BIOMETRIC METHODOLOGY

Biometrics WILEY

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

Shaun Seaman[1] | Oliver Dukes[2] | Ruth Keogh[3] | Stijn Vansteelandt[2,3]

[1]MRC Biostatistics Unit, University of Cambridge, Institute of Public Health, Cambridge, UK

[2]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

[3]London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Shaun Seaman, MRC Biostatistics Unit, University of Cambridge, Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK.
Email: shaun.seaman@mrc-bsu.cam.ac.uk

Abstract

Accounting for time-varying confounding when assessing the 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 (AFTMs) 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 generalized linear models and describe two more efficient, double robust, closed-form estimators. All three estimators avoid the artificial recensoring of AFTMs 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.

KEYWORDS

accelerated failure time model, Aalen's additive 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 (eg, disease severity) are often affected by

earlier exposures (eg, 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 a history of exposure and confounders, at each timepoint (Robins *et al.*, 2000). The reason for this is twofold. First, such adjustment procedures eliminate the indirect effects of early exposures on survival that are mediated through those confounders. For example, it would be undesirable to eliminate the 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 noninformative 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 modeling 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 the 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, that is, 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.* 2011) 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 parameterize 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 generalizes Picciotto *et al.*'s model to continuous time and parameterizes 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 of being parameterized in continuous time, SNCSTMs can handle irregular measurement times and allow the 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, that is, the same for all individuals, and (for notational simplicity) are $1, 2, ..., K$, 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 $1, ..., K$, 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 ($k = 0, ..., K$). 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 $\bar{A}_{ki} = (A_{0i}, ..., A_{ki})^\top$, $\bar{L}_{ki} = (L_{0i}, ..., L_{ki})^\top$, and $A_{-1,i} \equiv \varnothing$. The causal ordering of the variables is $\{L_0, A_0, T \wedge 1, L_1, A_1, T \wedge 2, ..., L_K, A_K, T \wedge (K + 1)\}$, 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 onward had been set to zero by an intervention. We make the consistency assumption that $T_i = T_i(\bar{A}_{ki}, 0)$ with probability one for individuals with $A_{k+1,i} = \cdots = A_{Ki} = 0$. Note $T_i(\bar{A}_{k-1,i}, 0) \geqslant k$ if and only if $T_i(\bar{A}_{li}, 0) \geqslant k$ for all $l = k, ..., K$, that is, 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): $T(\bar{A}_{k-1}, 0) \perp\!\!\!\perp A_k \mid \bar{L}_k, \bar{A}_{k-1}, T \geqslant k$ ($k = 0, ..., K$) (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

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

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 $k + 1$ onward (ie, $A_{k+1}, ..., A_K$) 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, that is, the effect of A_k not mediated through the later exposures $A_{k+1}, ..., A_l$. For example, if $\psi_{k(k)}, ..., \psi_{k(K)}$ are all positive and $A_k > 0$, then intervening to set $A_k = 0$ is beneficial, that is, exposure is harmful. Conversely, if $\psi_{k(k)}, ..., \psi_{k(K)}$ 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

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

for $t \in [l, l + 1)$ (with $l = k, ..., K$), 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 $A_{k+1}, ..., A_l$ 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 $E\{dN_{(\bar{A}_{k-1}, 0)}(t) \mid \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\}$—which, by NUC, equals $E\{dN_{(\bar{A}_{k-1}, 0)}(t) \mid \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\}$—as a totally unspecified "baseline" hazard, rather than parameterizing its dependence on $\bar{A}_{k-1}$ and $\bar{L}_k$. One advantage of this is that the danger of incompatibility between Models $\mathcal{M}_0, ..., \mathcal{M}_K$ is avoided. To illustrate this danger, suppose it were assumed that $E\{dN(t) \mid \bar{A}_1, \bar{L}_1, T \geqslant t\} = \phi_{10}(t) + \phi_{1A_0}(t)A_0 + \phi_{1\bar{L}_1}(t)^\top \bar{L}_1 + \psi_{1(1)}A_1$ 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 $\mathcal{A}_k$ ($k = 0, ..., K$). Model $\mathcal{A}_k$ is a generalized linear model (GLM) for A_k given $\bar{A}_{k-1}, \bar{L}_k$ and $T \geqslant k$ with $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k, T \geqslant k)\} = \alpha_{k0}^\top H_k$, 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, For example, $H_k = (1, A_{k-1}, L_k^\top)^\top$. 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 to 4.3 consistently estimate $\psi_{k(l)}$ when Models $\mathcal{M}_k$ and $\mathcal{A}_k$ ($k = 0, ..., K$) 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 $\mathcal{A}_k, \mathcal{M}_k, ..., \mathcal{M}_K$ 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 $t \geqslant k$,

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

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

First we estimate $\psi_{k(k)}$ ($k = 0, ..., K$) 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 $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k, T \geqslant t)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)}(t - k)$ 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 $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k, Q)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)}(Q - k)$ to the resulting set of (up to $10n$) pseudo-individuals. A consistent estimate of $\psi_{k(k)}$ is then $\hat{\psi}_{k(k)}^{M1} = -\hat{\alpha}_{k(k)}/\phi_k$. 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 $\psi_{k(k+1)}$ ($k = 0, ..., K - 1$). When $t \in [k + 1, k + 2)$, Equation (3) is $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)\} = \alpha_{k0}^\top H_k + \alpha_{k(k)} + \alpha_{k(k+1)}(t - k - 1)$. 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, $\mathcal{M}_k, ..., \mathcal{M}_K$ imply that for $t \geqslant k + 1$,

$$\begin{aligned} P\{T(\bar{A}_k, 0) \geqslant t | \bar{A}_k, \bar{L}_k, T(\bar{A}_k, 0) \geqslant k\} \\ = E\{R(t)w_k(t) | \bar{A}_k, \bar{L}_k, T \geqslant k\}, \end{aligned} \tag{4}$$

where $w_k(t) = \prod_{j=k+1}^{K} \exp\{A_j v_j(t)^\top \psi_j\}$. 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 $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k, Q)\} = \alpha_{k0}^\top H_k + \alpha_{k(k+1)}(Q - k - 1)$ to a set of pseudo-individuals constructed as described above but using 10 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 $\hat{\psi}_{k(k)}^{M1}$. A consistent estimate of $\psi_{k(k+1)}$ is then $\hat{\psi}_{k(k+1)}^{M1} = -\hat{\alpha}_{k(k+1)}/\phi_k$.

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

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

to a set of pseudo-individuals constructed using 10 equally spaced values of t between l and $l + 1$ and using weights $w_k(Q)$, with $\psi_{j(m)}$ replaced by $\hat{\psi}_{j(m)}^{M1}$. For later reference, we denote the fitted value of $E(A_k | \bar{A}_{k-1}, \bar{L}_k, Q = t)$ thus

obtained as $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$. This is an estimate of $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$. Note that $\hat{\psi}_{j(m)}^{\mathrm{M1}}$ ($k < j \leqslant m \leqslant l$) must be calculated before $\hat{\psi}_{k(l)}^{\mathrm{M1}}$. 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 $\psi_{k(k)} = \cdots = \psi_{k(K)} = 0$, that is, 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 "stabilized") weights $w_k^*(Q)$ in place of $w_k(Q)$, where $w_k^*(t) = \exp\left\{\sum_{j=k+1}^{K} \Delta_{j(k)}^* v_j(t)^\top \psi_j\right\}$ and $\Delta_{j(k)}^* = A_j - E(A_j | \bar{A}_{k-1}, \bar{L}_k, T \geqslant j)$ ($j = k + 1, ..., K$). 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 $E(A_j | \bar{A}_{k-1}, \bar{L}_k, T \geqslant j)$ ($j = k + 1, ..., K$) 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 $\psi_{k(k)}$ ($k = 0, ..., K$) by solving unbiased estimating equation

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

where $\Delta_k(t) = A_k - E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$. The expectation $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$ 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 the 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 $(\bar{A}_{k-1}, \bar{L}_k)$ 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 $E\{dN_{(\bar{A}_{k-1}, 0)}(t) | \bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geqslant t\} = \gamma_{k(k)}^\top H_k$ 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 $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant k)$ is a constant given $(\bar{A}_{k-1}, \bar{L}_k)$, 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 $\sum_{i=1}^{n} \int_{k+1}^{k+2} R_i(t) \exp\{A_{k+1,i}\psi_{k+1(k+1)}(t - k - 1)\} \Delta_{ki}(t)[dN_i(t) - \{A_{k+1,i}\psi_{k+1(k+1)} + \Delta_{ki}(k + 1)\, \psi_{k(k+1)}\}dt] = 0$. The unknown $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geqslant t)$ and $\psi_{k+1(k+1)}$ are replaced by $\hat{e}_{k(k+1)}(\bar{A}_{k-1}, \bar{L}_k, t)$ and $\hat{\psi}_{k+1(k+1)}^{\mathrm{M2}}$. 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 $\hat{\psi}_{k(l)}^{M2}$ of $\psi_{k(l)}$ ($l \geq k$) is obtained by solving

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

$$\tag{7}$$

with $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geq t)$ replaced by $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ and $\psi_{j(l)}$ ($j > k$) replaced by $\hat{\psi}_{j(l)}^{M2}$; this requires that $\psi_{j(m)}$ ($k < j \leq m \leq l$) be estimated before $\psi_{k(l)}$. The estimator $\hat{\psi}_{k(l)}^{M2}$ 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 $\hat{\psi}_{k(l)}^{M2}$ 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 $\mathcal{B}_{k(l)}$ ($l \geq k$) be defined by the restriction $E\{dN_{(\bar{A}_{k-1},0)}(t)|\bar{A}_k, \bar{L}_k, T(\bar{A}_{k-1}, 0) \geq t\} = \{\gamma_{k(l)}^{\top} H_k - e_{k(l)}^*(\bar{A}_{k-1}, \bar{L}_k, k) \psi_{k(l)}\}dt$ $\forall t \in [l, l+1)$, where $\gamma_{k(l)}$ are unknown parameters. $\hat{\psi}_{k(l)}^{M2}$ is consistent if (a) $\mathcal{M}_k, ..., \mathcal{M}_l$, (b) either $\mathcal{A}_k$ or $\mathcal{B}_{k(l)}$, and (c) for each $j = k+1, ..., l$, either $\mathcal{A}_j$ or all of $\mathcal{B}_{j(j)}, ..., \mathcal{B}_{j(l)}$ 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 $e_{k(l)}^*(\bar{A}_{k-1}, \bar{L}_k, k) \psi_{k(l)}$ is a linear function of H_k, it can be omitted. As in method 1, efficiency may be gained by using stabilized weights $w_{ki}^*(t)$ in place of $w_{ki}(t)$ in Equation (7). Also, to make $\hat{\psi}_{k(l)}^{M2}$ invariant to additive transformations of $A_{k+1}, ..., A_l$, 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 $E(A_k | \bar{A}_{k-1}, \bar{L}_k, T(\bar{A}_k, 0) \geq t)$ 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 $t \geq k$, we calculate $\hat{e}_{k(l)}(\bar{A}_{k-1}, \bar{L}_k, t)$ by fitting the GLM $g\{E(A_k | \bar{A}_{k-1}, \bar{L}_k)\} =$

$\alpha_{k0}(t)^{\top}H_k$ to the set of individuals with $T \geq t$, 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 "$\hat{\psi}_{\text{TV PS-DR}}$" and, on the basis of results from a simulation study, recommended it over three alternatives. As in method 2, we can use stabilized 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, that is, 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, that is, to vary from one individual to another, and effect modification, that is, 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 ($k = 1, ..., K$), and let $S_{0i} = 0$ ($i = 1, ..., n$) and $\bar{S}_i = (S_{1i}, ..., S_{Ki})^{\top}$. 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 L_0, and we modify NUC to be $T(\bar{A}_{k-1}, 0) \perp\!\!\!\perp A_k | \bar{L}_k, \bar{A}_{k-1}, \bar{S}, T \geq S_k (k = 0, ..., K)$. 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 $Z_{k(l)} = (1, Z_{k(l)}^{int\top})^{\top}$, where $Z_{k(l)}^{int}$ is a known (possibly vector) function of $(\bar{A}_{k-1}, \bar{L}_k, \bar{S})$ ("int" stands for "interactions"), and let $Z_k = (Z_{k(k)}^{\top}, ..., Z_{k(K)}^{\top})^{\top}$.

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

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

where $v_k(t, Z_k, \bar{S})$ equals $((t - S_k)Z_{k(k)}^\top, 0, ..., 0)^\top$ if $t \in [S_k, S_{k+1})$, equals $((S_{k+1} - S_k)Z_{k(k)}^\top, (t - S_{k+1})Z_{k(k+1)}^\top, 0, ..., 0)^\top$ if $t \in [S_{k+1}, S_{k+2})$, and equals $((S_{k+1} - S_k)Z_{k(k)}^\top, (S_{k+2} - S_{k+1}) Z_{k(k+1)}^\top, (t - S_{k+2})Z_{k(k+2)}^\top, 0, ..., 0)^\top$ if $t \in [S_{k+2}, S_{k+3})$, and so forth. If $S_k = k$ and $Z_{k(l)} = 1$, Equation (8) reduces to Equation (1). Model $\mathcal{M}_k$ can also be written as $E\{dN_{(\bar{A}_{k-1}, 0)}(t) | \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_{k-1}, 0) \geqslant t\} = E\{dN_{(\bar{A}_k, 0)}(t) | \bar{A}_k, \bar{L}_k, \bar{S}, T(\bar{A}_k, 0) \geqslant t\} - A_k \psi_{k(l)}^\top Z_{k(l)} dt$ 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 to 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 $E\{dN_C(t) | C \geqslant t, \bar{A}_{\lfloor \tilde{T} \rfloor}, \bar{L}_{\lfloor \tilde{T} \rfloor}, \bar{S}, \tilde{T} > t, \tilde{T}\} = \lambda(t, \bar{A}_{\lfloor t \rfloor}, \bar{L}_{\lfloor t \rfloor}, \bar{S}) \forall t$, 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 the bias and efficiency of the methods. There were $K + 1 = 4$ visits and two time-dependent confounders (ie, dim $(L_k) = 2$). These and the exposure were generated as: $L_0 \sim N((0, 0), \Sigma)$, $A_0 \sim N(3 + (0.2, 0.1)^\top L_0, 0.9^2)$, $L_k \sim N(\Omega L_{k-1} + (0.1, 0.05)^\top A_{k-1}, \Sigma)$ 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, $\mathcal{M}_k$ $(k = 0, ..., K)$ is correctly specified with no effect modification (ie, $Z_{k(l)} = 1$) 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 data sets, 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 to 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 data sets. 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 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 to 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 *et al.*, 2019), but its effect on survival has not been studied. Baseline visit

was defined as an individual's first visit during 2008 to 2015, and there were up to $K = 8$ follow-up visits. The (irregular) visit times were defined as years after baseline visit; the median time between visits was 1.00 years (interquartile range 0.93–1.07). 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 nonadministrative 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 nonadministrative censorings (including 36 transplants). Of those who died, 74 (63) were treated (untreated) at the time of death. Total follow-up was 12 380 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

TABLE 1 Means ($\times 10$) and SEs ($\times 10$) of parameter estimates when $n = 1000$, visits are regular and the only censoring is administrative

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

Note: "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.

TABLE 2 Means (×10) and SEs (×10) of parameter estimates when $n = 1000$, visits are regular and censoring is random

| 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

Note: "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.

k and conditional on confounder history $\bar{L}_k$. More generally, $\exp\left(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\right)$ 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, and not assigned), and time-varying variables FEV$_1$%, body mass index, days of IV antibiotic use, and binary indicators for four infections (*Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Burkholderia cepacia complex*, and *Aspergillus*), CF-related diabetes, smoking, and use of other mucoactive treatments and oxygen therapy. The same variables (and treatment) were included in models for the inverse probability of censoring weights.

Figure 1A shows estimates of $\exp\left(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\right)$ from method 2. These suggest that starting treatment now rather than waiting may cause a small decrease in the probability of survival, at least for the first 5 years: $\exp\left(-\sum_{l=k}^{k+m-1}\psi_{k(l)}\right) = 0.997, 0.996, 0.997, 0.994,$ and 0.988 for $m = 1, ..., 5$, respectively. However, the confidence intervals (obtained by bootstrapping) include 1, that is, 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

TABLE 3 Coverage (%) of 95% bootstrap confidence intervals for methods 1, 2, and 1cw (ie, 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
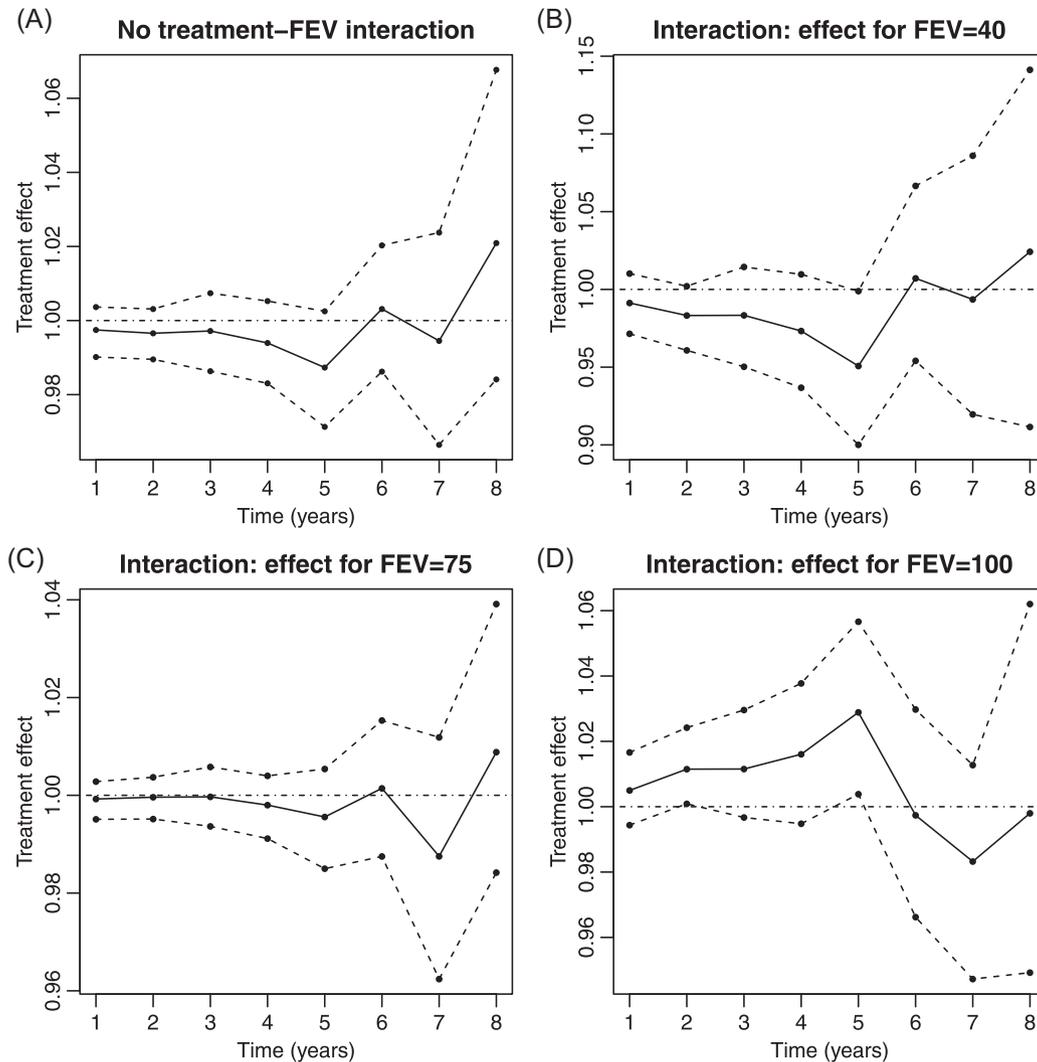
FIGURE 1 Estimates of the ratio of the survival probabilities when treatment is initiated immediately compared to initiation being delayed by one year. A: from the model with no interaction. B, C and D: from the model with interaction between treatment and FEV$_1$%

were between 4 and 9 times wider than those from method 2.

For illustration, we also fitted an SNCSTM with an interaction between treatment and FEV$_1$%. Figure 1B to 1D shows the estimated ratios of survival probabilities for three values of 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 FEV$_1$%=100, that is, starting treatment now may be better than waiting for patients with high 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, that is, 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 to 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 AFTMs, 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, that is, $P\{T(0) \geqslant t\}$, 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 $P\{T(0) \geqslant t\} = E[R(t)\prod_{j=0}^{K}\exp\{A_j v_j(t, Z_j, \bar{S})^{\top}\psi_j\}]$. If there is censoring before time t, $R_i(t)$ should be inversely weighted by an estimate of $P(C_i \geqslant t | \bar{A}_{\lfloor t \rfloor i}, \bar{L}_{\lfloor t \rfloor i}, \bar{S}_i)$.

A limitation is that, like other additive hazards models, the SNCSTM does not constrain hazards to be nonnegative, 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 appeal 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 discretize time so as to approximate continuous time and (as Picciotto *et al.* note) to reformulate it as a model for the 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 the 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.

ORCID

Shaun Seaman 🅳 http://orcid.org/0000-0003-3726-5937
Oliver Dukes 🅳 http://orcid.org/0000-0002-9145-3325
Ruth Keogh 🅳 http://orcid.org/0000-0001-6504-3253
Stijn Vansteelandt 🅳 http://orcid.org/0000-0002-4207-8733

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. *et al.* (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.