Published in final edited form as:

J Am Stat Assoc. 2019 ; 114(526): 938–947. doi:10.1080/01621459.2018.1469993.

Inverse probability weighted estimation of risk under representative interventions in observational studies

Jessica G. Young,

Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute

Roger W. Logan,

Department of Epidemiology, Harvard T.H. Chan School of Public Health

James M. Robins,

Department of Epidemiology, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Miguel A. Hernán

Department of Epidemiology, Department of Biostatistics, Harvard T.H. Chan School of Public Health, Harvard-MIT Division of Health Sciences and Technology

Abstract

Researchers are often interested in using observational data to estimate the effect on a health outcome of maintaining a continuous treatment within a pre-specified range over time; e.g. “always exercise at least 30 minutes per day”. There may be many precise interventions that could achieve this range. In this paper we consider representative interventions. These are special cases of random dynamic interventions; interventions under which treatment at each time is assigned according to a random draw from a distribution that may depend on a subject’s measured past. Estimators of risk under representative interventions on a time-varying treatment have previously been described based on g-estimation of structural nested cumulative failure time models. In this paper, we consider an alternative approach based on inverse probability weighting (IPW) of marginal structural models. In particular, we show that the risk under a representative intervention on a time-varying continuous treatment can be consistently estimated via computationally simple IPW methods traditionally used for deterministic static (i.e. “nonrandom” and “nondynamic”) interventions for binary treatments. We present an application of IPW in this setting to estimate the 28-year risk of coronary heart disease under various representative interventions on lifestyle behaviors in the Nurses Health Study.

Keywords

causal inference; marginal structural models; g-formula; longitudinal data; survival analysis

1 Introduction

We are often interested in using observational data to estimate the distribution of an outcome under a hypothetical intervention that ensures a time-varying treatment is maintained within

a pre-specified range. For example, suppose we want to estimate the 20-year risk of coronary heart disease (CHD) if all subjects exercised at least 30 minutes per day. Many interventions can be proposed to maintain this range. A simple one is “For each subject, set exercise to exactly x minutes per day” for any fixed choice of $x \geq 30$ in all follow-up intervals k . A more complex intervention is “For each subject, randomly choose daily exercise in interval k from the observed distribution of daily exercise in interval k amongst individuals who, in the observational study, (i) had the same exercise and confounder history the subject has under intervention and (ii) exercised at least 30 minutes per day in interval k ” for all k . Taubman [2007], Taubman et al. [2008] and Picciotto et al. [2012] referred to this intervention as an example of a *representative intervention*.

Representative interventions can be more generally stated as follows: “For a given subject, randomly assign treatment in interval k from the observed distribution of treatment in interval k amongst individuals who, in the observational study, (i) had the same treatment and confounder history the subject has under intervention and (ii) had treatment within the pre-specified range in interval k ” for all k . A representative intervention depends on the distribution of the observational data under analysis.

A representative intervention is a special case of a *random dynamic intervention* which is an intervention under which treatment assignment is non-deterministic and may depend on a subject’s measured past. In this paper, we show that the risk under a representative intervention on a time-varying continuous treatment can be consistently estimated via popular and computationally simple inverse probability weighting (IPW) methods for *deterministic static* (i.e. “nonrandom” and “nondynamic”) interventions on binary time-varying treatments. This result was previously noted by Taubman [2007] and Taubman et al. [2008], and, in the special case of a point treatment, by Stitelman et al. [2010]. Our formalization generalizes to the setting where treatment and confounders vary over time. As we show, the proposed IPW estimator follows directly from the identification result that the g-formula [Robins, 1986] for risk had, contrary to fact, all subjects followed a representative intervention has an equivalent IPW representation that does not depend on the distribution of treatment at each time but, instead, on the probability that treatment at each time is within the pre-specified range given the measured past.

Picciotto et al. [2012] described an alternative approach to estimating the effects of representative interventions via g-estimation of structural nested cumulative failure time models (SNCFTMs). They illustrated this approach in the Nurses Health Study to estimate the cumulative risk of coronary heart disease (CHD) under representative interventions on various lifestyle behaviors. Here we present an illustrative application of the proposed IPW approach to answer similar questions, also in the Nurses Health Study.

2 Observed data structure

Consider an observational study where measurements of treatment, covariates and disease status are updated on each of $i = 1, \dots, n$ subjects at regular intervals (e.g. every two years) over a specified follow-up period (e.g. 28 years). Let $k = 0, \dots, K + 1$ denote each measurement interval with baseline measurements taken in interval $k = 0$ and interval $K + 1$

corresponding to the specified end of follow-up. Assume the following random variables are measured in each interval for each subject (subjects are assumed independent and identically distributed and thus we suppress a subject-specific i subscript).

Let Y_k be an indicator of failure from the disease outcome of interest (e.g. CHD) by interval k with $Y_0 \equiv 0$. Let A_k be a continuous treatment variable (e.g., daily minutes of exercise) measured in interval k and L_k a vector of time-varying covariates also measured in interval k (e.g., body mass index, smoking). L_0 may also include time-fixed baseline covariates (e.g., any family history of CHD prior to baseline).

We use lower-case letters to denote possible realizations of a random variable, e.g., a_k is a possible value of treatment A_k . We denote the history of a random variable using overbars; for example, $\bar{A}_k = (A_0, \dots, A_k)$ is the observed treatment history through interval k . By notational convention, we set $\bar{L}_{-1}$ and $\bar{A}_{-1}$ to be identically 0. Also by convention, if $Y_k = 1$, we set $L_k = A_k = 0$. Within each interval, we will rely on the ordering assumption (Y_k, L_k, A_k). To focus ideas, in the main text, we assume that censoring due to incomplete or loss to follow-up and death by competing risk for the disease of interest are absent. Extensions of methodology that allow for these complications in the observed data are described in Appendices C and D.

3 Hypothetical interventions that maintain treatment within a pre-specified range

Define $R_k = I(A_k \in \mathcal{A}'_k)$ where $\mathcal{A}'_k$ is the restricted subset of $\mathcal{A}_k$, the support of A_k , containing only treatment values within the pre-specified range, $k = 0, \dots, K$. For example, given a pre-specified range of “at least 30 minutes”, $R_k = I(A_k \geq 30)$. Hernán and VanderWeele [2011] referred to the indicator variable R_k as a *compound treatment* associated with A_k . For $r_k \in \{0, 1\}$, denote a hypothetical intervention or regime that guarantees $R_k = r_k$ in all intervals $k = 0, \dots, K$ as $\bar{r}_K \equiv \bar{r}$.

We consider a regime $\bar{r} = \bar{1}$, under which treatment A_k is maintained within the pre-specified range through K . When $\mathcal{A}'_k$ contains more than a single value, the regime $\bar{r} = \bar{1}$ may be consistent with many interventions on A_k . Here we will generally define an intervention on A_k as a random draw from an intervention density $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ that may at most depend on the measured treatment, covariate and survival history $(\bar{A}_{k-1}, \bar{L}_k, \bar{Y}_k)$ for any realized covariates histories $(\bar{l}_k, \bar{a}_{k-1})$ possible under that intervention. An intervention $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ will be consistent with the regime $\bar{r} = \bar{1}$ if for all k

$$\begin{aligned} f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) &> 0, \quad \text{for } a_k \in \mathcal{A}'_k \\ f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) &= 0, \quad \text{otherwise} \end{aligned} \quad (1)$$

We classify an intervention on A_k as *deterministic* if $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ may only equal zero or one for all k and $(\bar{a}_k, \bar{l}_k)$; otherwise we classify it as *random*. We may further classify the intervention as *static* if $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ does not depend on $(\bar{l}_k, \bar{a}_{k-1})$ for any k ; otherwise we classify it as *dynamic*. To fix ideas, consider the following four interventions, all consistent with a regime $\bar{r} = \bar{1}$ that maintains daily minutes of exercise at or above 30 minutes per day in all intervals:

1. Deterministic static: “Assign all subjects exactly 30 minutes of exercise per day in interval k ”, or $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = 1$ if $a_k = 30$ and is 0 otherwise. This intervention is equivalent to the example of the first paragraph of Section 1 selecting $x = 30$.
2. Deterministic dynamic: “If a subject’s body mass index (BMI) is below 25 kg/mm³ in interval k , then assign that subject 60 minutes of daily exercise in that interval; otherwise assign that subject 30 minutes”, or, for $L_{1,k}$ the component of L_k corresponding to BMI at k ,
 - if $I_{1,k} < 25$ then $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = 1$ when $a_k = 60$ and 0 otherwise
 - if $I_{1,k} \geq 25$ then $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = 1$ when $a_k = 30$ and 0 otherwise
3. Random static: “Randomly assign daily exercise for each subject in interval k from a distribution such that the probability of receiving 30 minutes is 0.8 and the probability of receiving 60 minutes is 0.2” or $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = 0.8$ if $a_k = 30$, 0.2 if $a_k = 60$ and 0 otherwise.
4. Random dynamic: “For a subject with covariate and treatment history $(\bar{l}_k, \bar{a}_{k-1})$ under intervention, assign daily exercise in interval k as a random draw from the observed distribution of daily exercise in interval k amongst individuals who, in the observational study, (i) had history $(\bar{l}_k, \bar{a}_{k-1})$ and (ii) exercised at least 30 minutes per day in interval k ” for all k or

$$f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = f^{obs}(a_k | R_k = 1, \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) \quad (2)$$

where $f^{obs}(a_k | R_k = 1, \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ is the observed conditional density of A_k amongst those with observed covariate and treatment history $(\bar{L}_k, \bar{A}_{k-1}) = (\bar{l}_k, \bar{a}_{k-1})$ and observed treatment at k (A_k) within the pre-specified range. Note that, unlike the previous three examples, the intervention in this example is defined by characteristics of the observational study in Section 2. Following Taubman et al. [2008] and Picciotto et al. [2012], we say that a random dynamic intervention characterized by the intervention density (2) is a representative intervention on the time-varying treatment A_k .

In the next section, we review a general set of conditions under which data from an observational study can be used to identify the risk by end of follow-up had all subjects in the study population, contrary to fact, been assigned treatment according to a user-chosen $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$. By the above, a representative intervention is one special case.

4 Identifying risk under hypothetical interventions with observational data

Let $g = (g_0, \dots, g_K)$ be a *deterministic* regime, strategy, or intervention (static or dynamic) on A_k characterized by the intervention density $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}^g, \bar{Y}_k = 0) = I(a_k = a_k^g)$, where $a_s^g = g_s(\bar{\ell}_s, \bar{a}_{s-1}^g)$ is any component of $\bar{a}_k^g = (a_0^g, \dots, a_k^g)$ and $\bar{a}_s^g$ is recursively defined by the function g_s of $(\bar{l}_s, \bar{a}_{s-1}^g)$, $s = 0, \dots, k$.

Let $\bar{Y}_{K+1}^g$, $\bar{A}_{K+1}^g$ and $\bar{L}_{K+1}^g$ represent the counterfactual event, treatment, and covariate histories, respectively, under a deterministic regime g . We now define three g -specific identifying conditions for each $k = 0, \dots, K$,

1. Consistency: If $\bar{A}_k = \bar{A}_k^g$ then $\bar{Y}_{k+1} = \bar{Y}_{k+1}^g$ and $\bar{L}_{k+1} = \bar{L}_{k+1}^g$.
2. Exchangeability:

$$(Y_{k+1}^g, \dots, Y_{K+1}^g) \perp\!\!\!\perp A_k | \bar{L}_k = \bar{l}_k, \bar{A}_{k-1} = \bar{a}_{k-1}^g, Y_k = 0 \quad (3)$$

The exchangeability condition (3) is expected to hold in an experiment where the treatment A_k is physically randomized at each k possibly depending on treatment and covariate history $(\bar{L}_k, \bar{A}_{k-1})$. This condition is not, however, guaranteed to hold in an observational study and cannot be empirically examined. Exchangeability is sometimes referred to as the assumption of “no unmeasured confounders” and $\bar{L}_k$ the “measured confounder history through k ”.

3. Positivity:

$$\begin{aligned} f_{\bar{A}_{k-1}, \bar{L}_k, Y_k}(\bar{a}_{k-1}^g, \bar{l}_k, 0) &\neq 0 \Rightarrow \\ f_{A_k | \bar{L}_k, \bar{A}_{k-1}, Y_k}(a_k^g | \bar{l}_k, \bar{a}_{k-1}^g, 0) &\equiv f^{obs}(a_k^g | \bar{l}_k, \bar{a}_{k-1}^g, \bar{Y}_k = 0) > 0 \text{ w.p. } 1. \end{aligned} \quad (4)$$

where $f^{obs}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ denotes the observed treatment density, i.e., the conditional density of treatment at k in the observational study evaluated at a particular $(\bar{a}_k, \bar{l}_k)$. Note that the observed treatment density is generally distinct from the left-hand side of (2) as the former is not restricted to those with the event $R_k = 1$.

Let $\mathcal{G}$ be the set of all deterministic interventions g on A_k (both static and dynamic). Given the three conditions above hold for all $g \in \mathcal{G}$ then the risk by $K+1$ had all subjects been assigned treatment according to $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ is equivalent to the g-formula [Robins, 1986]:

$$\sum_{\bar{a}_K} \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[Y_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{Y}_k = 0] \times \prod_{j=0}^k \{ \Pr[Y_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{Y}_{j-1} = 0] \times f(l_j | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{Y}_j = 0) \times f^{int}(a_j | \bar{l}_j, \bar{a}_{j-1}, \bar{Y}_j = 0) \} \quad (5)$$

where $f(l_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{Y}_k = 0)$ and $\Pr[Y_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{Y}_k = 0]$ are the observed joint density of the confounders at k and probability of the outcome by $k+1$, respectively, conditional on past treatment, confounders, and survival to k , with $\bar{l}_j$ the first $j+1$ components of $\bar{l}_k, j=0, \dots, k, k=0, \dots, K$. Note that, we use a summation symbol in (5) and elsewhere for notational simplicity. However, in general, when $(\bar{A}_k, \bar{L}_k)$ contains any continuous components, sums would be replaced with integrals.

One minus expression (5) is the survival by $K+1$ under an intervention on A_k characterized by $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0), k=0, \dots, K$. This survival can be written as a weighted average of deterministic survival probabilities associated with the deterministic regimes $g \in \mathcal{G}$ with weights defined in terms of the choice of $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ [Robins, 1986]. The appendix of Young et al. [2014] demonstrates this equivalence using a simplified numerical example in a low-dimensional setting. By this relationship, we may succinctly denote the survival by $K+1$ had, contrary to fact, all subjects been assigned treatment according to any $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ as $\sum_{g \in \mathcal{G}} wt(g) \Pr[Y_{K+1}^g = 0]$ where $wt(g)$ is the g -specific weight.

4.1 The special case of representative interventions

Consider the special case in which we select $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ as a function of some constant $\bar{r}$:

$$f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) = f^{obs}(a_k | R_k = r_k, \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) \quad (6)$$

$k=0, \dots, K$. The right-hand side of (6) is equivalent to the right-hand side of (2) when we select $r_k = 1, k=0, \dots, K$. When $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ is selected as (6), the g-formula (5) becomes

$$\begin{aligned}
& \sum_{\bar{a}_K} \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[Y_{k+1} = 1 \mid \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{Y}_k = 0] \times \\
& \prod_{j=0}^k \{ \Pr[Y_j = 0 \mid \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{Y}_{j-1} = 0] \times \\
& f(l_j \mid \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{Y}_j = 0) \times f^{obs}(a_j \mid R_j = r_j, \bar{l}_j, \bar{a}_{j-1}, \bar{Y}_j = 0) \}
\end{aligned} \tag{7}$$

Generally, the g-formula (5) characterized by a given choice of $f^{int}(a_k \mid \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ may be re-expressed in many ways. The choice of representation is arbitrary for computing this expression nonparametrically. However, the choice of representation will impact the properties and computational complexity of parametric and semi-parametric estimators in the face of high-dimensions and finite samples.

As we will see in the next section, in the special case where $f^{int}(a_k \mid \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ is selected as (6), the following alternative representation of (7) is particularly convenient for deriving a class of IPW estimators that are easily computed with off-the-shelf statistical software. Specifically, define V as any subset of L_0 , possibly the empty set $\emptyset$. The g-formula (7) characterized by an intervention density (6) for a fixed choice of constant $\bar{r}$ can be represented as

$$\sum_v \sum_{k=0}^K \lambda_k^{\bar{r}}(v) \prod_{j=0}^{k-1} [1 - \lambda_j^{\bar{r}}(v)] f(v) \tag{8}$$

where

$$\lambda_k^{\bar{r}}(V) = \frac{E[Y_{k+1}(1 - Y_k)W_k^{\bar{r}} \mid V]}{E[(1 - Y_k)W_k^{\bar{r}} \mid V]} \tag{9}$$

and

$$W_k^{\bar{r}} = \prod_{j=0}^k \frac{I(R_j = r_j)}{f(R_j \mid \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)}, \tag{10}$$

$k = 0, \dots, K$. A proof of the equivalence between (7) and (8) is given in Appendix A which relies only on the positivity condition (4). Given the stronger assumption that all three of the above identifying conditions hold, and given we select $\bar{r} = \bar{1}$, then expressions (7) and (8) *both* identify the counterfactual risk by $K+1$ had all subjects been assigned treatment according to a representative intervention with $\lambda_k^{\bar{r}} = \bar{1}(V)$ the associated discrete time

counterfactual hazard for the outcome in interval $k + 1$ (Y_{k+1}) conditional on V , $k = 0, \dots, K$.

Note that our use of $\lambda_k^{\bar{r}}(V)$ to denote the counterfactual outcome hazard under an intervention on A_k characterized by the intervention density (6) could be more precisely denoted $\lambda_k^{(6), \bar{r}}(V)$ to explicitly specify the intervention on A_k corresponding to $\bar{r}$ is that defined by (6). The index $\bar{r}$ here represents a constant vector value that can be varied. It will facilitate our presentation of a particular marginal structural model or MSM [Robins, 2000] introduced in the next section. We will continue to use the less precise $\lambda_k^{\bar{r}}(V)$ for notational simplicity.

Remark:

As above, under a representative intervention, treatment for all subjects at k is a random draw from (2). Under a stronger exchangeability assumption than (3), the g-formula (7) for $\bar{r} = \bar{1}$ may *also* identify the risk by $K + 1$ had all subjects been assigned treatment according to a *partial* representative intervention, a special case of an intervention that depends on the natural value of treatment at k ; i.e., the value of treatment that would have been observed at k were the intervention discontinued right before k [Richardson and Robins, 2013, Young et al., 2014]. A partial representative intervention at k assigns treatment such that: (i) if a subject's natural value of treatment at k is within the prespecified range then do not intervene; (ii) otherwise, if a subjects' natural value of treatment at k is outside this range then intervene such that her treatment A_k is randomly drawn from the intervention density (2). See also Appendix F for additional discussion of these types of interventions.

Following Richardson and Robins [2013], we write this stronger exchangeability condition as follows:

Joint Exchangeability Assumption: For all deterministic static regimes $g = \bar{a}$

$$\left(Y_{k+1}^{\bar{a}}, L_{k+1}^{\bar{a}}, A_{k+1}^{\bar{a}} \right) \perp\!\!\!\perp A_k \mid \bar{L}_k = \bar{l}_k, \bar{A}_{k-1} = \bar{a}_{k-1}, Y_k = 0, \quad (11)$$

where, for any variable $(H_1, \dots, H_{K+1})$, $H_{k+1} \equiv (H_{k+1}, \dots, H_{K+1})$ and $\left(Y_{k+1}^{\bar{a}}, L_{k+1}^{\bar{a}}, A_{k+1}^{\bar{a}} \right)$ are the counterfactual values associated with the regime $\bar{a}$. Here $A_{k+1}^{\bar{a}}$ are the natural values of treatment under regime $\bar{a}$. The exchangeability assumption (11) with $L_m = (V_m, Y_m)$ also serves as a correction to and replacement for equation (9) in Picciotto et al. [2012], which was incorrect as written. See Richardson and Robins [2013] as well as the Appendix of Young et al. [2014] for examples illustrating the relative stringency of this assumption compared with (3) using causal graphs. **End of remark.**

5 Inverse probability weighted estimation of risk under a representative intervention

In low-dimensional settings with large samples we might nonparametrically estimate expression (7), or equivalently (8), evaluated at $\bar{r} = \bar{1}$ by enumerating all possible histories $(\bar{l}_K, \bar{a}_K)$ consistent with this representative intervention, calculating each component proportion, and summing over all histories. However, such an approach is not feasible in high-dimensional settings in which K is large and/or $(\bar{L}_K, \bar{A}_K)$ can take on many levels. One approach to moving forward is to impose model constraints on aspects of the underlying observed data generating process in addition to the nonparametric identifying assumptions of Section 4. Different parametric assumptions give rise to different estimators.

For example, the parametric g-formula [Robins, 1986] imposes parametric models on some or all of the components of (7) and approximates the high-dimensional sum/integral using Monte Carlo simulation based on the estimated model coefficients. Details are provided in Appendix F. The parametric g-formula can be computationally intensive and may be especially prone to bias due to model misspecification. Some misspecification can be guaranteed in typical longitudinal settings when the null is true, a problem known as the *g-null paradox* [Robins and Wasserman, 1997]; i.e., suppose that (i) the g-formula nonparametrically identifies the risk associated with a hypothetical time-varying treatment intervention (i.e., the assumptions of Section 4 hold), (ii) measured time-varying confounders at any time k are themselves affected by past treatment and (iii) the null hypothesis of no time-varying treatment effect is true; that is, the g-formula does not depend on the choice of $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$. Given (i)-(iii), it is not possible to correctly characterize the components of the g-formula by parsimonious parametric models. See Section 1.3 of Robins and Wasserman [1997] for proof.

Alternatively, we might construct an estimator that imposes parametric models on the denominator of $w_k^{\bar{r}}$ as defined in (10) and on $\lambda_k^{\bar{r}}(V)$ as defined in (9) for some chosen subset of baseline covariates V . The latter will correspond to a special case of an MSM. This approach is also subject to bias due to model misspecification and can be inefficient. Further, the sampling distribution of such an estimator can be skewed and highly variable in the face of extreme estimated weights [Robins and Wang, 2000]. This approach has the advantage that it is not subject to the g-null paradox as conditions (i)-(iii) above do not guarantee misspecification of parametric models for either (9) or the denominator of (10). It also admits a computationally simple estimation algorithm involving little more than fitting a weighted pooled logistic regression model.

In this section, we describe this estimator. We further consider its similarities to previously described IPW methods for estimating the effects of static deterministic interventions on binary time-varying treatments, both algorithmically and in terms of its asymptotic behavior. In Appendix B, we also review alternative semiparametric estimators to simple IPW estimation as they apply to representative interventions including g-estimation of SNCFTMs and doubly-robust estimators.

5.1 Marginal structural models for representative interventions

Assume there exists a real-valued parameter ψ^* such that, for user-chosen V , the following holds for all $k, \bar{r}_k$ and V :

$$h\{\gamma(k, \bar{r}_k, V; \psi)\} = \frac{E[Y_{k+1}(1 - Y_k)W_k^{\bar{r}} | V]}{E[(1 - Y_k)W_k^{\bar{r}} | V]} \quad (12)$$

When $\psi = \psi^*$ with $h(\cdot)$ a known inverse link function constrained between zero and one (e.g. logit or probit link) and γ a function of $k, \bar{r}_k$ and V and differentiable with respect to ψ .

We will say that (12) is an MSM for the discrete time hazard at $k + 1$ conditional on V under an intervention $\bar{r}$ through k where treatment is assigned according to (6)

Given the MSM (12) holds, which will be true by definition when $\gamma(k, \bar{r}_k, V; \psi)$ is selected as a saturated function of $k, \bar{r}_k$ and V , (8) may be rewritten as

$$\sum_v \sum_{k=0}^K h\{\gamma(k, \bar{r}_k, v; \psi)\} \prod_{j=0}^{k-1} [1 - h\{\gamma(j, \bar{r}_j, v; \psi)\}] f(v) \quad (13)$$

when $\psi = \psi^*$.

Remark:

In practice, saturated models will not be practical and misspecification of the MSM in the implementation of the algorithm we present in the next section may introduce bias. In theory, an MSM would be unnecessary if (i) enough subjects had data within the prespecified range throughout all follow-up times and (ii) V were selected as $\emptyset$. While (i) may be the case in some applications, there are general practical disadvantages to (ii) as we discuss in the next section. Neugebauer and van der Laan [2003] proposed the idea of a “nonparametric MSM”/“working MSM” as a way to avoid the need for the MSM assumption in the context of deterministic static interventions even in typical settings where (i) and (ii) fail. These arguments immediately extend to MSMs for random dynamic interventions. We consider the tradeoffs of this alternative perspective in Appendix B. **End of remark.**

5.2 Inverse probability weighted estimation

Let $\hat{\psi}$ be the solution to the estimating equation

$$\sum_{i=1}^n \sum_{\bar{r}} \sum_{k=0}^K U_{i,k}^{\bar{r}}(\psi, \hat{\alpha}) = 0, \quad (14)$$

with respect to ψ . Again, suppressing the i subscript, define

$$U_k^{\bar{r}}(\psi, \hat{\alpha}) = [Y_{k+1} - h\{\gamma(k, \bar{r}_k, V; \psi)\}](1 - Y_k)W_k^{\bar{r}}(\hat{\alpha})q(k, \bar{r}_k, V) \quad (15)$$

where $q(k, \bar{r}_k, V)$ may be any user-selected vector function of, at most, $k, \bar{r}_k$ and V and

$$W_k^{\bar{r}}(\hat{\alpha}) = \prod_{j=0}^k \frac{I(R_j = r_j)}{f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0; \hat{\alpha})}, \quad (16)$$

where $f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0; \alpha)$ is a model for the denominator of (10) and $\hat{\alpha}$ is the MLE of α .

Following Robins [2000], if (i) the MSM (12) is correctly specified and (ii) the model $f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0; \alpha)$ is correctly specified for the denominator of (10) then we have

$$E[U_k^{\bar{r}}(\psi^*, \alpha^*)] = 0 \quad (17)$$

for all k and $\bar{r}_k$ with α^* the true value of α and the IPW estimator $\hat{\psi}$ consistent and asymptotically normal for ψ^* . Note that (17) will hold for any choice of $q(k, \bar{r}_k, V)$. This choice, however, will generally affect the variability of the weights in finite samples.

Here we will assume pooled logistic models for both $h\{\gamma(k, \bar{r}_k, V; \psi)\}$ and $f(R_k | \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0; \alpha)$; that is,

$$h\{\gamma(k, \bar{r}_k, V; \psi)\} = \text{expit}\{\gamma(k, \bar{r}_k, V; \psi)\} \quad (18)$$

and

$$f(R_k = 1 | \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0; \alpha) = \text{expit}\{\phi(k, \bar{L}_k, \bar{A}_{k-1}; \alpha)\} \quad (19)$$

with ϕ a specified function of $(k, \bar{L}_k, \bar{A}_{k-1})$ and differentiable with respect to and

$$\text{expit}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}.$$

Given the model (18), suppose we further choose $q(k, \bar{r}_k, V) = q^*(k, \bar{r}_k, V) \frac{\partial \gamma(k, \bar{r}_k, V; \psi)}{\partial \psi^T}$ where $q^*(k, \bar{r}_k, V)$ may, analogously, be any user-selected scalar function of, at most, $k, \bar{r}_k$ and V . It follows that a solution to (14) may be obtained with off-the-shelf software via the following algorithm applied to a subject-interval data set constructed such that each subject will have

$K^* + 1$ lines indexed by $k = 0, \dots, K^*$. If, for a given subject, $Y_{K+1} = 0$ then $K^* = K$; otherwise if, for any $m = 0, \dots, K$, $Y_m = 0$ and $Y_{m+1} = 1$ then $K^* = m$.

To fix ideas, we begin by simply selecting $q^*(k, \bar{r}_k, V) = 1$. This choice of $q^*(k, \bar{r}_k, V)$ gives so-called “unstabilized weights”. In practice, other choices of $q^*(k, \bar{r}_k, V)$ are preferable as they tend to substantially reduce the variability of the sample weights and thus the influence of a few observations with extreme weights. An alternative choice is considered below giving so-called “stabilized weights”. We note that the practical reason against choosing $V = \emptyset$ in the current setting where effect modification by V is not of primary interest [Robins, 2000] is that increasing the number of baseline covariates in V will tend to reduce the variability of the corresponding stabilized weights in finite samples. This reduction in variance will generally be at the expense of an increased risk of bias due to model misspecification, particularly when the choice of V is high-dimensional.

IPW estimation algorithm for ψ :

1. Using all subject-interval records in the data set, obtain $\hat{\alpha}$ by fitting a pooled logistic regression model (19) with dependent variable R_k and independent variables a specified function of $k = 0, \dots, K$ and $(\bar{L}_k, \bar{A}_{k-1})$ corresponding to the choice of $\phi(\cdot)$.
2. For each subject, attach a *compound treatment* weight to line k of the data set calculated as

$$\prod_{j=0}^k \left[\frac{I(R_j = 1)}{\text{expit}\{\phi(j, \bar{L}_j, \bar{A}_{j-1}; \hat{\alpha})\}} + \frac{I(R_j = 0)}{1 - \text{expit}\{\phi(j, \bar{L}_j, \bar{A}_{j-1}; \hat{\alpha})\}} \right] \quad (20)$$

3. Using all subject-interval records in the data set, obtain $\hat{\psi}$ by fitting a *weighted* pooled logistic regression model with weights defined as in the previous step, dependent variable Y_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and $(\bar{R}_k, V)$ corresponding to the choice of $\chi(\cdot)$.

Note that, in constructing the compound treatment weight (20) in the second step of the above algorithm, the indicator variable in the numerator of (16) is set to one for all subject-interval observations because the MSM (12) is a model over every possible $\bar{r}$ through a given time k . Therefore, any subject’s observed compound treatment history through k will be consistent with one of these.

Stabilized weights may be constructed by the alternative choice of q^*

$$q^*(k, \bar{r}_k, V) = \prod_{j=0}^k f(r_j | \bar{r}_{j-1}, V, Y_j = 0; \alpha^{num})$$

where $f(r_j | \bar{r}_{j-1}, V, Y_j = 0; \alpha^{num})$ is a parametric model for $f(R_j | \bar{R}_{j-1}, V, Y_j = 0)$ evaluated at $\bar{r}_j$. We might again assume a pooled logistic model

$f(r_j = 1 | \bar{r}_{j-1}, V, Y_j = 0; \alpha^{num}) = \text{expit}\{\phi^{num}(j, \bar{r}_j, V; \alpha^{num})\}$. The estimate $\hat{\psi}$ remains consistent for ψ^* if this model is misspecified because, again, (17) may hold for any choice of q^* as long as it is, at most, a function of k , $\bar{r}_k$ and V .

A solution $\hat{\psi}$ to the estimating equation (14) based on this alternative choice of q^* may be obtained by the algorithm given above but with the following modifications:

1. Prior to step 2, using all subject-interval records, obtain $\hat{\alpha}^{num}$ by fitting an additional pooled logistic regression model with dependent variable R_k and independent variables a specified function of $k = 0, \dots, K$ and $(\bar{R}_{k-1}, V)$ corresponding to the choice of $\phi^{num}(\cdot)$
2. In step 2, replace (20) with the stabilized compound treatment weight

$$\prod_{j=0}^k \frac{I(R_j = 1) \text{expit}\{\phi^{num}(j, \bar{R}_{j-1}, V; \hat{\alpha}^{num})\}}{\text{expit}\{\phi(j, \bar{L}_j, \bar{A}_{j-1}; \hat{\alpha})\}} + \frac{I(R_j = 0)(1 - \text{expit}\{\phi^{num}(j, \bar{R}_{j-1}, V; \hat{\alpha}^{num})\})}{1 - \text{expit}\{\phi(j, \bar{L}_j, \bar{A}_{j-1}; \hat{\alpha})\}} \quad (21)$$

Here we write the numerator of the stabilized weight (21) as a function of $\bar{R}_k$ and not $\bar{r}_k$ because the MSM (12) is a model over every possible $\bar{r}$ through a given time k . Again, any subject's observed compound treatment history through k will be consistent with one of these.

Our final IPW estimate of the g-formula for the risk by $K+1$ had treatment been assigned according to (6) can then be obtained by the plug-in estimator

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=0}^K \text{expit}\{\gamma(k, \bar{r}_k, V_i; \hat{\psi})\} \prod_{j=0}^{k-1} [1 - \text{expit}\{\gamma(j, \bar{r}_j, V_i; \hat{\psi})\}] \quad (22)$$

When we select $\bar{r} = \bar{1}$, (22) is an IPW estimate of the g-formula characterized by a representative intervention (2).

The above estimation algorithm will be equivalent to the widely used approach described by Hernán et al. [2000] to estimate the discrete hazard at k conditional on baseline covariates under any deterministic static intervention on a binary time-varying treatment by making the following one to one replacements in the current algorithm: (i) replace the binary compound treatment history vector $\bar{R}_K$ with the binary treatment history vector $\bar{A}_K$, (ii) replace the user selected intervention $\bar{r}_K$ with a static deterministic intervention indexed by $\bar{a} \equiv \bar{a}_K$ and (iii)

replace $\bar{W}_K = (\bar{L}_K, \bar{A}_{K-1})$ with $\bar{L}_K$. This equivalence follows from the equivalence between the g-formula indexed by an intervention (6) and that indexed by a static deterministic intervention $\bar{a}$ after these one to one replacements. See Appendix B.2.

Remark:

In contrast to classical results for IPW estimators of the effects of deterministic static interventions, Haneuse and Rotnitzky [2013] recently showed that variance estimators for IPW estimators of the effects of random dynamic interventions that fail to account for estimation of nuisance parameters can be anticonservative. They show that this may occur when the user-specified intervention is characterized by an intervention density $f^{int}(a_k | \bar{L}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ that depends on knowledge of the observed treatment density $f^{obs}(a_k | \bar{L}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$. As they discuss, the problem with such choices of intervention density is that the corresponding g-formula depends on the observed treatment density which, in that case, is the nuisance parameter associated with its corresponding IPW estimator. Two recent examples of random dynamic interventions considered in the literature have this complication. These include interventions that assign treatment according to a fixed “shift” in the observed treatment density [Díaz Muñoz and van der Laan, 2012] and “threshold interventions” [Taubman et al., 2009] also defined by a function of the observed treatment density [Young et al., 2014]. By contrast, for the intervention density (6) that we consider here, the g-formula (7) characterized by this intervention density is *not* a function of the nuisance parameter associated with the corresponding IPW estimator which is $f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)$. In this case, the classical result applies for IPW estimators associated with deterministic static interventions that the variance estimator for $\hat{\psi}$

$$\sum_{i=1}^n U_i(\hat{\psi}, \hat{\alpha})^2, \quad (23)$$

which does not take into account estimation of the (model-based) nuisance parameter α , is asymptotically conservative [van der Laan and Robins, 2002] provided the compound treatment model (19) indexed by is correctly specified for $f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)$ where

$$U_i(\hat{\psi}, \hat{\alpha}) = \sum_{\bar{r}} \sum_{k=0}^K U_{i,k}^{\bar{r}}(\hat{\psi}, \hat{\alpha}). \text{ See also Appendix B.2. } \mathbf{End\ of\ remark.}$$

6 Data Analysis

The Nurses Health Study is an observational study of US female nurses who were between 36 and 61 years old in 1982. Information on various lifestyle factors and newly diagnosed diseases is updated every two years via a mailed questionnaire [Colditz et al., 1997] and new diagnoses are confirmed via medical records. We used data from the 1982 questionnaire ($k = 0$) on $n = 78,743$ Nurses Health Study participants who returned the 1982 questionnaire and had no prior history of CHD. Follow-up data was used through the 2010 questionnaire ($K + 1 = 14$) in order to estimate the 28-year risk of CHD under the following time-varying interventions:

1. All subjects smoke exactly zero cigarettes per day
2. All subjects exercise at least 30 minutes per day
3. All subjects keep diet score within the top two quintiles of the observed data
4. All subjects consume at least 5 grams of alcohol per day
5. All subjects adhere to interventions 1 – 4

Note that intervention 1 is a special case where the pre-specified range contains a single value. Here, the g-formula characterized by a representative intervention is equivalent to that characterized by a static deterministic intervention where treatment is always set to this single value. Subjects were censored due to incomplete follow-up upon first failure to return a questionnaire at the expected time of return.

We used the IPW algorithm described in Section 5.2 with the addition of stabilized censoring weights as described in Appendix C and extensions to the competing risk setting as in Appendix D. Risk ratios were constructed relative to the estimated 28-year CHD risk under the “natural course”; i.e. under an intervention that “abolishes censoring” only. Details are provided in Appendix E.

We defined V in terms of the following baseline covariates: any family history of CHD, smoking history, prior oral contraceptive use, BMI at age 18, age at baseline, and 1980 measurements of BMI, daily physical activity, diet score and alcohol use. For $k = 0, \dots, K$, interval k confounders L_k included cancer, angina, coronary artery bypass grafting (CABG), osteoporosis, menopause, high cholesterol, stroke, diabetes and high blood pressure by that interval and any statin, aspirin, multivitamin and postmenopausal hormone use during that interval.

The remaining components of L_k and A_k were assigned intervention-specific definitions. For example, under intervention 1, A_k was defined in terms of interval k daily smoking habits while L_k additionally contained interval k measurements of daily exercise, diet score and alcohol use. By contrast, under intervention 2, A_k was defined in terms of interval k daily minutes of physical activity while L_k additionally contained interval k measurements of daily smoking habits, diet score and alcohol use. Under the joint intervention 5, A_k was defined as a vector of the interval k measurements of smoking, exercise, diet score and alcohol use, R_k a vector of the corresponding compound treatments and L_k all remaining interval k covariates. See Section 6.3 for analysis details for the joint intervention. Robins et al. [2004] considered conditions under which the exchangeability assumption (3) could hold simultaneously for all definitions of A_k and L_k .

6.1 Weight models

In constructing the denominator of the stabilized compound treatment weights (21) for interventions 1-4 we assumed:

$$\phi(k, \bar{L}_k, \bar{A}_{k-1}; \alpha) = \alpha_0 + \alpha_1^T g(k) + \phi^\sim(\bar{L}_k, \bar{A}_{k-1}; \alpha^\sim) \quad (24)$$

where $g(k) = \{I(k=0), I(k=1), \dots, I(k=13)\}$. Due to the high dimension and intervention-specific definition of $(\bar{L}_k, \bar{A}_k)$, we summarize in Tables G.1 and G.2 of Appendix G the functional form of $\phi(\cdot)$ of equation (24). Measurements of some behaviors were not measured for any subject in some follow-up intervals k . For these intervals, missing covariate values were carried forward from the last interval of actual measurement. By this, certain components of the vector of measured confounders at k (L_k) for a given effect will be last measured values of a covariate relative to interval k . This may result in unmeasured confounding as the exchangeability assumption (3) of Section 4 must hold for this definition of L_k that fails to contain the full history of that covariate. Further, for exposures not measured in all intervals, our effect estimates cannot be interpreted in terms of sustained interventions but rather interventions that occur at times in which the exposure is measured. Tables G.1 and G.2 also list the years assessed for each time-varying covariate.

In constructing the numerator of the stabilized compound treatment weights for interventions 1-4 we defined:

$$\phi^{num}(k, \bar{R}_{k-1}, V; \alpha^{num}) = \alpha_0^{num} + \alpha_1^{num} \top g(k) + \alpha_2^{num} \top V + \alpha_3^{num} \top H_{k-1} \quad (25)$$

with $H_{k-1} = (R_{k-1}, R_{k-2})$ for interventions defined by treatments actually measured in all intervals. Otherwise we defined $H_{k-1} = (R_{k-1}, R_{k-1} \times ts_{k-1})$ where ts_{k-1} is the number of intervals passed between $k-1$ and the last measurement of treatment.

The fitting of numerator and denominator models for the stabilized compound treatment weights was restricted only to interval records where A_k corresponded to an actual measurement and not a carried forward value. Here, $f(R_k | \bar{L}_k, \bar{A}_{k-1}, \bar{V}_k = 0) = 1$ for all subjects; that is, given any subject's observed treatment history $\bar{A}_{k-1}$, we observe her realized value of R_k with probability one. For this reason, the stabilized compound treatment weights were set to one for interval records corresponding to a carried forward value of A_k .

IPW estimates reported in Table 2 were based on stabilized weights truncated at the 99th percentile to reduce the influence of extreme weights [Cole and Hernán, 2008, Petersen et al., 2012]. The distributions of the stabilized weights prior to truncation are summarized in Table 1 by intervention. As expected, the distribution of the unstabilized weights were substantially more variable for our choice of V as were both stabilized and unstabilized weights when no covariates were included in V ; i.e. when we redefined $V = \emptyset$ (results not shown).

6.2 Marginal structural models

We assumed the following functional form for γ of the MSM (18) for interventions 1-4. Specifically:

$$\gamma(k, \bar{R}_k, V; \psi) = \gamma_0(k, V; \psi_0) + \gamma_1(k, \bar{R}_k, V; \psi_1) \quad (26)$$

where

$$\gamma_0(k, V; \psi_0) = \psi_{00} + \psi_{01}^T V + \psi_{02}^T g(k)$$

and

$$\gamma_1(k, \bar{R}_k, V; \psi_1) = \psi_{11} R_k + \psi_{12} R_{k-1} \quad (27)$$

6.3 Joint interventions

For the joint intervention 5, we define both the compound and actual time-varying treatments in terms of the vectors $R_k = (R_{1,k}, R_{2,k}, R_{3,k}, R_{4,k})$ and $A_k = (A_{1,k}, A_{2,k}, A_{3,k}, A_{4,k})$, respectively.

Estimation proceeds with the algorithm of Section 5.2 but replacing the pooled logistic model (19) with a model that relies on the factorization

$$\begin{aligned} f(R_k | \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0) &= f(R_{1,k} | \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0) \\ &\times f(R_{2,k} | R_{1,k}, \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0) \times f(R_{3,k} | R_{2,k}, R_{1,k}, \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0) \\ &\times f(R_{4,k} | R_{3,k}, R_{2,k}, R_{1,k}, \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0) \end{aligned} \quad (28)$$

Pooled logistic models were fit for each component of the product in (28) and the denominator of the stabilized compound treatment weight (21) redefined accordingly. The functional form of the MSM was as in (26) but with $\gamma_1(k, \bar{R}_k, V; \psi_1)$ replaced by

$$\sum_{j=1}^4 \gamma_{j,1}(k, \bar{R}_{j,k}, V; \psi_{j,1}). \text{ Analogous to (27), each } \gamma_{j,1} \text{ was selected as } \psi_{j,11} R_{j,k} + \psi_{j,12} R_{j,k-1}.$$

6.4 Results

Four percent of subjects had a recorded CHD event during the 28-year follow-up. Twenty-six percent of subjects were censored due to incomplete follow-up (after an average follow-up of 10 years) and 14 percent died of competing risk. Table 2 presents IPW estimates based on the specifications detailed above and in Appendices C, D and E. Ninety-five percent confidence intervals were obtained via 500 nonparametric bootstrap samples. Table 3 presents the sensitivity of the IPW estimates to the choice of truncation level for the stabilized weights: 99th percentile (as in Table 2), 95th percentile, and the fixed value 10. Table 4 further presents sensitivity of the IPW estimates to the choice of functional form of the MSM – specifically, the choice of $\gamma_1(k, \bar{R}_k, V; \psi_1)$ in the model (26) – along with risk estimates obtained via the parametric g-formula for the five interventions (details provided in Appendix F) which does not rely on an MSM assumption.

Relative risk estimates in Table 2 for individual interventions 1 through 4 relative to the natural course are similar to those given for these same comparisons in Table 2 of Picciotto et al. [2012] based on g-estimation of SCNFTMs. The g-estimates were based on the same

Nurses Health Study data set used here but restricted through 20-year (as opposed to 28-year) follow-up. The g-estimates of these four relative risks and 95% CIs based on a nonparametric bootstrap were 0.8(0.7, 0.8), 0.9(0.8, 1.), 1.0(0.9, 1.1) and 0.9(0.8, 1.0), respectively. Picciotto et al. [2012] did not estimate effects of joint interventions.

Tables 3 and 4 indicate that, for the individual interventions 1 through 4, IPW estimates are not particularly sensitive to the method of weight truncation nor the functional form of the MSM. Further, for these interventions, IPW and parametric g-formula risk estimates are qualitatively similar. However, the joint intervention risk estimates show greater sensitivity to arbitrary implementation choices for the IPW estimator and greater differences between IPW estimation and the parametric g-formula. In the case of the joint intervention 5, both the parametric g-formula and the IPW estimator are most heavily relying on extrapolation compared with interventions 1 through 4 as very few subjects are actually following this intervention at any time in the Nurses Health Study. In particular, the percentage of surviving subjects with behaviors consistent with the pre-specified ranges of the individual behavioral interventions 1-4 in any given interval ranged (over time) between 72 and 95, 13 and 21, one and 54, and 28 and 35 percent, respectively. By contrast, this percentage for the joint intervention 5 ranged between only zero and six percent.

7 Conclusion

In this paper, we showed that previously developed algorithms and theory for IPW estimators of failure risk under a deterministic static intervention on a time-varying binary treatment also apply to IPW estimators of failure risk under a representative intervention on a time-varying treatment; a special case of a random dynamic intervention that maintains a continuous treatment within a prespecified range. This work extends results given in Stitelman et al. [2010] for point treatments to the general setting of time-varying treatments and confounders. As noted by Stitelman et al. [2010], these results allow for a more precise interpretation of IPW analyses that replace a continuously measured treatment variable with a dichotomized indicator that treatment is within the prespecified range.

Due to the reality that few subjects may be following a given representative intervention of interest through each follow-up time, we considered estimators that rely on a correctly specified MSM. We used IPW of MSMs to estimate 28-year risk under representative interventions on time-varying behaviors in the Nurses Health Study and considered sensitivity of results to the choice of MSM. Results were particularly sensitive to the choice of MSM for joint interventions consistent with multiple “healthy” behaviors, followed by few subjects in the data set. Results were not particularly sensitive to this choice for individual behavioral interventions and similar to those we previously reported in the Nurses Health Study using g-estimation of SNCFTMs for the first 20-years of follow-up [Picciotto et al., 2012].

Acknowledgments

This study was supported by grants R01 HL080644, UM1 CA186107, R01 HL034594 and R01 HL088521 from the National Institutes of Health (Bethesda, MD). The authors thank the Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School. The authors are also grateful to Sarah Taubman and Murray Mittleman for their contributions and comments to early drafts of this manuscript.

Appendix A

In this section we prove that, provided the positivity condition (4) of the main text holds for all deterministic regimes g , expressions (7) and (8) are equivalent. We will rely on the following Lemma.

Lemma A.1

$$E[(1 - Y_j)W_j^{int} | V] = E[(1 - Y_{j-1})W_{j-1}^{int} | V] - E[Y_j(1 - Y_{j-1})W_{j-1}^{int} | V]$$

where

$$W_j^{int} = \prod_{m=0}^j \frac{f^{int}(A_m | \bar{L}_m, \bar{A}_{m-1}, \bar{Y}_m = 0)}{f^{obs}(A_m | \bar{L}_m, \bar{A}_{m-1}, \bar{Y}_m = 0)} \quad (\text{A.1})$$

$j = 0, \dots, k$ and $k = 0, \dots, K$.

Proof: First, given positivity (4) we have

$$\begin{aligned} E[(1 - Y_j)W_j^{int} | V] &= E\left[(1 - Y_j)W_{j-1}^{int} E\left[\frac{f^{int}(A_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)}{f^{obs}(A_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} | \bar{L}_j, \bar{A}_{j-1}\right] | V\right] \\ &= E[(1 - Y_j)W_{j-1}^{int} | V] \end{aligned}$$

Further, by the event $Y_j = 0$ implying the joint event $(Y_j = 0, Y_{j-1} = 0)$ we have

$$\begin{aligned} E[(1 - Y_j)W_{j-1}^{int} | V] &= E[(1 - Y_j)(1 - Y_{j-1})W_{j-1}^{int} | V] \\ &= E[(1 - Y_{j-1})W_{j-1}^{int} - Y_j(1 - Y_{j-1})W_{j-1}^{int} | V] \\ &= E[(1 - Y_{j-1})W_{j-1}^{int} | V] - E[Y_j(1 - Y_{j-1})W_{j-1}^{int} | V] \end{aligned}$$

Theorem A.2 Define

$$\lambda_k^{int}(V) = \frac{E[Y_{k+1}(1 - Y_k)W_k^{int} | V]}{E[(1 - Y_k)W_k^{int} | V]} \quad (\text{A.2})$$

It follows that the g -formula (5) characterized by $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ is equivalent to

$$\sum_v \sum_{k=0}^K \lambda_k^{int}(v) \prod_{j=0}^{k-1} [1 - \lambda_j^{int}(v)] f(v) \quad (\text{A.3})$$

Proof: By the definition (A.2), we can rewrite (A.3) as

$$\sum_v \sum_{k=0}^K E[Y_{k+1}(1 - Y_k)W_k^{int} | V = v] \times \prod_{j=1}^k \frac{E[(1 - Y_{j-1})W_{j-1}^{int} | V = v] - E[Y_j(1 - Y_{j-1})W_{j-1}^{int} | V = v]}{E[(1 - Y_j)W_j^{int} | V = v]} f(v) \quad (\text{A.4})$$

By Lemma A.1, (A.4) reduces to

$$\sum_v \sum_{k=0}^K E[Y_{k+1}(1 - Y_k)W_k^{int} | V = v] f(v) \quad (\text{A.5})$$

which, by the laws of expectations and positivity (4), is equivalent to the g-formula (5).

Our final result follows.

Corollary A.3 *The g-formula (7) is equivalent to (8).*

Proof: Theorem A.2 immediately gives equivalence between expression (7) and (A.3) with the general density function $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ in the numerator of the weight (A.1) selected as (6). It remains only to show that (A.3) characterized by (6) is equivalent to (8). This will hold if equivalence between (A.2) characterized by (6) and the right hand side of (9) holds which we now show.

In the following, X_k may be either $Y_{k+1}(1 - Y_k)$ or $(1 - Y_k)$. For $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ selected as (6) we have

$$\begin{aligned} E[X_k W_k^{int} | V] &= E \left[X_k \prod_{j=0}^k \frac{f^{obs}(A_j | R_j = r_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)}{f^{obs}(A_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} \mid V \right] \\ &= E \left[X_k \prod_{j=0}^k \frac{f^{obs}(A_j | R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0) I(R_j = r_j)}{f^{obs}(A_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} \mid V \right] \end{aligned} \quad (\text{A.6})$$

By R_j a coursening of A_j , the event A_j implies the joint event (A_j, R_j) . Thus (A.6) can be equivalently written as

$$\begin{aligned}
& \mathbb{E} \left[X_k \prod_{j=0}^k \frac{f^{obs}(A_j | R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0) I(R_j = r_j)}{f^{obs}(A_j | R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} \mid V \right] \\
&= \mathbb{E} \left[X_k \prod_{j=0}^k \frac{f^{obs}(A_j | R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0) I(R_j = r_j)}{f^{obs}(A_j | R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0) f(R_j, \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} \mid V \right] \\
&= \mathbb{E} \left[X_k \prod_{j=0}^k \frac{I(R_j = r_j)}{f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} \mid V \right]
\end{aligned}$$

Appendix B

B.1 G-estimation of structural nested cumulative failure time models

Picciotto et al. [2012] described g-estimation of structural nested cumulative failure time models (SNCFTMs) for estimating risk by an end of follow-up time under time-varying representative interventions, also with application to the Nurses Health Study. As in the case of IPW methods, this approach relies on a correctly specified model for $f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)$. In place of an MSM it relies on a correctly specified SNCFTM such that, for some real-valued parameter $\psi^{*,sncftm}$, the following holds for all $k = 0, \dots, k+1$, $m = 0, \dots, k$, and $(R_m, \bar{A}_{m-1}, \bar{L}_m)$:

$$\exp\{\gamma^{sncftm}(k, m, R_m, \bar{A}_{m-1}, \bar{L}_m; \psi^{sncftm})\} = \frac{\mathbb{E}[Y_k^{(\bar{R}_m, 0)} | \bar{A}_{m-1}, \bar{L}_m, Y_m = 0]}{\mathbb{E}[Y_k^{(\bar{R}_{m-1}, 0)} | \bar{A}_{m-1}, \bar{L}_m, Y_m = 0]} \quad (\text{B.1})$$

when $\psi^{sncftm} = \psi^{*,sncftm}$ and $Y_m = 0$ where $Y_k^{(\bar{R}_m, 0)}$ is the outcome a subject would have experienced by k had, contrary to fact, she followed an intervention (6) where at each time $s = 0, \dots, m$, r_s is set to the observed R_s and set to zero at all $m+1 = 0, \dots, k$. Similarly, $Y_k^{(\bar{R}_{m-1}, 0)}$ is the counterfactual associated with an intervention (6) where, at each time $s = 0, \dots, m-1$, r_s is set to the observed R_s and set to zero at all $m = 0, \dots, k$. Picciotto et al. [2012] provide an algorithm for estimating risk by end of follow-up under a representative intervention (2) as a function of the g-estimate of the SNCFTM parameter ψ^{sncftm} . The g-estimate is a solution to an estimating equation which may be obtained by optimization procedures such as Newton-Raphson or Nelder-Mead Simplex. We refer the reader to Picciotto et al. [2012] for details.

In addition to the nonparametric identifying assumptions of Section 4 in the main text, this algorithm for risk further relies on the constraint that the SNCFTM does not depend on the time-varying covariate history $(\bar{A}_{m-1}, \bar{L}_m)$; see proof in Appendix B of Picciotto et al. [2012]. Like the IPW methods considered in the main text, this approach does not suffer from the g-null paradox and further avoids the possibility of extreme weights. We have previously shown conditions for equivalence between an SNCFTM and MSM in the context

of survival settings and provided a simulation study comparing g-estimation and IPW under these conditions, including the condition that the SCNFTM does not depend on time-varying covariates [Young et al., 2008].

B.2 Doubly robust extensions of IPW

The algorithmic and theoretical similarities between IPW estimators of risk under representative interventions on a continuous time-varying treatment and deterministic static interventions on a binary time-varying treatment generalize to doubly-robust extensions of IPW that solve the efficient influence curve. This allows reliance on the extensive theory already developed by Robins and colleagues for deterministic static interventions on a binary treatment [Robins and Rotnitzky, 1992, Scharfstein et al., 1999, van der Laan and Robins, 2002, Bang and Robins, 2005, van der Laan and Gruber, 2012] in order to derive doubly-robust estimators in the context of representative interventions on a continuous treatment. Following this earlier work, Schnitzer et al. [2014] recently proposed a doubly-robust estimator of cumulative survival by $k + 1$ under a deterministic static intervention on a time-varying binary treatment $\bar{a}$ as nonparametrically identified by the complement of the g-formula (5) for $k = K$ and the intervention density chosen as $\mathbb{I}(A_k = a_k)$, $k = 0, \dots, K$. Denote this function of the data $S_k^{\bar{a}}$ with $S_K^{\bar{a}} \equiv S^{\bar{a}}$. Following the above cited work, the efficient influence curve for the parameter $S^{\bar{a}}$ is

$$D^{\bar{a}} = \sum_{k=K+1}^0 D_k^{\bar{a}} \quad (\text{B.2})$$

where

$$D_0^{\bar{a}} = Q_0 - S^{\bar{a}} \quad (\text{B.3})$$

and, for $k = K + 1, \dots, 1$

$$D_k^{\bar{a}} = \prod_{j=0}^{k-1} \frac{I(A_j = a_j)}{f_k^{obs}(A_j | \bar{L}_j, \bar{A}_{j-1}, \bar{Y}_j = 0)} [Q_k - Q_{k-1}] \quad (\text{B.4})$$

with $Q_{K+1} = 1 - Y_{K+1}$, $Q_K = \Pr[Y_{K+1} = 0 | \bar{L}_K, \bar{A}_K = \bar{a}_K, \bar{Y}_K]$ (equal to 0 for $Y_K = 1$) and, for $k = K, \dots, 1$ $Q_{k-1} = \mathbb{E}[Q_k | \bar{L}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{Y}_{k-1}]$ (equal to 0 for $Y_{k-1} = 1$). Let $f_k^{obs} \equiv f_k^{obs}(A_k | \bar{L}_k, \bar{A}_{k-1}, \bar{Y}_k = 0)$, $k = 0, \dots, K$. The estimator $\hat{S}^{\bar{a}}$ that solves $D^{\bar{a}} \equiv D^{\bar{a}}(S^{\bar{a}}, \hat{Q}_0, \dots, \hat{Q}_{K+1}, \hat{f}_0^{obs}, \dots, \hat{f}_K^{obs}) = 0$ with respect to $S^{\bar{a}}$ is doubly-robust in that this estimator is consistent for $S^{\bar{a}}$ if either $\hat{Q}_0, \dots, \hat{Q}_{K+1}$ are consistent for $Q_0, \dots, Q_{K+1}$ or $\hat{f}_0^{obs}, \dots, \hat{f}_K^{obs}$ are consistent for $f_0^{obs}, \dots, f_K^{obs}$. Petersen et al. [2014] recently gave a doubly-

robust estimator for the parameters of an MSM for $S_k^{\bar{a}}$ at any given $k = 0, \dots, K$ possibly within levels of a specified subset of baseline covariates V , to accommodate many realistic settings (as we have considered in the main text in the context of the simple IPW estimator), where few subjects may be following a single intervention of interest through any given follow-up time and an MSM is practically required to borrow information across interventions and times. This approach, which solves the efficient influence curve for the parameters of the MSM, also relies on a user-supplied choice of function q – at most a function of time, intervention and V . Orellana et al. [2010a,b] have described approaches for finding the efficient choice of q for doubly-robust estimators.

We now prove that a doubly robust estimator of survival by $K + 1$ under an intervention (6) on a time-varying continuous treatment – as nonparametrically identified by the complement of (7) under assumptions given in the main text – is equivalent to a doubly-robust estimator of $S^{\bar{a}}$ but with (i) the binary treatment history vector $\bar{A}_K$ in the latter replaced by the binary compound treatment history vector $\bar{R}_K$, (ii) the user-selected binary treatment intervention $\bar{a}_K$ in the latter replaced by a user-selected $\bar{r}_K$ and (iii) the confounder history $\bar{L}_K$ in the latter replaced by an augmented version including the continuous treatment history through $K - 1$ $\bar{W}_K = (\bar{L}_K, \bar{A}_{K-1})$. Note, this also serves as a proof of claims made in the Remark at the end of Section 5.2 in the main text on asymptotic properties of the simple IPW estimator for representative interventions.

Proof: By (5) and making $\bar{Y}_K$ an implicit component of $\bar{L}_K$, we can write $S^{\bar{a}}$ as

$$\sum_{l_K} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_K = \bar{a}_K] \prod_{k=0}^K f(l_k \mid \bar{l}_{k-1}, \bar{a}_{k-1}) \quad (\text{B.5})$$

With replacements (i) and (ii) (B.5) becomes

$$\sum_{l_K} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{R}_K = \bar{r}_K] \prod_{k=0}^K f(l_k \mid \bar{l}_{k-1}, \bar{r}_{k-1}) \quad (\text{B.6})$$

With the additional replacement (iii) (B.6) becomes

$$\begin{aligned}
& \sum_{\bar{w}_K} \Pr[Y_{K+1} = 0 \mid \bar{W}_K = \bar{w}_K, \bar{R}_K = \bar{r}_K] \prod_{k=0}^K f(w_k \mid \bar{w}_{k-1}, \bar{r}_{k-1}) = \\
& \sum_{l_K} \sum_{\bar{a}_{K-1}} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_{K-1} = \bar{a}_{K-1}, \bar{R}_K = \bar{r}_K] \\
& \prod_{k=0}^K f(l_k, a_{k-1} \mid \bar{l}_{k-1}, \bar{a}_{k-2}, \bar{r}_{k-1})
\end{aligned} \tag{B.7}$$

Denote this final expression (B.7) $S^{\bar{r}}$. Comparing expressions (B.5) and (B.7) it is clear that $S^{\bar{r}}$ is the g-formula associated with a static deterministic intervention on a binary treatment but where we call the “treatment” R_k (rather than A_k) and we call the “measured confounder history” $\bar{W}_k$ (rather than $\bar{L}_k$). To prove the desired result, it then remains only to show that $S^{\bar{r}}$ is equivalent to the g-formula for survival by $K+1$ under an intervention (6) as defined by the complement of (7). By probability laws, we have that $S^{\bar{r}}$ is equal to

$$\begin{aligned}
& \sum_{l_K} \sum_{\bar{a}_{K-1}} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_{K-1} = \bar{a}_{K-1}, \bar{R}_K = \bar{r}_K] \\
& \prod_{k=0}^K f(l_k \mid \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{r}_{k-1}) f(a_{k-1} \mid \bar{l}_{k-1}, \bar{a}_{k-2}, \bar{r}_{k-1}) = \\
& \sum_{l_K} \sum_{\bar{a}_K} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_K = \bar{a}_K, \bar{R}_K = \bar{r}_K] \\
& f(a_K \mid \bar{r}_K, \bar{l}_K, \bar{a}_{K-1}) \prod_{k=0}^K f(l_k \mid \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{r}_{k-1}) f(a_{k-1} \mid \bar{l}_{k-1}, \bar{a}_{k-2}, \bar{r}_{k-1}) = \\
& \sum_{l_K} \sum_{\bar{a}_K} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_K = \bar{a}_K, \bar{R}_K = \bar{r}_K] \\
& \prod_{k=0}^K f(a_k \mid \bar{l}_k, \bar{a}_{k-1}, \bar{r}_k) f(l_k \mid \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{r}_{k-1}) =
\end{aligned} \tag{B.8}$$

$$\begin{aligned}
& \sum_{l_K} \sum_{\bar{a}_K} \Pr[Y_{K+1} = 0 \mid \bar{L}_K = \bar{l}_K, \bar{A}_K = \bar{a}_K] \\
& \prod_{k=0}^K f(a_k \mid r_k, \bar{l}_k, \bar{a}_{k-1}) f(l_k \mid \bar{l}_{k-1}, \bar{a}_{k-1})
\end{aligned} \tag{B.9}$$

To understand the last equality, let us explicitly consider the case where we select $\bar{r}_K \equiv \bar{r} = \bar{1}$ (i.e., the case of representative interventions). In this case, any component of the sum (B.8) will be zero for any $\bar{a}_K$ with component $a_k \notin \mathcal{A}'_k$, $k = 0, \dots, K$ by the fact that this sum is weighted by $f(a_k | \bar{r}_k = \bar{1}_k, \bar{l}_k, \bar{a}_{k-1})$. In all remaining non-zero components of this sum—that is, the sum over all histories $\bar{a}_K$ consistent with $\bar{r}_K = \bar{1}$ —we have that the event $(\bar{A}_k = \bar{a}_k)$ equals the joint event $(\bar{A}_k = \bar{a}_k, \bar{R}_k = \bar{r}_k)$, $k = 0, \dots, K$.

Explicitly pulling $\bar{Y}_K$ out of $\bar{L}_K$, noting that if any component of $\bar{Y}_K$ is one then we must have $Y_{K+1} = 1$ and noting $f^{obs}(a_k | R_k = r_k, \bar{l}_k, \bar{a}_{k-1}) \equiv f(\bar{a}_k | r_k, \bar{l}_k, \bar{a}_{k-1})$ (B.9) can be equivalently written as

$$\sum_{\bar{l}_K} \sum_{\bar{a}_K} \Pr[Y_{K+1} = 0 | \bar{L}_K = \bar{l}_K, \bar{A}_K = \bar{a}_K, \bar{Y}_K = 0] \prod_{k=0}^K f^{obs}(a_k | R_k = r_k, \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0) \\ f(l_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{Y}_k = 0) \Pr[Y_k = 0 | \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_{k-1} = \bar{l}_{k-1}, \bar{Y}_{k-1} = 0] \quad (\text{B.10})$$

Expression (B.10) is the complement of (7) which, by results of the main text, is equal to the g-formula for risk by $K + 1$ under a representative intervention when we select $r_k = 1$, $k = 0, \dots, K$. *End Proof*

B.3 Working marginal structural models

Following Neugebauer and van der Laan [2003], as a way to avoid reliance on a correctly specified MSM, Petersen et al. [2014] formally redefined their parameter of interest, rather than as a counterfactual survival function, as a projection specific to the choice of MSM and the choice of function q onto the true data functional; i.e. the g-formula identifying this counterfactual survival under nonparametric identifying assumptions. They refer to this projection as a “working MSM”.

An advantage of this alternative framework is that the consistency and (q -specific) efficiency properties of the TMLE estimator of this projection hold even when the MSM is a misspecified model for this true data functional. However, if our goal is not to estimate this projection but rather causal effects of interventions, we do require that the MSM is correctly specified and misspecification of this model is a potential source of bias within the context of any estimator imposing an MSM constraint. Some authors have proposed different data adaptive strategies for this problem [Haight et al., 2010, Platt et al., 2013, Taguri and Matsuyama, 2013]. In our application in Section 6 in the main text, we consider sensitivity of our results to this assumption.

Appendix C

Define C_{k+1} as the indicator that a subject is censored (e.g. due to incomplete follow-up) by interval $k + 1$, $k = 0, \dots, K$ with $C_0 \equiv 0$ by definition. Further let $\bar{Z}_k = (\bar{A}_k, \bar{L}_k)$ for all k .

Following previous authors [Robins and Rotnitzky, 1995, Hernán et al., 2000], we can generalize the estimation algorithm of Section 5 of the main text to account for the presence of censoring by multiplying the stabilized compound treatment weight (21) of the main text by a stabilized censoring weight

$$\prod_{j=0}^k \frac{\text{expit}\{\phi^{C,num}(j, \bar{R}_j, V; \hat{\alpha}^{C,num})\}}{\text{expit}\{\phi^C(j, \bar{Z}_j; \hat{\alpha}^C)\}} \quad (\text{C.1})$$

where $\text{expit}\{\phi^{C,num}(j, \bar{R}_j, V; \alpha^{C,num})\}$ is a pooled logistic model for

$\Pr[C_{j+1} = 0 \mid \bar{R}_j, V, \bar{C}_j = \bar{Y}_j = 0]$ with $\phi^{C,num}$ a specified function of $(j, \bar{R}_j, V)$ differentiable with respect to $\alpha^{C,num}$ and $\hat{\alpha}^{C,num}$ the MLE of $\alpha^{C,num}$. Analogously, $\text{expit}\{\phi^C(j, \bar{Z}_j; \alpha^C)\}$ is a pooled logistic model for $\Pr[C_{j+1} = 0 \mid \bar{Z}_j, \bar{C}_j = \bar{Y}_j = 0]$ with ϕ^C a specified function of $(j, \bar{Z}_j)$ differentiable with respect to α^C and $\hat{\alpha}^C$ the MLE of α^C . Following arguments given in Section 5.2 of the main text, only the denominator model need be correctly specified.

The generalized algorithm is applied to a subject-interval data set with a modified definition of K^* : If, for a given subject, $C_{K+1} = Y_{K+1} = 0$ then $K^* = K$; otherwise if, for any $m = 0, \dots, K$, $C_m = Y_m = 0$ and either $Y_{m+1} = 1$ or $C_{m+1} = 1$ then $K^* = m$. Under this generalization, step 3 of the IPW estimation algorithm in Section 5.2 also must restrict to records with $C_{k+1} = 0$. This can automatically be achieved in many statistical software packages by setting Y_{k+1} to missing if $C_{k+1} = 1$ in the input data set to the algorithm on each line k .

Provided C_{k+1} and the event $C_{k+1} = 0$ are, respectively, considered implicit components of A_k and the event $\bar{A}_k = \bar{a}_k^0$ then, given the identifying assumptions of Section 4 of the main text, the IPW estimate (22) evaluated at $\bar{r} = \bar{1}$ with the addition of the stabilized censoring weight (C.1) will be consistent for the risk by $K + 1$ had all subjects followed a representative intervention *and* had censoring been abolished. Under this generalization of the identifying assumptions, we rely on a more general interpretation of exchangeability equivalent to the assumption of “no unmeasured confounding or selection bias” with $\bar{L}_k$ the “measured confounders and selection factors through k ”.

In constructing the denominator of the stabilized censoring weights (C.1) in the application presented in Section 6 of the main text, we assumed:

$$\phi^C(k, \bar{Z}_k; \alpha^C) = \alpha_0^C + \alpha_1^C \text{I}_{g(k)} + \phi^C(\bar{Z}_k; \alpha^C) \quad (\text{C.2})$$

where $g(k) = \{I(k=0), I(k=1), \dots, I(k=13)\}$. Due to the high dimension of $\bar{Z}_k$, we summarize in Tables G.1 and G.2 of Appendix G the functional form of $\bar{Z}_k$ as included in $\phi^{C\sim}(\cdot)$ of equation (C.2).

In constructing the numerator of the stabilized censoring weights we defined:

$$\phi^{C,num}(k, \bar{R}_k, V; \alpha^{C,num}) = \alpha_0^{C,num} + \alpha_1^{C,num} \mathbb{I}_{g(k)} + \alpha_2^{C,num} \mathbb{I}_V + \alpha_3^{C,num} \mathbb{I}_{H_k} \quad (C.3)$$

with $H_k = (R_k, R_{k-1}, R_{k-2})$ for interventions defined by treatments actually measured in all intervals. Otherwise, we defined $H_k = (R_k, R_k \times ts_k, R_{k-1}, R_{k-1} \times ts_{k-1})$ where ts_k is the number of intervals passed between k and the last treatment measurement.

Appendix D

In this section we describe a generalization of the IPW algorithm in the main text to allow for the presence of death by competing risk. The estimator described in this section is similar to that described by Moodie et al. [2014] for estimating the effects of deterministic static interventions on a time-varying binary treatment in the presence of competing risks.

Denote D_{k+1} as an indicator of death by competing risk by interval $k+1$, $k=0, \dots, K$ with $D_0=0$ by definition. The generalized algorithm described below that incorporates data on $\bar{D}_{K+1}$ is based on a modified causal estimand. We now consider what may be viewed as a discrete time subdistribution cumulative incidence function or CIF [Kalbfleisch and Prentice, 1980] for the event of interest by $K+1$ under an intervention where treatment at each k is assigned according to an intervention density $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0)$ and censoring at each k is abolished. This may be equivalently conceptualized as the risk of the event by $K+1$ had treatment and censoring undergone these interventions and death by competing risk still occurred according to its observed distribution. Heretofore, we will refer to this counterfactual quantity simply as the risk by $K+1$ under an intervention density $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0)$.

Following the appendix in Taubman et al. [2009], the g-formula associated with this risk can be written as

$$\begin{aligned} & \sum_{\bar{a}_K} \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[Y_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{Y}_k = \bar{D}_{k+1}, \bar{C}_{k+1} = 0] \times \\ & \prod_{j=0}^k \{ \Pr[D_{j+1} = 0 | \bar{L}_j = \bar{l}_j, \bar{A}_j = \bar{a}_j, \bar{Y}_j = \bar{D}_j = \bar{C}_{j+1} = 0] \times \\ & \Pr[Y_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{Y}_{j-1} = \bar{D}_j = \bar{C}_j = 0] \times \\ & f(\bar{l}_j | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{Y}_j = \bar{D}_j = \bar{C}_j = 0) \times f^{int}(a_j | \bar{l}_j, \bar{a}_{j-1}, \bar{Y}_j = \bar{D}_j = \bar{C}_j = 0) \} \end{aligned} \quad (D.1)$$

Following previous arguments, suppose we select

$$f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0) = f^{obs}(a_k | R_k = r_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0) \quad (D.2)$$

When $\bar{r} = \bar{l}$, (D.2) is a representative intervention in this generalized setting.

Taubman et al. [2009] described an algorithm that gives a parametric g-formula based estimator of expression (D.1) by imposing parametric models on each component. This algorithm is reviewed in Appendix F. As an alternative, a generalization of the IPW approach described in the main text can be derived from the following alternative representation of expression (D.1) with $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0)$ selected as (D.2):

$$\sum_v \sum_{k=0}^K \lambda_k^{(\bar{r}, \bar{c}=0), 1}(v) \{1 - \lambda_k^{(\bar{r}, \bar{c}=0), 2}(v)\} \prod_{j=0}^{k-1} [\{1 - \lambda_j^{(\bar{r}, \bar{c}=0), 1}(v)\} \{1 - \lambda_j^{(\bar{r}, \bar{c}=0), 2}(v)\}] f(v) \quad (D.3)$$

where

$$\lambda_k^{(\bar{r}, \bar{c}=0), 1}(V) = \frac{E[Y_{k+1}(1 - D_{k+1})(1 - Y_k)W_k^{\bar{r}, \bar{c}=0} | V]}{E[(1 - D_{k+1})(1 - Y_k)W_k^{\bar{r}, \bar{c}=0} | V]}, \quad (D.4)$$

$$\lambda_k^{(\bar{r}, \bar{c}=0), 2}(V) = \frac{E[D_{k+1}(1 - Y_k)(1 - D_k)W_k^{\bar{r}, \bar{c}=0} | V]}{E[(1 - Y_k)(1 - D_k)W_k^{\bar{r}, \bar{c}=0} | V]}, \quad (D.5)$$

and

$$W_k^{\bar{r}, \bar{c}=0} = \prod_{j=0}^k \frac{I(C_{j+1} = 0)I(R_j = r_j)}{\Pr[C_{j+1} = 0 | \bar{L}_j, \bar{A}_j, \bar{D}_j = \bar{C}_j = \bar{Y}_j = 0]f(R_j | \bar{L}_j, \bar{A}_{j-1}, \bar{D}_j = \bar{C}_j = \bar{Y}_j = 0)} \quad (D.6)$$

Let $h\{\gamma(k, \bar{r}_k, V, \psi)\}$ and $h^D\{\gamma^D(k, \bar{r}_k, V, \psi^D)\}$ denote MSMs for (D.4) and (D.5) indexed by the parameters ψ and ψ^D , respectively. In the analysis of Section 6 in the main text, in addition to the model (18), we assume:

$$h^D\{\gamma^D(k, \bar{r}_k, V; \psi^D)\} = \text{expit}\{\gamma(k, \bar{r}_k, V; \psi^D)\}$$

for all $k = 0, \dots, K$ with γ selected as in (26) of Section 6.

Given these MSMs and models for the densities in the denominator of (D.6) are correctly specified, the consistency of the following estimation algorithm follows a straightforward generalization of the arguments of Section 5 based on a class of unbiased estimating equations analogous to (14). Again, we choose amongst these in such a way that gives stabilized compound treatment and censoring weights.

The algorithm is applied to a subject-interval data set where the definition of K^* is further modified such that: If, for a given subject, $C_{K+1} = D_{K+1} = Y_{K+1} = 0$ then $K^* = K$; otherwise if, for any $m = 0, \dots, K$, $D_m = C_m = Y_m = 0$ and *either* $D_{m+1} = 1$ or $C_{m+1} = 1$ or $Y_{m+1} = 1$ then $K^* = m$. We guaranteed risk sets were constructed in line with the conditional probabilities given in expression (D.1) by coding the input data set such that: for subjects with $C_{k+1} = 1$ both D_{k+1} and Y_{k+1} are set to missing values; otherwise, for subjects with $C_{k+1} = 0$ and $D_{k+1} = 1$ then Y_{k+1} is set to a missing value.

Generalized IPW estimation algorithm:

1. Calculate and attach a stabilized compound treatment and censoring weight as in Sections 5.2 of the main text and Appendix C to each subject-interval record.
2. Using all subject-interval records
 - (a) obtain $\hat{\psi}$ by fitting a *weighted* pooled logistic regression model with stabilized weights defined as in the previous step, dependent variable Y_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and $(\bar{R}_k, V)$ corresponding to the choice of $\gamma(\cdot)$
 - (b) obtain $\hat{\psi}^D$ by fitting a *weighted* pooled logistic regression model with stabilized weights defined as in the previous step, dependent variable D_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and $(\bar{R}_k, V)$ corresponding to the choice of $\gamma^D(\cdot)$

Our final IPW estimate is analogously obtained by the plug-in estimator

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=0}^K \text{expit}\{\gamma(k, \bar{r}_k, V_i; \hat{\psi})\} \{1 - \text{expit}\{\gamma^D(k, \bar{r}_k, V_i; \hat{\psi}^D)\}\} \times \prod_{j=0}^{k-1} [\{1 - \text{expit}\{\gamma(j, \bar{r}_j, V_i; \hat{\psi})\}\} \{1 - \text{expit}\{\gamma^D(j, \bar{r}_j, V_i; \hat{\psi}^D)\}\}] \quad (\text{D.7})$$

When we select $\bar{r} = \bar{1}$, (D.7) is an IPW estimate of the g-formula for the risk by $K + 1$ under a representative intervention in the presence of death by competing risk.

A generalization of the IPW estimator of the natural course risk given in Appendix E to the natural course subdistribution CIF can be derived using the arguments above. Here we replace the intervention density (D.2) with the observed treatment density $f^{obs}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = \bar{C}_k = \bar{Y}_k = 0)$ further conditioned on surviving competing risk.

Appendix E

In the application presented in Section 6 of the main text, we consider causal effect measures that compare the risk by $K+1$ had treatment been assigned according to a representative intervention and had censoring been abolished to that had *only* censoring been abolished. We define the latter intervention as *the natural course*. The g-formula associated with the risk by $K+1$ under the natural course is equivalent to expression (5) of the main text with $f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ selected as the observed treatment density $f^{obs}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{Y}_k = 0)$ and all terms additionally conditioned on the event $\bar{C}_k = 0$. For additional details and identifying conditions see [Young et al., 2011].

To derive an IPW estimator of the natural course risk, we analogously consider the following alternative representation of the g-formula associated with the natural course

$$\sum_v \sum_{k=0}^K \lambda_k^{\bar{c}=0}(v) \prod_{j=0}^{k-1} [1 - \lambda_j^{\bar{c}=0}(v)] f(v) \quad (\text{E.1})$$

where $\lambda_k^{\bar{c}=0}(v)$ is equivalent to the right hand side of equation (9) of the main text but with $w_k^{\bar{r}}$ replaced by

$$w_k^{\bar{c}=0} = \prod_{j=0}^k \frac{I(C_{j+1} = 0)}{\Pr[C_{j+1} = 0 | \bar{Z}_j, \bar{C}_j = \bar{Y}_j = 0]} \quad (\text{E.2})$$

Let $m\{\eta(k, V; \beta)\}$ denote an MSM for $\lambda_k^{\bar{c}=0}(v)$ indexed by a parameter β with $\eta(\cdot)$ a function of k and V , differentiable with respect to β . Assume the pooled logistic model in the denominator of the stabilized censoring weight (C.1) is correctly specified for the censoring mechanism $\Pr[C_{k+1} = 0 | \bar{Z}_k, \bar{C}_k = \bar{Y}_k = 0]$. Also assume the pooled logistic model

$$m\{\eta(k, V; \beta)\} = \text{expit}\{\eta(k, V, \beta)\}$$

Following previous arguments, in the absence of model misspecification, a consistent IPW estimator may be derived as the solution to an unbiased estimating equation chosen to allow a stabilized censoring weight. Here we define the numerator in terms of the pooled logistic model

$$\Pr[C_{k+1} = 0 \mid V, \bar{C}_k = \bar{Y}_k = 0; \alpha^{C, num'}] = \text{expit}\{\phi^{C, num'}(k, V; \alpha^{C, num'})\} \quad (\text{E.3})$$

Note that (E.3) differs from the model imposed in the numerator of (C.1) in that it may, at most, be a function of k and V and may not be a function of $\bar{R}_k$. As above, the estimator will remain consistent even if the numerator model is misspecified.

The following IPW estimation algorithm may be used to consistently estimate the MSM parameter β in a subject-interval data set with K^* defined as in Appendix C:

IPW estimation algorithm for β :

1. Using all subject-interval records in the data set:
 - (a) obtain $\hat{\alpha}^{C, num'}$ by fitting a pooled logistic regression model with dependent variable C_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and V corresponding to the choice of $\phi^{C, num'}(\cdot)$ of (E.3).
 - (b) obtain $\hat{\alpha}^C$ by fitting a pooled logistic regression model with dependent variable C_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and $\bar{Z}_k$ corresponding to the choice of $\phi^C(\cdot)$ of (C.1) of Appendix C.
2. Attach a stabilized censoring weight to line k of the data set calculated as

$$\prod_{j=0}^k \frac{\text{expit}\{\phi^{C, num'}(j, V; \hat{\alpha}^{C, num'})\}}{\text{expit}\{\phi^C(j, \bar{A}_j; \bar{L}_j, \hat{\alpha}^C)\}} \quad (\text{E.4})$$

This weight differs from (C.1) by its numerator.

3. Using all subject-interval records, obtain $\hat{\beta}$ by fitting a *weighted* pooled logistic regression model with stabilized weights defined as in the previous step, dependent variable Y_{k+1} and independent variables a specified function of $k = 0, \dots, K$ and V corresponding to the choice of $\eta(\cdot)$

Our final IPW estimate of the g-formula for the risk by $K + 1$ under the natural course is

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=0}^K \text{expit}\{\eta(k, V_i; \hat{\beta})\} \prod_{j=0}^{k-1} [1 - \text{expit}\{\eta(j, V_i; \hat{\beta})\}] \quad (\text{E.5})$$

In the analysis of Section 6 of the main text, assumption (C.2) of Appendix C was also used in constructing the denominator of the stabilized censoring weight (E.4) for the natural course risk. In constructing the numerator of the stabilized censoring weight (E.4) for the natural course we defined:

$$\phi^{C,num'}(k, V; \alpha^{C,num'}) = \alpha_0^{C,num'} + \alpha_1^{C,num'} \mathbb{I}_{g(k)} + \alpha_2^{C,num'} \mathbb{I}_V \quad (\text{E.6})$$

We assumed the MSM:

$$\eta(k, V; \beta) = \beta_0 + \beta_1^T V + \beta_2^T g(k)$$

The natural course risk estimator was also generalized to allow for the presence of death by competing risks as described in Appendix D.

Appendix F

Robins et al. [2004] introduced the *extended g-formula* to estimate from observational data the risk of failure under hypothetical interventions wherein a subject's treatment at time k is assigned based on the *natural value* of treatment at k ; that is, the value of treatment that would have been observed at k were the intervention discontinued right before k . An example is: "If a subject's natural value of treatment at k is below a particular threshold (or above in the case of a harmful exposure) then set treatment to this threshold value. Otherwise, do not intervene on this subject at k ." Taubman et al. [2008] referred to this special case of an intervention that depends on the natural value of treatment as a *threshold intervention*. We also gave the example of partial representative interventions in the main text. Like partial representative interventions and representative interventions, threshold interventions guarantee that a continuous treatment is maintained within a pre-specified range throughout follow-up.

Taubman et al. [2009] applied the parametric extended g-formula to estimate the 20-year risk of CHD in the Nurses Health Study under threshold interventions on behaviors consistent with interventions 1-5 listed in Section 6 of the main text. Under intervention 1, the g-formula characterized by a representative intervention is equivalent to that characterized by either a threshold intervention or the deterministic static intervention "set treatment to zero at all times" as the pre-specified range contains only a single value.

Here we review the general algorithm applied by Taubman et al. [2009]. As in the generalization of the IPW algorithm in Appendix D, this algorithm estimates the subdistribution cumulative incidence function in the presence of death by competing risk.

Denote the natural value of treatment at k in the observational study A_k^* which is equivalent to the treatment A_k in the observational study but not guaranteed equivalent under intervention [Richardson and Robins, 2013, Young et al., 2014]. Further, denote an intervention density that may depend on the natural value of treatment as

$f^d(a_k | a_k^*, \bar{L}_k, \bar{a}_{k-1}, \bar{C}_k = \bar{D}_k = \bar{Y}_k = 0)$. Let $Z_k^* = (L_k, A_k^*)$ such that

$$f(z_k^* | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{c}_k = \bar{D}_k = \bar{Y}_k = 0) = \prod_{j=1}^p m_{j,k}^{z_k^*}$$

where $(m_{1,k}^*, \dots, m_{p,k}^*)$ are conditional densities based on the factorization associated with an arbitrary permutation of the p components of $Z_k^* = 0, \dots, K$. Define $h^{user}(\bar{a}_k, a_k^*, \bar{l}_k)$ as a user-chosen intervention density. This may correspond to a choice of

$f^{int}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{c}_k = \bar{D}_k = \bar{Y}_k = 0)$ that does not depend on the natural value of treatment such as any of the examples considered in Section 3 or to a choice of

$f^d(a_k | a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{c}_k = \bar{D}_k = \bar{Y}_k = 0)$ such as the threshold interventions considered by Taubman et al. [2009].

For a user-chosen value of K , $h^{user}(\bar{a}_k, a_k^*, \bar{l}_k)$ and permutation of $Z_k^* = 0, \dots, K$ we apply the following algorithm to a subject-interval input data set constructed in the same manner as that required for the IPW estimation algorithm described in Appendix D:

1. Using all subject interval records:
 - (a) If $k > 0$, fit parametric models for the conditional densities $m_{j,k}^{z_k^*}$, $j = 1, \dots, p$ indexed by parameter $\beta_{\sim}$.
 - (b) Fit a parametric model for the conditional probability of the outcome $\Pr[Y_{k+1} = 1 | \bar{l}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{C}_{k+1} = \bar{D}_{k+1} = \bar{Y}_k = 0]$ indexed by parameter $\theta_{\sim}$.
 - (c) Fit a parametric model for the conditional probability of the competing risk $\Pr[D_{k+1} = 1 | \bar{l}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{C}_{k+1} = \bar{D}_k = \bar{Y}_k = 0]$ indexed by parameter θ^D .
2. For $k = 0, \dots, K$ and $v = 1, \dots, n$:
 - (a) If $k = 0$, set $z_{0,v}^*$ to the observed values of Z_0^* for subject v . Otherwise, if $k > 0$, recursively draw $z_{k,v}^*$ from the nested conditional densities estimated in step 1.a. based on previously drawn confounders through $k-1$ $\bar{l}_{k-1,v}$ and assigned treatment $\bar{a}_{k-1,v}$ under the user-chosen intervention.
 - (b) Assign the treatment $a_{k,v}$ according to the user-chosen intervention. For example, for $h^{user}(\bar{a}_k, a_k^*, \bar{l}_k)$ chosen as a threshold intervention we set $a_{k,v} = 30$ if $a_{k,v}^* \leq 30$ and otherwise set $a_{k,v} = a_{k,v}^*$.

- (c) Estimate the probability of the outcome by $k + 1$ given survival to k for the v^{th} simulated treatment and confounder history $(\bar{a}_{k,v}, \bar{l}_{k,v})$ based on the estimated coefficients $\hat{\theta}^{\sim}$ from step 1.b.
- (d) Estimate the probability of death by competing risk by $k + 1$ given survival to k for the v^{th} simulated treatment and confounder history $(\bar{a}_{k,v}, \bar{l}_{k,v})$ based on the estimated coefficients $\hat{\theta}^D \sim$ from step 1.c.

3. Compute the intervention risk estimate as

$$\begin{aligned} & \frac{1}{n} \sum_{v=1}^n \sum_{k=0}^K \Pr[Y_{k+1} = 1 \mid \bar{L}_k = \bar{l}_{k,v}, \bar{A}_k = \bar{a}_{k,v}, \bar{C}_{k+1} = \bar{D}_{k+1} = \bar{Y}_k = 0; \hat{\theta}^{\sim}] \times \\ & \prod_{j=0}^k \{1 - \Pr[D_{j+1} = 1 \mid \bar{L}_j = \bar{l}_{j,v}, \bar{A}_j = \bar{a}_{j,v}, \bar{C}_{j+1} = \bar{D}_j = \bar{Y}_j = 0; \hat{\theta}^D \sim]\} \times \\ & \{1 - \Pr[Y_j = 1 \mid \bar{L}_{j-1} = \bar{l}_{j-1,v}, \bar{A}_{j-1} = \bar{a}_{j-1,v}, \bar{C}_j = \bar{D}_j = \bar{Y}_{j-1} = 0; \hat{\theta}^{\sim}]\} \end{aligned} \quad (F.1)$$

where

$$\Pr[Y_{k+1} = 1 \mid \bar{L}_k = \bar{l}_{k,v}, \bar{A}_k = \bar{a}_{k,v}, \bar{C}_{k+1} = \bar{D}_{k+1} = \bar{Y}_k = 0; \hat{\theta}^{\sim}]$$

is obtained in step 2.c. and

$$\Pr[D_{k+1} = 1 \mid \bar{L}_k = \bar{l}_{k,v}, \bar{A}_k = \bar{a}_{k,v}, \bar{C}_{k+1} = \bar{D}_k = \bar{Y}_k = 0; \hat{\theta}^D \sim]$$

is obtained in step 2.d. for each $k = 0, \dots, K$. Note that for we have suppressed throughout the dependence of $\bar{a}_{k,v}$ and $\bar{l}_{k,v}$ on $\hat{\beta}^{\sim}$.

As discussed in Young et al. [2011], both steps 1. and 2. may be modified to avoid reliance on parametric models for histories such that a priori subject matter knowledge on the observed data structure is available. If $h^{user}(\bar{a}_k, \bar{a}_k^*, \bar{l}_k)$ is selected as a representative intervention, an additional parametric model for $f^{obs}(a_k \mid R_k = 1, \bar{l}_k, \bar{a}_{k-1}, \bar{C}_k = \bar{D}_k = \bar{Y}_k = 0)$ can be fit and the estimated parameters of this model used to assign treatment in step 2.b. Comparability to IPW estimates under representative interventions depends on correct specification of this model as well.

Table F.1 compares parametric g-formula estimates of the risk by $K + 1$ under threshold and representative interventions consistent with interventions 2 through 5 based on the above algorithm using the data set described in Section 6. Parametric models used for step 1 of the

algorithm are similar to those described in Taubman et al. [2009]. Threshold intervention results in Table F.1 differ from those presented in Taubman et al. [2009] based on extension of the follow-up from 20 to 28 years and correction of some coding errors.

Table F.1:

Comparison of parametric g-formula estimates of risk by 28-year follow-up in the Nurses Health Study under representative and threshold interventions both consistent with interventions 2 – 5 . Risk ratios (RRs) are constructed relative to the natural course.

	Intervention	Intervention density	Risk(%)	RR
2	Exercise 30 min/day	representative	4.3	0.9
			threshold 3.9	0.8
3	Diet in top two quintiles	representative	4.6	0.9
			threshold 4.6	0.9
4	Consume 5 g/day alcohol	representative	4.4	0.9
			threshold 4.6	0.9
5	Joint 1 – 4	representative	3.0	0.6
			threshold 2.9	0.6

Appendix G

Table G.1:

Functional forms of components of $\bar{L}_k$ used in $\phi(\cdot)$ and $\phi^{C\sim}(\cdot)$ of the stabilized compound treatment and censoring weight denominator models (24) of the main text and (C.2) of Appendix C, respectively, and years assessed for each of the time-varying covariates never acting as treatments. Abbreviations: “inter”=interactions with time since covariate last measured; “ind”=indicator variable; “cat”=categories.

Time-varying covariate	Function of “History” [*] (component of $\bar{L}_k$)	Years measured
cancer	current, lag1, lag2 (ind)	All
BMI	current, lag1, lag2 (5 cat)	All
statins	current, lag1, (ind, inter)	‘88, ‘94-on
angina	current, lag1, lag2 (ind)	All
diabetes	current, lag1, lag2 (ind)	All
CABG	current, lag1, lag2 (ind)	All
osteoporosis	current, lag1, lag2 (ind)	All
multivitamins	current, lag1, lag2 (ind)	All
menopause	current, lag1, lag2 (ind)	All
aspirin	current, lag1, lag2 (ind)	All
stroke	current, lag1, lag2 (ind)	All
high blood pressure	current, lag1, lag2 (ind)	All
post-menopausal hormone use	current, lag1, lag2 (ind)	All

^{*}“History” is relative to k : “current” is the interval k measurement, “lag1” the interval $k - 1$ measurement, “lag2” the interval $k - 2$ measurement.

Table G.2:

Functional forms of $(\bar{L}_k, \bar{A}_{k-1})$ and $\bar{Z}_k$ used in $\phi(\cdot)$ and $\phi^{C\sim}(\cdot)$ of the stabilized compound treatment and censoring weight denominator models (24) of the main text and (C.2) of Appendix C, respectively, and years assessed for each time-varying covariate that acts as either a treatment or confounder, depending on the intervention. Abbreviations: “inter”=interactions with time since covariate last measured; “ind”=indicator variable; “cat”=categories.

Time-varying covariate	Function of “History” when acting as “confounder” (component of $\bar{Z}_k$ or $\bar{L}_k$)	Function of “History” when acting as “treatment” (component of $\bar{A}_{k-1}$)	Years measured
cigarettes	current, lag1, lag2 (4 cat)	lag1, lag2 (4 cat) **	All
exercise	current, lag1, (5 cat, inter)	lag1, (5 cat, inter)	‘80, ‘82, ‘86, ‘88, ‘92-on
alcohol	current, lag1, (3 cat, inter)	lag1, (3 cat, inter)	‘80, ‘84, ‘86, ‘90, ‘94, ‘98
diet	current, lag1, (4 cat, inter)	lag1, (4 cat, inter)	‘80, ‘84, ‘86, ‘90, ‘94, ‘98

* “History” is relative to k : “current” is the interval k measurement, “lag1” the interval $k-1$ measurement, “lag2” the interval $k-2$ measurement.

** Separate models were fit for the compound treatment components of the numerator and denominator of the stabilized weight under intervention 1 relative to interval k for subjects who smoked versus did not smoke in interval $k-1$. For the model fit to non-smokers in $k-1$, no function of the interval $k-1$ smoking measurement was included as a predictor as all subjects had the same value of zero cigarettes per day.

References

- Bang H and Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:692–972, 2005. [PubMed: 16135020]
- Haight TJ, Wang Y, van der Laan MJ, and Tager IB. A cross-validation deletion-substitution-addition model selection algorithm: Application to marginal structural models. *Computational Statistics and Data Analysis*, 54(12):3080–3094, 2010. [PubMed: 25505354]
- Hernán MA, Brumback B, and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5): 561–570, 2000. [PubMed: 10955409]
- Kalbfleisch JD and Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley, 1980.
- Moodie, Stephens DA, and Klein MB A marginal structural model for multiple-outcome survival data: assessing the impact of injection drug use on several causes of death in the Canadian Co-infection Cohort. *Statistics in Medicine*, 33(8):1409–25, 2014. [PubMed: 24272681]
- Neugebauer R and van der Laan MJ. Locally efficient estimation of nonparametric causal effects on mean outcomes in longitudinal studies U.C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 134, 2003 URL <http://biostats.bepress.com/ucbbiostat/paper134>.
- Orellana L, Rotnitzky A, and Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: Main Content. *International Journal of Biostatistics*, 6, 2010a. Article 7.
- Orellana L, Rotnitzky A, and Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: Proofs and Additional Results. *International Journal of Biostatistics*, 6, 2010b. Article 8.
- Petersen M, Schwab J, , Gruber S, Blaser N, Schomaker M, and van der Laan MJ. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of Causal Inference*, 2(2):147–185, 2014. [PubMed: 25909047]

- Picciotto S, Hernán MA, Page JH, Young JG, and Robins JM. Structural nested cumulative failure time models to estimate the effects of interventions. *Journal of the American Statistical Association*, 107(499):886–900, 2012.
- Platt RW, Brookhart MA, Cole SR, Westreich D, and Schisterman EF. An information criterion for marginal structural models. *Statistics in Medicine*, 32(8):1383–1393, 2013. [PubMed: 22972662]
- Richardson TS and Robins JM. Single World Intervention Graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Center for Statistics and the Social Sciences, University of Washington Series, 2013 URL <http://www.csss.washington.edu/Papers/>. Working Paper Number 128.
- Robins JM and Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90:122–129, 1995.
- Robins JM, Hernán MA, and Siebert U. Effects of multiple interventions In Ezzati M, Lopez AD, Rodgers A, and Murray CJL, editors, *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health Organization, 2004.
- Robins JM and Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers In *AIDS Epidemiology – Methodological Issues*, pages 297–331. New York: Springer, 1992.
- Scharfstein DO, Rotnitzky A, and Robins JM. Adjusting for non ignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94:1096–1120, 1999.
- Schnitzer ME, Moodie EM, van der Laan MJ, Platt RW, and Klein MB. Modeling the impact of Hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation. *Biometrics*, 70:144–152, 2014. [PubMed: 24571372]
- Taguri M and Matsuyama Y. Comments on “An information criterion for marginal structural models”. *Statistics in Medicine*, 32(20):3590–3591, 2013. [PubMed: 23943549]
- Taubman SL, Mittleman MA, Robins JM, and Hernán MA. Alternative approaches to estimating the effects of hypothetical interventions In *JSM Proceedings, Health Policy Statistics Section*. Alexandria, VA: American Statistical Association, 2008.
- Taubman SL, Robins JM, Mittleman MA, and Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology*, 38(6):1599–611, 2009. [PubMed: 19389875]
- van der Laan MJ and Robins JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer, 2002.
- van der Laan MJ and Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics*, 8(Article 9), 2012.
- Young JG, Hernán MA, Picciotto S, and Robins JM. Equivalence between structural models for the effect of a time-varying exposure on survival. *Lifetime Data Analysis*, 2008.
- Young JG, Cain LE, Robins JM, O'Reilly EJ, and Hernán MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences*, 3(1):119–143, 2011. [PubMed: 24039638]
- Young JG, Hernán MA, and Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods*, 3(1):1–19, 2014. [PubMed: 25866704]

References

- Colditz GA, Manson JE, and Hankinson SE. The Nurses Health Study: 20-year contribution to the understanding of health among women. *Journal of Women's Health*, 6:49–62, 1997.
- Cole SR and Hernán MA. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6):656–64, 2008. [PubMed: 18682488]
- Díaz Muñoz I and van der Laan MJ. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68:541–549, 2012. [PubMed: 21977966]

- Haneuse S and Rotnitzky A. Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32:5260–5277, 2013. [PubMed: 23913589]
- Hernán MA and VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology*, 22(3):368–377, 2011. [PubMed: 21399502]
- Hernán MA, Brumback B, and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5): 561–570, 2000. [PubMed: 10955409]
- Neugebauer R and van der Laan MJ. Locally efficient estimation of nonparametric causal effects on mean outcomes in longitudinal studies. U.C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 134, 2003 URL <http://biostats.bepress.com/ucbbiostat/paper134>.
- Petersen ML, Porter KE, Gruber S, Wang Y, and van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 21(1):31–54, 2012. [PubMed: 21030422]
- Picciotto S, Hernán MA, Page JH, Young JG, and Robins JM. Structural nested cumulative failure time models to estimate the effects of interventions. *Journal of the American Statistical Association*, 107(499):886–900, 2012.
- Richardson TS and Robins JM. Single World Intervention Graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Center for Statistics and the Social Sciences, University of Washington Series, 2013 URL <http://www.csss.washington.edu/Papers/>. Working Paper Number 128.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986. [Errata (1987) in *Computers and Mathematics with Applications* 14, 917–921. Addendum (1987) in *Computers and Mathematics with Applications* 14, 923–945. Errata (1987) to addendum in *Computers and Mathematics with Applications* 18, 477.].
- Robins JM. Marginal structural models versus structural nested models as tools for causal inference In *Statistical Models in Epidemiology*, pages 95–133. New York: Springer, 2000.
- Robins JM and Wasserman L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs In Geiger D and Shenoy P, editors, *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 409–420. San Francisco: Morgan Kaufmann, 1997.
- Robins JM, Hernán MA, and Siebert U. Effects of multiple interventions In Ezzati M, Lopez AD, Rodgers A, and Murray CJL, editors, *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*. Geneva: World Health Organization, 2004.
- Robins JM and Wang N. Inference for imputation estimators. *Biometrika*, 87:113–124, 2000.
- Stitelman OM, Hubbard AE, and Jewell NP. The impact of coarsening the explanatory variable of interest in making causal inferences: Implicit assumptions behind dichotomizing variables U.C. Berkeley Division of Biostatistics Working Paper Series, 2010 URL <http://biostats.bepress.com/ucbbiostat/paper264>. Working Paper 264.
- Taubman S. Estimating the effects of hypothetical interventions in longitudinal data. 2007 Doctoral Dissertation.
- Taubman SL, Mittleman MA, Robins JM, and Hernán MA. Alternative approaches to estimating the effects of hypothetical interventions In *JSM Proceedings, Health Policy Statistics Section*. Alexandria, VA: American Statistical Association, 2008.
- Taubman SL, Robins JM, Mittleman MA, and Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology*, 38(6):1599–611, 2009. [PubMed: 19389875]
- van der Laan MJ and Robins JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer, 2002.
- Young JG, Cain LE, Robins JM, O'Reilly EJ, and Hernán MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences*, 3(1):119–143, 2011. [PubMed: 24039638]

Young JG, Hernán MA, and Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods*, 3(1):1–19, 2014. [PubMed: 25866704]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Summary of the distribution of stabilized weights prior to truncation used in construction of the five IPW behavioral intervention risk estimates presented in Table 2.

Intervention	Mean	Max	Median	95 percentile	99 percentile
1 Never smoke	1.3	9×10^4	0.96	1.5	2.2
2 Exercise 30 min/day	1.0	1×10^2	0.95	1.7	2.9
3 Diet in top two quintiles	1.0	1×10^2	0.91	2.2	4.6
4 Consume 5 g/day alcohol	1.0	7×10^1	0.95	1.6	3.2
5 Joint 1 – 4	1.2	3.0×10^4	0.85	2.6	6.1

Table 2:

IPW estimates of risk by 28-year follow-up in the Nurses Health Study under representative interventions consistent with interventions 1 – 5. Risk ratios (RRs) are constructed relative to the natural course and 95% confidence intervals based on 500 bootstrap samples.

	Intervention	Risk(%)	95% CI	RR	95% CI
0	Natural Course	5.0	4.8, 5.2	1	–
1	Never smoke	4.2	4.0, 4.4	0.8	0.8, 0.9
2	Exercise 30 min/day	4.3	3.7, 4.9	0.9	0.7, 1.0
3	Diet in top two quintiles	5.0	4.5, 5.5	1.0	0.9, 1.1
4	Consume 5 g/day alcohol	5.0	4.5, 5.6	1.0	0.9, 1.1
5	Joint 1 – 4	3.9	3.0, 4.8	0.8	0.6, 0.9

Table 3:

Sensitivity of IPW estimates of 28-year CHD risk under the five behavioral inter-ventions to different truncation levels.

Intervention	99 percentile	95 percentile	Fixed value 10
1 Never smoke	4.2	4.2	4.2
2 Exercise 30 min/day	4.3	4.3	4.3
3 Diet in top two quintiles	5.0	4.7	5.1
4 Consume 5 g/day alcohol	5.0	4.8	5.2
5 Joint 1 – 4	3.9	3.5	4.1

Table 4:

Sensitivity to model choices: empirical comparison of estimates of 28-year CHD risk under the five behavioral interventions using IPW under different choices of $\gamma_1(k, \bar{R}_k, V; \psi_1)$ in the model (26) and using the parametric g-formula indexed by a representative intervention.

Intervention	MSM 1	MSM 2	MSM 3	MSM 4	parametric g-formula**
1 Never smoke	4.2	4.1	4.1	4.0	4.2
2 Exercise 30 min/day	4.3	4.8	4.8	4.9	4.3
3 Diet in top two quintiles	5.0	4.9	5.1	5.1	4.6
4 Consume 5 g/day alcohol	5.0	4.9	4.9	4.8	4.4
5 Joint 1 – 4	3.9	4.1	4.2	4.6	3.0

* Four choices of $\gamma_1(k, \bar{R}_k, V; \psi_1)$:

¹MSM 1: $\gamma_1(k, \bar{R}_k, V; \psi_1)$ defined by (27) as in the IPW estimates of Table 2

²MSM 2: $\gamma_1(k, \bar{R}_k, V; \psi_1) = \psi_{11}R_k + \psi_{12}cumavg(\bar{R}_{k+1})$ where $cumavg(\bar{R}_{k+1}) = \frac{\sum_{m=0}^{k-1} R_m}{k}$ if $k > 0$ and zero otherwise.

³MSM 3: $\gamma_1(k, \bar{R}_k, V; \psi_1) = \psi_{11}R_k + \psi_{12}cumavg(\bar{R}_{k-1}) + \psi_{13}^T V_{sub} R_k + \psi_{14}^T V_{sub} cumavg(\bar{R}_{k-1})$ where V_{sub} is a subset of the baseline covariates V

⁴MSM 4: $\gamma_1(k, \bar{R}_k, V; \psi_1) = \psi_{11}R_k + \psi_{12}cumavg(\bar{R}_{k-1}) + \psi_{13}^T g(k)R_k + \psi_{14}^T g(k)cumavg(\bar{R}_{k-1})$

** The parametric g-formula estimate of the natural course risk by 28 years was the same as that obtained via IPW (5.0%) as described in Appendix E. A nonparametric estimate of this risk, valid under the more restrictive assumption of “no selection bias” due to censoring (see Young et al. [2011] for details), was 4.9%.