Published in final edited form as:

J Am Stat Assoc. 2012 ; 107(499): . doi:10.1080/01621459.2012.682532.

Structural Nested Cumulative Failure Time Models to Estimate the Effects of Interventions

Sally Picciotto [Assistant Researcher],

Research Fellow in the Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115

Division of Environmental Health Sciences, UC Berkeley School of Public Health, Berkeley, CA 94720

Miguel A. Hernán [Professor],

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115

John H. Page [Assistant Professor, Research Fellow],

Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115

Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115

Jessica G. Young [Research Associate], and

Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115

James M. Robins

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115

Abstract

In the presence of time-varying confounders affected by prior treatment, standard statistical methods for failure time analysis may be biased. Methods that correctly adjust for this type of covariate include the parametric g-formula, inverse probability weighted estimation of marginal structural Cox proportional hazards models, and g-estimation of structural nested accelerated failure time models. In this article, we propose a novel method to estimate the causal effect of a time-dependent treatment on failure in the presence of informative right-censoring and time-dependent confounders that may be affected by past treatment: g-estimation of structural nested cumulative failure time models (SNCFTMs). An SNCFTM considers the conditional effect of a final treatment at time m on the outcome at each later time k by modeling the ratio of two counterfactual cumulative risks at time k under treatment regimes that differ only at time m . Inverse probability weights are used to adjust for informative censoring. We also present a procedure that, under certain “no-interaction” conditions, uses the g-estimates of the model parameters to calculate unconditional cumulative risks under nondynamic (static) treatment regimes. The procedure is illustrated with an example using data from a longitudinal cohort study, in which the “treatments” are healthy behaviors and the outcome is coronary heart disease.

Keywords

Causal inference; Coronary heart disease; Epidemiology; G-estimation; Inverse probability weighting

1. INTRODUCTION

A common approach to estimating the causal effect of a time-varying treatment on failure time is to model the hazard of failure as a function of past treatment and covariate history (Kalbeisch and Prentice 1980). However, this approach may be biased even under the causal null whether or not one adjusts for the past history of measured covariates in the analysis when (a) there exists a time-dependent risk factor for failure that also predicts subsequent treatment, and (b) past treatment history predicts subsequent risk factor level (Robins 1986; Hernán, Hernández-Díaz, and Robins 2004). For example, when interested in the effect of a dietary intervention on the risk of coronary heart disease (CHD), diabetes is such a risk factor because (a) diabetes is a risk factor for CHD and receiving a diagnosis of diabetes may lead to changes in diet, and (b) the risk of diabetes is affected by an individual's prior diet history.

In the context of survival analysis, Robins and colleagues have developed several methods to deal with the problem of time dependent confounders that are themselves affected by previous treatment. These include the parametric g-formula (Robins 1986; Taubman et al. 2009), inverse probability weighted (IPW) estimation of marginal structural Cox proportional hazard models (Cox MSMs; Robins 1999; Hernán, Brumback, and Robins 2001), and g-estimation of structural nested accelerated failure time models (SNAFTMs; Robins 1998; Hernán et al. 2005).

In this article, we propose a new approach to estimate the causal effect of a time-dependent treatment on failure in the presence of informative right-censoring and time-dependent confounders that may be affected by past treatment. This approach is based on a new class of structural models called structural nested cumulative failure time models (SNCFTMs), previously discussed by Page (2005; unpublished Sc.D. dissertation) and briefly in a simulation study by Young et al. (2010). An SNCFTM can be viewed as a special case of the multivariate structural nested mean model (Robins 1994). The use of SNCFTMs requires that, for all values of the measured covariates, the conditional cumulative probability of failure under the possible treatment regimes satisfies a particular type of rare failure assumption, formally defined in Section 3.

In this "rare failure" context, the SNCFTM has advantages compared with Cox MSMs, SNAFTMs, and the parametric g-formula. Like the Cox MSM, but in contrast to the SNAFTM, the SNCFTM admits unbiased estimating equations that are differentiable in the model parameters and thus are easily solved. Like the SNAFTM, but in contrast to the Cox MSM, a correctly specified SNCFTM does not require a positivity assumption (i.e., it is not necessary for both treated and untreated to occur in every stratum of covariates); furthermore, it naturally allows one to investigate both quantitative and qualitative effect modification by evolving time-dependent covariates. Finally, the SNCFTM is dependent on fewer parametric assumptions than are required in applications of the parametric g-formula.

Besides describing the SNCFTM, this article presents an application to estimate the effects of lifestyle interventions on the 20-year risk of CHD among women in the Nurses' Health Study. For concreteness, we introduce the notation and method in the context of our

application. Although we restrict attention to time-varying binary treatments, the extension to continuous and/or multivariate treatments is straightforward (Robins 1994).

2. STUDY POPULATION AND NOTATION

The Nurses' Health Study is a prospective cohort study of female nurses followed since 1976, when participants were 30–55 years old. Data are collected in mailed questionnaires every 2 years to update information on potential risk factors and newly diagnosed diseases (Colditz, Manson, and Hankinson 1997). Let $k = 0, 1, \dots, K + 1$ where $k = 0$ denotes the time of return of the 1982 questionnaire and time k is $2k$ years later, with the last questionnaire returned at time K . Follow-up ends at $K + 1 = 10$ (20 years). For brevity, we use “by time k ” to mean “before the start of interval k ,” and “reported at time k ” to mean “reported on the questionnaire returned at the start of interval k .”

As in the literature by Stampfer et al. (2000) and Taubman et al. (2009), we define CHD as confirmed first myocardial infarction (fatal or nonfatal). Let Y_k indicate CHD (0: no, 1: yes) by time k for $k = 1, 2, \dots, K + 1$. Our study population consists of 78,746 women who were free of CHD ($Y_0 = 0$) and had complete baseline covariate data in 1982. During the follow-up, 2319 women experienced a CHD event, 5616 died from other causes, and 16,818 were censored when they first failed to return a questionnaire (Taubman et al. 2009). We explain how to handle censoring in Section 6.

We estimated the effects of separate interventions on cigarette smoking, exercise, diet score (as defined by Stampfer et al. 2000), and alcohol intake. The assessments of these variables have been validated (Colditz et al. 1986; Salvini et al. 1989; Wolf et al. 1994). We quantify these effects from a public health, rather than biological, perspective. Following Taubman and colleagues (2009), we estimated the effects of four hypothetical interventions sustained from 1982 to 2002:

- (1) avoid smoking;
- (2) exercise at least 30 min/day;
- (3) keep diet score in a range corresponding to the top two quintiles of the observed data; and
- (4) consume at least 5 g/day of alcohol.

Although there is ongoing debate about the potential nonlinearity of the effect of high levels of alcohol consumption on CHD risk, the choice of 5 g/day as a threshold for alcohol was made to be consistent with the definition of the low-risk group described by Stampfer et al. (2000). Very few women in the Nurses' Health Study report heavy drinking; less than 6% of observations correspond to at least two drinks (27.4 g alcohol) per day.

Our specific goal is to estimate the effect of representative interventions or regimes as defined by Taubman et al. (2008). In a representative regime, the actual daily exercise duration, diet score, or amount of alcohol assigned at time m is randomly drawn from the observed distribution of the variable intervened on (e.g., daily exercise duration), conditional on past covariate history, among those subjects who, in the observed data, have followed the intervention (e.g., exercised at least 30 min/day) through time m (see Section 4 for a formal definition). Note that our estimates of the effect of a hypothetical intervention under a representative regime cannot be extrapolated to cohorts whose observed data distribution differs from that of the Nurses' Health Study. We nonetheless chose to study representative regimes to achieve our goal of comparing our estimates with those obtained in previous analyses that used alternative statistical methods.

For each of the above interventions, we denote by A_m the observed value at time m of the risk factor to be intervened on (smoking, exercise, diet score, or alcohol intake). Let $\mathbf{L}_m$ be the observed vector of covariates reported at time m , where $\mathbf{L}_0$ includes time-invariant (fixed or baseline) covariates. We always include the survival indicator $S_j = 1 - Y_j$ in $\mathbf{L}_j$. We view A_{m-1} as a component of $\mathbf{V}_m = (A_{m-1}, \mathbf{L}_m)$; an arbitrary single component of $\mathbf{V}_m$ (either A_{m-1} or one of the components of $\mathbf{L}_m$) will be denoted X_m . Table 1 lists the variables included in $\mathbf{V}_0$ and $\mathbf{V}_m$ in our application. For each intervention, we then defined an observed binary time-varying variable R_m to indicate whether a woman's reported A_m (for $m \geq 0$) was consistent with the intervention under consideration. For example, when estimating the effects of the exercise intervention, $R_m = 1$ for those who exercise at least 30 min/day at time $m \geq 0$, and $R_m = 0$ otherwise. That is, $R_m = I(A_m \geq 30)$ for $m \geq 0$ and $R_m = 0$ for $m < 0$. We refer to R_m as the "treatment" indicator, and thus "being treated" in the observed data ($R_m = 1$) means "behaving consistently with the intervention." Note that for the "avoid smoking" intervention, $R_m = 1$ denotes no smoking at time m . At each time m , we assume the temporal ordering $Y_m, \mathbf{L}_m, R_m, A_m$ (Robins et al. 2004). Our dataset includes $K + 1$ observations for every uncensored individual. Further, since $Y_k = 1$ denotes failure before k , $Y_k = 1$ implies that $Y_j = 1$, for all $j, k - j \leq K + 1$. When $Y_j = 1$, we define, by convention, $\mathbf{L}_j = R_j = 0$.

Overbars represent history from time 0, so $\mathbf{R}_m^- = (R_0, \dots, R_m)$ and $\mathbf{V}_m^- = (\mathbf{V}_0, \dots, \mathbf{V}_m)$. By convention, $\mathbf{R}_{-1}^- = \mathbf{V}_{-1}^- = 0$. To denote the full histories of treatment and covariates from time 0 to the end of the study, we use $\mathbf{R} = (R_0, \dots, R_K)$ and $\mathbf{V} = (\mathbf{V}_0, \dots, \mathbf{V}_K)$, respectively. The observed covariate and treatment history $(\mathbf{V}_m, \mathbf{R}_m)$ can be written $(\mathbf{V}_m, R_m)$ since each prior treatment R_j for $j < m$ is determined by A_j , which is a component of $\mathbf{V}_{j+1}$. We assume that the data observed on our n study participants are n independent, identically distributed realizations of a vector $(\bar{\mathbf{Y}}_{K+1}, \bar{\mathbf{V}}_K, R_K, A_K)$. In general, we use m to refer to a time when treatment was measured and k to refer to a time when the survival outcome was measured, with $m < k$. Where capital letters represent random variables and the corresponding lower-case letters represent their realizations.

Let a treatment regime $\mathbf{g} \equiv \bar{\mathbf{g}}_K \equiv (g_0, \dots, g_K)$ be a vector of functions where g_m assigns to each covariate history $\mathbf{v}_m^-$, a treatment r_m to subjects who survive until time m (i.e., those with $y_m = 0$), and, by convention, 0 to those who have failed (i.e., those with $y_m = 1$). We define a nondynamic regime $\mathbf{g} \equiv \bar{\mathbf{g}}_K = \mathbf{r}_K^-$ to be one where for those with $y_m = 0$, $\mathbf{g}_m(\mathbf{v}_m^-)$ equals the same r_m for all $\mathbf{v}_m^-$. The four hypothetical interventions described above are nondynamic regimes that assign the same value of r_m to all subjects with $y_m = 0$ at every time point m from 1982 to 2002 (until failure). The regime "Always treat" (when $y_m = 0$) is represented by $\mathbf{g} = (1, 1, \dots, 1) = \mathbf{1}^-$ while the regime "Never treat" is $\mathbf{g} = \mathbf{0}^-$. Another regime that will be used below is the regime "Receive the treatment actually received through time m , but no treatment thereafter" denoted by $\mathbf{g} = (\mathbf{R}_m, \mathbf{0})$, where $\mathbf{0}$ (when following $\mathbf{R}_m$) represents no treatment from time $m + 1$ up to the end of the study. For any treatment regime $\mathbf{g}$, $Y_{k,\mathbf{g}}$ denotes the counterfactual (or potential) outcome by time k under regime $\mathbf{g}$. For example, $Y_{k,\mathbf{g}} = 0$ indicates that a woman would have been free of CHD at time k if she had followed treatment regime $\mathbf{g}$. More generally, $\mathbf{V}_{m,\mathbf{g}} = (A_{m-1,\mathbf{g}}, \mathbf{L}_{m,\mathbf{g}})$ is the counterfactual $\mathbf{V}_m$.

We will see below that for any nondynamic regime $\mathbf{g}$, under appropriate consistency and exchangeability assumptions, the distribution of $\bar{\mathbf{Y}}_{K+1,\mathbf{g}}$ under the above definition is identified and equals the distribution of $\bar{\mathbf{Y}}_{K+1,\mathbf{g}}$ under the corresponding representative regime.

3. STRUCTURAL NESTED CUMULATIVE FAILURE TIME MODELS

Let $E[Y_{k,g}|\bar{\mathbf{V}}_m, R_m, Y_m = 0]$ be the counterfactual risk of developing the outcome by time k , among those free of the outcome at time $m < k$, given their observed covariate and treatment history up to time m , had they followed treatment regime $\mathbf{g}$. An SNCFTM models the ratio of two such risks under different regimes $\mathbf{g}$ conditional on covariate and treatment history $(\bar{\mathbf{V}}_m, R_m)$. The numerator of the ratio is the conditional risk had everyone received their actual treatment history $\bar{\mathbf{R}}_m$ through time m and no treatment from time $m + 1$ to the end of the study, that is, $\mathbf{g} = (\bar{\mathbf{R}}_m, \mathbf{0})$. The denominator of the ratio is the conditional risk had everyone received their actual treatment history $\bar{\mathbf{R}}_{m-1}$ through time $m - 1$ and no treatment from time m to the end of the study (i.e., $\mathbf{g} = (\bar{\mathbf{R}}_{m-1}, \mathbf{0})$). Adopt the convention that $R_m = 0$ for those who have $Y_m = 1$. Then, the general form of the model is

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi^* \right) \right] = \frac{E \left[Y_{k, \mathbf{g} = (\bar{\mathbf{R}}_m, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0 \right]}{E \left[Y_{k, \mathbf{g} = (\bar{\mathbf{R}}_{m-1}, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0 \right]} \quad (1)$$

$$= 1 \text{ if } Y_m = 1 \quad (2)$$

where $\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi)$ is a function of treatment and covariate history indexed by the (possibly vector-valued) parameter ψ whose unknown true value is ψ^* . Note that $R_m = 0$ implies $\exp[\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi^*)] = 1$.

An SNCFTM for the effect of the intervention “Require everyone to exercise at least 30 min/day at each time m ” on the risk at time k compares the conditional CHD risk at each time k under two regimes: a regime that does not modify anyone’s observed exercise indicator (either at least 30 min/day or less than 30 min/day) through time m , and then makes everyone exercise less than 30 min/day thereafter (i.e., sets the indicator to 0), and the nearly identical regime that sets the indicator to 0 starting one interval sooner.

We refer to $\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi)$ as the “blip function” because the two treatment regimes being compared in Equation (1) can differ only at time m , and thus an SNCFTM models the effect of a final “blip” of treatment at time m on the outcome at time k . For consistency, the exponentiated blip function $\exp[\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi)]$ must be 1 when $R_m = 0$, because then the two treatment regimes being compared are identical. The model should also be chosen so that $\exp[\gamma_k(\bar{\mathbf{V}}_m, R_m; 0)] = 1$ when there is no effect of treatment at time m on outcome at time k . An example of a blip function is

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = \exp [\psi R_m], \quad (3)$$

so $\psi = 0$ corresponds to no effect, $\psi < 0$ to beneficial effect, and $\psi > 0$ to harmful effect.

The blip function we used in our analysis, which may be more realistic in some contexts, is

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = 1 + \frac{\exp [\psi R_m] - 1}{k - m}. \quad (4)$$

The effect of R_m under model (4) is the same as that under model (3) at time $k = m + 1$, and then diminishes as the time elapsed since m , $k - m$, increases. Under certain conditions, this model is consistent with data generated under a particular SNAFTM (Young et al. 2008).

Other possible blip functions might include a product term between treatment R_m and a component of treatment history $\mathbf{R}_{m-1}$, for example,

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = 1 + \frac{\exp [(\psi_1 + \psi_2 R_{m-1}) \times R_m] - 1}{k - m}, \quad (5)$$

a product term between treatment R_m and a baseline covariate X_0 in $\mathbf{V}_0$,

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = 1 + \frac{\exp [(\psi_1 + \psi_2 X_0) \times R_m] - 1}{k - m}, \quad (6)$$

or a product term between treatment R_m and a component X_m of the time-varying covariate vector $\mathbf{V}_m$,

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = 1 + \frac{\exp [(\psi_1 + \psi_2 X_m) \times R_m] - 1}{k - m}. \quad (7)$$

Page (2005) considers additional functional forms for the blip function.

Remark 1 (Restriction to rare failures). Note that the conditional expectations of the potential outcomes (failures) necessarily lie within the interval $[0, 1]$. Thus, $\exp[\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m; \psi^*)]$ is bounded by

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi^* \right) \right] \leq \frac{1}{E \left[Y_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0 \right]}.$$

We do not impose this constraint in fitting the SNCFTM model because estimation of ψ^* would then require a correct model for $E[Y_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0]$. Instead, we will estimate ψ^* under a semiparametric model that leaves the conditional expectation in the denominator completely unspecified, and thus protects our estimates against misspecification of the model for this conditional expectation. As a consequence, our semiparametric SNCFTM should only be used for studies in which the cumulative probability of failure is small. As seen later, our methodology allows us to obtain estimates of $E[Y_{K+1,g=(\mathbf{0})}]$ and, in certain cases, $E[Y_{K+1,g=(\mathbf{r}_K^*)}]$ for all $\mathbf{r}_K^*$ without requiring a model for the $E[Y_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0]$. If these estimates are all much less than 1, we have some, but not conclusive, evidence, for the rare failure assumption.

If we substitute the counterfactual survival indicators $S_{k,g=(\bar{\mathbf{R}}_m, \mathbf{0})} = 1 - R_{k,g=(\bar{\mathbf{R}}_m, \mathbf{0})}$ and $S_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} = 1 - R_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}$ for $R_{k,g=(\bar{\mathbf{R}}_m, \mathbf{0})}$ and $S_{k,g=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}$ in Eq (1), we obtain the structural nested multiplicative survival time model (SNMSTM) considered by Martinussen et al. (2011, p. 776). We chose to fit an SNCFTM rather than a SNMSTM because, with rare failures, unconstrained estimation of an SNMSTM may result in estimated survival probabilities exceeding 1.

The next two sections describe how to obtain a consistent estimate of ψ^* via g-estimation, and how the parameter ψ^* can, under a certain restrictive “no-interaction” condition, be used to calculate the absolute risk of failure for nondynamic treatment regimes.

4. IDENTIFICATION, REPRESENTATIVE REGIMES, AND G-ESTIMATION

This section describes assumptions under which the mean of $\bar{Y}_{K+1,g}$ is identified for all regimes g . We then show how the method of g-estimation can be used to estimate the parameter ψ^* of the SNCFTM when these assumptions hold.

4.1 Assumptions and Definitions

Let $\bar{g}_m(\bar{\mathbf{v}}_m) \equiv (g_0(\bar{\mathbf{v}}_0), \dots, g_m(\bar{\mathbf{v}}_m)) = \bar{\mathbf{r}}_m$ for any regime g that specifies the treatment $\bar{\mathbf{r}}_m$ through time m and consider the following identifying assumptions:

Assumption 1 (Consistency). For $m = 0, \dots, K$, and any regime g

$$\bar{g}_m(\bar{\mathbf{V}}_m) = \bar{\mathbf{R}}_m \Rightarrow Y_{k,g} = Y_k, \text{ for } k \leq m+1.$$

That is, if a subject has an observed treatment history through m equal to that prescribed by regime g , then her observed outcome history through $m+1$ will equal her counterfactual outcome history under regime g through that time.

Assumption 2 (Exchangeability). For any treatment regime g , and for all $m \leq K$,

$$\underline{Y}_{m+1,g} \perp\!\!\!\perp \bar{\mathbf{R}}_m | \bar{\mathbf{V}}_m, Y_m = 0 \quad (8)$$

$$\underline{Y}_{m+1,g} \perp\!\!\!\perp \bar{\mathbf{A}}_m | \bar{\mathbf{V}}_m, Y_m = 0, \bar{\mathbf{R}}_m \quad (9)$$

where $\underline{Y}_{m+1,g} = (Y_{m+1,g}, Y_{m+2,g}, \dots, Y_{K+1,g})$ and $\perp\!\!\!\perp$ denotes statistical independence. Note these two conditions are equivalent to the single condition that

$$\underline{Y}_{m+1,g} \perp\!\!\!\perp \bar{\mathbf{A}}_m | \bar{\mathbf{V}}_m, Y_m = 0$$

since, in the observed data, $\bar{\mathbf{R}}_m$ is a deterministic function of the $\bar{\mathbf{A}}_m$.

The exchangeability assumptions in Equations (8) and (9) are expected to hold in a sequentially randomized trial in which at each time m treatment A_m is based on physical randomization with the randomization probabilities allowed to depend on the measured past $\bar{\mathbf{V}}_m$. In an observational study, they cannot be guaranteed to hold and cannot be empirically tested from the data.

Assumption 3 (Positivity). If $f_{\bar{\mathbf{V}}_m}(\bar{\mathbf{v}}_m | Y_m = 0) \Pr[Y_m = 0] > 0$, then

$$\Pr \left[A_m = a_m | \bar{\mathbf{V}}_m = \bar{\mathbf{v}}_m, Y_m = 0 \right] > 0 \text{ for all } a_m,$$

where $f_{\bar{\mathbf{V}}_m}(\bar{\mathbf{v}}_m | Y_m = 0)$ is the conditional density of $\bar{\mathbf{V}}_m$ evaluated at $\bar{\mathbf{v}}_m$.

Assumptions 1 and 3 and Equation (8) nonparametrically identify the distribution of $\bar{Y}_{K+1,\mathbf{g}}$ for all regimes $\mathbf{g}$ (Robins 1994). However, once we impose an SNCFTM, the model is no longer nonparametric and the positivity assumption is not required for identification. Therefore, in the absence of positivity, the estimates rely on some degree of extrapolation from the model.

All of the formal results in this article refer to the means of the $\bar{Y}_{K+1,\mathbf{g}}$. However, when Assumption 2 (i.e., both Equations 8 and 9) holds, the distribution of $\bar{Y}_{K+1,\mathbf{g}}$ for any nondynamic regime $\mathbf{g} = \mathbf{r}_K$ has an additional interpretation as the distribution of $\bar{Y}_{K+1,\mathbf{g}_{\text{rep}}(\mathbf{g})}$ under the representative regime $\mathbf{g}_{\text{rep}} \equiv \mathbf{g}_{\text{rep}}(\mathbf{g})$ defined as follows (Taubman et al. 2008):

Definition 2. The representative regime $\mathbf{g}_{\text{rep}}(\mathbf{g})$ corresponding to a nondynamic regime $\mathbf{g} = \mathbf{r}_K$ is the random regime in which, for each subject with a counterfactual history $(\mathbf{I}_m, \dots, \mathbf{I}_{m-1})$, (a) R_m is set to r_m and (b) A_m is a random draw a_m from the observed conditional distribution of A_m given $\mathbf{L}_m = \mathbf{I}_m, \dots, \mathbf{I}_{m-1} = \mathbf{I}_{m-1}, \mathbf{R}_m = \mathbf{r}_m$.

A Clarifying Remark. Note that the notation for the counterfactual outcome under regime $\mathbf{g}$, $Y_{k,\mathbf{g}}$, can be replaced by a mathematically equivalent, but more cumbersome, notation that makes it transparent that the counterfactual outcome is also a function of the particular value of treatment A_m that the subject would receive if assigned to treatment R_m . (Hernán and VanderWeele 2011) By definition, under regime $\mathbf{g} = \mathbf{r}_K$ for each subject we (a) set (manipulate) the indicator of following the intervention at m to r_m and then (b) a precise value of A_m consistent with r_m is then chosen by each subject without further intervention. In contrast, under the representative regime, clause (b) is replaced by the following clause (b') we also intervene on A_m by drawing its value at random from the aforementioned distribution. In both cases, it is the value of the A_m that is the causal determinant of later failure. If Equation (9) holds, the choice of A_m in regime $\mathbf{g} = \mathbf{r}_K$ is unconfounded; as a consequence (b) and (b') are equivalent in the sense that, for $\mathbf{g} = \mathbf{r}_K$, $\bar{Y}_{K+1,\mathbf{g}_{\text{rep}}(\mathbf{g})}$ and $\bar{Y}_{K+1,\mathbf{g}}$ have the same distribution. Since it would be quite unusual for Equation (9) not to hold when Equation (8) holds, the possibility that, for $\mathbf{g} = \mathbf{r}_K$, the means of $\bar{Y}_{K+1,\mathbf{g}_{\text{rep}}(\mathbf{g})}$ and $\bar{Y}_{K+1,\mathbf{g}}$ differ is mainly of theoretical interest. Therefore, readers can safely ignore the possibility of difference without compromising their understanding of the remainder of the article.

4.2 G-Estimation

To describe g-estimation, define

$$H_{m,k}(\psi) = Y_k \exp \left\{ - \sum_{j=m}^{k-1} \gamma_{j,k} \left(\bar{\mathbf{V}}_j, R_j; \psi \right) \right\} \text{ if } Y_m = 0 \\ = 1 \text{ if } Y_m = 1.$$

Under our identifying assumptions, this quantity has the same mean as the counterfactual outcome $Y_{k,\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}$ conditional on $\bar{\mathbf{V}}_m$ when $\psi = \psi^*$ (Robins 1994; Page 2005). Therefore,

$$E \left[Y_{k,\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} \mid \bar{\mathbf{V}}_m, R_m \right] = E \left[H_{m,k}(\psi) \mid \bar{\mathbf{V}}_m, R_m \right] \\ = E \left[H_{m,k}(\psi) \mid \bar{\mathbf{V}}_m \right]$$

by Assumptions 1 and 2. We now use the above equality to describe the theory underlying g-estimation of the parameter ψ of the SNCFTM.

For our parameter vector ψ of dimension d , we consider estimating functions of the form

$$U(\psi; \mathbf{B}, \mathbf{J}) \equiv \sum_{m=0}^K (1 - Y_m) \sum_{k=m+1}^{K+1} \mathbf{D}(\bar{\mathbf{V}}_m, R_m, k)' \times \left[H_{m,k}(\psi) - B(\bar{\mathbf{V}}_m, k) \right],$$

where $B(\bar{\mathbf{v}}_m, k)$ is a user-supplied real-valued function, $\mathbf{J}(\bar{\mathbf{v}}_m, r_m, k)$ is a user-supplied $(1 \times d)$ -dimensional function with $k < m$, $\mathbf{D}(\bar{\mathbf{V}}_m, R_m, k) = \mathbf{J}(\bar{\mathbf{V}}_m, R_m, k) - E[\mathbf{J}(\bar{\mathbf{V}}_m, R_m, k) | \bar{\mathbf{V}}_m, Y_m = 0]$, and $'$ denotes the transpose. If $f(R_m | \bar{\mathbf{V}}_m, Y_m = 0)$ is known for each m , then under Assumption 2 and the SNCFTM $\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m; \psi)$, it follows that

$$E[U(\psi^*; B, \mathbf{J})] = 0,$$

where ψ^* is the true value of ψ . More generally, consider the class of estimating functions

$$\Lambda_{\text{nuis}}^\perp(\psi) = \left\{ U(\psi; \mathbf{B}, \mathbf{J}) \mid \mathbf{D}(\bar{\mathbf{V}}_m, R_m, k) \text{ and } B(\bar{\mathbf{V}}_m, k) \text{ have finite variance} \right\}.$$

It follows from theorem A3.1 in the article by Robins (1994) that $\Lambda_{\text{nuis}}^\perp(\psi^*)$ is the orthogonal complement of the nuisance tangent space for ψ^* under semiparametric model χ defined by the known $f(R_m | \bar{\mathbf{V}}_m, Y_m = 0)$ and the SNCFTM $\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m; \psi)$. It is well known that the influence functions of regular asymptotically linear (RAL) estimators of ψ^* are elements of the orthogonal complement to the nuisance tangent space (Bickel et al. 2003;

van der Laan and Robins 2003). Thus, if $P_n(G)$ denotes $\frac{1}{n} \sum_{i=1}^n G_i$ for any G , then any RAL estimator $\tilde{\psi}$ of ψ^* under semiparametric model χ is asymptotically equivalent to a solution $\hat{\psi}(B, \mathbf{J})$ of an estimating equation of the form

$$P_n[U(\psi; \mathbf{B}, \mathbf{J})] = 0 \quad (10)$$

for some choice of functions B and $\mathbf{J}$. That is, $n^{1/2} \{ \tilde{\psi} - \hat{\psi}(B, \mathbf{J}) \}$ converges to zero in probability.

When $f(R_m | \bar{\mathbf{V}}_m, Y_m = 0)$ is unknown, as in observational studies, we will replace it by an estimate to compute $E[\mathbf{J}(\bar{\mathbf{V}}_m, R_m, k) | \bar{\mathbf{V}}_m, Y_m = 0]$. If $\bar{\mathbf{V}}_m$ has two or more continuous components or many discrete components, we cannot hope to estimate $E[\mathbf{J}(\bar{\mathbf{V}}_m, R_m, k) | \bar{\mathbf{V}}_m, Y_m = 0]$ using nonparametric smoothing due to the curse of dimensionality. Therefore, we estimate it under a parametric model. If, as in our example, R_m is dichotomous, then we can restrict consideration to functions $\mathbf{J}(\bar{\mathbf{V}}_m, R_m, k) = \mathbf{J}^*(\bar{\mathbf{V}}_m, k)' R_m$, yielding $\mathbf{D}(\bar{\mathbf{V}}_m, R_m, k) = \mathbf{J}^*(\bar{\mathbf{V}}_m, k) \{ R_m - E[R_m | \bar{\mathbf{V}}_m, Y_m = 0] \}$. Here $\mathbf{J}^*(\bar{\mathbf{V}}_m, k)$ is a $(d \times 1)$ -dimensional function. Hence, it suffices to postulate a parametric model for treatment,

$$E \left[R_m | \bar{\mathbf{V}}_m, Y_m = 0 \right] = E \left[R_m | \bar{\mathbf{V}}_m, Y_m = 0; \alpha^* \right], \quad (11)$$

where α^* is an unknown vector of parameters to be estimated and $E[R_m|\bar{\mathbf{V}}_m, Y_m=0; \alpha]$ is a known function. The estimator $\hat{\psi}(B, \mathbf{J}^*, \hat{\alpha})$ solving $P_n[U(\psi; \mathbf{B}, \mathbf{J}^*, \hat{\alpha})] = 0$ where

$$U(\psi; \mathbf{B}, \mathbf{J}^*, \hat{\alpha}) \equiv \sum_{m=0}^K (1 - Y_m) \times \left\{ R_m - E \left[R_m | \bar{\mathbf{V}}_m, Y_m=0; \hat{\alpha} \right] \right\} \times \sum_{k=m+1}^{K+1} \mathbf{J}^* \left(\bar{\mathbf{V}}_m, k \right) \times \left[H_{m,k}(\psi) - B \left(\bar{\mathbf{V}}_m, k \right) \right]$$

will be consistent and asymptotically normal (CAN) if the model for $E[R_m|\bar{\mathbf{V}}_m, Y_m=0]$ is correct.

For simplicity of computation in our application, we chose $B(\bar{\mathbf{V}}_m, k) = 0$, and $\mathbf{J}^*(\bar{\mathbf{V}}_m, k) = 1$ for one-parameter models (3) and (4), $(1, R_{m-1})'$ for model (5), and $(1, X_0)'$ for model (6).

We estimated $E[R_m|\bar{\mathbf{V}}_m, Y_m=0]$ by $E \left[R_m | \bar{\mathbf{V}}_m, Y_m=0; \hat{\alpha} \right]$ where $\hat{\alpha}$ is the maximum likelihood estimate based on the pooled logistic model $\text{logit} \Pr[R_m = 1 | \bar{\mathbf{V}}_m, Y_m=0; \alpha] = \alpha' \mathbf{v}(\bar{\mathbf{V}}_m)$ where $\mathbf{v}(\bar{\mathbf{V}}_m)$ is a function of the covariates listed in Table 1. Thus, we estimated

$\left\{ \frac{e^{-\hat{\alpha}' \mathbf{v}(\bar{\mathbf{V}}_m)}}{1 + e^{-\hat{\alpha}' \mathbf{v}(\bar{\mathbf{V}}_m)}} \right\}^{-1}$. For each covariate X_m in $\bar{\mathbf{V}}_m$ that was not measured at every time m , the function $\mathbf{v}(\bar{\mathbf{V}}_m)$ included a product term between the most recent measurement of that component X_t for $t < m$ and the time since that measurement $m - t$, as described by Taubman et al. (2009). For each covariate in $\bar{\mathbf{V}}_m$ that was measured at every time, the function $\mathbf{v}(\bar{\mathbf{V}}_m)$ included its values at times m and $m - 1$, unless these measurements were so collinear as to result in unstable coefficients or convergence problems, in which case only the value at time m was included. The exclusion of the measurements at time $m - 1$ did not materially affect our estimates.

The estimator $\hat{\psi}(\hat{\alpha})$ that solves $P_n[U(\psi; \hat{\alpha})] = 0$ is our g-estimate, where $U(\psi; \alpha^*) = U(\psi; B, \mathbf{J})$ and $B, \mathbf{J}$ were chosen as above. To solve this estimating equation, we minimized a quadratic form obtained by multiplying the sum over all individuals of $U(\psi; \hat{\alpha})$ by the generalized inverse of the empirical covariance matrix by the transpose of the sum of $U(\psi; \hat{\alpha})$. We conducted a grid search using 100 points within a user-defined interval for each component of the parameter (by default, -1.5 to 1.5) to determine the initial value for the optimization procedure. Then we used the Newton–Raphson method to find the minimum. When a minimum was not obtained within the region, the search band was widened until a minimum was found. In the case of model (3), Newton–Raphson did not converge and the Nelder–Mead simplex method was used instead. The estimation procedure was repeated in 200 nonparametric bootstrap samples to obtain 95% confidence intervals (CIs) for the g-estimate.

We note that our estimator $\hat{\psi}(\hat{\alpha})$ did not require the positivity assumption. It follows that, under a nonsaturated SNCFTM, ψ^* and thus

$$\frac{E \left[Y_{k, \mathbf{g} = \left(\bar{\mathbf{R}}_m, 0 \right)} | \bar{\mathbf{V}}_m, R_m, Y_m=0 \right]}{E \left[Y_{k, \mathbf{g} = \left(\bar{\mathbf{R}}_{m-1}, 0 \right)} | \bar{\mathbf{V}}_m, R_m, Y_m=0 \right]}$$

can be consistently estimated even though positivity does not hold.

When we replace Y_k by $S_k = 1 - Y_k$ in the definition of $H_{m,k}(\psi)$ and redefine $H_{m,k}(\psi)$ as 0 when $Y_m = 1$, the above G-estimation procedure consistently estimates the parameter vector of an SNMSTM.

5. COMPUTING THE MARGINAL RISK UNDER NONDYNAMIC INTERVENTIONS

Equation (1) and our estimate $\hat{\psi}$ of ψ^* provide an estimate $\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \hat{\psi} \right) \right]$ of the contrast between the conditional means of $Y_{k,\mathbf{g}=(\bar{\mathbf{R}}_m, \mathbf{0})}$ and $Y_{k,\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}$ given $\bar{\mathbf{V}}_m, R_m, Y_m = 0$. To consider the interventions' effects at the population level, we are more interested in comparisons between the unconditional means of $Y_{k,\mathbf{g}}, Y_{k,\mathbf{g}=\mathbf{0}}$, and Y_k for various choices of $\mathbf{g}$. Below we show that $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ equals the expectation of an observable random variable when $\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi^*)$ is known. This raises the question whether, for certain other choices of $\mathbf{g}$, $E[Y_{k,\mathbf{g}}]$ is a function of the joint distribution of the data only through the $\alpha_k(\bar{\mathbf{V}}_m, R_m; \psi^*)$ and $E[Y_{k,\mathbf{g}=\mathbf{0}}]$. In general, the answer is no except in the special case in which (a) there is no treatment time-varying covariate interaction on the multiplicative cumulative failure scale, that is,

$$\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi^* \right) = \gamma_k \left(\mathbf{V}_0, \bar{R}_m; \psi^* \right)$$

does not depend on $\bar{\mathbf{V}}_m$ for all $m > 0$, and (b) the regime $\mathbf{g}$ of interest is a “baseline-adjusted” nondynamic regime $\mathbf{g} = \mathbf{r}_K^-(\mathbf{v}_0)$; that is, at each time m , the treatment r_m assigned by $\mathbf{g}$ to subjects with $Y_m = 0$ does not depend on covariate history $\bar{\mathbf{v}}_m$ except through the baseline covariates $\mathbf{v}_0$. Note that (a) is satisfied by blip functions (3)–(6) but not by blip function (7), and that (b) is satisfied by all of the interventions we considered in our application. Condition (a) can be tested by specifying a model $\gamma_k(\bar{\mathbf{V}}_m, R_m; \psi)$ depending on $\bar{\mathbf{V}}_m$ with a subvector of ψ , say, $\psi_{m,int}$, encoding the dependence on $\mathbf{V}_1, \dots, \mathbf{V}_m$ and then using g-estimation to test the hypothesis that $\psi_{m,int}^* = 0$. Below, for ease of notation, we restrict attention to nondynamic regimes $\mathbf{g} = \mathbf{r}_K^-$. Extension to “baseline-adjusted” nondynamic regimes $\mathbf{g} = \mathbf{r}_K^-(\mathbf{v}_0)$ is straightforward.

Outside of the special case where (a) and (b) hold, estimation of $E[Y_{k,\mathbf{g}}]$ requires that one estimate the nuisance functions $E[Y_{k,\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})} | \bar{\mathbf{V}}_m, R_m, Y_m = 0]$ and the density of $\bar{\mathbf{V}}_m$ given the past, which cannot be accomplished in a robust manner, would add substantially to the computational burden, and would make estimation more dependent on correct specification of additional models (Robins et al. 1992). Nonetheless, subject to these limitations, estimation of $E[Y_{k,\mathbf{g}}]$ for any regime $\mathbf{g}$ is possible by using the approach provided in the appendix of the article by Robins (1994) and in section 8 of the article by Robins, Rotnitzky, and Scharfstein (1999).

Below we first describe the estimation of $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ for each time k . To do so, the SNCFTM model is used to mathematically “remove” the effect of each subject’s nonzero treatments, starting at the end of the study period ($K + 1$) and working backward to the beginning ($k = 0$). We refer to this process as “blipping down.”

We then describe the estimation of $E[Y_{k,\mathbf{g}=\mathbf{r}_K^-}]$ under the additional assumption that (a) above holds. For concreteness, we suppose the regime $\mathbf{g}$ is the regime “Always treat,” that is, $\mathbf{g} = \mathbf{1}^-$. Then, given the estimates of the $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ obtained by “blipping down,” we use the

SNCFTM to construct estimates of $E[Y_{k,\mathbf{g}=1}]$, by starting at the first time point ($k = 0$) and working forward to the end of the study period ($K + 1$). We refer to this process as “blipping up” (Robins 1994; Page 2005).

5.1 Blipping Down

The quantity

$$H_{j,k}(\psi^*) = Y_j + (1 - Y_j) Y_k \prod_{m=j}^{k-1} \exp \left[-\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi^* \right) \right]$$

has the same conditional mean given $\bar{\mathbf{V}}_j, R_j = r_j$ as the counterfactual $Y_{k,\mathbf{g}=(\mathbf{r}_{j-1},0)}$ under the regime in which treatment is withheld starting at time j (Page 2005). For $j = 0$, $H_{0,k}(\psi^*)$ has the same mean as the cumulative risk at time k under the intervention “Never treat.” Note that

$$H_{0,k}(\psi^*) = Y_k \prod_{m=0}^{k-1} \exp \left[-\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi^* \right) \right], \quad (12)$$

because $Y_0 = 0$ for all participants. Thus, $P_n \left[H_{0,k}(\hat{\psi}) \right]$ consistently estimates the cumulative risk $E[Y_{k,\mathbf{g}=0}]$ at time k under “Never treat.” Indeed, Robins (1994) showed that $P_n \left[H_{0,k}(\hat{\psi}) \right]$ is consistent under the weaker exchangeability condition that $\mathbf{Y}_{m+1,\mathbf{g}=0} \perp\!\!\!\perp R_m | \bar{\mathbf{V}}_m, Y_m = 0$.

When, as in our application, blip function (4) is used,

$$H_{0,k}(\psi) = Y_k \prod_{m=0}^{k-1} \frac{k - m}{k - m + \exp[\psi R_m] - 1},$$

where $\frac{k - m}{k - m + \exp[\psi R_m] - 1}$ is the reciprocal of the exponentiated blip function (4). If the treatment is exercising at least 30 min/day, $P_n \left[H_{0,k}(\hat{\psi}) \right]$ is the estimated cumulative risk at time k if everyone had exercised less than 30 min/day throughout the period 1982–2002. As noted above, this blipping down procedure can be applied to any blip function, including those in Equations (3)–(7), regardless of whether (a) above holds.

5.2 Blipping Up

We next consider blipping up—a curiously more complicated computation. In Remark 3 below, we explain why the computation is so complex. We first consider one-dimensional parameter models like blip functions (3) and (4). When (a) above and the exchangeability assumption (Equation (8)) both hold, we can “blip up” from $E[H_{0,k}(\psi^*)]$ to the risk under any specified treatment regime $\mathbf{g} = \mathbf{r}$ with $\mathbf{r} \equiv \mathbf{r}_K$ by the formula

$$\begin{aligned} E \left[Y_{k,\mathbf{g}=\mathbf{r}} \right] &= \sum_{j=0}^{k-1} E \left[Y_{k-j,\mathbf{g}=\mathbf{0}} \right] \times t_{\mathbf{r}}(k, j, -1) \\ &= \sum_{j=0}^{k-1} E \left[H_{0,k-j}(\psi^*) \right] \times t_{\mathbf{r}}(k, j, -1), \end{aligned} \quad (13)$$

where the coefficients $t_{\mathbf{r}}(k, j, i)$ for $i = -1, 0, 1, \dots, k-2$ are recursively defined as

$$\begin{aligned} t_{\mathbf{r}}(k, 0, k-2) &= e_{\mathbf{r}}(k, k-1) \\ t_{\mathbf{r}}(k, j, k-1-[s+1]) &= t_{\mathbf{r}}(k, j, k-1-s) \times e_{\mathbf{r}}(k-j, k-1-s), \text{ for } j < s \\ t_{\mathbf{r}}(k, s, k-1-[s+1]) &= \left[1 - \sum_{j=0}^{s-1} t_{\mathbf{r}}(k, j, k-1-s) \right] \times e_{\mathbf{r}}(k-s, k-1-s) \end{aligned}$$

with $j \leq s$ and $e_{\mathbf{r}}(p, m) = \exp[\gamma_p(\mathbf{r}_m, \psi^*)]$. For example, under the regime “Always treat,” we have $r_m = 1$ for all $m > 0$, so $e_{\mathbf{r}}(p, m) = \exp[\psi^*]$ for blip function (3), and $e_{\mathbf{r}}(p, m) = 1 + (\exp[\psi^*] - 1)/(p - m)$ for blip function (4) (see Appendix B for a proof of Equation (13)).

After algebraic rearrangements and applying the definition of the blip function and the recursive definitions of $t_{\mathbf{r}}(k, j, i)$, Equation (13) amounts to the following: the cumulative probability of failure by time k under the treatment regime $\mathbf{g} = \mathbf{r}$ is a sum of probabilities of failure at each time $j < k$ (conditional on survival through $j-1$), weighted by cumulative probability of survival through $j-1$, wherein the blip function for the treatment at time $k-1$ is only multiplied by the final term in the sum (as only those who survive to time $k-1$ can be treated at that time). However, blip functions for earlier treatments are used to obtain the

other terms in the sum. It follows that, given estimates $\gamma_k(\mathbf{V}_0, \bar{R}_m; \hat{\psi})$ and $P_n[H_{0,k}(\hat{\psi})]$ computed above, we obtain a “substitution” estimator of $E[Y_{k,\mathbf{g}=\mathbf{r}}]$ by substituting these estimates for their respective estimands in Equation (13).

Remark 3. Robins (1994) and Robins, Rotnitzky, and Scharfstein (1999) provided the following simple formula for $E[Y_{k,\mathbf{g}=\mathbf{r}}]$ under a nondynamic regime when (a) holds:

$$E[Y_{k,\mathbf{g}=\mathbf{r}}] = E[Y_{k,\mathbf{g}=\mathbf{0}}] \exp \left[\sum_{m=0}^{k-1} \gamma_k(r_m; \psi^*) \right]$$

instead of Equation (13). The explanation for the difference is that the definition of a nondynamic regime for a failure time outcome used in this article is subtly different from that for the nonfailure time outcomes studied by Robins (1994) and Robins, Rotnitzky, and Scharfstein (1999). Indeed, under the definition in these articles, our nondynamic failure time regime is the dynamic regime: if without failure at time m , take treatment r_m , otherwise take treatment 0. This regime is dynamic because the treatment at time m depends on the value of an evolving time-dependent covariate—the indicator of survival at time m .

However, under a SNMSTM, the counterfactual survival probability $E[S_{k,\mathbf{g}=\mathbf{r}}]$ is given by the right-hand side of the previous display with $S_{k,\mathbf{g}=\mathbf{0}} = 1 - Y_{k,\mathbf{g}=\mathbf{0}}$ substituted for $Y_{k,\mathbf{g}=\mathbf{0}}$.

We now consider two-dimensional parameter models that depend on a prior treatment (blip function (5)) or a time-invariant covariate X_0 (blip function (6)).

For blip function (5), including an interaction term with treatment at the previous time, we use Equation (13) as stated, where $t_{\mathbf{r}}(k, j, -1)$ is now calculated using

$$e_{\mathbf{r}}(p, m) = 1 + \frac{\exp[(\psi_1 + \psi_2 r_{m-1}) \times r_m] - 1}{p - m},$$

and $r_{-1} = 0$.

For blip function (6), we first blip down separately in each stratum of X_0 to compute the conditional means $E[S_{k,\mathbf{g}=\mathbf{r}}|X_0=x_0]$ for each k . We then blip up using a conditional version of Equation (13),

$$E \left[Y_{k,\mathbf{g}=\mathbf{r}} | X_0=x_0 \right] = \sum_{j=0}^{k-1} E \left[H_{0,k-j}(\psi^*) | X_0=x_0 \right] \times t_{\mathbf{r}}(k, j, -1),$$

where $e_{\mathbf{r}}(p, m)$ becomes $e_{\mathbf{r},x_0}(p, m) = \exp[\gamma_p(x_0, r_m; \psi^*)]$ for each x_0 . At each time k , the marginal risk under the treatment regime $\mathbf{g} = \mathbf{r}$ is the weighted average

$$E \left[Y_{k,\mathbf{g}=\mathbf{r}} \right] = \sum_{x_0} w_{x_0} \times E \left[Y_{k,\mathbf{g}=\mathbf{r}} | X_0=x_0 \right], \quad (14)$$

where w_{x_0} is the unconditional probability that $X_0 = x_0$.

So far, we have described one method to estimate the treatment effect $E[Y_{k,\mathbf{g}=\mathbf{1}}] - E[Y_{k,\mathbf{g}=\mathbf{0}}]$ under condition (a) above: use g-estimation to obtain $\hat{\psi}$, estimate $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ by

$P_n \left[H_{0,k}(\hat{\psi}) \right]$, and finally estimate $E[Y_{k,\mathbf{g}=\mathbf{1}}]$ with the “substitution” estimator obtained from blipping up.

However, there is also a second, inequivalent method: estimate $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ by $P_n \left[H_{0,k}(\hat{\psi}) \right]$ as above; then estimate $E[Y_{k,\mathbf{g}=\mathbf{1}}]$ as follows. Define $R_m^\dagger = 1 - R_m$ and substitute $R_m^\dagger$ for R_m in Equation (12) and elsewhere. Define $\hat{\psi}^\dagger$ and $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right]$ to be $\hat{\psi}$ and $P_n \left[H_{0,k}(\hat{\psi}) \right]$, except based on $R_m^\dagger$ rather than R_m . Then $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right]$ estimates the cumulative risk $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ at time k and $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right] - P_n \left[H_{0,k}(\hat{\psi}) \right]$ is an estimate of $E[Y_{k,\mathbf{g}=\mathbf{1}}] - E[Y_{k,\mathbf{g}=\mathbf{0}}]$. A major advantage of the second method is that, in contrast to the first method, the estimate of $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ does not require assumption (a), since both $P_n \left[H_{0,k}(\hat{\psi}) \right]$ and $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right]$ are obtained by blipping down. A major disadvantage of the second method is that, in contrast to the first method, the second method is logically inconsistent for the following reason.

The estimators $P_n \left[H_{0,k}(\hat{\psi}) \right]$ and $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right]$ are based on different structural models. The estimate $P_n \left[H_{0,k}(\hat{\psi}) \right]$ is consistent for $E[Y_{k,\mathbf{g}=\mathbf{0}}]$ only if the SNCFTM of Equation (1) holds. In contrast, the estimate $P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right]$ is consistent for $E[Y_{k,\mathbf{g}=\mathbf{1}}]$ only if the model

$$\exp \left[\gamma_k \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = \frac{E \left[Y_{k,\mathbf{g}=\left(\bar{\mathbf{R}}_m, \perp \right)} | \bar{\mathbf{V}}_m, R_m, Y_m=0 \right]}{E \left[Y_{k,\mathbf{g}=\left(\bar{\mathbf{R}}_{m-1}, \perp \right)} | \bar{\mathbf{V}}_m, R_m, Y_m=0 \right]} \text{ for some } \psi$$

holds. However, if the null hypothesis that, for all regimes $\mathbf{g}$, $E[Y_{k,\mathbf{g}}] = E[Y_k]$ is false and, as in this article, the dimension of ψ is not large, these two structural models are generally

incompatible. Hence, at least one must be misspecified. Thus,

$P_n \left[H_{0,k}^\dagger(\hat{\psi}^\dagger) \right] - P_n \left[H_{0,k}(\hat{\psi}) \right]$ cannot be consistent for $E[Y_{k,g=1}] - E[Y_{k,g=0}]$. Of course, in truth both models will be somewhat misspecified and assumption (a) is never exactly true. In light of this, researchers may want to estimate $E[Y_{k,g=1}] - E[Y_{k,g=0}]$ both methods and see if the difference in the estimates is substantively unimportant.

Relationship to Counterfactual Hazard Models. In Appendix A we derive assumptions under which $\exp[\gamma_k(\mathbf{V}_m, R_m; \boldsymbol{\psi}^*)]$ approximates a counterfactual hazard ratio. We also describe why, when these assumptions do not hold, estimation of the ratio of counterfactual hazards under the structural nested hazard ratio model defined in the appendix is much less robust than g-estimation of the ratio of the counterfactual cumulative risks under an SNCFTM.

6. CENSORING

Selection bias due to censoring may be adjusted for by inverse probability weighting as follows. Let C_k be an indicator for whether an individual is censored (i.e., dead from a competing cause or lost to follow-up) at time k . We define the inverse probability weights

$$W_{m,k} = \prod_{j=m}^{k-1} \Pr \left[C_{j+1}=0 | C_j=0, Y_j=0, \bar{\mathbf{V}}_j, A_j \right]^{-1},$$

where $\Pr[C_{j+1}=0 | C_j=0, Y_j=0, \bar{\mathbf{V}}_j, A_j]$ is the probability of remaining uncensored through time $j+1$ conditional on surviving uncensored, and treatment and covariate history, at time j .

We can use g-estimation to consistently estimate the parameter $\boldsymbol{\psi}^*$ of the SNCFTM by using the weighted estimating equation (Robins, Rotnitzky, and Zhao 1995)

$$\begin{aligned} U(\boldsymbol{\psi}; \mathbf{B}, \mathbf{J}^*, \hat{\alpha}, \hat{v}) &\equiv \sum_{m=0}^K (1 - Y_m) \left\{ R_m - E \left[R_m | \bar{\mathbf{V}}_m, Y_m=0; \hat{\alpha} \right] \right\} \\ &\times \sum_{k=m+1}^{K+1} (1 - C_k) \mathbf{J}^* \left(\bar{\mathbf{V}}_m, k \right) \left[\widehat{W}_{m,k} H_{m,k}(\boldsymbol{\psi}) - B \left(\bar{\mathbf{V}}_m, k \right) \right] \end{aligned}$$

under Assumptions 1–3 with R_m replaced by $(R_m, C_{m+1}=0)$ and $Y_m=0$ by $(Y_m=0, C_m=0)$.

The estimated inverse probability weights $\widehat{W}_{m,k}$ were obtained by fitting a logistic model:

$$\Pr \left[C_{j+1}=0 | C_j=0, Y_j=0, \bar{\mathbf{V}}_j, A_j; \theta_0 \right] = \left(1 + \exp \left[-\theta'_0 \mathbf{z} \left(\bar{\mathbf{V}}_j, A_j \right) \right] \right)^{-1}$$

where $\mathbf{z}(\bar{\mathbf{V}}_j, A_j)$ is a vector of indicator variables for the categories of the covariates listed in Table 1. We fit separate models for censoring by death and censoring from other causes, and multiplied the estimated probabilities from each model at each time k to obtain the overall probability of remaining uncensored.

In our application, recall that $\mathbf{J}^*(\bar{\mathbf{V}}_m, k) = 1$ for one-parameter models (3) and (4), $(1, R_{m-1})'$ for model (5), and $(1, X_0)'$ for model (6); $B(\bar{\mathbf{V}}_m, k) = 0$.

For each intervention, the estimator $\hat{\psi}(\hat{\alpha}, \hat{\theta})$ that solved $P_n[U(\psi; \hat{\alpha}, \hat{\theta})] = 0$ was our g-estimate. We used the same minimization procedures described above, and nonparametric bootstraps based on 200 samples to obtain a 95% CI for the g-estimate.

In addition, the blipping down procedure was slightly modified: the risk $E[Y_{k,g=0}]$ at each time k under no treatment is estimated as the weighted average of $H_{0,k}(\hat{\psi})$ using the estimated weights $\hat{W}_{0,k}$. However, the blipping up procedure remains unchanged since, after the censored data has been used to obtain estimates of ψ and the $E[Y_{k,g=0}]$, it is not further used in the estimation of $E[Y_{k,g=1}]$.

7. DATA ANALYSIS

In the Nurses' Health Study, the g-estimates $\hat{\psi}$ (bootstrap 95% CIs) for blip function (4) were -0.668 ($-0.800, -0.565$) for the smoking intervention, -0.069 ($-0.244, 0.089$) for the exercise intervention, -0.034 ($-0.153, 0.101$) for the diet intervention, and -0.152 ($-0.310, -0.016$) for the alcohol intervention.

The observed 20-year risk of CHD, computed via a cumulative incidence method that took into account the subjects who died from other causes (Gooley et al. 1999), was 3.50%. We used the g-estimates $\hat{\psi}$ to calculate the 20-year CHD risks under the four "Always treat" interventions ($\mathbf{g} = \mathbf{1}$) by blipping down and up as described in Section 5. Table 2 displays these risks and their 95% CIs. Following the approach by Taubman et al. (2009), we computed risk ratios and risk differences (Table 2) by comparing the risks under each intervention with the risk under no intervention ($\mathbf{g} = \mathbf{R}$, observed treatments), which was estimated, via a weighted Kaplan-Meier method using the estimated weights $\hat{W}_{0,k}$, to be 4.16% (95% CI, 3.97–4.34). Note that this is the theoretical risk if nobody had died of other causes or been lost to follow-up. None of the four interventions resulted in a higher 20-year CHD risk than no intervention. The estimated risk ratios ranged from 0.80 for avoiding smoking to 0.97 for the dietary intervention. The bootstrap estimate of bias for the risk difference (absolute value of the difference between the point estimate and the bootstrap average) was 0.05 for each intervention. When we used the alternative method to calculate the risk under "Always treat" (reversing the coding for treatment and blipping down only), results were nearly identical.

We chose these hypothetical interventions for comparability with previous analyses (Taubman et al. 2009). However, g-estimation of SNCFTMs can be generally applied to other interventions. For example, when the method is applied to the intervention "Keep diet score in a range corresponding to the top quintile of the observed data" (instead of the top two quintiles), the estimated risk ratio (95% CI) was 0.62 (0.40, 0.92).

We also used the g-estimate $\hat{\psi}$ to calculate the 20-year CHD risks under the corresponding "Never treat" interventions ($\mathbf{g} = \mathbf{0}$, in other words, the counterfactual risks obtained by blipping down), and the risk differences and risk ratios comparing each "Always treat" intervention with its corresponding "Never treat." For each intervention, the bootstrap estimate of bias for this risk difference (absolute value of the difference between the point estimate and the bootstrap average) was 0.09. As expected, the estimates of the risk ratios in Table 3 were further from the null than those in Table 2. The difference increased with the proportion of women whose observed data were consistent with the "Always treat" intervention. Thus, the risk ratio for "Always treat" versus "Never treat" ranged from 0.41

(compared with 0.80 for “Always treat” versus “No intervention”) for avoiding smoking to 0.96 (compared with 0.97) for the dietary intervention.

We repeated the analyses using the SNCFTMs defined by blip functions (3), (5), and (6). For blip function (6), we considered $X_0 = I(A_{-1} = a_{-1})$, where a_{-1} is 30 min/day for exercise, 4 (quintile) for diet score, or 5 g/day of alcohol, and $X_0 = I(A_{-1} = 0 \text{ cigarettes/day})$ for the smoking intervention. Figure 1 compares the risk ratios for “Always treat” versus “No intervention” for each of the four interventions using the four different models (3)–(6). The estimates for all four interventions were reasonably similar across all four models. The estimate for the exercise intervention for blip function (5) was stronger than it was for the other models, perhaps indicating (as supported by the signs of the parameters) that this model better captured the distinction between habitual exercisers and those who began exercising only when they realized they were at elevated risk for CHD.

In interval cohorts like the Nurses’ Health Study, L_m and A_m are measured simultaneously, so their temporal order cannot be determined. The ordering used to define our causal parameters above—that of L_m preceding A_m for any given intervention—is in line with the strong assumption that, while possibly associated, (a) no risk factor measured at time m causes any other risk factor in the same period, and (b) the common causes that correlate the treatments at each time m are independent of those at any time m' different from m (Taubman et al. 2009; Robins, Hernán, and Siebert 2004). To assess the sensitivity of our estimates to this assumption, we repeated the analysis for each structural model assuming that A_m occurred before L_m . Risk ratios were all in the same direction as those in our main analyses, and changed by no more than 0.09.

Software to implement SNCFTMs, along with documentation for its use, is available on the website: <http://www.hsph.harvard.edu/causal>.

8. CONCLUSION

This article describes SNCFTMs, a new class of structural models, and a practical example of its application to complex longitudinal data. SNCFTMs have a key advantage over the previously proposed SNAFTMs: the estimating function is smooth in the parameter and thus g-estimation of SNCFTMs is numerically stable and can be carried out using standard optimization routines (e.g., the Newton–Raphson procedure).

It is useful to fit a number of different structural models. If results fluctuate dramatically for different choices, it will be important to consider the biological plausibility of the model, its vulnerability to residual confounding and measurement error, and the appropriateness of the interventions that are being considered.

Three approaches for causal inference from observational data are g-estimation of structural nested models, inverse probability weighting of marginal structural models, and the parametric g-formula. The comparison of effect estimates obtained using different methods can provide a crosscheck and shed light on the validity of the different modeling assumptions needed for each method. For comparison purposes, Figure 2 shows the risk ratios (with 95% CI) estimated for “Always treat” versus “No intervention” for our four representative interventions as presented in Table 2 alongside those obtained by using the parametric g-formula as described by Taubman et al. (2008) and allowing for censoring by death. The risk ratios estimated from the SNCFTM and the parametric g-formula are in the same direction for all four interventions. Unsurprisingly, considering the vastly fewer modeling assumptions required, the CIs from the semiparametric SNCFTM ranged from about the same width (for the smoking intervention) to about twice as wide (for the diet intervention) as the CIs from the fully parametric g-formula.

Our application of SNCFTMs has some limitations. First, our estimates are only valid under the assumption of exchangeability conditional on the measured variables. Because the Nurses' Health Study is an observational one, this assumption is not guaranteed to hold. Second, the estimates may not be valid when the rare failure condition (Remark 1) fails to hold. In our application, there was no evidence that this rare failure assumption failed to hold. Third, the functions B and J^* that we chose for computational simplicity, described in Section 4, do not result in a doubly robust, locally efficient estimator (Page 2005). Fourth, we only considered structural models $\exp[\gamma_k(\mathbf{V}_m, R_m; \psi^*)]$ that depended on $\mathbf{V}_m$ only through $\mathbf{V}_0$ and were thus without effect modification by time-varying covariates. This allowed us to use the method in Section 5.2 to compute absolute risks. Fifth, we only considered binary treatment variables, because we aimed to compare the results from the SNCFTM with those from the parametric g-formula (Taubman et al. 2008).

Finally, we used inverse probability weighting to adjust for informative censoring due to both loss to follow-up and competing risks (i.e., death from other causes). Thus, our method theoretically estimates the risk if no one in the study population had been lost to follow-up or died from other causes. In contrast, Taubman et al. (2009) estimated the risk if no one in the population had been lost to follow-up, while allowing the subjects in the population to die from other causes according to the observed distribution of death. This different estimand explains, in part, the difference in the risk estimate under no intervention between our method (4.16) and that found by Taubman and collaborators (3.68). (For caveats on the use of inverse probability weighting to adjust for informative censoring due to competing risks see, for example, section 12.2 of the article by Robins 1986, along with Rubin 1998, and Hernán, Hernández-Díaz, and Robins 2004.)

Future extensions of this work will include the development of the software needed both to compute doubly robust, locally efficient estimators and to implement alternative methods to adjust for censoring by competing risks.

Acknowledgments

This research was funded by NIH grants R01 HL34594 and R01 HL080644. The authors thank Roger Logan for programming support, and JoAnn Manson, Frank Hu, and Walter Willett for helpful comments on the article.

APPENDIX A: APPROXIMATING THE HAZARD RATIO

In this appendix, we will show how the blip function $\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m; \psi)$ relates to the hazard ratio.

Remark 4. If we take k to be that instant of time after m , that is, $k = m + 1$, where the unit of time is very small, and further assume that the hazard ratio is constant over time, then Equation (1) rewritten as

$$\exp \left[\gamma_{m,m+1} \left(\bar{\mathbf{V}}_m, R_m; \psi \right) \right] = \frac{E \left[Y_{m+1, \mathbf{g} = \left(\bar{\mathbf{R}}_m, 0 \right)} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0 \right]}{E \left[Y_{m+1, \mathbf{g} = \left(\bar{\mathbf{R}}_{m-1}, 0 \right)} \mid \bar{\mathbf{V}}_m, R_m, Y_m = 0 \right]}$$

is an approximation to the conditional hazard ratio. This formulation does not require a rare failure assumption, but it does require that relative to the median survival time, say, the time intervals are measured in very small units.

For the rest of this section, we will consider time k as a continuous variable. Let $X \approx W$ mean that X and W are arbitrarily close. Since we are assuming a rare outcome, but the time intervals may not be small, we will prove the following theorem.

Theorem 5. Suppose $Y_m = 0$. Let $\lambda_{Y_{\mathbf{g}}}(k|\bar{\mathbf{V}}_m, R_m)$ represent the conditional counterfactual hazard of Y at time k under regime $\mathbf{g}$. Define the conditional hazard ratio at time k comparing treatment at time m , R_m , with no treatment at time m (0_m) among individuals with the same covariate and treatment history up to m , as follows:

$$HR(k|\bar{\mathbf{V}}_m, R_m) = \frac{\lambda_{Y_{\mathbf{g}}}(k|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_m, \underline{0})}}{\lambda_{Y_{\mathbf{g}}}(k|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_m, \underline{0})}}.$$

Under the (limiting) rare failure assumption that for all $\mathbf{g}$, $\Pr(Y_{\mathbf{g},K}|\bar{\mathbf{V}}_m, R_m) \approx 0$, if the conditional hazard ratio is constant over time, that is,

$$HR(k|\bar{\mathbf{V}}_m, R_m) = HR(\bar{\mathbf{V}}_m, R_m), \quad K > k \geq m,$$

then

$$\exp \left[\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m) \right] \approx HR(\bar{\mathbf{V}}_m, R_m),$$

which does not depend on k .

Proof. In Equation (1), the conditional expectation of Y is the cumulative distribution of failure up to k given $(\bar{\mathbf{V}}_m, R_m)$. Since the density at any time u is equal to the product of hazard at u and the survival up to time u , Equation (1) may be written as

$$e^{\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)} = \frac{\int_m^k \lambda_{Y_{\mathbf{g}}}(u|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_m, \underline{0})} S_{Y_{\mathbf{g}}}(u|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_m, \underline{0})} du}{\int_m^k \lambda_{Y_{\mathbf{g}}}(t|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \underline{0})} S_{Y_{\mathbf{g}}}(t|\bar{\mathbf{V}}_m, R_m)_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \underline{0})} dt}, \quad (\text{A.1})$$

where $S_{Y_{\mathbf{g}}}(u|\bar{\mathbf{V}}_m, R_m) = \Pr[Y_{u,\mathbf{g}} = 0|\bar{\mathbf{V}}_m, R_m, Y_{m,\mathbf{g}} = 0]$ is the conditional counterfactual survival function of $Y_{\mathbf{g}}$ at time u .

Under the rare failure assumption, $S_{Y_{\mathbf{g}=(\bar{\mathbf{R}}_m, \underline{0})}}(u|\bar{\mathbf{V}}_m, R_m)$ is arbitrarily close to 1 for all $u < K$. Thus, Equation (A.1) may be approximated by

$$e^{\gamma_{m,k}(\bar{V}_m, R_m)} \approx \frac{\int_m^k \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_m \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right) du}{\int_m^k \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(t | \bar{V}_m, R_m \right) dt}.$$

This approximation will be reasonable as long as $S_{Y_{\bar{g} = (\bar{R}_m, \underline{0})}}(u | \bar{V}_m, R_m)$ is greater than about 0.95. If we rewrite the numerator and rearrange as follows, we obtain:

$$\begin{aligned} e^{\gamma_{m,k}(\bar{V}_m, R_m)} &\approx \frac{\int_m^k \left[\frac{\lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_m \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right)}{\lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right)} \right] \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right) du}{\int_m^k \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(t | \bar{V}_m, R_m \right) dt} \\ &= \frac{\int_m^k HR \left(u | \bar{V}_m, R_m \right) \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right) du}{\int_m^k \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(t | \bar{V}_m, R_m \right) dt} \\ &= \int_m^k HR \left(u | \bar{V}_m, R_m \right) w \left(m, u, k | \bar{V}_m, R_m \right) du, \end{aligned}$$

where

$$w \left(m, u, k | \bar{V}_m, R_m \right) = \frac{\lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(u | \bar{V}_m, R_m \right)}{\int_m^k \lambda_Y \left(\bar{g} = \begin{pmatrix} \bar{R}_{m-1} \\ \underline{0} \end{pmatrix} \right) \left(t | \bar{V}_m, R_m \right) dt}.$$

If $HR(u | \bar{V}_m, R_m) = HR(\bar{V}_m, R_m)$ does not depend on u , then we can pull it out of the integral. Note that

$$\int_m^k w \left(m, u, k | \bar{V}_m, R_m \right) du = 1,$$

and therefore $e^{\gamma_{m,k}(\bar{V}_m, R_m)} \approx HR(\bar{V}_m, R_m)$. $\square$

The above proof also shows that if $HR(u | \bar{V}_m, R_m)$ depends on time u , then $e^{\gamma_{m,k}(\bar{V}_m, R_m)}$ depends on time k and is a weighted average of the $HR(u | \bar{V}_m, R_m)$, $m \leq u \leq k$ with u -specific weights proportional to the hazard $\lambda_{Y_{\bar{g} = (\bar{R}_{m-1}, \underline{0})}}(u | \bar{V}_m, R_m)$. Since we are modeling $e^{\gamma_{m,k}(\bar{V}_m, R_m)}$ but may be interested in hazard ratios, we invert this last relationship and derive $HR(u | \bar{V}_m, R_m)$ as a function of u when $\gamma_{m,k}(\bar{V}_m, R_m)$ depends on k .

Theorem 6. If $Y_m = 0$, then, under the rare failure assumption, for $k > m$,

$$HR\left(k|\bar{\mathbf{V}}_m, R_m\right) \approx e^{\gamma_{m,k}\left(\bar{\mathbf{V}}_m, R_m\right)} \left[\frac{\frac{\partial \gamma_{m,k}\left(\bar{\mathbf{V}}_m, R_m\right)}{\partial k}}{w\left(m, k, k|\bar{\mathbf{V}}_m, R_m\right)} + 1 \right].$$

Proof. In the proof of Theorem 5, we found

$$e^{\gamma_{m,k}\left(\bar{\mathbf{V}}_m, R_m\right)} \approx \int_m^k HR\left(u|\bar{\mathbf{V}}_m, R_m\right) w\left(m, u, k|\bar{\mathbf{V}}_m, R_m\right) du, \quad (\text{A.2})$$

where

$$w\left(m, u, k|\bar{\mathbf{V}}_m, R_m\right) = \frac{\lambda_Y\left(\bar{\mathbf{g}}=\left(\bar{\mathbf{R}}_{m-1}, \underline{0}\right)\right)\left(u|\bar{\mathbf{V}}_m, R_m\right)}{\int_m^k \lambda_Y\left(\bar{\mathbf{g}}=\left(\bar{\mathbf{R}}_{m-1}, \underline{0}\right)\right)\left(t|\bar{\mathbf{V}}_m, R_m\right) dt}$$

can be seen as a weight proportional to the hazard at time u under no treatment at time m .

Taking derivatives of Equation (A.2) with respect to k , using the chain rule on the left hand side and using Leibniz's rule on the right results in the following:

$$\begin{aligned} e^{\gamma_{m,k}\left(\bar{\mathbf{V}}_m, R_m\right)} \frac{\partial \gamma_{m,k}\left(\bar{\mathbf{V}}_m, R_m\right)}{\partial k} &\approx HR\left(k|\bar{\mathbf{V}}_m, R_m\right) w\left(m, k, k|\bar{\mathbf{V}}_m, R_m\right) \\ &- HR\left(m|\bar{\mathbf{V}}_m, R_m\right) w\left(m, m, k|\bar{\mathbf{V}}_m, R_m\right) \frac{\partial m}{\partial k} \\ &+ \int_m^k \frac{\partial}{\partial k} \left[HR\left(u|\bar{\mathbf{V}}_m, R_m\right) w\left(m, u, k|\bar{\mathbf{V}}_m, R_m\right) \right] du. \end{aligned}$$

The middle term is 0 because m does not depend on k . Next, inside the integral, pull the hazard ratio $HR(u|\bar{\mathbf{V}}_m, R_m)$ out of the derivative since it is constant with respect to k .

$$\int_m^k \frac{\partial}{\partial k} \left[HR\left(u|\bar{\mathbf{V}}_m, R_m\right) w\left(m, u, k|\bar{\mathbf{V}}_m, R_m\right) \right] du = \int_m^k HR\left(u|\bar{\mathbf{V}}_m, R_m\right) \frac{\partial}{\partial k} \left[w\left(m, u, k|\bar{\mathbf{V}}_m, R_m\right) \right] du$$

Now the numerator of $w(m, u, k|\bar{\mathbf{V}}_m, R_m)$ is also constant with respect to k , so

$$\begin{aligned}
& \frac{\partial}{\partial k} \left[w \left(m, u, k | \bar{\mathbf{V}}_m, R_m \right) \right] \\
&= - \lambda_{Y_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}} \left(u | \bar{\mathbf{V}}_m, R_m \right) \frac{\lambda_{Y_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}} \left(k | \bar{\mathbf{V}}_m, R_m \right)}{\left(\int_m^k \lambda_{Y_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}} \left(u | \bar{\mathbf{V}}_m, R_m \right) du \right)^2} \\
&= - w \left(m, u, k | \bar{\mathbf{V}}_m, R_m \right) w \left(m, k, k | \bar{\mathbf{V}}_m, R_m \right)
\end{aligned}$$

by the chain rule and the Fundamental Theorem of Calculus. Putting all of this together,

$$\begin{aligned}
& e^{\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)} \frac{\partial \gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)}{\partial k} \approx HR \left(k | \bar{\mathbf{V}}_m, R_m \right) w \left(m, k, k | \bar{\mathbf{V}}_m, R_m \right) \\
& - \int_m^k HR \left(u | \bar{\mathbf{V}}_m, R_m \right) w \left(m, k, k | \bar{\mathbf{V}}_m, R_m \right) du
\end{aligned}$$

Note that $w(m, k, k | \bar{\mathbf{V}}_m, R_m)$ does not depend on u and can thus be pulled out of the integral. The rest of the integral is the right hand side of (A.2). From here, using Equation (A.2) and rearranging gives the desired result:

$$HR \left(m, k | \bar{\mathbf{V}}_m, R_m \right) \approx e^{\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)} \left[\frac{\frac{\partial \gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)}{\partial k}}{w \left(m, k, k | \bar{\mathbf{V}}_m, R_m \right)} + 1 \right].$$

□

Note that if one were interested in $HR(k | \bar{\mathbf{V}}_m, R_m)$ rather than $e^{\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m)}$ as an effect measure, one might consider parametric models for $HR(k | \bar{\mathbf{V}}_m, R_m)$. Such a model is referred to as a structural nested hazard ratio model. However, in contrast with an SNCFTM, a structural nested hazard ratio model does not admit unbiased estimating functions even when treatment probabilities are known (as in a sequential randomized trial; Robins 2000).

These results only consider comparisons between conditional means of $Y_{\mathbf{g}=(\bar{\mathbf{R}}_m, \mathbf{0})}$ and $Y_{\mathbf{g}=(\bar{\mathbf{R}}_{m-1}, \mathbf{0})}$ given $(\bar{\mathbf{V}}_m, R_m)$. We ultimately would be more interested in comparisons between the unconditional means of $Y_{\mathbf{g}}$ and $Y_{\mathbf{g}=\mathbf{0}^-}$ for various choices of $\mathbf{g}$. Define $HR_{\mathbf{g}}(k) = \lambda_{Y_{\mathbf{g}}}(k) / \lambda_{Y_{\mathbf{g}=\mathbf{0}^-}}(k)$. The above proofs can also be used to show that if the outcome is rare, that is,

$$\Pr \left[Y_{K, \mathbf{g}=\mathbf{0}^-} = 1 | \bar{\mathbf{V}}_m, R_m \right] \approx 0$$

for all combinations of $(\bar{\mathbf{V}}_m, R_m)$, then for $k > K + 1$ and for all $\mathbf{g}$,

$$HR_g(k) \approx \{w(k, k)\}^{-1} \frac{\partial}{\partial k} \left\{ \frac{E[Y_{k,g}]}{E[Y_{k,g=0}]} \right\} + \left\{ \frac{E[Y_{k,g}]}{E[Y_{k,g=0}]} \right\},$$

and

$$\frac{E[Y_{k,g}]}{E[Y_{k,g=0}]} \approx \int_0^k HR_g(u) w(u, k) du$$

where, similar to above,

$$w(u, k) = \frac{\lambda_{Y_{k,g=0}}(u)}{\int_m^k \lambda_{Y_{k,g=0}}(t) dt}.$$

This raises the question of whether, for certain $\mathbf{g}$'s, $E[Y_{k,g}]$ is a function of the joint distribution of the data only through $\gamma_{m,k}(\mathbf{V}_m, R_m)$ and $E[Y_{k,g=0}]$. In general, the answer is no except in the special case in which

$$\gamma_{m,k}(\bar{\mathbf{V}}_m, R_m) = \gamma_{m,k}(\mathbf{V}_0, R_m)$$

does not depend on $\mathbf{V}_m$ for all $m > 0$, in which case we say there is no treatment covariate interaction on the multiplicative cumulative failure scale; this case is explored in detail in Section 5.2 and Appendix B.

APPENDIX B: PROOF OF EQUATION (13) (BLIPPING UP)

In this appendix, we provide the proof of Equation (13). We will need two lemmas.

Lemma 7. Suppose that there are no unmeasured confounders and that (possibly after stratification on a pre-intervention variable) the blip function $\gamma_{m,k}$ depends only on treatment. Let

$$b_{\mathbf{r}}(p, m|j, s, t) = E \left[Y_{p,g=\left(\bar{\mathbf{r}}_m, \bar{\mathbf{0}}\right)} \middle| \bar{\mathbf{V}}_j, \bar{r}_s, Y_t=0 \right].$$

Let m be an integer ≥ -1 , and use the notation $e_{\mathbf{r}}(p, m) = \exp[\gamma_p(\bar{\mathbf{r}}_m; \boldsymbol{\psi})]$. Then

$$b_{\mathbf{r}}(m+1, m|m-1, m-1, m) = b_{\mathbf{r}}(m+1, m-1|m-1, m-1, m) \times e_{\mathbf{r}}(m+1, m).$$

Proof. By the definition of the blip function γ_k , and the assumption that the blip function does not depend on time-varying covariates, $b_{\mathbf{r}}(m+1, m|m, m, m) = b_{\mathbf{r}}(m+1, m-1|m, m, m) \times e_{\mathbf{r}}(m+1, m)$. Under exchangeability, the actual treatment received at time m is

independent of the counterfactual outcome at time $m + 1$, given past treatment and covariate history, so on both sides we can ignore the R_m in the condition. Thus,

$$b_{\mathbf{r}}(m+1, m | m, m-1, m) = b_{\mathbf{r}}(m+1, m-1 | m, m-1, m) \times e_{\mathbf{r}}(m+1, m).$$

Then, integrating out $\mathbf{V}_m$, we conclude that

$$b_{\mathbf{r}}(m+1, m | m, m-1, m-1, m) = b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m) \times e_{\mathbf{r}}(m+1, m)$$

because the blip function depends only on treatment. $\square$

The second lemma contains the main substance of the proof of the theorem.

Lemma 8. Define $b_{\mathbf{r}}(p, m | j, s, t)$, $e_{\mathbf{r}}(p, m)$, and m as above. Let $s \leq m$ be a positive integer and $j \leq s$ be a nonnegative integer, where j indexes integer valued differences of earlier failure times relative to $m + 1$ and s indexes integer valued differences of earlier treatment occasions relative to $m + 1$. Define the functions $t_{\mathbf{r}}$ recursively as follows:

$$\begin{aligned} t_{\mathbf{r}}(k, 0, k-2) &= e_{\mathbf{r}}(k, k-1) \\ t_{\mathbf{r}}(k, j, k-1 - [s+1]) &= t_{\mathbf{r}}(k, j, k-1-s) \times e_{\mathbf{r}}(k-j, k-1-s), \text{ for } j < s \\ t_{\mathbf{r}}(k, s, k-1 - [s+1]) &= \left[1 - \sum_{j=0}^{s-1} t_{\mathbf{r}}(k, j, k-1-s) \right] \times e_{\mathbf{r}}(k-s, k-1-s) \end{aligned}$$

Then, under the assumptions of Lemma 7:

$$b_{\mathbf{r}}(m+1, m | m-s, m-s, m-s) = \sum_{j=0}^s b_{\mathbf{r}}(m+1-j, m-(s+1) | m-s, m-s, m-s) \times t_{\mathbf{r}}(m+1, j, m-(s+1))$$

and

$$\begin{aligned} b_{\mathbf{r}}(m+1, m | m-s-1, m-s-1, m-s) \\ = \sum_{j=0}^s b_{\mathbf{r}}(m+1-j, m-(s+1) | m-(s+1), m-(s+1), m-s) \\ \times t_{\mathbf{r}}(m+1, j, m-(s+1)). \end{aligned}$$

Proof. Note that for each value of s , the first equation implies the second, by the same reasoning that was used in Lemma 7. Specifically, on both sides of the first equation, we change the condition from “ $|m-s, m-s, m-s$ ” to “ $|m-s, m-(s+1), m-s$ ” using the assumption of no unmeasured confounders: the counterfactual outcome at a later time ($m+1$ or $m+1-j$) under a given treatment regime is independent of actual treatment received at an earlier time ($m-s$) given the past ($\mathbf{R}_{m-(s+1)}$), and then we integrate out $\mathbf{V}_{m-s}$ on both sides. This yields the second equation.

The proof that the first equation is true proceeds by induction.

Base case: Consider $s = 1$. We want to prove:

$$\begin{aligned}
 & b_{\mathbf{r}}(m+1, m | m-1, m-1, m-1) \\
 &= b_{\mathbf{r}}(m+1, m-2 | m-1, m-1, m-1) \\
 & \times t_{\mathbf{r}}(m+1, 0, m-2) \\
 & + b_{\mathbf{r}}(m, m-2 | m-1, m-1, m-1) \\
 & \times t_{\mathbf{r}}(m+1, 1, m-2).
 \end{aligned}$$

Note that by the recursive definition,

$$\begin{aligned}
 t_{\mathbf{r}}(m+1, 0, m-2) &= t_{\mathbf{r}}(m+1, 0, m-1) \times e_{\mathbf{r}}(m+1-0, m-1) \\
 &= e_{\mathbf{r}}(m+1, m) \times e_{\mathbf{r}}(m+1, m-1),
 \end{aligned}$$

and also

$$\begin{aligned}
 t_{\mathbf{r}}(m+1, 1, m-2) &= [1 - t_{\mathbf{r}}(m+1, 0, m-1)] \times e_{\mathbf{r}}(m+1-1, m-1) \\
 &= [1 - e_{\mathbf{r}}(m+1, m)] \times e_{\mathbf{r}}(m, m-1).
 \end{aligned}$$

Replacing these terms, the equation we are trying to prove becomes:

$$\begin{aligned}
 & b_{\mathbf{r}}(m+1, m | m-1, m-1, m-1) \\
 &= b_{\mathbf{r}}(m+1, m-2 | m-1, m-1, m-1) \\
 & \times e_{\mathbf{r}}(m+1, m-1) \\
 & + b_{\mathbf{r}}(m, m-2 | m-1, m-1, m-1) \\
 & \times [1 - e_{\mathbf{r}}(m+1, m)] \\
 & \times e_{\mathbf{r}}(m, m-1).
 \end{aligned}$$

Starting from the left hand side,

$$\begin{aligned}
 & b_{\mathbf{r}}(m+1, m | m-1, m-1, m-1) \\
 &= b_{\mathbf{r}}(m+1, m | m-1, m-1, m) \\
 & \times [1 - b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1)] \\
 & + b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1).
 \end{aligned} \tag{B.1}$$

This is because if events Q and T are mutually exclusive, then $\Pr[Q \vee T] = \Pr[Q|\neg T] \times \Pr[\neg T] + \Pr[T]$.

Here, in the population of people who survived up to time $m-1$ with a given covariate and treatment history up to then, $Q \vee T$ refers to the (counterfactual) event “failure between time $m-1$ and time $m+1$ under a regime whose final blip of treatment occurs at time m ,” which can occur in one of two mutually exclusive ways: T is the counterfactual event “failure between times $m-1$ and m under a final blip of treatment at time $m-1$ ” and Q is the counterfactual event “failure between times m and $m+1$ under a final blip of treatment at time m ,” which can only occur if T does not.

By Lemma 7, the first factor on the right hand side of Equation (B.1) can be rewritten to yield

$$\begin{aligned}
 &= e_{\mathbf{r}}(m+1, m) \times b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m) \\
 &\quad \times [1 - b_{\mathbf{r}}^-] [(m, m-1 | m-1, m-1, m-1)] \\
 &\quad + b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1).
 \end{aligned}$$

However, since $\Pr[Q|\neg T] \times \Pr[\neg T] = \Pr[Q \vee T] - \Pr[T]$, part of the product becomes a difference:

$$\begin{aligned}
 &b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m) \\
 &\quad \times [1 - b_{\mathbf{r}}^-(m, m-1 | m-1, m-1, m-1)] \\
 &= b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m-1) \\
 &\quad - b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1).
 \end{aligned}$$

So now we have

$$\begin{aligned}
 &b_{\mathbf{r}}^-(m+1, m | m-1, m-1, m-1) \\
 &= e_{\mathbf{r}}(m+1, m) \times [b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m-1) - b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1)] \\
 &\quad + b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1).
 \end{aligned}$$

Algebraic rearrangement yields:

$$= e_{\mathbf{r}}(m+1, m) \times b_{\mathbf{r}}(m+1, m-1 | m-1, m-1, m-1) + [1 - e_{\mathbf{r}}(m+1, m) \times b_{\mathbf{r}}(m, m-1 | m-1, m-1, m-1)]$$

which becomes, after applying the definition of the blip function again,

$$\begin{aligned}
 &= e_{\mathbf{r}}^-(m+1, m) \times e_{\mathbf{r}}^-(m+1, m-1) \\
 &\quad \times b_{\mathbf{r}}(m+1, m-2 | m-1, m-1, m-1) \\
 &\quad + [1 - e_{\mathbf{r}}(m+1, m)] \\
 &\quad \times e_{\mathbf{r}}(m, m-1) \times b_{\mathbf{r}}(m, m-2 | m-1, m-1, m-1).
 \end{aligned}$$

This is the right-hand side of the equation we wanted to prove. So the lemma holds for $s = 1$.

Induction Step: Suppose that the theorem holds for $s - 1$. In other words,

$$\begin{aligned}
 &b_{\mathbf{r}}(m+1, m | m - (s-1), m - (s-1), m - (s-1)) \\
 &= \sum_{j=0}^{s-1} b_{\mathbf{r}}(m+1-j, m-s | m - (s-1), m - (s-1), m - (s-1)) \\
 &\quad \times t_{\mathbf{r}}(m+1, j, m-s)
 \end{aligned}$$

and

$$b_{\mathbf{r}}(m+1, m | m-s, m-s, m-(s-1)) = \sum_{j=0}^{s-1} b_{\mathbf{r}}(m+1-j, m-s | m-s, m-s, m-(s-1)) \times t_{\mathbf{r}}(m+1, j, m-s). \quad (\text{B.2})$$

We want to prove that the theorem holds for s :

$$b_{\mathbf{r}}(m+1, m | m-s, m-s, m-s) = \sum_{j=0}^s b_{\mathbf{r}}(m+1-j, m-(s+1) | m-s, m-s, m-s) \times t_{\mathbf{r}}(m+1, j, m-(s+1)). \quad (\text{B.3})$$

Starting from the left hand side of Equation (B.3), the first step follows the same reasoning as in the base case:

$$\begin{aligned} & b_{\mathbf{r}}(m+1, m | m-s, m-s, m-s) \\ &= b_{\mathbf{r}}(m+1, m | m-s, m-s, m-(s-1)) \\ & \quad \times \{1 - b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s)\} \\ & \quad + b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s). \end{aligned}$$

Now substituting in for the first $b_{\mathbf{r}}$ using the second part of the induction hypothesis (Equation (B.2)), we obtain

$$\begin{aligned} &= \left\{ \sum_{j=0}^{s-1} b_{\mathbf{r}}(m+1-j, m-s | m-s, m-s, m-(s-1)) \times t_{\mathbf{r}}(m+1, j, m-s) \right\} \\ & \quad \times \{1 - b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s)\} \\ & \quad + b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s). \end{aligned}$$

Distributing the $\{1 - b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s)\}$ term into the sum and using the same reasoning as in the base case, this becomes:

$$\begin{aligned} &= \sum_{j=0}^{s-1} \left\{ t_{\mathbf{r}}(m+1, j, m-s) \times \left[b_{\mathbf{r}}(m+1-j, m-s | m-s, m-s, m-s) - b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s) \right] \right\} \\ & \quad + b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s). \end{aligned}$$

Using some algebra, we get

$$\begin{aligned} &= \left\{ \sum_{j=0}^{s-1} b_{\mathbf{r}}(m+1-j, m-s | m-s, m-s, m-s) \times t_{\mathbf{r}}(m+1, j, m-s) \right\} \\ & \quad + \left\{ 1 - \sum_{j=0}^{s-1} t_{\mathbf{r}}(m+1, j, m-s) \right\} \\ & \quad \times b_{\mathbf{r}}(m-(s-1), m-s | m-s, m-s, m-s). \end{aligned}$$

Applying the definition of the blip function both inside the sum and to the term at the end of the equation, we obtain:

$$\begin{aligned}
&= \sum_{j=0}^{s-1} e_{\mathbf{r}}^-(m+1-j, m-s) \\
&\quad \times b_{\mathbf{r}}^-(m+1-j, m-(s+1) | m-s, m-s, m-s) \\
&\quad \times t_{\mathbf{r}}^-(m+1, j, m-s) \\
&\quad + \left\{ 1 - \sum_{j=0}^{s-1} t_{\mathbf{r}}^-(m+1, j, m-s) \right\} \\
&\quad \times e_{\mathbf{r}}^-(m-s+1, m-s) \\
&\quad \times b_{\mathbf{r}}^-(m-(s-1), m-(s+1) | m-s, m-s, m-s).
\end{aligned}$$

Now, in the sum $j < s$, so $t_{\mathbf{r}}(m+1, j, m-(s+1)) = t_{\mathbf{r}}(m+1, j, m-s) \times e_{\mathbf{r}}(m+1-j, m-s)$. We have exactly that product as a coefficient of $b_{\mathbf{r}}(m+1-j, m-(s+1) | m-s, m-s, m-s)$. Meanwhile,

$$t_{\mathbf{r}}^-(m+1, s, m-(s+1)) = \left\{ 1 - \sum_{j=0}^{s-1} t_{\mathbf{r}}^-(m+1, j, m-s) \right\} \times e_{\mathbf{r}}^-(m-s+1, m-s),$$

which is the coefficient of $b_{\mathbf{r}}(m-(s-1), m-(s+1) | m-s, m-s, m-s)$. So our expression becomes:

$$\begin{aligned}
&b_{\mathbf{r}}^-(m+1, m | m-s, m-s, m-s) \\
&= \left\{ \sum_{j=0}^{s-1} b_{\mathbf{r}}^-(m+1-j, m-(s+1) | m-s, m-s, m-s) \times t_{\mathbf{r}}^-(m+1, j, m-(s+1)) \right\} \\
&\quad + t_{\mathbf{r}}^-(m+1, s, m-(s+1)) \\
&\quad \times b_{\mathbf{r}}^-(m-(s-1), m-(s+1) | m-s, m-s, m-s).
\end{aligned}$$

However, the last term can be incorporated into the sum, yielding

$$= \sum_{j=0}^s b_{\mathbf{r}}^-(m+1-j, m-(s+1) | m-s, m-s, m-s) \times t_{\mathbf{r}}^-(m+1, j, m-(s+1)).$$

This is the right-hand side of Equation (B.3), which is what we were trying to prove.

□

Theorem 9. Under exchangeability and no treatment-covariate interaction on the multiplicative cumulative failure scale (i.e., $\gamma_k(\mathbf{V}_0, \mathbf{R}_m, \boldsymbol{\psi})$ is a function only of treatment history and baseline covariates), for any nondynamic regime $\mathbf{g} = \mathbf{r}^-$ and $m \leq K$,

$$E \left[Y_{m+1, \mathbf{g}=\mathbf{r}^-} \right] = \sum_{j=0}^m E \left[Y_{m+1-j, \mathbf{g}=\mathbf{0}} \right] \times t_{\mathbf{r}}^-(m+1, j, -1).$$

Proof. The theorem follows directly from Lemma 8 if we set $s = m$. Then $m - s - 1 = -1$ and $m - s = 0$, so the second formula from Lemma 8 becomes

$$b_{\mathbf{r}}(m+1, m | -1, -1, 0) = \sum_{j=0}^m b_{\mathbf{r}}(m+1-j, -1 | -1, -1, 0) \times t_{\mathbf{r}}(m+1, j, -1),$$

which can be rewritten as

$$E \left[Y_{m+1, \mathbf{g}=\bar{\mathbf{r}}_m} | \bar{\mathbf{V}}_{-1}, R_{-1}, Y_0=0 \right] = \sum_{j=0}^m E \left[Y_{m+1-j, \mathbf{g}=\bar{\mathbf{0}}} | \bar{\mathbf{V}}_{-1}, R_{-1}, Y_0=0 \right] \times t_{\mathbf{r}}(m+1, j, -1).$$

The conditioning event, however, is true by definition, so it can be left out. Thus,

$$E \left[Y_{m+1, \mathbf{g}=\bar{\mathbf{r}}} \right] = \sum_{j=0}^m E \left[Y_{m+1-j, \mathbf{g}=\bar{\mathbf{0}}} \right] \times t_{\mathbf{r}}(m+1, j, -1).$$

□

REFERENCES

- Bickel, P.; Klaassen, C.; Ritov, Y.; Wellner, J. Efficient and Adaptive Estimation for Semiparametric Models. The Johns Hopkins University Press; Baltimore, MD: 1993. [5]
- Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-Year Contribution to the Understanding of Health Among Women. *Journal of Women's Health*. 1997; 6:49–62. [2].
- Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, Hennekens CH, Speizer FE. Validation of Questionnaire Information on Risk Factors and Disease Outcomes in a Prospective Cohort Study of Women. *American Journal of Epidemiology*. 1986; 123:894–900. [2]. [PubMed: 3962971]
- Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators. *Statistics in Medicine*. 1999; 18:695–706. [8]. [PubMed: 10204198]
- Hernán MA, Brumback B, Robins JM. Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. *Journal of the American Statistical Association*. 2001; 96:440–448. [1].
- Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural Accelerated Failure Time Models for Survival Analysis in Studies With Time-Varying Treatments. *Pharmacoepidemiology and Drug Safety*. 2005; 14:477–491. [1]. [PubMed: 15660442]
- Hernán MA, Hernández-Díaz S, Robins JM. A Structural Approach to Selection Bias. *Epidemiology*. 2004; 15:615–625. [1,10]. [PubMed: 15308962]
- Hernán MA, VanderWeele TJ. Compound Treatments and Transportability of Causal Inference. *Epidemiology*. 2011; 22(3):368–377. [5]. [PubMed: 21399502]
- Kalbeisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. Wiley; New York: 1980. [1]
- Martinussen T, Vansteelandt S, Gerster M, Hjelmberg JVB. Estimation of Direct Effects for Survival Data by Using the Aalen Additive Hazards Model. *Journal of the Royal Statistical Society, Series B*. 2011; 73:773–788. [4].
- Page, J. 2005. Doubly Robust Estimation: Structural Nested Cumulative Failure Time Models. Sc.D. dissertation, Departments of Epidemiology and Biostatistics, Harvard School of Public Health. [1,4,5,6,10]

- Robins JM. A New Approach to Causal Inference in Mortality Studies With Sustained Exposure Periods: Application to Control of the Healthy Worker Survivor Effect. *Mathematical Modelling*. 1986; 7:1393–1512. [1,10].
- Robins JM. Correcting for Non-Compliance in Randomized Trials Using Structural Nested Mean Models. *Communications in Statistics - Theory and Methods*. 1994; 23:2379–2412. [1,4,5,6,7].
- Robins, JM. Structural Nested Failure Time Models. In: Armitage, P.; Colton, T., editors. *Encyclopedia of Biostatistics*. Wiley; Chichester/New York: 1998. p. 4372-4389.[1]
- Robins JM. Association, Causation, and Marginal Structural Models. *Synthese*. 1999; 121:151–179. [1].
- Robins, JM. Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Vol. 116. Springer-Verlag; New York: 2000. p. 95-134.[12]
- Robins JM, Blevins D, Ritter G, Wulfsohn M. G-Estimation of the Effect of Prophylaxis Therapy for *Pneumocystis carinii* Pneumonia on the Survival of AIDS Patients. *Epidemiology*. 1992; 3:319–336. [6]. [PubMed: 1637895]
- Robins JM, Blevins D, Ritter G, Wulfsohn M. Errata to G-Estimation of the Effect of Prophylaxis Therapy for *Pneumocystis carinii* Pneumonia on the Survival of AIDS Patients. *Epidemiology*. 1993; 14:79–97. [xxxx].
- Robins, JM.; Hernán, MA.; Siebert, U. Effects of Multiple Interventions. In: Ezzati, M.; Lopez, AD.; Rodgers, A.; Murray, CJL., editors. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Vol. I. World Health Organization; Geneva: 2004. p. 2191-2230.[3,9]
- Robins, JM.; Rotnitzky, A.; Scharfstein, D. Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: Halloran, ME.; Berry, D., editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials*. Vol. 116. Springer-Verlag; New York: 1999. p. 1-92.[6,7]
- Robins JM, Rotnitzky A, Zhao LP. Analysis of Semi-parametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*. 1995; 90:106–121. [8].
- Rubin DB. More Powerful Randomization-Based p -Values in Double-Blind Trials With Noncompliance. *Statistics in Medicine*. 1998; 17:371–385. [10]. [PubMed: 9493260]
- Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Food-Based Validation of a Dietary Questionnaire: The Effects of Week-to-Week Variation in Food Consumption. *International Journal of Epidemiology*. 1989; 18:858–867. [2]. [PubMed: 2621022]
- Stampfer MJ, Hu FB, Manson JE, Rimm EB, Willett WC. Primary Prevention of Coronary Heart Disease in Women Through Diet and Lifestyle. *The New England Journal of Medicine*. 2000; 343:16–22. [2]. [PubMed: 10882764]
- Taubman, SL.; Robins, JM.; Mittleman, MA.; Hernán, MA. JSM Proceedings 2008, Health Policy Statistics Section. American Statistical Association; Alexandria, VA: 2008. Alternative Approaches to Estimating the Effects of Hypothetical Interventions. [2,4,10]
- Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on Risk Factors for Coronary Heart Disease: An Application of the Parametric g-Formula. *International Journal of Epidemiology*. 2009; 38:1599–1611. [1,2,6,9,10]. [PubMed: 19389875]
- van der Laan, MJ.; Robins, JM. *Unified Methods for Censored Longitudinal Data and Causality*. Springer; New York: 2003. [5]
- Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, Rosner B, Kriska A, Willett WC. Reproducibility and Validity of a Self-Administered Physical Activity Questionnaire. *International Journal of Epidemiology*. 1994; 23:991–999. [2]. [PubMed: 7860180]
- Young, JG.; Hernán, MA.; Picciotto, S.; Robins, JM. Proceedings of the Section on Statistics in Epidemiology. ASA; Alexandria, VA: 2008. Simulation From Structural Survival Models Under Complex Time-Varying Data Structures; p. 1130-1135.[3]
- Young JG, Hernán MA, Picciotto S, Robins JM. Relation Between Three Classes of Structural Models for the Effect of a Time-Varying Exposure on Survival. *Lifetime Data Analysis*. 2010; 16:71–84. [1]. [PubMed: 19894116]

Figure 1.

Comparison of results using four different structural models for the effect of four hypothetical interventions on 20-year risk of coronary heart disease in the Nurses' Health Study. Ratios of the risks for "Always treat" to the risks for "No intervention" (with 95% confidence intervals) using blip function (4) are represented by diamonds, blip function (3) by squares, blip function (5) by triangles, and blip function (6) by circles. Data points are grouped by intervention: from left to right, the interventions are on smoking, exercise, diet, and alcohol.

Figure 2.

Comparison of results using SNCFTM and parametric g-formula for the effect of four hypothetical interventions on 20-Year risk of coronary heart disease in the Nurses' Health Study. Ratios of the risks for "Always treat" to the risks for "No intervention" (with 95% confidence intervals) using SNCFTM with blip function (4) are represented by diamonds, and parametric g-formula by squares. Data points are grouped by intervention: from left to right, the interventions are on smoking, exercise, diet, and alcohol.

Table 1 Covariates included in the models for treatment (R_m) and for remaining uncensored ($C_{m+1} = 0$)

Time-varying covariates in V_m	Years assessed	No. of categories
Time period	All	10
Multivitamins use	Starting in 1980	2
Aspirin use	1980, 1982, 1984, 1988 on	2
Statin use	1988, 1994 on	2
Postmenopausal hormone use	All	2
Number of cigarettes ^a	All	5
Exercise ^a	1980, 1982, 1986, 1988, 1992 on	6
Diet score ^a	1980, 1984, 1986, 1990, 1994, 1998	5
Alcohol intake ^a	1980, 1984, 1986, 1990, 1994, 1998	4
High blood pressure	All	2
High cholesterol	All	2
Diabetes	All	2
Angina	All	2
Stroke	All	2
Coronary artery bypass graft	Starting in 1986	2
Cancer	All	2
Menopause	All	2
Osteoporosis	Starting in 1982	2
BMI	All	6
Baseline covariates (1980) in V_0		
Age		5
Parental MI before age 60		2
Smoking history		2
Oral contraceptive history		2
BMI at age 18		5
BMI		6
Exercise		6
Diet score		5
Alcohol intake		4

^aThis variable is omitted from the model for R_m when R_m represents behavior consistent with an intervention on this risk factor.

Table 2

Percent CHD risks under four “Always treat” interventions sustained over a 20-year period, risk differences (RD), and risk ratios (RR) comparing “Always treat” to “No intervention,” obtained from the SNCFTM, Nurses’ Health Study

Intervention	% Risk under “Always treat”	95% CI	“Always treat” vs. “No intervention”			
			RD	95% CI	RR	95% CI
Avoid smoking	3.32	3.10, 3.47	−0.84	−0.98, −0.72	0.80	0.77, 0.83
Exercise 30 min/day	3.84	3.08, 4.65	−0.33	−1.03, 0.47	0.92	0.75, 1.11
Diet in top two quintiles	4.03	3.58, 4.63	−0.13	−0.57, 0.43	0.97	0.86, 1.10
Consume 5 g/day alcohol	3.59	3.07, 4.15	−0.57	−1.07, −0.07	0.86	0.75, 0.98

Table 3

Percent CHD risks under four “Never treat” interventions sustained over a 20-year period, risk differences (RD), and risk ratios (RR) comparing “Always treat” to “Never treat,” obtained from the SNCFTM, Nurses’ Health Study

Intervention	% Risk under “Never treat”	95% CI	“Always treat” vs. “No treat”			
			RD	95% CI	RR	95% CI
Avoid smoking	8.16	7.23, 9.41	−4.84	−6.10, −3.88	0.41	0.35, 0.47
Exercise 30 min/day	4.20	3.98, 4.40	−0.37	−1.16, 0.52	0.91	0.73, 1.13
Diet in top two quintiles	4.21	3.92, 4.47	−0.18	−0.79, 0.59	0.96	0.82, 1.15
Consume 5 g/day alcohol	4.38	4.07, 4.72	−0.80	−1.57, −0.09	0.82	0.67, 0.98