

Jessica G. Young*, Miguel A. Hernán and James M. Robins

Identification, Estimation and Approximation of Risk under Interventions that Depend on the Natural Value of Treatment Using Observational Data

Abstract: Robins et al. (2004, Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva: World Health Organization) 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 . Several authors have parametrically applied the extended g-formula to estimate long-term failure risk under hypothetical interventions on time-varying behaviors in observational studies. For example, Taubman et al. (2009, Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology*, 38(6):1599–1611) used this approach to estimate the 20-year risk of coronary heart disease in the Nurses' Health Study under the hypothetical intervention "If a subject's natural value of exercise by the end of day k is less than 30 minutes, set her exercise on day k to exactly 30 minutes; otherwise, do not intervene on her on that day". Non-parametrically, the extended g-formula differs from the (non-extended) g-formula of Robins (1986, A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512) in that it is a function of (i) a user-specified intervention depending on the natural value of treatment and (ii) the distribution of natural treatment itself. Richardson and Robins (2013, <http://www.csss.washington.edu/Papers/>) recently defined a sufficient condition such that the extended g-formula may identify risk under an intervention that depends on the natural value of treatment, provided this expression is well-defined. In this paper, we complement this result by showing that the extended g-formula associated with an intervention depending on the natural value of treatment is algebraically equivalent to the (non-extended) g-formula associated with a particular random dynamic regime that does not depend on this value. Using previous results for random dynamic regimes, we show that this equivalence immediately gives a sufficient positivity condition that guarantees the extended g-formula is well-defined as well as semi-parametric alternatives to the parametric extended g-formula for estimation. Finally, given a hypothetical intervention that depends on the natural value of treatment, we define a plausible (implementable) approximation to this hypothetical intervention along with an untestable assumption that gives exact equivalence.

Keywords: causal inference; survival analysis; g-formula; random dynamic regimes; inverse-probability weighting

*Corresponding author: Jessica G. Young, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue Kresge Bldg, Boston, MA 02115, USA, E-mail: jyoung@hsph.harvard.edu

Miguel A. Hernán, Departments of Epidemiology and Biostatistics, Harvard School of Public Health and Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA, E-mail: mhernan@hsph.harvard.edu

James M. Robins, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA, USA, E-mail: robins@hsph.harvard.edu

1 Introduction

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 . Several authors (Robins et al. 2004; Taubman et al. 2009; Lajous et al. 2013; Danaei et al. 2013; García-Aymerich et al. 2014) have parametrically applied this approach to estimate the risk of failure in observational studies under hypothetical time-varying interventions of the following form: "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*. For example, Taubman et al. (2009) used the parametric extended g -formula to estimate the 20-year risk of coronary heart disease (CHD) in the Nurses' Health Study (NHS) under the following hypothetical threshold intervention on daily minutes of exercise on all days of follow-up "If a subject's natural value of exercise by the end of day k is less than 30 minutes, set her exercise on day k to exactly 30 minutes. Otherwise, do not intervene on this subject on day k ". Threshold interventions have the property that they guarantee a continuous treatment is maintained within a pre-specified range (e.g. at least 30 minutes per day) continually throughout the follow-up while minimizing the number of subjects requiring intervention at each time.

Non-parametrically, the extended g -formula differs from the (non-extended) g -formula of Robins (1986) in that it includes (i) a specific user-supplied intervention density that depends on the natural value of treatment at each k and (ii) the density of natural treatment itself at each k conditional on past measured confounders (Robins et al. 2004). Richardson and Robins (2013) recently defined a condition such that the extended g -formula non-parametrically identifies risk under an intervention that depends on the natural value of treatment associated with the user-supplied intervention density in (i), provided this expression is well-defined. In this paper, we complement this result by showing the algebraic equivalence between the extended g -formula associated with a user-supplied intervention density (i) and the (non-extended) g -formula associated with a particular *random dynamic regime* that does not depend on the natural value of treatment and may, at most, depend on the measured confounders.

Provided the identifying condition of Richardson and Robins (2013) holds, this algebraic equivalence gives

1. a sufficient positivity condition such that the extended g -formula is well-defined and thus non-parametrically identifies risk under an intervention that depends on the natural value of treatment in an observational study and
2. semi-parametric alternatives to the parametric extended g -formula for estimation.

Given this equivalence, these results follow immediately from previous work on identification and estimation of the effects of random dynamic regimes that do not depend on the natural value of treatment and may, at most, depend on the measured confounders. For example, see Robins (1986, 1997), Pearl (2000), Murphy et al. (2001), van der Laan et al. (2005), Hernán et al. (2006), Tian (2008), Dawid and Didelez (2008), Robins and Hernán (2009), Orellana et al. (2010a, 2010b), Cain et al. (2010), Stitelman et al. (2010), Dawid and Didelez (2010), Young et al. (2011), Picciotto et al. (2012) and Díaz Muñoz and van der Laan (2012).

Finally, there has been no consideration of the limits on physical implementation of interventions that depend on the natural value of treatment. For example, once we observe that a subject has exercised 20 minutes by the end of day k we cannot subsequently intervene and make her exercise any more (or any fewer) minutes by the end of that day. Therefore, given a hypothetical intervention that depends on the natural value of treatment, we define a plausible (implementable) approximation to this intervention. We also provide an untestable assumption that, when satisfied, would give exact equivalence.

The structure of the paper is as follows. In Section 2, we define the observational data structure of interest and give a classification of hypothetical interventions that do not depend on the natural value of treatment and may, at most, depend on the measured confounders, including random dynamic regimes. In Section 3, we review a set of conditions that non-parametrically identifies risk by the end of follow-up in the observational study under any hypothetical intervention within this classification by the (non-extended) g-formula. In Section 4, we show the algebraic equivalence between the extended g-formula associated with an intervention that depends on the natural value of treatment and the (non-extended) g-formula associated with a particular random dynamic regime. In Section 5, we review the parametric extended g-formula estimator and give a semi-parametric alternative that follows immediately from the results of Section 4 given previous semi-parametric results in the context of random dynamic regimes. In Section 6, we define a plausible approximation to an intervention that depends on the natural value of treatment and an assumption for exact equivalence.

2 A classification of interventions that do not depend on the natural value of treatment

Consider an observational study in which the following random variables are measured during each follow-up time (e.g. day) $k = 0, \dots, K + 1$ for each of $i = 1, \dots, n$ subjects. We assume subjects are independent and identically distributed and thus suppress the i subscript. Let D_k be an indicator of failure (e.g. CHD) by k , L_k a vector of measured confounders at the start of k (e.g. smoking, body mass index [BMI] and diet), and A_k the treatment observed during k (e.g. number of minutes of actual daily exercise). During any given time k , D_k precedes (L_k, 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 k . For notational convenience, we set \bar{L}_{-1} and \bar{A}_{-1} to be identically 0 and, by definition, $\bar{D}_0 = 0$. We use lower-case letters to denote possible realizations of a random variable, for example, a_k is a possible realization of treatment A_k . For simplicity, we assume that no subjects are lost to follow-up or die from competing risks and that all variables are perfectly measured. If a subject has failed by k , i.e. $D_k = 1$, then by convention, we will set $L_k = A_k = 0$.

Our goal is to estimate the risk of failure that would have been observed by the end of follow-up $K + 1$ had all subjects in this study population followed a hypothetical intervention or treatment regime. Generally define a treatment regime that does not depend on the natural value of treatment as a rule that assigns treatment at k as an independent draw from an intervention density $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ that may, at most, depend on $(\bar{a}_{k-1}, \bar{l}_k)$, $k = 0, \dots, K$ (Robins 1986).

Treatment regimes can be either *deterministic* or *random*. A regime is deterministic if $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ may only equal zero or one for all (\bar{a}_k, \bar{l}_k) and $k = 0, \dots, K$. Otherwise, it is random. In particular, we denote $g = (g_0, \dots, g_K)$ to be the deterministic regime associated with the intervention density defined by $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}^g, \bar{D}_k = 0) = 1$ if $a_k = a_k^g$ and 0 otherwise, where $a_s^g = g_s(\bar{l}_s, \bar{a}_{s-1}^g)$ 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 $(\bar{l}_s, \bar{a}_{s-1}^g)$, $s = 0, \dots, k$.

Treatment regimes can further be classified as *static* or *dynamic*. A deterministic regime g is *static* if a_k^g does not depend on any component of \bar{l}_k for all k . Otherwise g is *dynamic*. Analogously, a random regime may be classified as *static* if the intervention density $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ does not depend on any component of \bar{l}_k for all k . Otherwise, this random regime can be classified as *dynamic*. As noted by Picciotto et al. (2012), and as made explicit in our notation, treatment assignment under any regime $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ within the current classification depends on surviving to k (i.e. the event $D_k = 0$).

To fix ideas, let us consider some examples of treatment regimes in the context of interventions on daily exercise:

1. **Deterministic static regime:** “Set daily exercise to 30 minutes on every day k for all subjects” or $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 1$ if $a_k = 30$ and 0 otherwise for all $k = 0, \dots, K$. For this regime g , $a_k^g = 30$ for any k and confounder history \bar{l}_k .
2. **Deterministic dynamic regime:** “If a subject’s BMI at the start of day k is ≥ 25 , then set her exercise to exactly 30 minutes on that day. Otherwise, set her exercise to exactly 60 minutes” or, for $L_{1,k}$ the component of L_k corresponding to the day k BMI measurement,
 - if $l_{1,k} \geq 25$, then $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 1$ if $a_k = 30$ and 0 otherwise
 - if $l_{1,k} < 25$, then $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 1$ if $a_k = 60$ and 0 otherwise $k = 0, \dots, K$. For this regime g , $a_k^g = 30$ if $l_{1,k} \geq 25$ and $a_k^g = 60$ otherwise for all k .
3. **Random static regime:** “Randomly assign a subject’s exercise on day k such that the probability of receiving 30 minutes is 0.8 and the probability of receiving 60 minutes is 0.2” or $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 0.8$ if $a_k = 30$, 0.2 if $a_k = 60$, and 0 otherwise. The intervention density $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ may take on values between 0 and 1 but its value does not depend on \bar{l}_k for any k .
4. **Random dynamic regime:** “If a subject’s BMI at the start of day k is ≥ 25 , randomly assign her exercise on day k such that the probability of receiving 30 minutes is 0.8 and the probability of receiving 60 minutes is 0.2. Otherwise, set her exercise to 60 minutes on that day” or
 - if $l_{1,k} \geq 25$, then $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 0.8$ if $a_k = 30$, 0.2 if $a_k = 60$ and 0 otherwise
 - if $l_{1,k} < 25$, then $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 1$ if $a_k = 60$ and 0 otherwise $k = 0, \dots, K$. The intervention density $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ may take on values between 0 and 1 and its value depends on \bar{l}_k for some k .

3 Identifying risk under interventions that do not depend on the natural value of treatment

In observational studies, treatment is not under the control of the investigator but is assigned by some unknown treatment rule that generally differs from the hypothetical regime of interest $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. In this section, we will review a set of conditions under which data from an observational study can still be used to identify the risk had all subjects, contrary to fact, followed a treatment regime characterized by $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$.

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

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

$$(D_{k+1}^g, \dots, D_{K+1}^g) \prod A_k | \bar{L}_k = \bar{l}_k, \bar{A}_{k-1} = \bar{a}_{k-1}^g, D_k = 0 \quad [1]$$

Exchangeability [1] encodes the assumption that the measured history $(\bar{L}_k, \bar{A}_{k-1})$ is sufficient to control confounding for the effect of treatment at k on future outcomes. It is often referred to as the assumption of no unmeasured confounding and the vector \bar{L}_k the *measured confounder history* at k .

3. **Positivity:**

$$\begin{aligned} f_{\bar{A}_{k-1}, \bar{L}_k, D_k}(\bar{a}_{k-1}^g, \bar{l}_k, 0) &\neq 0 \Rightarrow \\ f_{\bar{A}_k | \bar{L}_k, \bar{A}_{k-1}, D_k}(a_k^g | \bar{l}_k, \bar{a}_{k-1}^g, 0) &\equiv f^{\text{obs}}(a_k^g | \bar{l}_k, \bar{a}_{k-1}^g, \bar{D}_k = 0) > 0 \text{ w.p.1.} \end{aligned} \quad [2]$$

where $f^{\text{obs}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ denotes the observed treatment density, that is, the conditional density of treatment at k in the observational study evaluated at a particular (\bar{a}_k, \bar{l}_k) .

Under the three g -specific identifying assumptions stated above for each deterministic regime $g \in \mathcal{G}$, where \mathcal{G} is the set of all deterministic regimes, the risk by $K + 1$ under an intervention characterized by any $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is equivalent to the g -formula (Robins 1986):

$$\begin{aligned} & \sum_{\bar{a}_K} \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{D}_k = 0] \\ & \times \prod_{j=0}^k \{ \Pr[D_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{D}_{j-1} = 0] \\ & \times f(l_j | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{D}_j = 0) \times f^{\text{int}}(a_j | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{D}_j = 0) \} \end{aligned} \quad [3]$$

where $f(l_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$ and $\Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{D}_k = 0]$ 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}_k the first $k + 1$ components of \bar{l}_K , $k = 0, \dots, K$. A proof of this equivalence under the current data structure and notation is provided in the appendix of Young et al. (2011) following Lemma 4.2 of Robins (1986).

One-minus expression [3] is equivalent to survival by $K + 1$ under a treatment regime characterized by $f^{\text{int}}(a_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$, $k = 0, \dots, K$. 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 $f^{\text{int}}(a_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$. Appendix A reviews this equivalence and provides a simplified numerical example in a low-dimensional setting. Note for a given choice of $f^{\text{int}}(a_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$, the three identifying assumptions need only hold for the subset of deterministic regimes g that contribute a non-zero weight to the weighted average.

In settings with high-dimensional confounders and/or multiple follow-up times, it will often be quite cumbersome (if not impossible) to list every deterministic regime in the set \mathcal{G} with non-zero weights corresponding to a particular choice of $f^{\text{int}}(a_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$. An exception is the case where $f^{\text{int}}(a_k | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$ is defined in terms of a single deterministic regime g . In this special case, all weight is given to this single deterministic regime and expression [3] reduces to:

$$\begin{aligned} & \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k^g, \bar{D}_k = 0] \\ & \times \prod_{j=0}^k \{ \Pr[D_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}^g, \bar{D}_{j-1} = 0] \\ & \times f(l_j | \bar{l}_{j-1}, \bar{a}_{j-1}^g, \bar{D}_j = 0) \} \end{aligned} \quad [4]$$

which may be more familiar to some readers.

4 Identifying risk under interventions that depend on the natural value of treatment

Given an intervention, define the natural value of treatment at k as the value of treatment that would have been observed at time k were the intervention discontinued right before k . We denote the natural value of treatment at k as A_k^* where, for notational simplicity, we suppress dependence on the associated

intervention. Thus far, we have only considered interventions that may, at most, depend on the measured confounders as classified in Section 2. We now extend our consideration to interventions that may also depend on the natural value of treatment at k . We shall represent such a hypothetical intervention by its intervention density, $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. An example of $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is the threshold intervention of Taubman et al. (2009) on daily exercise stated in Section 1 such that

$$\begin{aligned} \text{If } a_k^* \leq 30 \text{ then } f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= 1 \text{ if } a_k = 30 \text{ and } 0 \text{ o.w.} \\ \text{If } a_k^* > 30 \text{ then } f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= 1 \text{ if } a_k = a_k^* \text{ and } 0 \text{ o.w.} \end{aligned} \quad [5]$$

Note that, in an observational study, the natural value of treatment at k A_k^* is equivalent to the observed treatment A_k as no intervention has been made.

Robins et al. (2004) defined the *extended g-formula* for risk by $K + 1$ associated with an intervention density $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$:

$$\begin{aligned} &\sum_{\bar{a}_K} \sum_{\bar{a}^*_K} \sum_{\bar{l}_K} \sum_{k=0}^K \Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{D}_k = 0] \\ &\quad \times \prod_{j=0}^k \{ \Pr[D_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{D}_{j-1} = 0] \\ &\quad \times f^d(a_j|a_j^*, \bar{l}_j, \bar{a}_{j-1}, \bar{D}_j = 0) \times f(a_j^* | \bar{l}_j, \bar{a}_{j-1}, \bar{D}_j = 0) \\ &\quad \times f(l_j | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{D}_j = 0) \} \end{aligned} \quad [6]$$

where we stress that $f(a_k^* | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is the conditional density of $A_k = A_k^*$ in the observational study evaluated at a_k^* given past treatment, confounders and survival to k , $k = 0, \dots, K$. To emphasize this fact, we sometimes write this density as $f^{\text{obs}}(a_k^* | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$.

Richardson and Robins (2013) defined a condition such that expression [6] identifies from observational data the risk by $K + 1$ under a hypothetical intervention $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ provided this expression is well-defined. We can informally understand this condition as the assumption that A_k^* is not a confounder and has no effect on the outcome except through future treatment. We consider this condition more formally in the context of a simple example in Appendix B.

Consider one particular intervention that does *not* depend on A_k^* within the classification of Section 2 specifically chosen as

$$\begin{aligned} &f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) \\ &= \sum_{a_k^*} f^d(a_k | a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) f^{\text{obs}}(a_k^* | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) \end{aligned} \quad [7]$$

for any (\bar{a}_k, \bar{l}_k) . We will say that this choice of $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is an *implied* treatment rule because it is a marginalization of the user-supplied density $f^d(a_k | a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ over the observational data density of $A_k = A_k^*$. For this *particular* choice of $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, the extended g-formula [6] is equivalent to the (non-extended) g-formula [3]. This equivalence follows by the absence of A_k^* from the conditioning statement of the conditional probability of the outcome at any time $k + 1, \dots, K + 1$ in expression [6].

By this equivalence, it immediately follows that, with $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ defined by eq. [7], the positivity condition [2] of Section 3 guarantees that both the (non-extended) g-formula [3] and the extended g-formula [6] are well-defined. Note, again, for this $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, this condition need only hold for the subset of deterministic regimes g that contribute a non-zero weight to the associated weighted

average of deterministic regimes. Díaz Muñoz and van der Laan (2011, 2012) and Haneuse and Rotnitzky (2013) noted a similar result in the point treatment setting for random dynamic regimes that might be interpreted in terms of implied random dynamic regimes based on an explicit deterministic mechanism depending on the natural value of treatment. The regimes considered by these authors are discussed in Section 5.2.

The implied intervention density [7] is a function of the observed treatment density $f^{\text{obs}}(a_k^*|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, which is generally unknown in high-dimensional observational data (although, it may be estimated). Therefore, the implied $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ will also generally be unknown. For example, for $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ as defined in [5], the marginalization [7] evaluates to

$$\begin{aligned} f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= \Pr^{\text{obs}}(A_k \leq a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) \text{ if } a_k = 30, \\ f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= f^{\text{obs}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) \text{ if } a_k > 30, \\ f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= 0 \text{ if } a_k < 30. \end{aligned} \quad [8]$$

The implied rule [8] is a random dynamic regime by the classification given in Section 2 as $f^{\text{obs}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ will generally be a nondegenerate density.

Finally, while the extended g-formula [6] and the (non-extended) g-formula [3] associated with the random dynamic regime [7] require the same positivity condition by their equivalence, the conditions required for risk identification under an intervention $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ and under [7] are not generally equivalent. In particular, the identifying condition defined by Richardson and Robins (2013) for an intervention $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is generally more stringent than that required for the random dynamic mechanism [7], the latter of which is equivalent to the exchangeability condition [1] of Section 3. An exception is under the null; here the two conditions are equivalent. For details, see Section 5.6 of Richardson and Robins (2013) and Appendix B.

5 Estimating an intervention risk using observational data

In low-dimensional settings, we can non-parametrically estimate expression [3] by first enumerating all possible treatment and confounder histories under a specified intervention $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, calculating each component proportion, and then taking the overall sum. When $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is implied by the sum [7] then we must additionally enumerate all possible natural treatment histories and calculate this implied rule. In high-dimensional settings, such that K is large and/or there are continuously measured covariates, such an approach is not feasible. In this case, parametric or semi-parametric approaches may be used.

5.1 Parametric estimation

Robins (1986) described a parametric estimator of the (non-extended) g-formula given in expression [3] which involves parametrically modelling each component density and using Monte Carlo simulation to approximate the sum over all possible histories under an intervention that does not depend on the natural value of treatment as in Section 2. Robins et al. (2004) and Taubman et al. (2009) generalized this algorithm to allow for an intervention that depends on the natural value of treatment as in Section 4. Briefly, this more general approach involves the following steps:

1. Parametrically estimate the joint density of natural treatment and confounders at each follow-up time (except baseline) given survival and past treatment and confounders.
2. Parametrically estimate the probability of failure at each follow-up time given survival and past measured treatment and confounders.
3. Recursively, for each $k = 0, \dots, K$
 - (a) Set baseline confounders and natural treatment to the observed sample values. For $k > 0$, generate time k confounders and natural treatment based on the estimated model coefficients and previously generated treatment and confounders under intervention.
 - (b) Assign time k treatment under intervention based on the rule of interest which may be an explicitly specified $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ depending at most on the past measured confounders or an explicitly specified $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ depending on the natural value of treatment at k .
 - (c) Calculate the discrete failure hazard at $k + 1$ given *only* past generated treatment and confounders under the intervention (ignoring the natural treatment value).
4. Calculate the cumulative probability of failure by $K + 1$ using the $k + 1$ specific failure hazards for each generated treatment and confounder history under intervention.
5. Calculate the average cumulative probability of failure by $K + 1$ over all generated intervention histories.

Robins et al. (2004), Taubman et al. (2009), Lajous et al. (2013), Danaei et al. (2013) and García-Aymerich et al. (2014) have applied the above approach to estimate failure risk under time-varying threshold interventions on lifestyle factors that depend on the natural value of treatment in various observational studies including the NHS, the Offspring Framingham Heart Study and the Health Professionals Follow-up Study. A more technical description of this algorithm is given in Appendix C and may be implemented using a SAS macro publicly available at www.hsph.harvard.edu/causal/software.

This estimation algorithm effectively ignores that, for an intervention $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, the implied treatment rule depending only on the measured confounders is $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ as defined by the marginalization [7]. The natural value of treatment A_k^* is generated at each k regardless of whether the explicit intervention of interest depends on it or not. If the intervention does not depend on it, then A_k^* is generated but not used. Note that expression [3] can be rewritten as

$$\begin{aligned} & \sum_{\bar{a}_k} \sum_{\bar{a}_k} \sum_{\bar{l}_k} \sum_{k=0}^K \Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{D}_k = 0] \\ & \times \prod_{j=0}^k \{ \Pr[D_j = 0 | \bar{L}_{j-1} = \bar{l}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{D}_{j-1} = 0] \\ & \times h^{\text{user}}(\bar{a}_j, a_j^*, \bar{l}_j) \times f(z_j^* | \bar{l}_{j-1}, \bar{a}_{j-1}, \bar{D}_j = 0) \} \end{aligned} \quad [9]$$

where $f(z_k^* | \bar{l}_{k-1}, \bar{a}_{k-1}, \bar{D}_k = 0)$ is the joint density of Z_k^* , an arbitrarily ordered vector including A_k^* and L_k , conditional on survival and past treatment and confounder history. Here, $h^{\text{user}}(\bar{a}_k, a_k^*, \bar{l}_k)$ may be selected as an explicitly specified $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ that may at most depend on the measured confounder history as in the examples given in Section 2 or an explicitly specified $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. Under the latter choice, expression [9] is equivalent to expression [3] with $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ defined by eq. [7] and, thus (by the arguments of Section 4), also equivalent to the extended g-formula [6].

5.2 Semi-parametric estimation

The parametric g-formula may be subject to bias due to model misspecification and to the g-null paradox (Robins and Wasserman 1997). As an alternative, several authors have described semi-parametric estimators

of risk under explicitly specified random dynamic regimes that may, at most, depend on the measured confounder history (Murphy et al. 2001; Cain et al. 2010; Stitelman et al. 2010; Díaz Muñoz and van der Laan 2012). These approaches do not require specification of the likelihood and may be more robust to model misspecification. Here, we describe how an inverse-probability weighted (IPW) risk estimator can be extended to implied random dynamic regimes such as that defined by eq. [7].

Following Cain et al. (2010), consider the following IPW estimator of risk by $K + 1$ under an explicitly specified $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. Let $\hat{\psi}$ be the solution to the estimating equation

$$\sum_{i=1}^n \sum_{k=0}^K U_{i,k}(\psi, \hat{\alpha}) = 0 \quad [10]$$

with respect to ψ where

$$U_{i,k}(\psi, \hat{\alpha}) = (D_{i,k+1} - \lambda(k, \psi))(1 - D_{i,k})W_{i,k}(\hat{\alpha})$$

$\lambda(k, \psi)$ is a flexible function of k and the parameter vector ψ and

$$W_{i,k}(\alpha) = \frac{\prod_{j=0}^k f^{\text{int}}(A_{i,j}|\bar{L}_{i,j}, \bar{A}_{i,j-1}, \bar{D}_j = 0)}{\prod_{j=0}^k f^{\text{obs}}(A_{i,j}|\bar{L}_{i,j}, \bar{A}_{i,j-1}, \bar{D}_j = 0; \alpha)} \quad [11]$$

with $\hat{\alpha}$ the MLE of α given the model $f^{\text{obs}}(a_j|\bar{l}_j, \bar{a}_{j-1}, \bar{D}_j = 0; \alpha)$ for the observed treatment density as defined in eq. [2] with α_0 the true population value of α .

If this treatment model is correctly specified and there exists ψ_0 such that

$$\lambda(k, \psi_0) = \frac{E[D_{k+1}(1 - D_k)W_k(\alpha_0)]}{E[(1 - D_k)W_k(\alpha_0)]} \quad [12]$$

then we have

$$E[U_k(\psi_0, \alpha_0)] = 0 \quad [13]$$

for all k and the estimator $\hat{\psi}$ consistent for ψ_0 and asymptotically normal. Note that, under these assumptions, the g-formula [3] is equivalent to

$$\sum_{k=0}^K \lambda(k, \psi_0) \prod_{j=0}^{k-1} (1 - \lambda(j, \psi_0)) \quad [14]$$

The IPW estimator of expression [3] is then given by the plug-in estimator where ψ_0 in expression [14] is replaced by the IPW estimate $\hat{\psi}$. Analogous to Cain et al. (2010), we might impose a Cox marginal structural model if few individuals are following $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ to borrow information from individuals following other interventions (Robins 2000). Note that in the case where $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ corresponds to a single deterministic regime g , $\prod_{j=0}^k f^{\text{int}}(A_{i,j}|\bar{L}_{i,j}, \bar{A}_{i,j-1}, \bar{D}_j = 0)$ in the numerator of the weight [11] becomes $\prod_{j=0}^k I(A_{i,j} = A_j^g)$ which renders an estimating equation more familiar to some readers.

To extend the IPW estimator described above (and related semi-parametric approaches) to explicitly specified interventions of the form $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, we must replace $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ for $(\bar{A}_{i,k}, \bar{L}_{i,k}) = (\bar{a}_k, \bar{l}_k)$ in the weight [11] with the marginalization [7] for every possible treatment and confounder history observed in the data. Thus, in contrast to the parametric g-formula estimator described above, semi-parametric methods cannot “ignore” the fact that the explicit treatment rule of interest $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ implies the marginalization [7].

In general, the computational complexity of this marginalization will, of course, depend on the form of $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ and $f^{\text{obs}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. For example, the implied rule [8] requires knowledge of $f^{\text{obs}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ which must be estimated to calculate the denominator of the weights. If this is based on a parametric model then one must also estimate $\text{Pr}^{\text{obs}}(A_k \leq a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ based on that model, which will be used for the numerator of the weights for any subject with $A_k = 30$.

Other authors have considered semi-parametric estimators of risk under random dynamic regimes that might be interpreted in terms of implied random dynamic regimes based on an explicit deterministic mechanism depending on the natural value of treatment (Díaz Muñoz and van der Laan 2011, 2012; Haneuse and Rotnitzky 2013). For example, Díaz Muñoz and van der Laan (2012) considered various semi-parametric estimators of risk under a random dynamic regime on a point treatment that somehow shifts the observed treatment density by a certain amount. They allowed this shift to, at most, depend on values of the measured confounders, considering interventions on physical activity as a particular example.

Specifically, extending to our more general time-varying setting, this shift $\delta(\bar{l}_k, \bar{a}_{k-1})$ could be achieved by the following mechanism: “On each day k , if a subject with treatment and confounder history $(\bar{l}_k, \bar{a}_{k-1})$ has exercised a_k^* minutes under no intervention by the end of the day then have her, instead, exercise $a_k^* - \delta(\bar{l}_k, \bar{a}_{k-1})$ on that day”. If we fix $\delta(\bar{l}_k, \bar{a}_{k-1}) = -30$ for all $(\bar{l}_k, \bar{a}_{k-1})$, then this intervention maintains exercise at or above 30 minutes per day for all subjects and corresponds to a particular choice of $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, such that $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) = 1$ if $a_k = a_k^* - \delta(\bar{l}_k, \bar{a}_{k-1})$ and 0 otherwise.

For this choice of $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, the marginalization [7] is conveniently equivalent to $f^{\text{obs}}(a_k - \delta(\bar{l}_k, \bar{a}_{k-1})|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ for all values of a_k . As noted by Díaz Muñoz and van der Laan (2012), this choice of $f^{\text{int}}(a_k|\bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ may also render practical violations of positivity less influential on the performance of the estimators. See Petersen et al. (2012) for a detailed discussion of the potential influence of practical positivity violations on various estimators.

6 A plausible approximation to interventions that depend on the natural value of treatment

In the previous sections, we have considered hypothetical interventions at k that depend on the natural value of treatment also at k . Such interventions are generally not plausible in practice. For example, once an individual has exercised less than 30 minutes by the end of day k , she cannot, instead, have exercised 30 minutes by the end of that day. It follows that, even given “perfect” conditions (e.g. identifiability and no model misspecification) it is unclear how to use observational estimates associated with such interventions to inform real-world future policy or the design of future randomized experiments.

We might, however, approximate such interventions with a plausible (implementable) experiment. Let X_k be a subject’s stated *intention* with respect to treatment on day k measured at the start of that day (e.g. *intended* daily minutes of exercise at the start of day k). Given an intervention $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, denote $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ as a plausible approximation that assigns treatment according to the same rule as $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ at each k but replacing A_k^* with X_k .

For example, given the threshold intervention on exercise of Taubman et al. (2009) characterized by [5], a plausible approximation is “If a subject’s intention at the start of day k is to exercise less than 30 minutes on that day then ensure she exercises exactly 30 minutes by the end of day k . Otherwise, ensure she exercises her intended amount” or

$$\begin{aligned} \text{If } x_k \leq 30 \text{ then } f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= 1 \text{ if } a_k = 30 \text{ and } 0 \text{ o.w.} \\ \text{If } x_k > 30 \text{ then } f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0) &= 1 \text{ if } a_k = x_k \text{ and } 0 \text{ o.w.} \end{aligned} \quad [15]$$

Suppose treatment is assigned according to $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ and the following assumption held:

Natural value of treatment assumption: Under any intervention, for all k , every subject's intended minutes of exercise at the start of day k is equal to what her subsequent behavior would be on that day were the intervention based on intention discontinued right before k .

Under this assumption, the plausible rule $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is not an approximation but exactly equal to $f^d(a_k|a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$. Further, under the reasonable assumption that intention has no direct effect on the outcome except through future treatment, the risks by $K + 1$ under these two rules will be equivalent. Thus, all identification and estimation results of Sections 4 and 5 apply.

In an actual experiment where treatment is assigned according to $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$, it is impossible to empirically examine whether this assumption holds, even given X_k is measured. However, in an observational study, this relationship *can* be examined given X_k is measured. In particular, in an observational study (i.e. under no intervention), the natural value of treatment assumption implies that for each subject and all k

$$\text{If } X_k = x_k \text{ and } A_k^* = a_k^* \text{ then } x_k = a_k^* \quad [16]$$

Here, again, the natural value of treatment A_k^* is equivalent to the measured treatment A_k for all subjects as no intervention is made. Note that, while assumption [16] implies that the natural value of treatment assumption holds for the observational study, assumption [16] does not guarantee this assumption will hold under an intervention $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$.

Finally, we point out that when assumption [16] does not hold, $f^d(a_k|x_k, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is simply an example of a deterministic dynamic regime g by the classification given in Section 2 with L_k in Section 2 replaced with (X_k, L_k) . This deterministic dynamic regime g is specifically defined such that $a_k^g = 30$ if $x_k \leq 30$ and $a_k^g = x_k$ otherwise. Further, by the arguments of Section 3, given the assumptions of that section for this choice of g , risk under this regime is identified by the deterministic regime g -formula [4], again replacing L_k with (X_k, L_k) . Note that, in this setting, any of the conditional densities in the g -formula [4] may depend on \bar{X}_k without restriction.

By contrast, if assumption [16] holds in the observational study, positivity as defined in condition [2] immediately fails for this g . Specifically, $a_k^g \neq x_k$ for $x_k < 30$ because we must have $a_k^g \geq 30$ for all k and (\bar{x}_k, \bar{l}_k) under this definition of g . Therefore, given [16], $f^{\text{obs}}(a_k^g|\bar{x}_k, \bar{l}_k, \bar{a}_{k-1}^g, \bar{D}_k = 0) = 0$ whenever $x_k < 30$. As a consequence of this positivity violation, $\Pr[D_{k+1} = 1|\bar{X}_k = \bar{x}_k, \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k^g, \bar{D}_k = 0]$ in expression [4] is undefined for all histories such that $x_k < 30$, $k = 0, \dots, K$.

7 Conclusions

In this paper, we showed the equivalence between the extended g -formula associated with an intervention that depends on the natural value of treatment and the (non-extended) g -formula of Robins (1986) associated with a particular random dynamic regime that does not depend on this value. This equivalence immediately gives a sufficient positivity condition that guarantees the extended g -formula is well-defined. This positivity result, coupled with the results of Richardson and Robins (2013), now provides a formal causal framework for previously published applications of the parametric extended g -formula to estimate risk under threshold interventions in observational studies. It also immediately gives semi-parametric alternatives to the parametric extended g -formula. Finally, we considered limits on the practical implementation of threshold interventions along with possible real-world approximations.

The assumption of positivity is often informally described as the assumption that there are at least some subjects in the observational study who are observed to follow the hypothetical intervention of interest

within every possible level of the “past”. By this understanding, it would appear that positivity *must* be violated for the threshold intervention on exercise considered by Taubman et al. (2009). Specifically, no subject who exercised less than 30 minutes on day k can be following the intervention at k . Our positivity result makes clear that, given appropriate identification conditions, it is not necessary to observe such patterns in the observational study. It is only necessary to observe some individuals following the implied random dynamic regime [7].

Appendix A

Representing the g-formula characterized by a random dynamic regime as a weighted average of deterministic regimes

Given $g \in \mathcal{G}$, let $f_k(a_k^g | \bar{a}_{k-1}^g, \bar{l}_k)$ equal the intervention density $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ evaluated at \bar{a}_k^g . In the following, we assume L_k is discrete and we choose an ordering such that $\{l_{k,1}, \dots, l_{k,|\mathcal{L}_k|}\} = \mathcal{L}_k$, with \mathcal{L}_k the support of L_k and $|\mathcal{L}_k|$ its cardinality.

Let $q_k^g(\bar{l}_{k-1}) = \prod_{h=1}^{|\mathcal{L}_k|} f_k(a_k^g | \bar{a}_{k-1}^g, \bar{l}_{k-1}, l_{k,h})$. For $k = K-1, \dots, 0$, let

$$q_k^g(\bar{l}_{k-1}) = \prod_{h=1}^{|\mathcal{L}_k|} f_k(a_k^g | \bar{a}_{k-1}^g, \bar{l}_{k-1}, l_{k,h}) q_{k+1}^g(\bar{l}_{k-1}, l_{k,h})$$

Let $wt(g) = q_0^g$.

By Lemma 4.2 of Robins (1986), given a particular $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ $k \leq K$ defining $wt(g)$ as above for all $g \in \mathcal{G}$, then one-minus expression [3] equals

$$\sum_{g \in \mathcal{G}} wt(g) \text{ “Pr” } [D_{K+1}^g = 0]$$

where “Pr” $[D_{K+1}^g = 0]$ is equivalent to one-minus expression [4] and $\sum_{g \in \mathcal{G}} wt(g) = 1$.

Simplified numerical example

Figure 1 depicts a hypothetical sequentially randomized trial where treatment is assigned at each time based on a particular intervention density $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ by a structural tree graph (Robins 1986). For numerical simplicity, we will consider a short follow-up with $K = 1$ and all binary treatment and covariates. For additional simplicity, we will assume that no subject fails prior to the end of follow-up (i.e. $\bar{D}_1 = 0$ for all subjects). We will also assume that all subjects have the same value of the baseline covariate $L_0 = l_0$. The intervention density is defined by the probability of receiving a given level of treatment given the past read directly off the graph. These probabilities imply that $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ corresponds to a random dynamic regime. For example, following the top branch of the graph, the probability of receiving treatment at $k = 1$ given $(l_0, a_0 = 1, l_1 = 1)$ is $\frac{10}{20}$ or 0.5.

The survival probability for the disease of interest in this hypothetical sequentially randomized trial is simply the overall proportion of those who did not get the disease at the end of follow-up out of the total number at risk at baseline. Specifically, 52 subjects at the end have $D_2 = 0$ out of the 100 subjects at risk at baseline; thus survival in this hypothetical trial characterized by $f^{\text{int}}(a_k | \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ is $\frac{52}{100}$. We will now show $\frac{52}{100}$ is equivalent to a weighted average of the g-formula for survival over all deterministic regimes that it is possible to follow in this hypothetical sequentially randomized trial with weights defined as in the previous section.

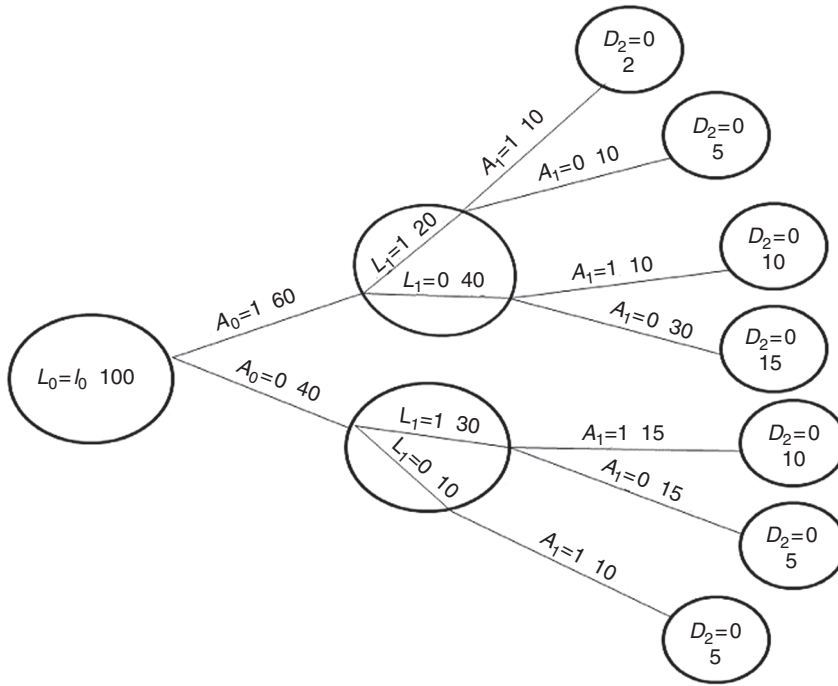


Figure 1 A hypothetical sequentially randomized trial for $K = 1$ and all binary (A_0, L_1, A_1)

First the set \mathcal{G} contains the following subset of deterministic regimes $(g_1, g_2, g_3, g_4, g_5, g_6)$:

- g_1 : $(a_0^{g_1}, a_1^{g_1}) = (0, 1)$; the static regime “do not treat at time 0; treat at time 1”
- g_2 : $(a_0^{g_2}, a_1^{g_2}) = (1, 0)$; the static regime “treat at time 0; do not treat at time 1”
- g_3 : $(a_0^{g_3}, a_1^{g_3}) = (1, 1)$; the static regime “always treat”
- g_4 : $(a_0^{g_4}, a_1^{g_4}) = (0, 1 - l_1)$; the dynamic regime “do not treat at time 0; if $l_1 = 1$ then do not treat at time 1; otherwise treat at time 1”
- g_5 : $(a_0^{g_5}, a_1^{g_5}) = (1, 1 - l_1)$; the dynamic regime “treat at time 0; if $l_1 = 1$ then do not treat at time 1; otherwise treat at time 1”
- g_6 : $(a_0^{g_6}, a_1^{g_6}) = (1, l_1)$; the dynamic regime “treat at time 0; if $l_1 = 0$ then do not treat at time 1; otherwise treat at time 1”

Note that \mathcal{G} contains additional deterministic regimes but we exclude these from the above subset as, for some covariate values, we observe no individuals following these regimes in the trial depicted in Figure 1. For example, in this trial, we observe no individuals who are untreated at both time 0 and time 1 with $L_1 = 0$. Any deterministic static or dynamic regime that allows this treatment and covariate pattern will contribute a zero weight by the definition of the previous section. Examples of deterministic regimes that would contribute a zero weight are $g_7 = (0, 0)$ and $g_8 = (0, l_1)$.

Using the definition of the previous section such that, again, $f_k(a_k^g | \bar{a}_{k-1}^g, \bar{l}_k) \equiv f^{\text{int}}(a_k^g | \bar{l}_k, \bar{a}_{k-1}^g, \bar{D}_k = 0)$, we define $wt(g)$ as follows for each g in the subset above:

$$wt(g) = f_0(a_0^g | l_0) \times f_1(a_1^g | l_1 = 0, a_0^g, l_0) \times f_1(a_1^g | l_1 = 1, a_0^g, l_0)$$

Specifically, we have by Figure 1:

- $wt(g_1) = \frac{40}{100} \times \frac{10}{10} \times \frac{15}{30} = \frac{1}{5}$
- $wt(g_2) = \frac{60}{100} \times \frac{30}{40} \times \frac{10}{20} = \frac{9}{40}$

$$\begin{aligned}
- \quad wt(g_3) &= \frac{60}{100} \times \frac{10}{40} \times \frac{10}{20} = \frac{3}{40} \\
- \quad wt(g_4) &= \frac{40}{100} \times \frac{10}{10} \times \frac{15}{30} = \frac{1}{5} \\
- \quad wt(g_5) &= \frac{60}{100} \times \frac{10}{40} \times \frac{10}{20} = \frac{3}{40} \\
- \quad wt(g_6) &= \frac{60}{100} \times \frac{30}{40} \times \frac{10}{20} = \frac{9}{40}
\end{aligned}$$

We leave to the reader to confirm that the sum of these weights is one.

Each “Pr” [$D_2^g = 0$] is defined by the g-formula

$$\sum_{l_1} \Pr[D_2 = 0 | \bar{A}_1 = \bar{a}_1^g, L_1 = l_1, L_0 = l_0] \times f(l_1 | a_0^g) \times f(l_0)$$

where $f(l_0) = 1$. Here, we evaluate this expression for all g in the subset:

$$\begin{aligned}
- \quad \text{“Pr” } [D_2^{g_1} = 0] &= \left(\frac{5}{10} \times \frac{10}{40} \right) + \left(\frac{10}{15} \times \frac{30}{40} \right) = \frac{5}{8} \\
- \quad \text{“Pr” } [D_2^{g_2} = 0] &= \left(\frac{15}{30} \times \frac{40}{60} \right) + \left(\frac{5}{10} \times \frac{20}{60} \right) = \frac{1}{2} \\
- \quad \text{“Pr” } [D_2^{g_3} = 0] &= \left(\frac{40}{60} \times \frac{10}{10} \right) + \left(\frac{2}{10} \times \frac{20}{60} \right) = \frac{44}{60} \\
- \quad \text{“Pr” } [D_2^{g_4} = 0] &= \left(\frac{5}{10} \times \frac{10}{40} \right) + \left(\frac{5}{15} \times \frac{30}{40} \right) = \frac{3}{8} \\
- \quad \text{“Pr” } [D_2^{g_5} = 0] &= \left(\frac{10}{10} \times \frac{40}{60} \right) + \left(\frac{5}{10} \times \frac{20}{60} \right) = \frac{5}{6} \\
- \quad \text{“Pr” } [D_2^{g_6} = 0] &= \left(\frac{15}{30} \times \frac{40}{60} \right) + \left(\frac{2}{10} \times \frac{20}{60} \right) = \frac{6}{15}
\end{aligned}$$

Finally we have that $\sum_{g \in \mathcal{G}} \text{“Pr”}[D_2^g = 0] wt(g)$ is equivalent to

$$\left(\frac{5}{8} \times \frac{1}{5} \right) + \left(\frac{1}{2} \times \frac{9}{40} \right) + \left(\frac{44}{60} \times \frac{3}{40} \right) + \left(\frac{3}{8} \times \frac{1}{5} \right) + \left(\frac{5}{6} \times \frac{3}{40} \right) + \left(\frac{6}{15} \times \frac{9}{40} \right) = \frac{52}{100}$$

Appendix B

Richardson and Robins (2013) defined a graphical condition based on a d-separation relation (i.e. checking for the absence of “backdoor paths”) that gives general identification for any intervention considered in the classification of Section 2 or an intervention that depends on the history of the natural value of treatment using the (non-extended) g-formula and extended g-formula, respectively. They further show that, given an appropriate consistency assumption, this graphical condition for identification implies an exchangeability condition analogous to condition [1] given in Section 3. In the restricted case, where the intervention does not depend on the history of the natural value of treatment, then this condition is equivalent to condition [1]. We refer the reader to Richardson and Robins (2013) for details of this more general exchangeability condition.

The d-separation condition of Richardson and Robins (2013) is applied to a transformation of a causal DAG (Spirtes et al. 1993; Pearl 2000) representing assumptions on the underlying data generating process that produced the data in the observational study. Richardson and Robins (2013) call this transformation a Single World Intervention Graph (SWIG). We now illustrate how to evaluate identification for different interventions on a time-varying treatment under a simple set of underlying observed data generating

assumptions using SWIGs. The examples given here are similar to examples depicted in figures 19 and 21 in Richardson and Robins (2013).

Remark on notation: In describing how to construct a SWIG associated with any hypothetical intervention under an assumed observed data generating mechanism we will adopt, for this section of the appendix only, the notation of Richardson and Robins (2013). This will create two inconsistencies with notation used in the main text which we now describe, along with our motivation behind this choice. Specifically, in this appendix, we will denote any hypothetical dynamic intervention as g which may, or may not, depend on the natural value of treatment. In the main text, this notation was reserved only for deterministic regimes (dynamic or static) that do not depend on the natural value of treatment. Further, we will change the meaning of one instance of counterfactual notation used in the main text. In particular, A_k^g was used in the main text to denote the counterfactual value of treatment assigned under an intervention g . Here, to be consistent with Richardson and Robins (2013), A_k^{+g} will be used to denote this counterfactual, and A_k^g will, alternatively, be used to denote the counterfactual *natural* value of treatment under g .

We chose not to adopt this more complex notational convention of Richardson and Robins (2013) in the main text as the primary results regarding positivity and semi-parametric estimation of the main text do not require formalization of a counterfactual natural value of treatment. This allows simpler notation in the main text that is consistent with previous work on interventions that do not depend on the natural value of treatment. It also allows a notational bridge to the motivating work by Robins et al. (2004) and Taubman et al. (2009). While we could have used notation fully consistent with the main text in this section of the appendix, we chose to adopt that of Richardson and Robins (2013), the foundational paper on SWIGs, in order to avoid confusion within the newly emerging literature on this topic. We now proceed with our examples.

Consider the simple time-varying observational study depicted in the causal DAG of Figure 2(i) where, as in the simplified numerical example of Appendix A, we assume a short follow-up ($K = 1$) and that no subject fails prior to the end of follow-up. In Figure 2(i), H_1 represents an unmeasured common cause of $A_0 = A_0^*$ and $A_1 = A_1^*$ and H_2 an unmeasured common cause of the covariate L and the outcome D .

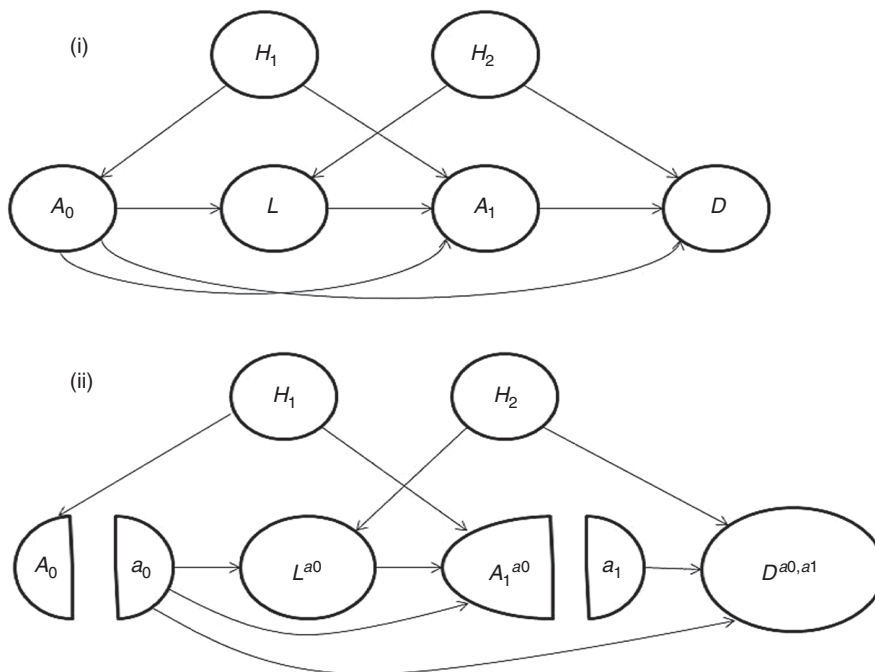


Figure 2 (i) A causal DAG representing underlying data generating assumptions for a simple time-varying observed data structure. (ii) A SWIG $\mathcal{G}(\bar{a})$ based on a transformation of the causal DAG in (i)

The d-separation condition of Richardson and Robins (2013) is evaluated for a given dynamic intervention g based on the following sets of transformations applied to a causal DAG:

1. Split each treatment node at k into two nodes with one node containing the natural value of treatment at k and the other a constant value a_k
2. Index all random variables after time 0 as counterfactuals under a static deterministic intervention \bar{a} , including the natural value of treatment.
3. All arrows out of the observed A_k on the original DAG should now be out of a_k and all arrows into the observed A_k on the original DAG should now be into the counterfactual natural value of treatment at k $A_k^{\bar{a}_{k-1}}$ (equivalent to the observed A_0 at baseline as no intervention has yet been made).

Figure 2(ii) depicts a SWIG derived from the causal DAG in Figure 2(i) under this first set of transformations. A SWIG constructed from these transformations is a non-dynamic SWIG denoted $\mathcal{G}(\bar{a})$.

To assess identification for a dynamic intervention g , we apply the following additional transformations:

1. Index all counterfactuals on $\mathcal{G}(\bar{a})$ by g rather than by \bar{a} or a subvector thereof
2. Replace each constant a_k with the counterfactual A_k^{+g}
3. Add dashed arrows from any variable temporally prior to A_k^{+g} into A_k^{+g} if treatment at k is assigned by this variable under the intervention g

A SWIG constructed by applying this second set of transformations to $\mathcal{G}(\bar{a})$ is a dynamic SWIG denoted $\mathcal{G}(g)$.

Richardson and Robins (2013) prove that a dynamic intervention g is identified if, for each time k , A_k^g and D^g are d-separated conditional on $\bar{A}_{k-1}^g, \bar{L}_k^g, \bar{A}_{k-1}^{+g}$ in $\mathcal{G}(g)$ once we apply the additional k -specific transformation of removing all dashed arrows out of A_k^g . This final transformation is only required to evaluate identification when g depends on the history of the natural value of treatment. Richardson and Robins (2013) define this last k -specific transformation of the SWIG $\mathcal{G}(g)$ as a new SWIG associated with what they term a *perturbed regime* at k . Richardson and Robins (2013) note that the aforementioned d-separation holds if and only if there is no unblocked backdoor path between A_k^g and D^g conditional on the same set of variables.

Figure 3 depicts two dynamic SWIGs created from transformations of the non-dynamic SWIG of Figure 2 (ii) which differ only by their dependence on the history of the natural value of treatment. The intervention under Figure 3(i) does not depend on any function of the history of the natural value of treatment by the absence of any dashed arrows from A_0 into either A_0^{+g} or A_1^{+g} and the absence of a dashed arrow from A_1^g into A_1^{+g} . By contrast, the intervention under Figure 3(ii) depends on this history by the presence of dashed arrows from A_0 into A_0^{+g} and A_1^g into A_1^{+g} .

By the d-separation condition of Richardson and Robins (2013), we can see that the intervention g in Figure 3(i), under which treatment assignment does not depend on the history of the natural value of treatment, is identified under our data generating assumptions. Specifically, there are no unblocked backdoor paths between A_0 and D^g . Further, conditional on L^g, A_0^{+g} and A_0 there is no unblocked backdoor path between A_1^g and D^g .

By contrast, we can see that the intervention g in Figure 3(ii), under which treatment assignment does depend on the history of the natural value of treatment, is not identified under our data generating assumptions. Again, applying the d-separation condition of Richardson and Robins (2013), following the transformation to the $k = 0$ perturbed regime (i.e. removal of the dashed arrow from A_0 into A_0^{+g}), we still have the unblocked backdoor path $A_0 \leftarrow H_1 \rightarrow A_1^g \rightarrow A_1^{+g} \rightarrow D^g$.

These examples illustrate that even given we have identification for an intervention that does not depend on the history of the natural value of treatment – for example, the random dynamic intervention [7] – it is not guaranteed that we will have identification for an intervention that does depend on some function of this history – for example, the threshold interventions of Taubman et al. (2009) – for all

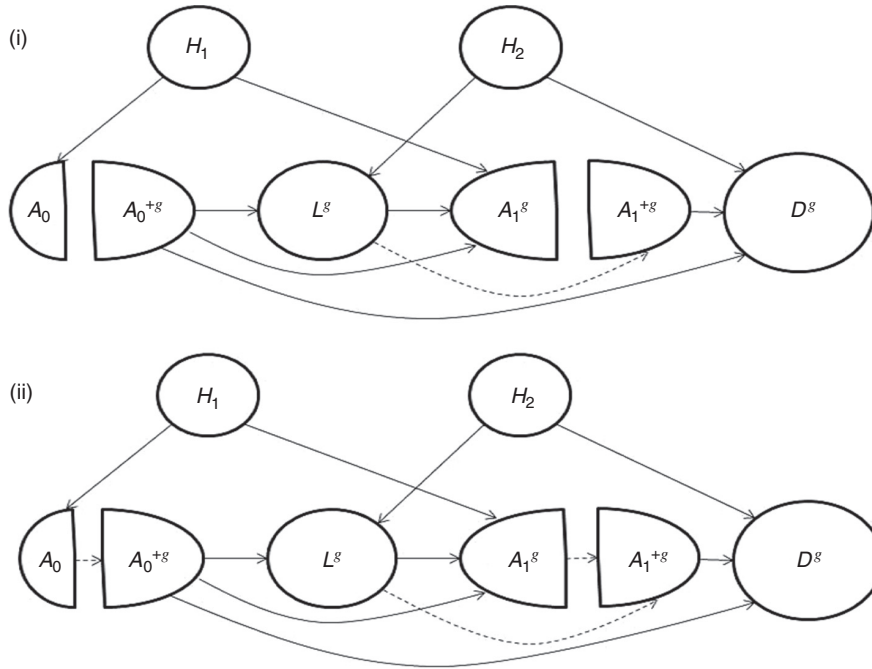


Figure 3 (i) A SWIG $\mathcal{G}(g)$ under which the intervention g does not depend on the history of the natural value of treatment. (ii) A SWIG $\mathcal{G}(g)$ under which the intervention g does depend on some function of this history

underlying observed data generating mechanisms. However, under additional restrictions on the original data generating assumptions depicted in Figure 2(i), we achieve identification for both of the dynamic regimes considered in Figure 3. For example, this would be the case under either of the following restrictions applied to our initial set of data generating assumptions in Figure 2(i):

1. The null is true (i.e. the arrows from A_0 and A_1 into D are removed).
2. The common cause H_1 of A_0 and A_1 is removed.

Appendix C

Let $Z_k^* = (Z_{1,k}^*, \dots, Z_{p,k}^*)$ be an arbitrary permutation of the p components in (L_k, A_k^*) , noting that

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

in expression [9] for any $k = 0, \dots, K$ where $(m_{1,k}^{z_k^*}, \dots, m_{p,k}^{z_k^*})$ are conditional densities based on the factorization implied by the user-selected permutation.

For user-chosen K and $h^{\text{user}}(\bar{a}_k, a_k^*, \bar{l}_k)$, $k = 0, \dots, K$ we do the following:

Step I: parametric modelling of conditional densities

Using the n individuals in the data set, for each $k = 0, \dots, K$:

1. If $k > 0$, fit parametric models for the conditional densities $m_{j,k}^{z_k^*}$, $j = 1, \dots, p$.
2. Fit a parametric model for the conditional probability of the outcome $\Pr[D_{k+1} = 1 | \bar{L}_k = \bar{l}_k, \bar{A}_k = \bar{a}_k, \bar{D}_k = 0]$

Step II: Monte Carlo simulation under the user-chosen $h^{\text{user}}(\bar{a}_k, a_k^*, \bar{l}_k)$

For $k = 0, \dots, K$ and $v = 1, \dots, n$:

1. 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 I.1 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.
2. Assign the treatment $a_{k,v}$ according to the user-chosen intervention. For example, for $h^{\text{user}}(\bar{a}_k, a_k^*, \bar{l}_k)$ chosen as $f^d(a_k | a_k^*, \bar{l}_k, \bar{a}_{k-1}, \bar{D}_k = 0)$ we set $a_{k,v} = 30$ if $a_{k,v}^* \leq 30$ and otherwise set $a_{k,v} = a_{k,v}^*$.
3. Estimate the probability of failure 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 from Step I.2.

STEP III: computation of disease risk by $k + 1$ under $h^{\text{user}}(\bar{a}_k, a_k^*, \bar{l}_k)$

Estimate expression [9], or equivalently expression [3], as

$$\frac{1}{n} \sum_{v=1}^n \sum_{k=0}^K \hat{\text{Pr}}[D_{k+1} = 1 | \bar{L}_k = \bar{l}_{k,v}, \bar{A}_k = \bar{a}_{k,v}, \bar{D}_k = 0] \times \prod_{j=0}^k \left\{ 1 - \hat{\text{Pr}}[D_j = 1 | \bar{L}_{j-1} = \bar{l}_{j-1,v}, \bar{A}_{j-1} = \bar{a}_{j-1,v}, \bar{D}_{j-1} = 0] \right\} \quad [17]$$

where

$$\hat{\text{Pr}}[D_{k+1} = 1 | \bar{L}_k = \bar{l}_{k,v}, \bar{A}_k = \bar{a}_{k,v}, \bar{D}_k = 0]$$

$k = 0, \dots, K$ is obtained in Step II.3.

As discussed in Young et al. (2011), both Steps I and II 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.

Funding: This research was funded by NIH grants R01 HL080644 and R37 AI032475.

References

- Cain, L. E., Robins, J. M., Lanoy, E., Logan, R., Costagliola, D., and Hernán, M. A. (2010). When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *International Journal of Biostatistics*, 6: Article 18.
- Danaei, G., Pan, A., Hu, F. B., and Hernán, M. A. (2013). Hypothetical lifestyle interventions in middle-aged women and risk of type 2 diabetes: a 24-year prospective study. *Epidemiology*, 24:122–128.
- Dawid, A. P. and Didelez, V. (2008). Identifying optimal sequential decisions. In: *Proceedings of the Twenty-Fourth Annual Conference on Uncertainty in Artificial Intelligence (UAI-08)*, D. McAllester and A. Nicholson (Eds.), 113–120. Corvallis, OR: AUAI Press.
- Dawid, A. P. and Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: a decision-theoretic overview. *Statistics Surveys*, 4:184–231.
- Díaz Muñoz, I. and van der Laan, M. J. (2011). Population intervention causal effects based on stochastic interventions. U.C. Berkeley Division of Biostatistics Working Paper Series. Paper 289. <http://www.bepress.com/ucbbiostat/paper289>
- Díaz Muñoz, I. and van der Laan, M. J. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68:541–549.
- García-Aymerich, J., Varraso, R., Danaei, G., Camargo, C. A., and Hernán, M. A. (2014). Incidence of adult-onset asthma after hypothetical interventions on body mass index and physical activity: an application of the parametric g-formula. *American Journal of Epidemiology*, 179(1):20–6.

- Haneuse, S. and Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30):5260–77.
- Hernán, M. A., Lanoy, E., Costagliola, D., and Robins, J. M. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology*, 98:237–242.
- Lajous, M., Willett, W. C., Robins, J. M., Young, J. G., Rimm, E., Mozaffarian, D., and Hernán, M. A. (2013). Changes in fish consumption in midlife and the risk of coronary heart disease in men and women. *American Journal of Epidemiology*, 178(3):382–391.
- Murphy, S. A., van der Laan, M. J., and Robins, J. M. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*, 96(456):1410–1423.
- Orellana, L., Rotnitzky, A., and Robins, J. M. (2010a). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *International Journal of Biostatistics*, 6:Article 7.
- Orellana, L., Rotnitzky, A., and Robins, J. M. (2010b). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part ii: proofs and additional results. *International Journal of Biostatistics*, 6:Article 8.
- Pearl, J. (2000). *Causality*. Cambridge, UK: Cambridge University Press.
- Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., and van der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, 210(1):31–54.
- Picciotto, S., Hernán, M. A., Page, J. H., Young, J. G., and Robins, J. M. (2012). Structural nested cumulative failure time models to estimate the effects of interventions. *Journal of the American Statistical Association*, 107(499):886–900.
- Richardson, T. S. and Robins J. M. (2013). Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper Number 128. <http://www.csss.washington.edu/Papers/>
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512. [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, J. M. (1997). Causal inference from complex longitudinal data. In: *Latent Variable Modeling and Applications to Causality*. Lecture Notes in Statistics 120, M. Berkane (Ed.), 69–117. New York: Springer.
- Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In: *Statistical Models in Epidemiology*, M. E. Halloran and D. Berry (Eds.), 95–133. New York: Springer.
- Robins, J. M. and Hernán, M. A. (2009). Estimation of the causal effects of time-varying exposures. In: *Advances in Longitudinal Data Analysis*, G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs (Eds.), 553–599. Boca Raton, FL: Chapman and Hall/CRC Press.
- Robins, J. M., Hernán, M. A., and Siebert, U. (2004). Effects of multiple interventions. In: *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors*, M. Ezzati, A. D. Lopez, A. Rodgers, and C. J. L. Murray (Eds.), 2191–2230. Geneva: World Health Organization.
- Robins, J. M. and Wasserman, L. (1997). Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, D. Geiger and P. Shenoy (Eds.), 409–420. San Francisco, CA: Morgan Kaufmann.
- Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction and Search*. New York: Springer.
- Stitelman, O. M., Hubbard, A. E., and Jewell, N. P. (2010). 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. Paper 264. <http://www.bepress.com/ucbbiostat/paper264>
- Taubman, S. L., Mittleman, M. A., Robins, J. M., and Hernán, M. A. (2008). Alternative approaches to estimating the effects of hypothetical interventions. In: *JSM Proceedings, Health Policy Statistics Section*, 4422–4426. Alexandria, VA: American Statistical Association.
- Taubman, S. L., Robins, J. M., Mittleman, M. A., and Hernán, M. A. (2009). Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *International Journal of Epidemiology*, 38(6):1599–1611.
- Tian, J. (2008). Identifying dynamic sequential plans. In: *Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, D. McAllester, P. Myllymaki (Eds.), 554–561. Corvallis, OR: AUAI Press.
- van der Laan, M. J., M. L. Petersen, and M. M. Joffe. (2005). History-Adjusted Marginal Structural Models and Statically-Optimal Dynamic Treatment Regimens. *International Journal of Biostatistics*, 10(1):Article 4.
- Young, J. G., Cain, L. E., Robins, J. M., O'Reilly, E. J., and Hernán, M. A. (2011). Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in Biosciences*. DOI: 10.1007/s12561-011-9040-7.