

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Prior Event Rate Ratio Adjustment for Hidden Confounding in Observational Studies of Treatment Effectiveness: A Pairwise Cox Likelihood Approach

Nan Xuan Lin^{a,b} and William Edward Henley^{b*}

Observational studies provide a rich source of information for assessing effectiveness of treatment interventions in many situations where it is not ethical or practical to perform randomized controlled trials. However, such studies are prone to bias from hidden (unmeasured) confounding. A promising approach to identifying and reducing the impact of unmeasured confounding is Prior Event Rate Ratio (PERR) adjustment, a quasi-experimental analytic method proposed in the context of electronic medical record database studies. In this paper we present a statistical framework for using a pairwise approach to PERR adjustment that removes bias inherent in the original PERR method. A flexible pairwise Cox likelihood function is derived and used to demonstrate the consistency of the simple and convenient PERR-ALT estimator. We show how to estimate standard errors and confidence intervals for treatment effect estimates based on the observed information, and provide R code to illustrate how to implement the method. Assumptions required for the pairwise approach (as well as PERR) are clarified, and the consequences of model misspecification are explored. Our results confirm the need for researchers to consider carefully the suitability of the method in the context of each problem. Extensions of the pairwise likelihood to more complex designs involving time-varying covariates or more than two periods are considered. We illustrate the application of the method using data from a longitudinal cohort study of enzyme replacement therapy for lysosomal storage disorders. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: Prior event rate ratio, pairwise Cox model, unmeasured confounding, observational study, treatment effect

1. Introduction

Observational studies, based on routinely collected patient data or data from population-based cohorts, offer a rich source of information for evaluating real-world effectiveness of medical treatments [1]. With the anticipated growth in implementation of electronic health record systems, there is increasing scope for using large observational datasets to inform the design of randomized trials and to address clinical questions for which trials are unlikely to be conducted due to ethical or logistical considerations. However, a major challenge in adopting this approach is the need to remove bias introduced due to confounding by indication or other biases due to the effect of unmeasured covariates [2]. Analyses that fail to account for relevant confounders may have important negative consequences for health policy and patient safety.

Tannen *et al.* [3] introduced prior event rate ratio (PERR) adjustment, a new quasi-experimental analytic approach to identifying and reducing hidden (unmeasured) confounding in the analysis of time-to-event data from clinical database studies.

^a Department of Mathematics and Information sciences, Northumbria University, Newcastle upon Tyne, NE2 1XE, UK.

^b Health Statistics Group, Institute of Health Research, University of Exeter Medical School, Exeter, EX1 2LU, UK.

* Correspondence to: William Edward Henley, Health Statistics Group, Institute of Health Research, University of Exeter Medical School, Exeter, EX1 2LU, UK.
E-mail: W.E.Henley@exeter.ac.uk

The PERR approach replicates a randomised trial by identifying a group of individuals from a clinical database using the same inclusion and exclusion criteria, a similar study time frame and a similar treatment regimen as the trial. The exposed group are the individuals who received treatment within the recruitment interval of the trial. The unexposed group are individuals not receiving the treatment within the recruitment window. Previous studies using the PERR method have generally matched unexposed patients to exposed patients on an index date (often the date that the exposed patients first received treatment). The method relies on a before-and-after design and assumes that the hazard ratio of the exposed to unexposed for a specific outcome before the start of the trial reflects the combined effect of all confounders independent of any influence of treatment. Let HR_s be the unadjusted hazard ratio of treated versus control during the study from a Cox regression model, and HR_p be the corresponding unadjusted hazard ratio of treated versus control in the prior period. To account for unmeasured confounding, the PERR adjusted hazard ratio is given by $HR_{PERR} = HR_s / HR_p$. The schematic of this method is displayed in Figure 1

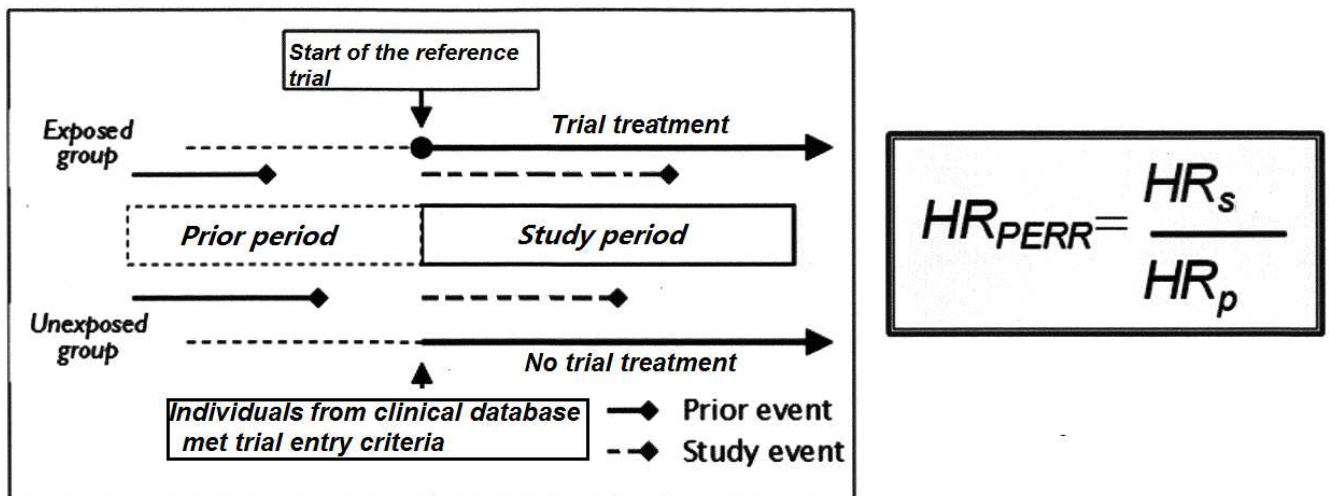


Figure 1. Schematic for the Prior Event Rate Ratio Adjustment method

Two studies by Tannen and colleagues reported good concordance between PERR results and those from published RCTs. First, observational data from the CPRD (Clinical Practice Research Datalink) were used to replicate the Scandinavian Simvastatin Survival Study of statin treatment for hypercholesterolemic subjects with coronary heart disease [4]; second, data from the same source were used to replicate two randomized trials of angiotensin-converting enzyme inhibitors in patients without congestive heart failure at high risk for cardiovascular disease [5]. Tannen *et al.* [5] conducted a broader validation study in which the PERR adjusted HR s were not significantly different from the trial HR s in 5 out of 8 comparisons of cardiovascular outcomes where unadjusted HR s were inconsistent with the trial results (suggesting the presence of unmeasured confounding). However, Yu *et al.* [6] noted that the original PERR method frequently led to attenuated treatment estimates in simulation experiments and introduced an alternative, referred to as PERR-ALT, based on a paired Cox model. This approach gave unbiased estimates in simulation studies where the unmeasured confounder effect did not vary temporally. More generally, both PERR methods performed well in reducing bias when the treatment effect was large compared to any confounder-treatment interaction. The original PERR estimator was more computationally stable than PERR-ALT when the event rate was low and the sample size was limited.

Given the promising results in these early studies, the PERR method is gaining acceptance as a useful approach for addressing hidden confounding when comparing or evaluating treatments in observational studies [2]. For example, PERR adjustment has been used to assess the incidence of *Campylobacter* and *Salmonella* infection in patients prescribed proton pump inhibitors (PPIs) compared with controls, using electronic health records from the Secure Anonymised Information Linkage (SAIL) databank [7]. Tannen *et al.* [8] proposed a strategy for performing comparative effectiveness research using the THIN database in which PERR adjustment is used to remove bias due to unmeasured confounding.

The aim of this paper is to set out a detailed statistical framework for using a pairwise approach to PERR adjustment and to address some of the methodological challenges. Yu *et al.* [6] present a simple and convenient formula for PERR-ALT adjustment. We extend their approach by deriving a flexible pairwise Cox likelihood function and using this to show that the PERR-ALT method is consistent, under relevant assumptions. The pairwise likelihood can be used to obtain consistent estimates of treatment effects, other measured covariates and period effects. We also consider ways in which the likelihood can be extended to relax the assumptions of the original PERR method. We address the issue of estimating standard errors and confidence intervals for the treatment effect estimates. Previous work has used a bootstrapping technique to estimate

confidence intervals for the PERR-ALT method because of difficulties in estimating the covariance between HR_s and HR_p [8]. We provide a direct approach based on the observed information matrix.

The paper is organized as follows. The definitions of prior and study event times, the hazard models and assumptions required for using the pairwise Cox method (as well as PERR and PERR-ALT) in the analysis of observational data are clarified in Section 2. In Section 3, we consider the nature of the bias in the original PERR method using asymptotic bias formulae. In Section 4, we propose a statistical test for detecting hidden confounding using prior event data. The pairwise Cox likelihood function is derived in Section 5 and formulae given for estimating standard errors. We also consider how the method could be applied in the context of crossover trials. Section 6 discusses situations in which the PERR method will be prone to bias. Extensions of the pairwise likelihood that permit more flexible modelling are considered in Section 7. The method is applied to data from a longitudinal cohort study of enzyme replacement therapy for lysosomal storage disorders in Section 8.

2. Definitions, assumptions and hazard model

We start by introducing some notation and setting out a general framework for quasi-experimental analytic studies using a pairwise Cox approach. Here random variables are denoted by upper case letters and their values are denoted by lower case letters.

In what follows we assume that there are two types of events of interest: the study event and the prior event. Usually, the study event is defined so that some individuals receive a treatment or exposure during or before the time of the study event (the exposed group) and some do not (the unexposed group). The prior event refers to an earlier event when no individuals are treated. However, these requirements are not compulsory. As illustrated below in an example related to treatment for diabetic retinopathy, there could be no order between the prior and study events and patients can be treated during or before the time of the prior event. We note that the approach can easily be generalised to more than two events as considered in Section 7.2.

Designing studies based on the pairwise Cox likelihood method requires the user to define the events and time origins carefully to reflect the requirements of the research problem. First, the prior and study events need to be defined. Usually, the prior and study events should be of the same nature so that the unknown confounders will have the same effects on the hazards of prior and study events. For example, this may be a reasonable assumption in a study investigating the protective effects of oestrogen replacement therapy on risk of myocardial infarction (MI) in post-menopausal women where previous MI is used as the prior event [8].

In order to define the event times used in the statistical analysis, the time origins when individuals first become at risk for the prior and study events need to be defined in the context of the research problem. This fixes the prior and study periods that provide the framework for the pairwise design. We define T_p^* for the variable of prior event time measured from the prior start point to a prior event and T_p^+ to be the corresponding variable for the prior censoring time. The prior start point is defined as the time origin when the individual is considered to be at risk of experiencing a prior event. The variable for the observed prior time is $T_p = \min(T_p^*, T_p^+)$ with the variable for the prior censoring indicator $\Delta_p = I(T_p^* \leq T_p^+)$, where $I(T_p^* \leq T_p^+)$ is an indicator function such that $\Delta_p = I(T_p^* \leq T_p^+) = 1$ if $T_p^* \leq T_p^+$ and 0 otherwise. We call the time span that the individual is at the risk of a prior event (or being observed for a prior event) the prior period, which is measured from the prior start point to the observed time T_p . The definition of prior period means that an individual is at the risk of prior event (or being observed for a prior event) until the prior event is observed ($T_p = T_p^*$) or right censored ($T_p = T_p^+$). Similar definitions apply to the study start point, study period and the variables $(T_s^*, T_s^+, T_s, \Delta_s)$ for the study event, censoring and observed times and censoring indicator respectively. Note that it is generally not appropriate when using the pairwise and PERR methods to measure T_s from the time of receiving treatment as this is often not the time when an individual starts to be under the risk of a study event (see below).

To illustrate the process of model formulation, we consider the example of assessing the effectiveness of influenza vaccination in older adults in the UK using electronic medical record data. Influenza is a major cause of illness and death amongst the elderly, and reliable estimates of vaccine effectiveness are important for informed vaccine policies and programmes [9]. To apply the pairwise Cox method we first need to define the prior and study events: if the focus is on more serious outcomes, then one possible event would be hospitalization for an influenza-related illness. We make the simplifying assumption that individuals are at risk of influenza during the winter season (from 1st October to 31st March each year) but not during the summer season (1st April to 30th September each year). Suppose data are available from the 2013/14 and 2014/15 influenza seasons. The prior and study time origins are 1st October 2013 and 1st October 2014 respectively (see Fig. 2).

Let $X = (X_1, \dots, X_q)^{tr}$ be q measured covariates in the study period, where tr is the notation of transpose, and C be the column vector for hidden covariates (or confounders). Conditional on X and C , (T_p, T_s) and (T_p^+, T_s^+) are assumed



Figure 2. Timeline for an illustrative study of influenza vaccination effectiveness using the pairwise cox method

to be independent. The independent censoring assumption is sensible and has been widely used in multiple events models such as Prentice, Williams and Peterson [11] and Wei, Lin and Weissfeld [12] (hereafter referred to as PWP and WLW). For example, in the influenza vaccination study, if patients are not infected with flu from Oct 2013 to Apr 2015, the prior and study events are right censored with $T_p^+ = T_s^+ = 6$ months (i.e. the length of flu season). In this case, the censoring is similar to administrative censoring.

In the presence of death, if death is related to the prior or study events, a competing risk model will be needed [10]. In this paper, we focus on the case where death is not related to the prior or study events and can be regarded as right censoring. If the individual died after the prior event and before the study period, $T_{si} = 0$ and $\Delta_{si} = 0$. We will show that this kind of data has no contribution to the pairwise Cox likelihood.

Let $(T_{pi}, T_{pi}^*, T_{pi}^+, \Delta_{pi}, T_{si}, T_{si}^*, T_{si}^+, \Delta_{si}, X_i, C_i)$ be n independent replicates of $(T_p, T_p^*, T_p^+, \Delta_p, T_s, T_s^*, T_s^+, \Delta_s, X, C)$. It is assumed that the hazard of prior event at $T_{pi}^* = t$ is

$$h_{pi}(t) = I(t \leq T_{pi})h_{0p}(t) \exp(\beta C_i) \tag{1}$$

where $h_{0p}(t)$ is the unspecified baseline hazard for the prior event, β is the vector of coefficients for C_i and $I(t \leq T_{pi})$ indicates whether t is within the prior period, i.e. whether the individual is at the risk of a prior event (or being observed for a prior event) until the prior event is observed or a censorship occurs.

Similarly, we assumed the hazard of a study event at $T_{si}^* = t$ is

$$h_{si}(t) = I(t \leq T_{si})h_{0s}(t) \exp(\theta X_i + \beta C_i) \tag{2}$$

where $h_{0s}(t)$ is the unspecified baseline hazard for the study event, $\theta = (\theta_1, \dots, \theta_q)$ is the vector of coefficients for X_i and $I(t \leq T_{si})$ is the indicator for the study period.

In (1) and (2), βC_i is assumed to be time invariant. In practice, the prior and study events need to be carefully defined to satisfy this assumption. As we said, usually the prior and study events should be of the same nature. As a consequence, the pairwise Cox likelihood approach is only applicable to problems where the prior and study events are non-terminal events. In particular, the method cannot be applied when death is an outcome of interest.

We maintain that a careful clarification of research purpose is needed before one can choose a statistical model. Like the PWP and WLW models, the basic models (1)–(2) allow for the different events (prior and study) to have different baseline hazards (h_{0p} and h_{0s}). The WLW model usually operates with a common time origin while the basic models (1)–(2) allow the different event times to have different time origins. In some cases, the Andersen and Gill model is suitable if the events shared a common baseline hazard, while in other cases, it is reasonable to assume different baseline hazards for different events. For example in the study of influenza vaccination, it is reasonable to allow different baseline hazards, $h_{0p}(t)$ and $h_{0s}(t)$, for the prior and study events because of the unknown changes in conditions between the two flu seasons (eg due to differences in the circulating strains of influenza).

The basic models (1)–(2) can be used wherever the PWP, WLW and the Andersen and Gill [13] models are applicable. For example, if the prior and study events are the first and second events respectively, then T_{pi} and T_{si} are the times for the first and second events respectively. In this case, by adopting a time-varying treatment indicator $X_i(t)$, the basic models (1)–(2) are the same as

- the PWP model, if T_{si} is measured from the time point of the prior event
- the WLW model, if T_{si} is measured from the same time origin as T_{pi}
- the Andersen and Gill model, if T_{si} is measured from the same time origin of T_{pi} and we assume $h_0 = h_{0p} = h_{0s}$.

The definition of events and measurement of times are important when using the PERR/pairwise method. Returning to the oestrogen replacement therapy example, it is not appropriate to define the prior event as the first MI (no matter whether the patient is treated or not) and the study event as the first MI after the treatment, and measure T_s from τ_i , the treatment time for the i th individual. Consider, for example if the treatment has no effect ($\theta = 0$), a first MI after treatment at $T_{si} = t$ without a prior MI can be regarded as both the prior and study events and will consequently have two different hazards $h_{0p}(\tau_i + t) \exp(\beta c_i) \neq h_{0s}(t) \exp(\beta c_i)$ (the common baseline hazard h_{0s} could not be the same as $h_{0p}(\tau_i + t)$ which depends on individual i). For this example, it is more appropriate to define the prior event as the first MI and the

study event as the second MI. The time T_s can be measured from the first MI (like the PWP model) or from the prior start point (like the AG and WLW models). The pairwise approach can be easily generalized to the cases where the third, fourth or more MIs are of interest by using a time-varying treatment variable.

Another example is an observational study of the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy. A patient could potentially experience blindness in both eyes; we define the prior and study events to be left and right eye blindness respectively. In this example, there is no order between the prior and study events and the prior and study periods overlap. The right eye can be blind before the left eye. The patient starts to be at the risk of blindness for both eyes from the same time origin. A WLW-type model can be used:

$$\begin{aligned} h_{pi}(t) &= I(t \leq T_{pi})h_{p0}(t) \exp(\theta_p X_{ip} + \beta C_i) \\ h_{si}(t) &= I(t \leq T_{si})h_{s0}(t) \exp(\theta_s X_{is} + \beta C_i), \end{aligned}$$

where $X_{ip} = 1$ if the left eye was on treatment and 0 otherwise, $X_{is} = 1$ if the right eye was on treatment and 0 otherwise, and θ_p and θ_s are the treatment effect on the left and right eyes respectively. The process of observing the left eye blindness is stopped (i.e. $I(t \leq T_{pi}) = 0$) if the left eye is blind ($t = t_{pi}^*$) or the follow up ends ($t = t_{pi}^+$). The likelihood function for this example is given in (16).

For the influenza vaccination example, the basic models (1)–(2) can be extended as

$$\begin{aligned} h_{pi}(t) &= I(t \leq T_{pi})h_{p0}(t) \exp(\theta X_{ip}(t) + \beta C_i) \\ h_{si}(t) &= I(t \leq T_{si})h_{s0}(t) \exp(\theta X_{is}(t) + \beta C_i), \end{aligned}$$

where T_{pi} and T_{si} are, respectively, measured from Octobers of 2013 and 2014 to the times of hospitalizations for an influenza-related illness or to Aprils of 2014 and 2015 (if the times are censored). The time origins are defined as the start of the 2013 and 2014 influenza seasons because these are the points at which patients first become at risk for a prior or study event respectively. We note that in this example, there is a gap in being under risk between the prior and study periods (during the summer period in 2014). In this type of problem, it is not appropriate to define the study time origin at the first event (like the PWP model) or at the start of the prior period (like the AG and WLW models). The definition of the time origins in this example requires the indicators of vaccinations $X_{ip}(t)$ and $X_{is}(t)$ to be time-varying: $X_{ip}(t) = 1$ if the individual i was vaccinated at or before the time $t_{pi} = t$ and 0 otherwise, $X_{is}(t) = 1$ if the individual i was vaccinated at or before the time $t_{si} = t$ and 0 otherwise. If there is a prior event at time t_{pi}^* , then $T_{pi} = t_{pi}^*$. The indicator for the prior period $I(t \leq T_{pi}) = 0$ for $t > t_{pi}^*$ reflecting the assumption that there is no risk of flu after a previous hospitalization related to influenza within the same flu season. This is because an individual contracting influenza develops antibodies to the circulating strain. If there is no prior event, the prior event time is censored and the observed prior time T_{pi} is equal to the length of the flu season, i.e. 6 (months). The indicator for the prior period $I(t \leq T_{pi}) = 0$ for $t > 6$ reflecting the assumption that there is no risk of flu after the flu season.

3. Bias in the PERR method

Suppose the true hazards are (1) and (2). The PERR method fits the following models for the prior and study event times:

$$I(t \leq T_{pi})h_{0p}^*(t) \exp(\theta_p^* X_i) \tag{3}$$

$$I(t \leq T_{si})h_{0s}^*(t) \exp(\theta_s^* X_i), \tag{4}$$

respectively. The parameters $\theta_p^* = \log(HR_p)$ and $\theta_s^* = \log(HR_s)$ are the log hazard ratios without adjustment for C_i for prior and study events respectively. The PERR adjusted log-HR is given by $\log(HR_{PERR}) = \theta_{PERR} = \theta_s^* - \theta_p^*$. The parameters θ_p^* and θ_s^* are estimated from the log partial likelihoods

$$\frac{1}{n} \sum_i^n \Delta_{ip} \left\{ \theta_p^* X_i - \log \sum_j^n I(T_{pi} \leq T_{pj}) \exp(\theta_p^* X_j) \right\} \tag{5}$$

$$\frac{1}{n} \sum_i^n \Delta_{is} \left\{ \theta_s^* X_i - \log \sum_j^n I(T_{si} \leq T_{sj}) \exp(\theta_s^* X_j) \right\} \tag{6}$$

respectively.

However, the PERR adjustment cannot entirely remove the bias from the estimate of the treatment effect, i.e. $\theta_{PERR} \neq \theta$. Lin *et al.* [14] studied the problem of hidden confounding in the Cox regression model and showed that

due to the nonlinearity of the Cox model, the resulting bias is a complicated combination of three sources: (a) bias due to omission of hidden covariates, even if they are not confounders, (b) bias due to censoring and (c) bias due to the hidden covariates being confounders. Here we conducted a simple simulation study to illustrate the impact of these three sources of bias on the performance of the PERR method. We generated 100000 paired times (t_p, t_s) for prior and study events from the models (1) and (2) respectively with $h_{0p}(t) = \theta = 1$, $h_{0s}(t) = \exp(1)$, $X \sim \text{Bin}(1, 0.5)$ and the log HR of the hidden covariate, β , taking 100 sequenced values from -10 to 10 . The data were fitted by the model (3) and (4). Since the sample size is large ($n = 100000$), the bias of the unadjusted estimate $\hat{\theta}_s^*$ for the naive Cox model and the bias of the PERR estimate $\hat{\theta}_{PERR}$ were approximated by $\hat{\theta}_s^* - \theta$ and $\hat{\theta}_p^* - \theta$ respectively.

We present bias curves in Fig. 3 showing the relationship between the bias of the PERR estimate and the value of β under three different scenarios to separate out the effect of the three potential sources of bias: Fig 3(a) shows the bias when there is a hidden balanced covariate $C \sim \text{Bin}(1, 0.5)$ in the absence of censoring; Fig 3(b) shows the bias when again $C \sim \text{Bin}(1, 0.5)$ but in addition both t_p and t_s are 50% censored; Fig 3(c) shows the bias when again t_p and t_s are 50% censored but C is now a hidden confounder with distribution $\text{Bin}(1, 0.3 + 0.4X)$ (ie the omitted covariate has the same marginal distribution as (a) and (b) but the conditional distribution depends on X).

The three figures show that PERR adjustment is biased even if the hidden covariate C is not a confounder. In Fig3 (a) and (b), it is hard to see the difference between the PERR method and the naive unadjusted Cox model because the bias curves for θ_s^* and θ_{PERR} overlap closely for all values of β : PERR adjustment fails to remove any of the bias attributable to censoring or to hidden covariates that are balanced across treatment groups. The figures show that the bias of PERR adjustment increases with the absolute value of the log-hazard ratio β of the hidden covariate. The direction of the bias is always the same: treatment effect estimates are attenuated. The PERR adjustment is only successful in removing the bias attributable to confounding, as shown in Fig3 (c). The unadjusted Cox model fails to remove the hidden confounding bias and the direction of the bias depends on the sign of β and the distribution of C conditional on X .

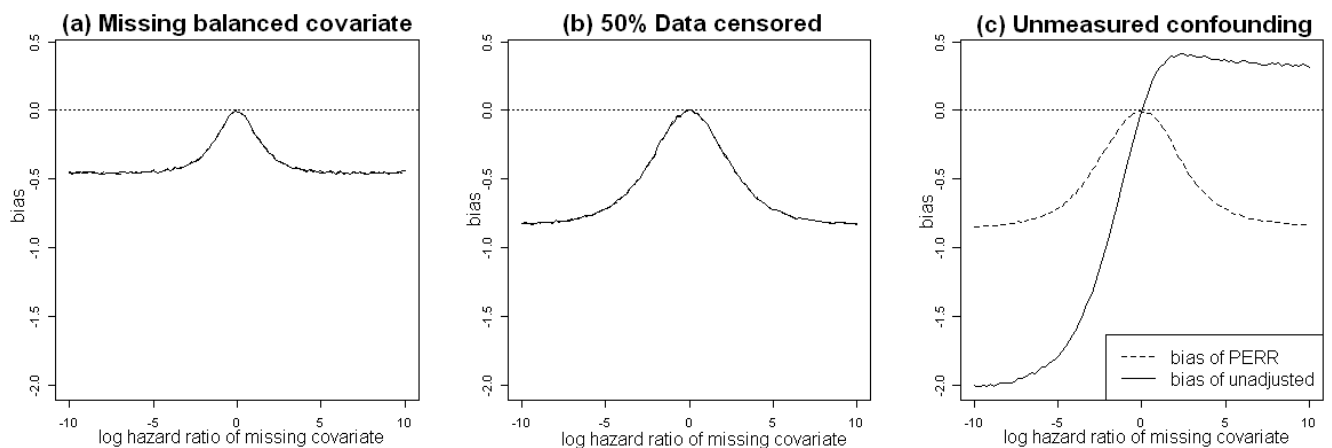


Figure 3. Bias curves for the PERR estimate θ_{PERR} and the unadjusted estimate θ^* of the naive Cox model in the presence of omitted covariates. Graphs based on simulated dataset with $n = 100,000$ and true log hazard ratio of treatment $\theta = 1$. An exponential baseline hazard was used with a rate parameter of $h_0(t) = 1$. The censoring mechanism was uniform.

Although the models (3) and (4) are misspecified, the maximal likelihood estimates $\hat{\theta}_p^*$ and $\hat{\theta}_s^*$ converge to well-defined constants as $n \rightarrow \infty$ [15]. The limiting values of $\hat{\theta}_p^*$ and $\hat{\theta}_s^*$ are sometimes called the “least false” value [16]. For simplicity of notation, we use the same θ_p^* and θ_s^* to denote these limits. By calculating the limits of the score functions of the likelihood (5) and (6) as $n \rightarrow \infty$ under the true models (1) and (2) and following the result of Lin *et al.* [14], it can be shown that the limits θ_p^* and θ_s^* are the solutions of the equations

$$\begin{aligned}
 0 = U_p(\theta_p^*) &= E \left(\Delta_p \left[X - \frac{E_{xc} \left\{ S_p^+(T_p|X) e^{\theta_p^* X} e^{-H_{0p}(T_p) e^{\beta C}} X \right\}}{E_{xc} \left\{ S_p^+(T_p|X) e^{\theta_p^* X} e^{-H_{0p}(T_p) e^{\beta C}} \right\}} \right] \right) \\
 0 = U_s(\theta_s^*) &= E \left(\Delta_s \left[X - \frac{E_{xc} \left\{ S_s^+(T_s|X) e^{\theta_s^* X} e^{-H_{0s}(T_s) e^{\theta X + \beta C}} X \right\}}{E_{xc} \left\{ S_s^+(T_s|X) e^{\theta_s^* X} e^{-H_{0s}(T_s) e^{\theta X + \beta C}} \right\}} \right] \right),
 \end{aligned} \tag{7}$$

where $H_{0p}(T_p) = \int_0^{T_p} h_{0p}(t) dt$, $H_{0s}(T_s) = \int_0^{T_s} h_{0s}(t) dt$, E_{xc} is the expectation with respect to X and C , and $S_p^+(T_p|X)$ and $S_s^+(T_s|X)$ are the survival functions of the prior and study censoring times conditional on X respectively.

The first-order Taylor series approximation for the bias of PERR adjustment is

$$\theta_{PERR} - \theta \approx I_s^{-1}(\theta)U_s(\theta) - I_p^{-1}(0)U_p(0),$$

where $I_s(\theta_s^*) = -\partial U_s(\theta_s^*)/\partial \theta_s^*$ and $I_p(\theta_p^*) = -\partial U_p(\theta_p^*)/\partial \theta_p^*$.

4. Use of prior event data to detect unmeasured confounding

A hypothesis test for detecting unmeasured confounding can be developed based on the first equation in (7). Suppose $C \perp X$ so that C is an unmeasured balanced covariate but not a confounder, the first equation in (7) reduces to

$$E(\Delta_p X) = E \left(\Delta_p \left[\frac{E_x \{ S_p^+(T_p|X) e^{\theta_p^* X} \}}{E_x \{ S_p^+(T_p|X) e^{\theta_p^* X} \}} \right] \right) \quad (8)$$

Under the independent censoring assumption $T_p^* \perp T_p^+ | (X, C)$ [12], the survival function for the prior censoring time can be written as $S_p^+(t|X, C)$. Since X is the set of covariates associated with the study period, $S_p^+(t|X, C) = S_p^+(t|C)$. With $C \perp X$, it follows

$$S_p^+(t|X) = \int S_p^+(t|X, c) f(c|X) dc = \int S_p^+(t|c) f(c) dc = S_p^+(t). \quad (9)$$

Therefore, (8) is further simplified to

$$E(X) = \frac{E(e^{\theta_p^* X})}{E(e^{\theta_p^* X})},$$

which has the solution $\theta_p^* = 0$.

This provides the basis for the following procedure to test for the presence of an unmeasured confounder using the prior event data:

1. Suppose the null hypothesis is

$$H_0 : C \perp X \quad \text{so that } C \text{ is an unmeasured balanced covariate but not a confounder}$$

2. To assess (9), fit the model (3) to the data t_p and x but with the censoring indicators reversed (i.e. by considering events as “censored” observations and censored observations as “events”). If the p-value for the estimate $\hat{\theta}_p^*$ is greater than 0.05, it indicates that (9) is deemed valid.
3. Fit the model (3) to t_p and x again but without reversing the censoring indicators. Under the null hypothesis H_0 and with (9), the distribution of the Wald test statistic for $\hat{\theta}_p^*$ can be approximated by the standard normal distribution $N(0, 1)$.
4. Calculate the p-value. Reject H_0 , if the p-value is less than 0.05.

Note that even if H_0 is not rejected, the pairwise method is still recommended since the unadjusted estimate $\hat{\theta}_s^*$ may still incorporate bias due to hidden balanced covariates and censoring, as illustrated in Fig. 3 (a) and (b).

5. Pairwise Cox Likelihood

For the hazards (1) and (2), we first write the baseline hazard ratio as

$$\frac{h_{0s}(t)}{h_{0p}(t)} = \exp\{\alpha(t)\}.$$

When evaluating effectiveness of treatments for many common chronic conditions such as cardiovascular disease, it is likely that $h_{0s}(t) > h_{0p}(t)$ and so $\exp\{\alpha(t)\} > 1$ for all $T_{pi} = T_{si} = t \geq 0$.

If we assume the baseline hazards $h_{0s}(t)$ and $h_{0p}(t)$ are proportional over time (analogous to the familiar Cox proportional hazards assumption), the log baseline hazard ratio is then equal to a constant $\alpha(t) = \alpha$. The parameter α can be regarded as a period effect. The hazards (1) and (2) now become

$$h_{pi}(t) = I(t \leq T_{pi}) h_{0p}(t) \exp(\beta C_i) \quad (10)$$

$$\begin{aligned} h_{si}(t) &= I(t \leq T_{si}) h_{0p}(t) \exp(\theta X_i + \beta C_i + \alpha) \quad (11) \\ &= I(t \leq T_{si}) h_{0pi}(t) \exp(\theta X_i + \alpha), \end{aligned}$$

where the hazards (10) and (11) now share the same term $h_{0pi}(t) = h_{0p}(t) \exp(\beta C_i)$, which can be regarded as an individual-specific baseline hazard. For n individuals, the full likelihood is

$$L_{full} = \prod_i^n \{h_{0pi}(T_{pi})\}^{\Delta_{pi}} e^{-H_{0pi}(T_{pi})} \{h_{0pi}(T_{si}) \exp(\theta X_i + \alpha)\}^{\Delta_{si}} e^{-H_{0pi}(T_{si}) e^{\theta X_i + \alpha}}, \quad (12)$$

where $H_{0pi}(t) = \int_0^t h_{0pi}(u) du$ is the cumulative hazard function. It is clear that we cannot obtain a consistent estimate of θ using (12), since the number of nuisance parameters $h_{0p1}(t), \dots, h_{0pn}(t)$ increases with the sample size n . We face the infinitely many nuisance parameters problem. The nuisance parameter $h_{0pi}(t)$ needs to be removed.

Suppose that θ and α are given, then as shown in Appendix A, the MLEs of $h_{0pi}(t)$ and $H_{0pi}(t)$ obtained from L_{full} are

$$h_{0pi}(T_{pi}) = \frac{\Delta_{pi}}{1 + P_i e^{\theta X_i + \alpha}}, \quad H_{0pi}(T_{pi}) = \frac{\Delta_{pi}}{1 + P_i e^{\theta X_i + \alpha}} + \frac{S_i \Delta_{si}}{S_i + e^{\theta X_i + \alpha}} \quad (13)$$

$$h_{0pi}(T_{si}) = \frac{\Delta_{si}}{S_i + e^{\theta X_i + \alpha}}, \quad H_{0pi}(T_{si}) = \frac{\Delta_{si}}{S_i + e^{\theta X_i + \alpha}} + \frac{P_i \Delta_{pi}}{1 + P_i e^{\theta X_i + \alpha}}, \quad (14)$$

where $P_i = I(T_{pi} \leq T_{si})$ and $S_i = I(T_{si} \leq T_{pi})$

Plug the results (13) and (14) back into (12) and we get the pairwise Cox likelihood

$$L(\theta, \alpha) = \prod_i^n \left(\frac{1}{1 + P_i e^{\theta X_i + \alpha}} \right)^{\Delta_{pi}} \left(\frac{e^{\theta X_i + \alpha}}{e^{\theta X_i + \alpha} + S_i} \right)^{\Delta_{si}}. \quad (15)$$

The likelihood (15) is now free of the unknown term $h_{0pi}(t) = h_{0p}(t) \exp(\beta C_i)$.

As shown in Web Appendix A, if we present the hazard model (10)–(11) and the likelihood (15) in the form of counting process, using the same techniques in the proofs of Gross and Huber [17], it can be shown that the estimates $\hat{\theta}$ and $\hat{\alpha}$ from (15) are consistent and asymptotically normal. The covariance matrix can be estimated by

$$\left[\sum_{i=1}^n e^{\hat{\theta} X_i + \hat{\alpha}} \left\{ \frac{\Delta_{si} S_i}{(e^{\hat{\theta} X_i + \hat{\alpha}} + S_i)^2} + \frac{\Delta_{pi} P_i}{(1 + P_i e^{\hat{\theta} X_i + \hat{\alpha}})^2} \right\} \begin{pmatrix} X_i X_i^{tr} & X_i \\ X_i^{tr} & 1 \end{pmatrix} \right]^{-1},$$

and the confidence interval can be constructed based on a normal approximation.

We can also show the consistency of the PERR-ALT adjustment in Yu *et al.* [6]. For the exposed group ($X_i = 1$), the hazards are

$$\begin{aligned} h_{pi}(t) &= I(t \leq T_{pi}) h_{p0}(t) \exp(\beta C_i) \\ h_{si}(t) &= I(t \leq T_{si}) h_{p0}(t) \exp(\theta + \beta C_i + \alpha). \end{aligned}$$

The pairwise likelihood is then

$$L(\theta + \alpha) = \prod_i \left(\frac{1}{1 + P_i e^{\theta + \alpha}} \right)^{\Delta_{pi}} \left(\frac{e^{\theta + \alpha}}{e^{\theta + \alpha} + S_i} \right)^{\Delta_{si}}.$$

Since $(\hat{\theta}, \hat{\alpha})$ are consistent for (θ, α) , the HR estimate for the exposed group (as denoted by $\hat{H}R_E$ in Yu *et al.* [6]), $\hat{H}R_E = \exp(\widehat{\theta + \alpha})$ is consistent for $\exp(\theta + \alpha)$.

For the unexposed group ($x_i = 0$), the hazards are

$$\begin{aligned} h_{pi}(t) &= I(t \leq T_{pi}) h_{p0}(t) \exp(\beta C_i) \\ h_{si}(t) &= I(t \leq T_{si}) h_{p0}(t) \exp(\beta C_i + \alpha), \end{aligned}$$

and the paired likelihood is

$$L(\alpha) = \prod_i \left(\frac{1}{1 + p_i e^{\alpha}} \right)^{\delta_{pi}} \left(\frac{e^{\alpha}}{e^{\alpha} + s_i} \right)^{\delta_{si}}.$$

Similarly, it can be shown that that the HR estimate for the unexposed group (as denoted by $\hat{H}R_{uE}$ in Yu *et al.* [6]), $\hat{H}R_{uE} = \exp(\hat{\alpha})$ is consistent for the baseline hazard ratio $\exp(\alpha)$.

Therefore, the PERR-ALT estimate $\hat{H}R_{PERR-ALT} = \hat{H}R_E / \hat{H}R_{uE}$ is consistent for the true treatment effect $\exp(\theta)$.

Note that the proportional baseline hazards assumption, $\log\{h_{0s}(t)/h_{0p}(t)\} = \alpha(t) = \alpha$, is not compulsory for the pairwise Cox method. The pairwise Cox likelihood is flexible. We can for example assume the log baseline hazard ratio is linear associated with time and thus specify $\alpha(t) = \alpha_0 + \alpha_1 t$. The pairwise likelihood is then

$$L(\theta, \alpha_0, \alpha_1) = \prod_i^n \left(\frac{1}{1 + p_i e^{\theta x_i + \alpha_0 + \alpha_1 t_{pi}}} \right)^{\delta_{pi}} \left(\frac{e^{\theta x_i + \alpha_0 + \alpha_1 t_{si}}}{e^{\theta x_i + \alpha_0 + \alpha_1 t_{si}} + s_i} \right)^{\delta_{si}}.$$

For the example of diabetic retinopathy, the likelihood function is

$$\prod_i^n \left(\frac{e^{\theta_p X_{ip}}}{e^{\theta_p X_{ip}} + P_i e^{\theta_s X_{is} + \alpha}} \right)^{\Delta_{pi}} \left(\frac{e^{\theta_s X_{is} + \alpha}}{e^{\theta_s X_{is} + \alpha} + S_i e^{\theta_p X_{ip}}} \right)^{\Delta_{si}}. \tag{16}$$

Since there is no biological differences between the left and right eyes, we can assume a common baseline hazard ($h_{0p}(t) = h_{0s}(t)$) and a common treatment effect on both eyes ($\theta_p = \theta_s = \theta$). Therefore, $\alpha = \log(h_{0s}(t)/h_{0p}) = 0$ and the likelihood becomes

$$\prod_i^n \left(\frac{1}{1 + P_i e^{\theta X_i}} \right)^{\Delta_{pi}} \left(\frac{e^{\theta X_i}}{e^{\theta X_i} + S_i} \right)^{\Delta_{si}}, \tag{17}$$

where $X_i = X_{is} - X_{ip}$. We can note that only patients receiving treatment on one of the eyes (i.e. $X_i = 1$ or -1) will contribute information to the likelihood (17).

5.1. Simulations

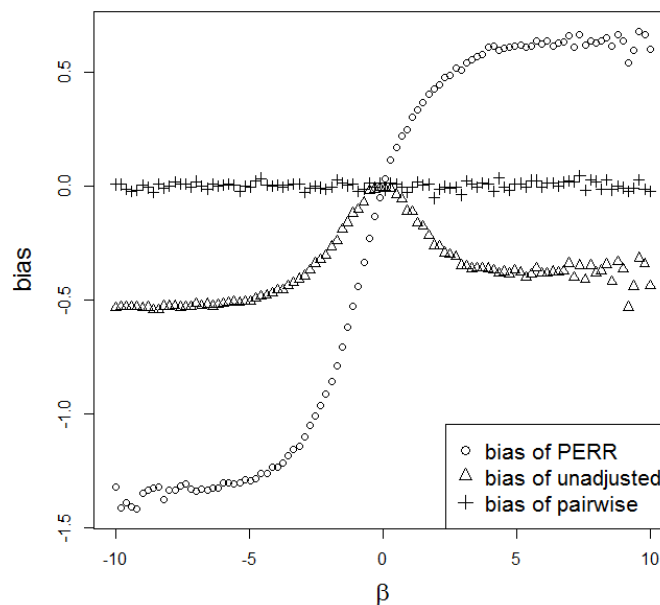


Figure 4. Simulation study of the bias of the unadjusted, PERR and pairwise methods in the case of a time-varying baseline hazard. Graphs based on simulated dataset of $n = 100,000$ paired event times. A Weibull baseline hazard $h_{0p} = 2t$ was used. The main effect of treatment was $\theta = 1$ and the period effect was $\alpha = -1$.

Figure 4 compares the biases of the unadjusted, PERR and pairwise methods in the case of a time-varying baseline hazard. We generated the prior and study event times by the Weibull models with $h_{0p}(t) = 2t, \theta = 1, \alpha = -1, \beta \in (-10, 10), X \sim B(1, 0.5), C \sim B(1, 0.3 + 0.4X)$ and both the prior and study data were 50% censored. The simulation results are similar to those in Figure 3 (c) and show that the pairwise method is consistent.

Table 1 gives some examples of the variances for the unadjusted, PERR and pairwise estimates under different censoring proportions for the prior and study periods. We generated t_p and t_s from (1) and (2) respectively with $h_{0p}(t) = \theta = \beta = 1, h_{0s}(t) = \exp(-0.5), X \sim Bin(1, 0.5)$ and $C|X \sim Bin(1, 0.3 + 0.4X)$. We set censoring proportions for prior and study periods as (10%, 50%, 90%). The average variance for each of the three methods and each value of the censoring proportion were calculated over 1000 simulation replications. For each replication, we set the number of events as 200 and the sample size as $200 / \{1 - \min(CP_p, CP_s)\}$, where CP_p and CP_s are the censoring proportions for prior and study periods respectively. Table 1 shows that generally the variances of the PERR and pairwise estimates are larger than those for the

Proportion of censoring in prior period	10%	10%	10%	50%	50%	50%	90%	90%	90%
Proportion of censoring in study period	10%	50%	90%	10%	50%	90%	10%	50%	90%
average variance for unadjusted estimate	0.03	0.06	0.53	0.03	0.04	0.24	0.04	0.03	0.02
average variance for PERR estimate	0.05	0.08	0.59	0.07	0.05	0.23	0.58	0.17	0.05
average variance for pairwise estimate	0.08	0.14	0.91	0.13	0.07	0.41	0.61	0.38	0.08
square root of the ratio between average variances for unadjusted and pairwise estimates	1.53	1.54	1.31	2.13	1.44	1.30	4.16	3.49	1.74

Table 1. Examples of the variances for the unadjusted, PERR and pairwise estimates under different censoring proportions in the prior and study periods

unadjusted estimates. The square root of the ratio between average variances shows how much the confidence interval of the pairwise method will be wider than that of the unadjusted method. It can be seen that the square root of variance ratio will generally increase with $CP_s - CP_p$.

5.2. Treatment of ties and left truncation

In the special case where the prior and study times are tied (ie $t_{pi}^* = t_{si}^*$), we assume the equality occurs because the measurement of event times is not accurate enough. The exact likelihood of the tied pair is hence the sum of the likelihoods in (13) if $t_{pi}^* > t_{si}^*$ and if $t_{pi}^* < t_{si}^*$. The sum is

$$\text{Exact likelihood} = L_i(\theta, \alpha)_{\text{if } t_{pi}^* > t_{si}^*} + L_i(\theta, \alpha)_{\text{if } t_{pi}^* < t_{si}^*} = \frac{1}{1 + e^{\theta X_i + \alpha}} + \frac{e^{\theta X_i + \alpha}}{e^{\theta X_i + \alpha} + 1} = 1.$$

Therefore, the tied pairs make no contribution to the full likelihood and should be removed before using the pairwise Cox method. If the tied pairs are not removed, the partial likelihood in (15) will be

$$\frac{e^{\theta X_i + \alpha}}{(1 + e^{\theta X_i + \alpha})^2} \neq 1, \tag{18}$$

which is not identical to the exact likelihood and thus estimates of θ and α will be biased

Another issue that will be relevant in some studies is that of left truncation. For example, this can arise when the time at which a patient first become at risk of a prior event is unknown. For left truncation, there are two options. One is to redefine the events with different starting points to avoid the problem. For example, it may be reasonable to define the prior origin as the date of the patient's first medical record after reaching a particular age (or the date of diagnosis), even though the patient may have been at risk before that time. However, this option is not always feasible and may change the interpretation of parameters, potentially leading to violation of the assumptions of the hazard models. The other option is to impose a distributional assumption on the time of left truncation [10]. The difficulty is that it is necessary to have reliable information on how to specify the distribution. More details can be found in Section 4.5 of the book by Cook and Lawless [10].

5.3. Applications in crossover trials

As Fig 3 (a) and (b) show, the unadjusted Cox model will give biased estimates of treatment effects, even in randomized trials, if needed covariates are omitted. As it is never possible to capture all component causes [18], this potential for bias is likely to be present to some degree in all randomised controlled trials. In particular, it can have practical consequences for effectiveness estimates in smaller trials.

The pairwise Cox method can be applied with the crossover trial design to remove the bias from hidden covariates. As we have shown, even if there are infinitely many background covariates and it is impossible to measure all of them, as long as their values and effects are constant in (1) and (2), the relevant terms will only rescale the individual-specific baseline hazard $h_{0pi}(t)$ and will be eliminated by the pairwise partial likelihood (15). However, as shown in Table 1, there is a trade-off between bias and variance when using the pairwise method with crossover trials.

6. Limitations of PERR adjustment

As with other methods for addressing hidden confounding, the validity of PERR adjustment based on the pairwise Cox likelihood relies on assumptions about the underlying causal model linking treatment (exposure) to outcomes. To

assist applied researchers in deciding when the PERR method is likely to be of use in practice, we now review the key assumptions of the method and consider robustness of the method to violations of these assumptions:

Assumption 1: Models (1) and (2) are correctly specified.

If the models are misspecified, for example if there is an interaction between treatment and the unmeasured confounder(s) or the value and effect of the unmeasured confounder are time-varying, the pairwise method will be biased. Uddin *et al.* [19] studied the performance of the PERR estimator when the effect of unmeasured confounder varies between periods, and showed that the PERR method gives biased estimates of the rate ratio but the bias is generally smaller than for conventional analyses. Figure 5 shows an example of the biases for the pairwise Cox likelihood method, the original PERR adjustment method and the standard Cox regression model for different degrees of confounder-treatment interaction. The paired event times t_p and t_s were generated from (1) and $h_{0p}(t) \exp(\theta X_i + \beta C_i + \alpha + \beta_1 X_i C_i)$ with $h_{0p}(t) = \theta = \alpha = \beta = 1$, $X \sim B(1, 0.5)$ and $C|X \sim B(1, 0.3 + 0.4X)$. The sample size was 100,000 and there was no censorship. The bias was estimated for different values of the interaction effect β_1 .

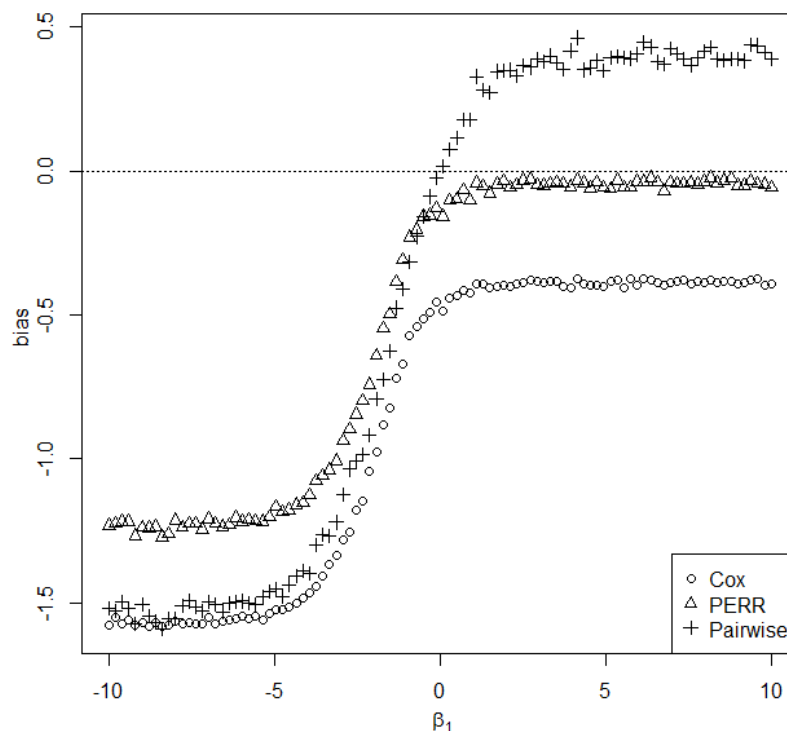


Figure 5. Simulation study of the bias of the pairwise Cox likelihood method and the original PERR adjustment method in the presence of an interaction between hidden confounder and treatment. Graphs based on simulated dataset of $n = 100,000$ paired event times. An exponential baseline hazard was used with a rate parameter of 1. The main effect of treatment was $\theta = 1$ and the period effect was $\alpha = 1$. β_1 denotes the size of the interaction effect.

It can be seen that when the true hazard model contains an interaction, the bias curve for the pairwise Cox method has a similar sigmoid shape to the curves for the Cox model (i.e. $h_{0s}^*(t)e^{\theta_s^* X}$) and the original PERR method. When the magnitude of the interaction effect is less than that of the main effect of treatment (ie $|\beta_1| < 1$, the pairwise method reduces the bias compared to the Cox model. However, the figure illustrates that there may be regions, corresponding to more extreme degrees of model misspecification, where the Cox model outperforms PERR adjustment (eg when $\beta_1 < -3$ in this example).

Assumption 2: Prior event occurrence should not influence likelihood of future treatment /exposure.

Gallagher *et al.* [20] and Uddin *et al.* [19] showed that the original PERR is biased when the likelihood of treatment is affected by the prior event occurrence. We assessed the performance of the pairwise method under this problem and found that it also gave biased estimates of treatment effects. An example of this bias is illustrated in Fig 6. The paired times t_p and t_s were generated from (1) and (2) with $h_{0p}(t) = \theta = \alpha = \beta = 1$. The sample size was 100,000. The censoring mechanism was uniform and both the prior and study event data were 50% censored. We generated

$$C \sim B(1, 0.5) \quad \text{and} \quad X_i \sim B(1, \text{expit}(C_i - \Delta_{pi})) \quad (19)$$

so that the treatment variable was associated with the unmeasured confounder and the prior censoring indicator δ_{pi} . Understanding the mechanism underlying the bias and developing possible solutions is an interesting open question in its own right.

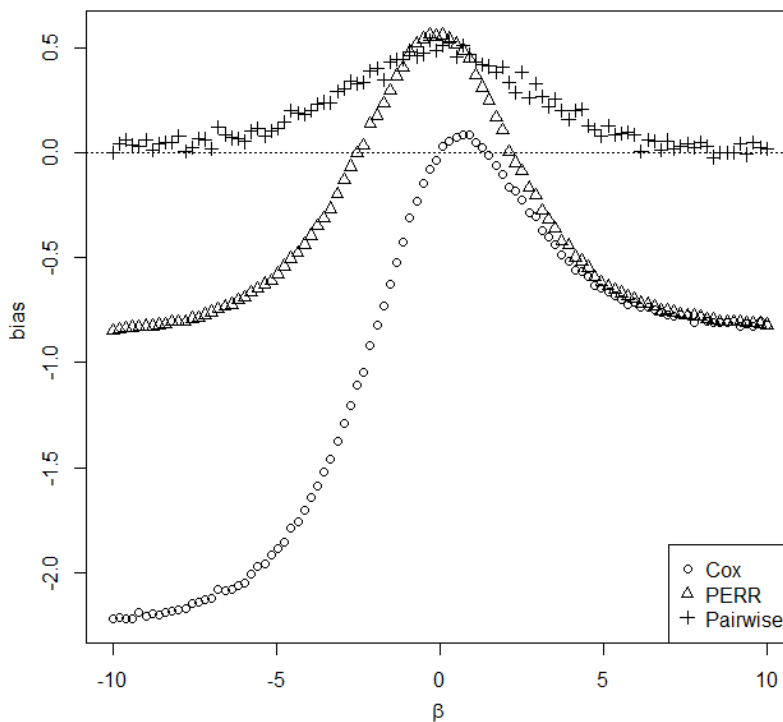


Figure 6. Simulation study of the bias of the pairwise Cox likelihood method and the original PERR adjustment method when prior events influence the likelihood of future receipt of treatment. Graphs based on simulated dataset of $n = 100,000$ paired times t_p and t_s generated from (1) and (2). An exponential baseline hazard was used with a rate parameter of 1. The censoring mechanism was uniform and both the prior and study event data were 50% censored. The main effect of treatment was $\theta = 1$ and the period effect was $\alpha = 1$. C and X_i were generated from (19) so that the treatment variable was associated with the unmeasured confounder and the prior censoring indicator δ_{pi} .

Remarks on the problem of differential case fatality: Gallagher *et al.* [20] also showed that the original PERR method is biased in the case of differential case fatality (i.e. where high risk patients are more likely to die before reaching exposure).

We conducted a simulation study to further explore the performance of the PERR and pairwise methods in the presence of differential case fatality (see Web Appendix B). The underlying bias of the PERR estimator (as shown in Section 3) was shown to change when case fatality was differential (both increases or decreases in bias were possible). However, we found that the pairwise method was unbiased in this situation. The reason is that the pairwise method compares the outcomes of the same patient before and after study and thus differential case fatality will not introduce bias. In contrast, for the PERR method, the distribution of the unmeasured confounder is changed when case fatality is differential. For example, suppose $C \sim \text{Bin}(1, 0.5)$ at the beginning of the prior period and the groups with $c = 0$ (low risk) and $c = 1$ (high risk) have differential case fatality of 10% and 50% respectively, then the distribution of C for the subjects in the PERR analysis becomes $\text{Bin}(1, \frac{0.5 \times 0.5}{0.5 \times 0.9 + 0.5 \times 0.5})$. As a result, differential case fatality changes the distribution of the unmeasured confounder and thus changes the bias of PERR. In fact, the bias of PERR in this case is the same as the bias in the absence of differential case fatality with $C \sim \text{Bin}(1, \frac{0.5 \times 0.5}{0.5 \times 0.9 + 0.5 \times 0.5})$. This change in the bias due to differential case fatality corresponds to the difference between the PERR biases for the two unmeasured confounding distributions (see Web Appendix B). We note that in the extreme case that all the subjects with $c = 1$ die before exposure and all the subjects with $c = 0$ survive to exposure, the PERR method (as well as the unadjusted Cox model) will be unbiased because all the subjects in the analysis have the same value of $c = 0$.

7. More general likelihoods

7.1. Time-varying covariates

The pairwise likelihood (15) can be extended to allow flexible modeling in more general situations. For example, in practice, we may have some covariates measured in the prior period, and their values and effects might change over time. Adopting a commonly-used model for time-varying covariates, the hazards can be written as

$$h_{pi}(t) = I(t \leq T_{pi})h_{0p}(t) \exp\{\theta(t)x_i(t) + \beta c_i\} \tag{20}$$

$$h_{si}(t) = I(t \leq T_{si})h_{0p}(t) \exp\{\theta(t + \tau_{pi})x_i(t + \tau_{pi}) + \beta c_i + \alpha\}, \tag{21}$$

where τ_{pi} is the time span between the starting points of the prior and study periods for the i th individual. The models (20) and (21) assume that the hazards at time $t_{pi} = t_{si} = t$ are only affected by the current value of the covariate and its effect at the time $t_{pi} = t_{si} = t$. Unbiased estimates of the model parameters can be obtained from

$$L(\theta(t), \alpha) = \prod_i^n \left\{ \frac{e^{\theta(t_{pi})x_i(t_{pi})}}{e^{\theta(t_{pi})x_i(t_{pi})} + p_i e^{\theta(t_{pi} + \tau_{pi})x_i(t_{pi} + \tau_{pi}) + \alpha}} \right\}^{\delta_{pi}} \left\{ \frac{e^{\theta(t_{si} + \tau_{pi})x_i(t_{si} + \tau_{pi}) + \alpha}}{e^{\theta(t_{si} + \tau_{pi})x_i(t_{si} + \tau_{pi}) + \alpha} + s_i e^{\theta(t_{si})x_i(t_{si})}} \right\}^{\delta_{si}} \tag{22}$$

The simple hazard models (10) and (11) and the likelihood (15) can be regarded as the special cases of (20), (21) and (22) respectively when

$$\theta(t)x_i(t) = \begin{cases} 0, & \text{for the hazard of prior event} \\ \theta x_i, & \text{for the hazard of study event} \end{cases} .$$

7.1.1. *An example with period-specific treatment effects* Suppose the time-varying covariates and effects take the values

$$\theta(t)X_i(t) = \begin{cases} \theta_p x_{pi}, & \text{for the hazard of prior event} \\ \theta_s x_{si}, & \text{for the hazard of study event} \end{cases} .$$

Such a model could be used for considering the effects of treatments, such as statins, for which effects on outcomes or side-effects may not become apparent until after an initial period on treatment. The corresponding hazards are

$$h_{pi}(t) = I(t \leq T_{pi})h_{0p}(t) \exp(\theta_p x_{pi} + \beta c_i) \tag{23}$$

$$h_{si}(t) = I(t \leq T_{si})h_{0p}(t) \exp(\theta_s x_{si} + \beta c_i + \alpha), \tag{24}$$

where θ_s is the parameter of interest. The likelihood is then

$$L(\theta_s, \theta_p, \alpha) = \prod_i^n \left\{ \frac{e^{\theta_p x_{pi}}}{e^{\theta_p x_{pi}} + p_i e^{\theta_s x_{si} + \alpha}} \right\}^{\delta_{pi}} \left\{ \frac{e^{\theta_s x_{si} + \alpha}}{e^{\theta_s x_{si} + \alpha} + s_i e^{\theta_p x_{pi}}} \right\}^{\delta_{si}} \tag{25}$$

A simulation study was conducted to compare the performance of the likelihood (25) with the unadjusted hazard model $h(t_{si}) = h_{0s}^*(t_{si}) \exp(\theta_s^* x_{si})$. We generated 100,000 paired event times (t_{pi}, t_{si}) from (23) and (24) with $h_{0p}(t) = \theta_s = 1$, $\theta_p = \alpha = -1$, $\beta = (-10, 10)$, $X_p \sim \text{Bin}(1, 0.5)$, $X_s \sim \text{Bin}(1, 0.7 - 0.4X_p)$ and $C \sim \text{Bin}(1, 0.3 + 0.4X_s)$. The censoring mechanism was uniform and both the prior and study event data were 50% censored. The biases of the pairwise method were calculated by $\hat{\theta}_s - \theta_s$, where $\hat{\theta}_s$ was obtained from (25). The bias of the unadjusted hazard ratio was $\hat{\theta}_s^* - \theta_s$.

The results presented in Figure 7 show that the pairwise method eliminates the bias in the unadjusted model. The PERR method was not considered here because it is not applicable in this case.

7.2. More than two periods

The likelihood can also be extended to the case of more than two periods. Suppose the hazard of the i th individual of experiencing the k th event is

$$h_{ik}(t) = I(t \leq T_{ik})h_0(t) \exp\{\beta c_i + \theta(t + \tau_{ik})x_i(t + \tau_{ik}) + \alpha_k\}, \quad k = 1, 2, \dots$$

where $h_0(t)$ is the baseline hazard of the first event, τ_{ik} is the time span between the starting points of the first period and the k th period and α_k is the log ratio of baseline hazards between the k th and the first events with $\alpha_1 = 0$.

For a sample of n individuals and the i th individual with m_i events, the partial likelihood is

$$L = \prod_{i=1}^n \prod_{k=1}^{m_i} \left(\frac{e^{\theta(t_{ik} + \tau_{ik})x_i(t_{ik} + \tau_{ik}) + \alpha_k}}{\sum_{j=1}^{m_i} I(t_{ik} \leq t_{ij}) e^{\theta(t_{ik} + \tau_{ij})x_i(t_{ik} + \tau_{ij}) + \alpha_j}} \right)^{\delta_{ik}} . \tag{26}$$

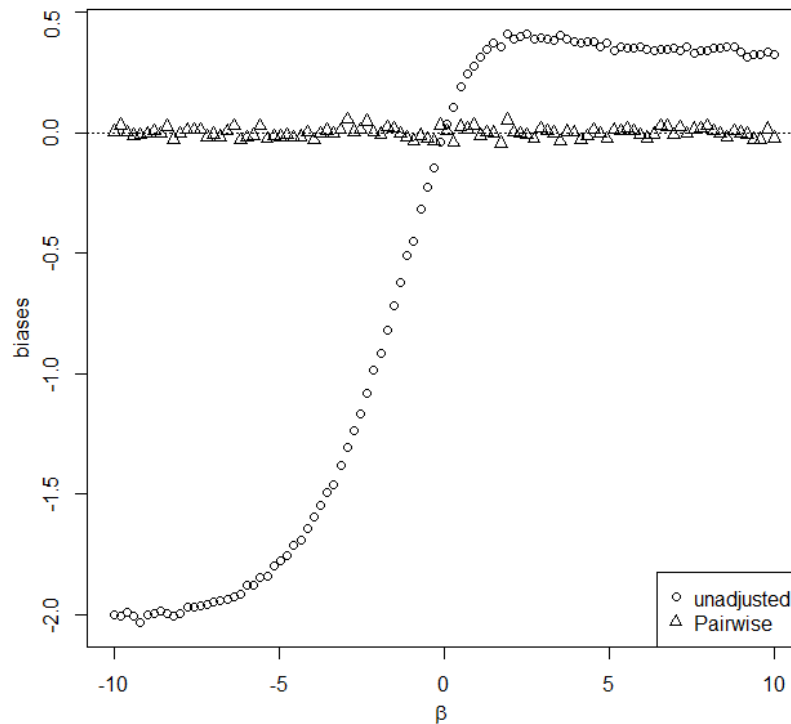


Figure 7. Simulation study of the pairwise method (25) and the unadjusted hazard model $h(t_{si}) = h_{0s}^*(t_{si}) \exp(\theta_s^* x_{si})$ in the presence of time varying covariates and effects. Graphs based on simulated dataset with $n = 100,000$ and log-hazard ratio of treatment $\theta_s = 1$ and $\theta_p = -1$. An exponential baseline hazard was used with a rate parameter of $h_0(t) = 1$. The censoring rate is 50% for both prior and study periods.

In the case of ties, (26) is a Breslow approximation to the exact likelihood of dealing tied times, which should work well if the ratio between the number of ties and m_i is small. However, if this condition is not satisfied, for example $m_i = 2$, the approximation can be poor as shown in (18). In this situation, the exact likelihood or Efrons approximation would be suggested.

8. Application: Comparative effectiveness of enzyme replacement therapies for patients with Fabry disease

8.1. Background and model specification

We consider application of the pairwise Cox approach to data from a longitudinal cohort study of enzyme replacement therapy for Fabry's disease. Fabry disease is a rare, inherited metabolic disorder resulting from absolute or partial deficiency of the enzyme α -galactosidase A and the progressive accumulation of undigested macromolecules in cells throughout the body, particularly in the heart, kidneys and nerve tissue [21]. There are two forms of enzyme replacement products approved for use in the US and Europe: agalsidase alfa and agalsidase beta. Studying the comparative efficacy and safety of these two therapies is of current clinical relevance as there is a tendency for the drug of choice to vary between treatment centres, often for historical reasons [21]. The rarity and severity of this condition has resulted in a lack of adequately powered randomised trials for making such comparisons, and so researchers have instead had to rely on observational studies, with the attendant risk of bias from hidden confounding. We fitted the pairwise Cox model to estimate comparative effectiveness of the two forms of ERT using data from the National Collaborative Study for Lysosomal Storage Disorders (NCS-LSD), a longitudinal cohort study collecting both prospective and historical data [22]. Here the focus is on illustrating usage of the methods rather than drawing firm conclusions about the treatments involved and so we refer to the treatments simply as therapy A and therapy B. Our analyses necessarily involve simplification of the issues that would be involved in substantive analyses of the source data and so the results we present should not be used to draw inferences for clinical practice.

Overall, a total of 211 adults with Fabry disease were being treated with ERT on recruitment to the NCS-LSD study. For this illustrative example, the outcome of interest was estimated GFR (eGFR), a measure of kidney function, and we analysed data for the 45 patients with at least two measures of eGFR, both before and after starting therapy. The data for

the 45 patients is provided in Web Appendix C and is available to download as a csv file. Of these patients, 26 received therapy A and 19 were on therapy B. Two potential confounding variables, age and gender, were available for this analysis. The patients on therapy A were younger and more likely to be female (mean age of starting therapy = 41.0 years; 35% female) than patients on therapy B (mean age of starting therapy = 48.1 years; 5% female). We were unable to consider other potential confounding variables, such as genotype and disease severity, due to a lack of consistency in the way that they were assessed and recorded. The aim of the pairwise Cox analysis was to account for these unmeasured differences. Given the modest sample size, the analysis can be seen as a pilot study to inform the design of a more extensive follow-up study.

To facilitate application of the pairwise Cox method, the multiple renal function measures were converted into time-to-event data by defining the prior and study events as the first times when eGFRs were greater than those in the first visits before treatment and the visits at or following treatment respectively (in the absence of measurement error, increases in eGFR correspond to improvements in kidney function). According to these definitions of the prior and study events, the start of the prior period was the first visit recorded in the database before treatment, and the corresponding start of the study period was the visit at or following the time of initiating ERT therapy. The observed prior and study event times, T_p and T_s , were measured from the start of the prior and study periods to the prior and study events (or to the times of receiving therapies and the end of study) respectively. The prior and study periods are terminated at T_p and T_s respectively. We note that defining the study origin as the visit at or following the point of starting ERT therapy was a consequence of the definition of study event. This helped simplify the problem for the purposes of this illustrative example. In practice, it is inappropriate to set the time of receiving treatment as the study origin, if the time of receiving treatment is not the time when an individual starts to be under the risk of the defined study event.

Let C_i be an unmeasured confounder, X_{i5} be the indicator of gender (0 = male, 1 = female), X_{i3} and X_{i4} be the ages in years at the beginning of the prior and study periods respectively, and X_{i1} and X_{i2} be the treatment indicators (no treatment: $X_{i1} = X_{i2} = 0$; Therapy A: $X_{i1} = 1, X_{i2} = 0$; Therapy B: $X_{i1} = 0, X_{i2} = 1$). The hazard functions of the prior and study events for the i th patient are assumed to be

$$h_{pi}(t) = I(t \leq T_{pi})h_{0p}(t)e^{\beta C_i + \theta_5 X_{i5} + \theta_3(X_{i3} + t)} \tag{27}$$

$$h_{si}(t) = I(t \leq T_{si})h_{0s}(t)e^{\beta C_i + \theta_5 X_{i5} + \theta_3(X_{i4} + t) + \theta_1 X_{i1} + \theta_2 X_{i2}} \tag{28}$$

$$= I(t \leq T_{si})h_{0pi}e^{(\theta_1 - \theta_2)X_{i1} + \alpha + \theta_2 + \theta_3(X_{i4} - X_{i3})} \tag{29}$$

where $\beta, \theta_5, \theta_3, \theta_1$, and θ_2 are the coefficients, and $\alpha = \log \{h_{0s}(t)/h_{0p}(t)\}$.

Note that in the hazard models (27)-(29), age is a time-varying covariate measured in both the prior and study periods and taking the values $X_{i3} + t$ and $X_{i4} + t$ at the times of prior and study events respectively. Since the prior and study events are of the same nature, i.e. renal function increases, the coefficient θ_5 (or θ_3) is assumed to be the same in the hazards (27) and (28). Otherwise, we would need to specify different θ_5 (or θ_3), such as using θ_{5p} in (27) and θ_{5s} in (28). As all the patients received one of the two types of ERT therapy and no one was untreated, $X_{i2} = 1 - X_{i1}$ and we are unable to estimate θ_1 and θ_2 . But the difference between Therapy A and Therapy B, $\theta_1 - \theta_2$, can be estimated.

8.2. Estimates under the pairwise, unadjusted Cox models and PERR

The pairwise likelihood is

$$L(\theta_1 - \theta_2, \alpha + \theta_2, \theta_3) = \prod_i^n \left\{ \frac{1}{1 + P_i e^{(\theta_1 - \theta_2)X_{i1} + \alpha + \theta_2 + \theta_3(X_{i4} - X_{i3})}} \right\}^{\Delta_{pi}} \left\{ \frac{e^{(\theta_1 - \theta_2)X_{i1} + \alpha + \theta_2 + \theta_3(X_{i4} - X_{i3})}}{e^{(\theta_1 - \theta_2)X_{i1} + \alpha + \theta_2 + \theta_3(X_{i4} - X_{i3})} + S_i} \right\}^{\Delta_{si}} \tag{30}$$

The likelihood of the Cox model fitted to the study data without adjustment for C_i is

$$L(\theta_5, \theta_3, (\theta_1 - \theta_2)) = \prod_i^n \left[\frac{e^{\theta_5 X_{i5} + \theta_3(X_{i4} + T_{si}) + (\theta_1 - \theta_2)X_{i1}}}{\sum_{j: T_{sj} \geq T_{si}} e^{\theta_5 X_{j5} + \theta_3(X_{j4} + T_{si}) + (\theta_1 - \theta_2)X_{j1}}} \right]^{\Delta_{si}} \tag{31}$$

The likelihood of the Cox model fitted to the prior data without adjustment for C_i is

$$L(\theta_5, \theta_3, (\theta_1 - \theta_2)) = \prod_i^n \left[\frac{e^{\theta_5 X_{i5} + \theta_3(X_{i3} + T_{pi}) + (\theta_1 - \theta_2)X_{i1}}}{\sum_{j: T_{pj} \geq T_{pi}} e^{\theta_5 X_{j5} + \theta_3(X_{j3} + T_{pi}) + (\theta_1 - \theta_2)X_{j1}}} \right]^{\Delta_{pi}} \tag{32}$$

The estimates of $e^{\theta_1 - \theta_2}$ under the pairwise model (30) and the unadjusted Cox models (31) and (32) are presented in the table below with 95% confidence intervals (CIs) obtained from the observed information. The R code for the

pairwise model (30) is provided in Appendix B. The PERR estimate is $1.25/0.93 = 1.34$ and the 95% CI was obtained by the bootstrap method. The estimates for the pairwise, unadjusted and PERR methods are consistently above 1, but the differences are not significant at the 5% level. Use of the pairwise method has led to an increased point estimate of 1.55, suggesting there may be the potential to uncover genuine differences between therapies in a larger study.

8.3. Using the prior data to detect unmeasured confounding

To make use of the method in Section 4, we first test whether the assumptions $(X_5, X_3) \perp X_1$ and $S_p^+(t|X_1, X_5, X_3) = S_p^+(t|X_5, X_3)$ hold. We assessed these by fitting a logistic regression to X_1 on X_3 and X_5 , and a Cox model to t_p on X_1, X_3, X_5 with the censoring indicator swapped. There was no significant evidence against either of these assumptions (the p-values for likelihood ratio tests are 0.14 and 0.88 respectively). Under these assumptions, the P-value of the unadjusted Cox model (32) provides a test for hidden confounding. In this example, there was no significant evidence of hidden confounding ($p = 0.86$), perhaps due to lack of power to detect an effect in this small sample. However, as mentioned in Section 4, the unadjusted Cox model and the PERR will underestimate the treatment effect, even if the unmeasured covariate is not a confounder. The pairwise method is still recommended in the absence of significant evidence of hidden confounding.

9. Discussion

This paper is concerned with identifying and reducing the impact of hidden confounding in observational studies of treatment effectiveness. Building on the work of Tannen and colleagues, we set out a general framework for quasi-experimental analysis using the PERR approach, and derived a flexible pairwise Cox likelihood function that can be used to estimate unbiased treatment estimates, under appropriate assumptions. The pairwise likelihood was used to demonstrate the consistency of the simple and convenient PERR-ALT estimator introduced by Yu *et al.* [6]. We showed how to estimate standard errors and confidence intervals for treatment effect estimates based on the pairwise model, and provided R code to illustrate how to implement the method. A simple test to detect unmeasured confounding was also developed based on data from a prior period.

Much of the previous work on PERR adjustment has focused on analysis of electronic medical record databases. While the growth of large-scale data registries makes this a particularly important application, the approach also has potential value in population based research studies. This was illustrated through an example in which the pairwise method was applied to a longitudinal cohort study combining prospective data collection with retrospective use of routine health records.

Designing future quasi-experimental studies that exploit the pairwise Cox method requires an understanding of the power of the method. We showed that the variance for the pairwise method is usually larger than that for a conventional Cox regression analysis. In the era of large electronic medical record databases, the additional variance is unlikely to be overly burdensome unless the focus is on narrow segments of the population.

One of the strengths of the pairwise Cox likelihood approach is its inherent flexibility. Like the conventional Cox model, the pairwise method can easily be extended to accommodate time-varying treatment covariates and coefficients. Furthermore, the form of $\alpha(t)$, corresponding to the period effect, can be chosen by comparing models with different forms of $\alpha(t)$, making it possible to relax the proportional baseline hazards assumption. In addition, we leave the user free to define the prior and study events for different research problems, without the need for the prior and study periods to be successive.

Several limitations of the PERR approach need to be considered by investigators conducting observational research. Yu *et al.* [6] highlight the fundamental requirement for prior events. Consequently, in its current form, the pairwise Cox likelihood cannot be used for evaluating treatment effects on mortality or other terminal events. The method also relies on strong assumptions about model specification including the absence of time-varying hidden confounders and confounder by treatment interaction. In practice, the method may be robust to realistic levels of departure from these assumptions [6] but, as we have demonstrated, in extreme cases the method can be as biased, or even more biased, than conventional approaches.

An additional challenge to the validity of the method arises when treatment allocation is associated with prior events, such as if a patient is switched to a new treatment following failure of an existing treatment. Consistent with the findings of Gallagher *et al.* [20], our simulations highlighted that both pairwise and PERR methods are likely to be biased in this situation. These findings reinforce the need to consider the context of the problem carefully when assessing the suitability of the PERR approach. In its current form, the pairwise method may be most useful where the decision to start treatment is not likely to be related to the risk of future events, for example in drug safety monitoring or evaluations of national vaccination programmes.

A number of other biostatistical methods for tackling hidden confounding have been proposed. Common approaches that provide estimates of causal effects include propensity scoring combined with regression calibration [23], instrumental variables [24] and regression discontinuity designs [25]. In practice, no one method is likely to be best in all problems and it is essential for investigators to carefully assess the potential biases in each proposed study, where possible tailoring the methods or combination of methods to address these biases [2]. Tannen *et al.* [5] provided preliminary evidence that the PERR approach can produce reliable results when replicating outcomes of cardiovascular trials, but further empirical studies are needed to establish the validity of the method for use in other clinical problems, and to determine the strengths and limitations of the PERR approach relative to other methods. Where possible this should include proof of concept studies to replicate results of randomized trials as well as clinically informed simulation studies.

Like the instrumental variable method which requires a suitable instrument and propensity score calibration which needs a suitable valuation study, the pairwise method can be only used when suitable data are available. For example, data on prior events will not be available in some problems, such as a primary prevention study of statin therapy for patients with no prior history of cardiovascular disease. However, the pairwise method is likely to become more accessible with the increasing availability of large electronic medical record databases, because such datasets can often provide the necessary volumes of longitudinal data (albeit of variable quality and completeness) before and after patients receive treatments.

In conclusion, the PERR and pairwise Cox methods offer a promising approach to addressing biases that can arise in observational studies due to lack of randomisation, and through further development could become a highly cost-effective way of using established datasets to answer questions about treatment effectiveness in clinical practice. The flexibility of the pairwise Cox likelihood offers a basis for generalizing the method but widespread adoption of the approach will require further progress in addressing the challenges of dealing with prior events that influence treatment, terminal events and the presence of time-varying confounding.

Acknowledgement

We thank Stuart Logan for providing access to the data on lysosomal storage disorders and for advice on application of the methods. This research was funded by the Medical Research Council [grant number G0902158]. William Henley received additional support from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

1. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; **312**: 1215-1218.
2. Alemayehu D, Alvir J, Jones B, Willke R. Statistical Issues with the Analysis of Non-Randomized Studies in comparative Effectiveness Research. *J Manag care Pharm* 2011; **17**: S22-S26.
3. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. *BMJ* 2009; **338**: b81.
4. Weiner MG, Xie D, Tannen RL. Replication of the Scandinavian Simvastatin Survival Study using a primary care medical record database prompted exploration of a new method to address unmeasured confounding. *Pharmacoepidemiol Drug Saf* 2008; **17**: 661-670.
5. Tannen RL, Weiner MG, Xie D. Replicated studies of two randomized trials of angiotensin converting enzyme inhibitors: further empiric validation of the prior event rate ratio to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf* 2009; **17**: 671-85.
6. Yu M, Xie D, Wang X, Weiner MG, Tannen RL. Prior event rate ratio (PERR) adjustment: Numerical studies of a statistical method to address unrecognized confounding in observational studies. *Pharmacoepidemiol Drug Saf* 2012; **21**: 60-68.
7. Brophy S, Jones KH, Rahman MA, Zhou S-M, John A, Atkinson MD, et al. Incidence of Campylobacter and Salmonella infections following first prescription for PPI a cohort study using routine data. *Am J Gastroenterol* 2013; **108**:1094-1100.
8. Tannen R, Xie D, Wang X, Yu M, Weiner MG. A new Comparative Effectiveness assessment strategy using the THIN database: comparison of the cardiac complications of pioglitazone and rosiglitazone. *Pharmacoepidemiol Drug Saf* 2013; **22**: 86-97.
9. Nichol KL, Nordin JD, Nelson DB, Mullooly JP, Hak E. Effectiveness of influenza vaccine in the community-dwelling elderly. *New England Journal of Medicine* 2007; **357**: 1373-1381.
10. Cook RJ, Lawless JF. The statistical analysis of recurrent events. *Springer Science & Business Media*, 2007.
11. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**: 373-379.
12. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**: 1065-1073
13. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Annals of Statistics* 1982; **10**: 1100-1120.
14. Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the Cox model with omitted covariates. *Biometrics* 2013; **69**: 850-860
15. Lin, DY, Wei, LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistics* 1989; **84**: 207-224.
16. Hjort, NL. On inference in parametric survival data models. *Int. Statist. Rev.* 1992; **60**: 335-387.
17. Gross TS, Huber C. Matched pair experiments: Cox and maximum likelihood estimation. *Scand J Stat* 1987; **14**: 27-41.

18. Gustafson P, McCandless LC, Levy AR, Richardson S. Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders. *Biometrics* 2010; **66**: 1129–1137.
19. Uddin MJ, Groenwold RH, van Staa TP, de Boer A, Belitser SV, Hoes AW, Roes KC, Klungel OH. Performance of prior event rate ratio adjustment method in pharmacoepidemiology: a simulation study. *Pharmacoepidemiology and drug safety* 2015; **24**: 468–477.
20. Gallagher A, De Vries F, Van Staa T. Prior event rate ratio adjustment: A magic bullet or more of the same? *Pharmacoepidemiology and Drug Safety*. *Pharmacoepidemiol Drug Saf* 2009; **18**: S14–S15.
21. Wyatt K, Henley W, Anderson L, Nikolaou V, Stein K, Klinger L, Hughes D, Waldek S, Lachmann R, Mehta A, Vellodi A, Logan S. The effectiveness and cost-effectiveness of enzyme and substrate replacement therapies: a longitudinal cohort study of people with lysosomal storage disorders. *Health Technol Assess* 2012; **16**: 1–543.
22. Henley WE, Anderson LJ, Wyatt KM, Nikolaou V, Anderson R, Logan S. The NCS-LSD cohort study: a description of the methods and analyses used to assess the long-term effectiveness of enzyme replacement therapy and substrate reduction therapy in patients with lysosomal storage disorders. *Journal of Inherited Metabolic Disease* 2014; 1–6.
23. Strmer T, Schneeweiss S, Rothman KJ, Avorn J, Glynn RJ. Performance of propensity score calibration a simulation study. *American journal of epidemiology* 2007; **165**, 1110–1118.
24. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. *JASA* 1996; **91**, 444–455.
25. O’Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen, I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ*, 2014; **349**, g5293.
26. Johansen S. An extension of Cox’s regression model. *International Statistical Review* 1983; **51**: 165–174.

Model	HR $e^{\widehat{\theta}_1 - \theta_2}$	95%CI	P-value
Pairwise (30)	1.55	(0.70, 3.14)	0.28
Unadjust (31)	1.25	(0.59, 2.66)	0.57
Unadjust prior (32)	0.93	(0.42, 2.05)	0.86
PERR	1.34	(0.31, 5.74)	0.69

Table 2. Estimates of the hazard ratio for the difference between ERT therapies using the Pairwise Cox method, the unadjusted Cox model and the PERR method, for the example data on patients with Fabry disease.

A. Estimate of $h_{0pi}(\cdot)$

The log full likelihood is

$$l_{full} = \sum_i^n [\Delta_{pi} \log \{h_{0pi}(T_{pi})\} - H_{0pi}(T_{pi}) + \Delta_{si} [\log \{h_{0pi}(T_{si})\} + \theta X_i + \alpha] - H_{0pi}(T_{si})e^{\theta X_i + \alpha}].$$

Using the argument similar to Johansen [26], we assume the hazard is zero, $h_{0pi}(t) = 0$, except at event times and the cumulative hazard is the summation of the hazards, i.e.

$$h_{0pi}(T_{pi}) = \Delta_{pi}h_{pi} \quad H_{0pi}(T_{pi}) = \Delta_{pi}h_{pi} + S_i\Delta_{si}h_{si} \quad (33)$$

$$h_{0pi}(T_{si}) = \Delta_{si}h_{si} \quad H_{0pi}(T_{si}) = \Delta_{si}h_{si} + P_i\Delta_{pi}h_{pi}. \quad (34)$$

Then l_{full} becomes

$$l_{full} = \sum_i^n [\Delta_{pi} \log(\Delta_{pi}h_{pi}) - \Delta_{pi}h_{pi} - S_i\Delta_{si}h_{si} + \Delta_{si} \{\log(\Delta_{si}h_{si}) + \theta X_i + \alpha\} - (\Delta_{si}h_{si} + P_i\Delta_{pi}h_{pi})e^{\theta X_i + \alpha}]$$

Keeping θ and α fixed, the MLEs of h_{pi} and h_{si} are

$$\frac{\partial l_{full}}{\partial h_{pi}} = \frac{\Delta_{pi}}{h_{pi}} - \Delta_{pi} - P_i\Delta_{pi}e^{\theta X_i + \alpha} \implies h_{pi} = \frac{1}{1 + P_i e^{\theta X_i + \alpha}}$$

$$\frac{\partial l_{full}}{\partial h_{si}} = -S_i\Delta_{si} + \frac{\Delta_{si}}{h_{si}} - \Delta_{si}e^{\theta X_i + \alpha} \implies h_{si} = \frac{1}{S_i + e^{\theta X_i + \alpha}}$$

These together with (33) and (34) lead to the results (13) and (14).

B. R code for the example

The log-likelihood function of (30) is

```
l<-function(theta_3, paitheta_2, theta_12){
sum(-dp*log(1+p*exp(theta_12*x1+paitheta_2+theta_3*(x_4-x_3)))
+ds*(theta_12*x1+paitheta_2+theta_3*(x_4-x_3)
-log(exp(theta_12*x1+paitheta_2+theta_3*(x_4-x_3))+s))
}
```

where θ_3 is θ_3 , paitheta_2 is $\alpha + \theta_2$, θ_{12} is $\theta_1 - \theta_2$, dp is δ_{pi} , p is p_i , x1 is X_{i1} , $\text{x}_4 - \text{x}_3 = X_{i4} - X_{i3}$, $\text{s} = s_i$.

The parameters θ_3 , $\alpha + \theta_2$ and $\theta_1 - \theta_2$ were estimated by minimize the negative log-likelihood:

```
lnlm<-function(p){
-1(p[1], p[2], p[3])
}
pairwise<-nlm(lnlm, c(0,0,0))
theta_3<-pairwise$estimate[1]
paitheta_2<-pairwise$estimate[3]
theta_12<-pairwise$estimate[2]
```

The following codes estimates the standard error $\hat{sd}(\hat{\theta}_1 - \hat{\theta}_2) = \text{sd_error}$:

```
A<-exp(theta_12*X_{i1}+paitheta_2+theta_3*(x_4-x_3))*
(ds*s/(exp(theta_12*X_{i1}+paitheta_2+theta_3*(x_4-x_3))+s)^2
+dp*p/(1+p*exp(theta_12*X_{i1}+paitheta_2+theta_3*(x_4-x_3)))^2)
```

```
I<-rep(0,9)
dim(I)<-c(3,3)
```

```
I[1,1]<-sum(A*(x_4-x_3)^2)
```

```
I[1,2]<-sum(A*(x_4-x_3))
I[2,1]<-I[1,2]
I[1,3]<-sum(A*(x_4-x_3)*X_{i1})
I[3,1]<-I[1,3]

I[2,2]<-sum(A)
I[2,3]<-sum(A*X_{i1})
I[3,2]<-I[2,3]
I[3,3]<-sum(A*X_{i1}^2)

sd_error<-sqrt(solve(I)[3,3])
```