Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Separable Effects for Causal Inference in the Presence of Competing Events

Mats J. Stensrud, Jessica G. Young, Vanessa Didelez, James M. Robins & Miguel A. Hernán

To cite this article: Mats J. Stensrud, Jessica G. Young, Vanessa Didelez, James M. Robins & Miguel A. Hernán (2022) Separable Effects for Causal Inference in the Presence of Competing Events, Journal of the American Statistical Association, 117:537, 175-183, DOI: 10.1080/01621459.2020.1765783

To link to this article: https://doi.org/10.1080/01621459.2020.1765783

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 117, NO. 537, 175–183: Theory and Methods
https://doi.org/10.1080/01621459.2020.1765783

Separable Effects for Causal Inference in the Presence of Competing Events

Mats J. Stensrud[a,b], Jessica G. Young[c], Vanessa Didelez[d,e], James M. Robins[a,f], and Miguel A. Hernán[a,f,g]

[a]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA; [b]Department of Biostatistics, University of Oslo, Oslo, Norway; [c]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [d]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany; [e]Faculty of Mathematics/Computer Science, University of Bremen, Bremen, Germany; [f]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA; [g]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA

ABSTRACT

In time-to-event settings, the presence of competing events complicates the definition of causal effects. Here we propose the new separable effects to study the causal effect of a treatment on an event of interest. The separable direct effect is the treatment effect on the event of interest not mediated by its effect on the competing event. The separable indirect effect is the treatment effect on the event of interest only through its effect on the competing event. Similar to Robins and Richardson's extended graphical approach for mediation analysis, the separable effects can only be identified under the assumption that the treatment can be decomposed into two distinct components that exert their effects through distinct causal pathways. Unlike existing definitions of causal effects in the presence of competing events, our estimands do not require cross-world contrasts or hypothetical interventions to prevent death. As an illustration, we apply our approach to a randomized clinical trial on estrogen therapy in individuals with prostate cancer. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received February 2019
Accepted April 2020

KEYWORDS
Competing risks; Direct and indirect effects; Failure time analysis; Lifetime and survival analysis; Separable effects

1. Introduction

A competing event is any event that makes it impossible for the event of interest to occur. For example, consider a randomized trial to estimate the effect of a new treatment on the 3-year risk of prostate cancer in which 1000 individuals with prostate cancer were assigned to the treatment and 1000 to placebo. All participants adhered to the protocol and remained under follow-up. After 3 years, 100 individuals in the treatment arm and 200 in the placebo arm died of prostate cancer. Also, 150 individuals in the treatment arm and 50 in the placebo arm died of other causes (e.g., cardiovascular disease). Death from cardiovascular disease is a competing event for death from prostate cancer: individuals who die of cardiovascular disease cannot subsequently die of prostate cancer. When competing events are present, several causal estimands may be considered to define the causal effect of treatment on a time-to-event outcome (Young et al. 2020).

Consider first the total treatment effect (Young et al. 2020) defined by the contrast of the cumulative incidence (risk) (Prentice et al. 1978; Andersen et al. 2012) of the event of interest under different treatment values. In our example, the total treatment effect on death from prostate cancer is the contrast of the cumulative incidence of death from prostate cancer under treatment, consistently estimated by $\frac{100}{1000}$, and under placebo, consistently estimated by $\frac{200}{1000}$. Therefore, the estimate of the total treatment effect on the additive scale is $\frac{100}{1000} - \frac{200}{1000} = -0.1$, which indicates that treatment reduced the risk of death from prostate cancer.

However, in our trial, the interpretation of the total treatment effect on the event of interest is difficult because the treatment also increased the risk of the competing event. The estimate of the total effect of treatment on the competing event is $\frac{150}{1000} - \frac{50}{1000} = 0.1$ on the additive scale. Thus, it is possible that the beneficial effect of treatment on death from prostate cancer is simply a consequence of the harmful effect of treatment on death from other causes: when more people die from other causes, fewer people can die from prostate cancer. Note that this problem of interpretation cannot be solved by considering contrasts of hazard functions, such as cause-specific and subdistribution hazards, because these estimands are defined conditional on a post-treatment event (survival) and therefore do not generally have a causal interpretation (Hernán 2010; Young et al. 2020).

One way to deal with this problem is to consider a second causal estimand on the risk scale: the (controlled) direct effect of treatment on the event of interest had competing events been eliminated. This estimand corresponds to defining the competing events as censoring events (Young et al. 2020), and is sometimes denoted the marginal (net) distribution function. Unlike the total effect, identification of the controlled direct effect requires untestable assumptions even in an ideal randomized trial with perfect adherence and no loss to follow-up (Young et al. 2020). Also, this causal estimand often introduces a new conceptual challenge: the direct effect is not sufficiently well-defined because there is no scientific agreement as to which hypothetical intervention, if any, would eliminate the competing

CONTACT Mats J. Stensrud *mstensrud@hsph.harvard.edu* Department of Epidemiology, Harvard Chan School of Public Health, Harvard University, Boston, MA, USA.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.

events (Hernán 2016). For example, in our prostate cancer trial, no intervention has ever been proposed that can prevent all deaths from causes other than prostate cancer. As a byproduct of the ill-defined intervention to prevent competing events, effect estimates cannot be empirically verified—not even in principle—in a randomized experiment.

A third causal estimand is the survivor average causal effect (SACE) (Robins 1986), which is the total treatment effect (on the risk scale) in the principal stratum of patients who would never experience the competing event under either level of treatment (Robins 1986; Frangakis and Rubin 2002; Young et al. 2020). Unlike the total effect, the presence of competing events is not a problem when interpreting the SACE, because the SACE is restricted to subjects who do not experience competing events. However, identification of the SACE requires strong untestable assumptions, for example, about cross-world counterfactuals, even in a perfectly executed trial. Also, the SACE could never, even in principle, be confirmed in a real-world experiment as it will never be possible to observe the status of the competing event for the same individual under two different levels of treatment.

The problems of the previous estimands can be overcome in settings in which the treatment exerts its effect on the event of interest and its effect on the competing event through different causal pathways. Here, we define the separable direct and indirect effects for settings with competing events. Like the controlled direct effect and the SACE, identification of separable effects relies on untestable assumptions even when the treatment is randomized. However, unlike the controlled direct effect and the SACE, separable effects do not require conceptual interventions on competing events or knowledge of cross-world counterfactuals; the separable effects are well-defined if we can articulate a hypothetical decomposition of the treatment into two components. Therefore, in principle, they may be verified in a future experiment. Our definitions of separable effects and conditions for identifiability follow from the work of Robins and Richardson (2010) and Didelez (2019) on mediation: the pure (natural) direct effects (Robins and Greenland 1992) are extensively used in mediation analyses, but they require untestable cross-world independence assumptions and are often difficult to interpret, for example, in survival settings. Robins and Richardson (2010) proposed an alternative causal estimand that overcomes these problems by considering a decomposed treatment: unlike the pure direct effects, the decomposed treatment effects can be identified under assumptions that are in principle empirically testable. Moreover, it was shown by Didelez (2019) that the decomposed treatment effects are sensible estimands in survival settings.

We have organized the article as follows. In Section 2, we describe the observed data structure. In Section 3, we present a conceptual treatment decomposition and provide explicit examples to fix ideas. In Section 4, we formulate the causal estimand and define the new separable effects. In Section 5, we present conditions that allow for identifiability of the separable effects. In Section 6, we give three different estimators for the separable effects that can be implemented with standard statistical models, and we use data from a randomized clinical trial to estimate a direct effect of estrogen therapy on prostate cancer mortality. In Section 7, we provide a final discussion of the new estimands.

2. Observed Data Structure

We consider a study in which individuals are randomly assigned to a binary treatment $A \in \{0,1\}$ at baseline (e.g., $A = 1$ if assigned to treatment and $A = 0$ if assigned to placebo). Let $L \in \mathcal{L}$ denote a vector of individual pretreatment characteristics. For each of equally spaced discrete time intervals $k \in \{0,1,\ldots,K+1\}$, let Y_k and D_k denote indicators of an event of interest and a competing event by interval k, respectively. In our example, Y_k denotes death due to prostate cancer and D_k death from other causes by interval k. We adopt the convention that D_k is measured just before Y_k. If an individual experiences the competing event at time k without a history of the event of interest ($D_k = 1, Y_{k-1} = 0$), then all future values of the event of interest are zero. We can approximate a continuous time setting by choosing time intervals that are arbitrary small.

By definition, $D_0 \equiv Y_0 \equiv 0$, that is, no individual experiences any event during the initial interval. We use overbars to denote the history of a random variable, such that $\bar{Y}_k = (Y_1, Y_2, \ldots, Y_k)$ is the history of the event of interest through interval k. Similarly, we use underbars to denote future values of a random variable, such that $\underline{Y}_k = (Y_k, Y_{k+1}, \ldots, Y_{K+1})$. We assume full adherence to the assigned treatment without loss of generality, and until Section 5.4, no loss to follow-up.

3. Decomposition of Treatment Effects

Suppose that treatment A can be conceptualized as having two binary components that act through different causal pathways: one component A_Y that affects the event of interest Y_k and one component A_D that affects the competing event D_k. This hypothetical decomposition of A can be formally described by the following conditions.

Suppose that A and the two components A_Y and A_D are deterministically related in the observed data,

$$A \equiv A_D \equiv A_Y, \tag{1}$$

but we can conceive hypothetical interventions that set A_D and A_Y to different values. For $k \in \{0,\ldots,K\}$, let Y^a_{k+1} be an individual's indicator of the event of interest at time $k+1$ when, possibly contrary to fact, A is set to the value $a \in \{0,1\}$. Similarly, let Y^{a_Y,a_D}_{k+1} be this outcome when, possibly contrary to fact, A_Y is set to a_Y and A_D is set to a_D, where $a_Y, a_D \in \{0,1\}$. We require that setting $A = a$ is equivalent to setting both A_Y and A_D to a, that is,

$$\begin{aligned} Y^{a_Y=a,a_D=a}_{k+1} &= Y^a_{k+1}, \\ D^{a_Y=a,a_D=a}_{k+1} &= D^a_{k+1}, \text{ for } k \in \{0,K\}. \end{aligned} \tag{2}$$

The assumption that A_D only exerts effects on Y_{k+1} through its effect on $\overline{D}_{k+1}$ can be stated as

$$\begin{aligned} Y^{a_Y,a_D=1}_k = D^{a_Y,a_D=1}_{k+1} = Y^{a_Y,a_D=0}_k = D^{a_Y,a_D=0}_{k+1} = 0 \implies \\ Y^{a_Y,a_D=0}_{k+1} = Y^{a_Y,a_D=1}_{k+1}, \quad \text{for } a_Y \in \{0,1\}, \end{aligned} \tag{3}$$

and, similarly, the assumption that A_Y only exerts effects on D_{k+1} through its effect on $\overline{Y}_k$ can be stated as

$$\begin{aligned} Y^{a_Y=1,a_D}_k = D^{a_Y=1,a_D}_k = Y^{a_Y=0,a_D}_k = D^{a_Y=0,a_D}_k = 0 \implies \\ D^{a_Y=1,a_D}_{k+1} = D^{a_Y=0,a_D}_{k+1}, \quad \text{for } a_D \in \{0,1\}. \end{aligned} \tag{4}$$

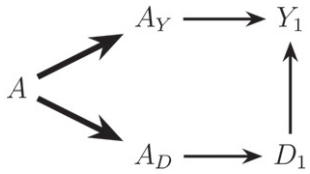

Figure 1. Directed acyclic graph for a trial with a randomized baseline treatment A, such that A_Y and A_D are deterministic functions (bold arrows) of A, competing event D_1 and event of interest Y_1 at time 1 of follow-up, with D_1 measured just before Y_1.

The causal diagram in Figure 1 represents this decomposition in a setting with a single time point. The bold arrows represent the deterministic relation (1). Our decomposition conditions do not preclude the existence of multiple forms of decompositions of A. However, every decomposition of A into two distinct components must be justified by subject-matter knowledge. Let us consider two examples.

3.1. Diethylstilbestrol and Prostate Cancer Mortality

In our prostate cancer example, we assume that A can be decomposed into a component A_Y that directly affects death from prostate cancer and a component A_D that directly affects death from other causes. Suppose that treatment $A = 0$ is placebo and $A = 1$ is diethylstilbestrol (DES), an estrogen which is thought to reduce mortality due to prostate cancer by suppressing testosterone production and to increase cardiovascular mortality through estrogen-induced synthesis of coagulation factors (Turo et al. 2014).

We could then consider a hypothetical treatment that has the same direct effect as DES on prostate cancer mortality, but lacks any effect on mortality from other causes; that is, the same effect as the A_Y component of DES when the A_D component is removed. Real-life treatments similar to such a hypothetical treatment are luteinizing hormone releasing hormone (LHRH) antagonists or orchidectomy (castration), which can stop testosterone production but, unlike estrogen, do not increase cardiovascular risk.

Also, we could consider a hypothetical treatment that has the same direct effect as DES on mortality from other causes, but that lacks any effect on prostate cancer mortality; that is, the same effect as the A_D component of DES when the A_Y component is removed. In practice, a drug that contains not only DES but also testosterone may resemble this hypothetical treatment, as the additional testosterone component can nullify the testosterone suppression that is induced by DES.

3.2. Statins and Dementia

Consider a study to quantify the effect of statins on dementia. Statins reduce cardiovascular mortality by lowering the cholesterol production in the liver. As dementia may develop due to microvascular events in the small cerebral arteries, lowering cholesterol may also reduce the risk of dementia. When studying the effect of statins on dementia, death will be a competing event.

Because statins appear to reduce mortality and dementia through the same mechanism, that is, lowering the cholesterol levels in the blood, decomposing A into the distinct components A_Y and A_D would be difficult. One possibility might be to leverage the distinct localization of the microvessels in the brain: we could bioengineer a cholesterol transporter, which is surgically implanted to shuttle cholesterol particles from the distal cerebral arteries directly to the large cerebral veins, circumventing the cerebral microvessels. That is, if Y_k and D_k denote dementia and death, respectively, then carriers of the transporter will have the A_Y component of statins on dementia, but they will lack the A_D component of statins on mortality. Robins and Richardson (2010, sec. 5.2) discussed the construction of plausible interventions in a mediation context, using nicotine in cigarettes as an example.

3.3. Practical Considerations

Whenever the decomposition of treatment A into A_Y and A_D is possible in principle, regardless of whether it is possible in practice at this time in history, the effects of A_Y and A_D are well-defined. Therefore, in both examples above, we described well-defined effects even though the decomposition of treatment may be practically possible in the prostate cancer example but not in the statin example.

However, caution is required when considering treatment decompositions that, as in the statins example, are possible in principle but not in practice. The problem is that practically impossible decompositions make it hard to evaluate the identifiability conditions for the effects of each component. As described in Section 5, the identification of the separable effects is based on the unverifiable condition that A_Y and A_D are treatment components actually operating in the data (Hernán 2016), such that A_Y has no direct effect on D_k and that A_D has no direct effect on Y_k. When relying on convoluted treatment decompositions, as in our statins example, we may be less confident that these conditions hold in the data. Of course, if these conditions are violated, our effect estimates may differ from those that would be obtained in a future experiment in which both components A_Y and A_D are randomly assigned.

On the other hand, a careful definition of treatment decomposition may help ground scientific conversations even if the decomposition is not yet possible. For example, it is debated whether statins have a protective effect on dementia (Power et al. 2015). To clarify the notion of a "protective effect" it would be helpful to consider a hypothetical trial in which subjects were randomly assigned to the cholesterol transporter or placebo.

4. Definition of Separable Effects

We can now define the *separable direct effects of treatment on the event of interest* as the contrasts

$$\Pr(Y_{k+1}^{a_Y=1,a_D}=1) \text{ versus } \Pr(Y_{k+1}^{a_Y=0,a_D}=1),$$

for $a_D = 1$ or $a_D = 0$; that is, the effect of the component of treatment that affects the event of interest A_Y when the component of treatment that affects the competing event A_D is set at a constant value a_D.

Analogously, we can define the *separable indirect effects of treatment on the event of interest* as the contrasts

$$\Pr(Y_{k+1}^{a_Y,a_D=1}=1) \text{ versus } \Pr(Y_{k+1}^{a_Y,a_D=0}=1),$$

for $a_Y = 1$ and $a_Y = 0$; that is, the effect of the component of treatment that affects the competing event A_D when the component of treatment that affects the event of interest A_Y is set at a constant value a_Y. In other words, the separable indirect effects are functions of the treatment component A_D that affects the competing event D_{k+1}, and the separable indirect effects arise because the competing event makes it impossible for the event of interest to occur.

From (2) we find that the sum of separable direct and indirect effects (on the additive scale) equals the total effect, for example,

$$\begin{aligned}&[\Pr(Y_{k+1}^{a_Y=1,a_D=1}=1)-\Pr(Y_{k+1}^{a_Y=0,a_D=1}=1)]\\&+[\Pr(Y_{k+1}^{a_Y=0,a_D=1}=1)-\Pr(Y_{k+1}^{a_Y=0,a_D=0}=1)]\\&=\Pr(Y_{k+1}^{a=1}=1)-\Pr(Y_{k+1}^{a=0}=1).\end{aligned}$$

To provide intuition about the magnitude of the separable effects, we describe four illustrative scenarios in Appendix A in the supplementary materials.

5. Identification of Separable Effects

The identification of the separable effects requires the identification of the quantities

$$\Pr(Y_{k+1}^{a_Y,a_D}=1), \tag{5}$$

where $a_Y, a_D \in \{0,1\}$. Identifying these quantities would be straightforward if each of the treatment components could be separately intervened upon, that is, if we could conduct a randomized experiment with four possible treatment arms defined by the four combinations of values of A_Y and A_D. However, when using data from a study like that of Section 2, in which only the treatment A is randomized, we only observe two out of the four treatment arms in a hypothetical trial in which A_Y and A_D were randomized. As a result, we need additional untestable conditions to identify (5). This conceptualization of the treatment decomposition in terms of a 4-arm randomized experiment was originally proposed by James Robins during a presentation at the UK Causal Inference Conference in London, April 2016. Since then, Robins and others have often publicly discussed this conceptualization in the context of mediation analysis, which is isomorphic to the context with competing events discussed here.

5.1. Identifiability Conditions

First, we need exchangeability conditional on the measured covariates L,

$$\bar{Y}_{K+1}^a, \bar{D}_{K+1}^a \perp\!\!\!\perp A \mid L \text{ for } a \in \{0,1\},$$

where time $K+1$ is the end of the study. This exchangeability condition is expected to hold when $A \equiv A_Y \equiv A_D$ is randomized.

Second, consistency, such that if $A = a$, then

$$\begin{aligned}Y_{k+1}^a &= Y_{k+1}\\ D_{k+1}^a &= D_{k+1},\end{aligned}$$

for $a \in \{0,1\}$ at all times $k \in \{0,\ldots,K\}$. If any subject has data history consistent with the intervention under a counterfactual scenario, then the consistency assumption ensures that the observed outcome is equal to the counterfactual outcome.

Third, positivity such that

$$\begin{aligned}&\Pr(L=l)>0 \implies\\&\quad \Pr(A=a \mid L=l)>0 \text{ for } a\in\{0,1\}, &(6)\\&\Pr(D_{k+1}=Y_k=0, L=l)>0 \implies\\&\quad \Pr(A=a|D_{k+1}=Y_k=0,L=l)>0 \text{ for } a\in\{0,1\} \text{ and}\\&\quad k\in\{0,\ldots,K\}, &(7)\end{aligned}$$

where (6) is the usual positivity condition under interventions on A and (7) ensures that among those event-free through each follow-up time, there exist individuals with $A = 1$ and individuals with $A = 0$. However, our estimand is based on hypothetical intervention on both A_Y and A_D, and our positivity conditions do not ensure the stricter condition that

$$\begin{aligned}&\Pr(L=l)>0 \implies\\&\quad \Pr(A_Y=a_Y, A_D=a_D \mid L=l)>0 \text{ for } a_Y, a_D\in\{0,1\},\end{aligned}$$

which, indeed, will be violated when $a_Y \neq a_D$ in our setting where $A \equiv A_Y \equiv A_D$.

To allow for identifiability under our positivity condition in (6), we introduce two conditions that are related to conditions described by Didelez in a mediation setting (Didelez 2019).

5.1.1. Dismissible Component Condition 1

$$\begin{aligned}\Delta\mathbf{1}: &\Pr(Y_{k+1}^{a_Y,a_D=1}=1 \mid Y_k^{a_Y,a_D=1}=0, D_{k+1}^{a_Y,a_D=1}=0, L=l)\\&=\Pr(Y_{k+1}^{a_Y,a_D=0}=1 \mid Y_k^{a_Y,a_D=0}=0, D_{k+1}^{a_Y,a_D=0}=0, L=l),\end{aligned}$$

for $a_Y \in \{0,1\}$ at all times $k \in \{0,\ldots,K\}$. That is, the counterfactual (discrete-time) hazards of the event of interest are equal under all values of A_D.

5.1.2. Dismissible Component Condition 2

$$\begin{aligned}\Delta\mathbf{2}: &\Pr(D_{k+1}^{a_Y=1,a_D}=1 \mid Y_k^{a_Y=1,a_D}=0, D_k^{a_Y=1,a_D}=0, L=l)\\&=\Pr(D_{k+1}^{a_Y=0,a_D}=1 \mid Y_k^{a_Y=0,a_D}=0, D_k^{a_Y=0,a_D}=0, L=l),\end{aligned}$$

for $a_D \in \{0,1\}$ at all times $k \in \{0,\ldots,K\}$. That is, the counterfactual (discrete-time) hazard functions of the competing event are equal under all values of A_Y. The dismissible component conditions are analogous to identification conditions for path specific effects (Shpitser 2013).

By considering a hypothetical trial in which both A_Y and A_D are randomized, we can define conditional independencies that imply the dismissible component conditions, and these conditional independencies can be read off of causal DAGs directly, see Appendix B in the supplementary materials for details.

The dismissible component conditions ensure that we can adjust for common causes of D_k and $Y_{k'}$ for all $k, k' \in \{1,\ldots,K+1\}$. In particular, an unmeasured common cause

of D_1 and Y_1, such as U_{YD} in Figure 2, violates $\Delta 1$ and $\Delta 2$. In our prostate cancer example, suppose that smoking is a common cause of death from prostate cancer (Y_k) and death from other causes (D_k). Then, if smoking is an unmeasured variable (such as U_{YD} in Figure 2), the dismissible component conditions will be violated.

However, the presence of unmeasured causes U_Y of Y_k and unmeasured causes U_D of D_k, as shown in Figure 3, does not violate $\Delta 1$ and $\Delta 2$ (see Appendix E in the supplementary materials for details); it just implies that contrasts of the hazard terms in (8) cannot be causally interpreted (Hernán 2010; Stensrud et al. 2017; Young et al. 2020), which is analogous to the mediation setting in Didelez (2019, Figure 6). For this reason, we have defined our causal estimands as contrasts of risks rather than as contrasts of hazards. Furthermore, adjusting for a measured common cause of Y_k and D_k, such as L in Figure 4, allows identification under $\Delta 1$ and $\Delta 2$. In subsequent figures, we have omitted the variables U_Y and U_D to avoid clutter, but our results are valid in the presence of U_Y and U_D. We have also omitted an arrow from L to A, but this arrow would not invalidate our results. Furthermore, we have intentionally omitted arrows from D_k to Y_s for $k < s$, as these arrows are redundant in our setting where the competing event is a terminating event that precludes the event of interest at all subsequent times. Finally, note that if the dismissible component conditions hold on a coarser scale (say, daily), then they will in general also hold on a finer scale (say, hourly), but the reverse is not true. This is analogous to any setting where measurements of time-varying covariates are needed to identify causal effects.

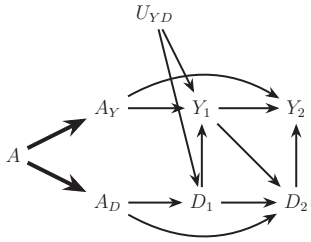

Figure 2. Extension of the causal directed acyclic graph in Figure 1 which includes an unmeasured common cause U_{YD} which violates conditions $\Delta 1$ and $\Delta 2$.

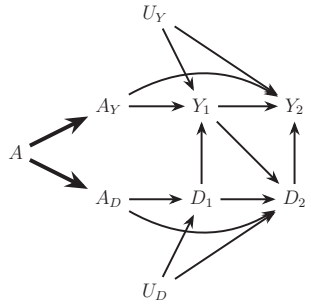

Figure 3. Extension of the directed acyclic graph in Figure 1 which includes unmeasured common causes U_Y and U_D, which are expected to exist but do not violate conditions $\Delta 1$ and $\Delta 2$.

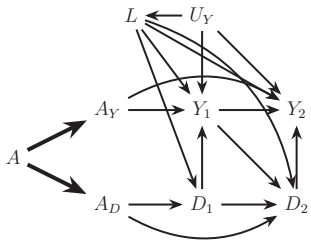

Figure 4. Conditions $\Delta 1$ and $\Delta 2$ can hold if the common cause L of Y_k and D_k, $k \in \{1, 2\}$, is measured.

The dismissible component conditions are not empirically verifiable in a trial in which the entire treatment A, but neither of its components A_Y and A_D, is intervened upon. However, both conditions could be tested in a trial in which A_Y and A_D were randomly assigned.

5.2. Identification Formula

Under the identifiability conditions in Section 5.1, we identify $\Pr(Y_{k+1}^{a_Y,a_D} = 1)$ from the following g-functional (Robins 1986) of the observed data described in Section 2,

$$\sum_l \Big[\sum_{s=0}^{k} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = 0, A = a_Y, L = l) \prod_{j=0}^{s} \big[\Pr(D_{j+1} = 0 \mid D_j = Y_j = 0, A = a_D, L = l) \times \Pr(Y_j = 0 \mid D_j = Y_{j-1} = 0, A = a_Y, L = l) \big] \Big] \Pr(L = l), \tag{8}$$

see Appendix C in the supplementary materials for proof.

5.3. Intuition on the Identification Formula (8) and Falsifiability of the Separable Effects

Identification formula (8) can be intuitively motivated as follows: consider an experiment G in which both A_Y and A_D are randomly assigned such that $\Pr(A_Y = a_Y, A_D = a_D) > 0$ for all $a_D, a_Y \in \{0, 1\}$. In the experiment G, $\Pr(Y_{k+1}^{a_Y,a_D} = 0) = \Pr(Y_{k+1} = 1 \mid A_Y = a_Y, A_D = a_D)$ by randomization. By the laws of probability $\Pr(Y_{k+1} = 1 \mid A_Y = a_Y, A_D = a_D)$ can in turn be re-expressed as

$$\sum_l \Big[\sum_{s=0}^{k} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = 0, A_Y = a_Y, A_D = a_D, L = l) \prod_{j=0}^{s} \big[\Pr(D_{j+1} = 0 \mid D_j = Y_j = 0, A_Y = a_Y, A_D = a_D, L = l) \times \Pr(Y_j = 0 \mid D_j = Y_{j-1} = 0, A_Y = a_Y, A_D = a_D, L = l) \big] \Big] \times \Pr(L = l). \tag{9}$$

Formula (8) can be obtained by applying the dismissible component conditions to the terms in (9). These additional conditions are needed for identification in our current study because, unlike in G, only A was randomized in our current study and not the separate components A_Y and A_D. If the experiment G is

actually conducted in the future, then the separable effect estimates obtained from (8) in our current study can be confirmed by comparing them to estimates of $\Pr(Y_{k+1} = 0 \mid A_Y = a_Y, A_D = a_D)$ from G (Robins and Richardson 2010).

Note that (8) can also be read off of a single world intervention graph (SWIG) (Richardson and Robins 2013) that satisfies the dismissible component conditions, as suggested in Figure 5, illustrating that the separable effects are single-world quantities that are empirically testable in principle. This is in contrast to alternative approaches from mediation analysis that require additional, untestable cross-world independence assumptions (Robins and Richardson 2010).

5.4. Separable Effects in the Presence of Censoring

We consider a subject to be censored at time $k+1$ if the subject remained under follow-up and was event-free until k, but we have no information about the subject's events at $k+1$ or later (Young et al. 2020). That is, censoring is a type of event that does not make it impossible for the event of interest to occur and we assume that censoring can in principle be prevented (Young et al. 2020). When the censoring is independent of future counterfactual events given L, as illustrated in Figure 6, we can identify the separable effects from

$$\begin{aligned}
&\sum_l \Big[\sum_{s=0}^{k} \Pr(Y_{s+1} = 1 \mid D_{s+1} = Y_s = \bar{C}_{s+1} = 0, \\
&\quad A = a_Y, L = l) \\
&\prod_{j=0}^{s} \big[\Pr(D_{j+1} = 0 \mid D_j = Y_j = \bar{C}_{j+1} = 0, A = a_D, L = l) \\
&\times \Pr(Y_j = 0 \mid D_j = Y_{j-1} = \bar{C}_j = 0, A = a_Y, L = l) \big] \Big] \\
&\times \Pr(L = l), \qquad (10)
\end{aligned}$$

where C_k is an indicator of being censored at k, see Appendix C in the supplementary materials for details. Alternatively, the identification formula can be derived by drawing a SWIG for the scenario of interest, as suggested in Figure 7. Hereafter we will use $\nu_{a_Y,a_D,k}$ to denote the g-formula (8).

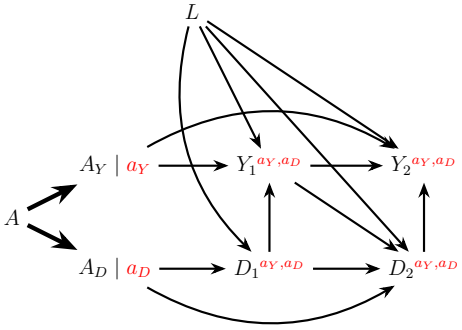

Figure 5. Single world intervention graph (SWIG) that describes a scenario with interventions on A_Y and A_D.The unmeasured common causes U_Y and U_D have been omitted to avoid clutter.

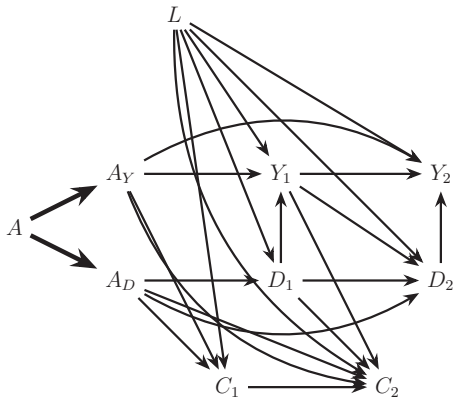

Figure 6. Directed acyclic graph with loss to follow-up (C_k). We have omitted arrows from C_k into Y_k and from C_k into D_k for $k \in \{1, 2\}$.

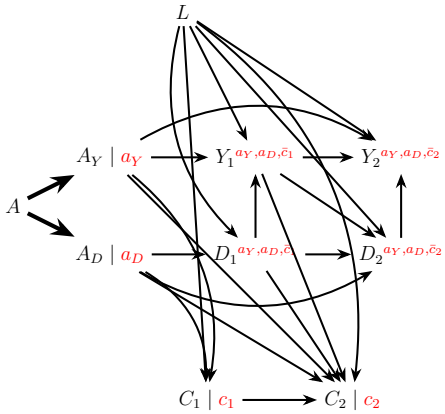

Figure 7. Single world intervention graph (SWIG) that describes a scenario with interventions on A_Y, A_D, and $\bar{C}_2$. We have omitted arrows from C_k into Y_k and from C_k into D_k for $k \in \{1, 2\}$.

5.5. Alternative Representations of the Identification Formula

The g-formula (10) can also be expressed as

$$\begin{aligned}
&\sum_{s=0}^{k} \mathbb{E}[W_{C,s}(a_Y) W_{D,s}(a_Y, a_D)(1 - Y_s) \\
&\quad \times (1 - D_{s+1}) Y_{s+1} \mid A = a_Y], \qquad (11)
\end{aligned}$$

where

$$\begin{aligned}
&W_{D,s}(a_Y, a_D) \\
&= \frac{\prod_{j=0}^{s} \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L, A = a_D)}{\prod_{j=0}^{s} \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L, A = a_Y)}, \\
&W_{C,s}(a_D) \\
&= \frac{I(\bar{C}_{s+1} = 0)}{\prod_{j=0}^{s} \Pr(\bar{C}_{j+1} = 0 \mid \bar{C}_j = D_j = Y_j = 0, L, A = a_D)},
\end{aligned}$$

see Appendix D in the supplementary materials for details. Furthermore, another representation of (10) is

$$\begin{aligned}
&\sum_{s=0}^{k} \mathbb{E}\{W_{C,s}(a_D) W_{Y,s}(a_D, a_Y)(1 - Y_s) \\
&\quad \times (1 - D_{s+1}) Y_{s+1} \mid A = a_D\}, \qquad (12)
\end{aligned}$$

where $W_{C,s}(a_D)$ is defined as in (11) and

$$W_{Y,s}(a_D, a_Y) = \frac{\Pr(Y_{s+1} = 1 \mid \bar{C}_{s+1} = D_{s+1} = Y_s = 0, L, A = a_Y)}{\Pr(Y_{s+1} = 1 \mid \bar{C}_{s+1} = D_{s+1} = Y_s = 0, L, A = a_D)} \times \frac{\prod_{j=0}^{s-1} \Pr(Y_{j+1} = 0 \mid \bar{C}_{j+1} = D_{j+1} = Y_j = 0, L, A = a_Y)}{\prod_{j=0}^{s-1} \Pr(Y_{j+1} = 0 \mid \bar{C}_{j+1} = D_{j+1} = Y_j = 0, L, A = a_D)},$$

as formally shown in Appendix D in the supplementary materials. Note that in settings without censoring, $W_{C,s}(a) \equiv 1, a \in \{0, 1\}$. Representations (11) and (12) motivate inverse probability (IP) weighted estimators of the separable effects, as described in Section 6.

6. Estimation of Separable Effects

To estimate the separable effects, we emphasize that (8) and (10) are functionals of (discrete-time) hazard functions and the density of L. Indeed, $\Pr(Y_{k+1} = 1 \mid D_{k+1} = Y_k = \bar{C}_{k+1} = 0, A = a, L = l)$ and $\Pr(D_{k+1} = 0 \mid D_k = Y_k = \bar{C}_{k+1} = 0, A = a, L = l)$ are often denoted "cause specific hazard functions" in the statistical literature. Though the term "cause specific" is confusing because the causal interpretation of these hazard functions is ambiguous (Young et al. 2020), we can nevertheless estimate these functions using classical statistical models. Provided that these hazard models are correctly specified, along with $\widehat{\Pr}(L = l)$ (Young et al. 2011), we can consistently estimate (10) using a parametric g-formula estimator (Robins 1986). However, we can also derive weighted estimators that rely on fewer model assumptions.

6.1. Inverse Probability Weighted Estimators

Motivated by the alternative g-formula representation (11), define

$$\hat{W}_{D,k,i}(a_Y, a_D; \hat{\eta}_D) = \frac{\prod_{j=0}^{k} \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L_i, A = a_D; \hat{\eta}_D)}{\prod_{j=0}^{k} \Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L_i, A = a_Y; \hat{\eta}_D)},$$

$$\hat{W}_{C,k,i}(a_D; \hat{\eta}_C) = \frac{I(\bar{C}_{k+1} = 0)}{\prod_{j=0}^{k} \Pr(\bar{C}_{j+1} = 0 \mid \bar{C}_j = D_j = Y_j = 0, L_i, A = a_D; \hat{\eta}_C)},$$

where $\Pr(D_{j+1} = 0 \mid \bar{C}_{j+1} = D_j = Y_j = 0, L, A = a_D; \eta_D)$ is a parametric model for the numerator (and denominator) of $W_{D,k}(a_Y, a_D)$ indexed by parameter η_D, and $\hat{\eta}_D$ is a consistent estimator of η_D (e.g., the MLE), and the terms in $\hat{W}_{C,k,i}(a_D; \hat{\eta}_C)$ are defined similarly, where $\hat{\eta}_C$ is a consistent estimator of η_C.

Let $\eta_1 = (\eta_D, \eta_C)$, and define the estimator $\hat{\nu}_{1,a_Y,a_D,k}$ of $\nu_{a_Y,a_D,k}$ as the solution to the estimating equation $\sum_{i=1}^{n} U_{1,k,i}(\nu_{a_Y,a_D,k}, \hat{\eta}_1) = 0$ with respect to $\nu_{a_Y,a_D,k}$ with

$$U_{1,k,i}(\nu_{a_Y,a_D,k}, \hat{\eta}_1) = I(A_i = a_Y)\Big[\sum_{s=0}^{k}\{\hat{W}_{1,s,i}(a_Y, a_D; \hat{\eta}_1) Y_{s+1,i}(1 - Y_{s,i}) \times (1 - D_{s+1,i})\} - \nu_{a_Y,a_D,k}\Big],$$

and $\hat{W}_{1,s,i}(a_Y, a_D; \hat{\eta}_1) = \hat{W}_{D,s,i}(a_Y, a_D; \hat{\eta}_D)\hat{W}_{C,s,i}(a_Y; \hat{\eta}_C)$.

Then, $\hat{\nu}_{1,a_Y,a_D,k}$ is a consistent estimator for $\nu_{a_Y,a_D,k}$ if the models indexed by elements in η_1 are correctly specified and $\hat{\eta}_1$ is a consistent estimator for η_1, which follows because (10) and (11) are equal. For example, we can use conventional statistical models for binary outcomes, such as pooled logistic regression models, to estimate the weights $W_{D,k}(a_Y, a_D)$ and $W_{C,k}(a_Y)$.

Analogous to $\hat{\nu}_{1,a_Y,a_D,k}$, we can derive an estimator based on (12). Suppose

$$\hat{W}_{Y,k,i}(a_D, a_Y; \hat{\eta}_Y) = \frac{\Pr(Y_{k+1} = 1 \mid \bar{C}_{k+1} = D_{k+1} = Y_k = 0, L_i, A = a_Y; \hat{\eta}_Y)}{\Pr(Y_{k+1} = 1 \mid \bar{C}_{j+1} = D_{k+1} = Y_k = 0, L_i, A = a_D; \hat{\eta}_Y)} \times \frac{\prod_{j=0}^{k-1} \Pr(Y_{j+1} = 0 \mid \bar{C}_{j+1} = D_{j+1} = Y_j = 0, L_i, A = a_Y; \hat{\eta}_Y)}{\prod_{j=0}^{k-1} \Pr(Y_{j+1} = 0 \mid \bar{C}_{j+1} = D_{j+1} = Y_j = 0, L_i, A = a_D; \hat{\eta}_Y)},$$

where the terms in $\hat{W}_{Y,k,i}(a_D, a_Y; \hat{\eta}_Y)$ are statistical models for binary outcomes and $\hat{\eta}_Y$ is a consistent estimator for η_Y.

Let $\eta_2 = (\eta_Y, \eta_C)$, and define the estimator $\hat{\nu}_{2,a_Y,a_D,k}$ of $\nu_{a_Y,a_D,k}$ as the solution to the estimating equation $\sum_{i=1}^{n} U_{2,k,i}(\nu_{a_Y,a_D,k}, \hat{\eta}_2) = 0$ with respect to $\nu_{a_Y,a_D,k}$, where

$$U_{2,k,i}(\nu_{a_Y,a_D,k}, \hat{\eta}_2) = I(A_i = a_D)\Big[\sum_{s=0}^{k}\{\hat{W}_{2,s,i}(a_Y, a_D; \hat{\eta}_2) Y_{s+1,i}(1 - Y_{s,i}) \times (1 - D_{s+1,i})\} - \nu_{a_Y,a_D,k}\Big],$$

and $\hat{W}_{2,s,i}(a_Y, a_D; \hat{\eta}_2) = \hat{W}_{C,s,i}(a_D; \hat{\eta}_C)\hat{W}_{Y,s,i}(a_D, a_Y; \hat{\eta}_Y)$. Analogous to the estimator based on (11), provided that the models indexed by elements in η_2 are correctly specified and $\hat{\eta}_2$ is a consistent estimator for η_2, then consistency of $\hat{\nu}_{2,a_Y,a_D,k}$ for $\nu_{a_Y,a_D,k}$ follows because (10) and (12) are equal.

In the next section, we use this approach to analyze a randomized trial on prostate cancer therapy. In Appendix F in the supplementary materials, we present simulations, suggesting that the estimators perform satisfactorily in finite samples. The simulations also illustrate that the separable effect can be substantially different than the total effect, and that the estimators may be biased if the dismissible component conditions are violated.

6.2. Example: A Randomized Trial of Prostate Cancer

Consider, as described in Section 3.1, a hypothetical drug that has the same direct effect as DES on prostate cancer mortality (same A_Y component), but lacks any effect on mortality due to other causes (opposite A_D component). Then we can define separable direct effects of treatment DES on prostate

Table 1. Estimates of cumulative incidence after 3 years of follow-up.

Estimand	G-formula estimate (95%CI)	IP weighted estimate (95%CI)
$\Pr(Y_{36}^{a=1} = 1)$	0.14 (0.08–0.20)	0.17 (0.10, 0.24)
$\Pr(Y_{36}^{a_Y=1,a_D=0} = 1)$	0.15 (0.09–0.21)	0.18 (0.10, 0.26)
$\Pr(Y_{36}^{a=0} = 1)$	0.21 (0.15–0.28)	0.23 (0.17, 0.35)

cancer mortality Y_k in the presence of competing mortality D_k from other causes. We estimated these separable effects using a parametric g-formula estimator and, for simplicity, one of the IP weighted estimators ($\hat{\nu}_{1,a_Y,a_D,k}$). We used publicly available data from a randomized trial (*http://biostat.mc.vanderbilt.edu/DataSets*) (Byar and Green 1980) that has been used in several methodological articles on competing events (Harrell, Lee, and Mark 1996; Kay 1986; Fine 1999; Varadhan et al. 2010). In total, 502 patients were assigned to four different treatment arms. We restrict our analysis to the placebo arm (127 patients) and the high-dose DES arm (125 patients).

To implement the parametric g-formula estimator, we used pooled logistic regression models to estimate the terms in (10), in which daily activity function, age group, hemoglobin level and previous cardiovascular disease were included as covariates (L in Figure 6), that is,

$$\begin{aligned}&\text{logit}[\Pr(Y_k = 1 \mid D_k = Y_{k-1} = \bar{C}_k = 0, A, L)]\\&\quad = \theta_{0,k} + \theta_1 A + \theta_2 Ak + \theta_3 Ak^2 + \theta_4' L, \qquad (13)\\&\text{logit}[\Pr(D_k = 1 \mid D_{k-1} = Y_{k-1} = \bar{C}_k = 0, A, L)]\\&\quad = \beta_{0,k} + \beta_1 A + \beta_2 Ak + \beta_3 Ak^2 + \beta_4' L, \qquad (14)\end{aligned}$$

where $\theta_{0,k}$ and $\beta_{0,k}$ are time-varying intercepts modeled as cubic polynomials. To allow time-varying treatment effects, we included $\theta_2, \theta_3, \beta_2$ and β_3.

To implement the IP weighted estimator $\hat{\nu}_{1,a_Y,a_D,k}$, we only require the model (14) (similarly, we would only require the model (13) to implement $\hat{\nu}_{2,a_Y,a_D,k}$).

Both the parametric g-formula and IP weighted estimator gave cumulative incidence estimates under the hypothetical drug that were similar, but not identical, to those under DES treatment. Table 1 displays estimates of the 3-year cumulative incidence and 95% bootstrap confidence intervals based on both estimators and Figure 8(B) shows cumulative incidence curves from the IP weighted estimator (R code is provided in the supplementary material).

Our analysis suggests that DES mostly reduces prostate cancer mortality via testosterone suppression because the estimate of the separable indirect effect on 3-year mortality is close to zero. Using either the parametric g-formula or the IP weighted estimator, the estimate of the additive indirect effect after 3 years of follow-up is 0.01 ($0.15 - 0.14 = 0.01$ and $0.18 - 0.17 = 0.01$), which can be interpreted as the reduction in prostate cancer mortality under DES compared with placebo that is due to the DES effect on mortality from other causes. That is, the total effect of DES on prostate cancer mortality is not simply a consequence of a harmful effect on death from other causes.

The validity of our estimates relies on the assumption that L is sufficient to adjust for the common causes of Y_k and D_k. This assumption would be violated if other factors, such as unmeasured comorbidities, exert effects on both Y_k and D_k. Also, our approach relies on the absence of time-varying common causes of the event of interest and the competing event. In future work, we will generalize our approach to allow for time-varying covariates.

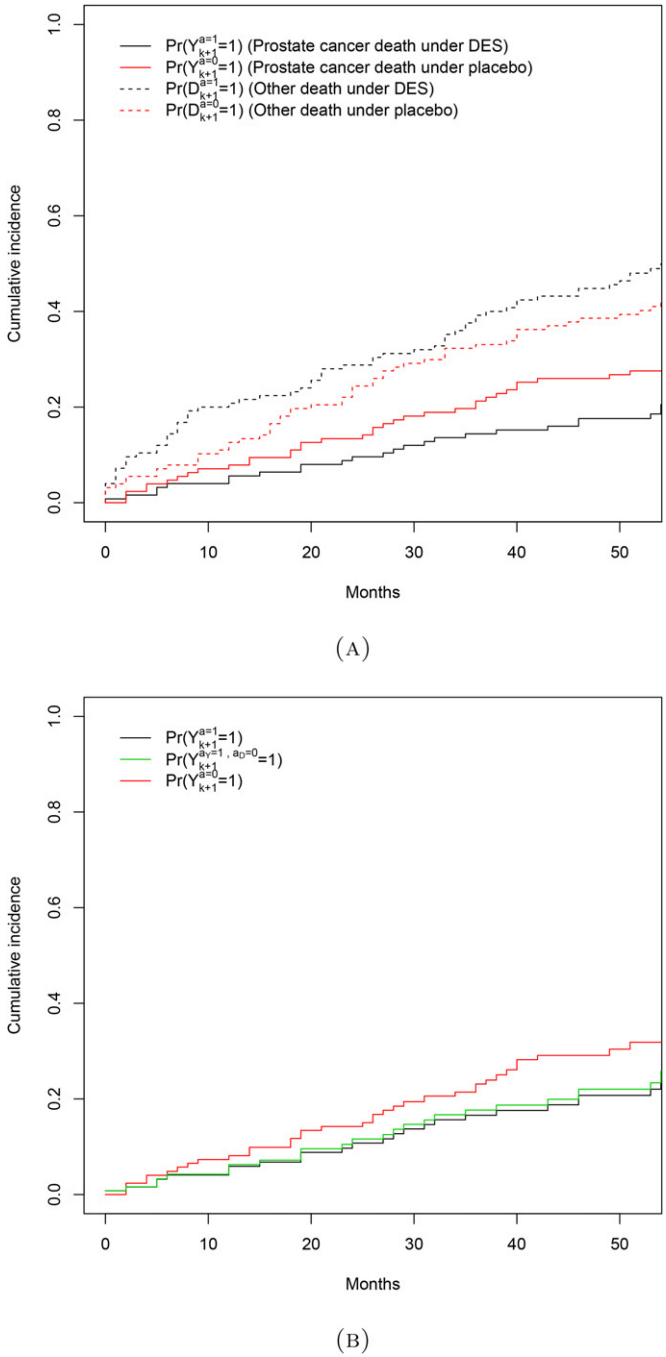

Figure 8. Estimated cumulative incidence of (A) death from prostate cancer and death from other causes under DES (black) and placebo (red) using the Aalen–Johansen estimator, and (B) death from prostate cancer under DES (black), placebo (red), and the hypothetical treatment where the effect on death of other causes is removed (green) using the inverse probability weighted estimator $\hat{\nu}_{1,a_Y,a_D,k}$.

7. Discussion

We have defined separable effects as new estimands to promote causal reasoning in competing event settings. The separable effects are motivated by hypothetical interventions, in which a

time-fixed treatment is decomposed into distinct components, and each component can be assigned different values.

Therefore, to define and interpret the separable effects, investigators must use their subject-matter knowledge to explicitly articulate a hypothetical decomposition of the treatment. An explicit consideration of this decomposition helps assess the plausibility of the assumptions and guides the design of future experiments to empirically verify the effects (Robins and Richardson 2010).

Classical statistical estimands fail to provide the same information as the separable effects (see Young et al. (2020) for a detailed discussion of interpretation and identification of counterfactual contrasts in classical estimands for competing event settings). In particular, the cumulative incidence functions of the event of interest and the competing event do not clarify the mechanism by which treatment exerts effects on the event of interest, even if these outcomes are considered jointly in an ideal randomized trial. Furthermore, estimands on the hazard scale, for example, subdistribution hazards and cause-specific hazards, do not have a straightforward causal interpretation and thus cannot solve the problem (Hernán 2010; Young et al. 2020).

Identification of separable effects requires, even in a perfectly executed randomized trial, adjustment for pretreatment variables that are common causes of the event of interest and the competing event. However, this strong condition is also needed for the causal interpretation of analysis of trials targeting conventional estimands such as controlled direct effects or counterfactual contrasts of hazard functions (Young et al. 2020).

For simplicity, we have considered settings in which the treatment A is randomly assigned. For example, we illustrated the application of standard time-to-event methods to estimate the separable effects in a prostate cancer randomized trial. However, our approach can be easily extended to analyses of observational studies under the additional assumption of no unmeasured confounding for the effect of treatment on both the competing event and the event of interest.

Finally, the idea of separable effects is not only relevant to settings in which the outcome of interest is a time-to-event. Many practical settings involve intermediate outcomes that are ill-defined after the occurrence of a terminating event. For example, we may be interested in treatment effects on outcomes such as quality of life or cognitive function, and these outcomes are meaningless after death. We aim to study separable effects in such settings in future research.

Supplementary Materials

The online supplementary materials contain the proofs of the identification formulas and simulation studies. Furthermore, we describe conditional independencies that imply the dismissible component conditions. Computer code in R is provided as separate files.

Funding

This work was funded by NIH grant R37 AI102634. M.J.S. was also supported by an ASISA Fellowship and the Research Council of Norway, grant NFR239956/F20.

References

Andersen, P. K., Geskus, R. B., de Witte, T., and Putter, H. (2012), "Competing Risks in Epidemiology: Possibilities and Pitfalls," *International Journal of Epidemiology*, 41, 861–870. [175]

Byar, D., and Green, S. (1980), "The Choice of Treatment for Cancer Patients Based on Covariate Information," *Bulletin du Cancer*, 67, 477–490. [182]

Didelez, V. (2019), "Defining Causal Mediation With a Longitudinal Mediator and a Survival Outcome," *Lifetime Data Analysis*, 25, 593–610. [176,178,179]

Fine, J. (1999), "Analysing Competing Risks Data With Transformation Models," *Journal of the Royal Statistical Society*, Series B, 61, 817–830. [182]

Frangakis, C. E., and Rubin, D. B. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29. [176]

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996), "Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors," *Statistics in Medicine*, 15, 361–387. [182]

Hernán, M. A. (2010), "The Hazards of Hazard Ratios," *Epidemiology*, 21, 13. [175,179,183]

——— (2016), "Does Water Kill? A Call for Less Casual Causal Inferences," *Annals of Epidemiology*, 26, 674–680. [176,177]

Kay, R. (1986), "Treatment Effects in Competing-Risks Analysis of Prostate Cancer Data," *Biometrics*, 42, 203–211. [182]

Power, M. C., Weuve, J., Sharrett, A. R., Blacker, D., and Gottesman, R. F. (2015), "Statins, Cognition, and Dementia—Systematic Review and Methodological Commentary," *Nature Reviews Neurology*, 11, 220. [177]

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Jr., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, 541–554. [175]

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 128. [180]

Robins, J. M. (1986), "A New Approach to Causal Inference in Mortality Studies With a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect," *Mathematical Modelling*, 7, 1393–1512. [176,179,181]

Robins, J. M., and Greenland, S. (1992), "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155. [176]

Robins, J. M., and Richardson, T. S. (2010), "Alternative Graphical Causal Models and the Identification of Direct Effects," in *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, eds. P. E. Shrout, K. M. Keyes, and K. Ornstein (Oxford: Oxford University Press), pp. 103–158. [176,177,180,183]

Shpitser, I. (2013), "Counterfactual Graphical Models for Longitudinal Mediation Analysis With Unobserved Confounding," *Cognitive Science*, 37, 1011–1035. [178]

Stensrud, M. J., Valberg, M., Røysland, K., and Aalen, O. O. (2017), "Exploring Selection Bias by Causal Frailty Models: The Magnitude Matters," *Epidemiology*, 28, 379–386. [179]

Turo, R., Smolski, M., Esler, R., Kujawa, M. L., Bromage, S. J., Oakley, N., Adeyoju, A., Brown, S. C., Brough, R., Sinclair, A., and Collins, G. N. (2014), "Diethylstilboestrol for the Treatment of Prostate Cancer: Past, Present and Future," *Scandinavian Journal of Urology*, 48, 4–14. [177]

Varadhan, R., Weiss, C. O., Segal, J. B., Wu, A. W., Scharfstein, D., and Boyd, C. (2010), "Evaluating Health Outcomes in the Presence of Competing Risks: A Review of Statistical Methods and Clinical Applications," *Medical Care*, 48, S96–S105. [182]

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*, 3, 119. [181]

Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., and Hernán, M. A. (2020), "A Causal Framework for Classical Statistical Estimands in Failure-Time Settings With Competing Events," *Statistics in Medicine*, 39, 1199–1236. [175,176,179,180,181,183]