Lifetime Data Anal (2013) 19:279–296
DOI 10.1007/s10985-013-9242-z

On collapsibility and confounding bias in Cox and Aalen regression models

Torben Martinussen · Stijn Vansteelandt

Received: 2 July 2012 / Accepted: 3 January 2013 / Published online: 18 January 2013

Abstract We study the situation where it is of interest to estimate the effect of an exposure variable X on a survival time response T in the presence of confounding by measured variables Z. Quantifying the amount of confounding is complicated by the non-collapsibility or non-linearity of typical effect measures in survival analysis: survival analyses with or without adjustment for Z typically infer different effect estimands of a different magnitude, even when Z is not associated with the exposure, and henceforth not a confounder of the association between exposure and survival time. We show that, interestingly, the exposure coefficient indexing the Aalen additive hazards model is not subject to such non-collapsibility, unlike the corresponding coefficient indexing the Cox model, so that simple measures of the amount of confounding bias are obtainable for the Aalen hazards model, but not for the Cox model. We argue that various other desirable properties can be ascribed to the Aalen model as a result of this collapsibility. This work generalizes recent work by Janes et al. (Biostatistics 11:572–582, 2010).

T. Martinussen (✉)
Department of Biostatistics, University of Copenhagen, Øster Farimagsgade 5B,
1014 Copenhagen K, Denmark
e-mail: tma@sund.ku.dk

S. Vansteelandt
Department Applied Mathematics and Computer Science, Ghent University,
Krijgslaan 281, S9, 9000 Gent, Belgium

S. Vansteelandt
Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine,
Keppel Street, London WC1E 7HT, UK
e-mail: stijn.vansteelandt@ugent.be

Keywords Aalen's additive model · Causal effect · Cox model · Collapsibility · Confounding

1 Introduction

Epidemiologists have a longstanding interest in quantifying the magnitude of bias in exposure-disease associations, which is induced by the omission of a specific confounding variable Z (see e.g., Miettinen 1972). Understanding the extent of such bias is important to better enable the evaluation of other studies on exposure-disease associations. For instance, to gauge the quality of the results when control for relevant factors was not made. Such knowledge is also useful in the planning of studies. For instance, financial and logistic constraints, as well as quality considerations, often restrain the number of potential confounding variables one can collect per individual. To boost the precision of the exposure-disease association, one may wish to prioritize data collection on strong outcome predictors. We instead recommend to prioritize data collection on important confounding variables in line with our viewpoint that the concern for bias surpasses efficiency concerns. Routine assessment of the amount of confounding has therefore been recommended (Miettinen 1972).

Quantifying the amount of confounding is a delicate problem, which forms the source of much confusion in the statistical and epidemiological literatures. The reason is that the standard approach to control for confounders works by including them in a regression analysis of outcome on exposure, but that the resulting adjusted association may differ from the unadjusted association even in the absence of confounding bias; likewise, confounding bias may be present even when the adjusted and unadjusted associations are identical (Miettinen and Cook 1981; Greenland and Robins 1986; Greenland et al. 1999). That the difference between certain adjusted and unadjusted association measures cannot be equated with confounding bias, is due to their non-linearity (Janes et al. 2010), which is more commonly referred to as non-collapsibility (Greenland et al. 1999).

Understanding whether association measures are collapsible (i.e., that the adjusted and unadjusted association measures are the same when control is made for factors Z that are not associated with the exposure, and hence do not confound the exposure-outcome association) is of interest in its own right, for various reasons. First, the adjusted association between exposure and outcome (controlling for a sufficient set of confounders) measures a subgroup-specific effect, which may differ from the corresponding population-averaged effect when the association measure is not collapsible. When, as often, the interest lies in the potential public health impact of the exposure, primarily the latter effect is of interest. Second, different empirical studies typically control for different covariate sets. When sufficient control for confounders is made in all these studies (in the sense that each of these covariate sets includes a sufficient set to control for confounding), then collapsible association measures will, by definition, be better comparable across studies (Tsiatis et al. 2000). Disregarding other sources of bias that may occur (e.g., due to model misspecification), heterogeneity between studies may then be interpreted as the result of varying effect sizes across populations. Similar conclusions cannot always be made for non-collapsible association measures.

Importantly, note that the use of collapsible association measures is also of interest in the absence of confounding bias. For instance, effect measures obtained from different randomized clinical trials may differ across studies when, as usual, different studies consider somewhat different patient populations or when they control for different sets of baseline covariates to increase power. This may again happen as a result of non-collapsibility, even in the absence of effect modification. In summary, collapsible effect measures tend to be better transportable to other study populations.

Greenland et al. (1999) defined the amount of confounding bias to be the difference between the *population-averaged* exposure effect (adjusted for confounding) and the raw association between exposure and outcome. Because both association measures refer to the overall study population, this measure of confounding bias is not subject to non-collapsibility. Janes et al. (2010) promoted this idea and used G-computation (Robins 1986)—a form of direct standardization (Vansteelandt and Keiding 2011)—to calculate the population-averaged exposure effect, adjusting for confounding, in the context of a binary response Y which lends itself to logistic regression. In this article, we study the same problem but with a (possibly right-censored) survival time response T instead of a binary response. We focus in particular on two popular models for a survival time response, namely the Cox model (Cox 1972) and Aalen's additive hazards model (Aalen 1980, 1989, and also Aalen et al. 2008). Our key result is that there is no non-linearity or non-collapsibility of the parameters indexing the Aalen model, unlike those indexing the Cox model. We argue that various desirable properties can be ascribed to this collapsibility, including the existence of simple measures of confounding bias under the Aalen model. In contrast, the non-collapsibility of the Cox model implies that the simple *change in estimate* (Rothman et al. 2008) approach is problematic for this model. We therefore show how to properly assess confounding bias in the Cox model, which to the best of our knowledge has never been dealt with in a rigorous way for this model. This is illustrated in the worked applications of Sects. 5.2 and 5.3.

2 Confounding bias

We let T denote the survival time of interest and let X be the exposure variable, which we assume for now to be binary. We return to the continuous exposure case in Sect. 7. The hazard function of T given X and a third variable Z (that might be a vector) is denoted by $\lambda(t|X = x, Z = z)$. Later we will specialize to the Cox model and the Aalen additive hazards model, but we keep the notation general for now. We let g be an arbitrary association measure (e.g., $g(a, b) = log(a/b)$ for the log relative risk). The marginal association is defined as

$$\mathcal{T}_m(t) = g(\lambda(t|x = 1), \lambda(t|x = 0)),$$

while the marginal exposure effect (causal effect) is defined as

$$\mathcal{T}(t) = g(\lambda(t|\hat{x} = 1), \lambda(t|\hat{x} = 0)),$$

where $\hat{x} = x$ is short for $do(X = x)$; that is, the do-operator of Pearl, see Pearl (2000, p. 70), and Sect. 3 on how to calculate distributions under the $do(X = x)$-operation. In particular, $\lambda(t|\hat{x} = x)$ denotes the hazard function that would be realized at time t if X were uniformly set to x in the population.

We define

$$\Delta(t) = \mathcal{T}(t) - \mathcal{T}_m(t)$$

to be the amount of confounding bias. If $\Delta(t) \neq 0$, then we say that there is confounding because the marginal exposure effect differs from the marginal association.

Througout, we will assume that the covariate set Z is sufficient to adjust for confounding, so that $\lambda(t|\hat{x} = x, Z) = \lambda(t|X = x, Z)$. Given this assumption, it is common in routine practice to measure the amount of confounding of the association between X and T in terms of a comparison of their conditional association (given Z) and their marginal association. That is, $\Delta_{si}(t) = \mathcal{T}_c(t) - \mathcal{T}_m(t)$ (with "si" for simple), where

$$\mathcal{T}_c(t) = g(\lambda(t|x = 1, Z), \lambda(t|x = 0, Z))$$

encodes the conditional association. However, this interpretation may be misleading because

$$\Delta_{si}(t) = \Delta(t) + \Delta_{nl}(t),$$

where $\Delta_{nl}(t) = \mathcal{T}_c(t) - \mathcal{T}(t)$ (with "nl" denoting non-linear) may differ from zero even in the absence of confounding (even if X is randomly assigned). Indeed, $\Delta_{nl}(t) \neq 0$ reflects a difference between subgroup effects and population-averaged effects which is typical of nonlinear association measures and is commonly referred to as non-collapsibility (Greenland et al. 1999) or non-linearity (Janes et al. 2010).

3 Assessing the amount of confounding bias

To calculate the amount of confounding we make use of the G-computation formula (Robins 1986; Pearl 2000) for the distribution, $P(T|\hat{x} = x)$, that would have been observed under an intervention, setting the exposure to x. In our setting it reads

$$P(T|\hat{x} = x) = \int P(T|X = x, Z = z) dF_Z(z),$$

where F_Z denotes the marginal distribution function of Z. The corresponding survival function is given by

$$P(T > t|\hat{x} = x) = \int e^{-\Lambda(t|x,z)} dF_Z(z),$$

where $\Lambda(t|x,z) = \int_0^t \lambda(s|x,z)\,ds$ is the cumulative conditional hazard function. We call this the causal survival function. Similarly, one can obtain an expression for the causal density function, and then finally for the causal hazard function:

$$\lambda(t|\hat{x}=x) = \frac{\int e^{-\Lambda(t|x,z)}\lambda(t|x,z)dF_Z(z)}{\int e^{-\Lambda(t|x,z)}dF_Z(z)}. \tag{1}$$

We will now study the above defined quantities under the Aalen additive hazards model and under Cox's proportional hazards model where we take $g(a,b) = a - b$ and $g(a,b) = \log(a/b)$, respectively.

We allow for right-censoring, letting C denote the censoring time, and assume that T and C are conditionally independent given both (X,Z) and X, respectively. We observe the first time either failure or censoring occurs, $U = \min(T,C)$, and an indicator of whether it is censoring or failure that occurs, $\delta = I(T \leq C)$. The data consist of n independent replicates of $\{U_i, \delta_i, X_i, Z_i\}$, $i = 1,\ldots,n$, and we consider the time interval $[0,\tau]$, where $\tau < \infty$ is the endpoint of the study.

We now look at the suggested models separately.

3.1 The Aalen additive hazards model

Suppose that the Aalen additive hazards model

$$\lambda(t|x,z) = \beta_0(t) + \beta_X(t)x + \beta_Z(t)z, \tag{2}$$

holds, where $\beta(t) = \{\beta_0(t), \beta_X(t), \beta_Z(t)\}^T$ is a locally integrable function. Let $B_0(t) = \int_0^t \beta_0(s)\,ds$ and similarly with $B_X(t)$ and $B_Z(t)$. The marginal (unadjusted) hazard function under this model is given by

$$\lambda(t|x) = \beta_0(t) + \beta_X(t)x + \beta_Z(t)\frac{E(e^{-B_Z(t)Z}Z|X=x)}{E(e^{-B_Z(t)Z}|X=x)}, \tag{3}$$

see Aalen (1989) for similar calculations. In contrast, the causal hazard function becomes

$$\lambda(t|\hat{x}=x) = \tilde{\beta}_0(t) + \beta_X(t)x,$$

where

$$\tilde{\beta}_0(t) = \beta_0(t) + \beta_Z(t)\frac{E(e^{-B_Z(t)Z}Z)}{E(e^{-B_Z(t)Z})}.$$

It is thus seen that the additive hazards structure is preserved and that the marginal exposure effect is given by

$$\mathcal{T}(t) = g(\lambda(t|\hat{x}=1), \lambda(t|\hat{x}=0)) = \beta_X(t).$$

This is also the conditional exposure effect, demonstrating that non-linearity is not present for this model. It further follows that $\mathcal{T}(t) \neq \mathcal{T}_m(t)$ whenever $\beta_Z(t) \neq 0$, and Z and X are not independent.

To assess the amount of confounding bias, we may thus compare the cumulative exposure effect based on the conditional model, $B_X(t) = \int_0^t \beta_X(s)\,ds$, to the cumulative effect based on the marginal hazard function given by (3).

Let

$$l(x, B_Z(t)) = \frac{E(e^{-B_Z(t)Z}Z|X = x)}{E(e^{-B_Z(t)Z}|X = x)}. \tag{4}$$

Since X is binary we may write (3) as

$$\lambda(t|X = x) = \beta_0^m(t) + \beta_X^m(t)x, \tag{5}$$

where

$$\beta_0^m(t) = \beta_0(t) + l(0, B_Z(t)), \quad \beta_X^m(t) = \beta_X(t) + \beta_Z(t)\{l(1, B_Z(t)) - l(0, B_Z(t))\}.$$

The amount of confounding bias on the cumulative scale can then be estimated by

$$\hat{B}_X(t) - \hat{B}_X^m(t), \tag{6}$$

where $\hat{B}_X(t)$ and $\hat{B}_X^m(t)$ denote the usual Aalen least squares estimators from the conditional and marginal models, that is models (2) and (5).

3.2 The Cox model

If instead we assume the Cox proportional hazards model

$$\lambda(t|x, z) = \lambda_0(t)e^{\beta_X x + \beta_Z z},$$

where $\lambda_0(t)$ denotes the (locally integrable) baseline hazard function, then the marginal hazard function is given by $\lambda(t|x) = \lambda_0(t)e^{\beta_X x} j\{x, \theta(t)\}$, where

$$j\{x, \theta(t)\} = \frac{E[S_1\{x, \theta(t); Z\}|X = x]}{E[S_0\{x, \theta(t); Z\}|X = x]}$$

with

$$S_0\{x, \theta(t); z\} = e^{-\Lambda_0(t)\exp(\beta_X x + \beta_Z z)}, \; S_1\{x, \theta(t); z\} = S_0\{x, \theta(t); z\}e^{\beta_Z z},$$

and $\theta(t) = \{\beta_X, \beta_Z, \Lambda_0(t)\}^T$. The causal hazard function now becomes

$$\lambda(t|\hat{x} = x) = \lambda_0(t)e^{\beta_X x}h\{x, \theta(t)\},$$

where

$$h\{x,\theta(t)\} = \frac{E[S_1\{x,\theta(t);Z\}]}{E[S_0\{x,\theta(t);Z\}]}.$$

Hence, the marginal exposure effect is given by

$$\mathcal{T}(t) = \beta_X + \log\left[\frac{h\{1,\theta(t)\}}{h\{0,\theta(t)\}}\right]$$

so, in general, $\mathcal{T}(t) \neq \beta_X$ even when Z and X are independent, and therefore non-linearity is present. It again follows that confounding is present, $\mathcal{T}(t) \neq \mathcal{T}_m(t)$, whenever $\beta_z \neq 0$, and Z and X are not independent. Note also that, because of the aforementioned non-linearity, the Cox model may not hold for both the conditional and marginal relationship, which further emphasizes that naïve measures of the amount of confounding can be misleading.

We let

$$\hat{h}\{x,\theta(t)\} = \frac{\sum_{i=1}^n S_1\{x,\theta(t);Z_i\}}{\sum_{i=1}^n S_0\{x,\theta(t);Z_i\}},\ \hat{j}\{x,\theta(t)\} = \frac{\sum_{i=1}^n S_1\{x,\theta(t);Z_i\}I(X_i=x)}{\sum_{i=1}^n S_0\{x,\theta(t);Z_i\}I(X_i=x)}$$

denote the empirical estimates. It then follows that the amount of confounding bias can be estimated by

$$\hat{\Delta}(t) = \log\left(\frac{\hat{h}\{1,\hat{\theta}(t)\}}{\hat{h}\{0,\hat{\theta}(t)\}}\right) - \log\left(\frac{\hat{j}\{1,\hat{\theta}(t)\}}{\hat{j}\{0,\hat{\theta}(t)\}}\right), \tag{7}$$

where $\hat{\theta}(t) = \{\hat{\beta}_X, \hat{\beta}_Z, \hat{\Lambda}_0(t)\}^T$ with $(\hat{\beta}_X, \hat{\beta}_Z)$ being the Cox partial likelihood estimates and with $\hat{\Lambda}_0(t)$ the Breslow estimator. The causal effect of X is estimated by

$$\hat{\mathcal{T}}(t) = \hat{\beta}_X + \log\left[\frac{\hat{h}\{1,\hat{\theta}(t)\}}{\hat{h}\{0,\hat{\theta}(t)\}}\right]. \tag{8}$$

4 Large sample properties of the estimators

Large sample properties for the estimator given in (6) under the Aalen model are straightforward to derive as it is a difference between two Aalen least squares estimators. To be specific, we have that

$$n^{1/2}[\hat{B}_X(t) - \hat{B}_X^m(t) - \{B_X(t) - B_X^m(t)\}] = n^{-1/2}\sum_{i=1}^n \epsilon_i^B(t) + o_p(1),$$

where $\epsilon_i^B(t)$ are zero-mean iid terms. More details and the expression for $\epsilon_i^B(t)$ are given in the Appendix.

We now focus on the large sample properties of the amount of confounding bias (7) based on the Cox model. Define $W_n(t) = n^{1/2}\{\hat{\Delta}(t) - \Delta(t)\}$. We show in the Appendix that $W_n(t) = n^{-1/2}\sum_{i=1}^n \epsilon_i^{\Delta}(t) + o_p(1)$, where $\{\epsilon_i^{\Delta}(t)\}_{i=1,\ldots,n}$ are zero-mean iid processes. The expression for $\epsilon_i^{\Delta}(t)$ is given in the Appendix. Hence, the process $W_n(t)$ converges in distribution to a Gaussian process with variance $\Sigma(t)$ that can be estimated consistently by $\hat{\Sigma}(t) = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{\Delta}(t)^2$, where $\hat{\epsilon}_i^{\Delta}(t)$ is obtained by plugging in empirical quantities for the unknowns in $\epsilon_i^{\Delta}(t)$. Finally, we give the large sample results for the estimator of the causal effect, $\hat{\mathcal{T}}(t)$, under the Cox model. In the Appendix, it is shown that

$$n^{1/2}\left\{\hat{\mathcal{T}}(t) - \mathcal{T}(t)\right\} = n^{-1/2}\sum_{i=1}^{n} \epsilon_i^{\mathcal{T}}(t) + o_p(1),$$

where $\{\epsilon_i^{\mathcal{T}}(t)\}_{i=1,\ldots,n}$ are zero-mean iid processes. The expression for $\epsilon_i^{\mathcal{T}}(t)$ is given in the Appendix. The variance of the limit process can thus be consistently estimated by $n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{\mathcal{T}}(t)^2$, where $\hat{\epsilon}_i^{\mathcal{T}}(t)$ is obtained by plugging in empirical quantities for the unknowns in $\epsilon_i^{\mathcal{T}}(t)$. It now also follows that $n^{1/2}\{\hat{\Delta}_{nl}(t) - \Delta_{nl}(t)\} = n^{-1/2}\sum_{i=1}^n \epsilon_i^{\Delta_{nl}}(t) + o_p(1)$ where $\{\epsilon_i^{\Delta_{nl}}(t)\}_{i=1,\ldots,n}$ are zero-mean iid processes. The expression for $\epsilon_i^{\Delta_{nl}}(t)$ is given in the Appendix. Hence we can also easily estimate the variance of $\hat{\Delta}_{nl}(t)$.

5 Numerical results

5.1 Simulation study

We did a small simulation study to study the impact of the size of $\lambda_0(t)$ on the amount of non-linearity for the Cox model. The confounder variable, Z, was taken to be normally distributed around zero with standard deviation 0.5, and X was binary with $\text{logit}\{P(X = 1|Z = z)\} = 0.3z$. The survival times were generated according to the Cox proportional hazards model, $\lambda(t|x, z) = \lambda_0(t)e^{\beta_X x + \beta_Z z}$, with $\beta_X = \beta_Z = \log(2)$. The baseline hazard function $\lambda_0(t)$ was set to 0.1 and 0.1/8 and censoring was so that 20 and 75 % were censored under the two different values for $\lambda_0(t)$. Sample size was set to 5,000 and one run was made for the two different scenarios. Figure 1 displays the true $\Delta(t)$ (full line), $\hat{\Delta}(t)$ (dashed line) and $\hat{\Delta}_{si}(t)$ (dotted line) for the two different values of $\lambda_0(t) = 0.1, 0.1/8$ (left and right display, respectively). It is clear that the estimate $\hat{\Delta}(t)$ is close to $\Delta(t)$ and also that the degree of non-linearity (difference between $\hat{\Delta}_{si}(t)$ and $\hat{\Delta}(t)$) is noticeable for the scenario with $\lambda_0(t) = 0.1$ while it diminishes when lowering $\lambda_0(t)$ to 0.1/8.

Simulations were also run to investigate the performance of the variance estimators in terms of 95 %-coverage probabilities. They were found to be satisfactory (see Table 1). For instance, using the above simulation setting with $\lambda_0(t)$ set to 0.1/8, and with the sample size set to 300, the 95 %-coverage probabilities computed for the estimator $\hat{\mathcal{T}}(t)$ at time points $t = 5$, 10 and 15 was, based on 1,000 simulation runs, calculated to 94.8, 94.7 and 95.1, respectively.

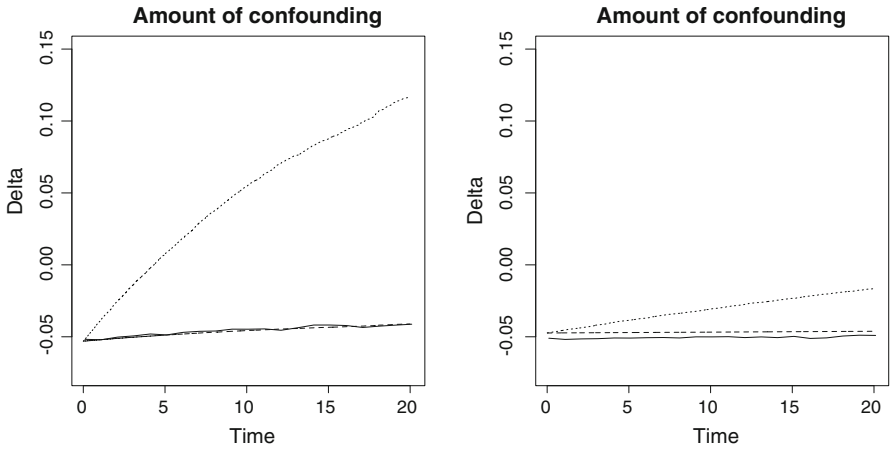

Fig. 1 True $\Delta(t)$ (*full line*), $\hat{\Delta}(t)$ (*dashed line*) and $\hat{\Delta}_{si}(t)$ (*dotted line*) for the values of $\lambda_0(t) = 0.1, 0.1/8$, *left* and *right* display, respectively

Table 1 Bias (Mean $\hat{\Delta}(t) - \Delta(t)$), average estimated standard error (SSE($\hat{\Delta}(t)$)), empirical standard error (SEE($\hat{\Delta}(t)$)) and coverage probability of 95 % pointwise confidence intervals (CP($\hat{\Delta}(t)$)), in function of sample size n and for two different settings of baseline mortality

$(n, \lambda_0(t))$		$t = 5$	$t = 10$	$t = 15$
(100,0.1/8)	Mean $\hat{\Delta}(t) - \Delta(t)$	0.001	−0.001	−0.002
	SSE($\hat{\Delta}(t)$)	0.087	0.083	0.081
	SEE($\hat{\Delta}(t)$)	0.092	0.087	0.086
	95 % CP($\hat{\Delta}(t)$)	0.959	0.965	0.964
(200,0.1/8)	Mean $\hat{\Delta}(t) - \Delta(t)$	−0.004	−0.006	−0.008
	SSE($\hat{\Delta}(t)$)	0.059	0.057	0.055
	SEE($\hat{\Delta}(t)$)	0.059	0.057	0.056
	95 % CP($\hat{\Delta}(t)$)	0.936	0.943	0.954
(100,0.1)	Mean $\hat{\Delta}(t) - \Delta(t)$	0.000	0.001	0.002
	SSE($\hat{\Delta}(t)$)	0.070	0.073	0.080
	SEE($\hat{\Delta}(t)$)	0.072	0.073	0.078
	95 % CP($\hat{\Delta}(t)$)	0.961	0.961	0.961
(200,0.1)	Mean $\hat{\Delta}(t) - \Delta(t)$	−0.001	−0.001	−0.001
	SSE($\hat{\Delta}(t)$)	0.059	0.058	0.057
	SEE($\hat{\Delta}(t)$)	0.058	0.059	0.065
	95 % CP($\hat{\Delta}(t)$)	0.937	0.940	0.951

5.2 Application to pneumonia data

In this subsection we consider data on time to hospitalized pneumonia in young children (Klein and Moeschberger 2003, p. 14). The interest centers on whether or not the

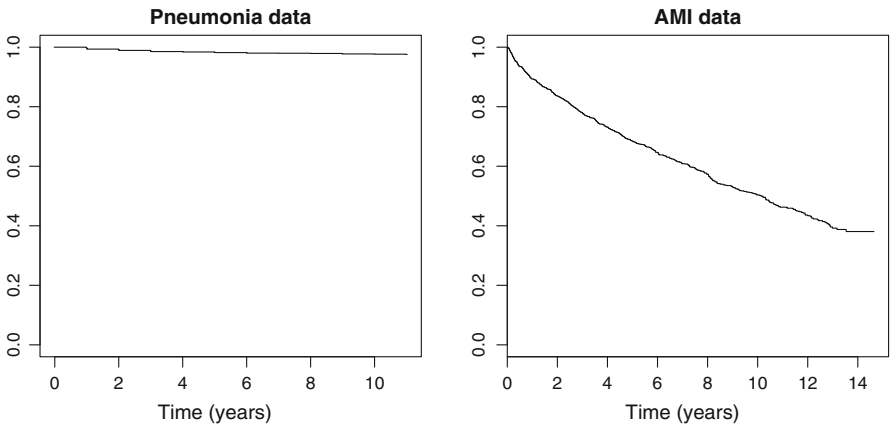

Fig. 2 Kaplan–Meier estimator for pneumonia data (*left*) and AMI data (*right*)

mother's feeding choice (breast feeding versus never breast fed) protected the infant against hospitalized pneumonia in the first year of life. There are data on 3,470 infants with a total of 73 events. We thus consider the effect of breast feeding: $X = 1$ if child was breast fed at birth; $X = 0$ if child was never breast fed. As there is an overrepresentation of smokers among those not giving breast feeding and since smoking during pregnancy is also a risk factor for pneumonia we use it as a confounding variable, Z, with $Z = 1$ (smoking yes) and $Z = 0$ (smoking no). For these data, the population-averaged causal effect estimate $\mathcal{T}(t)$ of breast feeding, and the estimate from the conditional Cox model are almost identical and hence the non-linearity problem for the Cox mode does not seem to be present for these data. This is in contrast to what was seen in the simulations, and is mainly due to hospitalized pneumonia being a rare phenomenon—the size of $\lambda_0(t)$ is estimated roughly to be around 0.0025 (see also Fig. 2, left display, where the Kaplan–Meier estimator is depicted).

5.3 Application to acute myocardial infarction data

In this subsection we consider survival of patients after an acute myocardial infarction event (AMI). The study was carried out at the University Clinical Centre in Ljubljana, where 1,040 patients were followed for up to 14 years. The end point was death from any cause, as gathering cause-specific death information had proved impossible to carry out. These data were also considered in Stare et al. (2005), and contain several variables recorded at the time of admission.

We will here concentrate on the effect of aspirin (1 = yes, 0 = no). There is missing information on aspirin for 20 patients. For some reason aspirin was more likely to be given to younger patients (see Table 2), and since age is a highly significant predictor of death, it is of interest to study the effect of aspirin adjusting for age as a confounder. As opposed to the pneumonia data, more events are seen here (Kaplan–Meier estimates are shown in Fig. 2), and therefore non-linearity might be an issue when applying the Cox model.

Table 2 AMI data

Cross-tabulation of aspirin and age

	Age (years) 24 – 53	54 – 61	62 – 70	71 – 95
Aspirin, no	53	62	81	110
Aspirin, yes	214	190	170	140

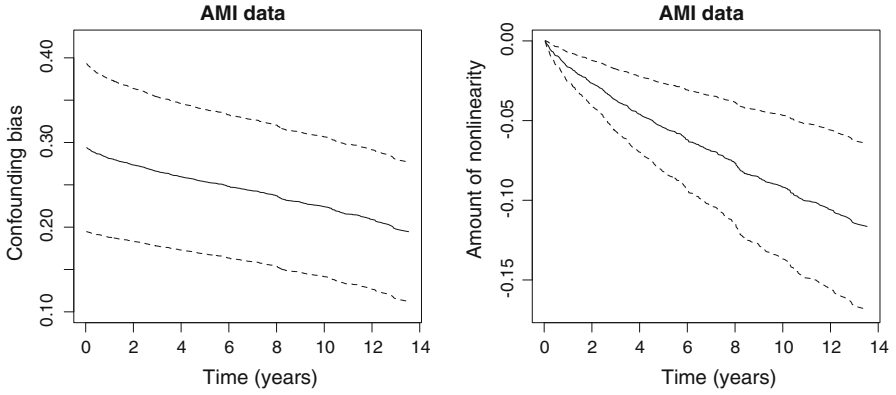

Fig. 3 AMI data. Effect of aspirin. *Left display* Estimate of confounding bias, $\hat{\Delta}(t)$ (*full line*) and 95 % confidence intervals (*hashed lines*). *Right display* Amount of non-linearity, $\hat{\mathcal{T}}_c - \hat{\mathcal{T}}(t)$ (*full line*), and 95 % confidence intervals (*hashed lines*)

Goodness-of-fit procedures (Martinussen and Scheike 2006) revealed that these data are adequately fitted by the Cox model (results not shown). As is seen from Fig. 3 (left display), confounding bias is indeed seen for these data as the difference between the marginal exposure effect of aspirin and the corresponding raw marginal effect is around 0.25. Hence it is reasonable to adjust for age when judging the effect of aspirin. As is seen from the right display of Fig. 3, there is a noticeable difference between the conditional effect estimated from the Cox model where we adjust for age, and the estimated population-averaged causal effect, $\hat{\mathcal{T}}(t)$.

6 Multiple confounders

Often, adjustment for several covariates is needed to control for confounding of the association between X and T. In such settings, the proposed measures continue to be useful to express how much confounding can be ascribed to all of these covariates jointly as Z may be multivariate. It may additionally be of interest to assess how much confounding can be ascribed to one of these covariates. We therefore propose a measure of confounding bias due to a scalar covariate Z under the assumption that (Z, W), where W is a covariate vector, is a set of covariates which are sufficient to control for confounding of the association between X and T. The amount of confounding due to Z may be defined as

$$\mathcal{T}_Z(t) - \mathcal{T}(t),$$

where $T_Z(t)$ denotes the marginal exposure effect that would have been obtained, had Z not been available. In other words, what is the change in causal effect estimate due to inclusion of Z. One thus needs to calculate $\mathcal{T}_Z(t)$ as $g(\lambda_Z(t|\hat{x}=1), \lambda_Z(t|\hat{x}=0))$, where $\lambda_Z(t|\hat{x}=x)$ is the hazard function corresponding to the distribution

$$P_Z(T|\hat{X}=x) = \int P(T|X=x, W=w) dF_W(w).$$

Below we deal with this problem assuming the Aalen additive hazards model

$$\lambda(t|x,z,w) = \beta_0(t) + \beta_X(t)x + \beta_W(t)w + \beta_Z(t)z.$$

Under this model it is easy to show that

$$\lambda_Z(t|\hat{x}=x) = \beta_0(t) + \beta_X(t)x + h(t;x),$$

where

$$h(t;x) = \frac{\int e^{-B_W(t)w - B_Z^*(t;B_Z,x,w)} \{\beta_W(t)w + \beta_Z(t) l(B_Z(t);x,w)\} dF_W(w)}{\int e^{-B_W(t)w - B_Z^*(t;B_Z,x,w)} dF_W(w)}$$

with

$$l(B_Z(t);x,w) = \frac{E(e^{-B_Z(t)Z} Z|X=x, W=w)}{E(e^{-B_Z(t)Z}|X=x, W=w)},$$
$$B_Z^*(t;B_Z,x,w) = \int_0^t l(B_Z(s);x,w) dB_Z(s).$$

Hence, $\mathcal{T}(t) = \beta_X(t)$ and $\mathcal{T}_Z(t) = \beta_X(t) + j(t)$, $j(t) = h(t;1) - h(t;0)$. The amount of confounding bias due to Z, on cumulated scale, is therefore $J(t) = H(t;1) - H(t;0)$, where

$$H(t;x) = \int_0^t h(s;x)\, ds.$$

An estimate of $H(t;x)$ is given by

$$\hat{H}(t;x) = \int_0^t \frac{\int e^{-\hat{B}_W(t)w - B_Z^*(t;\hat{B}_Z,x,w)} w d\hat{F}_W(w)}{\int e^{-\hat{B}_W(t)w - B_Z^*(t;\hat{B}_Z,x,w)} d\hat{F}_W(w)} d\hat{B}_W(s)$$
$$- \int_0^t \frac{\int e^{-\hat{B}_W(t)w - B_Z^*(t;\hat{B}_Z,x,w)} l(\hat{B}_Z(t);x,w) d\hat{F}_W(w)}{\int e^{-\hat{B}_W(t)w - B_Z^*(t;\hat{B}_Z,x,w)} d\hat{F}_W(w)} d\hat{B}_Z(s)$$

To be able to calculate this estimator we need a model for the conditional distribution of Z given $(X = x, W = w)$ to calculate $l(B_Z(t); x, w)$. Let $V = (X, W)$. If one takes $Z|V = v \sim N(\mu(v, \theta), \sigma^2)$ then

$$l(B_Z(t); v) = \mu(v, \theta) - \sigma^2 B_Z(t).$$

If Z is binary with $E(Z|V = v) = \text{expit}(\mu(v, \theta))$, one gets

$$l(B_Z(t); v) = \text{expit}(\mu(x) - B_Z(t)).$$

In both cases estimation is feasible. Large sample properties can be developed and hypotheses like $H : j(t) = j$ and $H : j(t) = 0$ may be investigated.

7 Discussion

In this paper, we have shown that the exposure coefficients under the Aalen additive hazards models, unlike those of the Cox proportional hazards model, can be interpreted as population-averaged effects in the case of no interaction between X and Z. Provided all relevant confounders have been adjusted for, the former coefficients thus succeed better to reflect what effect on the outcome would be attained in a study where the exposure were randomly assigned (and compliance with randomized assignment were perfect). Collapsibility of the association measures indexing the Aalen additive hazards model further forms the root cause as to why G-estimation is applicable to adjust for time-varying confounding in these models (Martinussen et al. 2011) and as to why direct exposure effects can be relatively easily separated from indirect exposure effects in these models (Lange and Hansen 2011). Finally, this property also makes simulation under marginal structural additive hazards models, such as $\lambda(t|\hat{x} = x) = \beta_0(t) + \beta_X(t)x$, straightforward as one can simulate from the corresponding additive hazard model for $\lambda(T|x, z)$ instead. For marginal structural Cox models, this is much less straightforward because the intricate relationship between the causal and the conditional hazard function, see (1), makes it difficult to find conditional hazard functions that satisfy the restrictions of a given population-averaged causal hazard function. One exception is when $\lambda(t|x, z) = z\lambda^*(t|x)$. Then,

$$\lambda(t|\hat{x} = x) = -\frac{d}{dt} \log\left(\phi_Z(\Lambda^*(t|x))\right),$$

where $\phi_Z(v) = E(e^{-Zv})$ is the Laplace transform, from which $\Lambda^*(t|x) = \phi_Z^{-1}(e^{-\Lambda(t|\hat{x}=x)})$. With a specific choice of Z's distribution the right-hand side of the latter display can then be calculated. Taking for example $Z \sim \Gamma(1/\theta, \theta)$ leads to

$$\lambda(t|x, z) = ze^{\theta\Lambda_0(t)e^{\beta x}} e^{\beta x}\lambda_0(t). \tag{9}$$

We conclude that generating $Z \sim \Gamma(1/\theta, \theta)$ and T under a hazard model for $\lambda(t|x, z)$ given by (9), implies the causal hazard function $\lambda(t|\hat{x} = x) = \lambda_0(t)e^{\beta x}$, thus suggesting a way of generating data from this marginal structural Cox model.

We have used G-computation to estimate the amount of confounding bias. Alternatively, we could have used inverse probability weighted estimators for marginal structural models (Hernán et al. 2000) to arrive at estimators of the causal hazard function, or related doubly robust estimators. An advantage of G-computation is that it tends to give more stable estimators by avoiding inverse probability weights, and that the standard models on which it relies can be readily validated using techniques described for instance in Martinussen and Scheike (2006); a disadvantage is that it is relatively more prone to model extrapolation.

In this paper we restricted the suggested methodology to the binary exposure case. We now briefly discuss the case where the exposure X is continuous assuming the additive hazards model (2). With $l(x, B_Z(t))$ defined in (4), the amount of confounding bias on hazard function scale is given by

$$\begin{aligned}&\lambda(t|\hat{X} = x + 1) - \lambda(t|\hat{X} = x) - \{\lambda(t|x + 1) - \lambda(t|x)\}\\ &= \beta_X(t) - [\beta_X(t) + \beta_Z(t)\{l(x + 1, B_Z(t)) - l(x, B_Z(t))\}]\\ &= -\beta_Z(t)\{l(x + 1, B_Z(t)) - l(x, B_Z(t))\},\end{aligned}$$

or on the cumulative scale,

$$-\int_0^t \{l(x + 1, B_Z(s)) - l(x, B_Z(s))\} dB_Z(s).$$

This can be estimated by plugging in the estimate $\hat{B}_Z(t)$ obtained from the fit of model (2) once we have an expression for the function g. If we assume that $Z|X = x \sim N(\mu(x), \sigma^2)$ then

$$l(x, B_Z(t)) = \mu(x) - \sigma^2 B_Z(t)$$

or if Z is binary with $E(Z|X = x) = \text{expit}(\mu(x))$, with $\text{expit}(a) = e^a/(1 + e^a)$, then

$$l(x, B_Z(t)) = \text{expit}(\mu(x) - B_Z(t)).$$

The case with a multidimensional Z may be worked out under a normality assumption but it might be more satisfying to solve the problem non-parametrically. One might attempt the approach where $l(x, B_Z(t))$ is estimated by

$$\frac{\sum_i Z_i e^{-B_Z(t)Z_i} W_x(X_i)}{\sum_i e^{-B_Z(t)Z_i} W_x(X_i)},$$

where $W_x(X_i)$ assigns heavy weight to subject i if X_i is close to x, and small weight otherwise (k-nearest neighbor for example). A detailed study of this approach is beyond the scope of this note, however.

The suggested approach can be used in the case where the hazard function fully parameterize the probability distribution. It would be interesting to try to extend the methodology to the situation where the hazard only partially parameterize the probability distribution, which would be the case in a competing risk scenario. This constitutes a topic for future research.

Acknowledgments The second author was supported by IAP research network grant nr. P06/03 from the Belgian government (Belgian Science Policy).

Appendix: Large sample properties

Aalen model

Let us first look at the Aalen model. Let $Y(t)$ be the $n \times 3$-matrix with ith row equal to $Y_i(t)(1, X_i, Z_i)$ where $Y_i(t)$ is the at-risk indicator, and let $Y^m(t)$ be the $n \times 2$-matrix with ith row equal to $Y_i(t)(1, X_i)$. We let $Y^-(t) = \{Y^T(t)Y(t)\}^{-1}Y^T(t)$ and similarly with $\{Y^m\}^-(t)$. Also let $A = (0, 1, 0)$ and $A^m = (0, 1)$, and let $N(t)$ denote the vector-counting process $(N_1(t), \ldots, N_n(t))^T$ with $N_i(t) = I(U_i \leq t, \delta_i)$. Since, with $B^m(t) = \{B_0^m(t), B_X^m(t)\}$,

$$\hat{B}_X(t) = \int_0^t AY^-(s)dN(s), \quad \hat{B}_X^m(t) = \int_0^t A^m\{Y^m\}^-(s)dN(s),$$

using the ordinary least squares inverses, it follows that

$$\epsilon_i^B(t) = \int_0^t AR^{-1}(s)Y_i(s)dM_i(s) - \int_0^t A^m\{R^m\}^{-1}(s)Y_i^m(s)dM_i^m(s),$$

where M_i and M_i^m denote the counting process martingales with respect to the marginal (conditioning on X) and the conditional (conditioning on X and Z) filtrations. In the latter display, $R(t)$ is the limit in probability of $n^{-1}Y^T(t)Y(t)$ and similarly with $R^m(t)$.

Cox model

We here sketch the proofs of the asymptotic results for the Cox model. The process $W_n(t)$ consists of differences like

$$\begin{aligned} n^{1/2}[\log[\hat{S}_j\{x, \hat{\theta}(t)\}] - \log[S_j\{x, \theta(t)\}]] &= \frac{n^{1/2}[\hat{S}_j\{x, \hat{\theta}(t)\} - S_j\{x, \theta(t)\}]}{S_j\{x, \theta(t)\}} + o_p(1) \\ n^{1/2}[\log[\hat{S}_j^x\{x, \hat{\theta}(t)\}] - \log[S_j^x\{x, \theta(t)\}]] &= \frac{n^{1/2}[\hat{S}_j^x\{x, \hat{\theta}(t)\} - S_j^x\{x, \theta(t)\}]}{S_j^x\{x, \theta(t)\}} + o_p(1), \end{aligned} \tag{10}$$

where

$$S_j\{x,\theta(t)\} = E[S_j\{x,\theta(t);Z\}],\ \hat{S}_j\{x,\theta(t)\} = n^{-1}\sum_{i=1}^{n} S_j\{x,\theta(t);Z_i\},$$

and

$$\begin{aligned} S_j^x\{x,\theta(t)\} &= E[S_j\{x,\theta(t);Z\}|X=x],\ \hat{S}_j^x\{x,\theta(t)\} \\ &= n_x^{-1}\sum_{i=1}^{n} S_j\{x,\theta(t);Z_i\}I(X_i=x) \end{aligned}$$

with $n_x = \sum_{i=1}^{n} I(X_i = x)$ and x being fixed. We now first show that

$$n^{1/2}[\hat{S}_j\{x,\hat{\theta}(t)\} - S_j\{x,\theta(t)\}] = n^{-1/2}\sum_{i=1}^{n} \tilde{\epsilon}_i^{S_j}(t,x) + o_p(1) \tag{11}$$

where $\left\{\tilde{\epsilon}_i^{S_j}(t)\right\}_i$ are zero-mean iid terms. This is similar to the proof in Chen et al. (2010, p. 724–725). Let $\mathcal{P}_n$ and P denote the empirical measure and the distribution under the true model, respectively. We write $\int f dQ$ as Qf, where Q here denotes a measure. With this notation, the left hand side of (11) can then be written as

$$\begin{aligned} &n^{1/2}(\mathcal{P}_n - P)[S_j\{x,\theta(t);Z\} - S_j\{x,\theta(t)\}] + Pn^{1/2}[S_j\{x,\hat{\theta}(t);Z\} \\ &-S_j\{x,\theta(t);Z\}] + n^{1/2}(\mathcal{P}_n - P)[S_j\{x,\hat{\theta}(t);Z\} - S_j\{x,\theta(t);Z\}], \end{aligned} \tag{12}$$

where the expectations $\mathcal{P}_n S_j\{x,\hat{\theta}(t);Z\}$ and $PS_j\{x,\hat{\theta}(t);Z\}$ are taken with respect to Z. As in Chen et al. (2010), the third term in (12) can be shown to converge to zero in probability using Lemma 19.24 of van der Vaart (1998) and because of the asymptotic properties of the partial likelihood estimator of the Cox model. The first term in (12) is already of the desired form so we can concentrate on the second term. We have

$$n^{1/2}[S_j\{x,\hat{\theta}(t);z\} - S_j\{x,\theta(t);z\}] = \dot{S}_j\{x,\theta(t);z\}n^{-1/2}\sum_{i=1}^{n}\epsilon_i^{\theta}(t) + o_p(1)$$

with $\dot{S}_j$ denoting the derivative of S_j with respect to the second (vector)-argument, and where $\{\epsilon_i^{\theta}(t)\}_i$ is the iid decomposition corresponding to $n^{1/2}\{\hat{\theta}(t)-\theta(t)\}$; expressions for these can be found in Martinussen and Scheike (2006) Ch. 7, and their empirical counterparts can be retrieved from the `cox.aalen`-function in the R-package `timereg`. Hence, (11) is fulfilled and therefore also

$$n^{1/2}\left[\log[\hat{S}_j\{x,\hat{\theta}(t)\}] - \log[S_j\{x,\theta(t)\}]\right] = n^{-1/2}\sum_{i=1}^{n}\epsilon_i^{S_j}(t,x) + o_p(1),$$

where

$$\epsilon_i^{S_j}(t,x) = \left[S_j\{x,\theta(t);Z_i\} - S_j\{x,\theta(t)\} + E[\dot{S}_j\{x,\theta(t);Z\}]\epsilon_i^{\theta}(t)\right]/S_j\{x,\theta(t)\}$$

are zero-mean iid terms. To handle (10) we let $n_x^* = n^{-1}\sum_{i=1}^n I(X_i = x)$ and assume that its limit in probability, p_x, say, is larger than zero. Now, rewrite the numerator of the right hand side of (10) as

$$n^{1/2}[\tilde{S}_j^x\{x,\hat{\theta}(t)\} - S_j^x\{x,\theta(t)\}] - \left[n^{-1/2}\sum_{i=1}^n \epsilon_i^{p_x}\right]S_j^x\{x,\theta(t)\}p_x^{-1} + o_p(1), \quad (13)$$

where

$$\tilde{S}_j^x\{x,\theta(t)\} = n^{-1}\sum_{i=1}^n S_j\{x,\theta(t);Z_i\}I(X_i = x)p_x^{-1},\ \epsilon_i^{p_x} = I(X_i = x) - p_x.$$

The first term in (13) can be dealt with as we did with the left hand side of (11), and the second term in (13) is already on the desired form. Thus

$$n^{1/2}\left[\log[\hat{S}_j^x\{x,\hat{\theta}(t)\}] - \log[S_j^x\{x,\theta(t)\}]\right] = \sum_{i=1}^n \epsilon_i^{S_j^x}(t,x) + o_p(1),$$

where

$$\epsilon_i^{S_j^x}(t,x) = \Big[S_j^x\{x,\theta(t);X_i,Z_i\} - S_j^x\{x,\theta(t)\} + E[\dot{S}_j^x\{x,\theta(t);X,Z\}]\epsilon_i^{\theta}(t) - \epsilon_i^{p_x}S_j^x\{x,\theta(t)\}p_x^{-1}\Big]/S_j^x\{x,\theta(t)\}$$

are zero-mean iid terms with $S_j^x\{x,\theta(t);X,Z\} = S_j\{x,\theta(t);Z\}I(X = x)p_x^{-1}$, and with $\dot{S}_j^x\{x,\theta(t);X,Z\}$ being its derivative with respect to the second (vector)-argument. Hence, $W_n(t) = n^{-1/2}\sum_{i=1}^n \epsilon_i^{\Delta}(t) + o_p(1)$, with

$$\epsilon_i^{\Delta}(t) = \epsilon_i^{\Delta_1}(t) - \epsilon_i^{\Delta_2}(t) \quad (14)$$

being zero-mean iid terms. In (14),

$$\epsilon_i^{\Delta_1}(t) = \epsilon_i^{S_1}(t,1) - \epsilon_i^{S_0}(t,1) + \epsilon_i^{S_0}(t,0) - \epsilon_i^{S_1}(t,0)$$
$$\epsilon_i^{\Delta_2}(t) = \epsilon_i^{S_1^1}(t,1) - \epsilon_i^{S_0^1}(t,1) + \epsilon_i^{S_0^0}(t,0) - \epsilon_i^{S_1^0}(t,0).$$

It readily follows that $\epsilon_i^{\mathcal{T}}(t) = \epsilon_i^{\Delta_1}(t) + \epsilon_i^{\beta_X}$, where $\epsilon_i^{\beta_X}$ is the first component of $\epsilon_i^{\theta}(t)$. Also $\epsilon_i^{\Delta_{nl}}(t) = \epsilon_i^{\Delta_1}(t)$.

References

Aalen OO (1980) A model for non-parametric regression analysis of counting processes. In: Klonecki W, Kozek A, Rosinski J (eds) Lecture notes in statistics-2: mathematical statistics and probability theory. Springer, New York, pp 1–25

Aalen OO (1989) A linear regression model for the analysis of life times. Stat Med 8:907–925

Aalen OO, Borgan Ø, Gjessing H (2008) Event history analysis: a process point of view. Springer, New York

Chen L, Lin DY, Zeng D (2010) Attributable fraction functions for censored event times. Biometrika 97:713–726

Cox DR (1972) Regression models and life-tables. J R Stat Soc Ser B 34:406–424

Greenland S, Robins JM (1986) Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 15:413–418

Greenland S, Robins JM, Pearl J (1999) Confounding and collapsibility in causal inference. Stat Sci 14:29–46

Hernán M, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11:561–570

Janes H, Dominici F, Zeger S (2010) On quantifying the magnitude of confounding. Biostatistics 11:572–582

Klein J, Moeschberger M (2003) Survival analysis: techniques for censored and truncated data. Springer, New York

Lange T, Hansen JV (2011) Direct and indirect effects in a survival context. Epidemiology 22:575–581

Martinussen T, Scheike TH (2006) Dynamic regression models for survival data. Springer, New York

Martinussen T, Vansteelandt S, Gerster M (2011) Estimation of direct effects for survival data using the Aalen additive hazards model. J R Stat Soc Ser B 73:773–788

Miettinen OS (1972) Components of crude risk ratio. Am J Epidemiol 96:168–172

Miettinen OS, Cook EF (1981) Confounding: essence and detection. Am J Epidemiol 114:593–603

Pearl J (2000) Causality: models, reasoning, and inference. Cambridge University Press, Cambridge

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. Math Model 7:1393–1512

Rothman KJ, Greenland S, Lash TL (2008) Modern epidemiology. Lippincott Williams & Wilkins, Philadelphia

Stare J, Henderson R, Pohar M (2005) An individual measure of relative survival. Appl Statist 54:115–126

Tsiatis AA, Davidian M, Zhang M, Lu X (2000) Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. Stat Med 27:4658–4677

van der Vaart AW (1998) Asymptotic statistics. Cambridge University Press, Cambridge

Vansteelandt S, Keiding N (2011) Invited commentary: G-computation—lost in translation? Am J Epidemiol 173:739–742