Adjusting for time-varying confounders in survival analysis using structural nested cumulative survival time models

Shaun Seaman

MRC Biostatistics Unit, University of Cambridge, Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, U.K.

shaun.seaman@mrc-bsu.cam.ac.uk.

Oliver Dukes

Ghent University, Ghent, Belgium

Ruth Keogh

London School of Hygiene and Tropical Medicine, London, U.K.

Stijn Vansteelandt

Ghent University, Ghent, Belgium

and London School of Hygiene and Tropical Medicine, London, U.K.

Running title: Structural nested cumulative survival time models

Abstract

Accounting for time-varying confounding when assessing causal effects of time-varying exposures on survival time is challenging. Standard survival methods that incorporate time-varying confounders as covariates generally yield biased effect estimates. Estimators using weighting by inverse probability of exposure can be unstable when confounders are highly predictive of exposure or the exposure is continuous. Structural nested accelerated failure time models require artificial recensoring, which can cause estimation difficulties. Here, we introduce the structural nested cumulative survival time model (SNCSTM). This model assumes that intervening to set exposure at time t to zero has an additive effect on the subsequent conditional hazard given exposure and confounder histories when all subsequent exposures have already been set to zero. We show how to fit it using standard software for generalised linear models and describe two more efficient, double robust, closed-form estimators. All three estimators avoid the artificial recensoring of accelerated failure time models and the instability of estimators that use weighting by the inverse probability of exposure. We examine the performance of our estimators using a simulation study and illustrate their use on data from the UK Cystic Fibrosis Registry. The SNCSTM is compared with a recently proposed structural nested cumulative failure time model, and several advantages of the former are identified.

Key words: Aalen's additive model; accelerated failure time model; g-estimation; marginal structural model; survival data.

1. Introduction

Observational studies that attempt to assess the effect of a time-varying exposure on a survival outcome typically suffer from time-varying confounding bias. Such bias is the result of time-varying factors that both influence exposure and are associated with survival, thereby distorting the association between the two. For example, studies of the effect of hospital-acquired pneumonia on time to death (since hospital admission) in critically ill patients are confounded by disease severity, because disease severity influences susceptibility to pneumonia infection and is strongly associated with mortality (Bekaert et al., 2010). Time-varying confounders (e.g. disease severity) are often affected by earlier exposures (e.g. pneumonia infection). This induces feedback relationships between exposures and confounders over time that cannot be untangled via traditional survival analysis regression methods that adjust for time-varying covariates, such as history of exposure and confounders, at each timepoint (Robins et al., 2000). The reason for this is two-fold. First, such adjustment procedures eliminate indirect effects of early exposures on survival that are mediated through those confounders. For example, it would be undesirable to eliminate effects of hospital-acquired pneumonia on survival that are mediated through disease severity, as scientific interest is primarily in the overall effect of infection. Second, such adjustment procedures are prone to collider-stratification biases that can render exposure and outcome dependent even in the absence of an exposure effect. See Daniel et al. (2013) for a review of these difficulties.

Time-varying confounding has received much attention in the causal inference literature. For survival time outcomes, the two main approaches are based on structural nested accelerated failure time models (AFTMs) (Robins and Tsiatis, 1991; Robins and Greenland, 1994) and marginal structural models (MSMs) (Robins et al., 2000). The latter approach is more popular, because of its greater simplicity and flexibility. In particular, accounting for non-informative censoring in MSMs does not, unlike in structural nested AFTMs, require an 'artificial recensoring' procedure in which originally uncensored subjects may become censored. Avoiding this recensoring is advantageous, because recensoring causes information loss, which can result in poor estimators and difficulties solving the estimating equations (Joffe et al., 2012). However, fitting MSMs relies on inverse weighting by the probability of exposure, which has it own drawback: estimators prone to large finite-sample bias and variance when confounders are highly predictive of the exposure, or when the exposure is continuous or discrete with many levels.

More recently, Young et al. (2010) and Picciotto et al. (2012) proposed a new class of discrete-time structural nested cumulative failure time models, which parameterize the effect of the exposure at each time t on the outcome at each later time in terms of the ratio of two (possibly) counterfactual cumulative failure risks at that later time under exposure regimes that differ only at time t. Their procedure has the desirable properties of structural nested AFTMs — viz. by avoiding inverse probability weighting, it handles continuous exposures without estimators being subject to large bias and variance, and it allows modelling of effect modification by time-varying covariates — while avoiding the need for artificial recensoring.

Here, we use developments by Martinussen et al. (2011) and Dukes et al. (2019) (hereafter DMTV). The former showed how to adjust for time-varying confounding when effects of exposure and confounders are parameterized on the additive hazard scale. They focused on the simple setting where interest is in estimating the direct effect of a binary baseline exposure on a survival outcome, i.e. the effect not mediated by a given intermediate variable, and where there are no baseline confounders. DMTV proposed an additive hazards model for the effect of a baseline exposure on survival time conditional on baseline confounders and derived the efficient score when (as assumed by Martinussen et al.) the confounders act additively on the hazard; this additivity assumption is not needed for consistency of their estimators.

Here, we propose a novel class of semiparametric structural nested cumulative survival time models (SNCSTMs), of which the models of Martinussen et al. (2011) and DMTV are special cases, and propose three estimators of its parameters. Our model allows baseline and time-varying confounders, binary or continuous exposure, any number of exposure measurement times and the option of constraining exposure effects to be common at different times; it does not parameterise the effects of confounders on the baseline hazard. It also allows investigation of exposure effect modification by time-varying factors. The SNCSTM is closely related to Picciotto et al.'s model, and our estimators share the forementioned desirable properties of the latter. The SNCSTM generalises Picciotto et al.'s model to continuous time and parameterises relative survival risks instead of failure risks. Our approach has several advantages over that of Picciotto et al. One of our estimators (Method 1) can be calculated using GLM software. Our other two estimators (Methods 2 and 3) are more efficient, double robust and available in closed form. All three estimators automatically handle random censoring. Also, because parameterised in continuous time, SNCSTMs can handle irregular measurement times and allow interpretation of parameters in terms of hazards.

We define notation and state fundamental assumptions in Section 2. A simple version of our SNCSTM is introduced in Section 3. In Section 4, we propose three methods for estimating its parameters. The general SNCSTM is described in Section 5. In Section 6, we discuss random censoring. A simulation study is described in Section 7. Section 8 describes an analysis of data from the UK Cystic Fibrosis (CF) Registry, looking at the effect of treatment with DNase on survival in people with CF. We conclude with a discussion in Section 9.

2. Notation and assumptions

Consider a study in which, for each of n subjects, a time-varying exposure and vector of possibly time-varying confounders are measured at time zero and at up to K follow-up visits. Until Section 5 we assume the follow-up times are regular, i.e. the same for all individuals, and (for notational simplicity) are $1, 2, \ldots, K$, and that all individuals are administratively censored at time $K+1$. Until Section 6 we assume there is no censoring apart from this administrative censoring. If visits are regular but not at times $1, \ldots, K$, or if administrative censoring occurs at a time different from $K+1$ or not at all, this can easily be accommodated by rescaling the time variable within each interval between consecutive visits.

Let T_i denote individual i's failure time, and A_{ki} and L_{ki} denote, respectively, his exposure and vector of confounders measured at time k ($k = 0, \ldots, K$). Let $R_i(t) = I(T_i \geqslant t)$ be the at-risk indicator. If individual i fails before his kth visit, A_{ki} and L_{ki} are defined as zero. Let $\bar{A}_{ki} = (A_{0i}, \ldots, A_{ki})^\top$, $\bar{L}_{ki} = (L_{0i}, \ldots, L_{ki})^\top$ and $A_{-1,i} \equiv \emptyset$. The causal ordering of the variables is $\{L_0, A_0, T \wedge 1, L_1, A_1, T \wedge 2, \ldots, L_K, A_K, T \wedge (K+1)\}$, where $x \wedge y$ means the minimum of x and y.

Define $T_i(\bar{A}_{ki}, 0)$ as individual i's (possibly) counterfactual failure time that would have applied if his exposures up to the kth visit had been as observed and his exposures from the $(k+1)$th visit onwards had been set to zero by an intervention. We make the consistency assumption that $T_i = T(\bar{A}_{k,i}, 0)$ with probability one for individuals with $A_{k+1,i} = \ldots = A_{Ki} = 0$. Note $T(\bar{A}_{k-1,i}, 0) \geqslant k$ if and only if $T(\bar{A}_{li}, 0) \geqslant k$ for all $l = k, \ldots, K$, i.e. intervening on an exposure can only affect survival after the time of that exposure. It follows that events $\{T_i \geqslant t\}$ and $\{T_i(A_{ki}, 0) \geqslant t\}$ are equivalent when $t \in [k, k+1)$. We assume $(\bar{A}_{Ki}, \bar{L}_{Ki}, T_i)$ $(i = 1, ..., n)$ are i.i.d and henceforth omit the subscript i unless needed.

We make the following sequential no unmeasured confounders assumption (NUC): $T(\bar{A}_{k-1}, 0) \perp\!\!\!\perp A_k \mid \bar{L}_k, \bar{A}_{k-1}, T \geqslant k$ $(k = 0, \ldots, K)$ (Robins, 1986). That is, among individuals who are still alive (or event-free) at time k, the assigned exposure A_k at time k may depend on $\bar{L}_k$ and $\bar{A}_{k-1}$, but given these, has no residual dependence on the remaining lifetime that would apply if all future exposures were set to zero.

3. Structural Nested Cumulative Survival Time Model (SNCSTM)

We first introduce a simple version of the SNCSTM that does not allow for exposure effect modification. The more general SNCSTM is described in Section 5.

For each $k = 0, \ldots, K$, let $\mathcal{M}_k$ be the model defined by the restriction

$$\frac{P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T \geqslant k\}}{P\{T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T \geqslant k\}} = \exp\{-A_k v_k(t)^\top \psi_k\}, \tag{1}$$

$\forall t \geqslant k$, where $\psi_k = (\psi_{k(k)}, \psi_{k(k+1)}, \ldots, \psi_{k(K)})^\top$ is a vector of of $K - k + 1$ unknown parameters, and $v_k(t)$ equals $(t - k, 0, \ldots, 0)^\top$ if $t \in [k, k+1)$, equals $(1, t - k - 1, 0, \ldots, 0)^\top$ if $t \in [k+1, k+2)$, and equals $(1, 1, t-k-2, 0, \ldots, 0)^\top$ if $t \in [k+2, k+3)$, etc. So, for any $k \leqslant l \leqslant t < l + 1$, $v_k(t)^\top \psi_k$ equals $\psi_{k(k)} + \ldots \psi_{k(l-1)} + \psi_{k(l)}(t - l)$.

Equation (1) means that among the survivors in the population at the kth visit time, in the stratum defined by any given $(\bar{A}_k, \bar{L}_k)$ the proportion who survive to a later time t when exposures from visit $k + 1$ onwards (i.e. $A_{k+1}, \ldots, A_K$) have already been set to zero would be multiplied by $\exp\{A_k v_k(t)^\top \psi_k\}$ if exposure A_k were also set to zero. Hence, $v_k(t)^\top \psi_k$ is the (controlled) direct effect of A_k on the probability of survival to time t given survival to visit k, i.e. the effect of A_k not mediated through the later exposures $A_{k+1}, \ldots, A_l$. E.g., if $\psi_{k(k)}, \ldots, \psi_{k(K)}$ are all positive and $A_k > 0$, then intervening to set $A_k = 0$ is beneficial, i.e. exposure is harmful. Conversely, if $\psi_{k(k)}, \ldots, \psi_{k(K)}$ are all negative, exposure is beneficial. This SNCSTM assumes the direct effect $v_k(t)^\top \psi_k$ is the same for any history $(\bar{A}_{k-1}, \bar{L}_k)$. In Section 5 we extend the SNCSTM to allow the effect to depend on the history.

By taking logs of each side of equation (1) and differentiating with respect to t, it can be shown that Model $\mathcal{M}_k$ can also be written as

$$\begin{aligned} &E\left\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\right\} \\ &\quad = E\left\{dN_{(\bar{A}_k,0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant t\right\} - A_k\psi_{k(l)}\, dt \end{aligned} \tag{2}$$

for $t \in [l, l+1)$ (with $l = k, \ldots, K$), where $N_{(\bar{A}_k,0)}(t) = I\{T(\bar{A}_k, 0) \leqslant t\}$ is the counting process for $T(\bar{A}_k, 0)$. Equation (2) can be interpreted as follows. In a stratum defined by $(\bar{A}_k, \bar{L}_k)$ and $T \geqslant k$, the hazard of failure at any time between visits l and $l+1$ ($l \geqslant k$) when $A_{k+1}, \ldots, A_l$ have already been set equal to zero would be reduced by $A_k\psi_{k(l)}$ if A_k were also set to zero.

Note that Model $\mathcal{M}_k$ treats $E\left\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\right\}$ — which, by NUC, equals $E\left\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\right\}$ — as a totally unspecified 'baseline' hazard, rather than parameterising its dependence on $\bar{A}_{k-1}$ and $\bar{L}_k$. One advantage of this is that the danger of incompatibility between Models $\mathcal{M}_0, \ldots \mathcal{M}_K$ is avoided. To illustrate this danger, suppose it were assumed that $E\left\{dN(t) \mid \bar{A}_1, \bar{L}_1, T \geqslant t\right\} = \phi_{10}(t) + \phi_{1A_0}(t)A_0 + \phi_{1\bar{L}_1}(t)^\top \bar{L}_1 + \psi_{1(1)}A_1$ for all $t \geqslant 1$. This, together with NUC, implies $\mathcal{M}_1$ holds. However, it also implies a restriction on the association between A_0 and T, a restriction which might conflict with that of $\mathcal{M}_0$. Such conflict would be the result of there being no coherent overall model.

4. Estimation methods

In order to estimate $\psi_{k(l)}$, we introduce nuisance Models $\mathcal{A}_k$ ($k = 0, \ldots, K$). Model $\mathcal{A}_k$ is a generalised linear model (GLM) for A_k given $\bar{A}_{k-1}$, $\bar{L}_k$ and $T \geqslant k$ with $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant k)\} = \alpha_{k0}^\top H_k$, where α_{k0} is an unknown finite-dimensional parameter and $H_k = H_k(\bar{A}_{k-1}, \bar{L}_k)$ is a known vector function of $(\bar{A}_{k-1}, \bar{L}_k)$ whose first element equals 1, e.g. $H_k = (1, A_{k-1}, L_k^\top)^\top$. The dispersion parameter ϕ_k is assumed not to depend on $\bar{A}_{k-1}$ or $\bar{L}_k$, and g is the canonical link function. The methods described in Sections 4.1–4.3 consistently estimate $\psi_{k(l)}$ when Models $\mathcal{M}_k$ and $\mathcal{A}_k$ $(k = 0, \ldots, K)$ are correctly specified. Method 1 can be applied using standard GLM software. Methods 2 and 3 improve on Method 1 by using more efficient estimators that are closely related to that described by DMTV in the setting of a single baseline exposure. Method 3 gives consistent estimation under slightly weaker conditions than Method 2, but is more computationally intensive.

4.1 *Method 1: fitting the GLM implied by Models $\mathcal{M}_k$ and $\mathcal{A}_k$*

Model $\mathcal{A}_k$ states that A_k given $\bar{A}_{k-1}$, $\bar{L}_k$ and $T \geqslant k$ obeys a GLM. Bayes' rule shows (see Web Appendix A) that Models $\mathcal{A}_k, \mathcal{M}_k, \ldots, \mathcal{M}_K$ and NUC together imply that, for any $t \geqslant k$, A_k given $\bar{A}_{k-1}$, $\bar{L}_k$ and $T(\bar{A}_k, 0) \geqslant t$ obeys the same GLM but with the intercept shifted by a function of t. Specifically, for $t \geqslant k$,

$$g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)\} = \alpha_{k0}^\top H_k + \alpha_k^\top v_k(t), \tag{3}$$

where $\alpha_k = (\alpha_{k(k)}, \ldots, \alpha_{k(K)})^\top$ and $\alpha_{k(l)} = -\psi_{k(l)}\phi_k$ $(l = k, \ldots, K)$. Our first estimation method for $\psi_{k(l)}$ involves fitting this GLM to estimate $\alpha_{k(l)}$ and calculating $\psi_{k(l)} = -\alpha_{k(l)}/\phi_k$. We now explain in more detail.

First we estimate $\psi_{k(k)}$ $(k = 0, \ldots, K)$ as follows. For $t \in [k, k+1)$, events $\{T(\bar{A}_k, 0) \geqslant t\}$ and $\{T \geqslant t\}$ are equivalent, and so equation (3) implies $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant t)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)}(t-k)$ for any $t \in [k, k+1)$. Hence, a consistent estimate $\hat{\alpha}_{k(k)}$ of $\alpha_{k(k)}$ can be obtained as follows. For each of a number (we used 10) of equally spaced values of t between k and $k+1$ (including k and $k+1$), identify the set of individuals with $T \geqslant t$ and, for each of these individuals, create a copy (a 'pseudo-individual') with the same value of $(\bar{A}_K, \bar{L}_K)$ and with new random variable Q equal to t. Fit the GLM $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, Q)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)}(Q-k)$ to the resulting set of (up to $10n$) pseudo-individuals. A consistent estimate of $\psi_{k(k)}$ is then $\hat{\psi}_{k(k)}^{\mathrm{M1}} = -\hat{\alpha}_{k(k)}/\phi_k$. When ϕ_k is unknown, it can be estimated by fitting

Model $\mathcal{A}_k$ to those of the original n individuals with $T \geqslant k$. In the simulation study of Section 7, we also tried using 50 values of t to construct the set of pseudo-individuals instead of 10, but found this made very little difference to the estimates.

Next we estimate $\psi_{k(k+1)}$ $(k = 0, \ldots, K-1)$. When $t \in [k+1, k+2)$, equation (3) is $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)} + \alpha_{k(k+1)}(t-k-1)$. If $T_i(\bar{A}_{ki}, 0)$ were known for all i, $\psi_{k(k+1)}$ could be estimated just as $\psi_{k(k)}$ was, but it is not. However, as shown in Web Appendix B, $\mathcal{M}_k, \ldots, \mathcal{M}_K$ imply that for $t \geqslant k+1$,

$$P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant k\} = E\left\{R(t)w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant k\right\}, \qquad (4)$$

where $w_k(t) = \prod_{j=k+1}^{K} \exp\left\{A_j v_j(t)^\top \psi_j\right\}$. That is, within the population stratum defined by any given value of $(\bar{A}_k, \bar{L}_k)$ and by $T(\bar{A}_k, 0) \geqslant k$ (or equivalently $T \geqslant k$), the proportion of individuals with $T(\bar{A}_k, 0) \geqslant t$ is equal to the proportion of individuals with $T \geqslant t$ after weighting each individual by $w_k(t)$. Remembering that the first element of H_k equals one for all individuals, it follows that a consistent estimate $\hat{\alpha}_{k(k+1)}$ of $\alpha_{k(k+1)}$ can be obtained by fitting the GLM $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, Q)\} = \alpha_{k0}^\top H_k + \alpha_{k(k+1)}(Q-k-1)$ to a set of pseudo-individuals constructed as described above but using ten values of t between $k+1$ and $k+2$ (rather than k and $k+1$) and using weights $w_k(Q) = \exp\{A_{k+1}\psi_{k+1(k+1)}(Q-k-1)\}$. The weights $w_k(Q)$ depend on $\psi_{k+1(k+1)}$, which is unknown, and so we replace it by its previously obtained estimate $\hat{\psi}_{k(k)}^{\mathrm{M1}}$. A consistent estimate of $\psi_{k(k+1)}$ is then $\hat{\psi}_{k(k+1)}^{\mathrm{M1}} = -\hat{\alpha}_{k(k+1)}/\phi_k$.

In general, $\psi_{k(l)}$ $(0 \leqslant k \leqslant l \leqslant K)$ is estimated by $\hat{\psi}_{k(l)}^{\mathrm{M1}} = -\hat{\alpha}_{k(l)}/\phi_k$, where $\hat{\alpha}_{k(l)}$ is the estimate of $\alpha_{k(l)}$ obtained by fitting the GLM

$$g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, Q)\} = \alpha_{k0}^\top H_k + \alpha_{k(l)}(Q-l) \qquad (5)$$

to a set of pseudo-individuals constructed using ten equally spaced values of t between l and $l+1$ and using weights $w_k(Q)$, with $\psi_{j(m)}$ replaced by $\hat{\psi}_{j(m)}^{\mathrm{M1}}$. For later reference, we denote the fitted value of $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, Q = t)$ thus obtained as $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$. This is an estimate of $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$. Note that $\hat{\psi}^{\rm M1}_{j(m)}$ $(k < j \leqslant m \leqslant l)$ must be calculated before $\hat{\psi}^{\rm M1}_{k(l)}$. If ϕ_k is unknown, it is estimated by fitting Model $\mathcal{A}_k$ to the original individuals with $T \geqslant k$.

Although this estimation procedure involves weights $w_k(t)$, these are different from the inverse probability of exposure weights used to fit MSMs and do not suffer the same instability that can plague the latter weights. In particular, if $\psi_{k(k)} = \ldots = \psi_{k(K)} = 0$, i.e. A_k has no direct effect on survival, then $w_k(t) = 1$. The variance of the weights can be reduced by using modified (or 'stabilised') weights $w^*_k(Q)$ in place of $w_k(Q)$, where $w^*_k(t) = \exp\left\{\sum_{j=k+1}^K \Delta^*_{j(k)} \upsilon_j(t)^\top \psi_j\right\}$ and $\Delta^*_{j(k)} = A_j - E(A_j \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant j)$ $(j = k+1, \ldots, K)$. This may improve efficiency, especially when A_j is precisely predicted by $(\bar{A}_{k-1}, \bar{L}_k)$. The ratio $w^*_k(Q)/w_k(Q)$ depends only on $\bar{A}_{k-1}$ and $\bar{L}_k$, and as model (5) is conditional on these, $\hat{\alpha}_{k(l)}$ remains consistent. Since $E(A_j \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant j)$ $(j = k+1, \ldots, K)$ is unknown, a working model $\mathcal{C}_{j(k)}$ is specified for it and its parameters estimated from the set of individuals still at risk at time j. Note that $\mathcal{C}_{j(k)}$ does not need to be correctly specified for $\hat{\psi}_{k(l)}$ to be consistent; indeed $\mathcal{C}_{j(k)}$ need not be compatible with $\mathcal{A}_k$.

4.2 *Method 2: g-estimation*

The principle underlying the following estimator of $\psi_{k(l)}$ is that after removing the effects of A_k and later exposures from the increment in the counting process $N(t) = I(T \geqslant t)$, NUC implies that the resulting 'blipped down' increment at any time $t \geqslant k$ is independent of A_k conditional on $\bar{A}_{k-1}$ and $\bar{L}_k$ and being still at risk.

First estimate $\psi_{k(k)}$ $(k = 0, \ldots, K)$ by solving unbiased estimating equation

$$\sum_{i=1}^n \int_k^{k+1} R_i(t) \Delta_{ki}(t) \left\{ dN_i(t) - A_{ki}\, \psi_{k(k)}\, dt \right\} = 0, \qquad (6)$$

where $\Delta_k(t) = A_k - E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$. The expectation $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$ is unknown, so we replace it by $\hat{e}_{k(k)}(\bar{A}_{k-1}, \bar{L}_k, t)$, obtained exactly as in Method 1. The next paragraph provides a rationale for equation (6).

NUC implies that counting process $N_{(\bar{A}_{k-1},0)}(t) = I(T(\bar{A}_{k-1},0) \leqslant t)$ for $T(\bar{A}_{k-1},0)$ is conditionally independent of A_k given $\bar{A}_{k-1}$, $\bar{L}_k$ and $T(\bar{A}_{k-1},0) \geqslant k$. We do not observe $N_{(\bar{A}_{k-1},0)}(t)$, but equation (2) relates $N_{(\bar{A}_{k-1},0)}(t)$ to $N_{(\bar{A}_k,0)}(t)$, the counting process for $T(\bar{A}_k,0)$, and we do observe $N_{(\bar{A}_k,0)}(t)$ when $t \in [k,k+1)$, because then it equals $N(t) = I(T \leqslant t)$, the counting process for the observed failure time T. In particular, equation (2) implies that, for any $t \in [k,k+1)$ and conditional on $(\bar{A}_k,\bar{L}_k)$, the expected increment in $N_{(\bar{A}_{k-1},0)}(t)$ during short time interval $(t,t+\delta]$ given $T(\bar{A}_{k-1},0) \geqslant t$ can be unbiasedly estimated by the corresponding mean of the observed increments in $N(t)$ minus $A_k\psi_{k(k)}\delta$ among the survivors at time t. Hence, the adjusted observed increment $N(t+\delta) - N(t) - A_k\psi_{k(k)}\delta$ should be uncorrelated with A_k given $(\bar{A}_{k-1},\bar{L}_{k-1})$ and $T \geqslant t$.

DMTV derived the semiparametric efficient estimating equation for $\psi_{k(k)}$ under Model $\mathcal{M}_k$ assuming known distribution of A_k given $(\bar{A}_{k-1},\bar{L}_k)$ and $t \geqslant k$. This equation involves inverse weighting by the hazard function; such weighting also features in efficient estimating equations of other additive hazards models. In practice, accurate estimation of the hazard function is difficult and increases the computational complexity of the procedure, and so this weighting is commonly omitted by standard fitting procedures for additive hazards models. Results of DMTV imply (see Web Appendix C) that if this is done with the semiparametric efficient equation for $\psi_{k(k)}$ under Model $\mathcal{M}_k$ and if $E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k,\bar{L}_k,T(\bar{A}_{k-1},0) \geqslant t\} = \gamma_{k(k)}^{\top}H_k$ for all $t \in [k,k+1)$, the result is equation (6).

To make equation (6) invariant to additive transformations of A_k, we replace $A_{ki}\psi_{k(k)}$ by $\Delta_{ki}(k)\psi_{k(k)}$. Since $E(A_k \mid \bar{A}_{k-1},\bar{L}_k,T(\bar{A}_k,0) \geqslant k)$ is a constant given $(\bar{A}_{k-1},\bar{L}_{k-1})$, this does not affect the unbiasedness of the estimating equations. Let $\hat{\psi}_{k(k)}^{\mathrm{M2}}$ denote the resulting estimator of $\psi_{k(k)}$.

Next estimate $\psi_{k(k+1)}$ using estimating equation $\sum_{i=1}^{n}\int_{k+1}^{k+2} R_i(t)\exp\{A_{k+1,i}\psi_{k+1(k+1)}$ $(t-k-1)\}\Delta_{ki}(t)\left[dN_i(t)-\left\{A_{k+1,i}\psi_{k+1(k+1)}+\Delta_{ki}(k+1)\ \psi_{k(k+1)}\right\}dt\right]=0$. The unknown $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$ and $\psi_{k+1(k+1)}$ are replaced by $\hat{e}_{k(k+1)}(\bar{A}_{k-1}, \bar{L}_k, t)$ and $\hat{\psi}^{\mathrm{M2}}_{k+1(k+1)}$. The next paragraph provides a rationale for this estimating equation.

Again we exploit the conditional independence of $N_{(\bar{A}_{k-1},0)}(t)$ and A_k (NUC) and the relation between $N_{(\bar{A}_{k-1},0)}(t)$ and $N_{(\bar{A}_k,0)}(t)$, but now over time interval $[k+1, k+2)$. An added complication is that $N_{(\bar{A}_k,0)}(t)$ is not observed when $t > k+1$. However, we know from equation (2) that when $t \in [k+1, k+2)$ the intensities of $N_{(\bar{A}_k,0)}(t)$ and $N(t) = N_{(\bar{A}_{k+1},0)}(t)$ differ by $A_{k+1}\psi_{k+1(k+1)}$ and (as noted in Section 4.1) there are $w_k(t) = \exp\{A_{k+1}\ \psi_{k+1(k+1)}(t-k-1)\}$ times as many individuals with $T(\bar{A}_k, 0) \geqslant t$ in the population as there are with $T(\bar{A}_{k+1}, 0) \geqslant t$. So, we can unbiasedly estimate the expected increment in $N_{(\bar{A}_{k-1},0)}(t)$ over small interval $[t, t+\delta)$ as the weighted mean of the increments in $N(t)$ minus $(A_{k+1}\psi_{k+1(k+1)} + A_k\psi_{k(k+1)})\delta$ with weights $\exp\{A_{k+1}\psi_{k+1(k+1)}(t-k-1)\}$. This justifies the above estimating equation but with $A_{ki}\psi_{k(k+1)}$ in place of $\Delta_{ki}(k+1)\psi_{k(k+1)}$. We use $\Delta_{ki}(k+1)\psi_{k(k+1)}$ instead for the same reason that we replaced $A_{ki}\psi_{k(k)}$ by $\Delta_{ki}(k)\psi_{k(k)}$ in equation (6).

In general, the consistent estimator $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ of $\psi_{k(l)}$ $(l \geqslant k)$ is obtained by solving

$$\sum_{i=1}^{n}\int_{l}^{l+1} R_i(t)w_{ki}(t)\ \Delta_{ki}(t) \times \left[dN_i(t) - \left\{\sum_{j=k+1}^{l} A_{ji}\psi_{j(l)} + \Delta_{ki}(l)\ \psi_{k(l)}\right\}\ dt\right] = 0 \qquad (7)$$

with $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant t)$ replaced by $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ and $\psi_{j(l)}$ $(j > k)$ replaced by $\hat{\psi}^{\mathrm{M2}}_{j(l)}$; this requires that $\psi_{j(m)}$ $(k < j \leqslant m \leqslant l)$ be estimated before $\psi_{k(l)}$. The estimator $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ is available in closed form (see Web Appendix E for formulae when $g(.)$ is the identity or logistic link function).

In Web Appendix F we prove $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ is double robust in the following sense. Let

$e^*_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ denote the limit as $n \to \infty$ of $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$, and let Model $\mathcal{B}_{k(l)}$ ($l \geqslant k$) be defined by the restriction $E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\} = \left\{\gamma_{k(l)}^\top H_k - e^*_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, k)\ \psi_{k(l)}\right\} dt\ \forall t \in [l, l+1)$, where $\gamma_{k(l)}$ are unknown parameters. $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ is consistent if 1) $\mathcal{M}_k, \ldots, \mathcal{M}_l$, 2) either $\mathcal{A}_k$ or $\mathcal{B}_{k(l)}$, and 3) for each $j = k+1, \ldots, l$, either $\mathcal{A}_j$ or all of $\mathcal{B}_{j(j)}, \ldots, \mathcal{B}_{j(l)}$ are correctly specified. The term $e^*_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, k)\psi_{k(l)}$ in Model $\mathcal{B}_{k(l)}$ arises because of the use of $\Delta_k(l)\psi_{k(l)}$, rather than $A_k\psi_{k(l)}$, in equation (7) (see proof). Note that if $\psi_{k(l)} = 0$ or $\mathcal{A}_k$ is a linear regression, so that $e^*_{k(l)}(A_k, \bar{L}_k, k)\ \psi_{k(l)}$ is a linear function of H_k, it can be omitted. As in Method 1, efficiency may be gained by using stabilised weights $w^*_{ki}(t)$ in place of $w_{ki}(t)$ in equation (7). Also, to make $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ invariant to additive transformations of $A_{k+1}, \ldots, A_l$, the term $A_{ji}\psi_{j(l)}$ can be replaced by $\Delta^*_{j(k),i}\psi_{j(l)}$.

4.3 *Method 3: improved g-estimation*

If we use a different estimator $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ of $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$ for the $\Delta_k(t)$ and $\Delta_k(l)$ terms in equation (7), then the estimator solving (7) remains consistent under a more general version of Model $\mathcal{B}_{k(l)}$. In Methods 1 and 2, $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ is calculated by fitting a single GLM to a set of pseudo-individuals, with time since lth visit, $Q-l$, as a covariate. In Method 3, we instead fit a separate GLM at each time since the lth visit. That is, for any $t \geqslant 0$, we calculate $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ by fitting the GLM $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k)\} = \alpha_{k0}(t)^\top H_k$ to the set of individuals with $T \geqslant t$, using weights $w_k(t)$. This set changes only at times t at which an individual leaves the risk set, and so the GLM needs to be fitted only at these times. This is the approach taken by DMTV, who denoted the resulting estimator of $\psi_{k(k)}$ as "$\hat{\psi}_{\mathrm{TVPS-DR}}$" and, on the basis of results from a simulation study, recommended it over three alternatives. As in Method 2, we can use stabilised weights and replace $A_j\psi_{j(l)}$ by $\Delta^*_{j(k)}\psi_{j(l)}$. As shown in Web Appendix

F, Method 3 has the same double robustness property as Method 2 but with the parameters $\gamma_{k(l)}$ in Model $\mathcal{B}_{k(l)}$ now allowed to be a function of $t-l$.

4.4 *Constraining exposure effects*

In some applications, it may be desirable to impose the constraint that $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ for all k, k', m, i.e. the effect of exposure measured at one visit k on the hazard m visits later is the same for all k. This reduces the number of parameters and, as we see in Section 7, increases the precision of their estimates. In Web Appendix G we explain how estimation may be performed under this constraint. See Vansteelandt and Sjolander (2016) for how to impose other constraints.

5. The general SNCSTM

In this section, we extend the SNCSTM to allow visit times to be irregular, i.e. to vary from one individual to another, and effect modification, i.e. the effect of exposure on survival to depend on the exposure and confounder histories.

Let S_{ki} denote the time of individual i's kth follow-up visit ($k = 1, \ldots, K$), and let $S_{0i} = 0$ ($i = 1, \ldots, n$) and $\bar{S}_i = (S_{1i}, \ldots, S_{Ki})^\top$. Until now, we have assumed $S_{ki} = k\ \forall i$. We assume visit times $\bar{S}$ are planned or randomly chosen at baseline using only baseline confounder information, i.e. L_0, and we modify NUC to be $T(\bar{A}_{k-1}, 0) \perp\!\!\!\perp A_k \mid \bar{L}_k, \bar{A}_{k-1}, \bar{S}, T \geqslant S_k$ ($k = 0, \ldots, K$). Also, let $S_{K+1,i}$ denote an administrative censoring time common to all individuals (until now, we assumed $S_{K+1,i} = K+1$). If there is no such time, let $S_{K+1,i} = \infty$. To allow effect modification, we define $Z_{k(l)} = (1, Z_{k(l)}^{\text{int}\top})^\top$, where $Z_{k(l)}^{\text{int}}$ is a known (possibly vector) function of $(\bar{A}_{k-1}, \bar{L}_k, \bar{S})$ ('int' stands for 'interactions'), and let $Z_k = (Z_{k(k)}^\top, \ldots, Z_{k(K)}^\top)^\top$.

For each $k = 0, \ldots, K$, let $\mathcal{M}_k$ be the model defined by the restriction

$$\frac{P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, \bar{S}, T \geqslant S_k\}}{P\{T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, \bar{S}, T \geqslant S_k\}} = \exp\{-A_k v_k(t, Z_k, \bar{S})^\top \psi_k\}, \qquad (8)$$

where $v_k(t, Z_k, \bar{S})$ equals $\big((t - S_k)Z_{k(k)}^\top, 0, \ldots, 0\big)^\top$ if $t \in [S_k, S_{k+1})$, equals $\big((S_{k+1} -$

$S_k)Z_{k(k)}^\top, (t-S_{k+1})Z_{k(k+1)}^\top, 0, \dots, 0\big)^\top$ if $t \in [S_{k+1}, S_{k+2})$, and equals $\big((S_{k+1}-S_k)Z_{k(k)}^\top,$ $(S_{k+2}-S_{k+1})Z_{k(k+1)}^\top, (t-S_{k+2})Z_{k(k+2)}^\top, 0, \dots, 0\big)^\top$ if $t \in [S_{k+2}, S_{k+3})$, etc. If $S_k = k$ and $Z_{k(l)} = 1$, equation (8) reduces to equation (1). Model $\mathcal{M}_k$ can also be written as $E\left\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_{k-1}, 0) \geqslant t\right\} = E\left\{dN_{(\bar{A}_k,0)}(t) \mid \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant t\right\}$ $-A_k\psi_{k(l)}^\top Z_{k(l)}\, dt$ for $t \in [S_l, S_{l+1})$.

The modifications to Methods 1 and 2 needed to fit the general SNCSTM are simple (see Web Appendix D). Modifying Method 3 is simple when visit times are regular; it is possible for irregular visit times, but is fiddly. In the simulation study reported in Section 7 we found little benefit from Method 3 relative to Method 2 when visit times were regular, and so did not implement it for irregular times.

6. Censoring

We now allow for censoring before the administrative censoring time. Let C_i and $\tilde{T}_i$ denote individual i's censoring and failure times, respectively. Redefine T_i and $N_i(t)$ as $T_i = \tilde{T}_i \wedge C_i$ and $N_i(t) = I(T_i \leqslant t,\ T_i < C_i)$; $R_i(t)$ is unchanged except that T_i has this new meaning. With these changes, Methods 1–3 remain valid, provided two further conditions hold (Vansteelandt and Sjolander, 2016). First, the censoring hazard does not depend on the exact failure time or future exposures or confounders. That is, the counting process, $N_C(t) = I(C \leqslant t)$, for the censoring time satisfies $E\{dN_C(t) \mid C \geqslant t, \bar{A}_{\lfloor \tilde{T} \rfloor}, \bar{L}_{\lfloor \tilde{T} \rfloor}, \bar{S}, \tilde{T} > t, \tilde{T}\} = \lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})\ \forall t$, where $\bar{A}_{\lfloor t \rfloor}$ and $\bar{L}_{\lfloor t \rfloor}$ are the exposure and confounder histories up to time t and $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})$ is some function only of $(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})$. The second condition, which can be weakened by using censoring weights (see Web Appendix H), is that $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}) = \lambda(t, L_0, \bar{S})$, so censoring depends only on baseline confounders.

7. Simulation study

We used a simulation study to investigate bias and efficiency of the methods. There were $K+1=4$ visits and two time-dependent confounders (i.e. $\dim(L_k)=2$). These and the exposure were generated as: $L_0 \sim N\left((0,0),\ \Sigma\right)$, $A_0 \sim N(3+(0.2,0.1)^\top L_0,\ \ 0.9^2)$, $L_k \sim N\left(\Omega L_{k-1}+(0.1,0.05)^\top A_{k-1},\ \Sigma\right)$ and $A_k \sim N(3+(0.1,0.05)^\top L_k,\ \ 0.7^2)$ $(k \geqslant 1)$, where $\Sigma = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}$ and $\Omega = \begin{bmatrix} 0.2 & 0.2 \\ 0.1 & 0.1 \end{bmatrix}$. The hazard of failure during the interval between the kth and $(k+1)$th visits was $0.34+(0.03,0.03)^\top L_k - 0.04A_k - 0.0145A_{k-1}I(k\geqslant 1) - 0.0055A_{k-2}I(k\geqslant 2) - 0.00245A_{k-3}I(k=3)$. For this data-generating mechanism, $\mathcal{M}_k$ $(k=0,\dots K)$ is correctly specified with no effect modification (i.e. $Z_{k(l)}=1$) and the true exposure effects are $\psi_{k(k)}=-0.04$, $\psi_{k(k+1)}=-0.01$, $\psi_{k(k+2)}=-0.004$ and $\psi_{k(k+3)}=-0.002$.

We considered three scenarios: two with regular and one with irregular visit times. For regular visits, $S_{ik}=k$. For irregular visits, inter-visit times $S_{k+1,i}-S_{ki}$ were independently uniformly distributed on $[0.5,1.5]$. There was administrative censoring at time 4. In one of the regular visit scenarios, there was no random censoring. In the other, and in the irregular visit scenario, there was an exponentially distributed random censoring time with mean 5. For the regular visit scenario without random censoring, the expected percentages of individuals observed to fail between visits 0 and 1, 1 and 2, 2 and 3, and between visit 3 and time 4 were 20%, 14%, 11% and 9%, respectively. For the regular and irregular visit scenarios with random censoring, these percentages were 18%, 10%, 6% and 4%, and the corresponding expected percentages of individuals censored were 16%, 11%, 8% and 5%. For each scenario, we generated 1000 datasets, each with $n=1000$ individuals. Estimation was done with and without the constraint, which is true here, that $\psi_{k(k+m)}=\psi_{k'(k'+m)}$.

Tables 1 and 2 show for the regular visit scenario without and with random censoring, respectively, the mean estimates and standard errors (SEs) for Methods 1–3. Results for the irregular visit scenario are in Web Appendix L. We see that all the estimators are approximately unbiased, though there is some bias for $\psi_{0(2)}$, $\psi_{0(3)}$ and $\psi_{1(3)}$, parameters for which there is relatively little information in the data. Comparing SEs, we see that Methods 2 and 3 give very similar results, and that these methods are more efficient than Method 1. This difference in efficiency is much greater when there is random censoring (it is even greater when visit times are irregular — see Web Appendix L). This may be because Method 1, unlike 2 and 3, does not distinguish between failure and censoring (or occurrence of next visit). Although Methods 2 and 3 use fitted values from the same GLM that is used in Method 1, the estimating equations for Methods 2 and 3 involve increments $dN(t)$, which equal one only when a failure occurs. For Methods 1 and 2, coverage of 95% bootstrap confidence intervals (using 1000 bootstraps) was close to 95% (see Table 3). Coverage was not evaluated for Method 3, as it is computationally intensive to bootstrap this method for 1000 simulated datasets. Imposing the constraint that $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ reduced SEs, as expected.

In this simulation study, censoring times are independent of exposures and confounders, and so censoring weights (Section 6) are not required for consistent estimation of the $\psi_{k(l)}$'s. However, applying Method 1 with censoring weights improved its efficiency (see Method 1cw in Tables 1 and 2), probably because chance associations between exposures and censoring events are reduced in the weighted sample. Coverage of bootstrap confidence intervals (Table 3) was close to 95% for most parameters, but there was overcoverage for some parameters. Using censoring weights had no effect on the efficiency of Method 2.

Web Appendix L shows results for $n = 250$ or for a shorter follow-up time with times between visits divided by four and administrative censoring at time 1 (and so fewer failures). These are qualitatively similar to the results in Tables 1 and 2, but with the relative inefficiency of Method 1 being even more marked in the scenarios with shorter follow-up time. Web Appendix L also describes a simulation study that demonstrates the double robustness of Methods 2 and 3.

8. Analysis of Cystic Fibrosis registry data

The UK Cystic Fibrosis (CF) Registry records health data on nearly all people with CF in the UK at designated approximately annual visits (Taylor-Robinson et al., 2018). To illustrate the use of the SNCSTM, we used data on 2386 individuals observed during 2008–2016 to investigate the causal effect of the drug DNase on survival. DNase has been found to have a beneficial effect on lung function, including using Registry data (Newsome S et al., 2019), but its effect on survival has not been studied. Baseline visit was defined as an individual's first visit during 2008–2015, and there were up to $K = 8$ follow-up visits. The (irregular) visit times were defined as years after baseline visit; median time between visits was 1.00 years (interquartile range 0.93 to 1.07). Individuals were defined as 'treated' if they had used DNase since the previous visit and 'untreated' otherwise. Individuals treated at a visit prior to their baseline visit were excluded, as were visits prior to age 18. Administrative censoring was applied at the end of 2016 and non-administrative censoring when an individual had a transplant or had not been seen for 18 months. The percentage of treated patients increased from 14% at the baseline visit to 52% at visit 8, and most patients who began using DNase continued to use it. There were 137 deaths during follow-up and 653 non-administrative censorings (including 36 transplants). Of those who died, 74 (63) were treated (untreated) at time of death. Total follow-up was 12380 person-years (py), and death rates while treated and untreated were, respectively, 0.019 (74/3930) and 0.0075 (63/8450) py^{-1}. The ratio of the probabilities of surviving for one year while treated and untreated is thus $0.981/0.9925 = 0.989$. However, this may be due to confounding: sicker patients being more likely to receive treatment.

We estimated the effect on survival of delaying initiation of treatment by one year. To do this, we (re)defined A_k as $A_k = 0$ for those treated at visit k, and $A_k = 1$ for those untreated. Now $\exp(-\psi_{k(k)})$ represents the multiplicative causal effect of intervening to start treatment at visit k rather than at visit $k+1$ on the probability of surviving for at least one year after visit k, among patients who survive to, and are untreated at, visit k and conditional on confounder history $\bar{L}_k$. More generally, $\exp\big(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\big)$ is the effect on the probability of surviving at least m years after visit k if visits are exactly annual. We imposed the constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$. (Potential) confounders at visit k were baseline variables sex, age and genotype class (low, high, not assigned), and time-varying variables $FEV_1\%$, body mass index, days of IV antibiotic use, and binary indicators for four infections (P. aeruginosa, S. aureus, B. cepacia complex, Aspergillus), CF-related diabetes, smoking, and use of other mucoactive treatments and oxygen therapy. The same variables (and treatment) were included in models for inverse probability of censoring weights.

Figure 1a shows estimates of $\exp\big(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\big)$ from Method 2. These suggest that starting treatment now rather than waiting may cause a small decrease in probability of survival, at least for the first five years: $\exp\big(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\big) = 0.997$, 0.996, 0.997, 0.994 and 0.988 for $m = 1, \ldots, 5$, respectively. However, the confidence intervals (obtained by bootstrapping) include 1, i.e. no treatment effect. This lack of a significant treatment effect may be because we have focused on a subset of the population (adults not previously treated with DNase) and/or because there are unmeasured confounders. As expected, Method 1 was very inefficient in this situation of irregular visits and substantial censoring. The confidence intervals were between 4 and 9 times wider than those from Method 2.

For illustration, we also fitted a SNCSTM with an interaction between treatment and $\mathrm{FEV}_1\%$. Figures 1b–d shows the estimated ratios of survival probabilities for three value of $\mathrm{FEV}_1\%$: 40, 75 and 100 (the 10th, 50th and 90th centiles of the distribution at baseline). Figure 1d suggests the ratio may actually be greater than 1 for $\mathrm{FEV}_1\% = 100$, i.e. starting treatment now may be better than waiting for patients with high $\mathrm{FEV}_1\%$. However, the interaction terms are not significant.

9. Discussion

One advantage of SNCSTMs is that, in contrast to MSMs, they can cope well with situations where the inverse probabilities of exposure are highly variable. Indeed, they can even be used when the so-called experimental treatment assignment assumption is violated, i.e. when some individuals are, on the basis of their time-varying covariate information, excluded from receiving particular exposure levels. For these individuals, $\Delta_i(t) = 0$, meaning they do not contribute to the estimating functions of Methods 1–3.

Another advantage of SNCSTMs is that they can be used to investigate time-varying modification of exposure effects on survival time. Although it is, in principle, possible to do this using structural nested AFT models, estimation difficulties caused by artificial recensoring mean that such models are usually kept simple and interactions are not explored.

The SNCSTM can also be used to estimate the counterfactual exposure-free survivor function, i.e. $P\{T(0) \geqslant t\}$, as $n^{-1}\sum_{i=1}^{n} R_i(t) \prod_{j=0}^{K} \exp\{A_{ji} v_j(t, Z_{ji}, \bar{S}_i)^\top \psi_j\}$. This is because equations (4) and (8) imply $P\{T(0) \geqslant t\} = E\big[R(t) \prod_{j=0}^{K} \exp\{A_j$

$v_j(t, Z_j, \bar{S})^\top \psi_j\}\big]$. If there is censoring before time t, $R_i(t)$ should be inversely weighted by an estimate of $P(C_i \geqslant t \mid \bar{A}_{\lfloor t \rfloor i}, \bar{L}_{\lfloor t \rfloor i}, \bar{S}_i)$.

A limitation is that, like other additive hazards models, the SNCSTM does not constrain hazards to be non-negative, and so does not exclude survival probabilities greater than one. Similarly, Picciotto et al.'s (2012) structural nested cumulative failure time model does not exclude failure probabilities greater than one.

Method 1 appears to be less efficient than Methods 2 and 3, but has the attraction that it can be applied using standard GLM software. In our simulation study, the efficiency loss was fairly small when the only censoring was administrative and visit times were regular. This method became much less competitive, however, when there was random censoring, and even more so when visit times were irregular. By not distinguishing between failure and censoring, Method 1 may also be more sensitive than Methods 2 and 3 to violation of the assumption that $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}) = \lambda(t, L_0, \bar{S})$. Of the three, Method 3 gives consistent estimation under the weakest assumptions. However, it needs more computation than Methods 1 and 2, especially when visit times are irregular and the exposure is binary. In our simulation study, Methods 2 and 3 performed similarly, and so the theoretical advantage of Method 3 may not be worth the extra computation. An R function for implementing our methods, with examples, is described in Web Appendix I.

DMTV discuss the close connection between their model for a point exposure (which is equivalent to the SNCSTM with $K = 0$) and Picciotto et al's (2012) cumulative failure time model. Although the latter is a discrete-time model for the probability of failure, it is easy to finely discretise time so as to approximate continuous time and (as Picciotto et al. note) to reformulate it as a model for probability of survival. As DMTV explain, a drawback of Picciotto et al.'s method is the difficulty of deriving the efficient estimating equation. This difficulty arises because their class of estimating functions uses correlated survival indicators. By instead using independent increments of a counting process, DMTV were able to derive the efficient estimating function. Methods 2 and 3 are extensions to time-varying exposures of DMTV's recommended method, and are therefore expected also to be more efficient than Picciotto et al.'s method. In Web Appendix J we elaborate on DMTV's discussion of Picciotto et al.'s model and reformulate it as a model for probability of survival. Tables 1 and 2 show mean estimates and SEs for the resulting Picciotto et al. estimator (described in Web Appendix J and denoted 'Method P' in tables). The SEs are larger than those of Methods 2 and 3, suggesting Methods 2 and 3 are indeed more efficient. Methods 2 and 3 also have the advantages of using closed-form estimators, handling random censoring automatically (because estimating functions are framed in terms of increments, which are observable up to the time of censoring), and being double robust. Picciotto et al. use an iterative Nelder-Mead algorithm, employ inverse probability of censoring weighting to handle random censoring, even when this censoring is completely at random, and their estimator is not double robust.

In Web Appendix K we outline how the SNCSTM can handle competing risks.

Acknowledgements

SRS is supported by MRC grant MC_UU_00002/10; OD by a Strategic Basic Research PhD grant from the Research Foundation Flanders (FWO); and RK by MRC grant MR/M014827/1. We thank people with cystic fibrosis and their families for consenting to their data being held in the UK Cystic Fibrosis (CF) Registry and NHS teams in CF centres and clinics for the input of data into the Registry. We also thank the UK Cystic Fibrosis Trust and the Registry Steering Committee for access to anonymized UK CF Registry data. The analyses presented in this paper use a data set resulting from a cleaning process undertaken by the CF Epidemiological Network, which was funded by a Strategic Research Centre Grant from the Cystic Fibrosis Trust.

References

Bekaert, M., Vansteelandt, S., and Mertens, K. (2010). Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. *Lifetime Data Analysis* **16,** 45–70.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M. G., and Sterne, J. (2013). Methods for dealing with time-dependent confounding. *Statistics in Medicine* **32,** 1584–1618.

Dukes, O., Martinussen, T., Tchetgen Tchetgen, E., and Vansteelandt, S. (2019). On doubly robust estimation of the hazard difference. *Biometrics* **75,** 100–109.

Joffe, M., Yang, W., and Feldman, H. (2012). G-estimation and artificial censoring: Problems, challenges, and applications. *Biometrics* **68,** 275–86.

Martinussen, T., Vansteelandt, S., Gerster, M., and Hjelmborg, J. (2011). Estimation of direct effects for survival data by using the Aalen additive hazards model. *Journal of the Royal Statistical Society: Series B* **73,** 773–788.

Newsome S, Daniel, R., Carr, S., Bilton, D., and Keogh, R. (2019). Investigating the effects of long-term dornase alfa use on lung function using registry data. *Journal of Cystic Fibrosis* **18,** 110–117.

Picciotto, S., Hernán, M., Page, J., Young, J., and Robins, J. (2012). Structural nested cumulative failure time models to estimate the effects of interventions. *Journal of the American Statistical Association* **107,** 866–900.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling* **7,** 1393–1512.

Robins, J. and Greenland, S. (1994). Adjusting for differential rates of prophylaxis therapy for PCP in high-dose versus low-dose AZT treatment arms in an aids randomized trial. *Journal of the American Statistical Association* **89,** 737–749.

Robins, J., Hernan, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11,** 550–560.

Robins, J. and Tsiatis, A. (1991). Correcting for non-compliance in randomized trials using rank-preserving structural failure time models. *Communications in Statistics* **20,** 2609–2631.

Taylor-Robinson, D., Archangelidi, O., Carr, S., Cosgriff, R., Gunn, E., Keogh, R., MacDougall, A., Newsome, S., Schluter, D., Stanojevic, S., Bilton, D., and CF-EpinNet Collaboration (2018). Data resource profile: the UK cystic fibrosis registry. *International Journal of Epidemiology* **47,** 9–10e.

Vansteelandt, S. and Sjolander, A. (2016). Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. *Epidemiologic Methods* **5,** 37–56.

Young, J., Hernan, M., Picciotto, S., and Robins, J. (2010). Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis* **16,** 71–84.

Supporting Information

Web Appendices, Web Tables and R function referenced in Sections 4–7 and 9 are available with this paper at the Biometrics website on Wiley Online Library.

Figure 1: Estimates of the ratio of the survival probabilities when treatment is initiated immediately compared to initiation being delayed by one year. a: from the model with no interaction. b, c and d: from the model with interaction between treatment and $FEV_1\%$.

Supporting Information for 'Adjusting for Time-Varying Confounders in Survival Analysis Using Structural Nested Cumulative Survival Time Models' by Seaman, Dukes, Keogh and Vansteelandt.

The proofs, estimators and inverse probability of censoring weights in these web appendices are for the general SNCSTM described in Section 5 of the article. Proofs, estimators and weights for the simple SNCSTM with regular visit times and no effect modification are just special cases of the proofs, estimators and weights given here. Specifically, for the simple SNCSTM, $\bar{S} = (1, 2, \ldots, K)$ and $Z_{k(l)} = 1$.

In these web appendices, we write $v_k(t)^\top \psi_k$ as $G_k(t)$. Mentions of equations (1)–(8) refer to equations that appear in the article.

A. Proof of equation (3)

Model $\mathcal{A}_k$ implies we can write the probability density of A_k given $\bar{A}_{k-1}, \bar{L}_k, \bar{S}, T \geqslant S_k$ as

$$f(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T \geqslant S_k) = b(A_k; \phi_k) \exp\{A_k \tau - c(\tau)\}/d(\phi)$$

for some functions $b(.)$, $c(.)$ and $d(.)$ and where $\tau = \alpha_{k0}^\top H_k$ is the linear predictor. To simplify notation, we shall omit the explicit conditioning on $\bar{S}$ and instead take it as implicit.

Using Bayes' Rule, Models $\mathcal{A}_k$ and $\mathcal{M}_k$ and the no-unmeasured confounders assumption, we have,

$$
\begin{aligned}
& f(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t) \\
& \propto f(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k) \times P(T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k) \\
& = f(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k) \times P(T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k) \\
& = b(A_k; \phi_k) \exp\{A_k \tau - c(\tau)\}/d(\phi) \\
& \qquad \times P(T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k) \exp\{-A_k G_k(t)\} \\
& = b(A_k; \phi_k) \exp\{A_k \tau - c(\tau)\}/d(\phi) \\
& \qquad \times P(T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k) \exp\{-A_k G_k(t)\} \\
& \propto b(A_k; \phi_k) \exp\{A_k \tau - c(\tau)\}/d(\phi) \times \exp\{-A_k G_k(t)\} \\
& = b(A_k; \phi_k) \exp\{A_k \tau^* - c(\tau)\}/d(\phi) \\
& \propto b(A_k; \phi_k) \exp\{A_k \tau^*\}/d(\phi) \\
& \propto b(A_k; \phi_k) \exp\{A_k \tau^* - c(\tau^*)\}/d(\phi)
\end{aligned}
$$

where $\tau^* = \tau - G_k(t) d(\phi)$.

B. Proof of equation (4)

First, we prove that

$$
\begin{aligned}
& P\{T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \\
& \qquad = E\left[R(t) \exp\left\{\sum_{j=k}^{K} A_j G_j(t)\right\} \mid \bar{A}_k, \bar{L}_k, \bar{S}, T \geqslant S_k\right]
\end{aligned} \tag{9}
$$

for $t \geqslant S_k$. To simplify notation, we shall omit the explicit conditioning on $\bar{S}$ and instead take it as implicit.

For $t \geqslant S_k$,

$$
\begin{aligned}
& P\{T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \\
& = P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \exp\{A_k G_k(t)\}
\end{aligned}
$$

Hence, for $S_k \leqslant t < S_{k+1}$,

$$
\begin{aligned}
& P\{T(\bar{A}_{k-1},0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\} && (10)\\
& = P\{T \geqslant t \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \\
& = E_T\{R(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \\
& = E_T\{R(t) \exp\{A_k G_k(t)\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} && (11)
\end{aligned}
$$

So, equation (9) has been proved for $S_k \leqslant t < S_{k+1}$.

Note that since (10) does not depend on A_k (by NUC), nor can (11). Hence, we can also write:

$$
P\{T(\bar{A}_{k-1},0) \geqslant t \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\} = E\left\{R(t) \exp\{A_k G_k(t)\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k\right\} \quad (12)
$$

Next, for $t \geqslant S_{k+1}$,

$$
\begin{aligned}
& P\{T(\bar{A}_{k-1},0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\} \\
& = P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\} \exp\{A_k G_k(t)\} \\
& = P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant S_k\} \exp\{A_k G_k(t)\} \\
& = P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant S_{k+1}\} \\
& \qquad \times P\{T(\bar{A}_k,0) \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant S_k\} \exp\{A_k G_k(t)\}
\end{aligned}
$$

Hence, for $S_{k+1} \leqslant t < S_{k+2}$,

$$
\begin{aligned}
&P\{T(\bar{A}_{k-1},0) \geqslant t \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\} \\
&= P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant S_{k+1}\} \\
&\qquad \times P\{T(\bar{A}_k,0) \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k,0) \geqslant S_k\} \exp\{A_k G_k(t)\} \\
&= P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T \geqslant S_{k+1}\} \\
&\qquad \times P\{T \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \\
&= E_{L_{k+1}}[P\{T(\bar{A}_k,0) \geqslant t \mid \bar{A}_k, \bar{L}_{k+1}, T \geqslant S_{k+1}\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_{k+1}] \\
&\qquad \times P\{T \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \\
&= E_{L_{k+1}}[E_{T,A_{k+1}}\{R(t)\exp\{A_{k+1}G_{k+1}(t)\} \mid \bar{A}_k, \bar{L}_{k+1}, T \geqslant S_{k+1}\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_{k+1}] \\
&\qquad \times P\{T \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \qquad \text{(using (12))} \\
&= E_{T,A_{k+1},L_{k+1}}\{R(t)\exp\{A_{k+1}G_{k+1}(t)\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_{k+1}\} \\
&\qquad \times P\{T \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \exp\{A_k G_k(t)\} \\
&= E_{T,A_{k+1},L_{k+1}}\{R(t)\exp\{A_k G_k(t) + A_{k+1}G_{k+1}(t)\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_{k+1}\} \\
&\qquad \times P\{T \geqslant S_{k+1} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \\
&= E_{T,A_{k+1},L_{k+1}}\{R(t)\exp\{A_k G_k(t) + A_{k+1}G_{k+1}(t)\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\}
\end{aligned}
$$

So, equation (9) has been proved for $S_{k+1} \leqslant t < S_{k+2}$.

Using induction, the same argument can be used to prove equation (9) for $S_{k+2} \leqslant t < S_{k+3}$, then for $S_{k+3} \leqslant t < S_{k+4}$, and so on.

Now, equation (4) follows from $\mathcal{M}_k$ and equation (9), because

$$
\begin{aligned}
&P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \\
&= P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \\
&= P\{T(\bar{A}_{k-1}, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \exp\{-A_k G_k(t)\} \\
&= E\left[R(t) \exp\left\{\sum_{j=k}^{K} A_j G_j(t)\right\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right] \exp\{-A_k G_k(t)\} \\
&= E\left[R(t) \exp\left\{\sum_{j=k+1}^{K} A_j G_j(t)\right\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right] \\
&= E\left\{R(t) w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right\}
\end{aligned}
$$

Note that the no unmeasured confounders assumption means that the left-hand side of equation (9) cannot depend on A_k. Hence, the right-hand side cannot depend on A_k either.

C. Relation between semiparametric efficient estimating equation for $\psi_{k(k)}$ and equation (6)

In their Section 3.1, DMTV derived the semiparametric efficient estimating equation for $\psi_{k(k)}$ when the conditional distribution of A_k given $(\bar{A}_{k-1}, \bar{L}_k)$ and $T \geqslant k$ is known. This estimating equation involves inverse weighting by the hazard function. When this inverse weighting is omitted, the semiparametric efficient estimating equation becomes

$$
\sum_{i=1}^{n} \int_{k}^{k+1} R_i(t) \Delta_{ki}(t) \left\{dN_i(t) - d\Omega_{ki}(t, \bar{A}_{ki}, \bar{L}_{ki}) - A_{ki} \psi_{k(k)} \, dt\right\} = 0, \tag{13}
$$

where $d\Omega_k(t, \bar{A}_k, \bar{L}_k) = E\{dN(t) - A_k \psi_{k(k)} \, dt \mid \bar{A}_k, \bar{L}_k, T \geqslant t\} = E\{dN_{(\bar{A}_{k-1}, 0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\}$ for $t \in [k, k+1)$. In their Section 4.1, DMTV showed that if $d\Omega_k(t, \bar{A}_k, \bar{L}_k) = \gamma_{k(k)}(t-k)^\top H_k$ for all $t \in [k, k+1)$ for some (possibly) time-varying parameter $\gamma_{k(k)}(t-k)$ and if the term $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant t)$ in $\Delta_k(t) = A_k - E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant t)$ is estimated by fitting a separate GLM at each time t (as we do in Method 3 — see our Section 4.3), then

$$\sum_{i=1}^{n} \int_{k}^{k+1} R_i(t)\Delta_{ki}(t)\; d\Omega_{ki}(t, \bar{A}_{ki}, \bar{L}_{ki}) = 0, \tag{14}$$

and so equation (13) reduces to equation (6). This result is also shown in our Web Appendix F. In Web Appendix F, we further show that if $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant t)$ is instead estimated by fitting a single GLM (as we do in Method 2), then equation (14) still holds, provided that $\gamma_{k(k)}(t-k) = \gamma_{k(k)}$ does not depend on t.

D. Estimation for the general SNCSTM

The following estimation methods, which are suitable for the general SNCSTM of Section 5, generalise those described in Section 4. They reduce to those described in Section 4 when visits times are regular with $S_k = k$ and there is no effect modification (i.e. $Z_{k(l)} = 1$).

For the general SNCSTM of Section 5, which allows for irregular visit times, $\Delta_k(t)$, $\hat{e}_{k(l)}$ and $e^*_{k(l)}$ depend, in general, on the visit times $\bar{S}$, and Model $\mathcal{A}_k$ is the GLM $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T \geqslant S_k)\} = \alpha_{k0}^{\top} H_k$.

D.1 *Method 1*

In Method 1, $\psi_{k(l)}$ is estimated by $\hat{\psi}^{\text{M1}}_{k(l)} = -\hat{\alpha}_{k(l)}/\phi_k$, where $\hat{\alpha}_{k(l)}$ is the estimate of $\alpha_{k(l)}$ given by fitting GLM

$$g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, Q)\} = \alpha_{k0}^{\top} H_k + \sum_{j=k}^{l-1} \alpha_{k(j)}^{\top} Z_{k(j)}(S_{j+1} - S_j) + \alpha_{k(l)}^{\top} Z_{k(l)}(Q - S_l)) \tag{15}$$

to a set of pseudo-individuals, using weights $w_k(Q)$, where $w_k(t) = \prod_{j=k+1}^{K} \exp\{A_j v_j(t, Z_k, \bar{S})^{\top}\psi_j\}$. When visit times are regular, this set is the same as in Section 3. Otherwise the rule for constructing the set is a little more complicated and is given in Section D.4. Let $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S}, t)$ denote the fitted value of $E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, Q = t)$.

D.2 *Method 2*

In Method 2, $\psi_{k(l)}$ $(l \geqslant k)$ is estimated as the solution, $\hat{\psi}^{\mathrm{M2}}_{k(l)}$, to estimating equations

$$\sum_{i=1}^{n} Z_{k(l),i} \int_{S_{li}}^{S_{l+1,i}} R_i(t) w_{ki}(t)\ \Delta_{ki}(t) \times \left[dN_i(t) - \left\{ \sum_{j=k+1}^{l} A_{ji} \psi^{\top}_{j(l)} Z_{j(l),i} + \Delta_{ki}(S_{li})\ \psi^{\top}_{k(l)} Z_{k(l),i} \right\} dt \right] = 0, \quad (16)$$

where $\Delta_k(t)$ is replaced by $A_k - \hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S}, t)$. Let Model $\mathcal{B}_{k(l)}$ $(l \geqslant k)$ be defined by $E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_{k-1}, 0) \geqslant t\} = \big\{\gamma^{\top}_{k(l)} H_k - e^{*}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S}, k)\ \psi^{\top}_{k(l)} Z_{k(l)}\big\} dt$ for all $t \in [S_l, S_{l+1})$.

As proved in Web Appendix F, estimator $\hat{\psi}^{\mathrm{M2}}_{k(l)}$ is consistent under the conditions 1–3 stated in Section 4.2 plus the extra condition that, unless all of $\mathcal{A}_j$ $(j = k, \ldots, l)$ are correctly specified or $Z_{k(l)} = 1$, additional covariates $Z^{\mathrm{int}}_{k(l)} * H_k$ are included in each of the GLMs of equation (15). Here, $X * Y$ denotes all pairwise interactions between X and Y.

When $\mathcal{A}_k$ is correctly specified, the true parameter values for these additional covariates $Z^{\mathrm{int}}_{k(l)} * H_k$ are zero and they can be omitted. In the analysis of the Cystic Fibrosis registry data that allowed for an interaction between treatment and FEV$_1$%, described in Section 8, $Z^{\mathrm{int}}_{k(l)} * H_k$ was omitted because its inclusion caused instability in the estimates of $\hat{\psi}^{\mathrm{M2}}_{k(l)}$.

When there is no effect modification or modification depends only on L_0 (i.e. $Z_{k(l)}$'s depend at most on L_0), stabilised weights can be used and $A_{ji}\psi^{\top}_{j(l)} Z_{j(l),i}$ in equation (16) can be replaced by $\Delta^{*}_{j(k),i}\psi^{\top}_{j(l)} Z_{j(l),i}$, where $\Delta^{*}_{j(k)} = A_j - E(A_j \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T_i \geqslant S_j)$.

D.3 *Method 3*

Modifying Method 3 is simple when visit times are regular: the GLM of Section 4.3 is just replaced by $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k)\} = \alpha_{k0}(t)^\top H_k + \sum_{j=k}^{l} \alpha_{k(j)}(t)^\top Z_{k(j)}^{\text{int}}$. As with Method 2, double robustness requires $Z_{k(l)}^{\text{int}} * H_k$ be added as covariates.

D.4 *General rule for construction of the set of pseudo-individuals*

To estimate $\psi_{k(l)}$, Methods 1 and 2 involve fitting a GLM for A_k given $\bar{A}_{k-1}, \bar{L}_{k-1}$, $\bar{S}$ and Q to a set of pseudo-individuals. The rule for constructing this set when the follow-up visit times are regular and equal to $1, 2, \ldots, K$ was described in Section 4.1 of the article. Here we describe the more general rule (of which that is a special case), which can be used even when visit times are irregular.

For any $t \geqslant 0$, let $\mathcal{I}_{k(l)}(t)$ denote the set of individuals with $T \geqslant S_k + t$ and $S_l \leqslant S_k + t < S_{l+1}$, i.e. those who t units after their kth visit are still at risk and have had their lth visit but not yet their $(l+1)$th visit. Let $q_{k(l)}^{\min}$ and $q_{k(l)}^{\max}$ denote, respectively, the minimum and maximum values of t such that the set $\mathcal{I}_{k(l)}(t)$ is not empty. For each of some number (we used 10) of equally spaced values of t between $q_{k(l)}^{\min}$ and $q_{k(l)}^{\max}$ (viz. $q_{k(l)}^{\min}, q_{k(l)}^{\min} + (q_{k(l)}^{\max} - q_{k(l)}^{\min})/9, q_{k(l)}^{\min} + 2(q_{k(l)}^{\max} - q_{k(l)}^{\min})/9, \ldots, q_{k(l)}^{\max}$), take the set $\mathcal{I}_{k(l)}(t)$ and for each individual i in this set, create a pseudo-individual with $Q = S_{ki} + t$ and the same value of $(\bar{A}_K, \bar{L}_K, \bar{S})$ as individual i. Let $\mathcal{P}_{k(l)}$ denote the resulting set of (up to $10n$) pseudo-individuals.

Note that in the special case of regular visit times, $q_{k(l)}^{\min} = S_l - S_k$ and $q_{k(l)}^{\max} = S_{l+1} - S_k$ (assuming there are still individuals at risk at time S_{l+1}). Therefore, each pseudo-individual has a value of Q equal to one of S_l, $S_l + (S_{l+1} - S_l)/9$, ..., S_{l+1}.

E. Closed form of estimator for Methods 2 and 3

E.1 *Method 2*

The estimator corresponding to equation (16) is

$$\hat{\psi}_{k(l)} = \left[\sum_{i=1}^{n} Z_{k(l),i} Z_{k(l),i}^{\top} \Delta_{ki}(S_{li}) \int_{S_{li}}^{S_{l+1,i}} R_i(t) w_{ki}(t)\ \Delta_{ki}(t)\ dt\right]^{-1} \times \left[\sum_{i=1}^{n} Z_{k(l),i} \int_{S_{li}}^{S_{l+1,i}} R_i(t) w_{ki}(t)\ \Delta_{ki}(t) \times \left\{dN_i(t) - \sum_{j=k+1}^{l} A_{ji}\ \psi_{j(l)}^{\top} Z_{j(l),i}\ dt\right\}\right] \tag{17}$$

Now,

$$\begin{aligned}
&\int_{S_l}^{S_{l+1}} R(t) w_k(t)\ \Delta_k(t) \left\{dN(t) - \sum_{j=k+1}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)} dt\right\} \\
&= R(T) I(S_l \leqslant T < S_{l+1}) w_k(T) \Delta_k(T) dN(T) \\
&\quad - \sum_{j=k+1}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)} R(S_l) w_k(S_l) \\
&\quad\quad \times \int_0^{(T \wedge S_{l+1}) - S_l} \exp\{I(l>k)\ A_l \psi_{k(l)}^{\top} Z_{k(l)}\ t\} \\
&\quad\quad\quad \times \left\{A_k - g^{-1}\left(g[E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant S_l\}] + \alpha_{k(l)}^{\top} Z_{k(l)} t\right)\right\}\ dt
\end{aligned} \tag{18}$$

Note that the term $A_k - g^{-1}\left(g[E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant S_l\}] + \alpha_{k(l)}^{\top} Z_{k(l)} t\right)$ in equation (18) is just $\Delta_k(t)$. This is because it follows from equation (3) of the article that

$$\begin{aligned}
\Delta_k(t) &= A_k - E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant t\} \\
&= A_k - g^{-1}\left(g[E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant t\}]\right) \\
&= A_k - g^{-1}\left\{\alpha_{k0}^{\top} H_k + \alpha_k^{\top} v_k(t)\right\} \\
&= A_k - g^{-1}\left\{\alpha_{k0}^{\top} H_k + \alpha_k^{\top} v_k(S_l) + \alpha_{k(l)}^{\top} Z_{k(l)} t\right\} \\
&= A_k - g^{-1}\left\{g[E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant S_l\}] + \alpha_{k(l)}^{\top} Z_{k(l)} t\right\}.
\end{aligned}$$

So, it is evident from equation (18) that we need to calculate integrals of the form

$$\int_0^y \exp(Bt)\left\{A_k - g^{-1}(E + Dt)\right\}\, dt \tag{19}$$

where, more specifically,

$$\begin{aligned}
y &= (T \wedge S_{l+1}) - S_l \\
B &= \begin{cases} A_l \psi_{k(l)}^\top Z_{k(l)} & \text{if} \quad l > k \\ 0 & \text{if} \quad l = k \end{cases} \\
E &= g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant S_l)\} \\
D &= \alpha_{k(l)}^\top Z_{k(l)}
\end{aligned}$$

When g is the identify link function, expression (19) becomes

$$\begin{aligned}
&\int_0^y \exp(Bt)(A_k - E - Dt)\, dt \\
&= \int_0^y \exp(Bt)\{\Delta_k(S_l) - Dt\}\, dt \\
= &\begin{cases} B^{-1}\exp(By)\{\Delta_k(S_l) - Dy\} - B^{-1}\Delta_k(S_l) + B^{-2}D\{\exp(By) - 1\} & \text{if } B \neq 0 \\ \Delta_k(S_l)y - Dy^2/2 & \text{if } B = 0 \end{cases}
\end{aligned}$$

When g is the logit link function, expression (19) becomes

$$\begin{aligned}
&\int_0^y \exp(Bt)\left\{A_k - \frac{\exp(E + Dt)}{1 + \exp(E + Dt)}\right\}\, dt \\
&= \begin{cases} A_k B^{-1}\{\exp(By) - 1\} - \int_0^y \frac{\exp\{E+Ft\}}{1+\exp(E+Dt)}\, dt & \text{if } B \neq 0 \\ A_k y - \int_0^y \frac{\exp\{E+Ft\}}{1+\exp(E+Ft)}\, dt & \text{if } B = 0 \end{cases}
\end{aligned}$$

where $F = B + D$.

In the special case where $B = 0$ (and so $F = D$) and $D \neq 0$,

$$\begin{aligned}
&\int_0^y \frac{\exp(E + Ft)}{1 + \exp(E + Dt)}\, dt \\
&= \int_0^y \frac{\exp(E + Dt)}{1 + \exp(E + Dt)}\, dt \\
&= D^{-1}\left[\log\{1 + \exp(E + Dy)\} - \log\{1 + \exp(E)\}\right]
\end{aligned}$$

When $F = 0$ and $D \neq 0$,

$$\int_0^y \frac{\exp(E + Ft)}{1 + \exp(E + Dt)}\,dt$$
$$= \exp(E)\int_0^y 1 - \frac{\exp(E + Dt)}{1 + \exp(E + Dt)}\,dt$$
$$= y\exp(E) - \exp(E)D^{-1}\left[\log\{1 + \exp(E + Dy)\} - \log\{1 + \exp(E)\}\right]$$

When $D = 0$ and $F \neq 0$,

$$\int_0^y \frac{\exp(E + Ft)}{1 + \exp(E + Dt)}\,dt = \frac{\exp(E)}{F\{1 + \exp(E)\}}\{\exp(Fy) - 1\}$$

When $F = D = 0$,

$$\int_0^y \frac{\exp(E + Ft)}{1 + \exp(E + Dt)}\,dt = \frac{\exp(E)}{1 + \exp(E)}y$$

When $F \neq 0$ and $D \neq 0$, numerical integration can be used.

E.2 *Method 3*

Rewrite the estimator of expression (17) as

$$\begin{aligned}\hat{\psi}_{k(l)} = &\left[\sum_{i=1}^{n} Z_{k(l),i}Z_{k(l),i}^{\top}\Delta_{ki}(S_{li})\right.\\ &\left.\times\int_0^{S_{l+1,i}-S_{li}} R_i(S_{li}+t)w_{ki}(S_{li}+t)\;\Delta_{ki}(S_{li}+t)\;dt\right]^{-1}\\ &\times\left[\sum_{i=1}^{n} Z_{k(l),i}\int_0^{S_{l+1,i}-S_{li}} R_i(S_{li}+t)w_{ki}(S_{li}+t)\;\Delta_{ki}(S_{li}+t)\right.\\ &\quad\left.\times\left\{dN_i(S_{li}+t) - \sum_{j=k+1}^{l} A_{ji}\;\psi_{j(l)}^{\top}Z_{j(l),i}\;dt\right\}\right]\end{aligned} \tag{20}$$

and rewrite $w_k(S_l + t)$ for $l > k$ and $t \in [0, S_{l+1} - S_l)$ as

$$w_k(S_l + t) = \exp\left\{\sum_{j=k+1}^{l-1} A_j \sum_{m=j}^{l-1}\psi_{j(m)}^{\top}Z_{j(m)}(S_{m+1} - S_m) + t\sum_{j=k+1}^{l} A_j\psi_{j(l)}^{\top}Z_{j(l)}\right\},$$

with $w_k(S_k + t) = 1$ and $t \in [0, S_{k+1} - S_k)$.

In Method 3, we fit a separate model for $E(A_{ki} \mid \bar{A}_{k-1,i}, \bar{L}_{ki}, \bar{S}_i, T(\bar{A}_{ki}, 0) \geqslant S_{ki} + t)$ for each value t at which the set $\mathcal{I}_k(t)$ (defined in Web Appendix D.4) changes. Henceforth we assume that visit times are regular. Then the set $\mathcal{I}_k(t)$ changes when one of the individuals fails, is censored or has their $(l+1)$th exposure at time

$S_k + t$, i.e. when $T \wedge S_{l+1} = S_k + t$. Let $e_0 = 0$ and let $\{e_1, \ldots, e_{Q_{kl}}\}$ denote the set of distinct values of $(T_i \wedge S_{l+1}) - S_l$ that are greater than or equal to zero. So, the fitted value of $E(A_{ki} \mid \bar{A}_{k-1,i}, \bar{L}_{ki}, \bar{S}_i, T(\bar{A}_{ki}, 0) \geqslant S_l + t)$ used for $\Delta_{ki}(S_l + t)$ in equation (20) is constant over each interval $t \in [e_q, e_{q+1})$ $(q = 0, \ldots, Q_{kl} - 1)$. Consequently, we can write, for $l = k$,

$$\begin{aligned}
&\int_0^{S_{l+1}-S_l} R(S_l + t) w_k(S_l + t)\ \Delta_k(S_l + t)\ dt \\
&= \sum_{q=0}^{Q_{kl}-1} R(S_l + e_q)\ \Delta_k(S_l + e_q)(e_{q+1} - e_q)
\end{aligned}$$

and for $k > l$,

$$\begin{aligned}
&\int_0^{S_{l+1}-S_l} R(S_l + t) w_k(S_l + t)\ \Delta_k(S_l + t)\ dt \\
&= \sum_{q=0}^{Q_{kl}-1} R(S_l + e_q)\ \Delta_k(S_l + e_q) \int_{e_q}^{e_{q+1}} w_k(S_l + t) dt \\
&= \sum_{q=0}^{Q_{kl}-1} R(S_l + e_q)\ \Delta_k(S_l + e_q) \\
&\quad \times \exp\left\{ \sum_{j=k+1}^{l-1} A_j \sum_{m=j}^{l-1} \psi_{j(m)}^\top Z_{j(m)} (S_{m+1} - S_m) \right\} \\
&\quad \times \int_{e_q}^{e_{q+1}} \exp\left\{ s \sum_{j=k+1}^{l} A_j \psi_{j(l)}^\top Z_{j(l)} \right\} ds
\end{aligned} \tag{21}$$

If $\sum_{j=k+1}^{l} A_j \psi_{j(l)}^\top Z_{j(l)} \neq 0$, then expression (21) equals

$$\begin{aligned}
&\exp\left\{ \sum_{j=k+1}^{l-1} A_j \sum_{m=j}^{l-1} \psi_{j(m)}^\top Z_{j(m)} (S_{m+1} - S_m) \right\} \left\{ \sum_{j=k+1}^{l} A_j \psi_{j(l)}^\top Z_{j(l)} \right\}^{-1} \\
&\times \sum_{q=0}^{Q_{kl}-1} R(S_l + e_q)\ \Delta_k(S_l + e_q) \\
&\times \left[\exp\left\{ e_{q+1} \sum_{j=k+1}^{l} A_j \psi_{j(l)}^\top Z_{j(l)} \right\} - \exp\left\{ e_q \sum_{j=k+1}^{l} A_j \psi_{j(l)}^\top Z_{j(l)} \right\} \right]
\end{aligned}$$

On the other hand, if $\sum_{j=k+1}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)} = 0$, then expression (21) equals

$$\exp\left\{ \sum_{j=k+1}^{l-1} A_j \sum_{m=j}^{l-1} \psi_{j(m)}^{\top} Z_{j(m)} (S_{m+1} - S_m) \right\}$$
$$\times \sum_{q=0}^{Q_{kl}-1} R(S_l + e_q)\, \Delta_k(S_l + e_q)(e_{q+1} - e_q)$$

F. Proof of double robustness of Methods 2 and 3

The basic results that justify Methods 2 and 3 are equation (9) and

$$E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_{k-1}, 0) \geqslant t\}$$
$$= \frac{E\left[R(t)w_k(t)\left\{dN(t) - \sum_{j=k}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)} dt\right\} \mid \bar{A}_k, \bar{L}_k, \bar{S}, T \geqslant S_k\right]}{E\left\{R(t)w_k(t) \mid \bar{A}_k, \bar{L}_k, \bar{S}, T \geqslant S_k\right\}} \quad (22)$$

for $l \geqslant k$ and $t \in [S_l, S_{l+1})$. We show below that equation (22) is implied by equation (9) and Models $\mathcal{M}_k, \ldots, \mathcal{M}_l$. Equation (22) means that within a stratum of the population defined by $(\bar{A}_k, \bar{L}_k, \bar{S})$ and by $T(\bar{A}_{k-1}, 0) \geqslant S_k$ (or equivalently, $T \geqslant S_k$) the counterfactual hazard when $A_k, \ldots, A_K$ are set to zero is equal to the actual hazard minus $\sum_{j=k}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)}$ after weighting individuals by $w_k(t)$. Note that, since the left-hand side of equation (22) does not depend on A_k (because of the no unmeasured confounders assumption), neither can the right-hand side.

We now prove equation (22). Again, we omit the explicit conditioning on $\bar{S}$. By taking logs of both sides of equation (9) and differentiating with respect to t and multiplying both sides by minus one, we obtain

$$E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\}$$
$$= E\left(R(t) \exp\left\{ \sum_{j=k}^{K} A_j G_j(t) \right\} \left[dN(t) - \frac{d}{dt}\left\{ \sum_{j=k}^{K} A_j G_j(t) \right\} dt \right] \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k \right)$$

It follows from this and equation (9) that

$$
\begin{aligned}
&E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\} \\
&= \frac{E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\}}{P\{T(\bar{A}_{k-1},0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant S_k\}} \\
&= \frac{E\left(R(t) \exp\left\{\sum_{j=k}^K A_j G_j(t)\right\} \left[dN(t) - \frac{d}{dt}\left\{\sum_{j=k}^K A_j G_j(t)\right\} dt\right] \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right)}{E\left\{R(t) \exp\left\{\sum_{j=k}^K A_j G_j(t)\right\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right\}} \\
&= \frac{E\left(R(t) \exp\left\{\sum_{j=k+1}^K A_j G_j(t)\right\} \left[dN(t) - \frac{d}{dt}\left\{\sum_{j=k}^K A_j G_j(t)\right\} dt\right] \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right)}{E\left\{R(t) \exp\left\{\sum_{j=k+1}^K A_j G_j(t)\right\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right\}} \\
&= \frac{E\left(R(t) w_k(t) \left[dN(t) - \frac{d}{dt}\left\{\sum_{j=k}^K A_j G_j(t)\right\} dt\right] \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right)}{E\left\{R(t) w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right\}}
\end{aligned}
$$

For $t \in [S_l, S_{l+1})$, $\frac{d}{dt}\left\{\sum_{j=k}^K A_j G_j(t)\right\} = \sum_{j=k}^l A_j \psi_{j(l)}^\top Z_{j(l)}$. Hence, equation (22) holds.

We now use this result to prove consistency of estimation for Method 2. Suppose that $\psi_{j(m)}$ ($k < j \leqslant m \leqslant l$) have already been consistently estimated by Method 2 and we are now estimating $\psi_{k(l)}$. Assume that Models $\mathcal{M}_k, \ldots, \mathcal{M}_l$ are correctly specified.

First, consider the following expression and suppose that $\mathcal{A}_k$ is correctly specified.

$$\int_{S_l}^{S_{l+1}} R(t) w_k(t)\ \Delta_k(t) \left\{dN(t) - \sum_{j=k}^l A_j \psi_{j(l)}^\top Z_{j(l)}\right\} dt \tag{23}$$

This is the same as the estimating function we use — i.e. the ith element of the left-hand side of equation (7) — (apart from the $Z_{k(l)}$ term) but with $\Delta_k(S_l)$ replaced by A_k.

Now, for any $t \in [S_l, S_{l+1})$, and using equation (22), we have

$$
\begin{aligned}
&E\left[R(t) w_k(t)\ \Delta_k(t) \left\{dN(t) - \sum_{j=k}^l A_j \psi_{j(l)}^\top Z_{j(l)}\right\} \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\right] \\
&= \Delta_k(t) E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\} \times E\{R(t) w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \\
&= \Delta_k(t) E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1},0) \geqslant t\} \times E\{R(t) w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\}
\end{aligned}
$$

Now take the expectation over A_k given $\bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k$. This yields

$$
\begin{aligned}
&E\left[R(t)w_k(t)\ \Delta_k(t)\left\{dN(t) - \sum_{j=k}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)}\right\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k\right] \\
&= E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\} \qquad (24) \\
&\qquad \times E[\Delta_k(t) E\{R(t)w_k(t) \mid \bar{A}_k, \bar{L}_k, T \geqslant S_k\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k] \\
&= E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\} \\
&\qquad \times E[\Delta_k(t) P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k] \\
&\qquad\qquad (25)
\end{aligned}
$$

Line (25) follows from equation (4).

If $\mathcal{A}_k$ is correctly specified,

$$
\begin{aligned}
&E\left[\Delta_k(t) P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k\right] \\
&\equiv E[\{A_k - E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)\} \\
&\quad \times P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k] \\
&= P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \\
&\quad \times \{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t) - E(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)\} \qquad (26) \\
&= 0 \qquad (27)
\end{aligned}
$$

Equation (26) follows because, for a general random variable X and a general event A, $E\{X\ P(A \mid X)\} = P(A)E(X \mid A)$. In conclusion, we have shown that equation (23) has expectation zero. Now, if we replace A_k in equation (23) by $\Delta_i(S_l)$, then we are simply adding a function of $(\bar{A}_{k-1}, \bar{L}_k)$ multiplied by $R(t)w_k(t)\Delta_k(t)$. It is straightforward to show that this extra term has expectation zero when $\mathcal{A}_k$ is correctly specified. This is because

$$
\begin{aligned}
&E\{R(t)w_k(t)\Delta_k(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \\
&= \Delta_k(t)\ E\{R(t)w_k(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \\
&= \Delta_k(t)\ P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\}
\end{aligned}
$$

using equation (4). Now, taking the conditional expectation of this over A_k given $(A_{k-1}, \bar{L}_k)$ and $T(\bar{A}_k, 0) \geqslant S_k$, we obtain

$$
\begin{aligned}
&E\{R(t)w_k(t)\Delta_k(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \\
&= E\left[\Delta_k(t)P\{T(\bar{A}_k, 0) \geqslant t \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant S_k\} \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant S_k\right],
\end{aligned}
$$

which we know, from equation (27), equals zero.

Likewise, when the $Z_{k(l)}$'s are functions only of L_0, changing A_j to $\Delta^{\dagger}_{j(k)} = A_j - E^{\dagger}(A_j \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T \geqslant S_k)$ $(j = k+1, \ldots, l)$ in the estimating function (23), where $E^{\dagger}(A_j \mid \bar{A}_{k-1}, \bar{L}_k, \bar{S}, T \geqslant S_k)$ denotes the limiting fitted value from Model $\mathcal{C}_{j(k)}$, simply adds a function of $(\bar{A}_{k-1}, \bar{L}_k)$ multiplied by $R(t)w_k(t)\Delta_k(t)$, and so the same is true. Finally, if the $Z_{k(l)}$'s are functions only of L_0 and we replace $w_k(t)$ by $w^*_k(t)$, we are simply multiplying the estimating function (23) by a function of $(\bar{A}_{k-1}, \bar{L}_k)$, and so it still has conditional expectation zero given $(\bar{A}_{k-1}, \bar{L}_k)$ and $T(\bar{A}_k, 0) \geqslant S_k$.

Now suppose that $\mathcal{A}_k$ may not be correctly specified but $\mathcal{B}_{k(l)}$ is correctly specified. Consider the following estimating equations for $\psi_{k(l)}$:

$$
\begin{aligned}
&\sum_{i=1}^{n} Z_{k(l),i} \int_{S_{li}}^{S_{l+1,i}} R_i(t)w_{ki}(t)\ \Delta_{ki}(t) \\
&\quad \times \left[dN_i(t) - \left\{\gamma_{k(l)}^{\top}\ h_k(\bar{A}_{k-1,i}, \bar{L}_{ki}, \bar{S}_i) + \sum_{j=k+1}^{l} A_{ji}\psi_{j(l)}^{\top} Z_{j(l),i} + \Delta_{ki}(S_{li})\psi_{k(l)}^{\top} Z_{k(l),i}\right\} dt\right] \\
&\qquad = 0
\end{aligned} \tag{28}
$$

for any given value of $\gamma_{k(l)}$, and with $\Delta_k(t)$ replaced by its estimate obtained as described in Section 4.2 of the article. These are the same as the estimating equations we use (i.e. equation (16)) except that they include the extra term $\gamma_{k(l)}^{\top} H_k$.

For any $t \in [S_l, S_{l+1})$, we have, using equation (22)

$$\begin{aligned}
& E\Big(R(t)w_k(t)\,\Delta_k(t)\,\Big[dN(t) - \Big\{\gamma_{k(l)}^{\top}\,h_k(\bar{A}_{k-1}, \bar{L}_k, \bar{S}) \\
&\qquad + \sum_{j=k+1}^{l} A_j \psi_{j(l)}^{\top} Z_{j(l)} + \Delta_k(S_l)\psi_{k(l)}^{\top} Z_{k(l)}\Big\}\,dt\Big] \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\Big) \\
&= \big[E\{dN_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\} - \gamma_{k(l)}^{\top}\,h_k(\bar{A}_{k-1}, \bar{L}_k, \bar{S}) \\
&\quad + E\{A_k \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\}\psi_{k(l)}^{\top} Z_{k(l)}\big] \\
&\quad \times E\{R(t)w_k(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant S_k\} \times \Delta_k(t)
\end{aligned}$$

which equals zero when Model $\mathcal{B}_{k(l)}$ holds. Hence, equations (28) are unbiased estimating equations for $\psi_{k(l)}$ when $\mathcal{B}_{k(l)}$ is correctly specified and $\gamma_{k(l)}$ equals its true value.

Finally, we shall show that, when either $Z_{k(l)} = 1$ or $Z_{k(l)}^{\text{int}} * H_k$ is included in GLM (15), then, because of the way that $E(A_{ki} \mid \bar{A}_{k-1,i}, \bar{L}_{ki}, T_i(\bar{A}_k, 0) \geqslant t)$ in $\Delta_{ki}(t)$ is estimated, we have

$$\sum_{i=1}^{n} Z_{k(l),i} \int_{S_{li}}^{S_{l+1,i}} R_i(t)w_{ki}(t)\,\Delta_{ki}(t) \times \gamma_{k(l)}^{\top} H_k\,dt = 0 \tag{29}$$

for any value of $\gamma_{k(l)}$, regardless of whether Model $\mathcal{A}_k$ or Model $\mathcal{B}_{k(l)}$ is correctly specified. This means that equations (28) reduce to equations (16), and thus our estimating equations (16) have expectation zero when either Model $\mathcal{A}_k$ or Model $\mathcal{B}_{k(l)}$ is correctly specified.

The reason why equation (29) holds is as follows. Let X denote any element of the vector $(H_k^\top, Z_{k(l)}^{\text{int}} * H_k^\top)^\top$. Then

$$
\begin{aligned}
&\sum_{i=1}^{n} \int_{S_{li}}^{S_{l+1,i}} R_i(t) w_{ki}(t)\ \Delta_{ki}(t)\ X_i\ dt \\
&= \sum_{i=1}^{n} \int_{S_{li}-S_{ki}}^{S_{l+1,i}-S_{ki}} R_i(S_{ki}+t) w_{ki}(S_{ki}+t)\ \Delta_{ki}(S_{ki}+t)\ X_i\ dt \\
&= \sum_{i=1}^{n} \int_{0}^{\infty} I(S_{li} \leqslant S_{ki}+t < S_{k+1,i}) R_i(S_{ki}+t) w_{ki}(S_{ki}+t)\ \Delta_{ki}(S_{ki}+t)\ X_i\ dt \\
&= \lim_{\delta \to \infty} \sum_{t \in \left\{ q_{k(l)}^{\min},\ q_{k(l)}^{\min} + (q_{k(l)}^{\max} - q_{k(l)}^{\min})/\delta,\ q_{k(l)}^{\min} + 2(q_{k(l)}^{\max} - q_{k(l)}^{\min})/\delta, \ldots,\ q_{k(l)}^{\max} \right\}} \sum_{i=1}^{n} I(S_{li} \leqslant S_{ki}+t < S_{k+1,i}) \\
&\qquad \times R_i(S_{ki}+t) w_{ki}(S_{ki}+t)\ \Delta_{ki}(S_{ki}+t)\ X_i \times \frac{q_{k(l)}^{\max} - q_{k(l)}^{\min}}{\delta} \\
&\approx \sum_{t \in \left\{ q_{k(l)}^{\min},\ q_{k(l)}^{\min} + (q_{k(l)}^{\max} - q_{k(l)}^{\min})/9,\ q_{k(l)}^{\min} + 2(q_{k(l)}^{\max} - q_{k(l)}^{\min})/9, \ldots,\ q_{k(l)}^{\max} \right\}} \sum_{i=1}^{n} I(S_{li} \leqslant S_{ki}+t < S_{k+1,i}) \\
&\qquad \times R_i(S_{ki}+t) w_{ki}(S_{ki}+t)\ \Delta_{ki}(S_{ki}+t)\ X_i \times \frac{q_{k(l)}^{\max} - q_{k(l)}^{\min}}{9}. \qquad (30)
\end{aligned}
$$

Expression (30) equals zero because it is one element of the score function vector (multiplied by $(q_{k(l)}^{\max} - q_{k(l)}^{\min})/9$) for the GLM of equation (15) fitted to the set $\mathcal{P}_{k(l)}$ with weights $w_k(Q)$. (The set $\mathcal{P}_{k(l)}$ was defined in Web Appendix D.4.)

The proof of double robustness of Method 3 is similar. The difference lies in the way that fitted values of $E(A_{ki} \mid \bar{A}_{k-1,i}, \bar{L}_{ki}, \bar{S}_i, T_i(\bar{A}_{ki}, 0) \geqslant t)$ are estimated. Suppose for simplicity that $Z_{k(l)} = 1$ and visit times are regular. For Method 3, the score equations corresponding to GLM $g\{E(A_k \mid \bar{A}_{k-1}, \bar{L}_k\} = \alpha_{k0}(t)^\top H_k$ fitted to $\mathcal{I}_k(t)$ using weights $w_k(S_k + t)$ are

$$
\sum_{i=1}^{n} R_i(S_{ki}+t) I(S_l \leqslant S_k + t < S_{l+1}) w_{ki}(S_k+t) \Delta_{ki}(S_k+t) H_{ki} = 0 \qquad (31)
$$

(The set $\mathcal{I}_k(t)$ was defined in Web Appendix D.4.) Hence, equation (31) holds for any $t \geqslant 0$. Consequently, when Method 3 is used,

$$\begin{aligned}
&\sum_{i=1}^{n} \int_{S_l}^{S_{l+1}} R_i(t) w_{ki}(t)\, \Delta_{ki}(t)\, \gamma_{k(l)}(t - S_{li})^\top H_{ki}\, dt \\
&= \sum_{i=1}^{n} \int_{0}^{\infty} R_i(S_{ki} + t) I(S_l \leqslant S_k + t < S_{l+1}) w_{ki}(S_k + t)\, \Delta_{ki}(S_k + t) \\
&\qquad \times \gamma_{k(l)}(t + S_k - S_l)^\top H_{ki}\, dt \\
&= \gamma_{k(l)}(t + S_k - S_l)^\top \int_{0}^{\infty} \sum_{i=1}^{n} R_i(S_{ki} + t) I(S_l \leqslant S_k + t < S_{l+1}) \\
&\qquad \times w_{ki}(S_k + t)\, \Delta_{ki}(S_k + t) H_{ki}\, dt \\
&= 0.
\end{aligned}$$

G. Constraining exposure effects

In this web appendix we explain how estimation of parameters can be performed under the constraint that $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ for all k, k', m.

For Method 1, a simple way to estimate $\psi_{k(k+m)}$ under this constraint is to calculate $\hat{\psi}^{\mathrm{M1}}_{k(k+m)}$ $(k = 0, \dots, K - m)$ separately as before and then, for each element of the vector $\hat{\psi}^{\mathrm{M1}}_{k(k+m)}$, calculate a weighted average of these $K-m+1$ estimates. Suitable weights are the reciprocals of the corresponding $K - m + 1$ variances estimated using a sandwich variance estimator that accounts for duplication of individuals as pseudo-individuals. It can be calculated using standard software and is a valid estimator of the variance of $\hat{\psi}^{\mathrm{M1}}_{k(k)}$, but not of $\hat{\psi}^{\mathrm{M1}}_{k(k+m)}$ when $m > 0$, because it ignores the uncertainty in $w_k(t)$ arising from estimating the ψ's. Nonetheless, it suffices for the purpose of averaging the $K - m + 1$ estimates of $\psi_{k(k+m)}$. For Methods 2 and 3, we simply sum the $K - m + 1$ estimating equations (7) for $\psi_{k(k+m)}$ $(k = 0, \dots, K - m)$ and solve the resulting single equation.

In Section 3, we assumed the regular visit times are $0, 1, \dots, K$, rescaling the time variable if necessary. Such rescaling may make it more reasonable to constrain

$\psi_{k(k+m)}$ to be a known multiple of $\psi_{k'(k'+m)}$. For example, if visits 1 and 2 are one and 13 months, respectively, after baseline, then $\psi_{0(0)}$ and $\psi_{1(1)}$ are measured in units of per-month and per-year, respectively. Since $\psi_{0(0)}$ and $\psi_{1(1)}/12$ are measured in the same units (per-month), one might constrain $\psi_{0(0)} = \psi_{1(1)}/12$. This requires only minor modification of the above procedures. However, using the more general SNCSTM described in Section 5 avoids the need to rescale time.

We have only considered one form of constraint on the exposure effects; Vansteelandt and Sjolander (2016) show how to impose other forms.

H. Inverse probability of censoring weighting

Assume that the first condition in Section 6 holds, namely, that

$$E\{dN_C(t) \mid C \geqslant t, \bar{A}_{\lfloor \tilde{T} \rfloor}, \bar{L}_{\lfloor \tilde{T} \rfloor}, \bar{S}, \tilde{T} > t, \tilde{T}\} = \lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}). \tag{32}$$

Let $w_k^C(t) = \exp\left\{\int_{S_k}^t \lambda(s, \bar{A}_{\lfloor s \rfloor}, \bar{L}_{\lfloor s \rfloor}, \bar{S})\, ds\right\}$. This is the inverse probability of remaining uncensored at time t. A parametric model for $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})$ is specified and its parameters (and hence $w_k^C(t)$) are estimated from the data. Now Methods 1–3 can be used with the weights $w_k(t)$ replaced by $w_k(t) \times w_k^C(t)$. If the assumptions sufficient for consistency in the absence of censoring (see Section 4) are satisfied, and equation (32) holds, and the model for $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})$ is correctly specified, then the resulting estimates of $\psi_{k(l)}$ are consistent.

More stable weights can be obtained by specifying, for each $k = 0, \ldots, K$, an additional parametric model for $\lambda_k(s, \bar{A}_{k-1}, \bar{L}_k, \bar{S}) = E\{dN_C(t) \mid C \geqslant t, \bar{A}_{k-1}, \bar{L}_k, \bar{S}, \tilde{T} > t\}$ $(t > S_k)$. This differs from the previous model in that it is conditional only on $(\bar{A}_{k-1}, \bar{L}_k)$. After estimating the parameters of this model, $w_k^C(t)$ is replaced by $w_k^{CS}(t) = \exp\left\{\int_{S_k}^t \lambda(s, \bar{A}_{\lfloor s \rfloor}, \bar{L}_{\lfloor s \rfloor}, \bar{S})\, ds \;- \int_{S_k}^t \lambda_k(s, \bar{A}_{k-1}, \bar{L}_k, \bar{S})\, ds\right\}$. Note that misspecification of this additional model does not affect consistency of the estimator of $\psi_{k(l)}$, and that if $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}) = \lambda(t, \bar{A}_{k-1}, \bar{L}_k, \bar{S})$ for all $t \in [S_k, S_{l+1})$, then $w_k^{CS}(t) = 1$ for $t \in [S_k, S_{l+1})$, i.e. no censoring weights are needed when estimating $\psi_{k(l)}$.

The formulae given in Web Appendix E for Method 2 are easily extended to handle inverse probability of censoring weighting, provided that the parametric models for censoring are proportional hazards models with a constant baseline hazard between visits. We now explain how this is done.

Above, we referred to two parametric models, one for $\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S})$ and an additional one for $\lambda_k(t, \bar{A}_{k-1}, \bar{L}_k, \bar{S})$, and said that the stabilised inverse probability of censoring weights are estimates of

$$w^{CS}(t) = \exp\left\{ \int_{S_{ki}}^{t} \lambda(s, \bar{A}_{\lfloor s \rfloor, i}, \bar{L}_{\lfloor s \rfloor, i}, \bar{S}_i)\, ds - \int_{S_{ki}}^{t} \lambda_k(s, \bar{A}_{k-1,i}, \bar{L}_{ki}, \bar{S}_i)\, ds \right\}.$$

(If unstabilised weights are used, $\lambda_k(s, \bar{A}_{k-1}, \bar{L}_k, \bar{S})$ is simply replaced by zero.)

For the first parametric model, we assume

$$\lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}) = \exp\{\beta_k^\top\, b_k(\bar{A}_k, \bar{L}_k, \bar{S})\}$$

for $t \in [S_k, S_{k+1})$, where $b_k(\bar{A}_k, \bar{L}_k, \bar{S})$ is a known vector function of $(\bar{A}_k, \bar{L}_k, \bar{S})$ whose first element equals one (this is an intercept term), and β_k is an unknown vector parameter. For the second parametric model, we assume

$$\lambda_k(t, \bar{A}_{k-1}, \bar{L}_k, \bar{S}) = \exp\{\beta_{k(l)}^\top\, b_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S})\}$$

for $t \in [S_l, S_{l+1})$, where $b_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S})$ is a known vector function of $(\bar{A}_{k-1}, \bar{L}_k, \bar{S})$ whose first element equals one, and $\beta_{k(l)}$ is an unknown vector parameter. Abbreviate $b_k(\bar{A}_k, \bar{L}_k, \bar{S})$ as b_k and $b_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, \bar{S})$ as $b_{k(l)}$. The inverse probability of censoring weight at time $t \in [S_l, S_{l+1})$ when estimating $\psi_{k(l)}$ is now

$$\begin{aligned} w_{k(l)}^{CS}(t) &= \prod_{j=k}^{l-1} \exp\left\{ \left(\beta_j^\top b_j - \beta_{k(j)}^\top b_{k(j)}\right) \times (S_{j+1} - S_j) \right\} \\ &\quad \times \exp\left\{ \left(\beta_l^\top b_l - \beta_{k(l)}^\top b_{k(j)}\right) \times (t - S_l) \right\} \end{aligned}$$

In particular,

$$w^{CS}_{k(l)}(S_l) = \prod_{j=k}^{l-1} \exp\left\{\left(\beta_j^\top b_j - \beta_{k(j)}^\top b_{k(j)}\right) \times (S_{j+1} - S_j)\right\}.$$

So, in equation (18), we should replace $w_k(T)$ by $w_k(T) \times w^{CS}_{k(l)}(T)$, replace $w_k(S_l)$ by $w_k(S_l) \times w^{CS}_{k(l)}(S_l)$, and replace $\exp\{I(l > k)\ A_l \psi_{k(l)}^\top Z_{k(l)}\ t\}$ by

$$\exp\left[\left\{I(l > k)\ A_l \psi_{k(l)}^\top Z_{k(l)} + (\beta_l^\top b_l - \beta_{k(l)}^\top b_{k(l)})\right\}\ t\right].$$

Similar modifications can be made to Method 3.

I. Software

Our R function *sncstm* can be used to apply Methods 1–3. Note that Method 3 is only implemented for regular visits and without inverse probability of censoring weights.

Two examples of the use of *sncstm* are provided in the files 'example1.r' and 'example2.r'.

The compulsory arguments of *sncstm* are as follows:

data : A data frame containing the following elements (here n is the number of individuals, $K+1$ is the number of visits, and p is the number of variables that are confounders in at least one of Models $\mathcal{A}_0, \ldots, \mathcal{A}_K$):

- tim — n-vector containing the failure or censoring time for each individual
- fail — n-vector containing the failure/censoring indicator for each individual (equals TRUE if fails and FALSE if censored)
- tau — $n \times (K+1)$ matrix whose $(i, k+1)$th entry is S_{ki}, the kth visit time for individual i (note that all the entries in the first column should equal zero, because $S_{0i} = 0$)

- A — $n \times (K+1)$ matrix whose ith row equals $(A_{0i}, A_{1i}, \ldots, A_{Ki})$, the treatments for individual i
- L — $n \times p$ matrix whose ith row contains the values for individual i of the p variables that are confounders in at least one of Models $\mathcal{A}_0, \ldots, \mathcal{A}_K$
- Z (optional) — if the optional argument useZ (see below) is not specified, then Z should not be specified either, but if useZ is specified, then Z should be a matrix; see Example 2 for details of how to specify Z when useZ is specified

useA : $(K+1)\times(K+1)$ matrix whose $(k+1)$th row indicates which of $A_0, A_1, \ldots, A_{k-1}$ to include as covariates in Model $\mathcal{A}_k$. If the $(k+1, j+1)$th element of useA equals TRUE, A_j is included in $\mathcal{A}_k$. If this element of useA equals FALSE, A_j is not included.

useL : $(K+1) \times p$ matrix whose $(k+1)$th row indicates which of the p confounders to include as covariates in Model $\mathcal{A}_k$. If the $(k+1, j+1)$th element of useL equals TRUE, the jth of the p confounders is included in $\mathcal{A}_k$. If this element of useL equals FALSE, the jth confounder is not included.

EXPOSETYPE : 'gaussian' if Model $\mathcal{A}_k$ is a linear regression; 'binomial' if $\mathcal{A}_k$ is a logistic regression.

The optional arguments of the *sncstm* function are as follows ('by default' means if the argument is not specified):

METHOD : Indicates which of the three estimation methods described in our article should be used to estimate the $\psi_{k(l)}$ parameters. By default, this equals 2, meaning that Method 2 is used. To use one of the other methods, set METHOD equal to 1 or 3.

CONSTRAIN : By default, this equals FALSE. If it equals TRUE, the constraint that $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ for all k, k', m is imposed.

NQUAD : When visit times are regular, this is the number of equally spaced values of t between S_k and S_{k+1} at which a pseudo-individual is created from each individual still at risk. When visit times are irregular, it is the number of equally spaced values of t between $q^{\min}_{k(l)}$ and $q^{\max}_{k(l)}$ (see Web Appendix D.4). By default, this equals 10, which is the number used for the main simulation study reported in our article. If the probability of failure between consecutive visits is large (e.g. $> 10\%$ of individuals still at risk at visit k fail before visit $k+1$) or if visit times are highly irregular (so that $q^{\max}_{k(l)} - q^{\min}_{k(l)}$ is very large), it may be desirable to increase NQUAD. The aim should be to choose NQUAD to be large enough such that any further increase in NQUAD has little impact on the estimates of $\psi_{k(l)}$. The NQUAD argument is ignored if METHOD=3.

STABILISE : By default, this equals FALSE, meaning that unstabilised weights $w_k(t)$ are used. When useZ (see below) is not specified (so that the SNCSTM assumes there is no effect modification), STABILISE can be set equal to TRUE, meaning that stabilised weights $w^*_k(t)$ are used instead of $w_k(t)$.

IPCW : By default, this equals FALSE, meaning there is no inverse probability of censoring weighting. Specify IPCW=TRUE to use inverse probability of censoring weighting.

useAcensor : If IPCW=TRUE, useAcensor is a $(K+1) \times (K+1)$ matrix, whose $(k+1, j+1)$th entry equals TRUE if A_j is included in the model for the hazard of censoring during time interval $[S_k, S_{k+1})$, and equals FALSE if it is not included. Ignored if IPCW=FALSE.

useLcensor : If IPCW=TRUE, useLcensor is a $(K+1) \times p$ matrix, whose $(k+1, j+1)$th entry equals TRUE if the jth variable in the L component of the data frame called 'data' (see above) is included in the model for the hazard of censoring during time interval $[S_k, S_{k+1})$, and equals FALSE if this variable is not included. Ignored if IPCW=FALSE.

admincens : The censoring model (i.e. the model used to calculate the inverse probability of censoring weights) considers censoring as the 'event' of interest and considers failure as a 'censoring'. By default, the censoring model treats all censorings as 'events'. However, one might not want inverse probability of censoring weights to adjust for administrative censorings. If so, adminstrative censorings need to be treated as 'censorings' rather than 'events' in the censoring model. To treat some censorings as 'censorings' rather than 'events' in the censoring model, specify admincens as a vector of length n whose ith entry equals TRUE if the censoring of individual i is to be treated by the censoring model as a 'censoring' and equals FALSE if it is to be treated as an 'event'. This argument is ignored if IPCW is not specified or if IPCW is specified to equal FALSE.

useZ : By default, there is assumed to be no effect modification, i.e. it assumes that the causal effect of A_k does not depend on the treatment or confounder histories $(\bar{A}_{k-1}, \bar{L}_k)$. This assumption corresponds to $Z_{k(l)} = 1$ in the SNCSTM. useZ can be used to indicate that the SNCSTM should instead allow the causal effect of A_k to depend on $(\bar{A}_{k-1}, \bar{L}_k)$. In that case, useZ should be a $(K+1) \times (K+1) \times (N_{\text{mod}} + 1)$ array, where N_{mod} is the number of effect modifiers of A_k. Note that if useZ is specified, its $(k, l, 1)$th entry should equal 1 for all k and l (so that a main effect of A_k is included in the SNCSTM) and the Z component of the data frame called 'data' (see above) should also be specified. See Example 2 for an example of how to specify useZ and Z.

RETURNBOOT : By default, this equals FALSE. To use the *sncstm* function with the *boot* function, in order to calculate bootstrap confidence intervals, set RETURNBOOT equal to TRUE. See Examples 1 and 2 for how to use the *boot* function.

EXPONCUM : By default, this equals FALSE. If it equals TRUE, the *sncstm* function returns not only the estimates of $\psi_{k(l)}$ $(0 \leqslant k \leqslant l \leqslant K)$ but also the corresponding estimates of $\sum_{j=k}^{l} \psi_{k(j)}$ and $\exp\left\{\sum_{j=k}^{l} \psi_{k(j)}\right\}$. These quantities are relevant to the calculation of relative probabilities of survival (see the analysis of the Cystic Fibrosis data in Section 8 of our article).

VERBOSE : By default, this equals FALSE. If it equals TRUE, the *sncstm* function will print some extra information.

J. Comparison with Picciotto et al.'s (2012) method

J.1 *The case of* $K = 0$

For simplicity, suppose that $K = 0$, that $Z_{0(0)} = 1$ (i.e. no effect modification), that there is no random censoring, and that all individuals still at risk at time 10 are administratively censored at that time.

To compare the SNCSTM with Picciotto et al.'s (2012) method, first reformulate Picciotto et al.'s original structural nested cumulative failure time model as a model for survival, rather than failure. Then treat the data as being the result of ten visits, at times $t = 0, \ldots, 9$, at each of which the exposure of an individual is the same (i.e. A_0), with the exposure effect being $\psi_{0(0)}$. Then Picciotto et al.'s estimating function (see their first equation after their equation (11)) is

$$\sum_{s=0}^{9} R(s)\{A_0 - \hat{E}(A_0 \mid \bar{A}_0^s, L_0, T \geqslant s)\}$$
$$\times \sum_{t=s+1}^{10} J(\bar{A}_0^s, L_0, t) \left\{\exp\left(\sum_{j=s}^{t-1} \psi_{0(0)} A_0\right) R(t) - B(\bar{A}_0^s, L_0, t)\right\} \qquad (33)$$

where $\bar{A}_0^s = A_0$ if $s > 0$ and is null otherwise, and where $J(\bar{A}_0^s, L_0, t)$ and $B(\bar{A}_0^s, L_0, t)$ are any given functions of $\bar{A}_0^s$, L_0 and t. Picciotto et al. use $J(\bar{A}_0^s, L_0, t) = 1$ and $B(\bar{A}_0^s, L_0, t) = 0$. Obviously, $\hat{E}(A_0 \mid \bar{A}_0^s, L_0, T \geqslant s) = A_0$ when $s > 0$. So, with $B(\bar{A}_0^s, L_0, t) = 0$, expression (33) reduces to

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} \sum_{t=1}^{10} J(L_0, t) \exp(\psi_{0(0)} A_0 t) R(t) \tag{34}$$

As Dukes and Vansteelandt (2018) explain, a drawback of Picciotto et al.'s method relative to Methods 2 and 3 is the difficulty of deriving the efficient choice of $J(L_0, t)$. This difficulty arises because of the correlation between the survival indicators $R(1), \ldots, R(10)$. Methods 2 and 3 are instead based on independent martingale increments, which makes it easier to derive efficient estimating equations.

J.2 *The case of* $K = 1$

Now consider the more complicated scenario where $K = 1$. Suppose that we still have $Z_{0(0)} = Z_{0(1)} = Z_{1(1)} = 1$ (i.e. no effect modification) and still there is no random censoring. Suppose that $S_1 = 10$, i.e. A_1 is measured at time 10 and that all individuals still at risk at time 20 are administratively censored at that time ($S_2 = 20$).

Picciotto et al.'s estimating function for $(\psi_{0(0)}, \psi_{1(1)}, \psi_{0(1)})$ is (using their notation

$H_{st} = H_{st}(\psi_{0(0)}, \psi_{1(1)}, \psi_{0(1)})$ defined on page 890)

$$\sum_{s=0}^{19} R(s) \Big[\{A_0 - \hat{E}(A_0 \mid \bar{A}_0^s, L_0, T \geqslant s)\} I(s \leqslant 9)$$

$$+ \{A_1 - \hat{E}(A_1 \mid \bar{A}_1^s, \bar{L}_1, T \geqslant s)\} I(s \geqslant 10) \Big]$$

$$\times \sum_{t=s+1}^{20} \big\{ J(A_0^s, L_0, t) I(s \leqslant 9) + J(\bar{A}_1^s, \bar{L}_1, t) I(s \geqslant 10) \big\} \times H_{st}$$

$$= \{A_0 - \hat{E}(A_0 \mid L_0)\} \sum_{t=1}^{20} J(L_0, t) H_{0t}$$

$$+ R(10) \{A_1 - \hat{E}(A_1 \mid A_0, \bar{L}_1, T \geqslant 10)\} \sum_{t=11}^{20} J(A_0, \bar{L}_1, t) H_{10,t}$$

where $\bar{A}_1^s = A_0$ for $s \leqslant 10$ and $\bar{A}_1^s = \bar{A}_1$ for $s \geqslant 11$. If we take $J(L_0, t) = J_{0(0)}$ for $t \leqslant 10$, $J(L_0, t) = J_{0(1)}$ for $t \geqslant 11$ and $J(A_0, \bar{L}_1, t) = J_{1(1)}$, then this becomes

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} J_{0(0)} \sum_{t=1}^{10} H_{0t} + \{A_0 - \hat{E}(A_0 \mid L_0)\} J_{0(1)} \sum_{t=11}^{20} H_{0t}$$

$$+ R(10) \{A_1 - \hat{E}(A_1 \mid A_0, \bar{L}_1, T \geqslant 10)\} J_{1(1)} \sum_{t=11}^{20} H_{10,t}$$

$$= \{A_0 - \hat{E}(A_0 \mid L_0)\} J_{0(0)} \sum_{t=1}^{10} \exp(\psi_{0(0)} A_0 t) R(t)$$

$$+ \{A_0 - \hat{E}(A_0 \mid L_0)\} J_{0(1)} \exp\{\psi_{0(0)} A_0 \times 10\} \sum_{t=11}^{20} \exp\{(\psi_{0(1)} A_0 + \psi_{1(1)} A_1)(t - 10)\} R(t)$$

$$+ \{A_1 - \hat{E}(A_1 \mid A_0, \bar{L}_1, T \geqslant 10)\} J_{1(1)} \sum_{t=11}^{20} \exp\{\psi_{1(1)} A_1 (t - 10)\} R(t)$$

If we choose $J_{0(0)} = (1, 0, 0)^\top$, $J_{1(1)} = (0, 1, 0)^\top$ and $J_{0(1)} = (0, 0, 1)^\top$, then this estimating function becomes a vector with the following three elements:

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} \sum_{t=1}^{10} \exp(A_0 \psi_{0(0)} t) R(t)$$

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} \exp\{A_0 \psi_{0(0)} \times 10\} \sum_{t=11}^{20} \exp\{(A_0 \psi_{0(1)} + A_1 \psi_{1(1)})(t - 10)\} R(t)$$

$$\{A_1 - \hat{E}(A_1 \mid A_0, \bar{L}_1, T \geqslant 10)\} \sum_{t=11}^{20} \exp\{A_1 \psi_{1(1)} (t - 10)\} R(t)$$

To impose the constraint that $\psi_{0(0)} = \psi_{1(1)}$, one may instead choose $J_{0(0)} = J_{1(1)} = (1, 0)^\top$ and $J_{0(1)} = (0, 1)^\top$.

J.3 *The general case of $K \geqslant 0$*

More generally, for $K \geqslant 0$, the estimating function is a vector with elements of the form

$$\{A_k - \hat{E}(A_k \mid \bar{A}_{k-1}, \bar{L}_k, T \geqslant 10k)\} \exp\left\{\sum_{j=k}^{l-1} A_j \sum_{m=j}^{l-1} \psi_{j(m)} \times 10\right\}$$
$$\times \sum_{t=10l+1}^{10l+10} \exp\left\{\sum_{j=k}^{l} A_j \psi_{j(l)}(t-10l)\right\} R(t) \qquad (0 \leqslant k \leqslant l \leqslant K)$$

when the constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}\ \forall k, k', m$ is not imposed, and analogously when the constraint is imposed.

J.4 *Structural nested cumulative failure time model with censoring*

When there is random (i.e. non-administrative) censoring, Picciotto et al. (2012) use inverse probability of censoring weights for their structural nested cumulative failure time model. These weights involve the conditional probability that an individual is censored by time $t+1$ given that he or she has survived and not been censored by time t and his or her treatment and confounder histories at time t. Note that this probability will, in general, depend on the treatment history at time t, even if censoring is completely at random, because individuals cease to be at risk of censoring when they fail. For example, suppose that $K = 0$, that there are no baseline confounders L_0 and that treatment A_0 increases the hazard of failure between times 0 and 1. Then the expected time that an individual is at risk of being censored is less for a treated individual than for an untreated individual. Suppose censoring is completely at random. Then the probability of being censored by time 1 is lower for treated individuals than for untreated individuals, because of the former group's smaller expected time at risk of censoring. For readers who are interested, we now provide a more specific example.

Let A_0 be binary with $P(A_0 = 1) = 0.5$ and let $T(a_0) \sim \exp(0.5+0.5a_0)$ $(a_0 = 0, 1)$

independently of A_0. The true value of $\psi_{0(0)}$ is given by

$$\begin{aligned}
\psi_{0(0)} &= \log\left[\frac{P\{T < 1 \mid A_0 = 1\}}{P\{T(0) < 1 \mid A_0 = 1\}}\right] \\
&= \log\left[\frac{P\{T(1) < 1\}}{P\{T(0) < 1\}}\right] \\
&= \log\left[\frac{1 - \exp(-1)}{1 - \exp(-0.5)}\right] \\
&= 0.474.
\end{aligned}$$

Picciotto et al.'s estimating function (see their first equation after their equation (11)) is

$$\{A_0 - E(A_0)\} \exp(-\psi_{0(0)} A_0) I(T \leqslant 1) \tag{35}$$

when there is no censoring. Expression (35) can be shown to have expectation zero when $\psi_{0(0)} = 0.474$.

Now suppose there is censoring, with C denoting an individual's censoring time. Picciotto et al. replace the term $I(T \leqslant 1)$ in expression (35) by $I(T \leqslant 1, C > T)$ (see their Section 6) and their estimating function becomes

$$\{A_0 - E(A_0)\} \exp(-\psi_{0(0)} A_0) I(T \leqslant 1, C > T) \tag{36}$$

This equals $-0.5 \times I\{T(0) \leqslant 1, C > T(0)\}$ if $A_0 = 0$ and $0.5 \exp(-\psi_{0(0)}) \times I\{T(1) \leqslant 1, C > T(1)\}$ if $A_0 = 1$.

Suppose that $C \sim \exp(5)$ independently of A_0, $T(0)$ and $T(1)$. Then

$$\begin{aligned}
E[I\{T(0) \leqslant 1, C > T(0)\}] &= P\{T(0) \leqslant 1, C > T(0)\} \\
&= \int_0^1 \exp\{-(0.5 + 5)t\} \times 0.5 \, dt \\
&= \frac{1 - \exp(-5.5)}{11}
\end{aligned}$$

and, similarly,

$$E[I\{T(1) \leqslant 1, C > T(1)\}] = \frac{1 - \exp(-6)}{6}$$

So, the expectation with respect to A_0, T and C of expression (36) is

$$\frac{1}{2}\left[-\frac{1-\exp(-5.5)}{11}\times\frac{1}{2}+\frac{\{1-\exp(-6)\}\exp(-\psi_{0(0)})}{6}\times\frac{1}{2}\right] \qquad (37)$$

This equals zero when

$$\psi_{0(0)} = -\log\left\{\frac{6}{11}\times\frac{1-\exp(-5.5)}{1-\exp(-6)}\right\} = 0.608$$

So, the estimator of $\psi_{0(0)}$ will converge asymptotically to 0.608, rather than to the true value, 0.474.

This (asymptotic) bias can be corrected by using inverse probability of censoring weighting. The weight for an individual with $A_0 = a_0$ is $1/P\{C > T(a_0)\} = \exp\{5\,T(a_0)\}$. So, the weighted estimating function is

$$\{A_0 - E(A_0)\}\exp(-\psi_{0(0)}A_0)I(T \leqslant 1, C > T)\times\exp(5T) \qquad (38)$$

To calculate the expectation of expression (38) at various values of $\psi_{0(0)}$, we simulated data on A_0, T and C for 10^7 individuals, calculated expression (38) for each individual, and averaged over the individuals. Figure 2 shows the result. Also shown in Figure 2 is the expectation of the unweighted estimating function. (This latter expectation is given by expression (37), but we also verified that the same result was obtained by calculating expression (36) for each of the 10^7 simulated individuals and averaging.) We see from Figure 2 that, unlike the unweighted estimating function, the weighted estimating function has expectation zero at $\psi_{0(0)} = 0.476$. This is very close to the true value of $\psi_{0(0)}$, viz. 0.474. The very small difference is likely to be due to the Monte Carlo error inherent in calculating the expectation by simulation.

J.5 *Censoring for the structural nested model for survival*

Now consider our situation, where Picciotto et al.'s structural nested cumulative failure time model has been reformulated as a model for survival, rather than for failure. In this reformulated model, no inverse probability of weighting is required when censoring is completely at random. We now explain why this is. Consider the estimating function of expression (34). The quantity $\{A_0 - \hat{E}(A_0 \mid L_0)\}J(L_0, t)\exp(\psi_{0(0)}A_0 t)R(t)$ is unknown if $R(t)$, the survival status at time t, is unknown. If we exclude individuals whose $R(t)$ is unknown, we should compensate for this exclusion by weighting each individual whose $R(t)$ is known by the inverse of the probability that his or her $R(t)$ is known. $R(t)$ is known if $T < C$ or $C > t$. So, if censoring is completely at random, with cumulative hazard $H(t)$, then the probability that $R(t)$ is known is $\exp\{-H(t \wedge T)\}$. In fact, this probability needs to be evaluated only for those individuals whose $R(t)$ is known and equals 1. This is because $\{A_0 - \hat{E}(A_0 \mid L_0)\}J(L_0, t)\exp(\psi_{0(0)}A_0 t)R(t) = 0$ when $R(t) = 0$. For an individual whose $R(t)$ is known to equal 1, T must be greater than t, meaning that $\exp\{-H(t \wedge T)\} = \exp\{-H(t)\}$. Since $\exp\{-H(t)\}$ is a constant (apart from depending on t), there is no need to weight by its inverse.

J.6 *Applying Picciotto et al.'s estimation method in simulation study*

We compared the performance of Picciotto et al.'s estimation method to our Methods 1–3 in the two regular visit scenarios of the simulation study of Section 7 of our article. In this simulation study, $S_1 = 1$, $S_2 = 2$, etc., rather than $S_1 = 10$, $S = 20$, etc. as was assumed above. So, the estimating equations become

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} \sum_{t=1}^{10} \exp(A_0\psi_{0(0)}t/10)R(t/10) = 0$$

$$\{A_0 - \hat{E}(A_0 \mid L_0)\} \exp\{A_0\psi_{0(0)}\} \sum_{t=11}^{20} \exp\{(A_0\psi_{0(1)} + A_1\psi_{1(1)})(t-10)/10\}R(t/10) = 0$$

$$\{A_1 - \hat{E}(A_1 \mid A_0, \bar{L}_1, T \geqslant 1)\} \sum_{t=11}^{20} \exp\{A_1\psi_{1(1)}(t-10)/10\}R(t/10) = 0$$

etc., when the constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}\ \forall k, k', m$ is not imposed, and analogously when the constraint is imposed. Inverse probability of censoring weighting was used in the scenario with random censoring, as described in Web Appendix J.4.

As Picciotto et al. (2012) note, these estimating equations cannot be solved using the Newton-Raphson algorithm. They used the Newson-Mead algorithm; we used a simple grid search with a fine grid.

Results are reported in Section 7 and Tables 1 and 2 of our article.

K. Competing risks

Suppose there are two competing causes of failure. Let $T(\bar{A}_k, 0)$ be the counterfactual failure time as previously defined, that is, it is the time from the start of the study until failure, regardless of the cause of failure. Let $N^{(j)}_{(\bar{A}_k,0)}(t)$ denote the counterfactual counting process indicator for a failure due to cause j ($j = 1, 2$). We consider the following semi-parametric additive cause-specific hazard model

$$
\begin{aligned}
&E\left\{dN^{(j)}_{(\bar{A}_{k-1},0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\right\} \\
&= E\left\{dN^{(j)}_{(\bar{A}_k,0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t\right\} - A_k \psi^{(j)}_{k(l)} dt,
\end{aligned}
$$

for $t \in [l, l+1)$ and $j = 1, 2$. Building on similar results in Martinussen and Vansteelandt (2018), the procedure proposed in our article is readily adjusted to the estimation of $\psi^{(j)}_{k(l)}$ ($j = 1, 2$) by redefining $w_k(t)$ as

$$
w_k(t) = \prod_{j=k+1}^{K} \exp\left\{A_k \psi^{(1)}_{k(l)} + A_k \psi^{(2)}_{k(l)}\right\}.
$$

This change accounts for the fact that the conditional mean of the at-risk indicator $R(t) = I(T \geqslant t)$ is influenced by the cause-specific hazards of both failure types. For instance, equation (7) becomes

$$
\sum_{i=1}^{n} \int_{l}^{l+1} R_i(t) w_{ki}(t) \Delta_{ki}(t) \left[dN^{(j)}_i(t) - \left\{ \sum_{s=k+1}^{l} A_{si} \psi^{(j)}_{s(l)} + \Delta_{ki}(l) \psi^{(j)}_{k(l)} \right\} dt \right] = 0,
$$

which must be solved jointly for $j = 1$ and $j = 2$.

Formulae very similar to those given by Martinussen and Vansteelandt (2018) can be used to convert these estimates of the parameters $\psi^{(j)}_{k(l)}$, which describe the causal effect of exposure on cause-specific hazards, into estimates of the causal effect of exposure on cumulative incidence.

L. Additional simulation studies

In Section 7 we showed results for $n = 1000$ in the two regular visit scenarios, one with no censoring and one with random censoring. Web Tables 4 and 5 show the results for the irregular visit scenario.

Web Tables 6–11 show the corresponding results for the three scenarios when $n = 250$.

Web Tables 12–17 show the corresponding results for $n = 1000$ with a shorter follow-up time. In this case, the visit times are divided by four and administrative censoring occurs at time 1. This means that the visit times are $S_k = k/4$ ($k = 0, \ldots, 4$) in the regular visit scenarios and the inter-visit times are $S_k - S_{k-1} \sim \text{Uniform}[0.5/4, 1.5/4]$ in the irregular visit scenario.

We also carried out a simple simulation study to illustrate the double robustness properties of Methods 2 and 3. We shall use $\mathcal{B}^*_{k(l)}$ to denote the modified version of Model $\mathcal{B}_{k(l)}$ allowed by Method 3. The original model, Model $\mathcal{B}_{k(l)}$, allowed by Method 2 assumes that the intercept and coefficients for $\bar{A}_{k-1}$ and $\bar{L}_k$ are constant over time. The modified model, Model $\mathcal{B}^*_{k(l)}$, is more general, allowing as it does the intercept and coefficients for $\bar{A}_{k-1}$ and $\bar{L}_k$ to vary over time. If Model $\mathcal{B}_{k(l)}$ is correctly specified, then so is Model $\mathcal{B}^*_{k(l)}$.

In this simulation study, we assumed $K = 1$, $S_0 = 0$, $S_1 = 1$, a single time-dependent, continuous confounder, and a continuous treatment. The hazard of $T(a_0, 0)$ given A_0 and L_0 during time interval $t \in [0, 1)$ was assumed to be

$$E\{dN_{(a_0,0)}(t) \mid A_0, L_0, T(a_0, 0) \geqslant t\} = \omega_0 + \gamma_{00}L_0 + \delta_{00}a_0 \qquad t \in [0, 1)$$

where $\omega_0 = 2.7$, $\gamma_{00} = 0.75$ and $\delta_{00} = -0.3$. This implies that Model $\mathcal{M}_0$ is correctly specified during the time interval $t \in [0, 1)$ and $\psi_{0(0)} = \delta_{00} = -0.3$. It also implies that Model $\mathcal{B}_{0(0)}$ is correctly specified, since $E\{dN_{(0)}(t) \mid A_0, L_0, T(0) \geqslant t\} = \omega_0 + \gamma_{00}L_0$ is linear in L_0 and the intercept ω_0 and coefficient γ_{00} of L_0 are constant in t.

Intervening on A_0 may change the value of L_1. So, let $L_1(a_0)$ denote the value of L_1 when A_0 is set by intervention to equal a_0. We shall assume that

$$L_1(a_0) \mid A_0 = a_0, L_0, T(a_0, 0) \geqslant 1 \sim N(1.6 - 0.5a_0, 0.5)$$

The hazard of $T(a_0, a_1)$ given $\bar{A}_1$, L_0 and $L_1(a_0)$ during time interval $t \in [1, 2)$ was assumed to be

$$\begin{aligned} &E\{dN_{(a_0,a_1)}(t) \mid \bar{A}_1 = (a_0, a_1), L_0, L(a_0), T(a_0, a_1) \geqslant t\} \\ &= \omega_1 + \gamma_{10}L_0 + \gamma_{11}L_1(a_0) + \delta_{10}a_0 + \delta_{11}a_1 + (t - 1)c \qquad t \in [1, 2) \end{aligned}$$

where $\omega_1 = 2.7$, $\gamma_{10} = 0$, $\gamma_{11} = 0.75$, $\delta_{10} = 0.275$, $\delta_{11} = -0.3$ and c is a constant. We shall consider two values of c: 0 and 9/32. This form of the hazard implies that Model $\mathcal{M}_1$ is correctly specified and $\psi_{1(1)} = \delta_{11} = -0.3$. Also, Model $\mathcal{B}_{0(0)}$ is correctly specified if $c = 0$, but not if $c = 9/32$ (because the intercept term $2.7 + (t - 1)c$ is then a function of t). However, Model $\mathcal{B}^*_{0(0)}$ is correctly specified regardless of whether $c = 0$ or $c = 9/32$ (because the intercept in that model is allowed to be a function of t).

We shall now show that: i) Model $\mathcal{M}_0$ is correctly specified during time interval

$t \in [1,2)$, with $\psi_{0(1)} = -0.1$; ii) Model $\mathcal{B}_{0(1)}$ is correctly specified if $c = 9/32$ but not if $c = 0$; and iii) Model $\mathcal{B}^*_{0(1)}$ is correctly specified whatever the value of c.

For $t \in [1,2)$,

$$\begin{aligned}
&P\{T(a_0,0) \geqslant t \mid A_0, L_0\} \\
&= P\{T(a_0,0) \geqslant 1 \mid A_0, L_0\} \\
&\quad \times E_{L_1(a_0)}\left[P\{T(a_0,0) \geqslant t \mid \bar{A}_1, L_1(a_0), T(a_0,0) \geqslant 1\} \mid A_0, L_0, T(a_0,0) \geqslant 1\right] \\
&= \exp\{-(\omega_0 + \gamma_{00}L_0 + \delta_{00}a_0)\} \\
&\quad \times E_{L_1(a_0)}\Bigg(\exp\Bigg[-\{\omega_1 + (t-1)c/2 + \gamma_{10}L_0 + \gamma_{11}L_1(a_0) + \delta_{10}a_0\}(t-1)\Bigg] \\
&\qquad\qquad \mid A_0, L_0, T(a_0,0) \geqslant 1\Bigg) \\
&= \exp\left[-\{\omega_0 + \gamma_{00}L_0 + \delta_{00}a_0\} - \{\omega_1 + (t-1)c/2 + \gamma_{10}L_0 + \delta_{10}a_0\}(t-1)\right] \\
&\quad \times E_{L_1(a_0)}\left[\exp\{-\gamma_{11}L_1(a_0)\,(t-1)\} \mid A_0, L_0, T(a_0,0) \geqslant 1\right] \qquad (39)
\end{aligned}$$

Moreover, because $L_1(a_0)$ is normally distributed given A_0, L_0 and $T(a_0,0) \geqslant 1$, and using the form of the moment generating function of a normal distribution, we have

$$\begin{aligned}
&E_{L_1(a_0)}\left[\exp\{-\gamma_{11}L_1(a_0)\,(t-1)\} \mid A_0, L_0, T(a_0,0) \geqslant 1\right] \\
&= \exp\Big[-E\{\gamma_{11}L_1(a_0)(t-1) \mid A_0, L_0, T(a_0,0) \geqslant 1\} \\
&\qquad + \frac{1}{2}\mathrm{Var}\{\gamma_{11}L_1(a_0)(t-1) \mid A_0, L_0, T(a_0,0) \geqslant 1\}\Big] \\
&= \exp\Big[-(t-1)\gamma_{11}E\{L_1(a_0) \mid A_0, L_0, T(a_0,0) \geqslant 1\} \\
&\qquad + \frac{1}{2}(t-1)^2\gamma_{11}^2\mathrm{Var}\{L_1(a_0) \mid A_0, L_0, T(a_0,0) \geqslant 1\}\Big] \qquad (40)
\end{aligned}$$

It follows from equations (39) and (40) that the conditional hazard of $T(a_0,0)$ given A_0 and L_0 during time interval $t \in [1,2)$ equals

$$\begin{aligned}
&\omega_1 + (t-1)c + \gamma_{10}L_0 + \delta_{10}a_0 + \gamma_{11}E\{L_1(a_0) \mid A_0, L_0, T(a_0,0) \geqslant 1\} \\
&-(t-1)\gamma_{11}^2\mathrm{Var}\{L_1(a_0) \mid A_0, L_0, T(a_0,0) \geqslant 1\} \qquad (41)
\end{aligned}$$

We see from the hazard (41) that Model $\mathcal{M}_0$ is correctly specified in the time interval $t \in [1,2)$, with

$$\begin{aligned}\psi_{0(1)} &= \delta_{10} + \gamma_{11} \times \frac{E\{L_1(a_0) \mid A_0, L_0\} - E\{L_1(0) \mid A_0, L_0\}}{a_0} \\ &= 0.275 + 0.75 \times (-0.5) \\ &= -0.1\end{aligned}$$

Moreover, if $c = \gamma_{11}^2 \mathrm{Var}\{L_1(a_0) \mid A_0, L_0, T(a_0, 0) \geqslant 1\} = 0.75^2 \times 0.5 = 9/32$, the hazard (41) reduces to

$$\begin{aligned}&\omega_1 + \gamma_{10} L_0 + \delta_{10} a_0 + \gamma_{11} E\{L_1(a_0) \mid A_0, L_0, T(a_0, 0) \geqslant 1\} \\ &= 2.7 + 0 \times L_0 + 0.275 a_0 + 0.75(1.6 - 0.5 a_0) \\ &= 3.9 - 0.1 a_0\end{aligned}$$

which shows that Model $\mathcal{B}_{0(1)}$ is correctly specified if $c = 9/32$ but not if $c = 0$. The more general model, Model $\mathcal{B}^*_{0(1)}$, is correctly specified whether $c = 0$ or $c = 9/32$.

We generated observed data from this model as follows. First, we need to specify how to generate L_0, A_0 and A_1.

For L_0, we assumed that $L_0 \sim N(0, 0.5)$.

The data-generating distribution of A_0 given L_0 was either $A_0 \mid L_0 \sim N(3 - L_0, 0.9^2)$ or $A_0 \mid L_0 \sim N(3.5 - 2I(L_0 \geqslant 0.75), 0.9^2)$, where $I(.)$ is the indicator function. We call these two data-generating models $\mathcal{A}_0^{\mathrm{gen}(1)}$ and $\mathcal{A}_0^{\mathrm{gen}(2)}$, respectively. The Model $\mathcal{A}_0$ that will be assumed when fitting the SNCSTM to these simulated data is a linear regression of A_0 on L_0. So, if the data-generating model is $\mathcal{A}_0^{\mathrm{gen}(1)}$ then the assumed Model $\mathcal{A}_0$ is correctly specified, but if it is $\mathcal{A}_0^{\mathrm{gen}(2)}$ then Model $\mathcal{A}_0$ is misspecified.

The hazard for the observed failure time T is $2.7 + 0.75 L_0 - 0.3 A_0$ during the time interval $t \in [0, 1)$.

To generate L_1 given A_0, L_1 and $T \geqslant 1$, we assumed

$$L_1 \mid A_0, L_0, T \geqslant 1 \sim N(1.6 - 0.5A_0, 0.5)$$

The true data-generating distribution of A_1 given A_0, $\bar{L}_1$ and $T \geqslant 1$ was either $A_1 \mid A_0, \bar{L}_1, T \geqslant 1 \sim N(3 - L_1, 0.9^2)$ (called model $\mathcal{A}_1^{\text{gen}(1)}$) or $A_1 \mid A_0, \bar{L}_1, T \geqslant 1 \sim N(3.5 - 2I(L_1 \geqslant 0.75), 0.9^2)$ (called model $\mathcal{A}_1^{\text{gen}(2)}$). The Model $\mathcal{A}_1$ that will be assumed when fitting the SNCSTM to these simulated data is a linear regression of A_1 on A_0 and $\bar{L}_1$. So, if the data-generating model is $\mathcal{A}_1^{\text{gen}(1)}$ then Model $\mathcal{A}_1$ is correctly specified, but if it is $\mathcal{A}_1^{\text{gen}(2)}$ then Model $\mathcal{A}_1$ is misspecified.

The hazard for the observed failure time T is $2.7+0.75L_1+0.275A_0-0.3A_1+(t-1)c$ during the time interval $t \in [1, 2)$.

We considered the following four scenarios:

(1) The data-generating models for A_0 and A_1 are $\mathcal{A}_0^{\text{gen}(1)}$ and $\mathcal{A}_1^{\text{gen}(1)}$. This means that Models $\mathcal{A}_0$ and $\mathcal{A}_1$ are correctly specified, and so all of Methods 1–3 should yield consistent estimates of $\psi_{0(0)}$, $\psi_{0(1)}$ and $\psi_{1(1)}$. Here we used $c = 0$.

(2) The data-generating models for A_0 and A_1 are $\mathcal{A}_0^{\text{gen}(1)}$ and $\mathcal{A}_1^{\text{gen}(2)}$. This means that $\mathcal{A}_0$ is correctly specified, but $\mathcal{A}_1$ is misspecified. Here, Method 1 should yield a consistent estimate of $\psi_{0(0)}$ but the estimates of $\psi_{0(1)}$ and $\psi_{1(1)}$ may be inconsistent. We used $c = 0$, which means Model $\mathcal{B}_{1(1)}$ is also correctly specified. The double robustness properties of Methods 2 and 3 should mean that these methods yield consistent estimates not only of $\psi_{0(0)}$ but also of $\psi_{0(1)}$ and $\psi_{1(1)}$.

(3) The data-generating models for A_0 and A_1 are $\mathcal{A}_0^{\text{gen}(2)}$ and $\mathcal{A}_1^{\text{gen}(1)}$. This means that $\mathcal{A}_0$ is misspecified, but $\mathcal{A}_1$ is correctly specified. Here, Method 1 should yield a consistent estimate of $\psi_{1(1)}$ but the estimates of $\psi_{0(0)}$ and $\psi_{0(1)}$ may be inconsistent. We used $c = 9/32$, which means Model $\mathcal{B}_{0(1)}$ is correctly specified.

Since Model $\mathcal{B}_{0(0)}$ is also correctly specified, the double robustness properties of Methods 2 and 3 should mean that these methods yield consistent estimates of $\psi_{0(0)}$, $\psi_{0(1)}$ and $\psi_{1(1)}$.

(4) The data-generating models for A_0 and A_1 are $\mathcal{A}_0^{\text{gen}(2)}$ and $\mathcal{A}_1^{\text{gen}(2)}$. This means that Models $\mathcal{A}_0$ and $\mathcal{A}_1$ are both misspecified. Here, Method 1 may yield inconsistent estimates of all three parameters. We used $c = 0$, which means that Models $\mathcal{B}_{1(1)}$ and $\mathcal{B}_{0(0)}$ are correctly specified. Method 2 should therefore yield consistent estimates of $\psi_{0(0)}$ and $\psi_{1(1)}$. However, Model $\mathcal{B}_{0(1)}$ is misspecified, meaning that the estimate of $\psi_{0(1)}$ from Method 2 may be inconsistent. Method 3 should yield consistent estimates of all three parameters, because Model $\mathcal{B}^*_{0(1)}$ is correctly specified.

For each of these four scenarios, we considered the case of no random censoring and the case of censoring completely at random with constant censoring hazard 0.5. Thus, there were a total of eight scenarios. In all eight scenarios, all individuals who had not failed or been censored prior to time $t = 2$ were administratively censored at that time. For each scenario, we generated 5000 simulated datasets, each of $n = 5000$ individuals.

Table 18 shows the mean of the parameter estimates over the 5000 simulated datasts when there is no censoring. Results for the random censoring scenarios were very similar. Monte Carlo standard errors for these means are shown in brackets. The results are as expected. That is, where we have predicted an estimator to be consistent, it is approximately unbiased, and where it has been predicted to be possibly inconsistent, it is generally biased. One exception is that the bias (if any) in the estimator of $\psi_{0(1)}$ from Method 2 is very small even in the fourth scenario, where we predicted this estimator was likely to be inconsistent. This is probably because the misspecification of Model $\mathcal{B}_{0(1)}$ is only minor, i.e. only the intercept term depends on t and this dependence is not large.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[Table 11 about here.]

[Table 12 about here.]

[Table 13 about here.]

[Table 14 about here.]

[Table 15 about here.]

[Table 16 about here.]

[Table 17 about here.]

[Table 18 about here.]

[Figure 1 about here.]

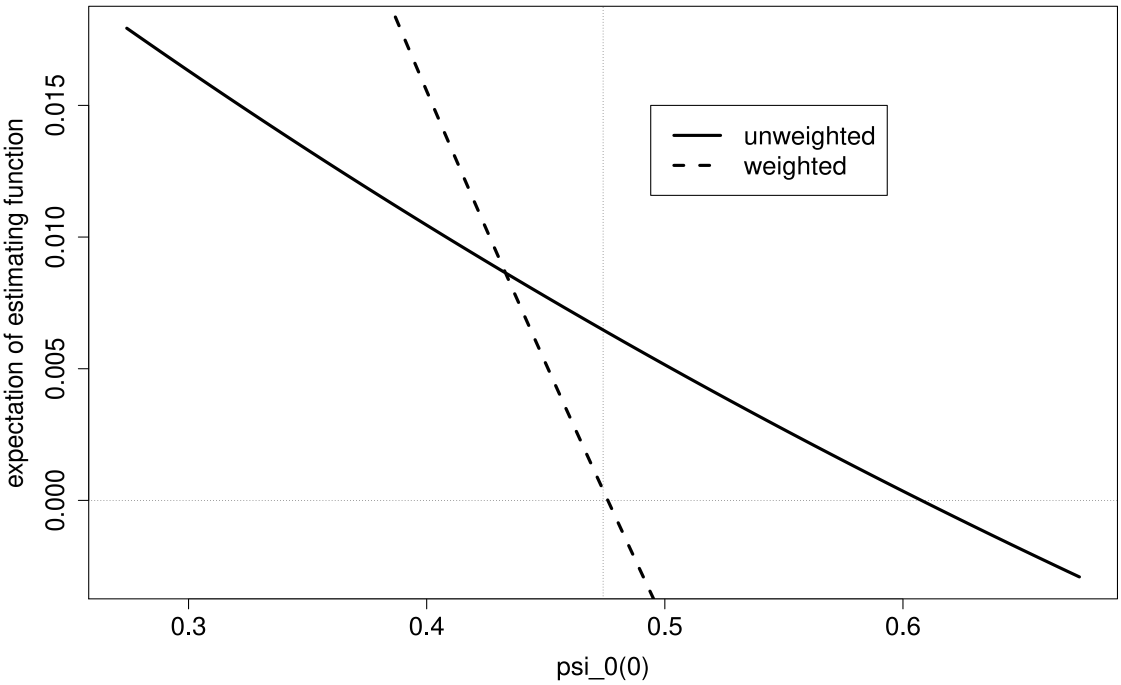

Figure 2. Expectations of unweighted (solid line) and weighted estimating function (broken line) as functions of $\psi_{0(0)}$. The true value of $\psi_{0(0)}$ is shown by the vertical dotted line.

Table 1

Means (×10) and SEs (×10) of parameter estimates when $n = 1000$, visits are regular and the only censoring is administrative. 'Mtd' is method ('P' is Picciotto et al.'s method — see Section 9) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
		Means									
1	no	0.393	0.098	0.031	0.025	0.391	0.096	0.034	0.403	0.098	0.383
2	no	0.396	0.100	0.032	0.024	0.394	0.097	0.033	0.408	0.100	0.392
3	no	0.395	0.100	0.031	0.023	0.392	0.096	0.033	0.406	0.099	0.388
P	no	0.394	0.107	0.030	0.021	0.394	0.094	0.049	0.408	0.102	0.387
1	yes	0.386	0.096	0.032	0.024	0.386	0.096	0.032	0.386	0.096	0.386
2	yes	0.397	0.099	0.032	0.023	0.397	0.099	0.032	0.397	0.099	0.397
3	yes	0.395	0.098	0.032	0.023	0.395	0.098	0.032	0.395	0.098	0.395
P	yes	0.394	0.104	0.030	0.029	0.394	0.104	0.030	0.394	0.104	0.394
		SEs									
1	no	0.177	0.187	0.199	0.218	0.243	0.254	0.260	0.251	0.273	0.272
2	no	0.169	0.180	0.191	0.204	0.237	0.246	0.253	0.240	0.262	0.267
3	no	0.169	0.179	0.190	0.204	0.236	0.245	0.252	0.239	0.260	0.265
P	no	0.196	0.290	0.349	0.397	0.265	0.376	0.452	0.270	0.384	0.300
1	yes	0.113	0.131	0.158	0.217	0.113	0.131	0.158	0.113	0.131	0.113
2	yes	0.109	0.129	0.151	0.203	0.109	0.129	0.151	0.109	0.129	0.109
3	yes	0.109	0.128	0.150	0.203	0.109	0.128	0.150	0.109	0.128	0.109
P	yes	0.126	0.206	0.306	0.494	0.126	0.206	0.306	0.126	0.206	0.126

Table 2

Means (×10) and SEs (×10) of parameter estimates when $n = 1000$, visits are regular and censoring is random. 'Mtd' is method ('1cw' is Method 1 with censoring weights; 'P' is Picciotto et al.'s method — see Section 9) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
		Means									
1	no	0.394	0.108	0.021	0.054	0.396	0.105	0.055	0.403	0.111	0.383
1cw	no	0.396	0.102	0.020	0.054	0.393	0.096	0.054	0.408	0.097	0.383
2	no	0.396	0.104	0.036	0.033	0.399	0.096	0.038	0.411	0.098	0.393
3	no	0.396	0.103	0.036	0.033	0.396	0.095	0.038	0.407	0.096	0.385
P	no	0.397	0.117	0.024	0.050	0.399	0.095	0.078	0.405	0.117	0.390
1	yes	0.391	0.106	0.031	0.053	0.391	0.106	0.031	0.391	0.106	0.391
1cw	yes	0.392	0.099	0.031	0.054	0.392	0.099	0.031	0.392	0.099	0.392
2	yes	0.398	0.099	0.037	0.032	0.398	0.099	0.037	0.398	0.099	0.398
3	yes	0.396	0.099	0.037	0.032	0.396	0.099	0.037	0.396	0.099	0.396
P	yes	0.395	0.108	0.035	0.051	0.395	0.108	0.035	0.395	0.108	0.395
		SEs									
1	no	0.265	0.313	0.372	0.467	0.400	0.483	0.569	0.462	0.563	0.577
1cw	no	0.201	0.234	0.373	0.469	0.298	0.346	0.572	0.348	0.424	0.406
2	no	0.180	0.211	0.252	0.304	0.276	0.313	0.380	0.317	0.385	0.373
3	no	0.180	0.211	0.251	0.303	0.275	0.310	0.375	0.314	0.380	0.367
P	no	0.219	0.389	0.571	0.728	0.334	0.557	0.855	0.385	0.652	0.457
1	yes	0.186	0.241	0.311	0.463	0.186	0.241	0.311	0.186	0.241	0.186
1cw	yes	0.140	0.179	0.313	0.465	0.140	0.179	0.313	0.140	0.179	0.140
2	yes	0.130	0.162	0.211	0.303	0.130	0.162	0.211	0.130	0.162	0.130
3	yes	0.130	0.161	0.210	0.301	0.130	0.161	0.210	0.130	0.161	0.130
P	yes	0.157	0.282	0.475	0.802	0.157	0.282	0.475	0.157	0.282	0.157

Table 3

Coverage (%) of 95% bootstrap confidence intervals for Methods 1, 2 and 1cw (i.e. Method 1 with censoring weights) when n=1000, visits are regular, either there is only administrative censoring or there is random censoring, and the constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is not imposed.

Mtd	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
	No censoring									
1	96.0	96.0	95.5	94.7	94.4	95.5	96.6	95.4	95.7	94.5
2	96.5	96.4	95.4	95.7	94.9	95.6	96.5	96.0	95.8	94.7
	Random censoring									
1	95.0	95.6	96.4	94.8	95.3	95.5	95.9	95.6	96.0	95.4
1cw	96.5	96.8	96.6	95.2	95.9	97.9	95.9	97.1	97.8	97.7
2	95.7	95.7	95.9	96.1	94.9	95.9	96.7	95.9	96.6	96.1

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.394	0.112	0.009	0.024	0.399	0.118	0.034	0.382	0.133	0.354
1cw	no	0.390	0.110	0.009	0.025	0.418	0.106	0.036	0.397	0.136	0.351
2	no	0.398	0.110	0.034	0.026	0.401	0.096	0.020	0.413	0.114	0.408
1	yes	0.392	0.116	0.026	0.024	0.392	0.116	0.026	0.392	0.116	0.392
1cw	yes	0.395	0.112	0.027	0.025	0.395	0.112	0.027	0.395	0.112	0.395
2	yes	0.401	0.107	0.029	0.027	0.401	0.107	0.029	0.401	0.107	0.401

Table 4

Means ($\times 10$) of parameter estimates when n=1000 and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.486	0.645	0.828	0.555	0.722	1.044	1.013	0.875	1.144	1.057
1cw	no	0.453	0.617	0.831	0.556	0.656	1.001	1.015	0.797	1.098	0.997
2	no	0.180	0.214	0.245	0.277	0.265	0.315	0.378	0.305	0.374	0.355
1	yes	0.355	0.490	0.646	0.543	0.355	0.490	0.646	0.355	0.490	0.355
1cw	yes	0.325	0.465	0.646	0.545	0.325	0.465	0.646	0.325	0.465	0.325
2	yes	0.125	0.161	0.204	0.273	0.125	0.161	0.204	0.125	0.161	0.125

Table 5

SEs ($\times 10$) of parameter estimates when n=1000 and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.396	0.098	0.034	0.037	0.397	0.118	0.005	0.393	0.100	0.375
2	no	0.405	0.100	0.042	0.037	0.423	0.128	0.013	0.423	0.106	0.408
3	no	0.403	0.099	0.042	0.037	0.417	0.125	0.014	0.412	0.103	0.393
1	yes	0.362	0.099	0.028	0.037	0.362	0.099	0.028	0.362	0.099	0.362
2	yes	0.410	0.109	0.037	0.036	0.410	0.109	0.037	0.410	0.109	0.410
3	yes	0.403	0.107	0.037	0.036	0.403	0.107	0.037	0.403	0.107	0.403

Table 6

Means (×10) of parameter estimates when n=250 and visits are regular and there is no random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ *is imposed.*

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.377	0.372	0.416	0.448	0.473	0.492	0.543	0.514	0.571	0.549
2	no	0.363	0.359	0.407	0.439	0.474	0.491	0.540	0.521	0.584	0.574
3	no	0.361	0.356	0.404	0.433	0.467	0.482	0.528	0.508	0.568	0.554
1	yes	0.227	0.249	0.304	0.443	0.227	0.249	0.304	0.227	0.249	0.227
2	yes	0.235	0.255	0.310	0.434	0.235	0.255	0.310	0.235	0.255	0.235
3	yes	0.232	0.251	0.306	0.429	0.232	0.251	0.306	0.232	0.251	0.232

Table 7

SEs ($\times 10$) of parameter estimates when n=250 and visits are regular and there is no random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.409	0.115	0.015	-0.003	0.388	0.076	-0.034	0.397	0.153	0.354
1cw	no	0.399	0.107	0.021	-0.005	0.373	0.105	-0.037	0.408	0.121	0.354
2	no	0.404	0.107	0.050	0.051	0.419	0.115	0.020	0.439	0.101	0.403
3	no	0.401	0.106	0.050	0.049	0.411	0.114	0.024	0.422	0.094	0.376
1	yes	0.382	0.108	0.005	-0.002	0.382	0.108	0.005	0.382	0.108	0.382
1cw	yes	0.373	0.105	0.010	-0.005	0.373	0.105	0.010	0.373	0.105	0.373
2	yes	0.409	0.108	0.049	0.046	0.409	0.108	0.049	0.409	0.108	0.409
3	yes	0.401	0.106	0.049	0.044	0.401	0.106	0.049	0.401	0.106	0.401

Table 8

Means ($\times 10$) of parameter estimates when n=250 and visits are regular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.544	0.638	0.765	0.948	0.840	0.949	1.197	0.967	1.161	1.203
1cw	no	0.431	0.482	0.778	0.974	0.633	0.711	1.230	0.741	0.926	0.911
2	no	0.387	0.422	0.524	0.671	0.585	0.659	0.866	0.707	0.885	0.898
3	no	0.385	0.418	0.516	0.655	0.575	0.641	0.832	0.683	0.830	0.840
1	yes	0.383	0.474	0.616	0.902	0.383	0.474	0.616	0.383	0.474	0.383
1cw	yes	0.294	0.355	0.632	0.938	0.294	0.355	0.632	0.294	0.355	0.294
2	yes	0.280	0.325	0.432	0.644	0.280	0.325	0.432	0.280	0.325	0.280
3	yes	0.276	0.317	0.424	0.630	0.276	0.317	0.424	0.276	0.317	0.276

Table 9

SEs (×10) of parameter estimates when n=250 and visits are regular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ *is imposed.*

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.448	0.074	0.043	-0.045	0.387	0.161	0.028	0.294	0.126	0.374
1cw	no	0.440	0.089	0.044	-0.049	0.381	0.146	0.025	0.289	0.124	0.400
2	no	0.423	0.090	0.043	-0.022	0.439	0.124	0.054	0.422	0.113	0.442
1	yes	0.403	0.117	0.030	-0.029	0.403	0.117	0.030	0.403	0.117	0.403
1cw	yes	0.401	0.120	0.030	-0.031	0.401	0.120	0.030	0.401	0.120	0.401
2	yes	0.425	0.099	0.045	-0.021	0.425	0.099	0.045	0.425	0.099	0.425

Table 10

Means (×10) of parameter estimates when n=250 and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ *is imposed.*

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.956	1.356	1.661	1.253	1.422	2.148	1.941	1.746	2.309	2.001
1cw	no	0.885	1.315	1.670	1.266	1.344	2.059	1.972	1.639	2.238	1.895
2	no	0.381	0.433	0.533	0.715	0.558	0.715	0.898	0.732	0.943	0.892
1	yes	0.695	1.039	1.234	1.135	0.695	1.039	1.234	0.695	1.039	0.695
1cw	yes	0.646	1.000	1.256	1.161	0.646	1.000	1.256	0.646	1.000	0.646
2	yes	0.262	0.337	0.450	0.687	0.262	0.337	0.450	0.262	0.337	0.262

Table 11

SEs (×10) of parameter estimates when n=250 and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.393	0.099	0.028	0.005	0.391	0.070	0.036	0.396	0.111	0.386
2	no	0.396	0.103	0.026	0.009	0.395	0.071	0.035	0.406	0.109	0.393
3	no	0.396	0.103	0.026	0.009	0.394	0.071	0.035	0.404	0.109	0.390
1	yes	0.369	0.089	0.030	0.004	0.369	0.089	0.030	0.369	0.089	0.369
2	yes	0.397	0.096	0.029	0.008	0.397	0.096	0.029	0.397	0.096	0.397
3	yes	0.396	0.096	0.029	0.008	0.396	0.096	0.029	0.396	0.096	0.396

Table 12

Means ($\times 10$) of parameter estimates when n=1000, times between visits are divided by four and visits are regular and there is no random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.355	0.339	0.318	0.326	0.433	0.428	0.428	0.431	0.427	0.422
2	no	0.342	0.323	0.310	0.312	0.417	0.407	0.415	0.409	0.407	0.411
3	no	0.341	0.322	0.310	0.312	0.416	0.405	0.413	0.407	0.405	0.409
1	yes	0.201	0.218	0.245	0.325	0.201	0.218	0.245	0.201	0.218	0.201
2	yes	0.200	0.216	0.243	0.311	0.200	0.216	0.243	0.200	0.216	0.200
3	yes	0.200	0.215	0.242	0.311	0.200	0.215	0.242	0.200	0.215	0.200

Table 13

SEs ($\times 10$) of parameter estimates when n=1000, times between visits are divided by four and visits are regular and there is no random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.391	0.093	0.027	0.021	0.394	0.068	0.028	0.399	0.088	0.396
1cw	no	0.396	0.100	0.027	0.020	0.398	0.063	0.028	0.408	0.099	0.400
2	no	0.394	0.104	0.028	0.009	0.398	0.073	0.044	0.404	0.104	0.401
3	no	0.394	0.104	0.027	0.009	0.396	0.073	0.044	0.402	0.104	0.398
1	yes	0.381	0.082	0.026	0.020	0.381	0.082	0.026	0.381	0.082	0.381
1cw	yes	0.387	0.087	0.027	0.019	0.387	0.087	0.027	0.387	0.087	0.387
2	yes	0.398	0.096	0.034	0.008	0.398	0.096	0.034	0.398	0.096	0.398
3	yes	0.396	0.096	0.034	0.008	0.396	0.096	0.034	0.396	0.096	0.396

Table 14

Means (×10) of parameter estimates when n=1000, times between visits are divided by four and visits are regular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	0.502	0.507	0.502	0.532	0.635	0.672	0.693	0.669	0.677	0.680
1cw	no	0.380	0.378	0.503	0.534	0.471	0.487	0.694	0.481	0.508	0.485
2	no	0.346	0.332	0.330	0.339	0.429	0.433	0.449	0.430	0.452	0.444
3	no	0.346	0.332	0.329	0.338	0.427	0.431	0.447	0.428	0.450	0.440
1	yes	0.302	0.339	0.402	0.530	0.302	0.339	0.402	0.302	0.339	0.302
1cw	yes	0.224	0.251	0.404	0.533	0.224	0.251	0.404	0.224	0.251	0.224
2	yes	0.206	0.226	0.260	0.336	0.206	0.226	0.260	0.206	0.226	0.206
3	yes	0.206	0.226	0.259	0.336	0.206	0.226	0.259	0.206	0.226	0.206

Table 15

SEs (×10) of parameter estimates when n=1000, times between visits are divided by four and visits are regular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ *is imposed.*

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
True		0.400	0.100	0.040	0.020	0.400	0.100	0.040	0.400	0.100	0.400
1	no	0.339	0.097	0.004	0.020	0.514	0.056	0.032	0.396	0.048	0.400
1cw	no	0.338	0.101	0.004	0.020	0.515	0.054	0.033	0.391	0.063	0.397
2	no	0.391	0.095	0.031	0.006	0.396	0.092	0.030	0.416	0.112	0.405
1	yes	0.402	0.075	0.019	0.020	0.402	0.075	0.019	0.402	0.075	0.402
1cw	yes	0.400	0.081	0.018	0.020	0.400	0.081	0.018	0.400	0.081	0.400
2	yes	0.399	0.099	0.030	0.007	0.399	0.099	0.030	0.399	0.099	0.399

Table 16

Means ($\times 10$) of parameter estimates when n=1000, times between visits are divided by four and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

Mtd	Con	$\psi_{0(0)}$	$\psi_{0(1)}$	$\psi_{0(2)}$	$\psi_{0(3)}$	$\psi_{1(1)}$	$\psi_{1(2)}$	$\psi_{1(3)}$	$\psi_{2(2)}$	$\psi_{2(3)}$	$\psi_{3(3)}$
1	no	1.589	1.923	2.026	0.753	1.967	2.670	2.217	2.197	2.695	2.458
1cw	no	1.559	1.894	2.031	0.755	1.920	2.629	2.220	2.162	2.642	2.413
2	no	0.342	0.347	0.324	0.356	0.420	0.441	0.461	0.453	0.429	0.444
1	yes	1.021	1.312	1.512	0.729	1.021	1.312	1.512	1.021	1.312	1.021
1cw	yes	0.999	1.298	1.515	0.733	0.999	1.298	1.515	0.999	1.298	0.999
2	yes	0.201	0.233	0.266	0.354	0.201	0.233	0.266	0.201	0.233	0.201

Table 17

SEs (×10) of parameter estimates when n=1000, times between visits are divided by four and visits are irregular and there is random censoring. 'Mtd' is method ('1cw' is Method 1 with censoring weighting) and 'Con' is whether constraint $\psi_{k(k+m)} = \psi_{k'(k'+m)}$ is imposed.

$\mathcal{A}_0$	$\mathcal{A}_1$	Mtd	$\psi_{0(0)}$		$\psi_{0(1)}$		$\psi_{1(1)}$	
	True		0.300		0.100		0.300	
correct	correct	1	0.293	(0.000)	0.101	(0.002)	0.298	(0.001)
		2	0.295	(0.000)	0.101	(0.001)	0.301	(0.001)
		3	0.295	(0.000)	0.101	(0.001)	0.300	(0.001)
correct	misspec	1	0.292	(0.000)	0.104	(0.001)	0.262	(0.001)
		2	0.294	(0.000)	0.100	(0.001)	0.301	(0.001)
		3	0.294	(0.000)	0.100	(0.001)	0.302	(0.001)
misspec	correct	1	0.262	(0.000)	0.087	(0.001)	0.298	(0.001)
		2	0.298	(0.000)	0.104	(0.001)	0.302	(0.001)
		3	0.300	(0.000)	0.104	(0.001)	0.301	(0.001)
misspec	misspec	1	0.262	(0.000)	0.088	(0.001)	0.262	(0.001)
		2	0.298	(0.000)	0.102	(0.001)	0.300	(0.001)
		3	0.300	(0.000)	0.101	(0.001)	0.301	(0.001)

Table 18

Results from simulation study to investigate double robustness of Methods 2 and 3. Mean estimates ($\times -1$) over 5000 simulated datasets are shown, along with Monte Carlo standard errors of these means (in brackets). 'Mtd' means method. Numbers in red indicate estimators expected to be inconsistent.