Manuscript type: Original Research article (Methodology paper)

Effect modification in settings with "truncation by death"

Authors: Bronner P. Gonçalves[1], Etsuji Suzuki[2, 3]

Affiliations:

[1] Department of Comparative Biomedical Sciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom

[2] Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan

[3] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

Correspondence:

Bronner P. Gonçalves, Department of Comparative Biomedical Sciences, University of Surrey, Daphne Jackson Rd, Guildford, United Kingdom (b.goncalves@surrey.ac.uk)

Etsuji Suzuki, Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan (etsuji-s@cc.okayama-u.ac.jp)

Running head: Effect modification and truncation by death

Conflict of interest statement: The authors declare no competing interests.

Source of Funding: BG is funded by a Senior Fellowship from the University of Surrey. ES is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP19KK0418 and JP23K09740). The funding sources had no involvement in study design; the collection, analysis, and interpretation of data; the writing of the article; and the decision to submit the article for publication.

Data and computing code availability: Not applicable.

Abstract

Epidemiologic studies recruiting individuals with higher-than-population average mortality can be affected by "truncation by death", whereby the outcome of interest (e.g. quality of life) is considered not to be defined for individuals who die before the end of follow-up. Here, we use the potential outcomes framework and principal stratification to derive conditions under which the survivor average causal effect, an estimand defined for the "always-survivors" stratum, is modified by a variable that represents a possible common cause of survival and the outcome of interest, and by a variable that only affects survival. Further, we show that this principal effect can be expressed as a weighted average of this treatment effect for individuals with each level of these variables, and that these weights depend not only on the relative frequencies of the levels in the total population, but also in the "always-survivors" principal stratum. The implications of this work for the transportability of the survivor average causal effect are also discussed.

Keywords: effect modification, principal stratification, causal inference, transportability

Manuscript

The scientific problem of "truncation by death", formally described nearly two decades ago [1], is present in studies where, for individuals who die before assessment, the outcome of interest is undefined. In this context, a frequently discussed example is one of a randomized trial where individuals are assigned to either active or control treatments and where outcome, e.g. quality of life, is assessed sometime after assignment. For individuals who die before the outcome is measured, quality of life is undefined and the outcome is then said to be truncated. Examples of epidemiologic studies where "truncation by death" is present include those with non-negligible mortality, e.g. studies in elderly population [2] and of cancer patients [3].

A naïve analysis of this type of data involves a comparison of the outcome by treatment arm conditional on observed survival. However, as pointed out by several authors [1, 4], since actual survivors in the exposed and unexposed groups represent non-fully overlapping sets of individuals, this comparison is not fair, and does not have a causal interpretation. Relatedly, in settings where "truncation by death" is present, it is often assumed that there are common, unmeasured, causes of the survival outcome and the outcome of interest [5], which explains why "selection bias" is often mentioned in discussing this type of analysis, that conditions on observed survival status.

An alternative analytic approach, with a causal interpretation, is based on the principal stratification framework [6] and involves estimation of effects in the population stratum of individuals who would survive under both treatment levels. Our aim, here, was to investigate, using this framework, modification of treatment effects on the survival outcome and on the outcome of interest. We consider two scenarios: first, we discuss effect modification by a variable that represents a common cause of survival and the outcome of interest, and in the second scenario, we study effect modification by a variable that only affects survival.

Notation and definition of the survivor average causal effect

Let A denote treatment (1 = active treatment, 0 = control treatment), which is assumed to be randomized. S denotes the survival outcome (1 = survival, 0 = death), and Y represents the outcome of interest, here a dichotomized version of a quality of life score, with $Y = 1$

corresponding to better quality of life, relative to $Y = 0$. Potential outcome variables for survival and quality of life are, respectively, S^a and Y^a, and correspond to the values that S and Y would take if treatment were set to level a. As implied above, if $S = 0$, Y is undefined. Throughout, we assume no loss to follow-up.

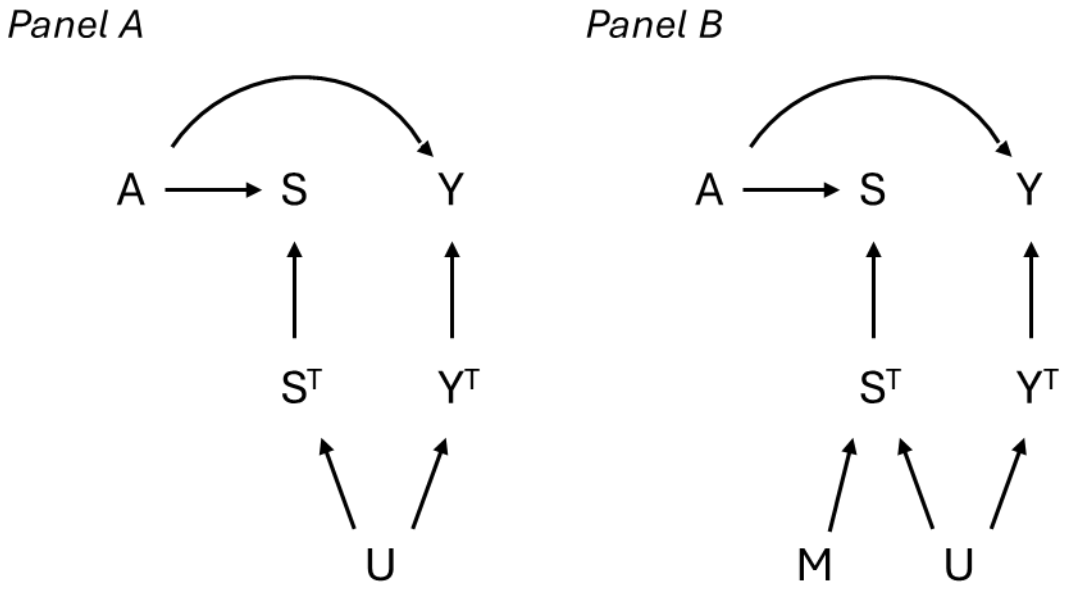
The study population can be partitioned into four principal strata of S, defined by the joint potential survival values: the "always-survivors" stratum ($S^1 = 1, S^0 = 1$), the "protectable" stratum ($S^1 = 1, S^0 = 0$), the "defiers" stratum ($S^1 = 0, S^0 = 1$), and the "never-survivors" stratum ($S^1 = 0, S^0 = 0$). When "truncation by death" is present, for the outcome of interest, one of the estimands with a valid causal interpretation corresponds to a comparison of the potential quality of life outcomes under the two treatment levels for the group of "always-survivors", $E[Y^1 - Y^0 | S^1 = 1, S^0 = 1]$, which is often referred to as the survivor average causal effect (SACE) (see Stensrud et al.[7] for another estimand with causal interpretation that could be studied in this setting; we discuss this estimand in the *Discussion* section). In contrast, causal estimands are not definable for the other principal strata, as the quality of life outcome would be undefined for at least one of the two levels of treatment.

Note that although, for simplicity, we assume a binary outcome, the fact that the outcome is undefined for individuals with the truncating event, that is, individuals who die before assessment, is more evident for quantitative variables that represent characteristics of participants at a particular follow-up time or age. Indeed, as indicated in the Appendix D in Young et al.[8], for binary outcomes that can be interpreted as indicators of failure from an event of interest by a particular time, death would be a competing event, and its occurrence before the event of interest may be indicated as $Y = 0$.

Treatment effect modification by a common cause of survival and quality of life

In this section, we consider the possible effect modification by a variable U that represents a common cause of survival and quality of life; variables with a structural role similar to that of U are often assumed to be present in "truncation by death" settings. This is illustrated in **Figure 1** (Panel A), that corresponds to an extended directed acyclic graph [5, 9], with response type variables S^{T} and Y^{T}; these variables are defined based on joint potential outcome variables: $S^{\mathrm{T}} = (S^1, S^0)$ and $Y^{\mathrm{T}} = (Y^1, Y^0)$.

Figure 1. Causal diagrams representing the settings considered here. In both panels, in addition to observable variables (A, treatment; S, survival; Y, quality of life; U, unspecified common cause of S and Y), the diagram also includes response type variables, S^{T} and Y^{T}. Panel A is used for the discussion in the section *Treatment effect modification by a common cause of survival and quality of life*, and Panel B illustrates the situation presented in the section *Effect modification by another cause of the survival outcome*; in the latter panel, the variable M corresponds to a measured effect modifier. As shown by VanderWeele and Robins [10], effect modifiers do not need to be direct causes of the outcome; the authors also described indirect effect modification, effect modification by proxy, and effect modification by common cause. We focus on the modification of the effect of A on S and of the SACE by U (Panel A), on the modification of A on S by M, and, after conditioning on S^{T} (by restricting the analysis to "always-survivors"), on the modification of the SACE by M (Panel B).



In potential outcomes notation, there is effect modification in the risk difference measure across strata of U for the effect of treatment A on survival S if the following holds [11]:

$$E[S^1 - S^0|U = 0] \neq E[S^1 - S^0|U = 1]$$

where, for simplicity, the variable U is assumed to have two levels, 0 and 1.

Since the response type variable S^{T}as well as the variable U are pre-treatment variables, a different way of conceiving of this is to assume potentially different proportions of survival outcome-based principal strata in groups defined by levels of U. From the law of total probability, the following equation holds:

$$\begin{aligned} E[S^1 - S^0] &= \sum\nolimits_u E[S^1 - S^0|U = u]\Pr(U = u) \\ &= E[S^1 - S^0|U = 0]w + E[S^1 - S^0|U = 1](1 - w) \end{aligned}$$

where $w = \Pr(U = 0)$.

After discussing effect modification for the survival outcome, we now consider the SACE, which is restricted to the "always-survivors", i.e., individuals with $(S^1 = 1, S^0 = 1)$. As the variable U has an effect on quality of life (**Figure 1**, Panel A), it might also represent an effect modifier of the SACE. In particular, there is effect modification in the risk difference measure across strata of U for the effect of treatment *A* on the quality of life *Y* if the following inequality holds:

$$E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), U = 0] \neq E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), U = 1],$$

where $E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), U = u]$ is the SACE for individuals with $U = u$ (SACE_u).

The SACE for the entire "always-survivors" stratum is expressed using SACE_u as:

$$\mathrm{SACE} = E[Y^1 - Y^0|S^1 = 1, S^0 = 1]$$

$$= \sum_u E[Y^1 - Y^0 | (S^1 = 1, S^0 = 1), U = u] \Pr(U = u | S^1 = 1, S^0 = 1)$$
$$= \text{SACE}_{u=0} w' + \text{SACE}_{u=1}(1 - w'),$$

where $w' = \Pr(U = 0 | S^1 = 1, S^0 = 1)$, which can be rewritten, by using Bayes rule, as

$$\frac{\Pr(S^1 = 1, S^0 = 1 | U = 0) \Pr(U = 0)}{\Pr(S^1 = 1, S^0 = 1 | U = 0) \Pr(U = 0) + \Pr(S^1 = 1, S^0 = 1 | U = 1) \Pr(U = 1)}.$$

Here, the weighting of the different levels of U is different from that for the survival outcome. In fact, for the SACE, the weight is a proportion of U in the "always-survivors". Put differently, the expression for w' suggests that for transporting effects in the context of "truncation by death", not only the frequency of the effect modifier in the total population, that is, $\Pr(U = u)$, should be taken into consideration, but also the U level- and population-specific frequencies of the "always-survivors" stratum, $\Pr(S^1 = 1, S^0 = 1 | U = u)$, which are not identifiable except under strong assumptions. It is worth noting that the SACE defined as a risk difference is a directly collapsible causal measure (see Section 3 in Colnet et al. [12], and references therein), as it can be written as a weighted average of subgroup-specific SACEs, with weights corresponding to proportions of subgroups in the target population (that is, "always-survivors" stratum).

Effect modification by another cause of the survival outcome

We now consider Panel B of **Figure 1**, where a variable M that only affects survival is explicitly drawn. There is effect modification in the risk difference measure across strata of M for the effect of treatment A on survival S if the following holds:

$$E[S^1 - S^0 | M = 0] \neq E[S^1 - S^0 | M = 1].$$

In analyses on the SACE, that are restricted to the "always-survivors" principal stratum (that is, to one level of the response variable S^{T}), the non-causal path $M \rightarrow S^{\text{T}} \leftarrow U \rightarrow Y^{\text{T}}$ becomes open. In other words, in the total population, M is d-separated with Y^{T}, and they are independent; however, when conditioning on S^{T}, M is d-connected with Y^{T}, and they are likely

dependent. Therefore, M might be an effect modifier of the SACE even if it does not have an effect on quality of life; the following inequality may thus hold:

$$E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = 0] \neq E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = 1],$$

where $\mathrm{E}[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = m]$ is the SACE for individuals with $M = m$ (SACE_m).

The SACE_m, in its turn, can be expressed using SACE_u as follows:

$$\begin{aligned}
\mathrm{SACE}_m &= E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = m] \\
&= \sum\nolimits_u E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = m, U = u]\Pr(U = u|(S^1 = 1, S^0 = 1), M = m) \\
&= \sum\nolimits_u E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), U = u]\Pr(U = u|(S^1 = 1, S^0 = 1), M = m) \quad (\because Y^T \\
&\qquad \perp\!\!\!\perp M|(S^T, U)) \\
&= \mathrm{SACE}_{u=0} w''_m + \mathrm{SACE}_{u=1}(1 - w''_m),
\end{aligned}$$

where $w''_m = \Pr(U = 0|(S^1 = 1, S^0 = 1), M = m)$, which can be rewritten by using Bayes rule as

$$\frac{\Pr(S^1 = 1, S^0 = 1|M = m, U = 0)\Pr(M = m|U = 0)\Pr(U = 0)}{\sum_u \Pr(S^1 = 1, S^0 = 1|M = m, U = u)\Pr(M = m|U = u)\Pr(U = u)}.$$

Note that, although SACE_m is influenced by U, SACE_u is not influenced by M, because $Y^T \perp\!\!\!\perp M|(S^T, U)$ holds.

The SACE for the entire "always-survivors" stratum is expressed using SACE_m as:

$$\begin{aligned}
\mathrm{SACE} &= E[Y^1 - Y^0|S^1 = 1, S^0 = 1] \\
&= \sum\nolimits_m E[Y^1 - Y^0|(S^1 = 1, S^0 = 1), M = m]\Pr(M = m|S^1 = 1, S^0 = 1) \\
&= \mathrm{SACE}_{m=0} w''' + \mathrm{SACE}_{m=1}(1 - w'''),
\end{aligned}$$

where $w''' = \Pr(M = 0|S^1 = 1, S^0 = 1)$, which can be rewritten by using Bayes rule as

$$\frac{\Pr(S^1 = 1, S^0 = 1|M = 0)\Pr(M = 0)}{\Pr(S^1 = 1, S^0 = 1|M = 0)\Pr(M = 0) + \Pr(S^1 = 1, S^0 = 1|M = 1)\Pr(M = 1)}.$$

See the *eAppendix* for discussion on the relationship between w', w''_m, and w'''. In the *eAppendix*, we also present similar derivations for the SACE when this estimand is defined as a risk ratio; note that in this case, the weights described above do not apply, and the relevant weights depend on the subset of the "always-survivors" stratum whose potential outcome under absence of exposure corresponds to presence of the outcome of interest, that is, individuals with $(S^1 = 1, S^0 = 1, Y^0 = 1)$. Finally, notice that although above we assume that M and U are binary variables, the weights that were derived are equally applicable when M and U are categorical variables.

Numerical example

In **Table 1**, we present a numerical, simulated example that illustrates effect modification by U and M. The approach used to generate the data in the example is described in the *eAppendix*; parameter values used in the simulation are presented in **Table S1**. The example represents a hypothetical drug trial of patients with cancer, with Y corresponding to quality of life 1 year after treatment initiation, and S, to 1-year survival; M could be, for instance, a genetic factor that affects survival but that does not lead to clinically apparent morbidity that might affect quality of life, and U could, for example, correspond to age at enrolment (below $[U = 1]$ or above $[U = 0]$ 60 years), and is assumed to affect both survival and quality of life. In **Table 1**, we present proportions of the survival outcome-based principal strata by levels of these variables; the proportions of the "always-survivors" and "never-survivors" strata are respectively highest and lowest when $M = 1$ and $U = 1$. Note that there is no association between M and U in the total population. Further, we show the proportions of quality of life-based principal strata for individuals in the "always-survivors" stratum. In **Table S2** (*eAppendix*), we show, for this example, the dependence between Y^T and M in the "always-survivors" stratum, and also that $Y^T \perp\!\!\!\perp M|(S^T, U)$.

As shown in **Table 2**, we can calculate the quantities defined in the previous sections to show that both U and M are effect modifiers for the effect of A on S. For example, the effect of A on S for individuals with $U = 0$ corresponds to the difference in proportions between those in either the "always-survivors" or "protectable" strata and those in the "always-survivors" or "defiers" strata: $\{(0.177 + 0.032 + 0.071 + 0.058) - (0.177 + 0.016 + 0.071 + 0.049)\}/0.50 = 0.050$. For $U = 1$, the effect of A on S can be similarly computed as 0.072.

Finally, we can calculate the SACE for different levels of U and M. Note that U level-specific SACE can be directly calculated from the four rightmost columns in **Table 1**; for each U level, numbers of individuals in different quality of life-based principal strata are obtained by adding corresponding numbers of those with $M = 0$ and of those with $M = 1$. Consistent with the pattern for the effect of A on S for individuals with $U = 0$ versus $U = 1$, the $\text{SACE}_{u=0}$ is smaller than the $\text{SACE}_{u=1}$. For M level-specific SACEs, the computation shown in **Table 2** uses the formula for SACE_m in the previous section; for example, the weight ${w''}_{m=0}$ to calculate $\text{SACE}_{m=0}$ was obtained as $\Pr(U = 0|(S^1 = 1, S^0 = 1), M = 0) = 0.071/(0.071 + 0.148)$. Note however that SACE_m can also be calculated using, for each (Y^1, Y^0)-based principal stratum, the corresponding sum over U levels. In **Table S3**, for illustration, we present the difference between M-level specific SACEs under different assumptions for the effects of M and U on S and S and Y, respectively.

Table 1. Numerical example of effect modification by variables M and U. Data presented in this table (proportions, based on 100,000 simulated patients) were generated using the approach described in the *eAppendix*.

	Survival outcome (S)								Quality of life (Y)					
		S^a			$U = 1$		$U = 0$		Y^a		$U = 1$		$U = 0$	
SY type	Response type of S	S^1	S^0	Total population	$M = 1$	$M = 0$	$M = 1$	$M = 0$	Y^1	Y^0	$M = 1$	$M = 0$	$M = 1$	$M = 0$
1	"Always-survivors"	1	1	0.62	0.226	0.148	0.177	0.071	1	1	0.036	0.023	0.051	0.020
2	"Always-survivors"	1	1						1	0	0.130	0.086	0.041	0.017
3	"Always-survivors"	1	1						0	1	0.024	0.015	0.034	0.014
4	"Always-survivors"	1	1						0	0	0.036	0.022	0.051	0.020
5	"Protectable"	1	0	0.15	0.016	0.045	0.032	0.058	1	Undefined	NS	NS	NS	NS
6	"Protectable"	1	0						0	Undefined	NS	NS	NS	NS
7	"Defiers"	0	1	0.09	0.004	0.021	0.016	0.049	Undefined	1	NS	NS	NS	NS
8	"Defiers"	0	1						Undefined	0	NS	NS	NS	NS
9	"Never-survivors"	0	0	0.14	0.007	0.034	0.024	0.072	Undefined	Undefined	NS	NS	NS	NS
Total				1	0.25	0.25	0.25	0.25						

Footnote: Our simulation assigned individuals to S-defined principal strata based on levels of M and U, and for those in the "always-survivors" stratum, assignment to Y-defined principal strata was dependent on U. To simplify the simulation approach, individuals in other S-defined principal strata were not assigned to Y-defined principal strata (in the table, NS = not simulated); note however that if our goal was to study identification and estimation, then all individuals would need to be assigned to Y-defined principal strata. Due to rounding, totals and subtotals might be slightly different from the corresponding sum.

Table 2. Calculation of the effect of variable *A* on survival *S* and of the SACE for the numerical example introduced in **Table 1**.

Estimands		Calculation in the numerical example
Effect of *A* on *S*		
Total	$E[S^1 - S^0]$	$(0.62 + 0.15) - (0.62 + 0.09) = 0.06$
$U = 0$	$E[S^1 - S^0 \mid U = 0]$	$\frac{(0.177 + 0.032 + 0.071 + 0.058) - (0.177 + 0.016 + 0.071 + 0.049)}{0.50} = 0.050$
$U = 1$	$E[S^1 - S^0 \mid U = 1]$	$\frac{(0.226 + 0.016 + 0.148 + 0.045) - (0.226 + 0.004 + 0.148 + 0.021)}{0.50} = 0.072$
$M = 0$	$E[S^1 - S^0 \mid M = 0]$	$\frac{(0.148 + 0.045 + 0.071 + 0.058) - (0.148 + 0.021 + 0.071 + 0.049)}{0.50} = 0.066$
$M = 1$	$E[S^1 - S^0 \mid M = 1]$	$\frac{(0.226 + 0.016 + 0.177 + 0.032) - (0.226 + 0.004 + 0.177 + 0.016)}{0.50} = 0.056$
Effect of *A* on *Y* among "always-survivors"		
SACE	$E[Y^1 - Y^0 \mid (S^1 = 1, S^0 = 1)]$	$\frac{(0.130 + 0.086 + 0.041 + 0.017) - (0.024 + 0.015 + 0.034 + 0.014)}{0.62} = 0.302$
$\text{SACE}_{u=0}$	$E[Y^1 - Y^0 \mid (S^1 = 1, S^0 = 1), U = 0]$	$\frac{(0.051 + 0.041 + 0.020 + 0.017) - (0.051 + 0.034 + 0.020 + 0.014)}{0.248} = 0.040$
$\text{SACE}_{u=1}$	$E[Y^1 - Y^0 \mid (S^1 = 1, S^0 = 1), U = 1]$	$\frac{(0.036 + 0.130 + 0.023 + 0.086) - (0.036 + 0.024 + 0.023 + 0.015)}{0.372} = 0.476$
$\text{SACE}_{m=0}$	$E[Y^1 - Y^0 \mid (S^1 = 1, S^0 = 1), M = 0]$	$0.040\left(\frac{0.071}{0.071 + 0.148}\right) + 0.476\left(\frac{0.148}{0.071 + 0.148}\right) = 0.335$
$\text{SACE}_{m=1}$	$E[Y^1 - Y^0 \mid (S^1 = 1, S^0 = 1), M = 1]$	$0.040\left(\frac{0.177}{0.177 + 0.226}\right) + 0.476\left(\frac{0.226}{0.177 + 0.226}\right) = 0.285$

Discussion

Settings with "truncation by death" present analytic and interpretative difficulties to investigators. Specifically, the possible existence of factors (here, U) that affect survival and the outcome of interest might lead to selection bias when analyses are conditioned on observed survival S, rather than on the membership to the "always-survivors" principal stratum [5]. Here, we showed that the SACE is a function of the SACE for each level of U (i.e., $\mathrm{E}[Y^1 - Y^0 | (S^1 = 1, S^0 = 1), U = u]$), and of the proportions of U in the "always-survivors" (i.e., $\Pr(U = u | S^1 = 1, S^0 = 1)$). One implication is that the SACE estimated in one population might not be transportable to another population even if population level frequencies of the effect modifier U (i.e., $\Pr(U = u)$) and U level-specific SACEs are the same in the two populations. We further showed that a variable M only affecting survival might also be a modifier of the SACE, if a common cause, U, of the survival and outcome of interest is present. This occurs because conditioning on the "always-survivors" principal stratum likely leads to a non-causal association between M and U. Furthermore, the formula for M level-specific SACEs implies that the transportability of these effects depends on U level-specific SACEs and on the relation between M and U in the "always-survivors" stratum. When using M level-specific SACEs estimated in one population in another population, investigators should thus consider whether these conditions are similar in the target population; or more precisely, concerning the relation between M and U, whether the proportion of individuals with $U = 0$ varies in a similar manner by levels of M in the "always-survivors" strata in the two populations.

It is important to emphasize that our goal was to formally define estimands in the strata of potential effect modifiers, and determine their contributions to the SACE for the entire population of "always-survivors". In fact, several approaches have been described for identification and estimation of the SACE: point identification might be possible under assumptions that are often considered stronger than those required for identification of standard average causal effects (that is, effects that are not defined in terms of principal strata) (see Chiba et al. [13] for different assumptions that can identify the SACE; see also Section 2 of Tchetgen Tchetgen [14] for a discussion of methods that have been proposed); when these assumptions are not plausible, then large sample bounds can be used [4]; finally, sensitivity analyses [15] that require assuming different values of non-identifiable parameters can also be

undertaken to determine the range of values for the SACE that are consistent with the data, assumptions and background knowledge. In many of these approaches, monotonicity of the effect of exposure on the post-treatment variable has been assumed; in the *eAppendix*, we illustrate how the monotonicity assumption could be used to identify the weights derived above. Investigators who will develop identification and estimation methods for the SACE might consider incorporating some of the results presented above, especially for applications where effect modification is suspected and determining the magnitude of the SACE in different subpopulations is useful. Note that, in addition to translation of our formal results into identification and estimation approaches, two other related directions for future research are: investigation of potentially truncated outcomes that are measured repeatedly, at multiple time points, which complexifies definitions of principal strata, potentially with time-varying exposures; and relaxing the assumption of no loss to follow-up

The work described here partially relates to previous studies that have considered effect modification in the context of "truncation by death". In the analysis by Wang and colleagues [3], the authors proposed an approach to identify the SACE that requires existence of variables that meet assumptions analogous to those used in instrumental variable analyses. To relax one of the identification assumptions, the authors proposed an alternative assumption that involves absence of modification (by the exposure of interest) of the effect of a baseline covariate, that does not correspond to the exposure of interest, on the outcome. In another study, Greene and colleagues [16] described a new estimand, balanced-SACE, that corresponds to an effect of the exposure on the outcome in a subset of the "always-survivors" stratum, with the subset defined in terms of survival times; in that study, the authors considered stratification by effect modifiers to limit bias. Note, however, that different from these previous studies, the current work focuses on formally describing the contribution (that is, the weights) of effect modifier stratum-specific SACEs to the SACE in the total "always-survivors" population.

When, then, should we investigate modification of the SACE? The answer partially relates to considerations on modification of standard average causal effects (i.e., effects that are not defined in terms of principal strata). For instance, for standard causal effects, a reason for investigating effect modification is to inform decision making related to resource allocation – in other words, to identify which subgroups of the total population would benefit most from the implementation of a particular intervention. On the other hand, as with other principal strata

effects, this motivation for studying effect modification, namely, to directly inform public health decision making, might be less often relevant in the context of SACE, since principal strata membership in the target population is not usually discernible. There are, however, two other reasons for investigating effect modification: first, quantifying the SACE in different strata of effect modifiers has, one could argue, explanatory value, as it would improve understanding of the underlying data generating process; second, when comparing the SACE in different populations (or relatedly when considering transportability of the SACE), estimating the SACE by levels of effect modifiers and understanding how stratum-specific SACEs are weighted would be scientifically valuable. For these reasons, reporting of the SACE, or of its bounds or of sensitivity analyses, by strata of potential effect modifiers would be informative.

While the current work uses the principal stratification framework, other approaches should also be considered in studies with "truncation by death". A notable example is the conditional separable effect described by Stensrud and colleagues [7]. The assumptions required for and the interpretation of results from this type of analysis are considerably different from those in the principal stratification framework. For instance, conditional separable effects require conceiving of modified versions of the treatment that can be interpreted as directly affecting the outcome of interest. Despite the important conceptual differences between the SACE and conditional separable effects, effect modification might also be relevant for the latter estimand. In fact, given that conditional separable effects are not restricted to population strata that are not discernible and refer to a subset of the population that may be observable in a future experiment, analyses of the modification of these effects could be more relevant to future decisions related to resource allocation than analyses of the modification of the SACE. Of note, as shown in Proposition 1 of the study by Stensrud and colleagues [7], if a monotonicity assumption, the modified treatment assumption (Condition 5 in their study), and full isolation (Conditions 2 and 7) hold, a particular conditional separable effect is equivalent to the SACE.

We conclude with a clarification of how our focus on effect modification in measure, more precisely risk difference measure, relates to effect modification in distribution, which was described by VanderWeele and corresponds to variation in the probability distribution of potential outcomes across strata of the effect modifier (see Definition 3 in VanderWeele [17]). As shown in Proposition 2 of VanderWeele [17], for a binary outcome variable (in our context,

the survival variable S or the outcome variable Y), effect modification in distribution is equivalent to effect modification in risk difference measure or risk ratio measure. Thus, effect modification in risk difference measure, as expressed in formulas above, implies effect modification in distribution, but not vice versa.

References

1. Rubin, D.B., *Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death.* Statistical Science, 2006. **21**(3): p. 299-309.
2. McGuinness, M.B., et al., *A comparison of methods to estimate the survivor average causal effect in the presence of missing data: a simulation study.* BMC Med Res Methodol, 2019. **19**(1): p. 223.
3. Wang, L., X.H. Zhou, and T.S. Richardson, *Identification and estimation of causal effects with outcomes truncated by death.* Biometrika, 2017. **104**(3): p. 597-612.
4. Zhang, J.L. and D.B. Rubin, *Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by "Death".* Journal of Educational and Behavioral Statistics, 2003. **28**(4): p. 353-368.
5. Suzuki, E., *Generalized causal measure: the beauty lies in its generality.* Epidemiology, 2015. **26**(4): p. 490-5.
6. Frangakis, C.E. and D.B. Rubin, *Principal stratification in causal inference.* Biometrics, 2002. **58**(1): p. 21-9.
7. Stensrud, M.J., et al., *Conditional Separable Effects.* Journal of the American Statistical Association, 2023. **118**(544): p. 2671-2683
8. Young, J.G., et al., *A causal framework for classical statistical estimands in failure-time settings with competing events.* Stat Med, 2020. **39**(8): p. 1199-1236.
9. Suzuki, E., et al., *A counterfactual approach to bias and effect modification in terms of response types.* BMC Med Res Methodol, 2013. **13**: p. 101.
10. VanderWeele, T.J. and J.M. Robins, *Four types of effect modification: a classification based on directed acyclic graphs.* Epidemiology, 2007. **18**(5): p. 561-8.
11. VanderWeele, T.J., *On the distinction between interaction and effect modification.* Epidemiology, 2009. **20**(6): p. 863-71.
12. Colnet, B., et al. *Risk ratio, odds ratio, risk difference... Which causal measure is easier to generalize?* 2024.
13. Chiba, Y., M. Taguri, and Y. Uemura, *On the Identification of the Survivor Average Causal Effect.* Journal of Biometrics & Biostatistics, 2011. **2**.
14. Tchetgen Tchetgen, E.J., *Identification and estimation of survivor average causal effects.* Stat Med, 2014. **33**(21): p. 3601-28.
15. Chiba, Y. and T.J. VanderWeele, *A simple method for principal strata effects when the outcome has been truncated due to death.* Am J Epidemiol, 2011. **173**(7): p. 745-51.
16. Greene, T., et al., *The balanced survivor average causal effect.* Int J Biostat, 2013. **9**(2): p. 291-306.
17. VanderWeele, T.J., *Confounding and effect modification: distribution and measure.* Epidemiol Methods, 2012. **1**(1): p. 55-82.