

A marginal structural model for multiple outcome survival data: Assessing the impact of intravenous drug use on several causes of death in the Canadian Co-infection Cohort

Erica E. M. Moodie^{a*}, David A. Stephens^b and Marina B. Klein^c

It is often the case that interest lies in the effect of an exposure on each of several distinct event-types. For example, we are motivated by an interest in the impact of recent intravenous drug use on deaths due to each of cancer, end-stage liver disease, and overdose in the Canadian Co-infection Cohort (CCC). We develop a marginal structural model that permits estimation of cause-specific hazards in situations where more than one cause of death is of interest. Marginal structural models allow for the causal effect of treatment on outcome to be estimated using inverse probability weighting under the assumption of no unmeasured confounding; these models are particularly useful in the presence of time-varying confounding variables which may also mediate the effect of exposures. An asymptotic variance estimator is derived, and a cumulative incidence function estimator is given. We compare the performance of the proposed marginal structural model for multiple outcome data to that of conventional competing risks models in simulated data, and demonstrate the use of the proposed approach in the CCC. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: Causal inference; Competing risks; Confounding; Failure-time data; Intermediate variables; Inverse probability weighting; Longitudinal data; Marginal structural models; multiple outcome data; Simulation; Survival analysis; Time-dependent confounding

1. Introduction

Assessment of the effect of treatment or exposure on multiple outcome survival data – time-to-event data where the terminal event can take one of a number of possible outcome types, also termed competing risks data – is typically complicated by the fact that the occurrence of one outcome precludes the occurrence of others. For example, in an HIV setting, patients may die of AIDS-related complications, end-stage liver disease, or other causes. Furthermore, in a longitudinal exposure setting, time-varying confounders of the exposure effect may also be mediators through which previous exposures act; for example, assessing the impact of intravenous drug use (IDU) on cause-specific mortality requires adjustment for confounding by time-varying factors such as CD4 cell count, viral load, and anti-retroviral therapy (ART) use, yet some or all of these variables may be affected by IDU and thus lie on the causal pathway between the exposure and the outcomes of interest. Thus there are two key challenges in the estimation of the effect of exposure. The first is the issue of how to account appropriately for, and understand, the multiple outcomes of interest (different causes of death), while the second challenge arises because there exist time-dependent confounding variables that are both affected

^aDepartment of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Ave. E., Montreal, QC, H3A 1A2, Canada ^bDepartment of Mathematics and Statistics, McGill University, 805 Sherbrooke Street W., Montreal, QC, H3A 2K6, Canada ^cDepartment of Medicine, McGill University Health Centre, 3650 Saint Urbain Street, Montreal, QC, H2X 2P4, Canada

* Correspondence to: Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Ave. E., Montreal, QC, H3A 1A2, Canada. E-mail: erica.moodie@mcgill.ca

by evolving exposure to intravenous drugs and are independent prognostic factors for one or more of the causes of death being studied.

The possibility of confounding of time-varying exposures was first addressed in the survival literature a quarter century ago. Robins proposed G-computation as a means of assessing the impact of a time-varying sequence of treatments on an outcome [1]. Keiding et al. found the parameters of an accelerated failure time model using G-estimation in order to account for the bias that can arise due to time-varying confounding and intermediate variables [2]. In that paper, the competing risk (death) to the outcome of interest (relapse) was treated as a censoring event, which was handled by inverse-probability-of-censoring weighting. More recently, marginal structural models (MSMs) have been used to estimate the effect of hospital-acquired infections on mortality in a cause-specific model, censoring when a competing event occurred [3]. The same data were used with a focus on mortality in an intensive care unit (so that all censoring was administrative); a marginal structural model was developed for intensive care unit death [4]. In this paper, we construct a multiple outcome MSM (MO-MSM) based on a discrete time survival formulation with competing risks via a cause-specific hazard formulation (see [5, Chapter 8] for example)

Our approach is similar in some aspects to the method outlined in [4] but the methods bear several key differences. For example, our method is semi-parametric and designed to provide cause-specific hazard ratios, while the method in [4] is non-parametric, aimed at describing the cumulative incidence function for a given single outcome type. In our motivating example, we are specifically interested in understanding the effect of an exposure on several different outcomes, while Bekaert and colleagues treat censoring as a competing risk as an ingenious means of avoiding instability of traditional MSM estimators in the presence of extreme weights. Finally, the approach of [4] is essentially restricted to monotone treatment/exposure patterns, whereas our approach is constructed to allow for more general patterns; in our example, the exposure of interest – in our example, intravenous drug use amongst co-infected HIV/Hepatitis C (HCV) subjects – is intermittent.

In this paper, we develop a semi-parametric approach for multiple outcome data to unbiasedly estimate the effect of a time-varying exposure in the presence of time-varying confounding covariates that act also as intermediates between exposure and outcomes. We use inverse probability weighting as the basis for an estimation procedure for marginal structural models [6–8]. In doing so, we produce an approach that can accommodate high-dimensionality in the confounders and avoid bias due to adjustment for intermediate variables (as a consequence of collider-stratification or blocking of indirect effects), whilst simultaneously avoiding bias that occurs when confounding variables are omitted from a regression model, provided all relevant covariates have been recorded.

We begin by describing the context for the problem, motivating our statistical developments with the data from the Canadian HIV/HCV Co-infection Cohort (CCC). In Section 3, we present models for the cause-specific hazard functions. We describe the semiparametric inference framework in Section 4, and demonstrate the large-sample properties of the estimators via simulation in Section 5. Finally, we illustrate the approach using data from the CCC in Section 6.

2. Setting: The effect of injection drug use on mortality in the Canadian HIV/HCV Co-infection Cohort

Considerable attention has been devoted to the HIV epidemic among injection drug users, yet few studies have examined the impact of injection (intravenous) drug use among HIV-positive individuals. Anti-retroviral therapy has led to a dramatic reduction in AIDS related deaths, however mortality in HCV co-infected individuals remains much higher than in people infected with HIV alone or than in the general population. The excess mortality appears to be driven in part by issues related to continued injection drug use (IDU) following HCV infection as well as mortality from liver disease and cancers. Active drug use is associated with lifestyle destabilization that could result in adverse health outcomes which may be mediated through lack of adherence to treatment or poorer immune functioning [9], or may act directly on the liver. It is precisely for this reason – that the effect of treatment may act through covariates such as CD4 cell count (a marker of health and immune function) which may also act as confounding variables (if, say, poorer health increases the likelihood of using injection drugs and also increases risk of death) – that marginal structural models are required. Conventional survival models are unable to appropriately account for time-varying confounding variables that lie on the causal pathway from exposure to an outcome event.

We investigate whether IDU is causally linked three causes of death: (a) death by cancer (other than of the liver), (b) death due to end-stage liver disease (ESLD), defined as decompensated cirrhosis of the liver or liver cancer, or (c) overdose. We utilize a competing risks marginal structural model framework. HIV/AIDS deaths were not considered a potential competing risk, but rather were treated as a censoring event since very few were observed. We allow for the effect of IDU to be mediated through interruptions in ART as well as current health status, which in a co-infected population can be well-summarized by CD4 cell count, HIV viral load, and the AST-to-platelet ratio index (a measure of liver damage).

The data we study were obtained from the Canadian HIV/HCV Co-infection Cohort [10]. The cohort began participant recruitment in 2003; there are currently 1119 patients enrolled from 16 sites across Canada. Cohort eligibility required that patients were at least 16 years old with documented HIV infection and evidence of HCV exposure. The study was approved by the research ethics boards of all participating institutions. All participants included in this analysis had virologic evidence of HCV infection and had at least one cohort visit between April 2003 and July 2012. Patients were censored at their last clinic visit prior to July 2012. The outcome variable is validated cause of death (see details below).

3. Notation and assumptions

3.1. Notation

Suppose we are interested in K different causes of death, and observe follow-up to time Y_i in $i = 1, \dots, n$ individuals. We wish to quantify the influence of exposure (or treatment), A , on outcome. We assume that time is treated as a discrete quantity, and consider the case of regularly scheduled follow-up visits, with inter-visit intervals labelled $j = 1, \dots, J$, with visit j occurring at precisely j time units, that is, interval j runs from time equals $j - 1$ to time equals j . Enrollment to the study occurs at time equals zero. The discrete time formulation is common in the analysis of time-to-event data in a causal inference setting (see, for example, [8, 11–13]), and is often necessary in practice, especially when data are recorded primarily via clinic visits. An individual i is observed for $n_i \leq J$ intervals. Let A_{ij} denote the time-varying indicator of exposure in interval j . Let L_{ij} denote the vector of prognostic and confounding factors measured at visit $j - 1$, and $V_i \subseteq L_{i1}$ denotes baseline covariates. We follow the convention in the marginal structural modelling literature of using an overbar to denote the history of a time-dependent variable, so that, for example, $\bar{A}_{i3} = (A_{i1}, A_{i2}, A_{i3})$ is the exposure history for individual i through to the third interval/fourth visit (visits at times 0, 1, 2, and 3). We further let $\bar{A}_i$ denote the sequence of exposures $(A_{i1}, A_{i2}, \dots, A_{in_i})$.

Suppose event (failure or censoring) times $Y_1, \dots, Y_n$ are observed independently on N individuals, and outcome variables $D_0, D_1, \dots, D_n \in \{0, 1, \dots, K\}$ are observed, with $D_i = k > 0$ meaning failure due to cause k , and $D_i = 0$ meaning censored/alive at the last point of follow-up. We may further consider ‘status’ in each interval, denoted D_{ij} , indicating whether an individual is alive or has died due to some cause k in interval j . We use the indicator C_{ij} to indicate whether individual i was censored in interval j . That is, $D_{ij} > 0$ if individual i dies of some cause between time point $j - 1$ and time point j . Similarly, $C_{ij} = 0$ if the individual is observed at visit j , but $C_{ij} = 1$ if the individual is not observed at visit j , having dropped out between time point $j - 1$ and time point j . Throughout, we use capital letters to represent random variables and lower case letters to represent realizations of the random variables. Where possible to do so, individual indexing will be suppressed to reduce notational complexity.

We aim to quantify causal effects, which we do using a structural model – i.e., a model for the hazard of failure due to cause k ($k = 1, \dots, K$) under treatment regime $\bar{a}$. This can be accomplished with the help of counterfactuals (also called potential outcomes). Specifically, let $Y_i(\bar{a})$ denote the (discrete) death or censoring time of subject i had he received exposure sequence $\bar{a}$ rather than his observed treatment history $\bar{A}_i$. We can similarly let $Y^k(\bar{a})$ denote the random variable representing an individual’s counterfactual survival time to death by cause k had he been exposed to the particular exposure pattern $\bar{a}$ rather than his observed exposure history. The subjects actual event or censoring time, Y_i , equals his counterfactual survival/censoring time under the observed exposure history $\bar{A}$: $Y_i = Y_i(\bar{A}_i)$; this is known as the axiom of consistency and must be assumed to be true.

3.2. Defining the cause-specific hazards

We model hazards which are *cause-specific failure probabilities* at time j , conditional on survival up to at least j . Following Kalbfleisch and Prentice [5, Chapter 8, p. 251], in our discrete time setting, we define the cause-specific hazard probability for cause k at time j by

$$h_{jk}(\bar{a}) = P[Y(\bar{a}) = j, D = k | Y(\bar{a}) \geq j] = \frac{P[Y(\bar{a}) = j, D = k]}{P[Y(\bar{a}) \geq j]}. \quad (1)$$

This probability, termed by some authors [14] the *subhazard*, records the conditional probability of death in interval j from cause k , given that the individual has survived at least until interval j . The total (across all causes) hazard probability, $h_j(\bar{a})$ is

$$h_j(\bar{a}) = P[Y(\bar{a}) = j | Y(\bar{a}) \geq j] = \sum_{k=1}^K P[Y(\bar{a}) = j, D = k | Y(\bar{a}) \geq j] = \sum_{k=1}^K h_{jk}(\bar{a})$$

and the cause-specific failure probability $p_{jk}(\bar{a}) = P[Y(\bar{a}) = j, D = k]$ as

$$p_{jk}(\bar{a}) = h_{jk}(\bar{a}) \prod_{l=1}^{j-1} (1 - h_l(\bar{a})) = h_{jk}(\bar{a}) \prod_{l=1}^{j-1} h_{l0}(\bar{a}). \quad (2)$$

Consequently, the cause-specific *cumulative incidence function* (CIF) (or *subdistribution function*) for cause k is defined as

$$P_{jk}(\bar{a}) = P[Y(\bar{a}) \leq j, D = k] = \sum_{l=1}^j p_{lk}(\bar{a}) = \sum_{l=1}^j h_{lk}(\bar{a}) \left\{ \prod_{r=1}^{l-1} h_{r0}(\bar{a}) \right\}. \quad (3)$$

Thus, in the discrete time formulation, it is straightforward to compute the cause-specific CIF from the cause-specific hazards; see the discussion of the continuous time case in Section 4.4. The *all-cause* CIF is then

$$P_j(\bar{a}) = \sum_{k=1}^K P_{jk}(\bar{a}). \quad (4)$$

Finally, let

$$\pi_k = P[D = k] = \sum_{j=1}^{\infty} P[Y(\bar{a}) = j, D = k] = \sum_{j=1}^{\infty} p_{jk} < 1$$

so that $\pi_1 + \dots + \pi_K = 1$. This, of course, confirms that the cause- k probabilities $\{p_{jk}, j = 1, \dots, \infty\}$ constitute a *subdistribution*. Finally, we let $\mathbf{h}_j(\bar{a}) = (h_{j0}(\bar{a}), h_{j1}(\bar{a}), \dots, h_{jK}(\bar{a}))^T$ denote the vector of cause-specific hazards for the counterfactual failure at time j , with

$$h_{j0}(\bar{a}) = 1 - \sum_{k=1}^K h_{jk}(\bar{a}).$$

and $\mathbf{H}(\bar{a}) = [\mathbf{h}_1(\bar{a}), \dots, \mathbf{h}_J(\bar{a})]$.

Our cause-specific competing risks formulation is reasonably standard, but not the only approach available. We describe alternative model formulations using the subdistribution hazard, and give a discussion of terminology, in §S1 of the Supplementary Material, and explain further our decision to use cause-specific hazard probabilities as the building blocks of the analysis.

3.3. Counterfactual representation and the causal model.

In Section 4, we consider models for the distribution of hazards for the counterfactual outcomes, $Y^k(\bar{a})$. These are known as *causal* (or *structural*), models which quantify the causal effect of exposure sequences on death. Marginal structural models posit a model for $\mathbf{h}_j(\bar{a})$ for each exposure pattern $\bar{a}$, although interest may lie in particular patterns such as “never exposed” and “always exposed” ($\bar{a} = (0, 0, \dots, 0)$) and $\bar{a} = (1, 1, \dots, 1)$, respectively) or a more complex pattern such as interrupting the exposure every third interval ($\bar{a} = (1, 1, 0, 1, \dots)$).

Marginal structural models may be contrasted with the typical associational models considered in survival analysis, which model the *observed* (rather than counterfactual) failure times as a function of exposures in the population. Associational models are biased for the true marginal effect of exposure on the population survival time whenever there is a time-dependent confounder, L_j , that is predicted by previous exposure. The bias occurs since an associational model that adjusts for the time-dependent confounder will block any effect of exposure that is mediated through L_j (tending to underestimate the effect of exposure) and may additionally introduce bias by opening a “back door” path through a latent variable such as underlying health status [15], while an associational model that does not adjust for L_j fails to control for confounding [7, 8, 16]. Note, however, that conventional models are not biased in the absence of time-varying confounding, and conventional conditional models may be preferred over marginal models if the primary purpose of the analysis is predictive.

3.4. Assumptions

In our discrete survival setting, a *marginal structural model* (MSM) is a model for the counterfactual outcomes based on the counterfactual hazards $\mathbf{h}_j(\bar{a})$. In the presence of only administrative censoring (e.g., calendar date rather than, say, an individual withdrawing consent from the study or becoming too ill to return for follow-up), the MSM may be fit without asymptotic bias provided three conditions are met:

1. At every interval j , the vector of variables L_j contains all confounders for the treatment and outcome relationship. This is known as the sequential randomization assumption, or no unmeasured confounders [7].

2. The model for the conditional probability of being treated given exposure and covariate history, $P(A_j = 1 | \bar{L}_j, \bar{A}_{j-1}, V)$, is correctly specified with respect to confounding variables (“instruments”, which predict exposure but not outcome, may be omitted). This model is used to weight the contribution of each individual until the time that they experience an event or are lost to follow-up.
3. Each exposure level occurs with positive probability for all covariate/exposure histories. This is called the *positivity* or *experimental treatment assignment* (ETA) assumption. See [17] for a thorough discussion of the assumption, and a diagnostic tool.

We allow for the potential for follow-up to end for reasons other than administrative. Denoting the model for the probability of not being censored by $P(C_j = 0 | \bar{L}_j, \bar{A}_{j-1}, V)$, we obtain the following decomposition:

$$P(A_j = 1, C_j = 0 | \bar{L}_j, \bar{A}_{j-1}, V) = P(A_j = 1 | \bar{L}_j, \bar{A}_{j-1}, V, C_j = 0) P(C_j = 0 | \bar{L}_j, \bar{A}_{j-1}, V).$$

It may also be the case that the probability of censoring depends on the outcome the subject would have had in the absence of censoring due to covariates which affect both the outcome and the probability of being censored. In this situation, two additional assumptions are required.

4. At every interval j , the vector of variables L_j also contains all common variables required to block selection bias due to censoring.
5. The model for the probability of being observed at interval j , $P(C_j = 0 | \bar{L}_j, \bar{A}_{j-1}, V)$, is correctly specified with respect to variables that predict both censoring and the response.

It is possible to check whether balance in covariates is achieved between the exposure groups within strata of the propensity score, however the presence of unmeasured confounders is not testable and relies instead on substance area knowledge.

4. Modelling and estimation

We begin by describing a multinomial likelihood approach in a setting where there are no confounders, as we might expect in a sequential multiple assignment randomized trial [18]. We then demonstrate how this can be recast in the MSM framework to handle time-varying confounding and mediation.

4.1. A likelihood approach for the no-confounding situation

Suppose that exposure is randomized at each interval (not conditional on any covariates). We can then construct a likelihood via the cause-specific hazard probabilities from (1) as which is given by

$$L(\mathbf{H}) = \prod_{i=1}^N L_i(\mathbf{H}) = \prod_{i=1}^N \prod_{j=1}^{n_i} \prod_{k=0}^K h_{ijk}^{\mathbb{I}[D_{ij}=k]}$$

[5, §8.2.3]. In the hazard parameterization, as hazards represent conditional probabilities, the likelihood is the product of **conditionally independent** contributions; a typical profile for an individual subject with $n_i = 5$, with failure of type 2 is $h_{10}h_{20}h_{30}h_{40}h_{52}$, thus **at each interval** j , the outcome is Multinomial($1, h_{j0}, h_{j1}, \dots, h_{jK}$) with the outcomes across i, j being mutually conditionally independent. Therefore, the model can be fitted using multi-category regression, treating the multiple intervals for a single individual as independent contributions. We may then consider, for example, a specification for the cause-specific hazard model which allows for a proportional modification to the odds-on-death based on a baseline hazard $h_{jk}, j = 1, \dots, J, k = 1, \dots, K$ on a scale relative to the hazard of surviving in interval j, h_{j0} .

We consider, for example, a specification for the cause-specific hazard model which allows for the hazard to be modified by covariates in a distinct fashion for each of the K causes, with observed data Y (follow-up time), D (status: censoring or cause of death), A (exposure), and L (covariates). Consider covariates $X_{j1}, \dots, X_{jp}$ that may modify the baseline hazard, which may include baseline covariates (a subset of L) and possibly time-varying exposure (A). We do not include other time-varying covariates in this set of variables, so as not to block any effect of treatment which may be mediated through these covariates. For individual i , let

$$\eta_{ijk} = \sum_{l=1}^p \beta_{kl} x_{ijl} \quad k = 1, \dots, K$$

and

$$\log \left(\frac{h_{ijk}}{h_{ij0}} \right) = \log \left(\frac{h_{ijk}}{1 - h_{ij1} - \dots - h_{ijK}} \right) = \eta_{ijk} + \log \left(\frac{h_{jk}}{h_{j0}} \right).$$

where h_{ijk} denotes the hazard individualized to subject i by the covariates in interval j , specific to cause k , and h_{jk} denotes the baseline hazard for that interval/cause. The model therefore corresponds to a modification by covariates of the odds-on-death defined by baseline hazard h_{jk} , $j = 1, \dots, J$, $k = 1, \dots, K$ on a scale relative to the hazard of surviving in interval j , h_{j0} . For example, for a single binary factor A , we might have that when $A_i = 1$

$$\log \left(\frac{h_{ijk}/h_{ij0}}{h_{jk}/h_{j0}} \right) = \psi$$

but that when $A_i = 1$, the log odds-on-death for cause k is zero. The model is analogous to the proportional hazards model in continuous time, and we term it a *proportional odds-on-death* model. The general relationship yields

$$h_{ijk} = \frac{\omega_{jk} \exp\{\eta_{ijk}\}}{1 + \sum_{k=1}^K \omega_{jk} \exp\{\eta_{ijk}\}}$$

where $\omega_{jk} = h_{jk}/h_{j0}$ for $k = 1, \dots, K$. See [19] for an in-depth treatment of the correspondence between multinomial regression and complex survival models.

4.2. A marginal structural model for multiple-outcomes data

Marginal structural models address the aforementioned shortcomings of conventional regression approach. A MSM is a model for $\mathbf{h}_{jk}(\bar{a})$, that is, it is a model for the marginal distribution of the counterfactual (rather than the observed) cause-specific survival function for type of death k . Specifically, the MSM models the dependence of $Y(\bar{a})$ on a sequence of exposures and yields a causally-interpretable estimate of the population distribution of the outcomes that would be observed if the entire population were to have a particular (static) pattern of exposure imposed upon it, conditional on the assumptions detailed in Section 3.4.

The causal estimand is a quantity in the marginal model derived from the conditional model described in Section 4.1. For example, for a single binary exposure factor A , we might wish to assess the marginal effect of exposure in a single interval, and consider the parameter

$$\log \left(\frac{h_{ijk}(1)/h_{ij0}(1)}{h_{jk}/h_{j0}} \right) = \psi$$

say. Alternatively, we might wish to consider the aggregate effect of exposure up to interval j , and address the parameter

$$\log \left(\frac{h_{ijk}(g(\bar{a}_j))/h_{ij0}(g(\bar{a}_j))}{h_{jk}/h_{j0}} \right) = \psi_g$$

for some function g , say

$$g(\bar{a}_j) = \sum_{l=1}^j a_l.$$

Note that, here, $\bar{a}_j$ corresponds to a potential or counterfactual exposure pattern.

Estimation of marginal model parameters is achieved using inverse probability weighting (IPW) to account for selective observation of individuals with particular exposure and censoring patterns. Stabilized weights,

$$w_j^e = \prod_{t=1}^j \frac{P(A_t = a_t | \bar{A}_{t-1}, V, C_t = 0)}{P(A_t = a_t | \bar{L}_t, \bar{A}_{t-1}, V, C_t = 0)},$$

are used to appropriately reweight the observed sample without the greater variability typically observed when using the unstabilized weights [7, 8]. Additionally, weights must be constructed to account for the censoring process

$$w_j^c = \begin{cases} \prod_{t=1}^j \frac{P(C_t = c_t | \bar{A}_{t-1}, V, C_{t-1} = 0)}{P(C_t = c_t | \bar{L}_t, \bar{A}_{t-1}, V, C_{t-1} = 0)} & c_1 = \dots = c_{t-1} = 0, c_t = 0, 1 \\ 0 & \text{otherwise} \end{cases}.$$

These weights are then multiplied to create a single weight for each individual at each observation interval: in the non-zero case, we have

$$\begin{aligned} w_j &= w_j^e w_j^c \\ &= \prod_{t=1}^j \frac{P(A_t = a_t | \bar{A}_{t-1}, V, C_t = 0) P(C_t = c_t | \bar{A}_{t-1}, V, C_{t-1} = 0)}{P(A_t = a_t | \bar{L}_t, \bar{A}_{t-1}, V, C_t = 0) P(C_t = c_t | \bar{L}_t, \bar{A}_{t-1}, V, C_{t-1} = 0)}. \end{aligned} \quad (5)$$

Weights are often modelled parametrically, with parameters α , and estimated via logistic or multinomial regression. The weights are used to modify the multinomial likelihood, creating a weighted likelihood

$$L^{\text{IPW}}(\mathbf{H}; \beta, \alpha) = \prod_{i=1}^N \prod_{j=1}^{n_i} \prod_{k=0}^K \exp \{w_{ij}(\alpha) \mathbb{I}[D_{ij} = k] \log h_{ijk}(\beta)\}, \quad (6)$$

which is then maximized to find estimates of the parameters β used to model the dependence of the hazard on the exposure. The effect of reweighting has been described as creating a pseudopopulation in which the exposure and the outcome are no longer confounded [7]; heuristically, the maximization of the weighted likelihood may be viewed as a regression carried on the individuals in this pseudopopulation.

4.3. Variance estimation

The asymptotic standard errors for the inverse-probability weighted likelihood estimator of the MO-MSM described above requires a correction to the usual multinomial model standard errors (which we refer to as the uncorrected standard errors) to account for estimation of parameters in the MSM weights.

Let $U(\beta, \alpha)$ be the estimating function based on the weighted likelihood $L^{\text{IPW}}(\mathbf{H}; \beta, \alpha)$ from (6), where β are the (p_b) parameters of interest and α represents the (p_a nuisance) parameters of the exposure and censoring models used to derive the weights. Let $\varphi = (\beta, \alpha)$. The estimating function is based on the $(p_a + p_b) \times 1$ system of equations which for individual i takes the form

$$U_i(\beta, \alpha) = \frac{\partial}{\partial \varphi} \left\{ \sum_{j=1}^{n_i} \sum_{k=0}^K w_{ij}(\alpha) \mathbb{I}[D_j = k] \log h_{ijk}(\beta) \right\}.$$

We subsequently drop the dependence on i whenever possible. The derivative in $U(\beta, \alpha)$ can be broken into two blocks: the first p_b elements take the form

$$U^\beta(\beta, \alpha) = \sum_{j=1}^{n_i} \sum_{k=0}^K w_j(\alpha) \mathbb{I}[D_j = k] \frac{\partial}{\partial \beta} \log h_{jk}(\beta) \quad (7)$$

whereas the last p_a elements take the form

$$U^\alpha(\beta, \alpha) = \sum_{j=1}^{n_i} \sum_{k=0}^K \frac{\partial}{\partial \alpha} w_j(\alpha) \mathbb{I}[D_j = k] \log h_{jk}(\beta) \quad (8)$$

with the obvious extension to the second and cross partial derivatives, $U^{\beta\beta}$, $U^{\alpha\beta}$ and $U^{\alpha\alpha}$.

In practice, the estimation is performed in two stages, first estimating α by $\hat{\alpha}$ using a separate model, and then estimating β using $U(\beta, \hat{\alpha})$. We may derive the variance of the resulting estimator $\hat{\beta}$ by performing a first order Taylor expansion of U about the probability limit, α^* , of $\hat{\alpha}$. Consider the first block of U , U^β ; with β fixed, we have the expansion

$$\begin{aligned} U_{\text{adj}}^\beta(\beta, \hat{\alpha}) &= U^\beta(\beta, \alpha^*) + \mathbb{E} [U^{\beta\alpha}(\beta, \alpha^*)] (\hat{\alpha} - \alpha^*) + o_p(1) \\ &= U^\beta(\beta, \alpha^*) - \mathbb{E} [U^{\beta\alpha}(\beta, \alpha^*)] (\mathbb{E} [\ell_\alpha^{\alpha\alpha}(\alpha^*)])^{-1} \ell_\alpha^\alpha(\alpha^*) + o_p(1) \end{aligned}$$

where ℓ_α^α is the score equation for the nuisance (exposure and censoring models), and $\ell_\alpha^{\alpha\alpha}$ is the matrix of derivatives of the score function. Then the estimating function has asymptotic variance

$$\mathbb{E}[U_{\text{adj}}^\beta(\beta, \alpha^*)^{\otimes 2}] = \mathbb{E}[U_{\text{adj}}^\beta(\beta, \alpha^*) U_{\text{adj}}^\beta(\beta, \alpha^*)^\top].$$

It follows that the asymptotic variance of the hazard linear predictor parameters $\hat{\beta}$ is

$$\text{Var}[\hat{\beta}] = \mathbb{E} \left[\left\{ \left(\mathbb{E} \left[\frac{\partial}{\partial \beta} U_{\text{adj}}^\beta(\beta, \alpha^*) \right] \right)^{-1} U_{\text{adj}}^\beta(\beta, \alpha^*) \right\}^{\otimes 2} \right],$$

which we term the corrected (sandwich) variance, and can be estimated by replacing α^* by $\hat{\alpha}$, and the expectations by expectations with respect to the empirical measure, evaluated at $(\hat{\beta}, \hat{\alpha})$. In contrast, note that the (asymptotic) uncorrected variance is obtained from

$$\mathbb{E} \left[\left\{ \left(\mathbb{E} \left[\frac{\partial}{\partial \beta} U(\beta, \alpha^*) \right] \right)^{-1} U(\beta, \alpha^*) \right\}^{\otimes 2} \right].$$

Further details are given in the Supplementary Material, §S2.

4.4. Estimating the cumulative incidence function

Some authors note that, in the continuous time case, cause-specific hazards cannot be readily interpreted in terms of the cumulative incidence function (CIF). For example, Beyersmann et al. [20] write

A Cox, i.e. proportional hazards analysis of all cause-specific hazards to study [...] risk factors, is feasible, but notoriously difficult to interpret in terms of the CIF: The CIF is an involved function of all cause-specific hazards, and a risk factor may have quite different effects on the different cause-specific hazards. This is different from the usual survival situation, where hazard-based analyses are directly interpretable in terms of the distribution function due to the one-to-one relationship between hazard and distribution function.

As we have seen, however, in the discrete case, equations (3) and (4) give explicit forms for the CIF derived from the collection of cause-specific hazards, and so an estimate of the CIF can be obtained by substituting in the estimates $\hat{h}_{jk}$ for h_{jk} . This implies that the influence of any confounder on one of the cause-specific hazards can be directly used in a straightforward fashion to study the influence on the CIF using plug-in procedures. Pointwise asymptotic variance estimates can then be computed using standard approximation methods (see Supplementary Material, §S2). In the continuous case, Cheng et al. [21] provide an analytic means of constructing the CIF with confidence bounds; the discrete case is largely similar.

Alternatively, simulation-based approaches can be used, in particular, Monte Carlo simulation of hypothetical new cohorts of subjects with given covariate settings under different exposure profiles, using point estimates of parameters estimated from the observed data. That is, beginning with a baseline covariate profile, data can be generated up to censoring or death using the parameter estimates to generate death status. For approximate variances, the Monte Carlo approach would draw samples from the asymptotic Normal approximation to the distribution of $\hat{\beta}$, and then compute the estimated CIF for each sample. Code to perform this Monte Carlo calculation is included in the Supplementary Materials, §S4.

5. Simulation

We investigated the performance of the proposed procedure in several simulations, with different sample sizes, parameter settings and numbers of failure types. We present the results for two failure causes; more extensive results for three causes of failure can be found in §S4 of the Supplementary Material.

Suppose there are N individuals followed for up to $J = 10$ intervals, with $K = 2$ possible causes of failure. We extend the data-generating approach of Young and colleagues [13], who developed a method of simulating data from a structural accelerated failure time model by generation of counterfactual survival times, and modifying these by exposures and covariates; see Appendix A for theoretical justification of the method. Note that this algorithm is essentially the same as the algorithm outlined in [22], albeit adjusted our the causal setting. The Young algorithm can be implemented in two ways that we outline below. Note that these algorithms simulate data in continuous time, which we then discretize to yield discrete time data. Young's algorithm (equivalent to simulation by the probability integral transform or cdf inversion method, as is detailed in the Appendix) can be easily converted to directly generate discrete failure times, but we omit details here, as the principles are identical to the continuous time version and results are very similar when the event rate is low.

Let Y denote the failure time, D_j the indicator for status in interval j (0 =alive or k =death due to cause k , $k \in \{1, \dots, K\}$). Further, let A_j denote a binary treatment during the interval j and L_j denote the time-varying confounder measured at the start of interval j . $Y(0)$ denotes the counterfactual survival time under the regime 'never exposed.' Data may then be then generated using either of the following algorithms:

1. **The total hazard approach:** In a continuous time model and in a cause-specific competing risk setting, the total hazard at time t , $\lambda(t)$, is given by

$$\lambda(t) = \sum_{k=1}^K \lambda_k(t)$$

where $\lambda_1(t), \dots, \lambda_K(t)$ are the cause-specific hazards. In our setting of piecewise constant hazards, $\lambda_k(t; A_j) = \exp\{\psi_k A_j\}$ for $j-1 \leq t < j$, so

$$\lambda(t; A_j) = \sum_{k=1}^K \exp\{\psi_k A_j\} \quad j-1 \leq t < j.$$

- (a) Generate $Y(0)$ from an Exponential distribution with $\lambda = 0.01$.

(b) Define $L_0 = A_0 = D_0 = 0$. Then for each $j \in \{1, \dots, 10\}$, repeat the following:

- Generate the confounder L_j from a binomial distribution with

$$\text{logit}[P(L_j = 1 | \bar{L}_{j-1}, \bar{A}_{j-1}, D_j = 0)] = \beta_0 + \beta_1 I[Y(0) < c] + \beta_2 A_{j-1} + \beta_3 L_{j-1}$$

where $c = 30$.

- Generate the treatment A_j from a binomial distribution with

$$\text{logit}[P(A_j = 1 | \bar{L}_{j-1}, \bar{A}_{j-1}, D_j = 0)] = \alpha_0 + \alpha_1 L_j + \alpha_2 L_{j-1} + \alpha_3 A_{j-1}.$$

- Generate D_{j+1} , and possibly Y , using the rules:

- If $Y(0) > \int_0^{j+1} \lambda(s; \bar{A}_{j+1}) ds$, then $D_{j+1} = 0$,
- Otherwise, if $Y(0) \leq \int_0^{j+1} \lambda(s; \bar{A}_{j+1}) ds$, the Y lies in interval $j + 1$ with

$$Y := j + \frac{1}{\lambda(t; A_j)} \left(Y(0) - \int_0^j \lambda(s; \bar{A}_{j+1}) ds \right) \quad j \leq t < j + 1$$

and then generate D_{j+1} from the discrete distribution on $\{1, 2, \dots, K\}$, with

$$P[D_{j+1} = k] = \frac{\lambda_k(Y; A_j)}{\lambda(Y; A_j)} \quad j \leq t < j + 1.$$

Thus, the counterfactual cause-specific failure-time model is generated from the inhomogeneous process with piecewise constant cause-specific hazards. Note that the choice is c in step (b) dictates the degree of variability in the indicator $I[Y(0) < c]$, which in turn affects the degree to which the counterfactual exposure-free survival affects the time-varying variable across the sample; with $c = 30$, $E\{I[Y(0) < c]\} = 0.26$. R code is provided in §S4 of the Supplementary Material.

2. **The mixture approach:** The data-generating algorithm can be applied directly using a mixture of distributions, applying the algorithm for each cause of death $k = 1, \dots, K$ on fractions $\pi_1, \dots, \pi_K$ of the sample where $\sum_{k=1}^K \pi_k = 1$. The latter approach is simply a fixed-probability mixture of distributions, and can be conceptualized as encoding an assumption that individuals in the population have an unknown and unmeasured susceptibility to dying by one of the causes of death. In the continuous time setting, the cause-specific subdistribution functions $F_k(t)$ can be written

$$F_k(t) = P[Y \leq t, D = k] = P[Y \leq t | D = k] P[D = k]$$

so simulating from the cause-specific model reduces to generating D , and then Y conditional on D . A similar example of a competing risks formulation via a mixture model is given in [14, Example 5.4].

We begin by demonstrating the comparability of the proposed approach and the unweighted multinomial model of Section 4.1 to a conventional competing risks approach, the Cox proportional hazards function using multiple outcomes and time-varying covariates; see Sections 3.7.1 and 8.4 of the book by Therneau and Grambsch [23]. Table 1 presents the results from a no-confounding scenario, where exposures were generated as though in a randomized trial (i.e. where $\text{logit}[P(A_j = 1 | \bar{L}_{j-1}, \bar{A}_{j-1}, D_j = 0)] = \text{logit}[P(A_j = 1)] = \alpha_0$) using the competing risks data-generation approach outlined above. Three fitted models were compared: (i) the extended Cox model, (ii) an unweighted multinomial regression, and (iii) the MO-MSM, estimated via IPW in a multinomial model using stabilized weights. We considered bias, variance, and mean squared error (MSE) for a sample size of 1000, with 5000 independent random samples. All three models yield virtually identical estimates and exhibited the same variability, indicating that the multinomial modelling approach (weighted or unweighted) is a reasonable modelling choice in the absence of time-varying confounding, and provides good estimates of the hazard ratio from a Cox model when the event rate is low.

Results in Table 2 present scenarios with time-varying confounding. In these simulations, parameters controlling exposure were $\alpha = (\log(2/7), 0.5, 0.5, \log(4))$, indicating a strong serial dependence of current treatment on past treatment (odds ratio of 4), and a more modest dependence on the time-varying confounder (odds ratios of 1.6 for current and past values of the confounder). The parameters controlling the time-varying confounder were set to $\beta = (\log(3/7), 2, \log(0.5), \log(1.5))$, so that the value depended strongly on the latent, exposure-free survival time and moderately on previous exposure and the value of the confounder in the last interval. Again, three models were fit: (i) an unweighted multinomial regression which does not adjust for the time-varying confounder, (ii) an unweighted multinomial regression which adjusts for the time-varying confounder by including it as a linear term in the regression model, and (iii) the MO-MSM. In these simulations, the mixture approach to data-generation was used; generating the data using the

Table 1. Simulation results from fitting the extended Cox model (Cox), a multinomial regression model (Multinom), and the multiple outcome marginal structural model (MO-MSM) for $n = 1000$ and various treatment effects in a no-confounding scenario: two causes of death.

	Cause of death 1			Cause of death 2		
	Cox	Multinom	MO-MSM	Cox	Multinom	MO-MSM
$(\psi_1, \psi_2) = (-0.1, 0.4)$						
$\hat{\psi}$	-0.098	-0.097	-0.098	0.415	0.416	0.415
$SD(\hat{\psi})$	0.301	0.301	0.301	0.268	0.268	0.267
Bias	0.002	0.003	0.002	0.015	0.016	0.015
MSE	0.090	0.091	0.091	0.072	0.072	0.072
$(\psi_1, \psi_2) = (0.3, 0.3)$						
$\hat{\psi}$	0.305	0.307	0.307	0.298	0.300	0.299
$SD(\hat{\psi})$	0.280	0.281	0.281	0.290	0.290	0.289
Bias	0.005	0.007	0.007	-0.002	0.000	-0.001
MSE	0.079	0.079	0.079	0.084	0.084	0.083
$(\psi_1, \psi_2) = (0, 0)$						
$\hat{\psi}$	0.005	0.005	0.004	0.004	0.004	0.003
$SD(\hat{\psi})$	0.300	0.300	0.299	0.301	0.302	0.302
Bias	0.005	0.005	0.004	0.004	0.004	0.003
MSE	0.090	0.090	0.089	0.091	0.091	0.091

total hazard approach produced the same conclusions (results not shown). We considered a range of parameters (ψ_1, ψ_2) , allowing for no effect of exposure on the outcome, identical effects, and effects of opposite signs.

The simulation results clearly demonstrate the bias of conventional model estimators in the presence of time-varying confounding, and the unbiasedness of the MO-MSM estimators. Of particular note is the considerable bias of the conventional models under the null setting of no treatment effect ($\psi_1 = \psi_2 = 0$). In the absence of time-varying confounding, the MO-MSM performed as well as Cox regression and the unweighted multinomial model. Thus, it appears that while MO-MSM estimators sometimes exhibit greater variability than the conventional model estimators, they are the preferred choice in moderate to large samples as evidenced by the small MSEs, particularly if time-varying confounding is present. Even in small samples, these estimators may be preferable due to the small bias, despite the larger MSE.

The distribution of weights used to estimate the marginal structural model are more variable in scenarios with no time-varying confounding than in those with. We additionally considered a range of value for $\pi_1, \pi_2 = 1 - \pi_1$; results are shown for the case of equal probabilities only as conclusions from the simulations with different mixing probabilities were the same. Additional simulations can be found in the Supplementary Material, §S5.

6. Example: The impact of IDU on mortality in an HIV/HCV co-infected cohort

We illustrate the methods using data from the multi-site prospective Canadian HIV/HCV Co-infection Cohort study. After providing informed consent, cohort participants underwent an initial evaluation followed by study visits which were scheduled at six month intervals. While exact dates of death are available for cohort participants, it is common practice in the MSM literature to discretize survival time [11, 12, 24–27]. We adopt the discrete-time approach, but note that it is also possible to implement MSMs for continuous survival time [28].

At each follow-up visit, participants completed a questionnaire and routine blood tests were conducted. At the time of analysis, there were 5736 observations on 1119 individuals recorded in the database who were HIV positive and had evidence of HCV infection. Time since co-infection is defined in the following fashion: for those individuals who were initially infected with HIV, it is the probable date of HCV infection; for those first infected with HCV, it is the probable date of HIV infection. Probable date of HCV infection is given by the earliest of the date at which the participant believes HCV was contracted, the date of the first positive HCV test, and the date of first injection drug use. Probable date of HIV infection is given by the earliest of the date at which the participant believes HIV was contracted and the date of the first positive HIV test. All reported deaths were verified and classified following the Coding of Death in HIV (CoDe) system (www.cphiv.dk/CoDe). Each time a participant was reported to have died, sites completed a detailed case report form which included all information related to the death (including death certificate information, autopsy reports if available and clinical diagnoses and events immediately preceding the death). Linkage to provincial vital statistics reports was

Table 2. Simulation results from fitting two conventional multinomial models and the marginal structural cause-specific hazard model for various sample sizes and treatment effects: two causes of death. Multinom 1 is a conventional multinomial regression model which does not adjust for time-varying confounders; Multinom 2 is a conventional multinomial regression model which adjusts for time-varying confounders; MO-MSM denotes the proposed cause-specific marginal structural model.

	Cause of death 1			Cause of death 2		
	Multinom 1	Multinom 2	MO-MSM	Multinom 1	Multinom 2	MO-MSM
$n = 250, (\psi_1, \psi_2) = (-0.1, 0.4)$						
$\hat{\psi}$	0.281	0.166	-0.137	1.000	0.847	0.572
$SD(\hat{\psi})$	4.698	4.368	4.4369	2.847	2.666	2.869
Bias	0.381	0.266	-0.037	0.600	0.447	0.172
% bias	-381.0	-266.4	36.9	150.0	111.8	42.9
MSE	22.216	19.152	19.687	8.464	7.306	8.262
$n = 500, (\psi_1, \psi_2) = (-0.1, 0.4)$						
$\hat{\psi}$	0.327	0.182	-0.100	0.842	0.700	0.422
$SD(\hat{\psi})$	0.459	0.462	0.513	0.403	0.406	0.450
Bias	0.427	0.282	0.000	0.442	0.300	0.022
% bias	-427.0	-281.8	0.3	110.5	74.9	5.5
MSE	0.393	0.293	0.263	0.358	0.255	0.203
$n = 1000, (\psi_1, \psi_2) = (-0.1, 0.4)$						
$\hat{\psi}$	0.335	0.189	-0.093	0.839	0.696	0.413
$SD(\hat{\psi})$	0.303	0.305	0.336	0.280	0.281	0.312
Bias	0.435	0.289	0.007	0.439	0.296	0.013
% bias	-434.7	-289.3	-7.3	109.8	73.9	3.2
MSE	0.281	0.177	0.113	0.271	0.166	0.097
$n = 1000, (\psi_1, \psi_2) = (0.3, 0.3)$						
$\hat{\psi}$	0.738	0.594	0.308	0.746	0.603	0.316
$SD(\hat{\psi})$	0.280	0.282	0.315	0.280	0.282	0.315
Bias	0.446	0.303	0.016	0.446	0.303	0.016
% bias	146.1	98.1	2.6	148.7	101.0	5.3
MSE	0.277	0.171	0.099	0.277	0.171	0.099
$n = 1000, (\psi_1, \psi_2) = (0, 0)$						
$\hat{\psi}$	0.427	0.280	0.001	0.439	0.292	0.010
$SD(\hat{\psi})$	0.292	0.294	0.321	0.299	0.301	0.333
MSE	0.267	0.164	0.103	0.282	0.176	0.111

performed in British Columbia, Alberta and Quebec and used to supplement data obtained in the case report forms and determine if any participants who had been lost to follow-up had died. The final determination of cause of death was made independently by two investigators and in the (two) cases where there were discrepancies, resolved by a third investigator.

6.1. Details of the analysis

Given the relatively small number of events, only a limited number of potential confounders could be safely included in outcome models. Preliminary modelling suggested that fitting a common baseline hazard across all causes of death was not appropriate. Three models were fit: an unweighted multinomial model adjusted only for baseline covariates, an unweighted multinomial model which included time-varying covariates, and a multiple outcome marginal structural model.

We fit complex treatment models in an effort to achieve the best possible balance of confounders between person-intervals in which IDU was and was not reported. Previous use of injection drugs is a very strong determinant of current use. For example, in the CCC, 92% of the person-intervals in which a participant reported no IDU in the previous interval were followed by another interval of no IDU; where IDU was reported, 67% reported continued IDU in the next interval. The models for the probability of IDU in a six-month period, and the probability of remaining in the study (i.e. for not being censored) used to construct the weights in the MO-MSM included the following covariates: age, male sex, IDU in the previous six month interval, time-varying CD4+ T-cell count (log cells/uL), time-varying HIV RNA (log copies/ml),

time-varying ART interruption (defined as the cessation of all anti-retrovirals for at least 14 days over the last six months), time-varying AST-to-platelet ratio index (APRI) score (a surrogate marker of liver fibrosis), time since co-infection, and cohort visit number. Age, log transformed CD4, log transformed viral load, APRI score, and time since co-infection were parameterized flexibly, using three-knot splines, and interactions between previous IDU and each of male sex and ART interruption were also included in the models. Weights were standardized using models that included age and time since co-infection (each with three-knot splines), visit as well as male sex, IDU in the previous six months, and their interaction.

It is clinically plausible that the impact of IDU on mortality from ESLD and possibly cancer is the result of prolonged rather than acute exposure. Hence we chose to examine the impact of three different measures of IDU exposure: (i) cumulative number of intervals of use while in the cohort; (ii) cumulative number of intervals of use in the previous 12 months (i.e. 0 for no use, 1 for IDU in one six-month period, 2 for use in both of the preceding six month periods); and (iii) any IDU in the past six months. All outcome models included baseline covariates as well as the exposure variable. The second of the unweighted multinomial models additionally adjusted for the time-varying variables CD4, viral load, ART interruptions, and APRI. The multinomial and the multiple outcome marginal structural model assumed a proportional odds parameterization.

There were 7016 scheduled visits for the 1119 individuals included in the analysis. However, nearly half of all participants (519, or 46%) missed at least one scheduled visit, so that the total number of visits observed was 5736 (82%). The missing data (both item- and visit-missingness) were handled via multiple imputation [29] using *mice* (Multivariate Imputation by Chained Equations) with default settings (e.g. using predictive mean matching) in R v. 2.13.0 [30], with the following covariates as predictors: visit, ever injected drugs, active Hepatitis B infection, nadir CD4 cell count, IDU in the last six months, CD4 cell count, age at cohort entry, current and baseline APRI, HIV viral load, time since probable date of HIV infection, time since co-infection, male gender, whether ART was interrupted in the last six months, and an indicator for subject to account for the correlation of measures made on the same individuals. Twenty-five imputed datasets were analyzed.

In this analysis, the time-origin was taken to be the time of co-infection. While many clinicians consider time on study to be a relevant origin, as it allows examination of the factors affecting mortality after entry into care amongst those who seek it, cohort entry is not in general a meaningful or useful origin. The time from co-infection to cohort entry is in many cases long (several years). We treat this pre-cohort time as a single interval. That is, we set time 0 to be occurrence of co-infection and time 1 to be cohort entry, so that the first interval in our analysis runs from time of co-infection to time of cohort entry. For the 210 participants who reported never injecting drugs, we deterministically impute a value of 0 (no IDU) for the pre-cohort interval. For subjects who report IDU prior to cohort entry, the exposure in the pre-cohort interval is multiply imputed with probability corresponding to the fraction of intervals in the cohort study in which injection drugs were used. While this approach is somewhat crude, there are, to date, no methods that have attempted to address the issue of exposures which have occurred between the true (meaningful) origin and the start of follow-up. A more sophisticated treatment of this is an interesting and difficult challenge that is outside the scope of this illustrative example.

6.2. Results and Comments

In total, 97 deaths were recorded: 11 of cancer, 25 of ESLD, 21 by overdose and a remaining 40 from other (26) or unknown (14) causes. Two, five, and ten of the cancer, ESLD, and overdose deaths, respectively, occurred in conjunction with the use of injection drugs in both the past six months and 6-12 months previously; a further three, five, and six deaths, respectively, were associated with IDU in one of the two previous six-month intervals.

Modelling results are presented in Table 3. Unsurprisingly, all three modeling approaches revealed a strong and statistically significant association between IDU (in any of the three exposure parameterizations) and death by overdose; the point estimates are generally similar across the models considered, but varied according to the definition of the exposure for death by overdose, where the impact of recent IDU was much more predictive than cumulative use. None of the models suggest a significant relationship effect of IDU and either cancer or ESLD. Estimation via a Cox model extended to handle multiple causes of death and time-varying covariates produced findings similar in magnitude to the unweighted multinomial model with time-dependent covariates (results not shown): for example, the hazard ratio (95% CI) for death by cancer, ESLD, and OD associated with IDU in the last six months are, respectively, 1.31 (0.43, 3.95), 1.19 (0.53, 2.67), and 5.16 (2.28, 11.66).

Figure 1 demonstrates the use of equation (3) to construct CIFs using the collection of cause-specific hazard estimates from the MO-MSM to contrast the cumulative risk of death by overdose from the model in which the hazard depends only on IDU in the last six months. Following the variance calculations presented in §S3 of the Supplementary Material, confidence intervals for the CIF were derived which fully accounted for the variability of the inverse weighting in the marginal structural model (see §S2 of the Supplementary Material). A priori, we decided to adjust models for the interval length, to account for the typically long pre-cohort interval. Including interval length lead to very wide confidence intervals for the CIF. Omitting this variable produced more narrower intervals, with relatively modest changes to the HR estimates.

Table 3. Canadian HIV/HCV Co-infection Cohort study: the hazard ratio (95% confidence interval) of mortality from cohort entry by various causes attributable to IDU. The models fit are: a conventional multinomial regression model which does not adjust for time-varying confounders (Multinom 1); a conventional multinomial regression model which adjusts for time-varying confounders (Multinom 2); and the proposed cause-specific marginal structural model (MO-MSM).

Outcome	Multinom 1	Multinom 2	MO-MSM
Cumulative use			
Cancer	1.02 (0.76, 1.38)	0.97 (0.70, 1.35)	0.87 (0.67, 1.13)
ESLD	1.01 (0.85, 1.21)	1.01 (0.84, 1.22)	0.95 (0.81, 1.12)
OD	1.38 (1.08, 1.75)	1.40 (1.09, 1.80)	1.18 (0.97, 1.44)
Cumulative use in the past 12 months			
Cancer	1.28 (0.63, 2.62)	1.06 (0.48, 2.35)	0.88 (0.30, 2.56)
ESLD	1.17 (0.73, 1.87)	1.14 (0.70, 1.84)	0.89 (0.51, 1.55)
OD	2.41 (1.46, 3.99)	2.52 (1.50, 4.25)	1.94 (1.04, 3.62)
Use the last six months			
Cancer	1.77 (0.57, 5.50)	1.31 (0.38, 4.49)	1.08 (0.18, 6.42)
ESLD	1.20 (0.54, 2.66)	1.14 (0.50, 2.59)	0.87 (0.32, 2.35)
OD	4.77 (2.12, 10.75)	5.04 (2.20, 11.57)	4.03 (1.59, 10.22)

For example, the impact of IDU in the last six months in the MO-MSM that includes interval length is 4.03 (1.59, 10.22), compared to the MO-MSM in which this variable is omitted is 4.27 (1.73, 10.54). Using the latter model, we found, for example, that for a 45-year old man who acquired HCV 20 years ago, the cumulative probability (95% CI) of dying of OD within two years of joining the cohort was 0.0116 (0.0015, 0.0856) among those who never used intravenous drugs, as compared to 0.0485 (0.0065, 0.2852) for someone who used intravenous drugs at least once in each six month period after joining the cohort.

While we have attempted to account for time (and exposure) prior to cohort entry, we have not accounted for truncation of those HIV-HCV co-infected Canadians who may have died prior to cohort entry. While ESLD may take many years to develop, the same is not true of death by overdose or even cancer. In fact, there is some evidence to suggest that chronic ‘stable’ users of injection drugs are at a lower risk of overdose mortality than individuals who stop and re-start injection drug use. It is therefore likely that our results underestimate the impact of IDU on death by overdose. The impact on cancer is more difficult to surmise.

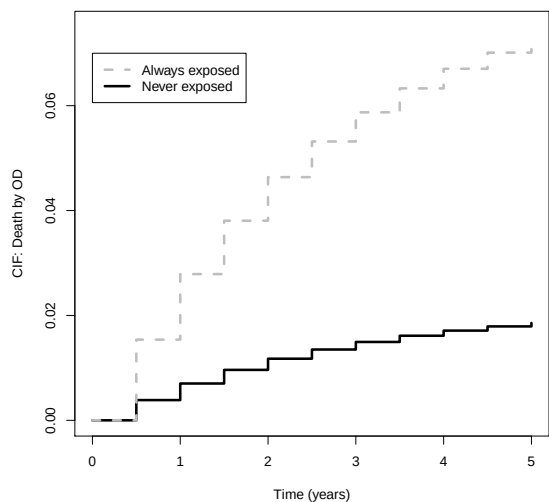


Figure 1. Cumulative Incidence Functions: Death by overdose amongst 45 year old males who always, or never, use intravenous drugs based on the model for IDU exposure in the previous six months.

7. Discussion

The causal inferential approach of using marginal structural models allows unbiased estimation of the hazard ratios of the cause-specific hazards in a competing risks setting. We have adapted the marginal modelling approach, which (conditional on several key assumptions being satisfied), attempts to answer the question that would be addressed in the (possibly infeasible or unethical) randomized trial that studies specific patterns of a time-varying exposure on each of several causes of death. The MO-MSM approach outlined above permits unbiased estimation of each of the cause-specific hazard ratios. We have shown how the predicted all-cause mortality cumulative incidence function can be estimated using estimated cause-specific hazards and an assumption of an constant hazard exposure-free failure time distribution.

In addition to presenting the first analysis of the effect of recent drug use on mortality in the Canadian Co-infection Cohort, we have detailed a general framework for the implementation of a methodology that is valid for both randomized and observational data. This approach provides a natural and necessary analytic tool to handle time-varying confounding and mediation in a multiple outcome setting. Additionally, we provided a general approach for simulating continuous-time failure data with multiple-outcomes from a marginal structural model.

A Simulation of survival times

A.1 Simulation via the Probability Integral Transform

Notation: For the simulation of a survival time for a single cause of failure according to a piecewise constant (time-varying) hazard $h(t)$, consider survivor function $S(t)$. Suppose m is defined by $m = \lfloor t \rfloor$, so that $m \leq t < m + 1$. We have

$$S(t) = \mathbb{P}[T > t] = \exp \left\{ - \int_0^t h(s) ds \right\} = \begin{cases} \exp \left\{ - \int_0^t h(s) ds \right\} & m = 0 \\ \exp \left\{ - \left[\sum_{j=0}^{m-1} \int_j^{j+1} h(s) ds + \int_m^t h(s) ds \right] \right\} & m \geq 1 \end{cases}$$

For interval j corresponding to $[j, j + 1)$, suppose $h(t) = \lambda \exp\{\psi_j x_j\}$ say,

$$\int_j^{j+1} h(s) ds = \lambda \exp\{\psi_j x_j\} \quad \& \quad \int_j^t h(s) ds = \lambda \exp\{\psi_j x_j\}(t - j).$$

Thus, for $m \leq t < m + 1$

$$S(t) = \exp \left\{ -\lambda \left[\sum_{j=0}^{m-1} \exp\{\psi_j x_j\} + \exp\{\psi_m x_m\}(t - m) \right] \right\} \quad (9)$$

where the summation is zero if $m = 0$. As $S(0) = 1$, we have for $m = 1, 2, \dots$,

$$S(m) = \exp \{ -\lambda \exp\{\psi_{m-1} x_{m-1}\} \} S(m-1)$$

and for $m \leq t < m + 1$

$$S(t) = \exp \{ -\lambda \exp\{\psi_m x_m\}(t - m) \} S(m)$$

Algorithm: Simulate T using the **Probability Integral Transform**: if $U \sim \text{Uniform}(0, 1)$, and $F_X(\cdot)$ is a cdf, then $X = F_X^{-1}(U)$ is distributed with cdf F_X . As $1 - U \sim \text{Uniform}(0, 1)$, this means that we can also simulate X by solving $S_X(X) = U$, where $S_X(x) = 1 - F_X(x)$.

1. Generate $u \sim \text{Uniform}(0, 1)$; set $s = 1$, $m = 0$, $S(0) = 1$.
2. While $u < s$ do
 - (a) set $S(m+1) := \exp \{ -\lambda \exp\{\psi_m x_m\} \} s$; set $s := S(m+1)$; set $Y_{m+1} = 0$
 - (b) if $u < s$ set $m := m + 1$
3. set $Y_{m+1} = 1$.
4. set $t = m - \exp\{ -\psi_m x_m \} \log(u/S(m))/\lambda$

This process generates a survival time from the inhomogeneous process with piecewise constant hazards $\lambda \exp\{\psi_m x_m\}$, $m = 0, 1, 2, \dots$

A.2 Young's Algorithm

Young's algorithm to simulate the potential survival outcome proceeds as follows:

1. Generate $T_0 \sim \text{Exp}(\lambda)$
2. For $m = 0, 1, 2, \dots$
 - (a) Generate A_m according to the treatment assignment mechanism.
 - (b) If $T_0 > \int_0^{m+1} \exp\{\psi A_j\} dj$, set $Y_{m+1} = 0$,
else if $T_0 \leq \int_0^{m+1} \exp\{\psi A_j\} dj$, set $Y_{m+1} = 1$ and

$$T = m + \left(T_0 - \int_0^m \exp\{\psi A_j\} dj \right) \exp\{-\psi A_m\} \quad (10)$$

In this formulation, T_0 corresponds to the untreated potential survival time, arising from a constant hazard, homogeneous process. Under treatment pattern $(A_1, A_2, \dots, A_m, \dots)$, the potential survival time arises from the inhomogeneous process with piecewise constant hazards $\lambda \exp\{\psi A_m\}$, $m = 1, 2, \dots$. If $\psi = 0$, or $A_m = 0$ for all m , the potential outcomes have the same distribution.

To see the equivalence with the Probability Integral Transform algorithm, suppose that $\psi_m \equiv \psi$ for all m , and note that

$$\int_0^m \exp\{\psi A_j\} dj \equiv \sum_{j=0}^{m-1} \exp\{\psi A_j\}.$$

Then, from (10), for $m < t \leq m + 1$

$$\begin{aligned} \mathbb{P}[T > t] &= \mathbb{P}\left[m + \left(T_0 - \int_0^m \exp\{\psi A_j\} dj\right) \exp\{-\psi A_m\} > t\right] \\ &= \mathbb{P}\left[T_0 > \int_0^m \exp\{\psi A_j\} dj + \exp\{\psi A_m\}(t - m)\right] \\ &= \exp\left\{-\lambda \left[\int_0^m \exp\{\psi A_j\} dj + \exp\{\psi A_m\}(t - m)\right]\right\} \\ &= \exp\left\{-\lambda \left[\sum_{j=0}^{m-1} \exp\{\psi A_j\} + \exp\{\psi A_m\}(t - m)\right]\right\} \end{aligned}$$

This is identical to (9). Therefore to generate counterfactual failure times reduces to probability integral transform (cdf inversion) simulation. Note that the effect of time-varying (non confounding) covariates can be introduced via the parameter λ .

The above (Probability Integral Transform/Young) algorithm can be used K times to create a mixture of types of death, as discussed on page 91 of Crowder (2001), or can be adapted so as to allow for K death types, as shown in the R code in the Supplementary Material, §S4.

Acknowledgement

Drs. Moodie and Stephens acknowledge support of the Natural Sciences and Engineering Research Council (NSERC) of Canada. Dr. Moodie is also supported by a Chercheur-Boursier career award from the Fonds de recherche en santé du Québec (FRSQ). Dr. Klein is supported by a Chercheur-National career award from the FRSQ.

The Canadian HIV/HCV Co-infection Cohort was funded Réseau SIDA/maladies infectieuses of the FRSQ, the Canadian Institutes of Health Research (CIHR, MOP-79529) and the CIHR Canadian HIV Trials Network (CTN222).

References

1. 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.

2. Keiding N, Filiberti M, Esbjerg S, Robins JM, Jacobsen N. The graft versus leukemia effect after bone marrow transplantation: A case study using structural nested failure time models. *Biometrics* 1999; **55**:23–28.
3. Vansteelandt S, Mertens K, Suetens C, Goetghebeur E. Marginal structural models for partial exposure regimes. *Biostatistics* 2009; **10**:46–59.
4. Bekaert M, Vansteelandt S, Mertens K. Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. *Lifetime Data Analysis* 2010; **16**:45–70.
5. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd edn. Wiley, New York, 2002.
6. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, Halloran E, Berry D, eds. Springer, New York, 1999; 95–134.
7. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**:550–60.
8. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**:561–70.
9. Lucas GM, Cheever LW, Chaisson RE, Moore RD. Detrimental effects of continued illicit drug use on the treatment of HIV-1 infection. *Journal of Acquired Immune Deficiency Syndromes* 2001; **27**:251–259.
10. Klein MB, Saeed S, Yang H, Cohen J, Conway B, Cooper C, Coté P, Cox J, Gill J, Haase D, Haider S, Montaner J, Pick N, Rachlis A, Rouleau D, Sandre R, Tyndall M, Walmsley S. Cohort profile: The Canadian HIV-Hepatitis C Co-infection Cohort study. *International Journal of Epidemiology* 2010; **39**:1162–1169.
11. Bryan J, Yu Z, van der Laan MJ. Analysis of longitudinal marginal structural models. *Biostatistics* 2004; **5**:361–380.
12. Westreich D, MacPhail P, Van Rie A, Malope-Kgokong B, Ive P, Rubel D, Boulmé R, Eron J, Sanne I. Effect of pulmonary tuberculosis on mortality in patients receiving HAART. *AIDS* 2009; **23**(6):707–715.
13. Young J, Hernán M, Picciotto S, Robins J. Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis* 2009; **16**(1):71–84.
14. Crowder M. *Classical Competing Risks*. Chapman & Hall/CRC, 2001.
15. Greenland S. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* 2003; **14**:300–306.
16. Robins JM. Causal inference from complex longitudinal data. In *Latent Variable Modeling and Applications to Causality*, Berkane M, ed. Springer, New York, 1997; 69–117.
17. Moore KL, Neugebauer RS, van der Laan MJ, Tager IB. Causal inference in epidemiological studies with strong confounding. Tech. rep., U.C. Berkeley Division of Biostatistics Working Paper Series, 2009.
18. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent e-health interventions. *American Journal of Preventive Medicine* 2007; **32**(5S):S112–118.
19. Jewell NP. Correspondences between regression models for complex binary outcomes and those for structured multivariate survival analyses. In *Advances in Statistical Modeling and Inference*, Nair V, ed. World Scientific Publishing Co Pte Ltd, Singapore, 2007; 45–64.
20. Beyersmann J, Dettenkofer M, Bertz H, Schumacher M. A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards. *Statistics in Medicine* 2007; **26**.
21. Cheng SC, Fine JP, Wei LJ. Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* 1998; **54**:219–228.
22. Beyersmann J, Latouche AL, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Statistics in Medicine* 2009; **28**(6):956–971. URL <http://dx.doi.org/10.1002/sim.3516>.

23. Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlang, New York, 2000.
24. Choi H, Hernán M, Seeger J, Robins J, Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *The Lancet* 2002; **359**(9313):1173–1177.
25. Cook N, Cole S, Hennekens C. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physicians' Health Study. *American Journal of Epidemiology* 2002; **155**(11):1045–1053.
26. Cole S, Hernán M, Robins J, Anastos K, Chmiel J, Detels R, Ervin C, Feldman J, Greenblatt R, Kingsley L, et al. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *American Journal of Epidemiology* 2003; **158**(7):687–694.
27. Sterne J, Hernán M, Ledergerber B, Tilling K, Weber R, Sendi P, Rickenbach M, Robins J, Egger M. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *The Lancet* 2005; **366**(9483):378–384.
28. Xiao Y, Abrahamowicz M, Moodie E. Accuracy of conventional and marginal structural Cox model estimators: A simulation study. *The International Journal of Biostatistics* 2010; **6**. Article 11.
29. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc., New York, 1987.
30. Van Buuren S, Oudshoorn CGM. Multivariate imputation by chained equations: Mice v1.0 user's manual, report pg/vgz/00.038. Tech. rep., TNO Prevention and Health, Leiden, 2000.

A marginal structural model for multiple outcome data: Assessing the impact of intravenous drug use on several causes of death in the Canadian Co-infection Cohort

Supplementary Material

Erica E. M. Moodie

David A. Stephens

Marina B. Klein

October 25, 2013

S1 Competing risks and multiple outcome survival analysis

Terminology and interpretation: There is, it seems, much confusion and debate in the literature concerning terminology (see, for example, [1–4] for an extended discussion, and [5] for a review of terminological variations). Our view coincides most closely with that of, for example, [5], where a multi-state counting process view is adopted. The multi-state view, as detailed in, for example [6], proposes a continuous time state process, $\{X(t)\}$, taking values on $\{0, 1, \dots, K\}$, with $X(0) = 0$, and transitions from state 0 into state k occurring at time-inhomogenous rate $\alpha_k(t)$, for $k = 1, \dots, K$. This model for the state-of-nature of a given individual is, essentially, a continuous time version of our discrete model; $\alpha_k(t)$ is the time-varying, cause-specific hazard at time t , whose value may be influenced by treatment and other covariates observed up to time t . Our model essentially assumes that $\alpha_k(t)$ is piecewise constant in t across discrete time intervals.

Modelling via the subdistribution hazard: The terminology and definitions used in Section 3.2 follows precisely those detailed in [7, Chapter 8] albeit translated to the discrete case. However the term *cause-specific hazard* has also be used in relation to the hazard defined directly from the cause- k failure probabilities from equation (2) in the main text, with all other failure types regarded as possible causes of censoring. Such a hazard might be considered if cause k was the principal focus of the analysis. Dropping counterfactual notation for simplicity, recall that the cause-specific hazard probability for cause k at time j is given by

$$h_{jk} = P[Y = j, D = k | Y \geq j] = \frac{P[Y = j, D = k]}{P[Y \geq j]} = \frac{p_{jk}}{\sum_{s=1}^K \sum_{l=j}^{\infty} p_{ls}}. \quad (\text{A.1})$$

In contrast with this, consider

$$\tilde{h}_{jk} = \frac{P[Y = j, D = k]}{P[Y \geq j, D = k]} = \frac{p_{jk}}{\sum_{l=j}^{\infty} p_{lk}} \quad (\text{A.2})$$

that is, the hazard defined directly in terms of the subdistribution probabilities for cause k , p_{jk} . The denominator can also be written

$$1 - P[Y < j] - P[Y \geq j, D \neq k].$$

As $P[Y = j, D = k]$ is a well-defined probability, we may write

$$\frac{P[Y = j, D = k]}{P[Y \geq j, D = k]} = \frac{P[Y = j|D = k] P[D = k]}{P[Y \geq j|D = k] P[D = k]} = \frac{P[Y = j|D = k]}{P[Y \geq j|D = k]} = \frac{\tilde{p}_{jk}}{\tilde{S}_{jk}} \quad (\text{A.3})$$

say, for the counterfactual ‘hazard’ probabilities for interval j considering only failures from cause k corresponding to the hazard in (A.2); see [8, p. 10 & Section 3.1]. The probability in equation (A.3) is referred to by some authors [9] as the *subdistribution hazard*; Crowder [8] terms it the *marginal hazard*. The probabilities in (A.3) can be estimated for cause k using standard approaches after weighting cases having other observed causes of death appropriately and retaining them in the risk set. The cause-specific CIF **for one cause** k can be then reconstructed using product-limit type estimation, and the dependence of the CIF on covariates can also be studied using regression modelling [9, 10]. For comprehensive details of the formulation and inference, see the textbook by Beyersmann and colleagues [11, Chapter 5].

Although such an approach is well-established and quite popular, in this paper we do not pursue it for several reasons. Most importantly, we wish to consider all causes of death simultaneously, not one in isolation; it is clearly not possible to compute the all-cause CIF using the individual marginal hazards, yet at the same time, as the collection of individual hazards are not estimated independently via the subdistribution approach, the variance of any estimator of the all-cause CIF is difficult to compute. Secondly, the usual objections to the Kalbfleisch-Prentice formulation of the cause-specific hazard concerning the inability to reproduce the cause-specific CIF due to the presence of intractable integrals do not apply to the discrete time case. Thirdly, in the discrete time setting, interpretation of the coefficients that appear in covariate-based hazard modification (see section 4) is also straightforward. Finally, the (effective) conditioning on the event $D = k$ in (A.3) and the interpretation of the hazard is – to us – conceptually problematic, as this event is in general not observed at time k .

S2 Variance estimation for the MSM

For individual i at interval j , we have

$$U_{ij}(\beta, \alpha) = \frac{\partial}{\partial \varphi} \left\{ \sum_{k=0}^K w_{ij}(\alpha) \mathbb{I}[D_{ij} = k] \log h_{ijk}(\beta) \right\} \quad (\text{A.1})$$

so that if $U = (U^\beta, U^\alpha)$, then

$$U_{ij}^\beta(\beta, \alpha) = \sum_{k=0}^K w_{ij}(\alpha) \mathbb{I}[D_{ij} = k] \frac{\partial}{\partial \beta} \log h_{ijk}(\beta) \quad (\text{A.2})$$

and

$$U_{ij}^{\beta\alpha}(\beta, \alpha^*) = \frac{\partial}{\partial \alpha} U_{ij}^\beta(\beta, \alpha^*) = \sum_{k=0}^K \frac{\partial w_{ij}(\alpha)}{\partial \alpha} \Big|_{\alpha=\alpha^*} \mathbb{I}[D_{ij} = k] \frac{\partial}{\partial \beta} \log h_{ijk}(\beta). \quad (\text{A.3})$$

Here, we place all β for all failure types in a single vector of length p_b ; $\beta = (\beta_1, \dots, \beta_K)^\top$, and

$$h_{ijk} = \frac{\exp\{\mathbf{x}_{ijk}\beta\}}{1 + \sum_{l=1}^K \exp\{\mathbf{x}_{ijl}\beta\}}$$

where for each k , $\mathbf{x}_{ijk}$ is the $(1 \times Kp_b)$ vector $\mathbf{x}_{ijk} = [0, 0, \dots, x_{ijk1}, \dots, x_{ijkp_b}, \dots, 0, 0]$.

First, we have for $k > 0$, in the multinomial model, for $k = 1, \dots, K$

$$\begin{aligned} \frac{\partial h_{ijk}(\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta} \left\{ \frac{\exp\{\mathbf{x}_{ijk}\beta\}}{1 + \exp\{\mathbf{x}_{ij1}\beta\} + \dots + \exp\{\mathbf{x}_{ijK}\beta\}} \right\} \\ &= h_{ijk}(\beta)\mathbf{x}_{ijk}^\top - h_{ijk}(\beta)h_{ij0}(\beta) \sum_{m=1}^K \exp\{\mathbf{x}_{ijm}\beta\}\mathbf{x}_{ijm}^\top = h_{ijk}(\beta) \left(\mathbf{x}_{ijk}^\top - \sum_{m=1}^K h_{ijm}(\beta)\mathbf{x}_{ijm}^\top \right) \end{aligned}$$

and $\exp\{\mathbf{x}_{ijm}\beta\} = h_{ijm}(\beta)/h_{ij0}(\beta)$. Hence

$$\frac{\partial}{\partial \beta} \log h_{ijk}(\beta) = \frac{1}{h_{ijk}(\beta)} \frac{\partial h_{ijk}(\beta)}{\partial \beta} = \left(\mathbf{x}_{ijk}^\top - \sum_{m=1}^K h_{ijm}(\beta)\mathbf{x}_{ijm}^\top \right)$$

Thus, if the parameters are not shared across different failure types, the contribution for those data for which $D_{ij} = k$, $k = 1, \dots, K$ is

$$\mathbf{x}_{ijk}^\top (1 - h_{ijk}(\beta))$$

for elements β_k of β , and

$$-\mathbf{x}_{ijl}^\top h_{ijl}(\beta)$$

for the elements β_l , $l \neq k$. When $k = 0$, the contribution for the whole β vector is

$$\frac{\partial}{\partial \beta} \log h_{ij0}(\beta) = \frac{1}{h_{ij0}(\beta)} \frac{\partial h_{ij0}(\beta)}{\partial \beta} = - \sum_{m=1}^K h_{ijm}(\beta)\mathbf{x}_{ijm}^\top = - \sum_{m=1}^K h_{ijm}(\beta)\mathbf{x}_{ijm}^\top.$$

This vector is essentially the sum of K contributions, one from each failure type, weighted by failure probability and then concatenated.

Secondly, dropping the dependence on i , we have that $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$ and

$$\begin{aligned} w_j(\alpha) &= \prod_{t=1}^j \frac{P(A_t = a_t | \bar{A}_{t-1}, V, C_t = 0, \alpha_1) P(C_t = c_t | \bar{A}_{t-1}, V, C_{t-1} = 0, \alpha_2)}{P(A_t = a_t | \bar{L}_t, \bar{A}_{t-1}, V, C_t = 0, \alpha_3) P(C_t = c_t | \bar{L}_t, \bar{A}_{t-1}, V, C_{t-1} = 0, \alpha_4)} \\ &= \prod_{t=1}^j \frac{(\exp\{a_t \mathbf{z}_{1t} \alpha_1\}) (1 + \exp\{\mathbf{z}_{3t} \alpha_3\}) (\exp\{c_t \mathbf{z}_{2t} \alpha_2\}) (1 + \exp\{\mathbf{z}_{4t} \alpha_4\})}{(\exp\{a_t \mathbf{z}_{3t} \alpha_3\}) (1 + \exp\{\mathbf{z}_{1t} \alpha_1\}) (\exp\{c_t \mathbf{z}_{4t} \alpha_4\}) (1 + \exp\{\mathbf{z}_{2t} \alpha_2\})} \end{aligned}$$

where these terms are defined using a logistic model, with

$$\begin{aligned} P(A_t = 1 | \bar{A}_{t-1}, V, C_t = 0, \alpha_1) &= \frac{\exp\{\mathbf{z}_{1t} \alpha_1\}}{1 + \exp\{\mathbf{z}_{1t} \alpha_1\}} \\ P(C_t = 1 | \bar{A}_{t-1}, V, C_{t-1} = 0, \alpha_2) &= \frac{\exp\{\mathbf{z}_{2t} \alpha_2\}}{1 + \exp\{\mathbf{z}_{2t} \alpha_2\}} \\ P(A_t = 1 | \bar{L}_t, \bar{A}_{t-1}, V, C_t = 0, \alpha_3) &= \frac{\exp\{\mathbf{z}_{3t} \alpha_3\}}{1 + \exp\{\mathbf{z}_{3t} \alpha_3\}} \\ P(C_t = 1 | \bar{L}_t, \bar{A}_{t-1}, V, C_{t-1} = 0, \alpha_4) &= \frac{\exp\{\mathbf{z}_{4t} \alpha_4\}}{1 + \exp\{\mathbf{z}_{4t} \alpha_4\}}. \end{aligned}$$

Then

$$\begin{aligned}\frac{\partial w_j(\alpha)}{\partial \alpha_1} &= w_j(\alpha) \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{1t}\alpha_1\}} \mathbf{z}_{1t}^\top \right] & \frac{\partial w_j(\alpha)}{\partial \alpha_2} &= w_j(\alpha) \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{2t}\alpha_2\}} \mathbf{z}_{2t}^\top \right] \\ \frac{\partial w_j(\alpha)}{\partial \alpha_3} &= w_j(\alpha) \sum_{t=1}^j \left[-\frac{1}{1 + \exp\{\mathbf{z}_{3t}\alpha_3\}} \mathbf{z}_{3t}^\top \right] & \frac{\partial w_j(\alpha)}{\partial \alpha_4} &= w_j(\alpha) \sum_{t=1}^j \left[-\frac{1}{1 + \exp\{\mathbf{z}_{4t}\alpha_4\}} \mathbf{z}_{4t}^\top \right].\end{aligned}\tag{A.4}$$

Thus

$$\begin{aligned}U_{ij}^\beta(\beta, \alpha^*) &= \sum_{k=0}^K w_{ij}(\alpha^*) \mathbb{I}[D_{ij} = k] \left(\mathbf{x}_{ijk}^\top - \sum_{m=1}^K h_{ijm}(\beta) \mathbf{x}_{ijm}^\top \right) & p_b \times 1 \\ U_{ij}^{\beta\alpha}(\beta, \alpha^*) &= \sum_{k=0}^K w_{ij}(\alpha^*) \mathbb{I}[D_{ij} = k] \left(\mathbf{x}_{ijk}^\top - \sum_{m=1}^K h_{ijm}(\beta) \mathbf{x}_{ijm}^\top \right) \mathbf{d}_{ij}^\top(\alpha^*) & p_b \times p_a\end{aligned}$$

where, generically,

$$\mathbf{d}_j(\alpha) = \begin{bmatrix} \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{1t}\alpha_1\}} \mathbf{z}_{1t}^\top \right] \\ \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{2t}\alpha_2\}} \mathbf{z}_{2t}^\top \right] \\ - \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{3t}\alpha_3\}} \mathbf{z}_{3t}^\top \right] \\ - \sum_{t=1}^j \left[\frac{1}{1 + \exp\{\mathbf{z}_{4t}\alpha_4\}} \mathbf{z}_{4t}^\top \right] \end{bmatrix} \quad p_a \times 1$$

leaving $U^{\beta\alpha}$ a $p_b \times p_a$ matrix.

The estimating function for α separates into four orthogonal components which are treated in an identical fashion. For α_1 , we have the logistic estimating equation

$$\dot{\ell}_{j\alpha_1}(\alpha_1) = \left(a_j - \frac{\exp\{\mathbf{z}_{1j}\alpha_1\}}{1 + \exp\{\mathbf{z}_{1j}\alpha_1\}} \right) \mathbf{z}_{1j}^\top \tag{A.5}$$

with derivative

$$\ddot{\ell}_{j\alpha_1}(\alpha_1) = -\frac{\exp\{\mathbf{z}_{1j}\alpha_1\}}{(1 + \exp\{\mathbf{z}_{1j}\alpha_1\})^2} \mathbf{z}_{1j} \mathbf{z}_{1j}^\top. \tag{A.6}$$

Similar results follow for α_2, α_3 and α_4 . The estimating equation is therefore computed via the estimating function which for each (i, j) takes the form.

$$U_{\text{adj}}^\beta(\beta, \alpha) = U^\beta(\beta, \alpha) - \mathbb{E} \left[U^{\beta\alpha}(\beta, \alpha) \right] \left(\mathbb{E} \left[\ddot{\ell}_\alpha(\alpha) \right] \right)^{-1} \dot{\ell}_\alpha(\alpha)$$

where $\mathbb{E} \left[U^{\beta\alpha}(\beta, \alpha) \right]$ is replaced by

$$\mathbb{P}_n \left[U^{\beta\alpha}(\beta, \hat{\alpha}) \right] = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} U_{ij}^{\beta\alpha}(\beta, \hat{\alpha}) = \hat{U}^{\beta\alpha}(\beta)$$

where $N = n_1 + \dots + n_n$, $\mathbb{E} [\ddot{\ell}_\alpha(\alpha)]$ is replaced by

$$\mathbb{P}_n [\ddot{\ell}_\alpha(\hat{\alpha})] = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} \ddot{\ell}_{ij\alpha}(\hat{\alpha}) = \Psi(\hat{\alpha})$$

say, yielding the estimating equation

$$\sum_{i=1}^N \sum_{j=1}^{n_i} \left\{ U_{ij}^\beta(\beta, \hat{\alpha}) - \hat{U}_{ij}^{\beta\alpha}(\beta) \Psi(\hat{\alpha})^{-1} \dot{\ell}_{ij\alpha}(\hat{\alpha}) \right\} = 0.$$

The estimating function based on $U_{\text{adj}}^\beta(\beta, \alpha^*)$ has asymptotic variance

$$\mathbb{E}[U_{\text{adj}}^\beta(\beta, \alpha^*)^{\otimes 2}] = \mathbb{E}[U_{\text{adj}}^\beta(\beta, \alpha^*) U_{\text{adj}}^\beta(\beta, \alpha^*)^\top]$$

yielding the variance for $\hat{\beta}$ as

$$\text{Var}[\hat{\beta}] = \mathbb{E} \left[\left\{ \left(\mathbb{E} \left[\frac{\partial}{\partial \beta} U_{\text{adj}}^\beta(\beta, \alpha^*) \right] \right)^{-1} U_{\text{adj}}^\beta(\beta, \alpha^*) \right\}^{\otimes 2} \right] = \mathbb{E} \left[\left\{ \left(\mathbb{E} [U_{\text{adj}}^{\beta\beta}(\beta, \alpha^*)] \right)^{-1} U_{\text{adj}}^\beta(\beta, \alpha^*) \right\}^{\otimes 2} \right].$$

In this expression, we have

$$U_{\text{adj}}^{\beta\beta}(\beta, \alpha^*) = \frac{\partial}{\partial \beta} U_{\text{adj}}^\beta(\beta, \alpha^*) = U^{\beta\beta}(\beta, \alpha^*) - \mathbb{E} [U^{\beta\alpha\beta}(\beta, \alpha^*)] \left(\mathbb{E} [\ddot{\ell}_\alpha(\alpha^*)] \right)^{-1} \dot{\ell}_\alpha(\alpha^*). \quad (\text{A.7})$$

For the first term in (A.7),

$$\begin{aligned} U_{ij}^{\beta\beta}(\beta, \alpha^*) &= - \sum_{k=0}^K w_{ij}(\alpha^*) \mathbb{I}[D_{ij} = k] \left\{ \sum_{m=1}^K h_{ijm}(\beta) \mathbf{x}_{ijm}^\top \left(\mathbf{x}_{ijm} - \sum_{m'=1}^K h_{ijm'}(\beta) \mathbf{x}_{ijm'} \right) \right\} \quad p_b \times p_b. \\ &= -w_{ij}(\alpha^*) \left\{ \sum_{m=1}^K h_{ijm}(\beta) \mathbf{x}_{ijm}^\top \left(\mathbf{x}_{ijm} - \sum_{m'=1}^K h_{ijm'}(\beta) \mathbf{x}_{ijm'} \right) \right\}. \end{aligned}$$

In the derivative of the second term, the derivative with respect to β_r is

$$U_{ij}^{\beta\alpha\beta_r}(\beta, \alpha^*) = - \sum_{k=0}^K w_{ij}(\alpha^*) \mathbb{I}[D_{ij} = k] \left\{ \sum_{m=1}^K \frac{\partial h_{ijm}(\beta)}{\partial \beta_r} \mathbf{x}_{ijm}^\top \right\} \mathbf{d}_{ij}^\top(\alpha^*) \quad p_b \times p_a.$$

Here, recall that

$$\frac{\partial h_{ijm}(\beta)}{\partial \beta_r} = h_{ijm}(\beta) \left(x_{ijmr} - \sum_{m'=1}^K h_{ijm'}(\beta) x_{ijm'r} \right) = \dot{h}_{ijmr}(\beta) \quad 1 \times 1$$

so that

$$U_{ij}^{\beta\alpha\beta_r}(\beta, \alpha^*) = - \sum_{k=0}^K w_{ij}(\alpha^*) \mathbb{I}[D_{ij} = k] \left\{ \sum_{m=1}^K \dot{h}_{ijmr}(\beta) \mathbf{x}_{ijm}^\top \right\} \mathbf{d}_{ij}^\top(\alpha^*) \quad p_b \times p_a.$$

whose expected value may be estimated consistently by

$$-\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} \sum_{k=0}^K w_{ij}(\hat{\alpha}) \mathbb{I}[D_{ij} = k] \left\{ \sum_{m=1}^K \dot{h}_{ijmr}(\hat{\beta}) \mathbf{x}_{ijm}^\top \right\} \mathbf{d}_{ij}^\top(\hat{\alpha}) = \hat{\mathbf{H}}_r$$

say. Note that the inner summation over m does not depend on k , so we have in fact that

$$\hat{\mathbf{H}}_r = -\frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{n_i} w_{ij}(\hat{\alpha}) \left\{ \sum_{m=1}^K \dot{h}_{ijmr}(\hat{\beta}) \mathbf{x}_{ijm}^\top \right\} \mathbf{d}_{ij}^\top(\hat{\alpha})$$

Thus, the second term in (A.7) may be estimated consistently by binding columnwise the components

$$\left[\hat{\mathbf{H}}_1 \Psi(\hat{\alpha})^{-1} \dot{\ell}_\alpha(\alpha), \dots, \hat{\mathbf{H}}_{p_b} \Psi(\hat{\alpha})^{-1} \dot{\ell}_\alpha(\hat{\alpha}) \right].$$

Thus the expected value of $U_{\text{adj}}^{\beta\beta}(\beta, \alpha)$ in (A.7) is consistently estimated by

$$\hat{U}_{\text{adj}}^{\beta\beta}(\hat{\beta}, \hat{\alpha}) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} \left\{ U_{ij}^{\beta\beta}(\hat{\beta}, \hat{\alpha}) - \left[\hat{\mathbf{H}}_1 \Psi(\hat{\alpha})^{-1} \dot{\ell}_{ij\alpha}(\hat{\alpha}), \dots, \hat{\mathbf{H}}_{p_b} \Psi(\hat{\alpha})^{-1} \dot{\ell}_{ij\alpha}(\hat{\alpha}) \right] \right\}.$$

Hence, $\text{Var}[\hat{\beta}]$ is consistently estimated by

$$\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} \left\{ \left[\hat{U}_{\text{adj}}^{\beta\beta}(\hat{\beta}, \hat{\alpha}) \right]^{-1} \left\{ U_{ij}^{\beta\beta}(\hat{\beta}, \hat{\alpha}) - \hat{U}_{ij}^{\beta\alpha}(\hat{\beta}) \Psi(\hat{\alpha})^{-1} \dot{\ell}_{ij\alpha}(\hat{\alpha}) \right\} \right\}^{\otimes 2}.$$

S3 Variance estimation for the CIF

Suppressing the dependence on treatment history $\bar{a}$, we have for the (counterfactual) cause-specific failure probabilities $p_{jk} = \text{P}[Y = j, D = k]$ the expression

$$p_{jk} = h_{jk} \prod_{l=1}^{j-1} (1 - h_l) = h_{jk} \prod_{l=1}^{j-1} h_{l0}$$

where

$$h_{j0} = 1 - \sum_{k=1}^K h_{jk} = 1 - h_j.$$

Recall that in our model

$$h_{jk} = \frac{\exp\{\mathbf{x}_{jk}\beta\}}{1 + \exp\{\mathbf{x}_{j1}\beta\} + \dots + \exp\{\mathbf{x}_{jK}\beta\}} \quad k = 1, \dots, K$$

with

$$h_{j0} = \frac{1}{1 + \exp\{\mathbf{x}_{j1}\beta\} + \dots + \exp\{\mathbf{x}_{jK}\beta\}}$$

for parameter vector $\beta = (\beta_1, \dots, \beta_d)^\top$, and each $\mathbf{x}_{jk} = (x_{jk1}, \dots, x_{jkd})$ a $d \times 1$ row vector containing relevant treatment and covariate values at interval j . Thus

$$\theta_{jk} = \log p_{jk} = \mathbf{x}_{jk}\beta - \sum_{l=1}^j \log(1 + \exp\{\mathbf{x}_{l1}\beta\} + \dots + \exp\{\mathbf{x}_{lK}\beta\})$$

and hence

$$\frac{\partial \theta_{jk}}{\partial \beta_r} = x_{jkr} - \sum_{l=1}^j \frac{\sum_{m=1}^K x_{lmr} \exp\{\mathbf{x}_{lm}\beta\}}{(1 + \exp\{\mathbf{x}_{l1}\beta\} + \dots + \exp\{\mathbf{x}_{lK}\beta\})} = x_{jkr} - \sum_{l=1}^j \sum_{m=1}^K x_{lmr} h_{lm}.$$

Thus the asymptotic variance of the vector $\theta_k = (\theta_{1k}, \dots, \theta_{Jk})$ can be computed routinely; in the usual asymptotic approximation $\hat{\beta} \sim \mathcal{AN}(\beta, \hat{\Sigma})$, thus by the Delta Method, $\hat{\theta}_k$ has asymptotic variance

$$V_1(\hat{\theta}_k) = (M(\hat{\theta}_k))^\top \hat{\Sigma} M(\hat{\theta}_k)$$

where $M(\theta_k)$ is the $d \times J$ matrix with (r, j) th element

$$\frac{\partial \theta_{jk}}{\partial \beta_r}.$$

Taking an exponential transformation of each element of θ_k yields $p_k = (p_{1k}, \dots, p_{Jk})^\top$, and the cumulative function

$$P_{jk} = \sum_{l=1}^j p_{lk}$$

can be obtained using the linear transform form $P_k = L_J p_k$ where L is the $J \times J$ matrix with zeros in the upper triangle, and ones elsewhere. The resulting asymptotic variance for P_k is thus $L_J^\top \mathbf{E}_k(\hat{\theta}_k) V_1(\hat{\theta}_k) \mathbf{E}_k(\hat{\theta}_k) L_J$ where $\mathbf{E}_k(\hat{\theta}_k)$ is the diagonal matrix with $\exp\{\theta_{jk}\} = p_{jk}$ in the (j, j) th position, $j = 1, \dots, J$.

An alternative strategy is to derive the asymptotic variance for the vector $\eta_k = (\eta_{1k}, \dots, \eta_{Jk})^\top$, where

$$\eta_{jk} = \log \left(\frac{\exp\{\theta_{jk}\}}{1 - \exp\{\theta_{jk}\}} \right) = \log \left(\frac{p_{jk}}{1 - p_{jk}} \right).$$

so that, for each j, k ,

$$\frac{\partial \eta}{\partial \theta} = \frac{1}{1 - \exp\{\theta\}} = \frac{1}{1 - p}.$$

The asymptotic variance of η_k is therefore

$$V_2(\hat{\theta}) = \mathbf{D}_k(\hat{\theta}) \hat{V}_1(\hat{\theta}_k) \mathbf{D}_k(\hat{\theta})$$

where $\mathbf{D}_k(\hat{\theta})$ is the $J \times J$ diagonal matrix with (j, j) th element $1/(1 - \hat{p}_{jk})$. This asymptotic Normal approximation can be used to construct a simultaneous 95% confidence interval for η_k , and hence p_k and P_k by transformation.

S4 R code

S4.1 Variance calculation

```
## This function extracts information from the fitted "multinom" object. It is based
## on the function "estfun.glm" written by Thomas Lumley, also used below.
estfun.multinom <- function(mn.obj)
{
  # Create X matrix from multinom object
  mf <- model.frame(mn.obj)
  xmat <- model.matrix(terms(mn.obj), mf)
  np <- dim(xmat)[2]
  # Calculate variance weight and residual, then return
  residX <- NULL
  for(i in 2:dim(residuals(mn.obj, "working"))[2])
    residX <- cbind(residX, as.vector(residuals(mn.obj, "working")[,i] * mn.obj$weights) * xmat)
  output <- residX
  output
}

## In the calculations below, all objects have been named so as to
## correspond as closely as possible to the derivations in S2 above.
# iptw.fit is the weighted multinomial object:
U <- estfun.multinom(iptw.fit)

# treat.fit.small and treat.fit.big are, respectively, objects resulting
# from the logistic regression used to fit the numerator
# and denominator models for the weighting; similarly for censoring.
l.dot.alpha1 <- estfun.glm(treat.fit.small)
l.dot.alpha2 <- estfun.glm(censor.fit.small)
l.dot.alpha3 <- estfun.glm(treat.fit.big)
l.dot.alpha4 <- estfun.glm(censor.fit.big)
l.dot <- cbind(l.dot.alpha1,l.dot.alpha2,l.dot.alpha3,l.dot.alpha4)

p.star1 <- dim(l.dot.alpha1)[2]
p.star2 <- dim(l.dot.alpha2)[2]
p.star3 <- dim(l.dot.alpha3)[2]
p.star4 <- dim(l.dot.alpha4)[2]
p.star <- p.star1+p.star2+p.star3+p.star4

l.dotdot <- matrix(0,nrow=p.star,ncol=p.star)
l.dotdot[1:p.star1,1:p.star1] <- summary(treat.fit.small)$cov.scaled
l.dotdot[(p.star1+1):(p.star1+p.star2),(p.star1+1):(p.star1+p.star2)] <-
  summary(censor.fit.small)$cov.scaled
l.dotdot[(p.star1+p.star2+1):(p.star1+p.star2+p.star3),(p.star1+p.star2+1):
```

```

      (p.star1+p.star2+p.star3)] <- summary(treat.fit.big)$cov.scaled
l.dotdot[(p.star1+p.star2+p.star3+1):p.star,(p.star1+p.star2+p.star3+1):p.star]
      <- summary(censor.fit.big)$cov.scaled

adjustment <- l.dotdot %*% t(l.dot)

bold.d1 <- NULL
temp <- (A-predict(treat.fit.small,type="response")) *
      model.matrix(terms(treat.fit.small),model.frame(treat.fit.small))
for(i in 1:dim(temp)[2])
      bold.d1 <- cbind(bold.d1,unlist(temp[,i],ID, cumsum))

bold.d2 <- NULL
temp <- (Censored-predict(censor.fit.small,type="response")) *
      model.matrix(terms(censor.fit.small),model.frame(censor.fit.small))
for(i in 1:dim(temp)[2])
      bold.d2 <- cbind(bold.d2,unlist(temp[,i],ID, cumsum))

bold.d3 <- NULL
temp <- -(A-predict(treat.fit.big,type="response")) *
      model.matrix(terms(treat.fit.big),model.frame(treat.fit.big))
for(i in 1:dim(temp)[2])
      bold.d3 <- cbind(bold.d3,unlist(temp[,i],ID, cumsum))

bold.d4 <- NULL
temp <- -(Censored-predict(censor.fit.big,type="response")) *
      model.matrix(terms(censor.fit.big),model.frame(censor.fit.big))
for(i in 1:dim(temp)[2])
      bold.d4 <- cbind(bold.d4,unlist(temp[,i],ID, cumsum))

bold.d <- cbind(bold.d1,bold.d2,bold.d3,bold.d4)
U.ba <- t(U) %*% bold.d
U.adj <- t(U) - U.ba %*% adjustment
cheese <- U.adj %*% t(U.adj)

x.mn <- model.matrix(terms(iptw.fit), model.frame(iptw.fit))
temp1 <- temp2 <- NULL
for(i in 2:dim(fitted(iptw.fit, "response"))[2]) {
      temp1 <- cbind(temp1, as.vector(fitted(iptw.fit, "response")[,i]) * x.mn)
      temp2 <- cbind(temp2, as.vector(fitted(iptw.fit, "response")[,i] * iptw.fit$weights) * x.mn)
}
bigX <- cbind(x.mn,x.mn,x.mn)
pb <- dim(x.mn)[2]

dbeta.U <- matrix(0,ncol=dim(U)[2],nrow=dim(U)[2])

```

```

for(i in 1:dim(U)[1]) {
  dbeta.U <- dbeta.U + temp1[i,] %*% t(temp2[i,])
  dbeta.diag <- -bigX[i,] %*% t(temp2[i,])
  dbeta.U[1:pb,1:pb] <- dbeta.U[1:pb,1:pb] + dbeta.diag[1:pb,1:pb]
  dbeta.U[(pb+1):(2*pb),(pb+1):(2*pb)] <-
    dbeta.U[(pb+1):(2*pb),(pb+1):(2*pb)] + dbeta.diag[(pb+1):(2*pb),(pb+1):(2*pb)]
  dbeta.U[(2*pb+1):(3*pb),(2*pb+1):(3*pb)] <- dbeta.U[(2*pb+1):(3*pb),(2*pb+1):(3*pb)] +
    dbeta.diag[(2*pb+1):(3*pb),(2*pb+1):(3*pb)]
}

U.bab <- matrix(0,ncol=dim(U)[2],nrow=dim(U)[2])
for(i in 1:dim(U)[1]) {
  U.bab <- U.bab + (temp1[i,] %*% t(bold.d[i,])) %*% (adjustment[,i] %*% t(temp2[i,]))
  U.bab.diag <- (-bigX[i,] %*% t(bold.d[i,])) %*% (adjustment[,i] %*% t(temp2[i,]))
  U.bab[1:pb,1:pb] <- U.bab[1:pb,1:pb] + U.bab.diag[1:pb,1:pb]
  U.bab[(pb+1):(2*pb),(pb+1):(2*pb)] <- U.bab[(pb+1):(2*pb),(pb+1):(2*pb)] +
    U.bab.diag[(pb+1):(2*pb),(pb+1):(2*pb)]
  U.bab[(2*pb+1):(3*pb),(2*pb+1):(3*pb)] <- U.bab[(2*pb+1):(3*pb),(2*pb+1):(3*pb)] +
    U.bab.diag[(2*pb+1):(3*pb),(2*pb+1):(3*pb)]
}

deriv.U.adj <- dbeta.U - U.bab
bread <- solve(deriv.U.adj)
MSMvar <- bread %*% cheese %*% bread

```

S4.2 Simulating data

```

##
### Data generation
##

n<-1000 # Number of subjects
N<-10 #number of intervals per subject
K<-3 # Number of causes of death

## This is the matrix of parameters of interest, possibly different
## at each interval
psi.mat<-matrix(0,nrow=K,ncol=N+1)

##Here are the effect sizes for the K=3 causes
psi.mat[1,]<--log(2)
psi.mat[2,]<--log(1.5)
psi.mat[3,]<--log(1)

##Here the (untreated) all-cause rate is set to lambda=0.01, with

```

```

##lambda/K per cause; muK=lambda is used in the algorithm.
lambda<-0.01
gamma.vec<-rep(log(lambda/K))
muK<-sum(exp(gamma.vec))

A<-L<-ID<-Y<-Z<-Tv<-Int<-ALast<-LLast<-LFirst<-numeric()
T0.vec<-T.vec<-Y.vec<-Z.vec<-rep(0,n)

##Here are the coefficients determining the
##mediation and treatment assignment mechanisms.
bevec<-c(log(3/7),2,log(0.5),log(1.5))
alvec<-c(log(2/7),0.5,0.5,log(4))

##cval is used as in Young's algorithm to introduce the confounding
cval<-30

##Begin the data-generation loop
for(i in 1:n){

  ##Generate the counterfactual (untreated) survival time
  T0<-rexp(1,lambda)
  Ival<-as.numeric(T0 < cval)

  ##Begin the interval-by-interval simulation
  m<-0
  mu.tot<-0
  A.vec<-L.vec<-ALast.vec<-LLast.vec<-LFirst.vec<-rep(0,N+1)

  ##Implement Young's algorithm with multiple causes
  ##Generate the survival time, then the cause

  while(muK*T0 > mu.tot & m <= N){
    if(m == 0){
      ##First interval
      eta<-bevec[1]+bevec[2]*Ival+bevec[3]*0+bevec[4]*0
      pval<-1/(1+exp(-eta))
      L.vec[m+1]<-rbinom(1,1,pval)

      eta<-alvec[1]+alvec[2]*L.vec[m+1]+alvec[3]*0+alvec[4]*0
      pval<-1/(1+exp(-eta))
      A.vec[m+1]<-rbinom(1,1,pval)
      ALast.vec[m+1]<-0;LLast.vec[m+1]<-0
      LFirst.vec<-rep(L.vec[m+1],N+1)
    }else{
      ##Subsequent intervals

```

```

    eta<-bevec[1]+bevec[2]*Ival+bevec[3]*A.vec[m]+
        bevec[4]*L.vec[m]
    pval<-1/(1+exp(-eta))
    L.vec[m+1]<-rbinom(1,1,pval)

    eta<-alvec[1]+alvec[2]*L.vec[m+1]+alvec[3]*L.vec[m]+
        alvec[4]*A.vec[m]
    pval<-1/(1+exp(-eta))
    A.vec[m+1]<-rbinom(1,1,pval)
    ALast.vec[m+1]<-A.vec[m];LLast.vec[m+1]<-L.vec[m]

}
muval<-sum(exp(gamma.vec+A.vec[m+1]*psi.mat[,m+1]))

##Tval is computed for each interval, but is overwritten
##until the final interval

Tval<-m+(muK*T0-mu.tot)/muval
mu.tot<-mu.tot+muval
m<-m+1
}

##After exiting the loop, the survival time has been generated as Tval
##Now need to generate the failure type.

if(m > N){
    ##In the case of censoring at tenth interval, no failure.
    Tval<-m-1
    Z.vec[i]<-0
}else{
    ##In the case of failure, use the ratio hazards to define the
    ##relevant multinomial distribution on the K causes.
    Z.vec[i]<-sample(c(1:K),1,prob=exp(gamma.vec+A.vec[m]*psi.mat[,m]))
}

##Store the outcomes

T0.vec[i]<-T0
T.vec[i]<-Tval
Y.vec[i]<-m-1

ID<-c(ID,rep(i,m))
Int<-c(Int,c(1:m))
A<-c(A,A.vec[1:m])
L<-c(L,L.vec[1:m])

```

```

    ALast<-c(ALast,ALast.vec[1:m])
    LLast<-c(LLast,LLast.vec[1:m])
    LFirst<-c(LFirst,LFirst.vec[1:m])
    Z<-c(Z,rep(0,m-1),Z.vec[i])
    tv<-c(1:m);tv[m]<-Tval
    Tv<-c(Tv,tv)
  }

DeathsK.df<-data.frame(ID,Int,Tv,A,ALast,L,LLast,LFirst,Z)

##Trim off the intervals beyond the Nth (loop goes one too far)

DeathsK.df<-DeathsK.df[DeathsK.df$Int<=N,]
write.csv(DeathsK.df,file="DeathsK.csv",row.names=FALSE)

```

S4.3 Fitting a predicted cumulative incidence function

```

##
### Compute the predicted CIF (1 minus all-cause survival)
### prediction plot by Monte Carlo
##

##First, compute the MSM-based estimates.

rm(list=ls())
y.dat<-read.csv("DeathsK.csv")
summary(y.dat)

expit<-function(logit){exp(logit)/(1+exp(logit))}

## Construct stabilized weights
denominator<-rep(NA, nrow(y.dat))
logit<-predict(glm(A~ALast+L+LLast, family=binomial, data=y.dat))
denominator[y.dat$A==1]<-expit(logit[y.dat$A==1])
denominator[y.dat$A==0]<-1-expit(logit[y.dat$A==0])
numerator<-rep(NA, nrow(y.dat))
logit<-predict(glm(A~ALast, family=binomial, data=y.dat))
numerator[y.dat$A==1]<-expit(logit[y.dat$A==1])
numerator[y.dat$A==0]<-1-expit(logit[y.dat$A==0])
stab.wt<-unlist(tapply(numerator/denominator,y.dat$ID,cumprod))
summary(stab.wt)

## Pooled multinomial regression MSM (weighted)
library(nnet)

```

```

modmsm<-multinom(Z~A, data=y.dat, weights=stab.wt, Hess = FALSE)
summary(modmsm)

## Forward simulation to predict all-cause CIF
## using n=10000 Monte Carlo samples
## Avoid censoring by setting N very large
n<-10000
N<-1500
K<-3

##Extract the fitted coefficients
psi.mat<-matrix(0,nrow=K,ncol=N)
psi.mat[1,<-coef(modmsm)[1,2]
psi.mat[2,<-coef(modmsm)[2,2]
psi.mat[3,<-coef(modmsm)[3,2]

##Here the fitted all-cause rate is computed
gamma.vec<-rep(coef(modmsm)[,1]);muK<-lambda<-sum(exp(gamma.vec))

T0.vec<-T.vec<-Y.vec<-Z.vec<-rep(0,n)
A<-ID<-Y<-Z<-Tv<-Int<-numeric()

##Again use Young's algorithm, but with treatment A set to 1 for
##all intervals; could use A=0 also.

for(i in 1:n){
  T0<-rexp(1,lambda)
  m<-0
  mu.tot<-0
  A.vec<-rep(0,N+1)
  while(muK*T0 > mu.tot & m <= N){
    A.vec[m+1]<-1
    muval<-sum(exp(gamma.vec+A.vec[m+1]*psi.mat[,m+1]))
    Tval<-m+(muK*T0-mu.tot)/muval
    mu.tot<-mu.tot+muval
    m<-m+1
  }
  if(m > N){
    Tval<-m-1
    Z.vec[i]<-0
  }else{
    Z.vec[i]<-sample(c(1:K),1,prob=exp(gamma.vec+A.vec[m]*psi.mat[,m]))
  }

  T0.vec[i]<-T0

```



```

T.vec[i]<-Tval
Y.vec[i]<-m-1

ID<-c(ID,rep(i,m))
Int<-c(Int,c(1:m))
A<-c(A,A.vec[1:m])
Z<-c(Z,rep(0,m-1),Z.vec[i])
tv<-c(1:m);tv[m]<-Tval
Tv<-c(Tv,tv)

}

plot(ecdf(T.vec),xlim=c(0,100),do.p=F)
lines(ecdf(T0.vec),col="gray")
legend(5,.3,col=c("black","gray"),
      legend=c("Always exposed","Never exposed"),lwd=1)

# under exponential model, can also calculate CIF analytically
mu.rate1<-exp(gamma.vec+psi.mat[,1])
mu.rate0<-exp(log(lambda))

xv<-seq(0,100,0.01)
y1<-1-exp(-sum(mu.rate1)*xv)
y0<-1-exp(-sum(mu.rate0)*xv)
lines(xv,y1,col="green")
lines(xv,y0,col="blue")

```

S5 Additional simulation results

S5.1 Finite-sample performance of the asymptotic variance estimator

We present a brief simulation using the data-generating models presented in Table 2 of the main manuscript to demonstrate the performance of the asymptotic variance estimator derived in §S2 above. In Table 1, we show the average effect estimate, its empirical standard deviation, and the square root of the mean variance estimator (denoted $\widehat{SE}(\psi_1)$). The empirical and asymptotic estimator of the variability agree closely, and coverage of 95% confidence intervals is very close to the nominal level.

Table 1: Simulation results from fitting the multiple outcome marginal structural model for various sample sizes and treatment effects in two causes of death over 1000 simulated data sets: assessing the performance of the variance estimator, $\widehat{SE}(\psi_1)$.

n	$\hat{\psi}_1$	<u>Cause of death 1</u>			$\hat{\psi}_2$	<u>Cause of death 2</u>		
		SD($\hat{\psi}_1$)	$\widehat{SE}(\hat{\psi}_1)$	Cover		SD($\hat{\psi}_2$)	$\widehat{SE}(\hat{\psi}_2)$	Cover
$(\psi_1, \psi_2)=(-0.1,0.4)$								
500	-0.078	0.525	0.488	94.5	0.440	0.473	0.437	93.7
1000	-0.104	0.343	0.340	96.0	0.408	0.308	0.304	94.6
2500	-0.106	0.208	0.213	95.4	0.407	0.184	0.191	96.2
$(\psi_1, \psi_2)=(0.3,0.3)$								
500	0.323	0.467	0.445	94.6	0.309	0.457	0.445	95.4
1000	0.319	0.317	0.311	94.9	0.288	0.326	0.312	94.2
2500	0.315	0.194	0.195	94.7	0.311	0.193	0.195	95.3
$(\psi_1, \psi_2)=(0,0)$								
500	0.020	0.491	0.477	94.9	0.023	0.487	0.474	95.1
1000	0.011	0.326	0.330	95.5	-0.008	0.335	0.330	95.4
2500	-0.005	0.204	0.208	95.5	0.002	0.202	0.207	95.5

S5.2 Three causes of failure

We performed simulations with three competing causes of death, considering a range of parameters (ψ_1, ψ_2, ψ_3) with a sample size of $n = 1000$. The data were generated using the modified Young algorithm in a manner similar to that for two causes of death, as outlined §5 of the main text. Results are presented in Table 2.

The simulations again demonstrate the bias of conventional model estimators, and indicate that the conventional model that does not adjust for the time-varying confounder (Model 1) exhibits the highest MSE at $n = 1000$. As in the simulations for two causes of death, the conventional model that adjusts for the

time-varying confounder (Model 2) is typically seriously biased (up to 144%), while the MO-MSM is not. Again, we note the considerable bias of the conventional models under the null setting of no treatment effect, and the unbiasedness of the MO-MSM.

In these simulations, the event rate was relatively low. In the scenario of no treatment effect – i.e. when $(\psi_1, \psi_2, \psi_3) = (0, 0, 0)$ – the mean number of events of each type was 31.7, with never more than 53 events of any type in a sample of size 1000. The highest event rate was observed with causal parameters set at $(\psi_1, \psi_2, \psi_3) = (1.0, 0.5, 0)$, where there were, on average, 57.4 failures from cause of death 1, 41.6 from cause of death 2, and 31.8 from the third cause of death.

S5.3 Three causes of failure: a more realistic setting

We performed a third simulation study which more closely resembles the Canadian Co-infection Cohort. The simulation study once again considered three outcomes (types of death) of interest, however the design differed from that of Section S5.2 in two key ways: (i) the baseline event rate was reduced, by setting λ_0 to 0.005, and (ii) there was a delay between diagnosis and cohort entry. It is this second feature that is of particular interest. We accomplished this by creating a “cohort entry” variable, which was a random draw from 1 to 5, with 1 indicating no delay between diagnosis and cohort entry and 5 a delay of two years (since each of the six-month intervals 1 through 4 were not observed). Note that for some individuals, an event may occur prior to cohort entry. For example, if a patient’s cohort entry time is 4 but an event occurs at time 2, this individual will never appear in the cohort. Two additional variables were made available for analysis in these simulations: time since infection and a binary indicator of whether any exposure was received prior to cohort entry (this variable is set to 0 for all individuals who enter the cohort at the time of diagnosis). Two “population” sizes were considered: 950 and 2500; the first of these was to generate data that were more similar to those in our motivating example, and the second was to examine large-sample properties of the estimators.

Note that because cohort entry time is generated from a uniform distribution, it is entirely unrelated to covariates, exposure, or the potential outcomes. We would therefore not expect to observe any bias arising from missing some (or even all) intervals for some individuals in models that are correctly specified.

Taking a population size of 950, the average sample size obtained was 930, and the range was 914-943 (recall that any individual whose event time fell before the cohort entry time failed to become a member of the cohort/sample). The mean (range) number of deaths by each of the three causes was 37 (20-58), 25 (10-45), and 19 (4-35), respectively, leaving, 850 (817-881) censored individuals. For the population of size 2500, the average sample size was 2448, with a range of 2423-2470. The mean (range) number of deaths by each of the three causes, respectively, was 97 (63-130), 66 (38-97), and 50 (26-74), respectively, leaving, 2236 (2177-2289) censored individuals.

Results are shown in Table 3 for six different models. We fit three conventional models: the first adjusted only for exposure; the second adjusted for exposure, time since infection, and prior exposure; and the third adjusted for all variables in the second model as well as the time-varying confounder. Three multiple-outcome marginal structural models were also fit, and these varied by the variables included in the response model: the first included only exposure; the second included exposure and time since infection; the third included exposure, time since infection, and exposure prior to cohort entry. In the smaller sample size, the MO-MSM that adjusted only for exposure performed the best in terms of both bias and variability, followed by the multinomial model that adjusted for only exposure. However as sample size increases, the MO-MSM that adjusted for both exposure and time since infection perform equally well, while the conventional

model continued to exhibit reasonably high bias (over 25% even for the parameters associated with the most frequently occurring outcome). It is not surprising that the models that adjust for exposure prior to cohort entry exhibits bias: as noted by Robins [12], “standard methods [...] that predict the mortality rate at time t using a summary of [exposure] history up to time t will produce biased estimates of the causal effect”. Adjusting prior exposure history without being able to account for variables that confound that exposure or lie between it and the outcome are indeed likely to result in bias.

We repeated the simulations using a later date of entry: we considered up to 30 intervals, where cohort entry occurred between visits 15 and 25 to ensure a higher degree of left-censoring (missed visits) and left-truncation (missed individuals/events). Taking a population size of 950, the average sample size obtained was 780.5, and the range was 749 to 812 (implying 15-21% of the 950 experienced events prior to the cohort entry date, and were therefore excluded from the cohort). For the MO-MSM that adjust only for prior exposure, the most competitive model, the percent bias for the parameter ψ for outcomes 1, 2, and 3, respectively, was 6.4, 7.5, and 17.2; that is, bias has decreased by further delaying cohort entry. We posit the following possible explanation: when visits are left-censored but few events are actually missed, we observed an upwards bias as there were many intervals in which no event occurred which were not observed, making the exposure appear more harmful than it truly is. By further delaying entry so that event times were not only left-censored, but its distribution was left-truncated, a counter-balancing bias has been introduced: by losing events, a downward bias is introduced. It is clear from these results that the impact of left-censoring and left-truncation of event times in the presence of time-varying exposures prior to cohort entry is complex. Our proposed solution is ad hoc and not unbiased, but in the absence of any methods to address this problem, it presents a reasonable solution both in terms of performance and feasibility.

Table 2: Simulation results from fitting two conventional multinomial models and the multiple-outcome marginal structural model for various treatment effects, $n = 1000$: three causes of death. Model 1 is a conventional multinomial regression model which does not adjust for time-varying confounders, Model 2 is a conventional multinomial regression model which adjusts for time-varying confounders, and MO-MSM denotes the proposed multiple-outcome marginal structural model.

	Cause of death 1			Cause of death 2			Cause of death 3		
	Model 1	Model 2	MO-MSM	Model 1	Model 2	MO-MSM	Model 1	Model 2	MO-MSM
$(\psi_1, \psi_2, \psi_3) = (1.0, 0.5, 0)$									
$\hat{\psi}$	1.452	1.314	1.024	0.949	0.809	0.523	0.447	0.307	0.011
SD($\hat{\psi}$)	0.317	0.319	0.355	0.337	0.339	0.435	0.375	0.378	0.557
Bias	0.452	0.314	0.024	0.449	0.309	0.023	0.447	0.307	0.011
% bias	45.2	31.4	2.4	89.7	61.8	4.5	-	-	-
MSE	0.305	0.200	0.127	0.315	0.210	0.190	0.341	0.237	0.311
$(\psi_1, \psi_2, \psi_3) = (-0.2, 0.2, 0.4)$									
$\hat{\psi}$	0.229	0.083	-0.204	0.633	0.489	0.208	0.829	0.684	0.414
SD($\hat{\psi}$)	0.399	0.402	0.444	0.364	0.366	0.473	0.342	0.344	0.503
Bias	0.429	0.283	-0.004	0.433	0.289	0.008	0.429	0.284	0.014
% bias	-214.6	-141.7	1.8	216.7	144.4	3.9	107.3	70.9	3.4
MSE	0.343	0.242	0.198	0.320	0.217	0.223	0.301	0.198	0.254
$(\psi_1, \psi_2, \psi_3) = (-0.6, 0, 1)$									
$\hat{\psi}$	-0.205	-0.349	-0.641	0.438	0.295	0.012	1.442	1.300	1.039
SD($\hat{\psi}$)	1.049	1.063	1.050	0.379	0.382	0.485	0.315	0.316	0.463
Bias	0.395	0.2512073	-0.041	0.438	0.295	0.012	0.442	0.300	0.039
% bias	-65.9	-41.9	6.8	-	-	-	44.2	30.0	3.9
MSE	1.257	1.193	1.104	0.336	0.233	0.236	0.294	0.190	0.216
$(\psi_1, \psi_2, \psi_3) = (0, 0, 0)$									
$\hat{\psi}$	0.427	0.281	0.002	0.424	0.277	0.006	0.426	0.280	0.016
SD($\hat{\psi}$)	0.372	0.374	0.414	0.378	0.380	0.485	0.375	0.378	0.553
MSE	0.321	0.219	0.171	0.322	0.221	0.235	0.323	0.222	0.306

References

1. Pintilie M. Analysing and interpreting competing risk data. *Statistics in Medicine* 2007; **26**(6):1360–1367. URL <http://dx.doi.org/10.1002/sim.2655>.
2. Latouche A, Beyersmann J, Fine JP. Comments on -Analysing and interpreting competing risk data- by M. Pintilie, *Statistics in Medicine* 2006. DOI: 10.1002/sim.2655. *Statistics in Medicine* 2007; **26**(19):3676–3679. URL <http://dx.doi.org/10.1002/sim.2823>.
3. Pintilie M. Author's reply. *Statistics in Medicine* 2007; **26**(19):3679–3680. URL <http://dx.doi.org/10.1002/sim.2822>.
4. Wolbers M, Koller M. Comments on -Analysing and interpreting competing risk data- (original article and author's reply) by Melania Pintilie, *Statistics in Medicine* 2007; **26**:1360-1367; DOI: 10.1002/sim.2655 (original article) and *Statistics in Medicine* 2007; DOI: 10.1002/sim.2822 (author's reply). *Statistics in Medicine* 2007; **26**(18):3521–3523. URL <http://dx.doi.org/10.1002/sim.2904>.
5. Beyersmann J, Latouche AL, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Statistics in Medicine* 2009; **28**(6):956–971. URL <http://dx.doi.org/10.1002/sim.3516>.
6. Andersen O P. and Borgan, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer Series in Statistics. Springer, New York, 1993.
7. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd edn. Wiley, New York, 2002.
8. Crowder M. *Classical Competing Risks*. Chapman & Hall/CRC, 2001.
9. Fine J, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 1999; **94**.
10. Geskus R. Cause-specific cumulative incidence estimation and the Fine and Gray model under both left truncation and right censoring. *Biometrics* 2011; **67**:39–49.
11. Beyersmann J, Schumacher M, Allignol A. *Competing Risks and Multistate Models with R*. Springer, New York, NY, 2012.
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.