Survival mediation analysis with the death-truncated mediator: The completeness of the survival mediation parameter

An-Shun Tai[1*], Chun-An Tsai[1**], Sheng-Hsuan Lin[1***]

1. Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan. Assembly Building I, 1001 University Road, Hsinchu, Taiwan 30010

[*] anshuntai@nctu.edu.tw

[**] beg2931as@gmail.com

[***] shenglin@stat.nctu.edu.tw

Corresponding author
Sheng-Hsuan Lin, MD, ScD
Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan
1001 University Road,
Hsinchu, Taiwan 30010
Cell: 886-3-5712121 ext 56822
E-mail: shenglin@stat.nctu.edu.tw

Summary

In medical research, the development of mediation analysis with a survival outcome has facilitated investigation into causal mechanisms. However, studies have not discussed the death-truncation problem for mediators, the problem being that conventional mediation parameters cannot be well-defined in the presence of a truncated mediator. In the present study, we systematically defined the completeness of causal effects to uncover the gap, in conventional causal definitions, between the survival and nonsurvival settings. We proposed three approaches to redefining the natural direct and indirect effects, which are generalized forms of the conventional causal effects for survival outcomes. Furthermore, we developed three statistical methods for the binary outcome of the survival status and formulated a Cox model for survival time. We performed simulations to demonstrate that the proposed methods are unbiased and robust. We also applied the proposed method to explore the effect of hepatitis C virus infection on mortality, as mediated through hepatitis B viral load.

Keywords: Cox proportional hazards model, Death-truncated mediator, Inverse odds ratio weighting, Inverse probability weighting, Regression-based method, Survival mediation analysis.

1. Introduction

1.1. Death-truncation problem

Mediation analysis is a technique used to investigate the mechanism of an already-confirmed causal effect. Several methods have been proposed for various settings, including binary outcomes, mixed model, time-varying settings, and multiple mediators (Huang and Cai, 2015; Lin et al., 2017a; Lin et al., 2017b; VanderWeele and Tchetgen Tchetgen, 2017; VanderWeele and Vansteelandt, 2010; VanderWeele and Vansteelandt, 2014; Zheng and van der Laan, 2012). In longitudinal studies, the problem of truncation by death arises when individuals die between follow-up visits. Thus, some variables may not be well-defined for dead individuals. The complete case approach is the conventional solution to this problem (Little, 1992; Little and Rubin, 2019); this method excludes individuals with death-truncated variables from the analysis. However, inference on the causal effects of exposure in the complete case approach could be biased even if experiments are randomized.

1.2. Literature review

To improve on this method, several models for causal inference have been proposed. Most studies have analyzed the causal effect of an exposure on nonsurvival outcomes truncated by death (Ding et al., 2011; Tchetgen Tchetgen, 2014; Wang, Zhou and Richardson, 2017; Zhang and Rubin, 2003). These studies have focused on estimating the survivor average causal effect (SACE), but they have omitted inference on causal mediation effects. In practice, SACE is not identifiable if further assumptions are not made (Zhang and Rubin, 2003). Sensitivity analysis is often performed to obtain a conservative estimate for SACE (Chiba and VanderWeele, 2011; Egleston et al., 2006; Gilbert, Bosch and Hudgens, 2003); alternatively, detailed covariate information for the identification process can be used (Ding et al., 2011; Tchetgen Tchetgen, 2014; Wang et al., 2017; Zhang and Rubin, 2003).

For causal mediation analysis with a survival outcome (referred to hereafter as survival mediation analysis), the problem of death truncation has received relatively little attention. Although methods for such mediation analysis have been adapted to survival outcomes (Fasanelli et al., 2019; Huang and Yang, 2017; Huang and Cai, 2015; Lange and Hansen, 2011; Tchetgen Tchetgen, 2011; VanderWeele, 2011), these methods require the assumption that the mediator is fully observed. If the mediator is death-truncated, the conventional mediation parameters are not well-defined, and these existing methods are therefore inappropriate for investigating the causal mechanism. Practically, this problem does not affect the total effect (TE), but it is critical for mediation analysis because the natural direct effect (NDE) and natural indirect effect (NIE) cannot be well-defined. To address this problem, two alternative formulations of mediation parameters have been separately proposed for the truncated mediator (Lin et al., 2017b; Zheng and van der Laan, 2017). Zheng and van der Laan proposed a random intervention formulation for the mediation parameter, based on a conditional mediator distribution with survival outcomes. By defining the conditional-intervention counterfactual, they formulated conditional mediation parameters in terms of time-varying variables to avoid the problem of death truncation. Similarly, Lin et al. adopted the interventional approach, where the intervention functions as an analogue of the set of causal effects in the survival setting, and the interventional analogue can be well-defined even if the mediator is truncated.

1.3. Unsolved problems and contributions of this study

Although two studies have proposed alternatives to conventional mediation parameters to remedy the problem of truncation in survival mediation analysis, three unsolved problems remain and should be addressed. First, in conventional causal definitions, it is unclear what the difference is between the survival setting and nonsurvival setting. Previous studies have suggested that conventional mediation parameters with survival outcomes can never be well-defined, but no mathematical proof for this suggestion has been proposed thus far. Second, current methodologies have their unique set of limitations. Specifically, the conditional mediation parameter proposed by Zheng and van der Laan requires information about each individual’s time-varying history to overcome the problem of undefined causal effects.

Therefore, the conditional mediation parameter cannot be applied to data sets that do not include time-varying covariates. Furthermore, strong sequential randomization assumptions are required for identification. As for the approach proposed by Lin et al., its limitation is that its interventional causal effects do not always sum up to the TE. The final problem is one pertaining to statistical inference: Zheng and van der Laan as well as Lin et al. have considered only the binary outcome of survival status rather than the single outcome of survival time. Therefore, the survival model must be extended—for example, to a Cox proportional hazards model.

To address these problems, we proposed three approaches to TE decomposition and comprehensively defined the mediation parameters for the death-truncated mediator. We also made appropriate assumptions to identify, through empirical data, the corresponding mediation parameter. By linking these proposed approaches, we formulated a theorem to illustrate the incompleteness of the conventional causal definitions in the survival setting. Additionally, based on the formula obtained using these approaches, we proposed three estimators using regression-based, inverse probability weighting (IPW), and inverse odds ratio weighting (IORW) methods to infer the causal effects. Binary survival status and survival time were both considered as outcome variables for modeling. The proposed estimators were illustrated by using Monte Carlo simulations and actual data sets.

1.4. Motivating example

This study was motivated by the Risk Evaluation of Viral Load Elevation and Associated Liver Disease/Cancer–Hepatitis B Virus (REVEAL-HBV) study—a community-based cohort study conducted in Taiwan to assess the effect of viral hepatitis on the development of hepatocellular carcinoma (HCC) (Chen et al., 2006). This study revealed that the viral loads of hepatitis B virus (HBV) and hepatitis C virus (HCV) play crucial roles in the development of HCC. Additionally, HCV inhibits HBV replication in patients with HBV/HCV coinfection. Thus, to understand the causal mechanism through which HCV affects the incidence of liver cancer and mortality, a mediation model is required to examine the effect of the HCV viral load on survival when mediated through the follow-up of HBV viral load among patients with HBV. However, in the REVEAL-HBV study, the follow-up of HBV viral loads for some patients with HBV-positive was truncated due to death. The truncation rate was 11.27%. The conventional causal mediation model, because it omits truncation events, can lead the researcher to misunderstand the causal mechanism through which HCV affects survival.

The remainder of this paper is organized as follows. In Section 2, we introduce the definitions and symbolism for mediation parameters and propose three approaches that address the problem of death truncation. In Section 3, we state the assumptions required and procedures for identifying the mediation parameters. In Section 4, we introduce three estimators for statistical inference based on the identified mediation parameters. In Section 5, we conduct a series of simulation studies to illustrate the performance of the proposed estimators by comparing them with the conventional complete case approach. An application to the investigation of the causal mechanism of HCV is illustrated in Section 6, and we discuss the strengths and limitations of the study in Section 7.

2. Notation and causal estimands

2.1. Notation and review of causal mediation analysis without previous death

Consider a longitudinal study, where T is the survival time and C_{ct} is the censoring time, where $\tilde{T} = \min(T, C_{ct})$ and $\delta = \mathrm{I}(T < C_{ct})$. In addition to survival time, we define an outcome indicator Y that indicates the binary survival status at the end of the follow-up period (1 represents survival and 0 represents death). Because survival time ($\tilde{T}$) and survival status (Y) are both of practical importance in medical research, we determine the causal estimands corresponding to survival time and survival status, separately. For the other variables, let A denote the exposure, M the mediator, C the baseline confounders, and Y_p the previous survival status during the period between A and M (1 represents survival and 0 represents death). The causal relationships between variables are described by a directed acyclic graph

(DAG), as illustrated in Figure 1. Notably, time-varying confounders cannot be included in the causal mechanism.

To conduct a causal mediation analysis, we further introduced counterfactual models for defining all effects (Robins and Greenland, 1992). Let $Y(a)$, $M(a)$, and $Y_p(a)$ denote the counterfactual values of Y, M, and Y_p, respectively, where $A = a$. Similarly, let $Y(a,m)$ denote the counterfactual of Y when $M = m$ and $A = a$. Additionally, let $Y(a,M(a^*))$ denote the counterfactual value of Y, where the exposure is set to a, and the mediator is set to the value it would take under exposure a^*. Subsequently, we make consistency and composition assumptions (Gibbard and Harper, 1978; Robins and Greenland, 1992; VanderWeele and Vansteelandt, 2009). According to the consistency assumption for $Y(a,m)$, the observed outcome Y is equal to $Y(a,m)$ when the observed values of A and M are a and m, respectively. For $M(a)$, this consistency assumption states that the observed mediator M is equal to $M(a)$ when the observed exposure is $A = a$. According to the composition assumption, $Y(a) = Y(a,M(a))$. Similarly, the counterfactual values of survival time, namely $T(a)$ and $T(a,M(a^*))$, follow the same definition.

For readability, we adopt survival status (Y) to illustrate the problem of conventional mediation analysis with previous death; a similar argument based on survival time (T) is provided in Web Appendix A. Conventionally, TE, NDE, and NIE are defined as follows based on the risk difference scale for the individual level (Pearl, 2001; Robins and Greenland, 1992):

$$\begin{aligned}
&\text{TE} = Y(1) - Y(0)\\
&\text{NDE} = Y\big(1, M(0)\big) - Y(0, M(0))\\
&\text{NIE} = \ Y\big(1, M(1)\big) - Y\big(1, M(0)\big) \qquad (1)
\end{aligned}$$

Let $\psi(a,a^*) \equiv E(Y(a,M(a^*)))$, where ψ is referred to as the mediation parameter (Pearl, 2001; Robins and Greenland, 1992) and is defined according to the expectation of the counterfactual value. Thus, the population level TE, NDE, and NIE can be defined as $\psi(1,1) - \psi(0,0)$, $\psi(1,0) - \psi(0,0)$, and $\psi(1,1) - \psi(1,0)$, respectively. In the Section 2.2, we discuss the problem encountered using this conventional definition when the mediator is truncated by previous death. In the subsequent sections, we consider the population level for identification and estimation.

2.2. Problem with using the conventional definition with previous death

Subjects can be assigned to the following four groups (also named principal strata) based on the status of Y_p under two counterfactual settings $(Y_p(1), Y_p(0))$ (Frangakis and Rubin, 2002; Wang et al., 2017): an (1) always-survivor group $(Y_p(1) = 1, Y_p(0) = 1)$: the subject always survives, regardless of exposure status; a (2) protected group $(Y_p(1) = 1, Y_p(0) = 0)$: the subject survives if exposed, but dies if not exposed; a (3) harmed group $(Y_p(1) = 0, Y_p(0) = 1)$: the subject dies if exposed, but survives if not exposed; and a (4) doomed group $(Y_p(1) = 0, Y_p(0) = 0)$: the subject always dies, regardless of exposure status. We denote the four groups as P_S, P_P, P_H, and P_D, respectively.

In the case of previous death, the composition assumption for Y is rewritten as $Y(a) = Y\left(a, Y_p(a), M\big(a, Y_p(a)\big)\right)$, and a further composition assumption is required for M, namely $M(a) = M\big(a, Y_p(a)\big)$. Thus, based on these new composition assumptions, the conventional causal effects in (1) can be rewritten as the follows:

$$\text{TE} = Y\left(1, Y_p(1), M\big(1, Y_p(1)\big)\right) - Y\left(0, Y_p(0), M\big(0, Y_p(0)\big)\right),$$
$$\text{NDE} = Y\left(1, Y_p(1), M\big(0, Y_p(0)\big)\right) - Y\left(0, Y_p(0), M\big(0, Y_p(0)\big)\right), \text{ and}$$
$$\text{NIE} = Y\left(1, Y_p(1), M\big(1, Y_p(1)\big)\right) - Y\left(1, Y_p(1), M\big(0, Y_p(0)\big)\right).$$
(2)

To clearly state the weakness of the conventional definition with regard to previous death in (2), we first define the completeness of causal effects as follows:

Definition 1. (Completeness of causal effects)
If the causal effects, namely TE, NDE, and NIE, are well-defined in all principal strata, then the formation of the defined causal effects is complete.

Based on *Definition 1*, we now show that the conventional definition of causal effects with previous death lacks completeness. First, in the always-survivor group (P_S), the counterfactual values of Y and M can be defined, and therefore all causal effects are well-defined. However, in the protected group (P_P) (i.e., $Y_p(0) = 0$ and $Y_p(1) = 1$), the counterfactual outcome $Y\left(1, Y_p(1), M\big(0, Y_p(0)\big)\right)$ is equal to $Y\left(1, y_p = 1, M\big(0, y_p = 0\big)\right)$. This cannot be defined because although its hypothetical status supposes no previous death, it includes $M\big(0, y_p = 0\big)$, which is the death-truncated mediator. Furthermore, $Y\left(0, Y_p(0), M\big(0, Y_p(0)\big)\right)$ also includes the death-truncated mediator, and according to its survival status, the individual is subject to previous death ($y_p = 0$), which implies that $Y\left(0, Y_p(0), M\big(0, Y_p(0)\big)\right)$ is always equal to zero. Consequently, we can define TE in P_P, but not in NDE and NIE. Similarly, in P_H, all counterfactual values of Y are either well-defined or zero, and in P_D, all counterfactuals are zero.

Table 1 illustrates the definition statuses of TE, NDE, and NIE in the four groups, where these definition statuses are such that TE is well-defined for all groups, but NDE and NIE are undefined in P_P. Therefore, the conventional causal definitions for NDE and NIE in this context of survival analysis are incomplete.

2.3. Three approaches to death-truncated mediation analysis

To address the incompleteness of the conventional causal definition, we formulated the following three approaches to redefine the NDE and NIE for the death-truncated mediator:

Approach 1: *Principle stratification*

In this approach, we maintain the conventional causal definition but define all effects only under a certain principal stratum, namely, the always-survivor group (P_S). This strategy is often used for estimating TE when the nonmortality outcome is truncated by death (Frangakis and Rubin, 2002; Zhang and Rubin, 2003). The mediation parameter for P_S is defined as $E(Y\big(a, M(a^*)\big)|P_S)$, which is well-defined. Following this approach, the conditional mediation parameters can only be identified with strong additional assumptions (Ding et al., 2011; Tchetgen Tchetgen, 2014; Wang et al., 2017); without these assumptions, only the boundary can be evaluated (Chiba and VanderWeele, 2011). Moreover, if we assume that the counterfactuals are homogeneously distributed across the principal strata, then the conditional causal effects can be interpreted as the average causal effect of the mediator on the outcome (Forastiere, Mattei and Ding, 2018).

Approach 2: *Decreasing monotonicity assumption for* Y_p *(i.e.,* $Y_p(1) \leq Y_p(0)$*); equivalently, assumption of no protected group*

In the second approach, we assume decreasing monotonicity. If the exposure leads to on the early death, we can assume that no individuals benefit from exposure (i.e., no individual belongs to P_P). Because the counterfactual values of Y in (2) are well-defined under the other three principal strata, all effects are well-defined under this assumption. However, if background knowledge indicates that exposure can have a protective effect against death for some individuals (i.e., the protected group includes some individuals), then this approach is not applicable. Notably, if we can assume the absence of a harmful effect, then the total direct effect and pure indirect effect can be identified in the presence of a protective effect as an alternative for effect decomposition. This argument is detailed in Web Appendix A.

Approach 3: *Death-truncated analogues of NDE and NIE* (i.e., NDE_{dt} *and* NIE_{dt})

Although *Approaches 1* and *2* have been widely used to estimate causal effects in survival analysis, these approaches cannot comprehensively solve the problem of death truncation. Therefore, in *Approach 3*, we adopt the sums of sets of path-specific effects (PSEs) as analogues for NDE and NIE, which are well-defined for all four groups. In the case with multiple mediators, a PSE is proposed to quantify the effect of the exposure on the outcome when mediated through a pathway comprised of the mediators of interest (Avin, Shpitser and Pearl, 2005; Daniel et al., 2015). Based on the definitions of the PSEs with Y_p and M as mediators, the TE from A to Y can be decomposed into four PSEs (Figure 1): (1) a PSE through Y_p only (i.e., $PSE_{A\to Y_p\to Y}$), (2) a PSE through M only (i.e., $PSE_{A\to M\to Y}$), (3) a PSE through neither Y_p nor M (i.e., $PSE_{A\to Y}$), and (4) a PSE through Y_p and then M (i.e., $PSE_{A\to Y_p\to M\to Y}$). This decomposition can be expressed as follows:

$$
\begin{aligned}
Y(1)-Y(0) = {} & \left\{Y\left(1,Y_p(1),M\left(1,Y_p(1)\right)\right)-Y\left(1,Y_p(1),M\left(0,Y_p(1)\right)\right)\right\} \\
& +\left\{Y\left(1,Y_p(1),M\left(0,Y_p(1)\right)\right)-Y\left(1,Y_p(1),M\left(0,Y_p(0)\right)\right)\right\} \\
& +\left\{Y\left(1,Y_p(1),M\left(0,Y_p(0)\right)\right)-Y\left(1,Y_p(0),M\left(0,Y_p(0)\right)\right)\right\} \\
& +\left\{Y\left(1,Y_p(0),M\left(0,Y_p(0)\right)\right)-Y\left(0,Y_p(0),M\left(0,Y_p(0)\right)\right)\right\} \\
= {} & PSE_{A\to M\to Y}+PSE_{A\to Y_p\to M\to Y}+PSE_{A\to Y_p\to Y}\ +PSE_{A\to Y}
\end{aligned}
$$

According to (2), NIE and NDE are equal to $PSE_{A\to M\to Y}+PSE_{A\to Y_p\to M\to Y}$ and $PSE_{A\to Y_p\to Y}+PSE_{A\to Y}$, respectively.

However, this effect decomposition is inappropriate in the presence of a death-truncated mediator, and an alternative effect decomposition is required. There are two reasons for why an alternative effect decomposition is required. First, the effects of paths $A\to Y_p\to Y$ and $A\to Y_p\to M\to Y$ cannot be separated in terms of identification and definitions. Because Y_p and Y are survival statuses measured at different times, the assumption that there are no unmeasured confounders between Y and Y_p does not hold. Therefore, we can only identify the effect through the pathways involving only Y_p (i.e., the combination of $A\to Y_p\to Y$ and $A\to Y_p\to M\to Y$) (Vanderweele, Vansteelandt and Robins, 2014). In Section 2.2, we show that $Y(1,Y_p(1),M(0,Y_p(1)))$ cannot be defined in P_P, indicating that $PSE_{A\to Y_p\to M\to Y}$ and $PSE_{A\to Y_p\to Y}$ cannot be well-defined separately. Second, NDE should be defined as a combination of effects through the pathway involving neither Y_p nor M (i.e.,$PSE_{A\to Y}$) and the pathway involving only Y_p (i.e., $PSE_{A\to Y_p\to M\to Y}+PSE_{A\to Y_p\to Y}$). In the counting process for survival time $N(t)=I(T>t)$, Y_p and Y are the survival statuses corresponding to two time points, denoted as $Y_p=dN(t_1)$ and $Y=dN(t_2)$, where $t_2>t_1$. Thus, the effects related to path $A\to Y_p$ can be regarded as a source contributing to the direct effect on the survival process. Moreover, the causal effect that passes through $Y_p\to M$ is meaningless, because, in our definition, $Y_p\to M$ represents the occurrence of a truncation event, which is deterministic rather than causal. Therefore, the path $A\to Y_p\to M\to Y$ can be more reasonably included in the direct effect than in the indirect effect.

Based on these two reasons, we propose the following alternative definitions of NDE and

NIE, namely death-truncated NDE $(\mathrm{NDE}_{\mathrm{dt}})$ and death-truncated NIE $(\mathrm{NIE}_{\mathrm{dt}})$, which are suitable for the case with a death-truncated mediator:

$$\begin{aligned}
\mathrm{NDE}_{\mathrm{dt}} &= PSE_{A\to Y_p\to M\to Y} + PSE_{A\to Y_p\to Y} + PSE_{A\to Y} \\
&= Y\left(1, Y_p(1), M\left(0, Y_p(1)\right)\right) - Y\left(0, Y_p(0), M\left(0, Y_p(0)\right)\right) \\
\mathrm{NIE}_{\mathrm{dt}} &= PSE_{A\to M\to Y} \\
&= Y\left(1, Y_p(1), M\left(1, Y_p(1)\right)\right) - Y\left(1, Y_p(1), M\left(0, Y_p(1)\right)\right)
\end{aligned}$$

(3)

In contrast to the conventional causal effects, the proposed formulations of death-truncated causal effects are complete (proof provided in Web Appendix A). Table 1 compares the definition statuses of the death-truncated and conventional causal effects.

In (3), $\mathrm{NDE}_{\mathrm{dt}}$ and $\mathrm{NIE}_{\mathrm{dt}}$ are identical to NDE and NIE in (2), respectively, when NDE and NIE are well-defined. This is stated in *Theorem 1.*

Theorem 1. (*Equivalence of death-truncated causal effects and conventional causal effects*) *In the presence of previous death, the death-truncated causal effects,* NDE_{dt} *and* NIE_{dt}*, are identical to conventional causal effects, NDE and NIE, respectively, in groups* P_S *(i.e.,* $Y_p(1) = 1, Y_p(0) = 1$*),* P_H *(i.e.,* $Y_p(1) = 0, Y_p(0) = 1$*), and* P_D *(i.e.,* $Y_p(1) = 0, Y_p(0) = 0$*).*

The proof of *Theorem 1* is provided in Web Appendix A. According to *Theorem 1* and the fact that NDE_{dt} and NIE_{dt} are complete, the proposed death-truncated causal effects are generalizations of the conventional causal effects. Finally, when death truncation occurs, let $\phi(a, a^*) \equiv E\big(Y(a, Y_p(a), M(a^*, Y_p(a)))\big)$, which is referred to as the survival mediation parameter. *Approach 3* provides novel causal estimands as the average causal effects at the population level; thus, the population level TE, $\mathrm{NDE}_{\mathrm{dt}}$, and $\mathrm{NIE}_{\mathrm{dt}}$ are defined as $\phi(1,1) - \phi(0,0)$, $\phi(1,0) - \phi(0,0)$, and $\phi(1,1) - \phi(1,0)$, respectively. Although these approaches are defined in terms of the expectation of time-invariant outcomes, we can extend these approaches to survival analysis by defining the survival mediation parameters as the hazard function or survival function on survival (Huang and Yang, 2017; Huang and Cai, 2015; Tchetgen Tchetgen, 2011).

Finally, we summarize the features of *Approaches 1* to *3* by considering the mediation parameters and effect decomposition. First, in *Approaches 1* and *2*, causal estimands are defined based on the conventional mediation parameter $\psi(a, a^*)$; by contrast, *Approach 3* adopts the survival mediation parameter $\phi(a, a^*)$ to define causal estimands. Second, the type of effect decomposition varied between these approaches: in *Approach 1*, TE was decomposed for the principle stratum (i.e., the always-survivor group P_S,), whereas in *Approaches 2* and *3*, TE was directly decomposed for the whole population, where *Approach 2* required a decreasing monotonicity assumption for Y_p. Furthermore, *Approaches 2* and *3* have identical statistical parameters, the proof of which is provided in Section 3. Because *Approach 1* focuses on the effect decomposition for P_S rather than the whole population, we only discuss identification for *Approaches 2* and *3* in Section 3.

3. Identification

As detailed in the previous section, the mediation parameters of *Approaches 2* and *3* are $\psi(a, a^*)$ and $\phi(a, a^*)$; to identify these parameters, five assumptions are required.

Assumption 1: *There is no unmeasured confounder between the exposure and overall survival status, including previous death status and final death status.*

$$(Y_p(a), Y(a, y_p = 1, m)) \coprod A\,|C$$

Assumption 2: *There is no unmeasured confounder between the mediator and final death status.*

$$Y(a, y_p = 1, m) \coprod M|A,\ \ y_p = 1, C$$

Assumption 3: *There is no unmeasured confounder between the exposure and the mediator.*

$$M(a, y_p = 1) \coprod A\,|C$$

Assumption 4: *Confounders between the mediator and overall survival status (previous death status and final death status) are not affected by previous covariates.*

$$(Y_p(a), Y(a, y_p = 1, m)) \coprod M(a', y_p = 1)\,|C$$

Assumption 5: *There is no unmeasured confounder between the mediator and previous death status.*

$$M(a, y_p = 1) \coprod Y_p|A\,, C$$

If there is no previous death (i.e., $Y_p(1) = Y_p(0) = 1$), *Assumptions 1* to *4* reduce to the four conventional assumptions in causal mediation analysis (VanderWeele and Vansteelandt, 2009) (i.e., $Y(a,m)) \coprod A\,|C$, $Y(a,m) \coprod M\,|A, C$, $M(a) \coprod A\,|C$, and $Y(a,m) \coprod M(a')\,|C$).

Moreover, *Assumption 5* is excluded because Y_p is always equal to zero. Accordingly, the assumptions required for identification and the proposed formation of causal effects in Section 2 are generalized versions of conventional causal models for the survival setting.

We can describe all assumptions in terms of a nonparametric structural equation model, according to which the data generation process is described as a function of previous variables and an error term:

$A = g_A(\varepsilon_A)$
$Y_p = g_p(A, U, \varepsilon_{Y_p})$
$M = g_M(A, \varepsilon_M)\ if\ Y_p = 1$; undefined $if\ Y_p = 0$
$Y = g_Y(A, Y_p, U, M, \varepsilon_Y)$

If C prevents confounding among A, M, and (Y_p, Y)—that is, ε_A, ε_M, and $(\varepsilon_{Y_p}, \varepsilon_Y)$ are independent—then all five assumptions are satisfied. The correspondence is presented in Web Appendix B. Both mediation parameters, namely $\phi(a, a^*)$ and $\psi(a, a^*)$, can be nonparametrically identified using the following two theorems.

Theorem 2. *Under positivity, consistency, and Assumptions 1 to 5,* $\phi(a, a^*)$ *can be identified as* $Q(a, a^*)$, *where*

$$Q(a, a^*) = \int_{m,c} E\big[Y|A = a, Y_p = 1, M = m, C = c\big]\ f(Y_p = 1|A = a, C = c) f_{M|A,C}\big(M = m|A = a^*, Y_p = 1, C = c\big) f(c) \quad dmdc$$

Theorem 3. *Under the decreasing monotonicity assumption for* Y_p $(Y_p(1) \leq Y_p(0))$, *positivity, consistency, and Assumptions 1 to 5, for* $a \geq a^*$, $\psi(a, a^*)$, *as defined in Approach 2, can be identified to be* $Q(a, a^*)$, *which is defined in Theorem 1.*

The proofs of *Theorems 2* and *3* are provided in Web Appendix B. According to these theorems, $\phi(a, a^*)$ and $\psi(a, a^*)$ are identified as the identical statistical parameter $Q(a, a^*)$, hereafter referred to as the survival mediation formula. The proposed survival mediation formula is an extension of Pearl's mediation formula (Pearl, 2001). Therefore, on the risk difference scale, TE, NDE (or $\mathrm{NDE_{dt}}$), and NIE (or $\mathrm{NIE_{dt}}$) for the binary survival status (Y) are exactly $Q(1,1) - Q(0,0)$, $Q(1,0) - Q(0,0)$, and $Q(1,1) - Q(1,0)$, respectively.

Consider the survival mediation parameter in terms of survival time (T). In conventional survival mediation analysis (Cho and Huang, 2019; Huang and Yang, 2017), the mediation parameter is defined as a log hazard; thus, in the presence of the truncated mediator, we defined the survival mediation parameter in terms of survival time as follows:

$$\phi_T(a,a^*) \equiv log\lambda\big(T(a,Y_p(a),M(a^*,Y_p(a)));\;\; t\big). \qquad (4)$$

The survival mediation parameter in (4) is identified in the following lemma.

Lemma 1. *Under positivity, consistency, and Assumptions 1 to 5,* $\phi_T(a,a^*)$ *can be identified as*

$$Q_T(a,a^*) = log(\vartheta_T^1(a,a^*)/\vartheta_T^2(a,a^*))$$

where

$$\vartheta_T^1(a,a^*) = \int_{m,c} \lambda(t|a,m,c,y_p=1)e^{-\Lambda(t|a,m,c,y_p=1)}\; f(Y_p=1|A=a,C = c)f_{M|A,C}\big(M=m|A=a^*,Y_p=1,C=c\big)f(c)\,dmdc,$$

and

$$\vartheta_T^2(a,a^*) = \int_{m,c} e^{-\Lambda(t|a,m,c,y_p=1)}\; f(Y_p=1|A=a,C = c)f_{M|A,C}\big(M=m|A=a^*,Y_p=1,C=c\big)f(c)\,dmdc.$$

In *Lemma 1*, $\lambda(t|a,m,c,y_p=1)$ is the conditional hazard function and $\Lambda(t|a,m,c,y_p=1)$ is the conditional cumulated hazard function. The identification assumptions and the proof are provided in Web Appendix B. To quantify $Q(a,a^*)$ and $Q_T(a,a^*)$, we propose estimators for statistical inference in Section 4.

4. Statistical inference for $Q(a,a^*)$ and $Q_T(a,a^*)$

4.1. Regression-based, IPW, and IORW methods for $Q(a,a^*)$

Based on the survival mediation formulas in *Theorems 2* and *3*, we propose three methods for estimating NDE_{dt} and NIE_{dt}: the regression-based, IPW, and IORW methods. The regression-based method is a common parametric mediation technique used to derive the analytic solution of the causal effects by assuming the appropriate regression of the variables. However, model misspecification results in parametric models that lack power for statistical inference. Therefore, we also developed two weighted methods that are more robust to estimate the causal effects. The three proposed methods are detailed in the following sections.

4.1.1. Regression-based estimator

In this method, the model distributions of Y, M, and Y_p should be specified. Thus, we assumed that the outcome Y and early survival status Y_p followed the logistic regression for the binary survival status. For the distribution of the mediator, we considered both a normal distribution for the continuous mediator and a logistic model for the binary mediator. For simplicity, the formula for the binary mediator is presented herein. The other cases are provided in Web Appendix C. The exposure is dichotomous. This model setting is a prominent case in medical research. Other model distributions are applied to this method through integration with the Monte Carlo approach, which is a type of G-computation (Robins, 1986). Table 2 presents the sequential constructions of the regression models of Y, M, and Y_p based on the DAG in Figure 1.

Based on this regression setting, the parameters $\Theta_1 \equiv \{\{\alpha\}, \{\beta\}, \{\theta\}\}$ are estimated using the maximum likelihood (ML) approach, and the estimator is denoted as $\widehat{\Theta}_1$. Subsequently, we derive the estimators of NDE and NIE through the regression-based method as $\widehat{NDE}_{dt}^{R} = \hat{Q}_R(1,0) - \hat{Q}_R(0,0)$ and $\widehat{NIE}_{dt}^{R} = \hat{Q}_R(1,1) - \hat{Q}_R(1,0)$, respectively, where $\hat{Q}_R(a, a^*)$ is the estimator obtained when using $\widehat{\Theta}_1$.

4.1.2. IPW estimator

In this section, we use the estimator obtained through the IPW method to evaluate NDE and NIE. Under consistency and exchangeability assumptions, Lange et al (2012) derived the inverse probability (IP) weight for mediation analysis without a truncated event (Lange and Hansen, 2011). Following the IPW method, the survival mediation formula $Q(a, a^*)$ can be rewritten as an expectation with respect to the outcome, mediator, and exposure, as stated in *Theorem 4.*

***Theorem 4* (*IPW estimation*)**

The survival mediation formula can be rewritten as

$$Q_{IPW}(a, a^*) = E[W(m, a, a^*) \times Y],$$

where $W(m, a, a^*)$ *is the weight and has the form*

$$\frac{f(M = m|A = a^*, Y_p = 1, C)I(A = a)I(Y_p = 1)}{f(M = m|A = a, Y_p = 1, C)f(A = a|C)}.$$

The proof of *Theorem 4* is provided in Web Appendix C. The weight in *Theorem 4* is equivalent to the conventional IP weight for mediation analysis without a truncated event constrained by $Y_p = 1$ (Lange and Hansen, 2011).

To calculate the IP weight, we assume models of the mediator and exposure as shown in Table 2. The ML approach is used to estimate $\Theta_2 \equiv \{\{\beta\}, \{\delta\}\}$ as $\widehat{\Theta}_2$. According to *Theorem 4*, the survival mediation formula of IPW can be derived by

$$\hat{Q}_{IPW}(a, a^*) = \mathbb{P}_n\left[\frac{\hat{f}(M = m|A = a^*, Y_p = 1, C = c, \hat{\beta}_0, \hat{\beta}_A, \hat{\beta}_C)I(A = a)I(Y_p = 1)}{\hat{f}(M = m|A = a, Y_p = 1, C = c, \hat{\beta}_0, \hat{\beta}_A, \hat{\beta}_C)\hat{f}(A = a|C = c, \hat{\delta}_0, \hat{\delta}_A)}Y\right]$$

where $\mathbb{P}_n[\cdot] = n^{-1}\sum_i[\cdot]_i$, and the direct and indirect effects of IPW are estimated by

$$\widehat{NDE}_{dt}^{IPW} = \hat{Q}_{IPW}(1,0) - \hat{Q}_{IPW}(0,0) \text{ and } \widehat{NIE}_{dt}^{IPW} = \hat{Q}_{IPW}(1,0) - \hat{Q}_{IPW}(0,0),$$

respectively. The IPW method has the advantage of fewer model assumptions for estimation compared with the regression-based method. However, specifying an appropriate model for the mediator remains challenging because the types of mediator measurement are various. To address this problem, we propose the IORW method, the details of which are presented in the following section.

4.1.3. IORW estimator

IORW was proposed by Tchetgen Tchetgen to improve parametric mediation techniques (Tchetgen Tchetgen, 2013). This approach leverages on the advantage of the invariance property of the odds ratio to define a new weight by replacing the conventional weight formed by the conditional distribution of the mediator with a regression of exposure on the mediator. The IORW method adopts the marginal structural model to define the causal effects and applies the estimating equation approach to calculate the estimates. For estimation using the IORW method, two parametric regression models are fitted, as shown in Table 2. Notably, the exposure in the IORW method is regressed on the mediator and the confounders, whereas, the exposure in the IPW method is regressed on the confounders only. The parameters $\boldsymbol{\alpha} =$

$\{\alpha_0, \alpha_A, \alpha_C\}$ and $\boldsymbol{\kappa} = \{\kappa_0, \kappa_M, \kappa_A\}$ of the regression model in Table 2 are estimated using the ML approach.

Following the two-step procedure for estimation in the study of IORW by (Tchetgen Tchetgen, 2013), we first derive the estimators of TE and NDE by using the weighted estimating equation (WEE) and IORW approaches, separately. NIE is then calculated by subtracting NDE from TE. To simplify the notation, we define the following two conditional expectations of the counterfactual outcome:

$$\phi(a, a^*|C = c) = E\big(Y(a, Y_p(a), M(a^*, Y_p(a)))\big|C = c\big), \text{ and}$$

$$\phi\big(a, a|C = c, Y_p = 1\big) = E\big(Y(a, Y_p(a), M(a^*, Y_p(a)))\big|C = c, Y_p = 1\big).$$

Weighted estimation of TE

First, we adopt the WEE approach to estimate TE. A regression model on $\phi\big(a, a|C = c, Y_p = 1\big)$ through a link function $g(\cdot)$ is defined as follows:

$$g^{-1}\left(\phi\big(a, a|C = c, Y_p = 1\big)\right) = \mu_{TE}(\boldsymbol{\eta}; a, c),$$

where $\mu_{TE}(\boldsymbol{\eta}; a, c) = \eta_0 + \eta_A a + \eta_C c + \eta_{AC} ac$ and $\boldsymbol{\eta} = \{\eta_0, \eta_A, \eta_C, \eta_{AC}\}$. For simplicity, we consider the identity link in this section. Suppose that the weight determined through the WEE approach is $w(a, c, \widehat{\boldsymbol{\alpha}}) = P(Y_p = 1|a, c, \widehat{\boldsymbol{\alpha}})$; thus, the WEE for any $\boldsymbol{\eta}$ can be written as follows:

$$U_{TE}(\boldsymbol{\eta}) = w(a, c, \widehat{\boldsymbol{\alpha}}) \times \Gamma_{TE}(a, c, \boldsymbol{\eta})\{Y - g(\mu_{TE}(\boldsymbol{\eta}; a, c))\},$$

where $\Gamma_{TE}(a, c, \boldsymbol{\eta}) = \partial g(\mu_{TE}(\boldsymbol{\eta}; a, c))/\partial\boldsymbol{\eta}$. Based on this estimating equation, the following theorem motivates our estimation strategy.

Theorem 5

Under the consistency assumption and Assumptions 1 to 5, $U_{TE}(\boldsymbol{\lambda})$ is an unbiased estimating equation conditioned on $C, Y_p = 1$ (i.e., $E_{Y,M|C,Y_p=1}\big(U_{TE}(\boldsymbol{\eta})\big) = 0$).

The proof of *Theorem 5* is provided in Web Appendix C. On the basis of this theorem, $\boldsymbol{\eta}$ is estimated by $\widehat{\boldsymbol{\eta}}$ which is the solution of the equation $\mathbb{P}_n[U_{TE}(\widehat{\boldsymbol{\eta}})] = 0$. As a result, TE can be directly derived through the following equation.

$$\widehat{TE}_{IORW} = \hat{\phi}(1{,}1) - \hat{\phi}(0{,}0) = E_C\left(\hat{\phi}(1{,}1|\boldsymbol{C}) - \hat{\phi}(0{,}0|\boldsymbol{C})\right)$$
$$= E_C\left(g(\mu_{TE}(\hat{\boldsymbol{\eta}}; 1, C))f_{Y_p}\left(y_p = 1\middle|1, C, \hat{\boldsymbol{\alpha}}\right) - g(\mu_{TE}(\hat{\boldsymbol{\eta}}; 0, C))f_{Y_p}\left(y_p = 1\middle|0, C, \hat{\boldsymbol{\alpha}}\right)\right).$$

IORW estimation of NDE

To estimate NDE, we fit the regression model on $\phi\left(a, 0\middle|C = c, Y_p = 1\right)$, where the exposure through the mediator is set as zero:

$$g^{-1}\left(\phi\left(a, 0\middle|C = c, Y_p = 1\right)\right) = \mu_{NDE}(\boldsymbol{\nu}; a, c),$$

where $\mu_{NDE}(\boldsymbol{\nu}; a, c) = \nu_0 + \nu_A a + \nu_C \boldsymbol{c} + \nu_{AC} a\boldsymbol{c}$ and $\boldsymbol{\nu} = \{\nu_0, \nu_A, \nu_C, \nu_{AC}\}$. To estimate the NDE, we apply the IORW to derive the estimator $\hat{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}$ under the survival mediation formula. Following the establishment of the odds ratio for the mediator and exposure in the study of IORW by Tchetgen Tchetgen (2013), the conditional odds ratio is modified as follows:

$$\mathrm{OR}\left(M, A\middle|C, Y_p = 1\right) \equiv \frac{f_{M|A,C,Y_p}\left(M\middle|A, Y_p = 1, C\right)f_{M|A,C,Y_p}\left(M = m_0\middle|A = a^*, Y_p = 1, C\right)}{f_{M|A,C,Y_p}\left(M = m_0\middle|A, Y_p = 1, C\right)f_{M|A,C,Y_p}\left(M\middle|A = a^*, Y_p = 1, C\right)p_{Y_p}\left(y_p = 1\middle|A, C, \alpha\right)}$$
$$= \frac{f_{A|M,C,Y_p}\left(A\middle|M, Y_p = 1, C\right)f_{A|M,C,Y_p}\left(A = a^*\middle|M = m_0, Y_p = 1, C\right)}{f_{A|M,C,Y_p}\left(A = a^*\middle|M, Y_p = 1, C\right)f_{A|M,C,Y_p}\left(A\middle|M = m_0, Y_p = 1, C\right)p_{Y_p}\left(y_p = 1\middle|A, C, \alpha\right)},$$

where $f_{A|M,C,Y_p}$ is the density function of *A* given (M, C, Y_p) and m_0 is a reference value for the mediator. Following (Tchetgen Tchetgen, 2013), we consider $m_0 = 0$ for the binary mediator. This equation follows from the invariance property of the odds ratio, and it motivates the strategy for modeling the exposure rather than modeling the mediator. By definition, the odds ratio is the parametric model $\mathrm{OR}\left(M, A\middle|C, y_p = 1, \boldsymbol{\alpha}, \boldsymbol{\kappa}\right)$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$ are the parameters of *A* given (M, C, Y_p) and Y_p respectively, and the estimators $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\kappa}}$ are derived using the ML estimation. Subsequently, the IORW estimating equation for any $\boldsymbol{\nu}$ is defined as follows:

$$U_{DE}(\boldsymbol{\nu}) = \mathrm{OR}\left(M = m, A = a\middle|C, y_p = 1, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\kappa}}\right)^{-1} \times \Gamma_{NDE}(a, c, \boldsymbol{\nu})\{Y - g(\mu_{NDE}(\boldsymbol{\nu}; a, c))\},$$

where $\Gamma_{NDE}(a, c, \boldsymbol{\nu}) = \partial g(\mu_{NDE}(\boldsymbol{\nu}; a, c))/\partial\boldsymbol{\nu}$. *Theorem 6* states that $U_{NDE}(\boldsymbol{\nu})$ is unbiased.

Theorem 6

Under the consistency assumption and the assumptions in Section 2, $U_{NDE}(\boldsymbol{\nu})$ *is unbiased when conditioned on* $C, Y_p = 1$ *(i.e.,* $E_{Y,M|C,Y_p=1}\left(U_{NDE}(\boldsymbol{\nu})\right) = 0$*).*

The proof of *Theorem 6* is provided in Web Appendix C. The estimator $\hat{\boldsymbol{\nu}}$ of $\boldsymbol{\nu}$ then solves the equation $\mathbb{P}_n[U_{NDE}(\hat{\boldsymbol{\nu}})] = 0$; thus, NDE_{dt} can be estimated through the following formula.

$$\widehat{NDE}_{dt}^{IORW} = \hat{\phi}(1,0) - \hat{\phi}(0,0) = E_C\left(\hat{\phi}(1,0|\boldsymbol{C}) - \hat{\phi}(0,0|\boldsymbol{C})\right)$$

$$= E_C\left(g(\mu_{NDE}(\hat{\boldsymbol{\nu}}; 1, C))f_{Y_p}\left(y_p = 1\middle|1, C, \hat{\boldsymbol{\alpha}}\right) - g(\mu_{NDE}(\hat{\boldsymbol{\nu}}; 0, C))f_{Y_p}\left(y_p = 1\middle|0, C, \hat{\boldsymbol{\alpha}}\right)\right).$$

Finally, $\widehat{NIE}_{dt}$ can be estimated by $\widehat{TE}_{IORW} - \widehat{NDE}_{dt}^{IORW}$.

4.2. Statistical inference for $\boldsymbol{Q_T(a, a^*)}$ by using Cox proportional hazards model

In this section, we adopt a Cox proportional hazards model and proposed a statistical method (hereafter referred to as the Cox model method) to infer the survival mediation formula presented in *Lemma 1*. Following the regression-based method described in Section 4.1.1, this method further requires distribution assumptions for M and Y_p. The complete model settings for the Cox model are listed in Table 2. In practice, the parameters of the log hazard model in Table 2 ($\boldsymbol{\gamma} = \{\gamma_0, \gamma_A, \gamma_M, \gamma_C\}$) can be estimated as $\hat{\boldsymbol{\gamma}}$ by using the partial likelihood method, and the parameters of M and Y_p are estimated as $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$, respectively, by using the ML approach. If the outcome of interest is relatively rare—the rare disease assumption—then the cumulated hazard $\Lambda(t|a, m, c, y_p = 1)$ is approximately zero. Therefore, $e^{-\Lambda(t|a,m,c,y_p=1)} \approx 0$, and the survival mediation formula can be approximated in terms of survival time $Q_T(a, a^*)$, as described in *Lemma 1*, by $log(\vartheta_T^a(a, a^*)/\vartheta_T^b(a, a^*))$, where

$$\vartheta_T^a(a, a^*) = \int_{m,c} \lambda(t|a, m, c, y_p = 1)\ f(Y_p = 1|A = a, C = c) \times$$
$$f_{M|A,C}(M = m|A = a^*, C = c)f(c)\ dmdc$$

and $\vartheta_T^b(a, a^*) = \int_{m,c}\ f(Y_p = 1|A = a, C = c)f_{M|A,C}(M = m|A = a^*, C = c)f(c)\, dmdc$.

Cho and Huang (2019) suggested that this approximation requires a cumulative disease rate of <10%.

NDE_{dt} and NIE_{dt} are estimated with respect to T as NDE_{dt}^T and NIE_{dt}^T, respectively, by adopting a substitution estimation approach. We obtain $\hat{Q}_T(a, a^*) \approx log(\hat{\vartheta}_T^a(a, a^*)/$

$\hat{\vartheta}_T^b(a,a^*)$) by substituting $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\gamma}}$ into $log(\vartheta_T^a(a,a^*)/\vartheta_T^b(a,a^*))$. Integration with respect to m and c in $\vartheta_T^a(a,a^*)$ and $\vartheta_T^b(a,a^*)$ can be approximated by using Monte Carlo integration. This results in the following estimators:

$$\widehat{NDE}_{dt}^T = \hat{Q}_T(1,0) - \hat{Q}_T(0,0) \text{ and } \widehat{NIE}_{dt}^T = \hat{Q}_T(1,1) - \hat{Q}_T(1,0).$$

When there is a generalized linear model for the mediator, the Cox model method can be performed using G-computation.

5. Simulation

Two simulation studies were conducted to assess the empirical biases and standard errors of the proposed estimators. In study 1, we assessed the regression-based, IP, and IORW methods in terms of binary survival status. In study 2, we evaluated the Cox model method in the survival setting. The details of study 2 are provided in Web Appendix D. In both studies, we also applied the complete case approach to the simulated data for comparison.

5.1. Three scenarios

Study 1 employed a sample of size 10,000 with one binary mediator. The data for each variable were simulated as follows:

$C \sim Bernoulli(p = 0.5)$

$\mathrm{A} \sim Bernoulli(p = 0.5)$

$Y_p \sim\ Bernoulli(\,p_1),\ p_1 = expit(\alpha_0 + \alpha_A A + \alpha_C C)$

$$M = \begin{cases} undefined,\ if\ Y_p = 0 \\ \sim Bernoulli(\,p_2),\ if\ Y_p = 1 \end{cases},\ p_2 = expit(\beta_0 + \beta_A A + \beta_C \mathrm{C}\,)$$

$$Y = \begin{cases} 0 \quad ,if\ Y_p = 0 \\ \sim Bernoulli(\,p_3),\ if\ Y_p = 1 \end{cases},\ p_3 = expit(\theta_0 + \theta_A A + \theta_M M + \theta_C \mathrm{C}\,)$$

The term $expit$ refers to the expit function, defined as expit(x) = 1/(1 + exp(−x)). The parameters were set as $\theta_A = 0.2$, $\theta_M = 0.5$, $\theta_C = 0.5$, $\beta_A = 0.2$, $\beta_C = 0.5$, $\alpha_A = 0.2$, and $\alpha_C = 0.5$. The values of β_0 and θ_0 were determined by solving $\mathrm{E}(M|\,Y_p = 1) = 0.4$ and $\mathrm{E}(Y|\,Y_p = 1) = 0.3$, which is an empirical setting. The remaining parameter α_0 was determined according to $\mathrm{P}(\,Y_p = 1)$. To investigate the model performance under various death-truncation rates (i.e., $1 - \mathrm{P}(\,Y_p = 1)$), we varied $\mathrm{P}(\,Y_p = 1)$ from 0.1 to 0.9 at increments of 0.1. Consequently, the data were simulated based on these parameter settings.

To evaluate the robustness of each method, we further considered three scenarios pertaining to model specification. In scenario 1, all models were correctly specified. By contrast, in scenarios 2 and 3, the models of Y and M, respectively, were misspecified. For each scenario, we performed 1000 repetitions, and we then calculated the bias, root empirical standard error (RESE), and root mean square error (RMSE) of estimates. Because the true values of estimators varied for the different probabilities of $Y_p = 1$, we used normalized absolute bias, normalized RESE, and normalized RMSE, which were divided by $P(Y_p = 1)$, to enable a fair comparison.

5.2 Result

The results of scenario 1 are illustrated in Figure 2. The estimations of NIE and NDE under the complete case approach were biased for a death-truncated mediator. By contrast, the three proposed methods precisely estimated all the causal effects. In scenario 2, the results of the regression-based, IPW, and IORW methods are presented in Figure 3(a) and Web Appendix D. The complete case approach was excluded from comparison in scenarios 2 and 3 because it produced biased results. Figure 3(a) shows the normalized biases of the estimators for the survival mediation formulas for various probabilities of $Y_p = 1$. The figure reveals that the regression-based method was more affected by the model misspecification of outcome Y than the other methods were. Subsequently, we explored the model performance when the model of the mediator M was misspecified. The results of scenario 3 are shown in Figure 3(b) and Web Appendix D. These results indicated that the regression-based and IPW methods failed to accurately estimate NIE, but the IPW method provided unbiased estimators for NDE and TE. This reflected the fact that the misspecification of model M only affected the estimation of NIE. Results for these three scenarios demonstrated the robustness of the IORW method.

6. Application

We applied the proposed methods to the motivating example, namely the REVEAL-HBV study (Chen et al., 2006). We were interested in the role that HBV viral loads play in the mechanism of the effect of HCV viral loads on mortality. Although HBV and HCV infections are two main causes of death among patients with liver disease, HCV is known to inhibit HBV replication in patients with HBV/HCV coinfection. Therefore, to investigate the causal mechanism, we conducted a survival mediation analysis where HCV and HBV were treated as the exposure and mediator, respectively. Because the follow-up of HBV viral loads for some patients with HBV were truncated in the data set due to death, we applied the proposed methods to infer the causal effects for the binary survival status and the survival time. For comparison, we also calculated the results by using the complete case approach. Sex, alcohol consumption, and age were considered as confounders in the analysis. Additional descriptions of the data and preprocessing methods are provided in Web Appendix E.

The results are listed in Table 3. We further adopted the bootstrapping method to calculate the 95% confidence intervals and the P values based on 1000 bootstrap samples. According to the results for all methods, the NDE and NIE estimates had opposite directions. Therefore, these methods reproduced the inhibition of HBV replication by HCV. However, for the binary survival status, the complete case approach exhibited less power to detect causal effects than the proposed methods did, especially for NDE. The complete case approach excluded the cases with a death-truncated mediator, and the excluded cases still contributed to the inference of NDE. Additionally, the results for the binary survival status revealed consistent estimations among the regression-based, IPW, and IORW methods. In the analysis of survival time, the results of the Cox model method were somewhat consistent with those of the complete case approach.

7. Discussion

In this study, we comprehensively illustrated the incompleteness of the conventional causal definitions in the presence of a death-truncated mediator and a survival outcome. We then proposed three approaches to redefine these causal definitions for completeness with a truncated mediator. In the establishment of new causal definitions, we noted a difference between the survival and nonsurvival settings in the conventional causal definitions. Moreover, we proved that the casual effects defined in the third approach are a generalization of the causal effects in the conventional definitions. Based on the proposed causal definitions, we developed three statistical methods for the binary survival status, in addition to a Cox model method for survival time. The regression-based method was limited by the problem of model misspecification, but it exhibited flexibility when computing the estimates through G-computation. The simulation study revealed that the IPW and IORW methods were more robust in model specification. However, extending these methods to cases with multiple mediators is challenging.

The proposed methods have three limitations that should be addressed in future studies. First, this study focused on survival outcomes. Although several methods for nonsurvival outcomes have been developed, they are applicable only to TE analysis. Methods for causal mediation analysis have not been developed for nonsurvival outcomes. Second, only one mediator can be used in the proposed method. However, because mechanisms are unlikely to be explained by a single mediator, methods allowing for multiple mediators should be developed. Third, this study's methods are restricted to time-invariant mediators. Future studies could extend the methods to permit time-to-event mediators, which can be competed by survival outcomes or even time-to-event exposures.

Acknowledgments

We thank Professor Hwai-I Yang (Genomics Research Center, Academia Sinica, Taipei, Taiwan) for suggesting revisions to refine the model and providing the example data set. We also thank Novi for the initial version of the regression-based method. This study was supported by a grant from the Ministry of Science and Technology in Taiwan (No. 109-2636-B-009 -001). This manuscript was edited by Wallace Academic Editing.

Reference

Avin, C., Shpitser, I., and Pearl, J. (2005). Identifiability of path-specific effects. *Department* *of Statistics, UCLA*.

Chen, C.-J., Yang, H.-I., Su, J.*, et al.* (2006). Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *Jama* **295**, 65-73.

Chiba, Y., and VanderWeele, T. J. (2011). A simple method for principal strata effects when the outcome has been truncated due to death. *American journal of epidemiology* **173**, 745-751.

Cho, S. H., and Huang, Y. T. (2019). Mediation analysis with causally ordered mediators using Cox proportional hazards model. *Statistics in medicine* **38**, 1566-1581.

Daniel, R., De Stavola, B., Cousens, S., and Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics* **71**, 1-14.

Ding, P., Geng, Z., Yan, W., and Zhou, X.-H. (2011). Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association* **106**, 1578-1591.

Egleston, B. L., Scharfstein, D. O., Freeman, E. E., and West, S. K. (2006). Causal inference for non-mortality outcomes in the presence of death. *Biostatistics* **8**, 526-545.

Fasanelli, F., Giraudo, M. T., Ricceri, F., Valeri, L., and Zugna, D. (2019). Marginal Time-Dependent Causal Effects in Mediation Analysis With Survival Data. *American journal of epidemiology* **188**, 967-974.

Forastiere, L., Mattei, A., and Ding, P. (2018). Principal ignorability in mediation analysis: through and beyond sequential ignorability. *Biometrika* **105**, 979-986.

Frangakis, C. E., and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21-29.

Gibbard, A., and Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In *Ifs*, 153-190: Springer.

Gilbert, P. B., Bosch, R. J., and Hudgens, M. G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics* **59**, 531-541.

Huang, Y.-T., and Yang, H.-I. (2017). Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology* **28**, 370-378.

Huang, Y. T., and Cai, T. (2015). Mediation analysis for survival data using semiparametric probit models. *Biometrics*.

Lange, T., and Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology* **22**, 575-581.

Lin, S. H., Young, J., Logan, R., Tchetgen Tchetgen, E. J., and VanderWeele, T. J. (2017a). Parametric Mediational g-Formula Approach to Mediation Analysis with Time-varying Exposures, Mediators, and Confounders. *Epidemiology* **28**, 266-274.

Lin, S. H., Young, J. G., Logan, R., and VanderWeele, T. J. (2017b). Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. *Stat Med* **36**, 4153-4166.

Little, R. J. (1992). Regression with missing X's: a review. *Journal of the American Statistical Association* **87**, 1227-1237.

Little, R. J., and Rubin, D. B. (2019). *Statistical analysis with missing data*: John Wiley & Sons.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, 411-420. San Francisco, CA, USA: Morgan kaufmann publishers Inc.

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.

Robins, J. M., and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143-155.

Tchetgen Tchetgen, E. J. (2011). On causal mediation analysis with a survival outcome. *The international journal of biostatistics* **7**, 1-38.

Tchetgen Tchetgen, E. J. (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistics in medicine* **32**, 4567-4580.

Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in medicine* **33**, 3601-3628.

VanderWeele, T., and Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface* **2**, 457-468.

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology (Cambridge, Mass.)* **22**, 582.

VanderWeele, T. J., and Tchetgen Tchetgen, E. (2017). Mediation Analysis with Time-Varying Exposures and Mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

VanderWeele, T. J., and Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American journal of epidemiology* **172**, 1339-1348.

VanderWeele, T. J., and Vansteelandt, S. (2014). Mediation Analysis with Multiple Mediators. *Epidemiol Method* **2**, 95-115.

Vanderweele, T. J., Vansteelandt, S., and Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology* **25**, 300-306.

Wang, L., Zhou, X.-H., and Richardson, T. S. (2017). Identification and estimation of causal effects with outcomes truncated by death. *Biometrika* **104**, 597-612.

Zhang, J. L., and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics* **28**, 353-368.

Zheng, W., and van der Laan, M. (2017). Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes. *Journal of Causal Inference* **5**, 20160006.

Zheng, W., and van der Laan, M. J. (2012). Causal mediation in a survival setting with time-dependent mediators.

Table 1. Definition statuses of total effect, natural direct effect, natural indirect effect, death-truncated natural direct effect, and death-truncated natural indirect effect for the four survival types.

Survival type (*Description*)	$Y_p(1)$	$Y_p(0)$	$M(1)$	$M(0)$	TE $Y(1,M(1))$ $-Y(0,M(0))$	NDE $Y(1,M(0))$ $-Y(0,M(0))$	NIE $Y(1,M(1))$ $-Y(1,M(0))$	NDE_{dt} $Y(1,M(0,Y_p(1)))$ $-Y(0,M(0,Y_p(0)))$	NIE_{dt} $Y(1,M(1,Y_p(1)))$ $-Y(1,M(0,Y_p(1)))$
Always-survivor (*The subject always survives, regardless of exposure status*)	1	1	*Well-defined*	*Well-defined*	*Well-defined*	*Well-defined*	*Well-defined*	*Well-defined*	*Well-defined*
Protected (*The subject survives if exposed, but dies if not exposed*)	1	0	*Well-defined*	*Undefined*	*Well-defined*	*Undefined*	*Undefined*	*Well-defined*	*Well-defined*
Harmed (*The subject dies if exposed, but survives if not exposed*)	0	1	*Undefined*	*Well-defined*	*Well-defined*	*Well-defined*	*Well-defined* (= 0)	*Well-defined*	*Well-defined* (= 0)
Doomed (*The subject always dies, regardless of exposure status*)	0	0	*Undefined*	*Undefined*	*Well-defined* (= 0)	*Well-defined* (= 0)	*Well-defined* (= 0)	*Well-defined* (= 0)	*Well-defined* (= 0)

TE: total effect; NDE: natural direct effect; NIE: natural indirect effect; NDE_{dt}: death-truncated NDE; NIE_{dt}: death-truncated NIE

Table 2. Model assumptions for the proposed methods.

Type of outcome	Proposed method	Model assumptions
Binary survival status (Y)	*Regression -based*	**Survival status** Y: $Y \vert A, M, C, Y_p = 1 \sim Bernoulli(\pi_y)$ where $logit(\pi_y) = \theta_0 + \theta_A A + \theta_M M + \theta_C C$, **Mediator** M: $M \vert A, C, Y_p = 1 \sim Bernoulli(\pi_M)$ where $logit(\pi_M) = \beta_0 + \beta_A A + \beta_C C$, and **Previous survival status** Y_p: $M \vert A, C \sim Bernoulli(\pi_{Y_p})$ where $logit(\pi_{Yp}) = \alpha_0 + \alpha_A A + \alpha_C C$.
	IPW	**Mediator** M: $M \vert A, C, Y_p = 1 \sim Bernoulli(\pi_M)$ where $logit(\pi_M) = \beta_0 + \beta_A A + \beta_C C$ and **Exposure** A: $A \vert C \sim Bernoulli(\pi_A)$ where $logit(\pi_A) = \delta_0 + \delta_C C$.
	IORW	**Previous survival status** Y_p: $M \vert A, C \sim Bernoulli(\pi_{Y_p})$ where $logit(\pi_{Yp}) = \alpha_0 + \alpha_A A + \alpha_C C$ and **Exposure** A: $A \vert M, C, Y_p = 1 \sim Bernoulli(\pi_A^{IORW})$ where $logit(\pi_A^{IORW}) = \kappa_0 + \kappa_M M + \kappa_A C$.
Survival time (T)	*Cox model*	**Survival model** T: $\log\left(\lambda_Y(t \vert A, M, C, Y_p = 1\ )\right) = log(\lambda_0(t)) + \gamma_0 + \gamma_A A + \gamma_M M + \gamma_C C$ where $\lambda_0(t)$ is the baseline hazard, **Mediator** M: $M \vert A, C, Y_p = 1 \sim Bernoulli(\pi_M)$ where $logit(\pi_M) = \beta_0 + \beta_A A + \beta_C C$, and **Previous survival status** Y_p: $M \vert A, C \sim Bernoulli(\pi_{Y_p})$ where $logit(\pi_{Yp}) = \alpha_0 + \alpha_A A + \alpha_C C$.

Table 3. Mechanism of the effect of hepatitis C virus infection on mortality, mediated through hepatitis B viral load.

Type of outcome	Method	Causal effects	Estimate	95% CI Lower Bound	95% CI Upper Bound	p-value
Binary survival status (Y) in risk difference scale	*Complete Case*	TE	-0.018	-0.057	0.020	0.356
		NDE	-0.029	-0.072	0.014	0.185
		NIE	0.011	0.002	0.020	0.022
	Regression-based	TE	-0.074	-0.078	-0.070	<0.001
		NDE	-0.083	-0.087	-0.079	<0.001
		NIE	0.009	0.004	0.014	<0.001
	IPW	TE	-0.071	-0.132	-0.009	0.025
		NDE	-0.093	-0.160	-0.025	0.008
		NIE	0.022	-0.006	0.050	0.129
	IORW	TE	-0.059	-0.104	-0.013	0.011
		NDE	-0.064	-0.109	-0.019	0.006
		NIE	0.005	0.005	0.006	<0.001
Survival time (T) in hazard ratio scale	*Complete Case*	TE	1.253	0.576	1.931	0.463
		NDE	1.491	0.695	2.287	0.226
		NIE	0.841	0.745	0.936	0.001
	Cox model	TE	1.154	0.494	1.814	0.647
		NDE	1.408	0.633	2.184	0.302
		NIE	0.820	0.723	0.916	<0.001

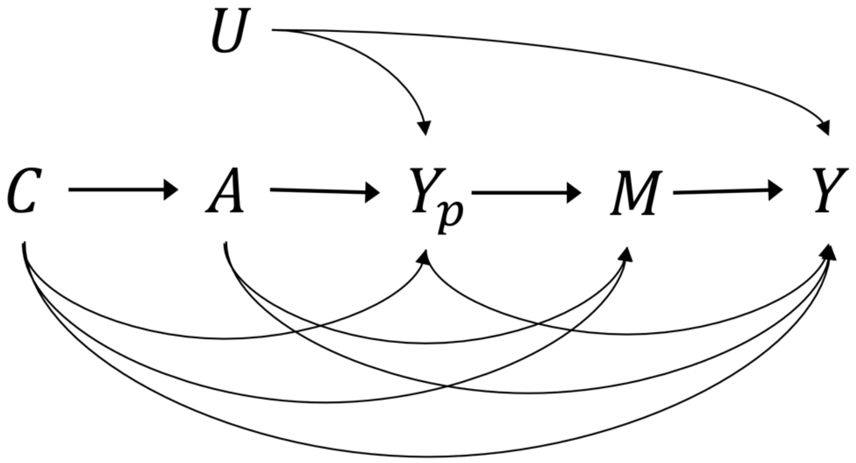

Figure 1. Direct acyclic graph of causal relationships between variables. A, M, Y_p, Y, and C denote the exposure, the mediator, the survival indicator between A and M, the survival outcome, and the baseline confounders, respectively. Y represents the survival status at the end of study. U is the unmeasured confounder between Y_p and Y. Y can be replaced by survival time T, if available.

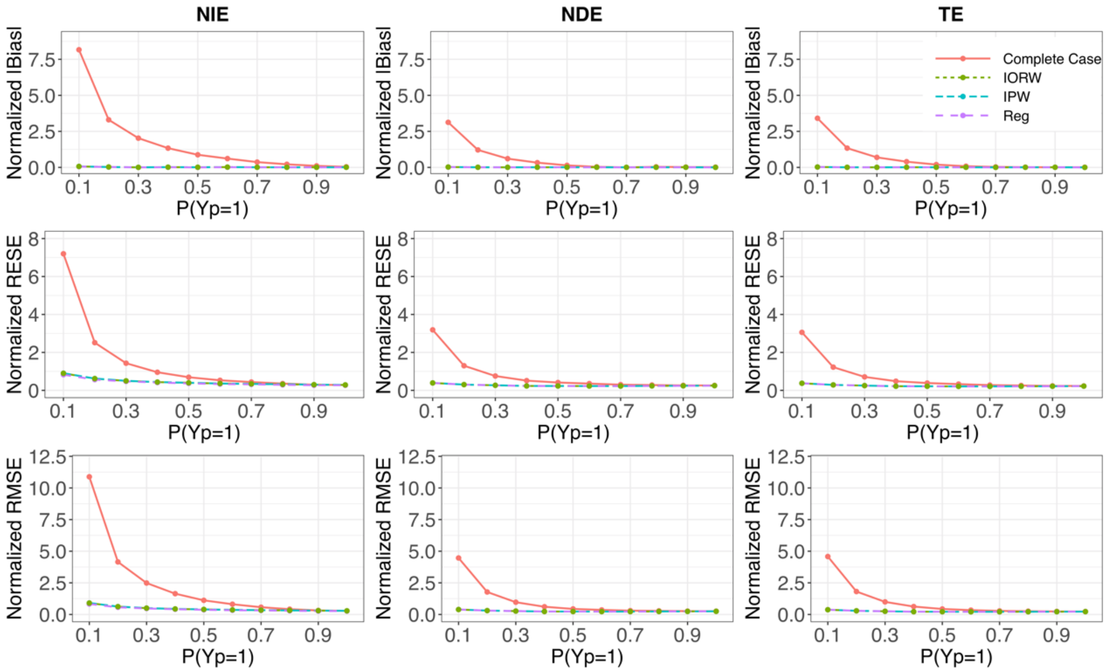

Figure 2. Performance evaluation for the methods under scenario 1. Rows represent measurements, and columns represent causal effects. The *x* axis of each plot represents the probability of $Y_p = 1$, and the *y* axis represents the quantity of measurements. The complete case approach, IORW, IPW, and Reg methods are indicated by red, green, blue, and purple lines, respectively. Abbreviations: NIE, nature indirect effect; NDE, nature direct effect; TE, total effect; RESE, root empirical standard error; RMSE, root mean square error; IORW, inverse odds ratio weighting; IPW, inverse probability weighting; Reg, regression-based method.

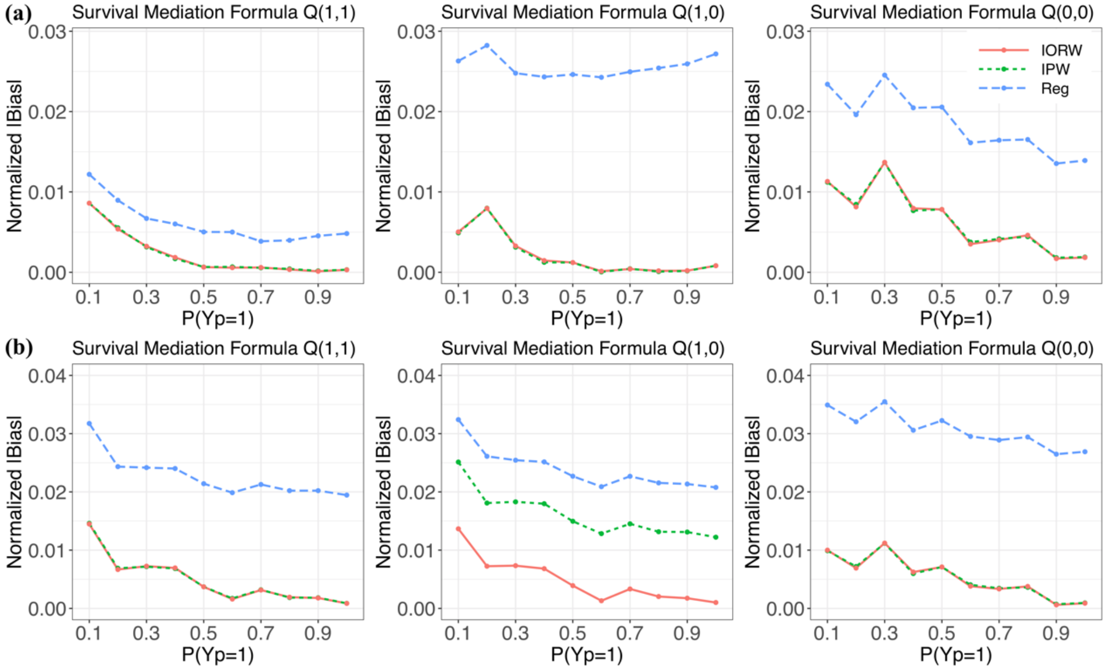

Figure 3. Absolute values of the normalized biases for the three methods for (a) scenario 2 and (b) scenario 3. The *x* axis of each plot represents the probability of $Y_p = 1$, and the *y* axis represents the absolute value of the normalized biases. The IORW, IPW, and Reg methods are indicated by red, green, and blue lines, respectively. Abbreviations: IORW, inverse odds ratio weighting; IPW, inverse probability weighting.