Relation between three classes of structural models for the effect of a time-varying exposure on survival

Noname manuscript No.
(will be inserted by the editor)

Relation between three classes of structural models for the effect of a time-varying exposure on survival

Jessica G. Young · Miguel A. Hernán · Sally Picciotto · James M. Robins

Received: date / Accepted: date

Abstract Standard methods for estimating the effect of a time-varying exposure on survival may be biased in the presence of time-dependent confounders themselves affected by prior exposure. This problem can be overcome by inverse probability weighted estimation of Marginal Structural Cox Models (Cox MSM), g-estimation of Structural Nested Accelerated Failure Time Models (SNAFTM) and g-estimation of Structural Nested Cumulative Failure Time Models (SNCFTM). In this paper, we describe a data generation mechanism that approximately satisfies a Cox MSM, an SNAFTM and an SNCFTM. Besides providing a procedure for data simulation, our formal description of a data generation mechanism that satisfies all three models allows one to assess the relative advantages and disadvantages of each modeling approach. A simulation study is also presented to compare effect estimates across the three models.

1 Introduction

Commonly used methods to estimate the effect of a time-varying treatment on mortality model the hazard at time t conditional on treatment and covariate history through time t (e.g., a Cox model) (Cox and Oakes, 1984). This standard approach, however, may be biased in the presence of a time-dependent covariate (Robins, 1986; Hernán et al, 2004) that is:

1. a time-dependent confounder, i.e., it affects both future risk of failure and treatment and
2. affected by past treatment.

This work was supported by NIH grant R01 HL080644.
The original publication is available at www.springerlink.com
see http://www.springerlink.com/content/g31gl43370611421/

Address(es) of author(s) should be given

As an example, consider an observational study of the effect of diet on risk of coronary heart disease. The time-varying covariate "diagnosis of diabetes" is a time-dependent confounder because a diagnosis of diabetes affects future dietary choices and is a risk factor for coronary heart disease. In addition, prior diet affects future risk of diabetes.

Robins and collaborators have developed methods to appropriately adjust for measured time-varying confounders that are affected by past treatment (for a review of these methods see Robins and Hernán (2009)). In the high-dimensional failure time setting, these methods include inverse probability weighting of marginal structural Cox models (Cox MSM) (Robins, 1998a; Hernán et al, 2000), g-estimation of structural nested accelerated failure time models (SNAFTM) (Robins et al, 1992, 1993; Hernán et al, 2005), and g-estimation of structural nested cumulative failure time models (SNCFTM) (Page et al, 2008; Picciotto et al, 2008, 2009).

This paper describes the relations between these three models. In previous work (Young et al, 2008) we described a data generation mechanism (with no modification of the treatment effect by time-varying covariates) that satisfied both a Cox MSM and an SNAFTM. In this paper, we describe a data generation mechanism that approximately satisfies a Cox MSM, an SNAFTM and an SNCFTM. Besides providing a procedure for data simulation, our formal description of a data generation mechanism that satisfies all three models allows one to assess the relative advantages and disadvantages of each modeling approach.

This paper is structured as follows. In §2 we describe the data structure of interest. In §3 we review general definitions of the SNAFTM, Cox MSM and SNCFTM and briefly describe associated estimation procedures and inference. In §4 we describe sufficient conditions for a data generation mechanism that satisfies all three models. In §5 we present results of a simulation study that compares estimators of the parameters of the three models both when using data generated under those sufficient conditions, and when using data in which the conditions are violated. In §6 we discuss our results.

2 Data structure and identifying assumptions

Consider a longitudinal study with n subjects and observation times $m = 0, 1, 2, ..., K+1$. Let T be a failure (death) time variable that may be either exactly observed or interval censored, Y_m the indicator for death by time m (1 = yes, 0 = no), V_m a vector of time-varying covariates measured at the start of the interval $[m, m+1)$, and A_m a treatment indicator (1 = yes, 0 = no) during the interval $(m, m+1]$. We use overbars to represent a variable's history, i.e., $\overline{V}_K = (V_0, V_1, ..., V_m, ..., V_K)$. By convention, (i) $Y_0 = 0$ and (ii) if $Y_m = 1$ then $V_m = 0$, $A_m = 0$, and $Y_{m+1} = 1$. Those who do not die before the last observation time $K+1$ are said to be administratively censored. The observed data consists of n i.i.d. copies of

$$O = \{\overline{L}_{K+1}, \overline{A}_K\}$$

where

$$\overline{L}_{K+1} = \{\overline{Y}_{K+1}, \overline{V}_K\}$$

$$\overline{Y}_{K+1} = (Y_0, ..., Y_{K+1})$$

if T is interval-censored and

$$\overline{Y}_{K+1} = (Y_u; 0 \leq u \leq K+1)$$

if T is exactly observed.

Let $g = \overline{a}$ for $\overline{a} \equiv \overline{a}_K$ in the support of $\overline{A}_K$ denote a (static or nondynamic) *treatment regime*. An example of a treatment regime is "treat continuously since baseline" or $g = (1, 1, ..., 1) = \overline{1}$. Let T_g and $\overline{V}_{K,g}$ represent the failure time and covariate history, respectively, a subject would have experienced had she, possibly contrary to fact, followed treatment regime $g = \overline{a}$. We say a subject follows treatment regime $g = \overline{a}$ if the subject takes treatment a_m at time m if alive at m. By convention, a subject takes treatment $a_m = 0$ if dead at m. Let $\overline{a}_m$ be the first m components of $\overline{a}$. The full data structure consists of the observed data O and the counterfactual data $\left(\overline{V}_{K,g}, T_g\right)$ for all $g = \overline{a}$. We can think of the observed data structure O as a missing data structure with $\left(\overline{V}_{K,g}, T_g\right)$ unobserved.

We assume the following three identifying assumptions (Robins and Hernán, 2009):

1. Consistency:
   Given $g = \overline{a}$, if $\overline{A}_m = \overline{a}_m$ then $\overline{Y}_{m+1,g} = \overline{Y}_{m+1}$ and $\overline{V}_{m+1,g} = \overline{V}_{m+1}$ where
   $$Y_{u,g} = I(T_g < u),$$
   $$\overline{Y}_{K+1,g} = (Y_{0,g}, ..., Y_{K+1,g})$$
   if T is interval-censored and
   $$\overline{Y}_{K+1,g} = (Y_{u,g}; 0 \leq u \leq K+1)$$
   if T is exactly observed. Also in the exactly observed case, if the above holds and either $T < m+1$ or $T_g < m+1$ it follows that $T_g = T$.
2. Conditional exchangeability: For any regime g and $m \in [0, K]$
   $$\left(T_g, \overline{V}_{K+1,g}\right) \coprod A_m | \overline{V}_m, \overline{A}_{m-1}, Y_m = 0$$
3. Positivity: $f_{\overline{A}_{m-1}, \overline{V}_m, Y_m}(\overline{a}_{m-1}, \overline{v}_m, 0) \neq 0 \implies \Pr(A_m = a_m | \overline{V}_m, \overline{A}_{m-1}, Y_m = 0) > 0$ w.p.1 for all a_m in the support of A_m, $0 \leq m \leq K$.

Informally, consistency is satisfied if the counterfactual outcomes are well defined, exchangeability if there is no unmeasured confounding, and positivity if there are subjects at all levels of exposure within levels of the measured confounders. See Young et al (2008) for a graphical representation of this data structure.

3 Model definitions, estimation and inference

3.1 Model definitions

Let T_0 be the counterfactual failure time under the treatment regime "never treat during the follow-up" or $g = \overline{0} \equiv \overline{0}_K = (0, 0, ...0)$. Let $T_{(\bar{a}_{m-1},0)}$ be the failure time under the regime "take treatment $\bar{a}_{m-1}$ through $m-1$ and then no more treatment" so $T_{\left(\overline{A}_{m-1},0\right)}$is the treatment regime "take treatment actually taken through $m-1$ and then no more treatment." We define three models:

An SNAFTM assumes

$$T_{(\bar{a}_{m-1},0)} \text{ and } \int_m^{T_{\bar{a}}} \exp\{\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\bar{a}}, \psi^*_{aft})\}dt$$
$$\text{have the same conditional distribution given}$$
$$\left(\overline{V}_m, \overline{A}_{m-1} = \bar{a}_{m-1}, T_{\bar{a}_{m-1}} > m\right); \tag{1}$$

an MSM assumes

$$\lambda_{T_{\bar{a}}}(t) = \lambda_0(t) \exp\{\gamma_{MSM}(t, \overline{a}_t, \psi^*_{msm})\}; \tag{2}$$

and an SNCFTM assumes

$$\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi^*_{cft})\} = \frac{E[Y_{k,g=(\overline{A}_m,0)}|\overline{V}_m, \overline{A}_m, Y_m = 0]}{E[Y_{k,g=(\overline{A}_{m-1},0)}|\overline{V}_m, \overline{A}_{m,}, Y_m = 0]}, \tag{3}$$

where $\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\bar{a}}, \psi_{aft})$ and $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm})$ are known functions, continuous in t and differentiable wrt to t except at $t = 1, 2, ..., K$; $\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})$ is a known function for $0 \leq m < k \leq K + 1$; $\lambda_{T_{\bar{a}}}(t)$ and $\lambda_0(t)$ are the hazard functions at t for $T_{\bar{a}}$ and T_0, respectively; and ψ^*_{aft}, ψ^*_{msm} and ψ^*_{cft} denote the unknown true values of the model parameters ψ_{aft}, ψ_{msm} and ψ_{cft}, respectively.

3.2 Estimation and inference

Briefly, estimating the parameters of the structural models defined above requires solving an estimating equation of the general form

$$\sum_{i=1}^{n} U_i(\psi, \widehat{\alpha}) = 0, \tag{4}$$

where ψ is ψ_{aft}, ψ_{msm}, or ψ_{cft} depending on which model is of interest, and $\widehat{\alpha}$ is a consistent estimator of the p-dimensional nuisance parameter α^* of a parametric model $\Pr(A_m = 1|\overline{V}_m, \overline{A}_{m-1}, Y_m = 0; \alpha)$ for the treatment mechanism $\Pr(A_m = 1|\overline{V}_m, \overline{A}_{m-1}, Y_m = 0)$, $0 \leq m \leq K$. The specific form of

$U(\psi, \widehat{\alpha})$ depends on the choice of model for the treatment mechanism and the choice of $\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\overline{a}}, \psi_{aft})$, $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm})$, or $\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})$. More efficient estimators of ψ exist which solve estimating equations with additional nuisance parameters.

If the model for the treatment mechanism is correctly specified and $\widehat{\alpha}$ is the MLE of α^*, then $\sqrt{n}(\widehat{\psi} - \psi^*) \longrightarrow N(0, \Sigma)$ where

$$\Sigma = \Lambda^{-1} \Gamma {\Lambda^{-1}}^T, \tag{5}$$

$\Lambda = \partial E[U(\psi, \alpha^*)]/\partial \psi^T_{\psi=\psi^*}$ when $U(\psi, \alpha^*)$ is a differentiable function of ψ, $S(\alpha^*)$ is the score for α evaluated at α^* and

$$\Gamma = var\{U(\psi^*, \alpha^*) - E[U(\psi^*, \alpha^*)S(\alpha^*)^T]E[S(\alpha^*)S(\alpha^*)^T]^{-1}S(\alpha^*)\}.$$

A consistent estimator of Σ is given by

$$\widehat{\Sigma} = \widehat{\Lambda}^{-1} \widehat{\Gamma} {\widehat{\Lambda}^{-1}}^T \tag{6}$$

where $\widehat{\Lambda} = P_n[\frac{\partial U(\psi, \widehat{\alpha})}{\partial \psi^T}|_{\psi=\widehat{\psi}}]$, $\widehat{\Gamma} = P_n\{AA^T\}$ and

$$A = U(\widehat{\psi}, \widehat{\alpha}) - P_n[U(\widehat{\psi}, \widehat{\alpha})S(\widehat{\alpha})^T]P_n[S(\widehat{\alpha})S(\widehat{\alpha})^T]^{-1}S(\widehat{\alpha}).$$

Differences in the specific form of $U(\psi, \widehat{\alpha})$ associated with each model result in varying degrees of computational complexity. For the Cox MSM, the inverse probability weighted estimator of ψ^*_{msm} that solves $\sum_{i=1}^n U_i(\psi_{msm}, \widehat{\alpha}) = 0$ can be computed using standard off-the-shelf software. Robust variance estimates that lead to conservative Wald confidence intervals for ψ^* are also straightforward to obtain using off-the-shelf software although, if desired, consistent estimates of the limiting variance estimates can be obtained from equation (6). In contrast, for the SNAFTM, the estimating equation $\sum_{i=1}^n U_i(\psi_{aft}, \widehat{\alpha}) = 0$ is non differentiable with respect to ψ_{aft} when there is administrative censoring (i.e., when not all subjects have failed by end of follow-up at $K + 1$) and so-called 'artificial' censoring is used to guarantee unbiasedness of the estimating function. As a consequence, solving $\sum_{i=1}^n U_i(\psi_{aft}, \widehat{\alpha}) = 0$ requires search-based algorithms (e.g., bisection method for one-dimensional ψ_{aft}, Nelder-Mead Simplex method in general). G-estimation of an SNCFTM is somewhat more computationally involved than inverse probability weighted estimation of a Cox MSM, but the estimating function $U(\psi_{cft}, \widehat{\alpha})$ is a continuously differentiable function of ψ_{cft}, even in the presence of administrative censoring. Thus, the estimating equation can be generally solved using a Newton-Raphson type procedure.

For more details on estimation of a Cox MSM, SNAFTM and SNCFTM see Hernán et al (2005), Hernán et al (2000) and Page et al (2008), respectively. For more on general inference for estimators obtained using estimating equation methodology see van der Laan and Robins (2002).

4 A data generation mechanism that satisfies all three models

The following theorem states sufficient conditions for the generation of data that satisfies an SNAFTM and a Cox MSM. Note that this is a special case of the more general theorem presented in Young et al (2008). Proofs of all theorems are presented in the appendix.

Theorem 1 *Suppose the counterfactual failure times $T_{\overline{a}}$ follow an SNAFTM (1) with $\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\overline{a}}, \psi_{aft}) = a_t \times \psi_{aft}$. Further assume that T_0 has an exponential distribution with hazard $\lambda_{T_0}(t) = \lambda_0$. Then the $T_{\overline{a}}$ also follow a Cox MSM with $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$ and $\psi^*_{msm} = \psi^*_{aft}$.*

Note that, in this case, $\exp\{\psi^*_{msm}\}$ is the hazard ratio comparing the regimes "always treat" vs. "never treat."

The next theorem provides conditions under which there is approximate equivalence between an SNAFTM and SNCFTM.

First, given the SNAFTM $\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\overline{a}}, \psi_{aft}) = a_t \times \psi_{aft}$, define for $u \geq m$

$$
\begin{aligned}
& h(u, \overline{A}_m, \psi_{aft}) \\
& = \int_0^m \exp\left(A_t \times \psi_{aft}\right) dt + \int_m^{\min(u, m+1)} \exp\left(A_t \times \psi_{aft}\right) dt \\
& + I\left(u > m+1\right)\left\{u - (m+1)\right\} \\
& = I\left(m \leq u < m+1\right)\left\{\sum_{l=0}^{m-1} e^{\psi_{aft} A_l} + (u-m)\, e^{\psi_{aft} A_m}\right\} \\
& + I\left(m+1 \leq u\right)\left\{\sum_{l=0}^{m} e^{\psi_{aft} A_l} + u - (m+1)\right\} \qquad (7)
\end{aligned}
$$

Note $\Pr\left(j < T_{g=(\overline{A}_m, 0)} < u | \overline{A}_m, V\right) = \Pr\left[h(j, \overline{A}_m, \psi^*_{aft}) < T_0 < h(u, \overline{A}_m, \psi^*_{aft}) | \overline{A}_m, V\right]$ under the above SNAFTM.

Theorem 2 *Suppose that the following assumptions hold in addition to those of Theorem 1 and §2:*

1. *the conditional distribution of V_m given $\overline{A}_{m-1}$, $\overline{V}_{m-1}, T_0$ depends on T_0 only through the function $I(T_0 < c)$ for a constant c such that $c > \max(\{h(K, \overline{A}_m), h(K, \overline{A}_{m-1})\}$ and*
2. *failure is rare in the sense that $S_{T_0}(t) \approx 1$ for $t < \max_{m \in \{0,\ldots,K\}} \{h(K, \overline{A}_m)\}$ where $S_{T_0}(t)$ is the survival function for T_0 at t, and $A \approx B$ means A and B are approximately equal.*

It then follows that the SNCFTM:

$$\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi^*_{cft})\} = 1 + \frac{e^{A_m \psi^*_{cft}} - 1}{k - m}$$

approximately holds with $\psi^*_{cft} = \psi^*_{aft}$.

When the probability of failure in any interval $(m, m+1)$ is small, $e^{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})}$ with $k = m+1$ approximates the conditional hazard ratio $\frac{\lambda_{T_{(\overline{A}_m,0)}}(t|\overline{V}_m,\overline{A}_m)}{\lambda_{T_{(\overline{A}_{m-1},0)}}(t|\overline{V}_m,\overline{A}_m)}$ at time $t \in (m, m+1]$ given $\overline{V}_m, \overline{A}_m$, for regime $g = (\overline{A}_m, 0)$ versus regime $g = (\overline{A}_{m-1}, 0)$ if, as we assume, the conditional hazard ratio is nearly constant in the interval $(m, m+1]$. Consider a correctly specified SNCFTM (3) with the form

$$\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})\} = 1 + \frac{e^{A_m\psi_{cft}} - 1}{k - m}$$

Under this model, $\exp\{\gamma_{CFT,m+1}(\overline{V}_m, \overline{A}_m, \psi_{cft})\} = e^{A_m\psi_{cft}}$. Note $\frac{\lambda_{T_{(\overline{A}_m,0)}}(t|\overline{V}_m,\overline{A}_m)}{\lambda_{T_{(\overline{A}_{m-1},0)}}(t|\overline{V}_m,\overline{A}_m)} = e^{A_m\psi_{cft}}$ does not imply $\frac{\lambda_{T_{(\overline{A}_m,0)}}(t)}{\lambda_{T_{(\overline{A}_{m-1},0)}}(t)} = e^{A_m\psi_{cft}}$. If it did we could conclude that the MSM $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$ was correctly specified with $\psi^*_{msm} = \psi^*_{cft}$. However, under the additional assumptions of Theorems 1 and 2, we can conclude that $\frac{\lambda_{T_{(\overline{A}_m,0)}}(t|\overline{V}_m,\overline{A}_m)}{\lambda_{T_{(\overline{A}_{m-1},0)}}(t|\overline{V}_m,\overline{A}_m)}$ approximates $\frac{\lambda_{T_{(\overline{A}_m,0)}}(t)}{\lambda_{T_{(\overline{A}_{m-1},0)}}(t)}$, and thus that the MSM $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$ with $\psi^*_{msm} = \psi^*_{cft}$ approximately holds.

The following theorem provides sufficient conditions for data generation that satisfies a SNAFTM, Cox MSM and SNCFTM in the special case where $\psi^*_{aft} = 0$.

Theorem 3 *Suppose the counterfactual failure times $T_{\overline{a}}$ follow a SNAFTM (1) with $\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\overline{a}}, \psi_{aft}) = a_t \times \psi_{aft}$ and $\psi^*_{aft} = 0$. Then:*

1. *the $T_{\overline{a}}$ follow a Cox MSM with $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$ and $\psi^*_{msm} = \psi^*_{aft}$ and*
2. *the following SNCFTM:*

$$\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi^*_{cft})\} = 1 + \frac{e^{A_m\psi^*_{cft}} - 1}{k - m}$$

holds with $\psi^*_{cft} = \psi^*_{aft}$.

5 Simulation study

We generated data consistent with the conditions stated in Theorems 1 and 2, and under the full data structure described in §2. The simulations consisted of 1000 samples, each with 2500 subjects and $K + 1 = 10$ observation times. Each sample was generated according to the general algorithm described in Young et al (2008) for SNAFTM data generation. Here, this algorithm was specifically implemented as follows:

For each of 2500 simulated subjects:

step 1: Simulate the counterfactual T_0 from an Exponential distribution with $\lambda_0 = 0.01$.

Define $V_{-1} = A_{-1} = Y_0 = 0$. Then for each $m \in [0, 9]$ implement steps 2-4:
step 2: Simulate V_m from $\text{logit}[\Pr(V_m = 1|\overline{V}_{m-1}, \overline{A}_{m-1}, T_0, Y_m = 0; \beta)] = \beta_0 + \beta_1 I(T_0 < c) + \beta_2 A_{m-1} + \beta_3 V_{m-1}$ where $\beta = \{\log(\frac{3}{7}), 2, \log(0.5), \log(1.5)\}$ and $c = 30$.
step 3: Simulate A_m from $\text{logit}[\Pr(A_m = 1|\overline{V}_m, \overline{A}_{m-1}, Y_m = 0; \alpha)] = \alpha_0 + \alpha_1 V_m + \alpha_2 V_{m-1} + \alpha_3 A_{m-1}$ where $\alpha = \{\log(\frac{2}{7}), 0.5, 0.5, \log(4)\}$.
step 4: Simulate Y_{m+1}, and possibly T, based on the following:

- if $T_0 > \int_0^{m+1} \exp\{\psi_{aft} \times A_j\} dj$ then $Y_{m+1} = 0$;
- else if $T_0 \leq \int_0^{m+1} \exp\{\psi_{aft} \times A_j\} dj$ then $Y_{m+1} = 1$ and $T \in (m, m+1]$ with $T = m + (T_0 - \int_0^m \exp\{\psi_{aft} \times A_j\} dj) \exp\{-\psi_{aft} \times A_m\}$.

Finally, redefine $V_l = 0, A_l = 0$ for $l > T$.

SAS code to implement the above algorithm is provided at www.hsph.harvard.edu/causal/software.htm.

Tables 1 through 3 display simulation results for the inverse probability weighted estimates $\hat{\psi}_{msm}$, and the g-estimates $\hat{\psi}_{cft}$ and $\hat{\psi}_{aft}$. The true value of the parameter ψ^*_{aft} was varied to be either -0.3, 0.0 or 0.3. Each table reports the mean of the model parameter estimates across Monte Carlo simulation samples (MC Mean), the difference between this mean and the true value of the parameter ψ^*_{aft} (Bias), variance of the model parameter estimates across samples (MC Var), the test statistic $T = \frac{\text{Bias}}{\sqrt{\text{MC Var}/2500}}$ and the two-sided p-value comparing T to a $N(0, 1)$ (p-value).

Results in Table 1 confirm that the estimators of ψ^*_{msm}, ψ^*_{cft} and ψ^*_{aft} are essentially unbiased when data are generated under the assumptions of Theorems 1 and 2.

Tables 2 and 3 display simulation results under a data generation mechanism in which the conditions of Theorems 1 and 2 are violated. Specifically, results presented in Table 2 are based on data generated as in Table 1, except with T_0 generated from a Weibull distribution with shape and scale parameters 2 and 0.02, respectively, which violates the condition that T_0 is exponentially distributed. Results presented in Table 3 differ from those of Table 1 in that $\lambda_0 = 0.1$ (as opposed to 0.01), which violates the rare disease condition defined in Theorem 2.

As expected, for $\psi^*_{aft} \neq 0$, the results reported in Table 2 confirm that violation of the exponential condition results in biased estimators of ψ^*_{msm} and ψ^*_{cft} as the data are no longer generated under a Cox MSM or an SNCFTM. Also as expected under $\psi^*_{aft} \neq 0$, violation of the rare disease condition results in biased estimators of ψ^*_{cft} (see Table 3).

In theory, the performance of the inverse probability weighted estimator $\hat{\psi}_{msm}$ of ψ^*_{msm} should be unaffected by violations of the rare disease condition. However, as is common practice (Hernán et al, 2000) we approximated $\hat{\psi}_{msm}$ via a weighted logistic regression model, which requires the rare disease condition in every time interval. This approximation may explain the poorer performance of the inverse probability weighted estimator that is seen in Table 3.

Table 1 Monte Carlo simulation results for estimators of the parameter of a Cox MSM, SNCFTM and SNAFTM when data are generated under the assumptions of §2 and Theorems 1 and 2 for various values of ψ^*_{aft} based on 1000 replicates, $n = 2500$ and $K + 1 = 10$.

ψ^*_{aft}	Model	MC Mean	Bias	MC Var	T	p-value
-0.3	Cox MSM	-0.301	-0.001	0.024	-0.15	0.88
	SNCFTM	-0.300	-0.000	0.060	0.00	1.00
	SNAFTM	-0.287	0.013	0.058	1.71	0.09
0.0	Cox MSM	0.000	0.000	0.020	0.14	0.88
	SNCFTM	-0.002	-0.002	0.046	-0.26	0.79
	SNAFTM	0.010	0.010	0.051	1.40	0.16
0.3	Cox MSM	0.302	0.002	0.018	0.50	0.62
	SNCFTM	0.294	-0.006	0.037	-0.99	0.32
	SNAFTM	0.302	0.002	0.047	0.27	0.77

Table 2 Monte Carlo simulation results for estimators of the parameter of a Cox MSM, SNCFTM and SNAFTM when data are generated as in Table 1 but with violation of the assumption that the T_0 are exponentially distributed. Here, the T_0 follow a Weibull distribution with shape and scale parameters 2 and 0.02, respectively.

ψ^*_{aft}	Model	MC Mean	Bias	MC Var	T	p-value
-0.3	Cox MSM	-0.364	-0.064	0.074	-7.47	< 0.0001
	SNCFTM	-0.467	-0.167	0.175	-12.64	< 0.0001
	SNAFTM	-0.300	-0.000	0.055	-0.02	0.98
0.0	Cox MSM	-0.001	-0.001	0.055	-0.16	0.88
	SNCFTM	-0.006	-0.006	0.083	-0.64	0.52
	SNAFTM	0.010	0.010	0.043	1.59	0.11
0.3	Cox MSM	0.358	0.058	0.044	8.66	< 0.0001
	SNCFTM	0.394	0.094	0.054	12.79	< 0.0001
	SNAFTM	0.301	0.001	0.037	0.242	0.81

Table 3 Monte Carlo simulation results for estimators of the parameter of a Cox MSM, SNCFTM and SNAFTM when data are generated as in Table 1 but with violation of the rare disease assumption ($\lambda_0 = 0.1$).

ψ^*_{aft}	Model	MC Mean	Bias	MC Var	T	p-value
-0.3	Cox MSM	-0.314	-0.014	0.004	-7.17	< 0.0001
	SNCFTM	-0.248	0.052	0.006	21.46	< 0.0001
	SNAFTM	-0.296	0.004	0.011	1.08	0.28
0.0	Cox MSM	0.001	0.001	0.003	0.58	0.56
	SNCFTM	-0.000	-0.000	0.005	-0.10	0.92
	SNAFTM	-0.000	-0.000	0.010	-0.03	0.98
0.3	Cox MSM	0.318	0.018	0.003	10.25	< 0.0001
	SNCFTM	0.245	-0.055	0.005	-25.31	< 0.0001
	SNAFTM	0.296	-0.004	0.011	-1.21	0.23

As expected based on Theorem 3, estimators of both ψ^*_{msm} and ψ^*_{cft} are unbiased when $\psi^*_{aft} = 0$ as shown in Tables 2 and 3.

6 Discussion

This paper defines sufficient conditions for a data generation mechanism to satisfy three structural failure time models: the SNAFTM, Cox MSM and SNCFTM. A simulation study where the data generation mechanism was (i) consistent with these conditions, and (ii) in violation of these conditions supports theoretical results regarding the sufficiency of these conditions. Our results also describe how to correctly simulate data from a SNCFTM with known parameter by generating data from a SNAFTM with known parameter. For simplicity, our discussion did not allow for right-censoring due to loss to follow-up or competing risks before $K + 1$, but estimating the model parameters in the presence of such censoring is straightforward under additional identifying assumptions as described in Hernán et al (2005), Hernán et al (2000) and Page et al (2008).

By generating data that satisfies all three models, we can evaluate the relative performance of the inverse probability weighted estimator of the Cox MSM and the g-estimators of the SNAFTM and SNCFTM under this limited data-generating mechanism. An interesting finding is that, as shown in Table 1, the widely used inverse probability weighted estimator of ψ^*_{msm} had similar or less bias, and a smaller variance, than the g-estimators of ψ^*_{aft} and ψ^*_{cft}, with the added advantage of being more easily computed.

As discussed in §3.2, the estimators studied in our simulations were simple to compute but non-optimal. Optimal estimators of parameters of structural nested models should be more efficient than those of marginal structural models under the assumption of no effect modification by past covariates (which is assumed in Theorems 1 and 2) (Robins and Hernán, 2009). Our simulation results suggest that, at least under this limited data-generating mechanism, non-optimal parameter estimates for the Cox MSM are actually more efficient than those of the SNCFTM or SNAFTM.

References

Cox DR, Oakes D (1984) Analysis of Survival Data. London: Chapman and Hall

Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11(5):561–570

Hernán MA, Hernández-Diáz S, Robins JM (2004) A structural approach to selection bias. Epidemiology 15:615–625

Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM (2005) Structural accelerated failure time models for survival analysis in studies with time-varying treatments. Pharmacoepidemiology and Drug Safety 14(7):477–491

van der Laan MJ, Robins JM (2002) Unified Methods for Censored Longitudinal Data and Causality. New York: Springer

Page J, Hernán MA, Robins JM (2008) Structural Nested Cumulative Failure Time Models. Tech. rep., Department of Epidemiology, Harvard School of Public Health

Picciotto S, Young J, Hernán MA (2008) G-estimation of structural nested cumulative failure time models. American Journal of Epidemiology 167(Suppl):S139

Picciotto S, Robins JM, Young J, Hernán MA (2009) Estimating absolute risks under hypothetical interventions using a structural nested cumulative failure time model. American Journal of Epidemiology 169(Suppl):S34

Robins JM (1986) A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. Mathematical Modelling 7:1393–1512

Robins JM (1998a) Marginal structural models. In: 1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science, American Statistical Association, pp 1–10

Robins JM (1998b) Structural nested failure time models. In: Armitage P, Colton T (eds) Encyclopedia of Biostatistics, Wiley, Chichester, pp 4372–4389

Robins JM, Hernán MA (2009) Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) Advances in Longitudinal Data Analysis, Boca Raton, FL: Chapman and Hall/CRC Press, pp 553–599

Robins JM, Blevins D, Ritter G, Wulfsohn M (1992) G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 3:319–336

Robins JM, Blevins D, Ritter G, Wulfsohn M (1993) Errata to g-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 4:189

Young JG, Hernán MA, Picciotto S, Robins JM (2008) Simulation from structural survival models under complex time-varying data structures. In: JSM Proceedings, Section on Statistics in Epidemiology, Denver, CO: American Statistical Association

7 Appendix

7.1 Proof of Theorem 1

Proof Without loss of generality, we can assume the SNAFTM is locally rank preserving in the sense that

$$T_{(\bar{a}_{m-1},0)} = \int_m^{T_{\bar{a}}} \exp\{\gamma_{AFT}(t, \overline{a}_t, \overline{V}_{t,\bar{a}}, \psi^*_{aft})\}dt,$$
$$T_0 = \int_0^{T_{\bar{a}}} \exp\left(a_t \times \psi^*_{aft}\right)dt, \tag{8}$$

since it is non-identifiable whether or not local rank preservation holds (Robins, 1998b). Thus

$$\begin{aligned}
&\Pr(T_{\bar{a}} > u) \\
&\quad = \Pr(T_0 > \int_0^u \exp\left(a_t \times \psi^*_{aft}\right)dt) \\
&\quad = \exp\left\{-\int_0^u \lambda_0 \exp\left(a_t \times \psi^*_{aft}\right)dt\right\}
\end{aligned}$$

Hence $\lambda_{T_{\bar{a}}}(u) = \{\partial/\partial u\}\left\{\int_0^u \lambda_0 \exp\left(a_t \times \psi^*_{aft}\right)dt\right\} = \lambda_0 \exp\left(a_u \times \psi^*_{aft}\right)$. It follows that $\psi^*_{aft} = \psi^*_{msm}$ for a Cox MSM with $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$.

7.2 Proof of Theorem 2

Proof By the definition of $Y_{j,g=(\overline{A}_m,0)}$ and $Y_{j,g=(\overline{A}_{m-1},0)}$for any $j \in [m, m+1, ..., k]$ and by consistency, we may rewrite the SNCFTM (3) as follows:

$$\begin{aligned}
\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})\} &= \frac{\Pr[T_{g=(\overline{A}_m,0)} < k | \overline{V}_m, \overline{A}_m, T_{g=(\overline{A}_m,0)} > m]}{\Pr[T_{g=(\overline{A}_{m-1},0)} < k | \overline{V}_m, \overline{A}_m, T_{g=(\overline{A}_m,0)} > m]} \\
&= \frac{\Pr[T_{g=(\overline{A}_m,0)} < k | \overline{V}_m, \overline{A}_m, T_{g=(\overline{A}_{m-1},0)} > m]}{\Pr[T_{g=(\overline{A}_{m-1},0)} < k | \overline{V}_m, \overline{A}_m, T_{g=(\overline{A}_{m-1},0)} > m]} \\
&= \frac{\Pr[k > T_{g=(\overline{A}_m,0)}, T_{g=(\overline{A}_{m-1},0)} > m, \overline{V}_m, \overline{A}_{m-1}]}{\Pr[k > T_{g=(\overline{A}_{m-1},0)} > m, \overline{V}_m, \overline{A}_{m-1}]} \\
&= \frac{\Pr[h(k, \overline{A}_m) > T_0 > h(m, \overline{A}_{m-1}), \overline{V}_m, \overline{A}_m]}{\Pr[h(k, \overline{A}_{m-1}) > T_0 > h(m, \overline{A}_{m-1}), \overline{V}_m, \overline{A}_m]}.
\end{aligned}$$

Now, with no loss of generality, explicitly writing out $\frac{\Pr[h(k,\overline{A}_m)>T_0>h(m,\overline{A}_{m-1}),\overline{V}_m,\overline{A}_m]}{\Pr[h(k,\overline{A}_{m-1})>T_0>h(m,\overline{A}_{m-1}),\overline{V}_m,\overline{A}_m]}$ under the locally rank preserving SNAFTM and noting that by assumption $f_{V_j|\overline{V}_{j-1},\overline{A}_{j-1},T_0}(V_j|\overline{V}_{j-1}, \overline{A}_{j-1}, u)$

is the same for all $u < \max\left\{h(K,\overline{A}_m), h(K,\overline{A}_{m-1})\right\}$, we obtain

$$\exp\{\gamma_{CFT,k}(\overline{V}_m,\overline{A}_m,\psi_{cft})\} =$$
$$\left\{\frac{\prod_{j=0}^{m} f_{A_j|\overline{V}_j,\overline{A}_{j-1}}(A_j|\overline{V}_j,\overline{A}_{j-1})}{\prod_{j=0}^{m} f_{A_j|\overline{V}_j,\overline{A}_{j-1}}(A_j|\overline{V}_j,\overline{A}_{j-1})}\right\}\times$$
$$\frac{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} f_{T_0}(u)\prod_{j=0}^{m} f_{V_j|\overline{V}_{j-1},\overline{A}_{j-1},T_0}(V_j|\overline{V}_{j-1},\overline{A}_{j-1},u)du}{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} f_{T_0}(u)\prod_{j=0}^{m} f_{V_j|\overline{V}_{j-1},\overline{A}_{j-1},T_0}(V|\overline{V}_{j-1},\overline{A}_{j-1},u)du}$$
$$= 1 + \frac{\int_{h(k,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} f_{T_0}(u)\prod_{j=0}^{m} f_{V_j|\overline{V}_{j-1},\overline{A}_{j-1},T_0}(V_j|\overline{V}_{j-1},\overline{A}_{j-1},u)du}{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} f_{T_0}(u)\prod_{j=0}^{m} f_{V_j|\overline{V}_{j-1},\overline{A}_{j-1},T_0}(V|\overline{V}_{j-1},\overline{A}_{j-1},u)du}$$
$$= 1 + \frac{\int_{h(k,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} f_{T_0}(u)du}{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} f_{T_0}(u)du}$$

$$= 1 + \frac{\int_{h(k,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} \lambda_{T_0}(u)S_{T_0}(u)du}{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} \lambda_{T_0}(u)S_{T_0}(u)du}$$
$$\approx 1 + \frac{\int_{h(k,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} \lambda_{T_0}(u)du}{\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} \lambda_{T_0}(u)du}$$

$$= 1 + \frac{\lambda_0\int_{h(k,\overline{A}_{m-1})}^{h(k,\overline{A}_m)} du}{\lambda_0\int_{h(m,\overline{A}_{m-1})}^{h(k,\overline{A}_{m-1})} du}$$
$$= 1 + \frac{h(k,\overline{A}_m) - h(k,\overline{A}_{m-1})}{h(k,\overline{A}_{m-1}) - h(m,\overline{A}_{m-1})}$$
$$= 1 + \frac{e^{\psi_{aft}A_m} - 1}{k-m}$$

7.3 Proof of Theorem 3

Proof By equation (8) and $\psi^*_{aft} = 0$ it follows that $T_0 = T_{\overline{a}}$ for any $\overline{a}$. Hence

$$\lambda_{T_{\overline{a}}}(t) = \lambda_{T_0}(t) = \lambda_{T_0}(t)\exp\{a_t \times \psi^*_{aft}\}$$

It follows that $\psi^*_{msm} = \psi^*_{aft}$ for a Cox MSM with $\gamma_{MSM}(t, \overline{a}_t, \psi_{msm}) = a_t \times \psi_{msm}$. Further, using equation (7), for $\psi^*_{aft} = 0$, $h(u, \bar{A}_m, \psi^*_{aft}) = u$ for any $\bar{A}_m$. Thus, we may rewrite the SNCFTM (3) as

$$\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi^*_{cft})\} = 1 = 1 + \frac{e^{A_m \psi^*_{aft}} - 1}{k - m}.$$

It follows that $\psi^*_{cft} = \psi^*_{aft}$ for a SNCFTM with $\exp\{\gamma_{CFT,k}(\overline{V}_m, \overline{A}_m, \psi_{cft})\} = 1 + \frac{e^{A_m \psi_{cft}} - 1}{k-m}$.