulative baseline hazard function $H_0(\cdot)$. But if times are not censored, $H_0(T)$ is an exponential variable with the rate $\mathrm{e}^{\theta x+\beta c}$ and (9) reduces to

$$E_z\left[\left\{(1-p)\mathrm{e}^{-\theta^*}\frac{E_{c|x=0}\left(\mathrm{e}^{-Z\mathrm{e}^{\beta C}}\right)}{E_{c|x=1}\left(\mathrm{e}^{-Z\mathrm{e}^{\theta+\beta C}}\right)} + p\right\}^{-1}\right] = 1, \quad (10)$$

where $Z = H_0(T) \sim Exp(\mathrm{e}^{\theta X+\beta C})$. As a result, the bias is independent of the form of $H_0(\cdot)$ in the absence of censorship.

When $E_{c|x=0}[\cdot]/E_{c|x=1}[\cdot] = 1$, C is not a confounder. In this case, Equation (10) shows that $\theta^* \neq \theta$ and, consequently, the MLE of the Cox model is still biased even if C is a balanced covariate. Bretagnolle and Huber-Carol (1985) studied the bias in this case and showed that the estimated effect is biased toward zero as $|\beta|$ increases. This is because the event times with $c \neq 0$ tend to zero as $\beta \to \infty$ and tend to ∞ as $\beta \to -\infty$. Consequently the subjects with $c \neq 0$ cannot provide information about θ in the limiting case. However, the subjects with $c = 0$ do still supply information about θ and hence the limit of $\hat{\theta}^*$ as $|\beta| \to \infty$ is not zero for binary C. An illustration of this explanation is given in Web Figure 1.
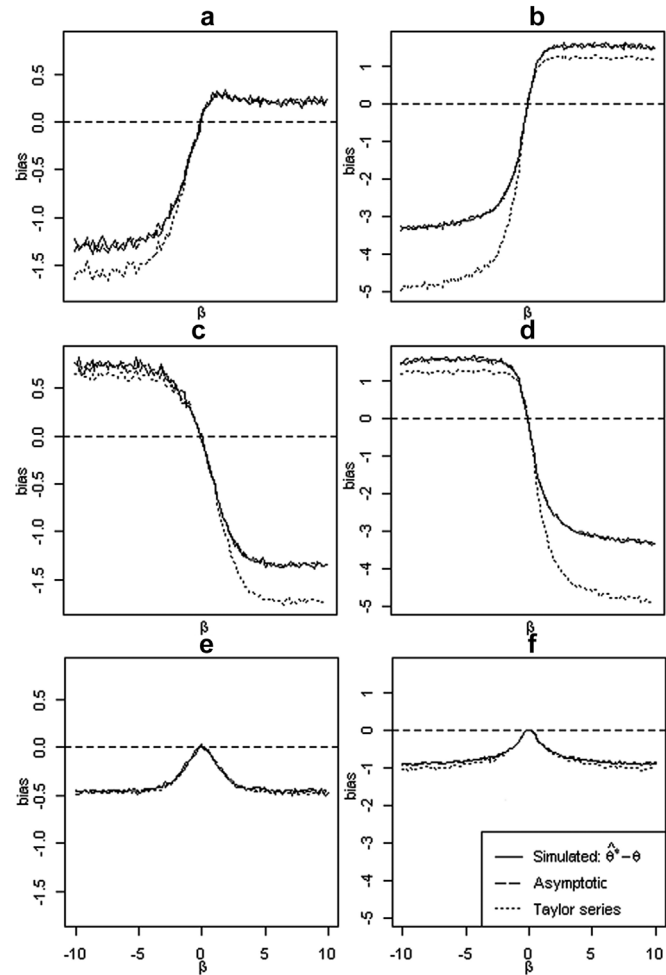
Figure 1. Comparison of simulated biases, asymptotic biases and first-order Taylor series approximations for different types of omitted covariate and censorship. Since θ^* is the asymptotic value of the MLE $\hat{\theta}^*$ and the sample size=10,000 is large, we calculated the simulated bias by $\hat{\theta}^* - \theta$. The asymptotic biases and Taylor series approximations were obtained from (9) and (11), respectively. Monte Carlo integration was used to approximate the expectations in formulae. (a) Binary confounder c: ($\rho_0 = 0.3, \rho_1 = 0.7$), censored; (b) Normal confounder c: ($\mu_0 = -1, \mu_1 = 1$), censored; (c) Binary confounder c: ($\rho_0 = 0.7, \rho_1 = 0.3$), censored; (d) Normal confounder c: ($\mu_0 = 1, \mu_1 = -1$), censored; (e) Binary balanced c: ($\rho_0 = \rho_1 = 0.5$), uncensored; (f) Normal balanced c: ($\mu_0 = \mu_1 = 0$), uncensored.

Following (5), the first-order Taylor series approximation is

$$\theta^* - \theta \approx \frac{\left\{(1-p)\varphi_0/\varphi_1 + p\right\}^{-1} - E_{OBS}\left\{V(\theta)\right\}}{E_{OBS}\left\{V(\theta) - V(\theta)^2\right\}}. \quad (11)$$

3.2. *Accuracy of Asymptotic Formulae and Taylor Series Approximations*

Figure 1 shows a comparison of the asymptotic and simulated biases and Taylor series approximation under the influence of different sources of bias. We generated 10,000 x from $B(1, 0.5)$.

The confounder C was generated from $B(1, \rho_x)$ for the binary confounder, and from $N(\mu_x, 1)$ for the normal confounder. The event times t were generated from model (1) with $h_0(\cdot) = 1$, $\theta = 1$ and β taking 100 sequence values from -10 to 10. For the censoring cases, we let $T^+ \sim uniform(0, \tau)$ with $\tau = 1$. The observed times were given by $\tilde{t} = \min(t, t^+)$.

Figure 1 shows that the simulated and asymptotic biases are seen to agree closely, confirming that these asymptotic formulas adequately describe the biases. The accuracies of the Taylor series approximations decrease as $|\theta^* - \theta|$ gets large, because the approximation error is of the order $O\left\{(\theta^* - \theta)^2\right\}$.

For more modest values of θ, for example 0.25 and 0.5, the biases will have similar patterns but be shifted up as $\theta \to 0$ (see Web Figures 2 and 3). In Web Figure 8, we let $\tau = 0.5 \exp(x)$ and $\tau = \exp(-3x)$ to allow the distribution of censoring to depend on treatment group. The figure illustrates how different choices of censoring function can impact on the biases.

3.3. *Bias of Omitting a Balanced Covariate in Randomized Studies*

Figure 1e and f show the biases when a balanced covariate is omitted. It is clear that omission of a relevant covariate leads to biased treatment estimates for the Cox model, even in randomized studies.

The reason is that the parameters θ and θ^* are measuring different features of the population. When we model the hazard as

$$h(t|x, c) = h_0(t) \exp(\theta x + \beta c) = \frac{f(t|x, c)}{S(t|x, c)},$$

the interpretation of $\exp(\theta) = h(t|x = 1, c)/h(t|x = 0, c)$ is the hazard ratio between $x = 1$ and $x = 0$ while the values of c are fixed. But in randomized studies (where we assume $f(c|x) = f(c)$), when we model the marginal hazard as

$$h(t|x) = h_0^*(t) \exp(\theta^* x) = \frac{f(t|x)}{S(t|x)} = \frac{\int_c f(t|x, c) f(c) \mathrm{d}c}{\int_c S(t|x, c) f(c) \mathrm{d}c}, \quad (12)$$

the interpretation of $\exp(\theta^*) = h(t|x = 1)/h(t|x = 0)$ is the hazard ratio between $x = 1$ and $x = 0$ while c is marginalized. Similarly, $h_0(t)$ is the hazard when $x = c = 0$, and $h_0^*(t)$ is the hazard when $x = 0$ and c is marginalized. The superscript $*$ emphasizes that they do not have the same interpretation.

When c is integrated out, the marginal hazards (12) for $x = 1$ and $x = 0$ are not proportional over time, and the MLE of θ^* represents an average over time of the log marginal hazard ratios between $x = 1$ and $x = 0$ (Lin and Wei, 1989). Therefore, it will lead to bias if we use a marginal hazard ratio $\exp(\theta^*)$ to estimate a hazard ratio $\exp(\theta)$. In randomized studies, as outlined in Section 3.1, usually $|\theta^*| < |\theta|$ and θ^* will attenuate to some limit between 0 and θ as $|\beta| \to \infty$.

3.4. *The Limits of Biases as $|\beta| \to \infty$*

One phenomenon that can be noticed from Figure 1 is that all biases increase with $|\beta|$ but always tend to some limits, no matter if C is a confounder or not. The reason is that the marginal hazard ratio has finite limits as β tends to ∞ and

$-\infty$. For example, for $C|x \sim B(1, \rho_x)$, the marginal hazard is

$$
\begin{aligned}
h(t|x) &= \frac{\int_c f(t|x, c) f(c|x) \mathrm{d}c}{\int_c S(t|x, c) f(c|x) \mathrm{d}c} \\
&= \frac{\rho_x h_0(t) e^{\theta x + \beta} e^{-H_0(t) e^{\theta x + \beta}} + (1 - \rho_x) h_0(t) e^{\theta x} e^{-H_0(t) e^{\theta x}}}{\rho_x e^{-H_0(t) e^{\theta x + \beta}} + (1 - \rho_x) e^{-H_0(t) e^{\theta x}}}.
\end{aligned}
$$

The ratio, $h(t|x = 1)/h(t|x = 0)$, tends to e^θ as $\beta \to \infty$ and

$$
\frac{(1 - \rho_1) h_0(t) e^\theta e^{-H_0(t) e^\theta}}{\rho_1 + (1 - \rho_1) e^{-H_0(t) e^\theta}} \Big/ \frac{(1 - \rho_0) h_0(t) e^{-H_0(t)}}{\rho_0 + (1 - \rho_0) e^{-H_0(t)}} \quad \text{as} \quad \beta \to -\infty.
$$

3.5. *The Effect of Censoring*

Figure 2a shows the effect of censoring on the bias of omitting a balanced covariate. The event times were generated from (1) with $h_0(t) = 1$, $\theta = 1$, $X \sim B(1, 0.5)$ and $C \sim B(1, 0.5)$. The possible censoring times t^+ were simulated from uniform $(0, \tau)$ with $\tau = \exp(\gamma_0 + \gamma_1 x)$.

Following the result A-4 in Web Appendix, the uncensoring probability can be written as

$$\varphi_{xc} = 1 - \int f^+(t|x, c) e^{-H_0(t) e^{\theta x + \beta c}} \mathrm{d}t,$$

where $f^+(t|x, c)$ is the density of possible censoring times.

Under the simulation settings, $H_0(t) = t$ and $f^+(t|x, c) = 1/\tau$. The probability of censoring conditional on x is thus

$$
\begin{aligned}
1 - \varphi_x = 1 - E_{c|x}(\varphi_{XC}) &= E_{c|x}\left\{\frac{1}{\tau} \int_0^\tau e^{-H_0(t) e^{\theta X + \beta C}} \mathrm{d}t\right\} \\
&= E_{c|x}\left\{\frac{1 - e^{-\tau e^{\theta X + \beta C}}}{\tau e^{\theta X + \beta C}}\right\} \quad (13)
\end{aligned}
$$

The values of γ_0 and γ_1 were then solved from (13) such that the probabilities of censoring were the same for $x = 0$ and $x = 1$ and could be 0%, 30%, 50%, 70%, and 90%. The number of event times n was fixed at 100,000 and the total sample size was $\tilde{n} = n/E_{xc}(\varphi_{XC})$.

Figure 2a shows that censoring influences the bias in two different ways. The bias increases as the censoring percentage increases from 0% to 50%, but decreases as the censoring percentage increases from 50% to 90%. The bias is plotted for a wider range of censoring percentages in Web Figure 4.

The reason for this inconsistent effect of censoring is as follows: when the censoring percentage increases (0–50%) and $\beta > 0$, the subjects with $c = 0$, which provide most of the information about θ, are likely to be censored, and consequently, the bias is increased. But as the censorship rate increases further (50–90%), almost all of the few events occur with $c = 1$ and almost all the times with $c = 0$ are censored. So nearly all the subjects supplying information about θ have the same value of $c = 1$ (Chastang, Byar, and Piantadosi, 1988). If the sample size is sufficiently large, the bias will tend to zero as the censoring percentage tends to 100%. A similar explanation applies for $\beta < 0$. An illustration of this explanation is given in Web Figure 5.
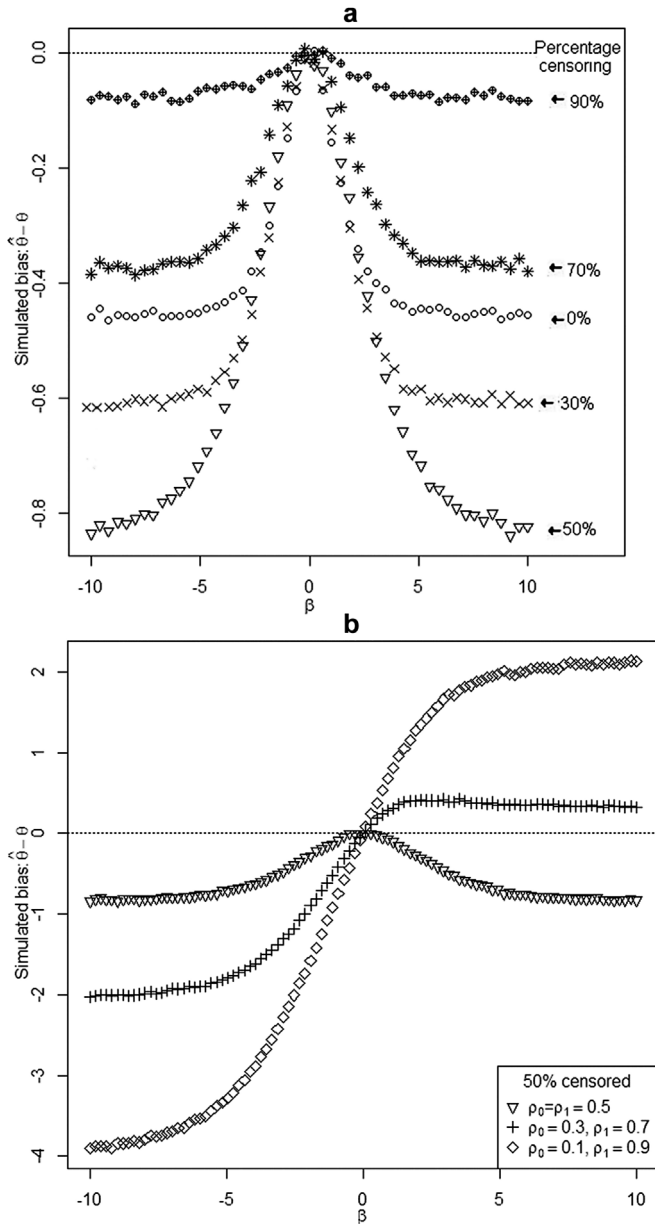
Figure 2. The effect of overall censoring and confounding on bias: (a) biases of omitting a balanced covariate where $0\% \rightarrow 90\%$ data are censored; (b) biases under different strengths of confounding, $(\rho_0, \rho_1) = (0.5, 0.5)$, $(0.3, 0.7)$ and $(0.1, 0.9)$ when 50% data are censored.

3.6. *The Effect of Confounding*

Of particular relevance to non-randomized studies, we considered the influence of different levels of confounding on the bias function when 50% of the data are censored (Fig. 2b). We generated $X \sim B(1, 0.5)$, $C|x \sim B(1, \rho_x)$ and consider three scenarios with $(\rho_0, \rho_1) = (0.5, 0.5)$, $(0.3, 0.7)$, $(0.1, 0.9)$. The difference $\rho_1 - \rho_0$ represents the imbalance of the distributions of $C|x \sim B(1, \rho_x)$ between $x = 1$ and $x = 0$ and so measures the strength of confounding. As $\rho_1 - \rho_0$ increases, it can be seen that the estimate is biased upwards for $\beta > 0$ and

downwards when $\beta < 0$. For the case $\rho_0 > \rho_1$, the bias would be affected in the other direction.

3.7. *The Effect of Additional Measured Covariates*

In practice, the analyst is likely to have access to additional measured covariates (possibly confounders) that would need to be adjusted for, in addition to the exposure variable X (and the unmeasured confounder, C).

Under the approach of Lin et al. (1998), an additional covariate Z does not affect the bias if the mean of C conditional on x and z is additive in x and z, that is $E(C|x, z) = q_1(x) + q_2(z)$ (VanderWeele, 2008). However, our simulation results in Figure 3a–c show that an additional covariate may introduce a small degree of bias when $|\beta|$ is large. We generated 100,000 $X \sim B(1, 0.5)$ and $C \sim B(1, 0.3 + 0.4x)$. The additional covariate Z was simulated from $B(1, 0.5)$, $B(1, 0.3 + 0.4x)$ and $B(1, 0.05 + 0.9x)$ for Figure 3a–c, respectively. Under these data-generating processes, $E(C|x, z) = E(C|x)$ and the additivity assumption is satisfied.

A sample of 100,000 survival times was generated from $h(t|x, c, z) = \exp(x + \beta c + \beta_2 z)$ and $T^+ \sim uniform(0, 1)$. The data were then fitted by the reduced model $h_0^*(t) \exp(\theta^* x + \beta_2^* z)$. It can be seen that the bias is not impacted by the distribution of Z, but is affected by β_2, when $|\beta|$ is large. The results were similar when we allowed censoring to depend on X and Z by assuming $T^+ \sim uniform(0, \exp(-\theta x - \beta_2 z))$ (see Web Figure 9).

It is then natural to investigate the influence of more than one additional covariate when $|\beta|$ is large. To simplify the problem, we examine the case where all the additional covariates are binary and independent of each other, with the same coefficient β_2. As the bias is only significant for large negative β, we set $\beta = -5$. The results displayed in Figure 3d, show the bias increases slightly with the number of covariates and the increments are not linear.

4. Sensitivity Analysis

The aim of our proposed method of sensitivity analysis is to assess how the point and interval estimators for θ or associated P-value would change given clinically plausible values of the sensitivity parameters β and $f(c|x)$.

4.1. *Point Estimate*

For a sample with $\tilde{n}$ observed times $\tilde{t}_1, \ldots, \tilde{t}_{\tilde{n}}$, of which n are uncensored $t_1, \ldots, t_n$, from (4) we have the relation between θ and θ^* approximately relies on the equations $0 = \widetilde{U}(\theta; \theta^*) = \widetilde{U} = \{\widetilde{U}_1, \ldots, \widetilde{U}_K\}^{tr}$ with

$$0 = \widetilde{U}_k = \frac{1}{n} \sum_{i=1}^{n}$$

$$\times \left[x_{ik} - \frac{\sum_{j=1}^{\tilde{n}} S^+(t_i|x_j) e^{x_j \theta^*} x_{jk} E_{c|x_j} \left\{ e^{-H_0(t_i) e^{x_j \theta + C\beta}} \right\}}{\sum_{j=1}^{\tilde{n}} S^+(t_i|x_j) e^{x_j \theta^*} E_{c|x_j} \left\{ e^{-H_0(t_i) e^{x_j \theta + C\beta}} \right\}} \right]$$

$$\text{for } k = 1, \ldots, K, \tag{14}$$

where the expectation $E_{c|x_j}(\cdot)$ can be calculated analytically or approximately with respect to $f(c|x)$.
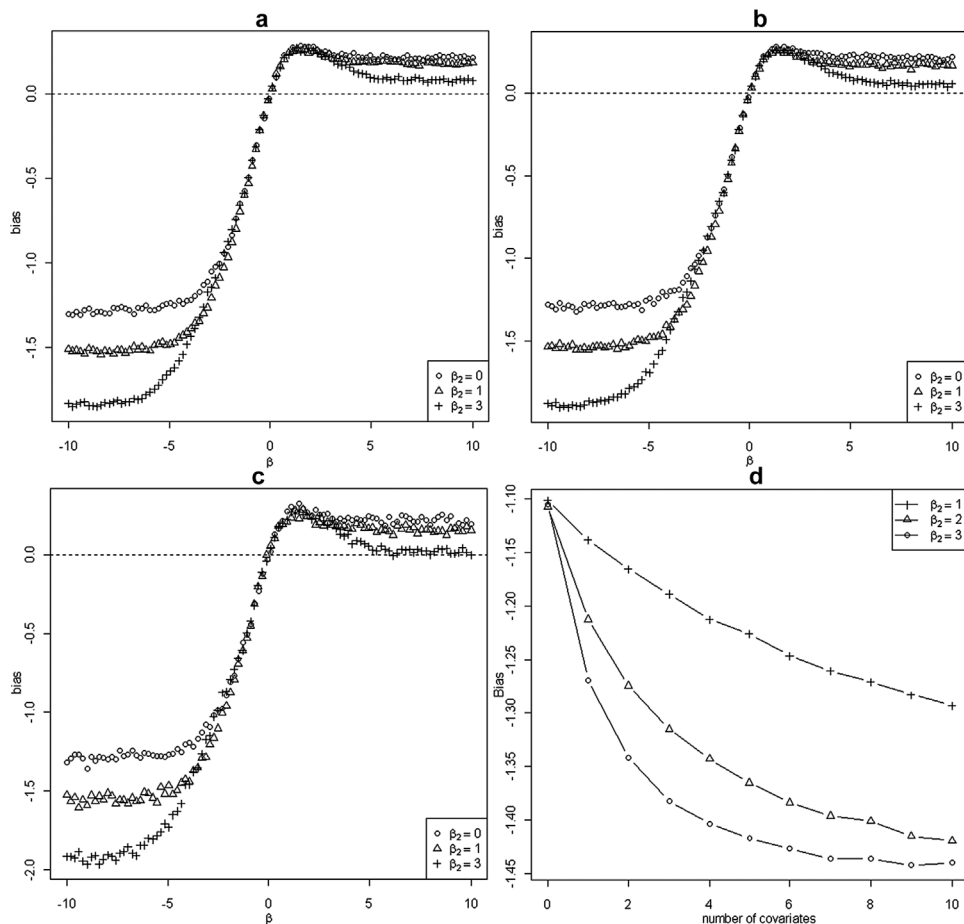
Figure 3. The effect of additional measured covariates on the simulated bias $\hat{\theta}^* - \theta$: (a) $Z|x \sim B(1, 0.5)$; (b) $Z|x \sim B(1, 0.3 + 0.4x)$; (c) $Z|x \sim B(1, 0.05 + 0.9x)$; (d) the effect of increasing the number of measured covariates on the simulated bias when $\beta_1 = -5$ and $\beta_2 = 1, 2$ and 3.

Write $\theta = R(\theta^*) = \arg_\theta \left\{ \widetilde{U}(\theta; \theta^*) = 0 \right\}$. Due to the functional invariance property of MLE, the point estimate of the true value θ is then $\hat{\theta} = R(\hat{\theta}^*)$. The function $R(\cdot)$ and its inverse $R^{-1}(\cdot)$ relate θ and θ^*, and play a key role in sensitivity analysis.

The baseline survivor function $e^{-H_0(t_i)}$ in (14) is estimated by solving

$$\hat{S}(t_i|x_i = 0) = E_{c|x_i=0}\left[e^{-\hat{H}_0(t_i)e^{C\beta}}\right],$$

where $\hat{S}(t_i|x_i = 0)$ is the Breslow (1972) estimator:

$$\hat{S}(t_i|x_i = 0) = \exp\left\{ -\sum_{m:t_m \leq t_i} \frac{1}{\sum_{j:\tilde{t}_j \geq t_m} e^{x_j \hat{\theta}^*}} \right\}.$$

The survival function of censoring can be also approximated by the Breslow (1972) estimator by considering events as "censored" observations and censored observations as "events" (Satten and Datta, 2001).

4.2. P-Values

In many applications, we are interested in evaluating the evidence the data give about a null hypothesis $H_0 : \theta = \theta_0$ (for example, that a hazard ratio equals one). Using $\theta = R(\theta^*)$, this null hypothesis is equivalent to $H_0 : \theta^* = R^{-1}(\theta_0)$ and the two-sided P-value is therefore

$$\text{P-value} = 2\Phi\left\{ -\left| \frac{\hat{\theta}^* - R^{-1}(\theta_0)}{\sigma^*} \right| \right\}, \tag{15}$$

where $\Phi\{\cdot\}$ is the cumulative distribution function of $N(0, 1)$ and σ^* is the standard error of $\hat{\theta}^*$.

4.3. Confidence Intervals

Since the distribution of $\hat{\theta} = R(\hat{\theta}^*)$ might be slightly skewed (see example in Web Figure 6), the traditional way of using standard error to calculate confidence intervals (CI) could be misleading. An alternative way is to construct CI by the highest density interval. To do this, we generate B bootstrap samples $\{\hat{\theta}^{*1}, \ldots, \hat{\theta}^{*B}\}$ from the multivariate normal distribution $N\{\hat{\theta}^*, \text{Cov}(\hat{\theta}^*)\}$, where $\text{Cov}(\hat{\theta}^*)$ is the covariance matrix of $\hat{\theta}^*$. The sample of the kth parameter $\hat{\theta}_k$, $(\hat{\theta}_k^1, \ldots, \hat{\theta}_k^B)$ is then obtained from $\hat{\theta}^b = R(\hat{\theta}^{*b})$ for $b = 1, \ldots, B$. The highest density interval of $\hat{\theta}_k$ can be computed from the sample $(\hat{\theta}_k^1, \ldots, \hat{\theta}_k^B)$ by using the `emp.hpd` function in the `R` package `TeachingDemos`.

Table 1
Simulated bias of point estimates and coverage of 95% confidence intervals for the hazard ratio associated with treatment under two methods of sensitivity analysis, when censoring is moderate. The equations (14) *and* (16) *were used to estimate θ and its confidence interval in the last two columns*

$\widetilde{n}$	τ	β	ρ_0	ρ_1	Fraction Censored (%)	unadjusted		Lin et al. (1998)		$\hat{\theta} = R(\hat{\theta}^*)$	
						Bias	Coverage (%)	Bias	Coverage (%)	Bias	Coverage (%)
100	0.56	1	0.1	0.9	50	0.78	39	0.01	97	−0.04	95
	0.57		0.3	0.7	51	0.35	83	−0.02	91	0.00	98
	0.58		0.5	0.5	51	−0.08	90	−0.08	90	−0.03	99
	0.35	2	0.1	0.9	50	1.38	3	−0.04	96	0.07	96
	0.36		0.3	0.7	50	0.42	78	−0.21	87	−0.06	97
	0.35		0.5	0.5	50	−0.24	82	−0.24	82	0.01	100
	0.20	3	0.1	0.9	50	1.71	0	−0.12	91	0.15	91
	0.22		0.3	0.7	49	0.36	84	−0.40	72	−0.11	99
	0.21		0.5	0.5	50	−0.44	68	−0.44	68	0.00	100
500	0.57	1	0.1	0.9	50	0.76	0	−0.02	95	0.00	95
	0.58		0.3	0.7	50	0.32	42	−0.06	90	−0.01	98
	0.57		0.5	0.5	50	−0.10	90	−0.10	90	−0.01	99
	0.34	2	0.1	0.9	50	1.27	0	−0.15	82	−0.04	92
	0.34		0.3	0.7	50	0.43	11	−0.20	70	0.03	99
	0.34		0.5	0.5	50	−0.30	44	−0.30	44	0.01	100
	0.20	3	0.1	0.9	50	1.65	0	−0.18	81	0.04	89
	0.20		0.3	0.7	50	0.38	21	−0.38	28	−0.02	100
	0.21		0.5	0.5	50	−0.48	4	−0.48	4	0.01	100
1000	0.57	1	0.1	0.9	50	0.73	0	−0.04	93	−0.02	96
	0.57		0.3	0.7	50	0.30	11	−0.07	89	0.01	99
	0.58		0.5	0.5	50	−0.10	80	−0.10	80	−0.01	100
	0.34	2	0.1	0.9	50	1.29	0	−0.13	80	0.00	94
	0.34		0.3	0.7	50	0.40	1	−0.23	41	0.00	99
	0.34		0.5	0.5	50	−0.30	7	−0.30	7	−0.01	100
	0.20	3	0.1	0.9	50	1.65	0	−0.18	65	0.03	90
	0.20		0.3	0.7	50	0.40	0	−0.36	6	0.00	99
	0.21		0.5	0.5	50	−0.49	0	−0.49	0	−0.01	100

However, the bootstrap method may become computationally inefficient, when the dimension of θ is high (e.g., $K \geq 7$). We thus give an approximation by using the confidence bounds of $\hat{\theta}^*$. Suppose we are interested in the parameter θ_k and its confidence interval $(C_L^{\theta_k}, C_U^{\theta_k})$. As shown in Section 3.7, the effect of additional measured covariates is negligible. It means that the solution of θ_k would not change appreciably if we ignore all the covariates except x_k in (14).

In addition, θ_k is usually a monotonically increasing function of θ_k^* in practice. Let $(C_L^{\theta_k^*}, C_U^{\theta_k^*})$ be the confidence interval of θ_k^*. The lower bound $C_L^{\theta_k}$ then can be estimated from the equation

$$0 = \frac{1}{n} \sum_{i=1}^{n}$$

$$\times \left[x_{ik} - \frac{\sum_{j=1}^{\tilde{n}} S^+(t_i|x_j) e^{x_{jk} C_L^{\theta_k^*}} x_{jk} E_{c|x_{jk}} \left\{ e^{-H_0(t_i) e^{x_{jk} C_L^{\theta_k}+C\beta}} \right\}}{\sum_{j=1}^{\tilde{n}} S^+(t_i|x_j) e^{x_{jk} C_L^{\theta_k^*}} E_{c|x_{jk}} \left\{ e^{-H_0(t_i) e^{x_{jk} C_L^{\theta_k}+C\beta}} \right\}} \right].$$

(16)

Similarly, $C_U^{\theta_k}$ can be obtained from the above equation by substituting $C_L^{\theta_k^*}$ by $C_U^{\theta_k^*}$. Our simulation shows that this approximation is sufficiently accurate and very efficient.

4.4. *Simulation Study*

Lin et al. (1998) proposed a simple method for sensitivity analysis. Here we conducted simulation studies to compare their method with our approach.

Table 1 shows the biases of point estimators and coverage of 95% CIs in 1,000 simulation replications, when given the true β and $f(c|x)$. To compare with the method of Lin et al. (1998), we used similar simulation settings to theirs: $\theta = 1$, $h_0(t) = 1$, $X \sim B(1, 0.5)$, $C|x \sim B(1, \rho_x)$ and τ was solved from (13) so as to ensure moderate levels of censorship (fraction censored was about 50%). It is clear that our proposed method gives almost unbiased point estimates and good coverage of confidence intervals. The method of Lin et al. (1998) gets worse as β increases, because it only addresses the bias attributable to confounding. The results for light (10%) and heavy (90%) censorships are presented in Web Tables 1 and 2, respectively. We note that both methods of sensitivity analysis gave biased treatment estimates when censoring was heavy and the sample size was small ($\tilde{n} = 100$). However, since the accuracy of

approximation (14) increases with the number of observed events, the proposed method is asymptotically unbiased irrespective of the censoring rate. The minimum sample size at which the method achieves approximately unbiased estimates increases with the censoring rate, and for a censoring rate as high as 90% is about $\tilde{n} = 1,000$.

5. Real Examples for Sensitivity Analysis

5.1. *Vitamin and Minerals Trial*

Ellis et al. (2008) conducted a randomized controlled trial assessing the effect of antioxidant and folinic acid supplementation on developmental outcomes for children with Down syndrome. Comparing infants allocated to folinic acid ($x = 0$) with those who were not ($x = 1$), the estimated hazard ratio for age of sitting was 1.25 (95% confidence interval 0.88–1.78). These results did not change appreciably after adjustment for area of residence, maternal ethnicity, birth weight, and social class.
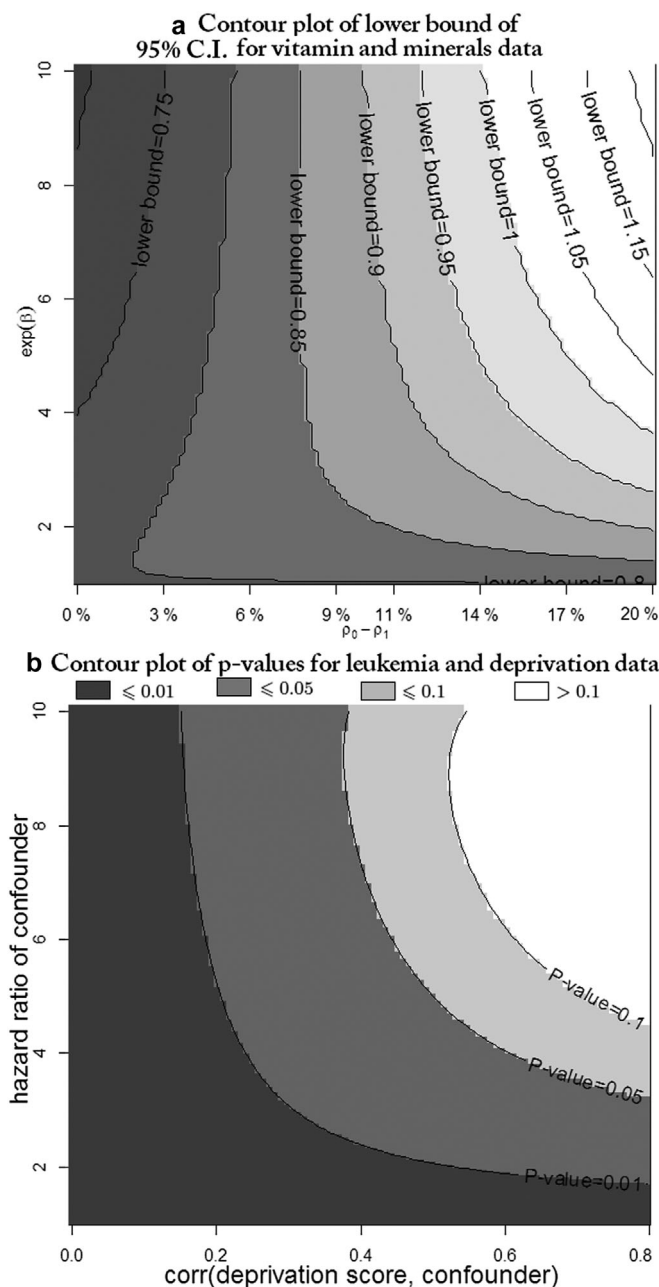
We now assess the impact on the treatment estimates for age at sitting of assuming a binary confounder, c, has been omitted from the model, where $C|x \sim B(1, \rho_x)$. As this is a randomized controlled trial and any random imbalance in the prevalence of the unmeasured confounder between treatment groups is likely to be small, we restrict $0 \leq \rho_0 - \rho_1 \leq 0.2$. Assuming the true prevalence of the omitted covariate for treatment groups combined is 0.5, the probability of a confounding effect $\rho_0 - \rho_1$ of more than 0.2 by chance is 0.02 for the trial sample of size $\tilde{n} = 138$.

Figure 4a shows the sensitivity of the lower limit of the confidence interval for the hazard ratio of folinic acid to adjustment for an unmeasured binary covariate of specified properties, where we set $\rho_0 + \rho_1 = 1$. For $\exp(\beta) = 10$, the difference in probabilities $\rho_0 - \rho_1$ must be > 0.14 for the treatment effect to become significant. The same conclusion can be obtained from the contour plot in Web Figure 3 which shows results of a similar sensitivity analysis for the P-value of the treatment estimate. The results for antioxidant supplementation in Web Table 3 show that the treatment effect is significant only when $\exp(\beta) = 10$ and $\rho_0 = 0.9$, $\rho_1 = 0.7$. Given the nature of the study design, the conditions required for the treatment effects to be significant are implausible, suggesting that the original findings of non-significance are robust to the presence of realistic levels of unmeasured confounding.

A simulation study was conducted with similar sample size and censoring rates to the vitamin and mineral trial, providing support for the validity of the treatment estimates presented in the sensitivity analysis (see Web Table 4). However, we note that in this illustrative application, the width of the confidence intervals suggests the sensitivity analysis, in common with the original analysis, lacks power to establish statistical significance for small studies.

5.2. *Leukaemia and Deprivation Study (Non-Randomized)*

Henderson, Shimakura, and Gorst (2002) analyzed the effect of a social deprivation score X (where lower values indicate less affluent areas) on the time in years since diagnosis with acute myeloid leukemia to death ($\tilde{n} = 1,043$). The estimated hazard ratio for a 1 point increase in x

a Contour plot of lower bound of 95% C.I. for vitamin and minerals data

b Contour plot of p-values for leukemia and deprivation data

Figure 4. Contour plots of sensitivity analysis results: (a) the lower bounds of the 95% confidence intervals (use (16)) for the hazard ratio of folinic acid on age of sitting for children with Down syndrome; (b)the P-values (use (15)) for the two-sided test that the log-hazard ratio of deprivation score $\theta = 0$.

was 1.03 (P-value $= 0.0012$) after adjustment for age, gender and white blood cell count, indicating that prognosis is less good if the patient lives in a more deprived residential location.

We now consider a potential unmeasured binary confounder C, which affects both survival time T and the deprivation score X. We generated c from $C|x \sim B\left\{1, \text{expit}(r_1 + r_2 x)\right\}$,

where (r_1, r_2) were solved from

$$
\begin{cases}
E(C) = E\left\{\operatorname{expit}(r_1 + r_2 X)\right\} = 0.5 \\
corr(X, C) = \dfrac{E(XC) - E(X)E(C)}{sd(X)sd(C)} \\
\qquad = \dfrac{E\left\{X\operatorname{expit}(r_1 + r_2 X)\right\} - 0.5E(X)}{0.5sd(X)}
\end{cases}
$$

such that the marginal distribution is $C \sim B(1, 0.5)$ and the desired $corr(X, C)$ is obtained.

Figure 4b shows the sensitivity of *P*-value for different choice of β and $corr(X, C)$. It shows that even if the correlation is strong, that is $corr(X, C) > 0.5$, the hazard ratio of the confounder needs to be > 4 for the hazard ratio of *x* to become non-significant at the 5% level. It seems unlikely that such an important covariate would be missed, suggesting that the original finding of a significant effect of deprivation score is robust to the presence of realistic levels of unmeasured confounding.

A simulation study was conducted with the same sample size ($\tilde{n} = 1043$), covariate *X* and censoring rate (15.7%) as this non-randomized study. To extend the range of scenarios considered, survival times were simulated assuming the true value of θ was 0 (i.e., assuming the continuous exposure has no effect on survival). Here the emphasis was on comparing the extent to which the sensitivity analysis methods avoid false rejection of the null hypothesis $H_0 : \theta = 0$. The results are summarized in Web Table 5 and provide further support for the validity of the proposed formulae when applied to data from non-randomized designs.

6. Discussion

We explored a general framework for assessing bias in treatment estimates from the Cox model with omitted covariates. Bias formulae based on asymptotic properties of the likelihood estimator were presented and validated in simulation experiments. The results showed that the confounding biases for censored survival data are typically complicated. However, the proposed approach made it possible to describe the influence of three different sources of bias: omission of a balanced covariate, data censoring and unmeasured confounding. Figure 5 characterises the sources of bias:

(a) In the absence of a missing covariate, the bias curve remains at zero (the solid line); when a balanced covariate is omitted, the effect is underestimated to a limit as $|\beta|$ increases (the dashed line).

(b) When the data are censored, the bias is maximized at 50% censoring but decreases with heavy censorship.

(c) When the missing covariate is a confounder, the shape of bias changes. If the association between *x* and *c* is positive, the limits increase on the right side but decrease on the left side, and hence the slope of bias increases. Conversely, if the association between *x* and *c* is negative, the limits decrease on the right side but increase on the left side.

Although the bias formula is applicable under a range of assumptions, this paper has focused on considering the sim-
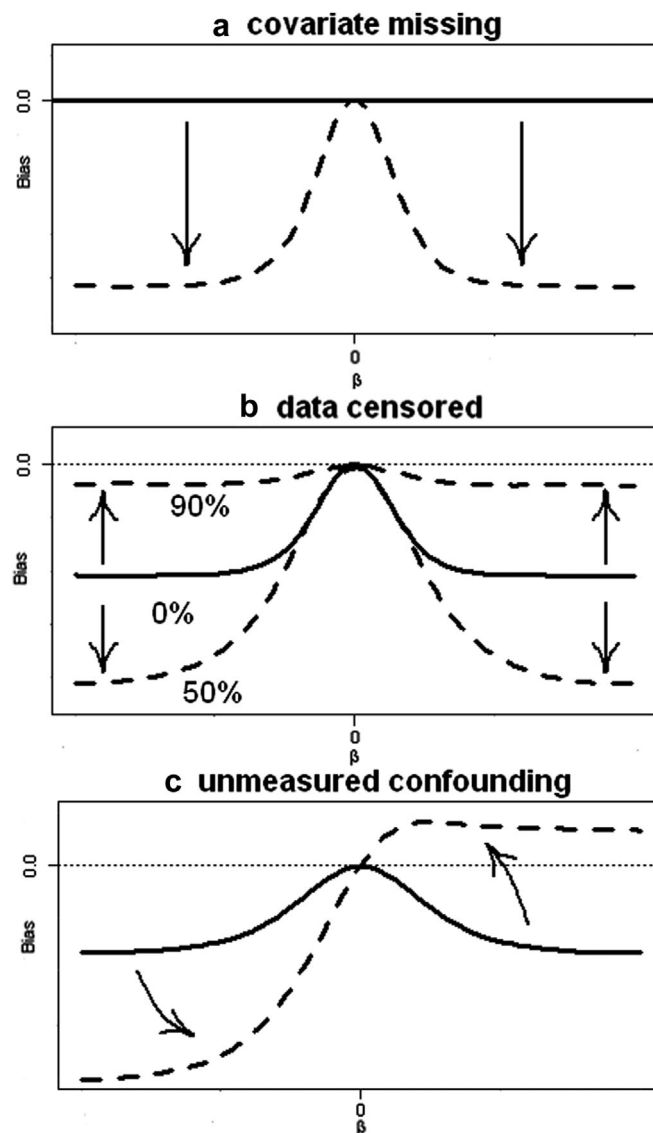
Figure 5. An illustration of the influence of the different sources of bias when estimating binary treatment effects from the Cox proportional hazards model with an omitted binary covariate. (a) solid: no missing data, no bias; dashed: bias due to omitting a balanced covariate. (b) solid: bias due to omitting a balanced covariate; dashed: bias due to omitting a balanced covariate and censoring. (c) solid: bias due to omitting a balanced covariate and censoring; dashed: bias due to omitting a confounder and censoring.

ple case of a binary exposure and a single unmeasured confounder. Further simulation work showed that the bias increased slightly in the presence of one or more measured confounders for large values of $|\beta|$. The extension to multiple unmeasured confounders is straightforward. If there are several missing covariates $c_1, \ldots, c_q$ with coefficients $\beta_1, \ldots, \beta_q$, then we can interpret *c* as the composite score, $\beta_1 c_1 +, \cdots, +\beta_q c_q$, with $\beta = 1$ (Lin et al., 1998). Lin et al. (1998) also argue that the choice of a single unmeasured confounder is a less severe

restriction when all the known confounders are adjusted for in the survival model.

The bias formula was used as the basis for proposing a new method to assess the sensitivity of estimates of treatment effects to omission of relevant covariates. Simulation experiments were conducted to compare the method with the approach of Lin et al. (1998), a special case of the proposed method when the rate of censoring is high. The method of Lin et al. (1998) has the benefit of ease of implementation, being based on a simple adjustment formula, but its relative performance deteriorates as the magnitude of $|\beta|$ increases. In contrast, the simulations indicate that the proposed method can provide sufficiently unbiased treatment estimates, and associated confidence intervals with good coverage, over a wide range of scenarios, when the true sensitivity parameters β and $f(c|x)$ are known.

Sensitivity analysis is a flexible approach to addressing omission of covariates that makes it possible to assess the impact of 'clinically plausible' levels of unmeasured confounding and other sources of bias on the treatment estimates (Groenwold, 2010). However, it does not provide a single precise estimate of treatment effectiveness nor does it help identify the nature of any bias from omitting covariates. A number of alternative strategies for tackling unmeasured confounding have been proposed that do attempt to provide explicit estimates of causal effects. An overview of these different methods was given in Aleyamehu et al (2011), including instrumental variables and the prior event rate ratio method (Tannen et al., 2009).

The method of sensitivity analysis proposed in this paper could be extended in a number of ways. First, incorporating adjustment for the propensity score into the sensitivity analysis would provide an efficient way of controlling for the effect of measured covariates (Rosenbaum, 1991). Other possible developments include consideration of specific distributional forms (both univariate and multivariate) for the unmeasured confounder(s) to provide special cases of the generic bias formulae for a wider range of common confounding models.

Omission of relevant covariates is a common source of bias when estimating treatment or exposure effects from survival data. Although we cannot directly adjust for unmeasured covariates, their potential impact can be assessed by means of sensitivity analyses. Indeed, Groenwold et al. (2010) argue that all analyses of causal associations in observational data should include an assessment of robustness to unmeasured confounding. The current study provides new tools for conducting sensitivity analysis for survival outcomes, with applicability to both randomized controlled trials and observational studies. Implementation of the methods requires numerical evaluation of the appropriate bias formulae. This can be achieved using Monte Carlo methods and illustrative R code is available on request from the authors.

7. Supplementary Materials

Web appendices, tables and figures referenced in Sections 2, 2.2, 3.1, 3.2, 3.5, 4.3, 4.4, 5.1 and 5.2 are available with this paper at the Biometrics website on Wiley Online Library.

References

Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal* **312**, 1215–1218.

Breslow, N. E. (1972). Discussion of the paper by D. R. Cox. *J R Statist, Soc B* **34**, 216–217.

Bretagnolle, J. and Huber-Carol, C. (1985). Sous-estimation des contrastes due a l'oubli de variables pertinentes dans le modele de Cox pour des durees de survie avec censure. *C.R.A.S.* **300**, 359–363.

Chastang, C., Byar, D., and Piantadosi, S. (1988). A quantitative study of the bias in estimating the treatment effect caused by omitting a balanced covariate in survival models. *Statistics in Medicine* **7**, 1243–1255.

Ellis, J. M., Tan, H. K., Gilbert, R. E., Muller, D. P., Henley, W. E., Moy, R., Pumphrey, R., Ani, C., Davies, S., Edwards, V., Green, H., Salt, A. and Logan, S. (2008). Supplementation with antioxidants and folinic acid for children with Down syndrome: Randomised controlled trial. *British Medical Journal* **336**, 594–597.

Gail, M., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with non-linear regressions and omitted covariates. *Biometrika* **71**, 431– 444.

Ghani, A. C., Henley, W. E., Donnelly, C. A., Mayer, S., and Anderson, R. M. (2001). Comparison of the effectiveness of non-nucleoside reverse transcriptase inhibitor-containing and protease inhibitor-containing regimens using observational databases. *AIDS* **15**, 1133–1142.

Groenwold, R. H., Nelson, D. B., Nichol, K. L., Hoes, A. W., and Hak, E. (2010). Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *International Journal of Epidemiology* **39**, 107–117.

Henderson, R., Shimakura, S., and Gorst, D. (2002). Modelling spatial variation in leukaemia survival data. *Journal of the American Statistical Association* **97**, 965–972.

Hernandez, A. V., Eijkemans, M. J. and Steyerberg, E. W. (2006). Randomized controlled trials with time-to-event outcomes: How much does prespecified covariate adjustment increase power? *Annals of Epidemiology* **16**, 41–48.

Klungel, O. H., Martens, E. P., Psaty, B. M., Grobbee, D. E., Sullivan, S. D., Stricker, B. H., Leufkens, H. G., and de Boer, A. (2004). Methods to assess intended effects of drug treatment in observational studies are reviewed. *Journal of Clinical Epidemiology* **57**, 1223–1231.

Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* **54**, 948–963.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistics* **84**, 207–224.

Rosenbaum, P. R. (1991). Discussing Hidden bias in observational studies. *Annals of Internal Medicine* **115**, 901–905.

Satten, G. A. and Datta, S. (2001). The Kaplan–Meier Estimator as an inverse-probability-of-censoring weighted average. *American Statistician* **55**, 207–210.

Tannen, R. L., Weiner, M. G. and Xie, D. (2009). Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *British Medical Journal* 338, b81.

VanderWeele, T. J. (2008). Sensitivity analysis: distributional assumptions and confounding assumptions. *Biometrics* **64**, 645–649.