Revista Colombiana de Estadística - Applied Statistics
January 2022, volume 45, issue 1, pp. 161–191
http://doi.org/10.15446/rce.v45n1.94553

Causal Mediation for Survival Data: A Unifying Approach via GLM

Mediación causal para datos de supervivencia: un enfoque unificador a través de GLM

Marcelo M. Taddeo[a], Leila D. Amorim[b]

Departmento de Estatística, Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, Brazil

Abstract

Mediation analysis has been receiving much attention from the scientific community in the last years, mainly due to its ability to disentangle causal pathways from exposures to outcomes. Particularly, causal mediation analysis for time-to-event outcomes has been widely discussed using accelerated failures times, Cox and Aalen models, with continuous or binary mediator. We derive general expressions for the Natural Direct Effect and Natural Indirect Effect for the time-to-event outcome when the mediator is modeled using generalized linear models, which includes existing procedures as particular cases. We also define a responsiveness measure to assess the variations in continuous exposures in the presence of mediation. We consider a community-based prospective cohort study that investigates the mediation of hepatitis B in the relationship between hepatitis C and liver cancer. We fit different models as well as distinct distributions and link functions associated to the mediator. We also notice that estimation of NDE and NIE using different models leads to non-contradictory conclusions despite their effect scales. The survival models provide a compelling framework that is appropriate to answer many research questions involving causal mediation analysis. The extensions through GLMs for the mediator may encompass a broad field of medical research, allowing the often necessary control for confounding.

***Key words*:** causal inference; generalized linear models; mediation; survival analysis.

Resumen

[a]Ph.D. E-mail: marcelo.magalhaes@ufba.br
[b]Ph.D. E-mail: leiladen@ufba.br

El análisis de mediación ha recibido mucha atención en los últimos años, principalmente debido a su capacidad para desenredar las vías causales. Particularmente, mediación causal para el tiempo hasta el evento se ha discutido ampliamente utilizando tiempos de falla acelerados, modelos de Cox y Aalen, con mediador continuo o binario. Derivamos expresiones generales para el efecto directo natural y el efecto indirecto natural para el el tiempo hasta el evento cuando el mediador se modela utilizando modelos lineales generalizados, que incluyen procedimientos existentes como casos particulares. Definimos una medida para evaluar variaciones en exposiciones continuas en presencia de mediación. Consideramos un estudio de cohorte prospectivo que investiga la mediación de la hepatitis B en la relación entre la hepatitis C y el cáncer de hígado. Encajamos diferentes modelos, así como distintas distribuciones y funciones de enlace. Todos los enfoques dan como resultado evaluaciones consistentes de los effectos considerando sus correspondientes escalas. Los modelos de supervivencia proporcionan un marco convincente apropiado para responder a muchas preguntas de investigación que involucran mediación causal. Las extensiones a través de GLM para el mediador pueden abarcar un amplio campo de investigación médica, lo que permite el control necesario para los factores de confusión.

***Palabras clave*:** análisis de supervivencia; inferencia causal; mediación; modelos lineales generalizados.

1. Introduction

Mediation analysis in the context of causal inference has been receiving much attention from the scientific community in the last years, especially in Epidemiology and Social Sciences, see Imai et al. (2010) and VanderWeele (2016). This is due to its ability to disentangle causal pathways from exposures/treatments to outcomes of interest, which results in a greater understanding of the underlying mechanism. In short, a mediator may be defined as a variable M "between" the exposure A and the outcome T, so that the effect of A on T splits into two parts: one transmitted directly from the exposure to the outcome and another transmitted indirectly through the mediator. Mediation models were initially proposed for continuous outcomes and mediators in cross sectional studies, see Baron & Kenny (1986) and MacKinnon (2008). Particularly, in the specific context of causal inference, formal definitions of decomposed effects into direct (without mediation) and indirect (with mediation) effects are found in Pearl (2001) and their estimation is discussed in Tchetgen Tchetgen & Shpitser (2012) among others. However, more recent advances have extended them to more complex models including survival outcomes, see Fulcher et al. (2017), Iacobucci (2012), Lange & Hansen (2011), Loyes et al. (2013), VanderWeele & Vansteelandt (2010), VanderWeele (2011) Vansteelandt et al. (2019); and allowing, among other features, multiple mediators, see Daniel et al. (2015), Fasanelli et al. (2019), Huang & Yang (2017); or time-dependent treatments and/or mediators, see Aalen et al. (2012), Aalen et al. (2020), Didelez (2019) and Lin et al. (2017).

Causal mediation analysis for survival outcomes has been widely discussed under accelerated failure times (AFT), Cox proportional hazards and Aalen additive hazards models, with or without an interaction between exposure and mediator, see Lange & Hansen (2011) and VanderWeele (2015). The usual approach assumes a binary exposure A and a linear or logistic regression model for the mediator M, see VanderWeele (2011) and VanderWeele (2015). Thus, there are still limitations when M does not follow linear or logistic models. In this paper, M is assumed to follow a generalized linear model (GLM), which includes the linear and logistic approaches as particular cases. We illustrate the proposed methodology by considering a community-based prospective cohort study to investigate the mediation of hepatitis B in the relatioship between hepatitis C and liver cancer. A mediation model for this data, based on the standard linear model, was proposed by Huang & Yang (2017) based on the existing scientific evidence. We reevaluate the mediation model considering distinct distributions and link functions for the mediator, jointly with varying models for the survival outcome (AFT, Cox and Aalen).

The article is organized as follows: in section 2 we present a brief review of causal mediation analysis for survival data. In section 3, we formally show general expressions for the Natural Direct Effect (NDE) and Natural Indirect Effect (NIE) for survival outcomes using AFT, Cox and Aalen models. We also specify the expressions for NDE and NIE for particular distributions. In section 4, we define responsiveness measures for the promptness of the outcome for changes in the continuous treatment levels. We apply this methodology for the mediation analysis of the liver cancer data to illustrate model specification and causal interpretation in section 5. Finally, methodology challenges and advances are discussed in section 6.

2. Mediation for Survival Data

2.1. Framework and Notation

Let T be the time to event and denote by T_a the potential response corresponding to treatment a, *i.e.* the time to event it would be observed if the treatment had been set at level a (regardless of the actual value of A). For binary treatments, we have T (the actual outcome), T_a and T_{a^*} (the potential outcomes had A been set equal to a and a^*, respectively). At the unit level, the causal effect would ideally be assessed, for each unit u, by the difference $T_a(u) - T_{a^*}(u)$. Since one of them is necessarily counterfactual, it is impossible to evaluate the causal effect by such difference. The solution usually adopted considers the expected differences $\mathrm{E}[T_a - T_{a^*}]$ or $\mathrm{E}[T_a - T_{a^*}|\boldsymbol{x}]$, where $\boldsymbol{x}$ is a vector of covariates representing the confounding between A and T.

In mediation analysis, the idea is to decompose the treatment effect into two parts corresponding to the paths $A \longrightarrow T$ and $A \longrightarrow M \longrightarrow T$. In doing so, it is usual to consider the potential responses $T_{aM_{a^*}}$, where possibly different treatments (a and a^*) are allowed to directly affect the response and the mediator.

The effect associated to the first path is called *Natural Direct Effect* and the one associated to the second path, *Natural Indirect Effect*. In the difference scale, they are respectively defined as

$$\text{NDE} = \text{E}[T_{aM_{a^*}} - T_{a^*M_{a^*}}] \quad \text{and} \quad \text{NIE} = \text{E}[T_{aM_a} - T_{aM_{a^*}}]. \tag{1}$$

Assuming the so-called *composition assumption* (VanderWeele & Vansteelandt, 2009): $T_a = T_{aM_a}$, referred to as (A0), we can write the *Total Effect* as $\text{TE} \equiv \text{E}[T_a - T_{a^*}] = \text{E}[T_{aM_a} - T_{a^*M_{a^*}}] = \text{NIE} + \text{NDE}$, regardless of the functional relation between the variables involved. Historically, before this counterfactual notation, there were two traditional approaches to estimate indirect effects known as "difference method" (considering two models for the outcome: with and without the mediator) and "product method" (defined by the product of the models' coefficients). They converge for the same results when the outcome and the mediator are continuous and a linear model is used, see Baron & Kenny (1986) and MacKinnon & Dwyer (1993). However, they diverge in many other situations, such as when analyzing binary outcomes or using the Cox model, see MacKinnon (2008) and VanderWeele (2015). Therefore, the validity of such approach (not based in counterfactual arguments) depends on the chosen model (MacKinnon et al., 1995). Unlike them, the counterfactual approach is a solution to the problem of establishing consistent definitions for direct, indirect and total effects.

2.2. Survival Models

For time-to-event data, the total effect can be decomposed in terms of natural direct and indirect effects through different scales depending on the chosen model, see VanderWeele (2011) and VanderWeele (2015). Throughout this paper we will focus on the following three models (Kalbfleisch & Prentice, 2002):

1. the *accelerated failure time (AFT) model*, in which

$$\log T = \beta_0 + \beta_1 a + \beta_2 m + \beta_3 am + \beta_4^\top \boldsymbol{x} + \gamma\varepsilon,$$

   with ε following an extreme value distribution and $\gamma > 0$, the scale parameter for the Weibull distribution. In this case, the natural direct and indirect effects are given by $\text{NDE}_{\text{aft}} = \log \text{E}[T_{aM_{a^*}}|\boldsymbol{x}] - \log \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}]$ and $\text{NIE}_{\text{aft}} = \log \text{E}[T_{aM_a}|\boldsymbol{x}] - \log \text{E}[T_{aM_{a^*}}|\boldsymbol{x}]$.

2. the *Aalen additive hazards model* (Aalen et al., 2008), in which

$$\lambda_T(t|a, m, \boldsymbol{x}) = \beta_0(t) + \beta_1(t)a + \beta_2(t)m + \beta_3(t)am + \beta_4(t)^\top \boldsymbol{x}, \tag{2}$$

   whose coefficients are typically time-dependent. In this case,

$$\text{NDE}_{\text{Aalen}}(t|\boldsymbol{x}) = \lambda_{T_{aM_{a^*}}}(t|\boldsymbol{x}) - \lambda_{T_{a^*M_{a^*}}}(t|\boldsymbol{x})$$

   and

$$\text{NIE}_{\text{Aalen}}(t|\boldsymbol{x}) = \lambda_{T_{aM_a}}(t|\boldsymbol{x}) - \lambda_{T_{aM_{a^*}}}(t|\boldsymbol{x}).$$

Another way to report these (time-dependent) effects is through their cumulative versions, which are given by $\text{CIE}_{\text{Aalen}}(t) = \int_0^t \text{NIE}_{\text{Aalen}}(s)ds$, $\text{CDE}_{\text{Aalen}}(t) = \int_0^t \text{NDE}_{\text{Aalen}}(s)ds$ and $CE_{\text{Aalen}}(t) = \int_0^t \text{TE}_{\text{Aalen}}(s)ds = \text{CIE}_{\text{Aalen}}(t) + \text{CDE}_{\text{Aalen}}(t)$.

3. the *Cox proportional hazards model*, where

$$\lambda_T(t|a, m, \boldsymbol{x}) = \lambda_o(t) \exp\left\{\beta_1 a + \beta_2 m + \beta_3 am + \beta_4^\top \boldsymbol{x}\right\},$$

and $\lambda_o(\cdot)$ is an unspecified baseline hazard function. Here,

$$\text{NDE}_{\text{Cox}}(t|\boldsymbol{x}) = \log \lambda_{T_{aM_{a^*}}}(t|\boldsymbol{x}) - \log \lambda_{T_{a^*M_{a^*}}}(t|\boldsymbol{x})$$

and

$$\text{NIE}_{\text{Cox}}(t|\boldsymbol{x}) = \log \lambda_{T_{aM_a}}(t|\boldsymbol{x}) - \log \lambda_{T_{aM_{a^*}}}(t|\boldsymbol{x}).$$

Accordingly, the exponentiated NDE and NIE stand for the ratio of expected counterfactual survival times in the AFT model and the counterfactual hazard ratio in the Cox model. Choosing any model is always related to different issues. For instance, the results of the Cox model are very easily interpreted. The Aalen model is particularly useful when the assumption of the proportionality of risks is not met. In addition, its coefficients may be time varying, which makes it more flexible than the standard Cox model. On the other hand, fully parametric models, such as the AFT model, are more efficient than semiparametric approaches when correctly specified.

An important aspect of causal inference is the determination of the necessary assumptions for identification of the effects of interest as well as their expressions in terms of the parameters of the models. It is well known that such assumptions are more stringent in the presence of mediators. Following Lange & Hansen (2011), we have the assumptions of no unmeasured counfounding between (A1) exposure and outcome: $A \perp\!\!\!\perp T_{am}|\boldsymbol{x}$, (A2) mediator and outcome: $M \perp\!\!\!\perp T_{am}|a, \boldsymbol{x}$, (A3) exposure and mediator: $A \perp\!\!\!\perp M_a|\boldsymbol{x}$; and the identifiability condition (A4) $M_{a^*} \perp\!\!\!\perp T_{am}|\boldsymbol{x}$. Under these conditions and the composition assumption (A0) together with the Aalen additive model (without the interaction term) and a Normal mediator, we have $\text{TE}_{\text{Aalen}} = [\beta_2(t)\zeta_1(a - a^*)] + \beta_1(t)(a - a^*) = \text{NIE}_{\text{Aalen}} + \text{NDE}_{\text{Aalen}}$. Similarly, but allowing interaction ($\beta_3 am$), VanderWeele (2011) determined the identification formulas for the AFT and Cox models. For instance, $\text{NDE}_{\text{Cox}} = \exp\{(\beta_1 + \beta_3(\zeta_0 + \zeta_1 a^* + \zeta_2^\top \boldsymbol{x} + \beta_2\sigma^2))(a - a^*) + 0.5\beta_3^2\sigma^2(a^2 - a^{*2})\}$ and $\text{NIE}_{\text{Cox}} = \exp\{(\beta_2 + \beta_3 a)\zeta_1(a - a^*)\}$. Similar expressions are found for the AFT model even when M is dichotomous. We refer the interested reader to VanderWeele & Vansteelandt (2010) or VanderWeele (2015) for more details.

In any case, the mediator is usually assumed to be either conditionally Normal with mean $\text{E}[M|a, \boldsymbol{x}] = \zeta_0 + \zeta_1 a + \zeta_2^\top \boldsymbol{x}$ and variance $\text{Var}(M|a, \boldsymbol{x}) = \sigma^2$, or binary and described by a logistic model. The aim of this article is to unify these approaches and, in addition, to allow the mediator to follow other distributions from the Exponential family, such as Poisson (if M is a count) and Gamma (if M is asymmetric, nonnegative and continuous).

3. Generalized Linear Models for Mediators

3.1. Notation for GLM

To represent the NDE and NIE for a wider class of mediators, assume that the mediator M takes value on $\mathcal{M}$ and let ν be a sigma-finite measure over $\mathcal{A}$, a σ-field over $\mathcal{M}$, such that $M \sim \mathrm{P}^M(\cdot|a, \boldsymbol{x})$, with $\mathrm{P}^M \ll \nu$, and let its density be given by $f(m|a, \boldsymbol{x}) = \frac{d\mathrm{P}}{d\nu}(m|a, \boldsymbol{x}) = \exp\{[\theta m - b(\theta)]/\phi + c(m; \phi)\}$, so that, conditionally on a and $\boldsymbol{x}$, the distribution of M belongs to the Exponential family. Here, ϕ stands for the scale parameter and $\theta = \theta(a, \boldsymbol{x})$. The conditional mean of M given a and $\boldsymbol{x}$, denoted by $\mu = \mu(a, \boldsymbol{x})$, can be written as $\mu = g^{-1}(\eta) = g^{-1}(\zeta_0 + \zeta_1 a + \zeta_2^\top \boldsymbol{x})$, where g is a suitable link function. In particular, $\theta(a, \boldsymbol{x}) = (g \circ b')^{-1}(\zeta_0 + \zeta_1 a + \zeta_2^\top \boldsymbol{x})$ and, when using canonical links, $\theta(a, \boldsymbol{x}) = \eta$. In the next three subsections, we describe the formulas of NDE and NIE under these conditions and assuming the time to event (T) to follow the AFT, Aalen and Cox models, respectively. We assume a binary treatment such that $A = a$ or $A = a^*$ and, for the sake of simplicity, we simply write $\theta(\boldsymbol{x}) \equiv \theta(a, \boldsymbol{x})$ and $\theta(a^*, \boldsymbol{x}) \equiv \theta^*(\boldsymbol{x})$.

3.2. Causal Effects in the Accelerated Failure Time Model

Under conditions (A0) – (A4) and the accelerated failure time model, we have

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = \mathrm{E}\left[e^{\gamma\varepsilon}\right] \exp\left[\beta_0 + \beta_1 a + \beta_4^\top \boldsymbol{x} + \frac{b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left(\theta^*(\boldsymbol{x})\right)}{\phi}\right]. \quad (3)$$

The proof is in the appendix $A1$. Now, it follows directly from (3) that

$$\mathrm{NIE}_{\mathrm{AFT}} = \frac{b\left[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - \left[b\left(\theta(\boldsymbol{x})\right) - b\left(\theta^*(\boldsymbol{x})\right)\right]}{\phi}, \quad (4)$$

and

$$\mathrm{NDE}_{\mathrm{AFT}} = \beta_1(a - a^*) + \frac{1}{\phi}\{b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a^*)\phi\right]\}. \quad (5)$$

Dependence on $\boldsymbol{x}$ can be removed by considering $\mathrm{E}[T_{aM_{a^*}}] = \mathrm{E}_X[\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{X}]]$. However, the expressions of the NIE and NDE become complicated, as terms in (3) do not cancel out.

Using the Normal distribution with $g(\mu) = \mu$, $\phi = \sigma^2$ and $b(\theta) = \theta^2/2$, we get $\mathrm{NIE}_{\mathrm{AFT}} = (\beta_2 + \beta_3 a)\zeta_1(a - a^*)$ and $\mathrm{NDE}_{\mathrm{AFT}} = \left[\beta_1 + \beta_3(\beta_2\sigma^2 + \theta^*(\boldsymbol{x}))\right](a - a^*) + 0.5\sigma^2\beta_3^2(a^2 - a^{*2})$, as in VanderWeele (2015). Using the Gamma distribution with parameterization $f(m) = (\nu/\mu)^\nu m^{\nu-1} e^{-\nu m/\mu}/\Gamma(\nu)$, the most usual link functions are (i) the reciprocal link $g(\mu) = 1/\mu$, for which $\theta(a, \boldsymbol{x}) = -\zeta_0 - \zeta_1 a - \zeta_2^\top \boldsymbol{x}$; (ii) the logarithmic link $g(\mu) = \log \mu$, for which $\theta(a, \boldsymbol{x}) = -\exp(-\zeta_0 - \zeta_1 a - \zeta_2^\top \boldsymbol{x})$ and (iii) the identity link $g(\mu) = \mu$, for which $\theta(a, \boldsymbol{x}) = -(\zeta_0 + \zeta_1 a + \zeta_2^\top \boldsymbol{x})^{-1}$. In any case, $\phi = \nu^{-1}$ and $b(\theta) = -\log(-\theta)$, so that $\mathrm{NIE}_{\mathrm{AFT}} = \nu\{\log[(\beta_2 + \beta_3 a)/\theta^*(\boldsymbol{x}) + \nu] - \log[(\beta_2 + \beta_3 a)/\theta(\boldsymbol{x}) + \nu]\}$ and $\mathrm{NDE}_{\mathrm{AFT}} = \beta_1(a - a^*) + \nu\{\log[(\beta_2 + \beta_3 a^*)/\theta^*(\boldsymbol{x}) + \nu] - \log[(\beta_2 + \beta_3 a)/\theta^*(\boldsymbol{x}) + \nu]\}$. Expressions for Binomial and Poisson distributions are in the Web Appendix B.

3.3. Causal Effects in the Additive Aalen Model

From the Aalen's framework and assumptions (A0) – (A4),

$$\lambda_{T_a M_{a^*}}(t) = \beta_0(t) + \beta_1(t)a + \mathrm{E}^X_{t,a,a^*}\left[\beta_4(t)^\top \boldsymbol{X} + (\beta_2(t) + \beta_3(t)a)\mu_{t,a,a^*}(\boldsymbol{X})\right], \quad (6)$$

where $\mu_{t,a,a^*}(\boldsymbol{x}) = \mathrm{E}_{\phi(B_2(t)+B_3(t)a)+\theta^*(\boldsymbol{x})}M$ and $\mathrm{E}^X_{t,a,a^*}[h(\boldsymbol{X})]$ is the expectation of $h(\boldsymbol{X})$ with respect to

$$q_{t,a,a^*}(\boldsymbol{x}) \propto \exp\left[B_4(t)^\top \boldsymbol{x} + \frac{b\left[\theta^*(\boldsymbol{x}) + (B_2(t) + B_3(t)a)\phi\right] - b\left[\theta^*(\boldsymbol{x})\right]}{\phi}\right], \quad (7)$$

where $q_{t,a,a^*} = dQ_{t,a,a^*}/d\mathrm{P}^X$ and $\propto$ means equality except for a proportionality constant. Proofs are found in Appendix *A*2. Notice that in the absence of interaction the measure Q_{t,a,a^*} does not depend on a, so that $dQ_{t,a,a^*}(\boldsymbol{x}) \equiv dQ_{t,a^*}(\boldsymbol{x})$. From (6), we have

$$\mathrm{NIE}_{\mathrm{Aalen}} = (\beta_2(t) + \beta_3(t)a)\mathrm{E}^X_{t,a,a^*}[\mu_{t,a,a}(\boldsymbol{X}) - \mu_{t,a,a^*}(\boldsymbol{X})], \quad (8)$$

and

$$\mathrm{NDE}_{\mathrm{Aalen}} = \beta_1(t)(a - a^*) + \mathcal{E}(t; a, a^*) - \mathcal{E}(t; a^*, a^*). \quad (9)$$

with $\mathcal{E}(t; a, a^*) = \mathrm{E}^X_{t,a,a^*}[\beta_4(t)^\top \boldsymbol{X} + (\beta_2(t) + \beta_3(t)a)\mu_{t,a,a^*}(\boldsymbol{X})]$. In particular, if there is no interaction, then $\mathrm{NDE}_{\mathrm{Aalen}} = \beta_1(t)(a - a^*)$, regardless of the mediator type.

For the sake of illustration we will ignore the interaction term between A and M just to keep formulas simpler. Using the Normal distribution, we get from (8) the standard formula $\mathrm{NIE}_{\mathrm{Aalen}} = \beta_2(t)\zeta_1(a - a^*)$, whereas using the Gamma distribution, it follows that $\mathrm{NIE}_{\mathrm{Aalen}} = \beta_2(t)\mathrm{E}^X_{t,a,a^*}[(B_2(t)/\nu + \theta^*(\boldsymbol{X}))^{-1} - (B_2(t)/\nu + \theta(\boldsymbol{X}))^{-1}]$, where $q_{t,a,a^*}(\boldsymbol{x}) \propto \left[\frac{\theta^*(\boldsymbol{x})}{B_2(t)/\nu + \theta^*(\boldsymbol{x})}\right]^\nu e^{B_4(t)^\top \boldsymbol{x}}$.

3.4. Causal Effects in the Proportional Hazard Rate (Cox) Model

If the conditions (A0)-(A4) hold and the event is relatively rare, so that the cummulative baseline hazard $\Lambda_o(t)$ is approximatelly equal to zero for all $t \geq 0$, as pointed out in VanderWeele (2011), then

$$\lambda_{T_a M_{a^*}}(t|\boldsymbol{x}) \approx \lambda_o(t) \exp\left\{\beta_1 a + \beta_4^\top \boldsymbol{x} + \frac{b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left[\theta^*(\boldsymbol{x})\right]}{\phi}\right\}. \quad (10)$$

The proof is in the Appendix A3. Hence,

$$\mathrm{NIE}_{\mathrm{Cox}} \approx \frac{b\left[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - \left(b\left[\theta(\boldsymbol{x})\right] - b\left[\theta^*(\boldsymbol{x})\right]\right)}{\phi}, \quad (11)$$

and

$$\mathrm{NDE}_{\mathrm{Cox}} \approx \beta_1(a - a^*) + \frac{1}{\phi}\left\{b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi\right] - b\left[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a^*)\phi\right]\right\}. \quad (12)$$

Though in different scales, the mathematical expressions for NDE_{Cox} and NIE_{Cox} are very similar to the corresponding expressions for NDE_{aft} and NIE_{aft} for rare events, no matter what is the mediator distribution in the Exponential Family. If there is no interaction between A and M, then $\text{NDE}_{\text{Cox}} = \beta_1(a - a^*)$ regardless of the mediator type. For settings with common outcomes and with Normal/binary mediator, a weighting approach using the proportional hazards model was described by Lange et al. (2012) and VanderWeele (2015).

3.5. Variance Assessment of the Direct and Indirect Effects

Let $\zeta = (\zeta_0, \zeta_1, \zeta_2)^\top$ and $\boldsymbol{Z_i} = (A_i, \boldsymbol{X}_i^\top)^\top$, where $i = 1, ..., n$ denotes the sample units. If $\widehat{\zeta}$ is the maximum likelihood estimator (MLE) of ζ and assuming independence among units, then it is well known that $\widehat{\zeta}$ is asymptotically Normal with mean ζ and variance $\Sigma_\zeta = \phi(\boldsymbol{Z}^\top \boldsymbol{W} \boldsymbol{Z})^{-1}$, where $\boldsymbol{Z}$ is the design matrix whose ith column is $\boldsymbol{Z}_i$ and $\boldsymbol{W} = diag(w_1, \ldots, w_n)$, with $w_i = 1/[b''(\theta_i)\eta_i^2]$ and $\theta_i = \theta(a_i, \boldsymbol{x}_i)$. We also let Σ_ϕ be the (asymptotic) variance of the estimator $\widehat{\phi}$ of ϕ. On the other hand, under the AFT and Cox models, $\widehat{\beta}$ is asymptotically Normal with mean β and variance $\Sigma_\beta = n^{-1}\Sigma(\beta, \tau)$. Here $\widehat{\beta}$ is the estimator of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^\top$ obtained by maximum likelihood in the AFT model and maximum partial likelihood in the Cox model. The elements τ and $\Sigma(\beta, \tau)$ denote, respectively, the monitoring period and asymptotic variance. For details on Σ, we refer the reader to Kalbfleisch & Prentice (2002), pages 172 to 181. Following the arguments in VanderWeele (2015), p. 467, and using the facts that the above estimators are (i) asymptotically unbiased and (ii) $\widehat{\zeta}$ and $\widehat{\phi}$ are orthogonal, we have for large n that the covariance matrix associated to β, ζ and ϕ is given by

$$\Sigma = \begin{pmatrix} \Sigma_\beta & 0 & 0 \\ 0 & \Sigma_\zeta & 0 \\ 0 & & \Sigma_\phi \end{pmatrix}.$$

Hence, using the Delta method we are able to find the corresponding (asymptotic) variance of each effect. For example, for NIE_{AFT}, the variance is given by $(\nabla\text{NIE}_{\text{AFT}})^\top \Sigma \nabla\text{NIE}_{\text{AFT}}$, where $\nabla\text{NIE}_{\text{AFT}}$ stands for the gradient of NIE_{AFT} with respect to (β, ζ, ϕ), which is given by

$$\nabla\text{NIE}_{\text{AFT}}$$
$$= \frac{1}{\phi} \begin{bmatrix} 0 \\ 0 \\ \phi\left(b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi]\right) \\ \phi\left(b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi]\right) a \\ 0 \\ b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - [b'(\theta(\boldsymbol{x})) - b'(\theta^*(\boldsymbol{x}))] \\ b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi]a - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi]a^* - [b'(\theta(\boldsymbol{x}))a - b'(\theta^*(\boldsymbol{x}))a^*] \\ \{b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - [b'(\theta(\boldsymbol{x})) - b'(\theta^*(\boldsymbol{x}))]\}\, \boldsymbol{x} \\ -\text{NIE}_{\text{AFT}} + \{b'[\theta(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi] - b'[\theta^*(\boldsymbol{x}) + (\beta_2 + \beta_3 a)\phi]\}\,(\beta_2 + \beta_3 a) \end{bmatrix}$$

when considering canonical links. The gradients associated to the other direct and indirect effects for the AFT and Cox models are similarly evaluated.

Finally, we note that the Delta method will deliver good estimates only if the sample size is large enough. Otherwise, one can use the bootstrap approach as done in the application section. Particularly, such approach is also feasible for the Aalen model.

3.6. Simulation Study

We conducted a simulation study to assess the effect of misspecification in the mediation model (wrong choice of the GLM for M) on the NIE estimation. For each subject, i, we generated data relative to (T_i, M_i, A_i), where T denotes the survival time generated by the Cox-exponential model as in Bender et al. (2005), such that $\lambda(t) = \lambda_0(t)\exp(\beta_1 A + \beta_2 M)$, $A \sim Bernoulli(0.5)$ and the linear predictor for M is given by $\zeta_0 + \zeta_1 A$, with $\zeta = (\zeta_0, \zeta_1)'$ varying according to the underlying mediator distribution used in the data generation process. Specifically, we assumed the distributions Normal, Gamma, Poisson and Bernoulli. Finally, we considered sample sizes equal to $n = 200$ and $n = 500$ and percentages of censoring equal to 0% and 25%. To assess the effect of model misspecification, we estimated the NIE and NDE using AFT, Cox and Aalen survival models always assuming a GLM with Normal distribution for the mediation model as the standard approach (regardless the true distribution). To perform this assessment we generated 2,000 samples under each configuration.

In Table 1, we present average and standard errors (in parentheses) for the NIE estimates for the settings aforementioned. The true NIE is given by $\text{NIE}_{\text{Distribution}}$ as defined in Sections 3.2, 3.3 and 3.4, respectively, for AFT, Aalen and Cox models. Notice that the by fitting a Normal GLM for the mediator, we are misspecifying it when the true underlying distributions for M are Gamma, Poisson and Bernoulli. The results show that the estimates when using the correctly specified Normal-GLM are quite close to the true NIE, with improved results for larger sample sizes ($n = 500$) and no censoring. The estimates deviate from the true NIE values when using other GLMs in all survival models. For instance, the estimates for the AFT model and underlying Gamma distribution are 1.56 and 1.57, respectively, for the settings with 25% censoring and sample sizes 200 and 500, while the true NIE should be 0.90.

The departures are also verified under the Bernoulli distribution for the mediator, particularly for smaller sample size ($n = 200$) and using the Aalen model, when the estimates varied betwen -0.11 and -0.12, but the true estimate should be -0,08. Due to the small NIE estimates when using the Poisson model (and moreover the use of three decimals for presenting them), the results are hard to interpret for all survival models.

To help us to assess the departures from the true NIE under misspecification of the GLM for the mediator, we present relative bias (%) for the NIE in Table 2. The results were remarkably interesting, clearly pointing out the increased magnitude of the biases when ignoring the proper GLM for the mediator. We can verify in

TABLE 1: Average NIE estimates and standard errors (in parentheses) using the GLM-Normal for modeling the mediator under different survival and underlying (true) mediator models.

Underlying Mediator Model	**Survival Models**					
	AFT		**Cox**		**Aalen**	
	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
Normal						
$(\zeta_o = 100, \zeta_1 = 10)$						
0% censoring	0.51	0.50	-0.52	-0.50	-0.024	-0.023
	(0.73)	(0.45)	(0.75)	(0.45)	(0.035)	(0.021)
25% censoring	0.52	0.48	-0.55	-0.51	-0.020	-0.020
	(0.79)	(0.48)	(0.85)	(0.51)	(0.034)	(0.020)
$\text{NIE}_{\text{Gauss}}$	0,50		-0,50		-0.022	
Gamma						
$(\zeta_o = 0.5, \zeta_1 = 3.0)$						
0% censoring	1.56	1.57	-1.58	-1.58	-0.37	-0.22
	(0.19)	(0.12)	(0.26)	(0.16)	(0.27)	(0.16)
25% censoring	1.56	1.57	-1.68	-1.66	-0.34	-0.19
	(0.20)	(0.13)	(0.31)	(0.18)	(0.24)	(0.14)
$\text{NIE}_{\text{Gamma}}$	0.90		-0.90		-4.46	
Poisson						
$(\zeta_o = 0.02, \zeta_1 = 0.06)$						
0% censoring	0.003	0.0030	-0.003	-0.0030	-0.02	-0.026
	(0.013)	(0.0069)	(0.013)	(0.0069)	(0.11)	(0.058)
25% censoring	0.002	0.0031	-0.002	-0.0033	-0.02	-0.024
	(0.014)	(0.0070)	(0.015)	(0.0074)	(0.11)	(0.054)
$\text{NIE}_{\text{Poisson}}$	0.003		-0.003		-0.024	
Bernoulli						
$(\zeta_o = 0.02, \zeta_1 = 0.90)$						
0% censoring	0.012	0.011	-0.012	-0.011	-0.12	-0.10
	(0.034)	(0.020)	(0.035)	(0.021)	(0.33)	(0.20)
25% censoring	0.012	0.010	-0.013	-0.010	-0.11	-0.08
	(0.037)	(0.023)	(0.039)	(0.024)	(0.32)	(0.19)
$\text{NIE}_{\text{Bernoulli}}$	0.011		-0.011		-0.08	

($\text{NIE}_{\text{Distribution}}$: true natural indirect effect when considering the actual underlying distribution for the mediator.)

our settings that the relative bias varies from 15.02% to over 87.17% when the underlying mediator is, respectively, Poisson and Gamma for the analyses using Cox models with $n = 200$. There is a substantial reduction on the relative bias for NIE estimation when sample size increases to 500, particularly for the Poisson and Bernoulli mediator distributions. For the Gamma GLM mediator, on the other hand, the relative bias does not change substantially with the increase in sample size for all survival models.

We also compared the AIC for the two models fitted for the mediator: the GLM for the underlying mediator model (TRUE) and the Gaussian model, i.e., under misspecification of the mediation model. In all cases, the AIC for the correct (TRUE underlying model) was lower than that obtained by fitting the Gaussian GLM, indicating that the criterion correctly indicates the proper model

TABLE 2: Relative bias for NIE using the GLM-Normal to model the mediator under different survival and underlying (true) mediator models

Underlying	**Survival Models**					
Mediator	**AFT**		**Cox**		**Aalen**	
Model	$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
Normal						
$(\zeta_o = 100, \zeta_1 = 10)$						
0% censoring	2.49%	-0.74%	-3.37%	0.21%	-11.26%	-6.15%
25% censoring	4.06%	-3.04%	-10.82%	-2.28%	-0.71%	6.94%
Gamma						
$(\zeta_o = 0.5, \zeta_1 = 3.0)$						
0% censoring	73.68%	74.74%	-76.07%	-76.30%	91.61%	95.10%
25% censoring	74.32%	75.03%	-87.17%	-85.16%	92.32%	95.67%
Poisson						
$(\zeta_o = 0.02, \zeta_1 = 0.06)$						
0% censoring	-16.29%	-1.74%	15.02%	-0.56%	13.70%	-8.09%
25% censoring	-20.07%	0.0079%	15.58%	-6.53%	26.77%	1.25%
Bernoulli						
$(\zeta_o = 0.02, \zeta_1 = 0.90)$						
0% censoring	14.21%	3.90%	-16.65%	-3.80%	-44.61%	-25.20%
25% censoring	13.57%	-6.05%	-20.61%	2.29%	-27.89%	-1.07%

for the mediator (Table 3). For instance, the AIC was equal to 649.06 and 682.85, respectively, for the GLMs with Bernoulli and Normal distributed mediator when its underlying (TRUE) distribution was Bernoulli.

TABLE 3: Average AIC for the mediator models using different GLMs

Underlying	**Sample Size**	
Mediator Model	$n = 200$	$n = 500$
Gamma		
True Gamma GLM	1,192.27	2,988.86
Misspecified Normal GLM	1,741.38	4,439.75
Poisson		
True Poisson GLM	536,60	1,336.03
Misspecified Normal GLM	581.24	1,446.77
Bernoulli		
True Bernoulli GLM	261.75	649.06
Misspecified Normal GLM	276.38	682.85

In summary, we showed that there might be substancial bias when ignoring the proper GLM for the mediator whatever the survival model (Aalen, Cox or AFT model) for NIE estimation is. Though not presented here, the impact on the NDE estimation is much smaller when using the GLM-Normal mediator model since its estimation does not depend on the GLM parameters associated to the mediator.

4. Responsiveness Measure for Continuous Treatments

Although it is usual to consider binary exposures and their causal contrast, it might be interesting to consider an alternative metric to describe the effect of small changes in a continuous exposure in the presence of mediation. Therefore, in this section we propose a measure that captures the responsiveness of the outcome, in different scales, to changes in the underlying (continuous) exposure level when mediation is present. More precisely, we look at the responsiveness, or derivative, of the potential responses due to small changes in the treatment level. It is worth mentioning that this is a local measure and, therefore, as will become clearer later, it is related to some pre-specified level of treatment. Although we consider, as in Section 3, three different scales, we will state the main definitions by using the expected time to event. Modifications to other scales should be trivial. Therefore, we define the *total responsiveness at* a^* (TR) as

$$\text{TR} = \frac{\partial}{\partial a}\{\text{E}[T_{aM_a}|\boldsymbol{x}] - \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}]\}\Big|_{a=a^*} = \lim_{h\to 0}\frac{\text{E}[T_{(a^*+h)M_{a^*+h}}|\boldsymbol{x}] - \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}]}{h},$$

and the *natural direct responsiveness at* a^* (NDR) as

$$\text{NDR} = \frac{\partial}{\partial a}\{\text{E}[T_{aM_{a^*}}|\boldsymbol{x}] - \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}]\}\Big|_{a=a^*} = \lim_{h\to 0}\frac{\text{E}[T_{(a^*+h)M_{a^*}}|\boldsymbol{x}] - \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}]}{h}.$$

Note that for the direct responsiveness, the treatment assigned to the mediator remains fixed at its "natural" level a^*. The *natural indirect responsiveness at* a^* (NIR), in turn, is simply defined in terms of TR and NDR as NIR = TR − NDR, i.e. the natural indirect responsiveness is the part of the total responsiveness that is not explained by the natural direct effect. However, if TR and NDR exist, then $\text{NIR} = \lim_{h\to 0}\{\text{E}[T_{(a^*+h)M_{a^*+h}}|\boldsymbol{x}] - \text{E}[T_{(a^*+h)M_{a^*}}|\boldsymbol{x}]\}/h$, so that

$$\text{NIR} = \frac{\partial}{\partial a}\{\text{E}[T_{aM_a}|\boldsymbol{x}] - \text{E}[T_{aM_{a^*}}|\boldsymbol{x}]\}\Big|_{a=a^*}.$$

As before, our goal is to study the identification of these measures in cases where the mediator's conditional distribution belongs to the Exponential Family.

4.1. Responsiveness Measures in the Accelerated Failure Rate Model

We already know that if (A0)–(A4) are satisfied, then $\text{E}[T_{aM_a}|\boldsymbol{x}]$ is identified by formula (3). Based on it and taking derivatives with respect to a, we get

$$\text{TR}_{\text{AFT}} = \left[\beta_1 + \frac{(\theta'(a^*,\boldsymbol{x}) + \beta_3\phi)\text{E}_{\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a^*)\phi}M - \theta'(a^*,\boldsymbol{x})\text{E}_{\theta(a^*,\boldsymbol{x})}M}{\phi}\right] \times \text{E}[T_{a^*M_{a^*}}|\boldsymbol{x}], \tag{13}$$

where $\theta'(a^*,\boldsymbol{x}) = \partial_a\theta(a^*,\boldsymbol{x})$. One particular issue that arises when using the identification formula (3) is the calculation of the expectation $\text{E}e^{\gamma\varepsilon}$, which depends

on the particular distribution associated to $\log T$. It is usual to choose ε so that it follows an extreme value distribution. This is done by assuming T to follow the Weibull distribution, see Kalbfleisch & Prentice (2002), p. 33. In this particular case, $\mathrm{E}e^{\gamma\varepsilon} = \Gamma(1+\gamma)$. Similarly, we get

$$\mathrm{NDR}_{\mathrm{AFT}} = \left(\beta_1 + \beta_3 \mathrm{E}_{\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a^*)\phi} M\right) \mathrm{E}[T_{a^* M_{a^*}}|\boldsymbol{x}],$$

and hence

$$\mathrm{NIR}_{\mathrm{AFT}} = \phi^{-1}\theta'(a^*,\boldsymbol{x}) \left(\mathrm{E}_{\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a^*)\phi} M - \mathrm{E}_{\theta(a^*,\boldsymbol{x})} M\right) \mathrm{E}[T_{a^* M_{a^*}}|\boldsymbol{x}].$$

In particular, if there is no interaction between A and M in the survival model, then the direct responsiveness simplifies to $\beta_1 \mathrm{E}[T_{a^* M_{a^*}}|\boldsymbol{x}]$. The same arguments would lead us to the responsiveness formulas associated to the Cox model, whose results are quite similar, except for the scale. In fact, by comparing the expressions (3) and (10), respectively, associated to $\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}]$ and $\lambda_{aM_{a^*}}(t)$, we see that both share the same structure (just with $\mathrm{E}[e^{\gamma\varepsilon}]$ in place of $\lambda_o(t)$ and vice versa). Hence, by mimicking the argument in the lines above, we get

$$\mathrm{TR}_{\mathrm{Cox}} = \left[\beta_1 + \frac{(\theta'(a^*,\boldsymbol{x}) + \beta_3\phi)\mathrm{E}_{\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a^*)\phi} M - \theta'(a^*,\boldsymbol{x})\mathrm{E}_{\theta(a^*,\boldsymbol{x})} M}{\phi}\right] \times \lambda_{a^* M_{a^*}}(t).$$

Notice that, in addition to the difference in scale, the responsiveness under the Cox model is also time dependent, *i.e.* $\mathrm{TR}_{\mathrm{Cox}} = \mathrm{TR}_{\mathrm{Cox}}(t)$. The corresponding direct and indirect responsiveness follow the same steps.

For the sake of illustration we only consider models without the interaction term. If M is Gaussian, then

$$\mathrm{NDR}_{\mathrm{AFT}} = \mathrm{E}\left[e^{\gamma\varepsilon}\right] \beta_1 \exp\left\{\beta_0 + \beta_1 a^* + \beta_2 \theta^*(\boldsymbol{x}) + \beta_4^\top \boldsymbol{x} + 0.5\sigma^2\beta_2^2\right\}$$

and $\mathrm{NIR}_{\mathrm{AFT}} = \theta'(a^*,\boldsymbol{x})[(\beta_2\theta(a^*,\boldsymbol{x}) + 0.5\sigma^2\beta_2^2)/\beta_1]\mathrm{NDR}_{\mathrm{AFT}}$. On the other hand, if M follows a Gamma distribution, then

$$\mathrm{NDR}_{\mathrm{AFT}} = \beta_1 \mathrm{E}\left[e^{\gamma\varepsilon}\right] \exp\left\{\beta_0 + \beta_1 a^* + \beta_4^\top \boldsymbol{x} + \nu \log \frac{\theta(a^*,\boldsymbol{x})}{\beta_2/\nu + \theta(a^*,\boldsymbol{x})}\right\}$$

and

$$\mathrm{NIR}_{\mathrm{AFT}} = \lambda'(a^*,\boldsymbol{x})\beta_2/\{\beta_1[\beta_2/\nu + \theta(a^*,\boldsymbol{x})]\}\mathrm{NDR}_{\mathrm{aft}},$$

with $\lambda(a^*,\boldsymbol{x}) = \log\theta(a^*,\boldsymbol{x})$. For specific links we have: $\lambda'(a^*,\boldsymbol{x}) = -\zeta_1$ for the log link, $\lambda'(a^*,\boldsymbol{x}) = -\zeta_1/\theta(a^*,\boldsymbol{x})$ for the reciprocal link, and $\lambda'(a^*,\boldsymbol{x}) = \zeta_1\theta(a^*,\boldsymbol{x})$ for the identity link function.

4.2. Responsiveness Measures in the Additive Aalen Model

By definition, the responsiveness associated with the Aalen model is defined in terms of the potential hazard functions $\lambda_{T_{aM_a}}$ and $\lambda_{T_{aM_{a^*}}}$. Unfortunately, in this particular case, the general expressions of total, direct and indirect sensitivities are

cumbersome and can be found in the Appendix C, formulas (A2), (A3) and (A4). In the following lines, however, we present the respective formulas for the simpler case in which there is no interaction between exposure and mediator in the survival model. As usual, the absence of interaction between A and M dramatically simplifies the direct responsiveness, which is given simply by $\text{NDR}_{\text{Aalen}} = \beta_1(t)$. The total responsiveness, in turn, is given by $TR_{\text{Aalen}} = \beta_1(t) + \Xi_2 + \Xi_3$, with

- $\Xi_2 = \phi^{-1}\beta_2(t)\text{E}^X_{t,a^*}\theta'(a^*, \boldsymbol{X})\text{Var}_{\phi B_2(t)+\theta(a^*,\boldsymbol{X})}M$;
- $\Xi_3 = \phi^{-1}\text{E}^X_{t,a^*}\left[\beta_4(t)^\top \boldsymbol{X} + \beta_2(t)\text{E}_{\phi B_2(t)+\theta(a^*,\boldsymbol{X})}M\right]\psi_{1,t}(a^*, \boldsymbol{X})$;

and $\psi_{1,t}$ as defined in Appendix B. Finally, the indirect responsiveness is derived from $\text{NDR}_{\text{Aalen}}$ and TR_{Aalen}, so that $\text{NIR}_{\text{Aalen}} = \Xi_2 + \Xi_3$ (with Ξ_2 and Ξ_3 as given above).

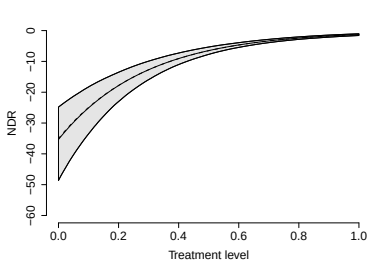
For the sake of illustration, we consider once again the Normal and Gamma distributed mediator without interaction between A and M in the survival (Aalen) model. In the Gaussian case, $\Xi_2 = \beta_2(t)\text{E}^X_{t,a^*}\theta'(a^*, \boldsymbol{X})$ and

$$\Xi_3 = B_2(t)\text{E}^X_{t,a^*}\left[\left(\beta_4(t)^\top \boldsymbol{X} + \beta_2(t)\theta(a^*, \boldsymbol{X})\right)\left(\theta'(a^*, \boldsymbol{X}) - \text{E}^X_{t,a^*}\theta'(a^*, \boldsymbol{X})\right)\right]$$

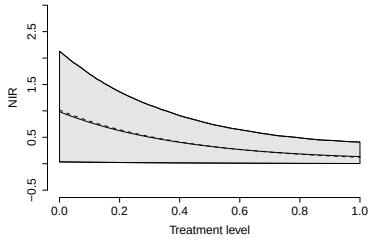
More specifically, if $\theta(a, \boldsymbol{x}) = \zeta_0 + \zeta_1 a + \zeta_2^\top \boldsymbol{x}$, then $\text{NIR}_{\text{Aalen}} = \zeta_1\beta_2(t)$ and $TR_{\text{Aalen}} = \beta_1(t) + \zeta_1\beta_2(t)$. For a Gamma mediator and no confounding, on the other hand, we have $\text{NIR}_{\text{Aalen}} = [\theta'(a^*)/(B_2(t)/\nu + \theta(a^*))^2]\beta_2(t)$, which can be specialized for different link functions by changing the function θ.

4.3. Simulation Study

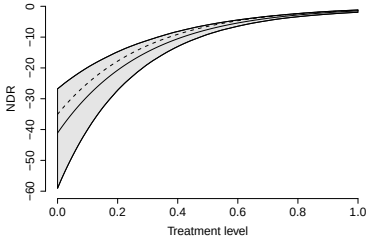
As in Section 3.6, we conducted a simulation study to assess the effect of misspecification in the mediation model (wrong choice of the GLM for M), but now regarding the NDR and NIR estimation. We adopted the same procedure, so that, for each subject we sampled (T_i, M_i, A_i), with T denoting the survival time generated by the Cox-exponential model as in Bender et al. (2005) and $\lambda(t) = \lambda_0(t)\exp(\beta_1 A + \beta_2 M)$, with $\beta_1 = 3.5$, $\beta_2 = -0.09$ and $\lambda(t) \equiv 11$. The treatment A was sampled from a Bernoulli(0.5) and the linear predictor for M was assumed to be $\zeta_0 + \zeta_1 A$, with $\zeta_0 = 0$, $\zeta_1 = -1$ and scale parameter equal to 1, associated to the Gamma distribution. The resulting direct and indirect responsiveness are illustrated in Figures 1 and 2 (dashed lines). They represent a situation in which an increase in treatment levels implies a direct decrease in the response variable, particularly in the vicinity of zero (*i.e.* small increases in treatment levels starting from zero result in greater changes than small increases in these levels in already treated individuals). On the other hand, the treatment has an indirect positive effect on the response, though such impact is reduced for higher doses. To illustrate the method, we estimated both NDR and NIR using the AFT model by assuming the Gamma (correctly specified) and the Gaussian (misspecified) distributions for the mediator as well as sample sizes 200 and 500 and percentages of censoring equal to 0% and 25%. To perform this assessment we generated 2,000 samples under each configuration.



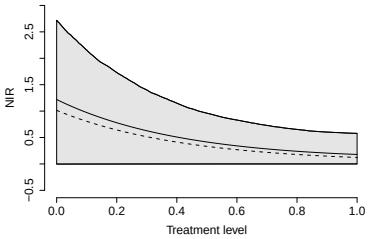
(a) NDR, $n = 200$, 0% censoring



(b) NIR, $n = 200$, 0% censoring



(c) NDR, $n = 200$, 25% censoring

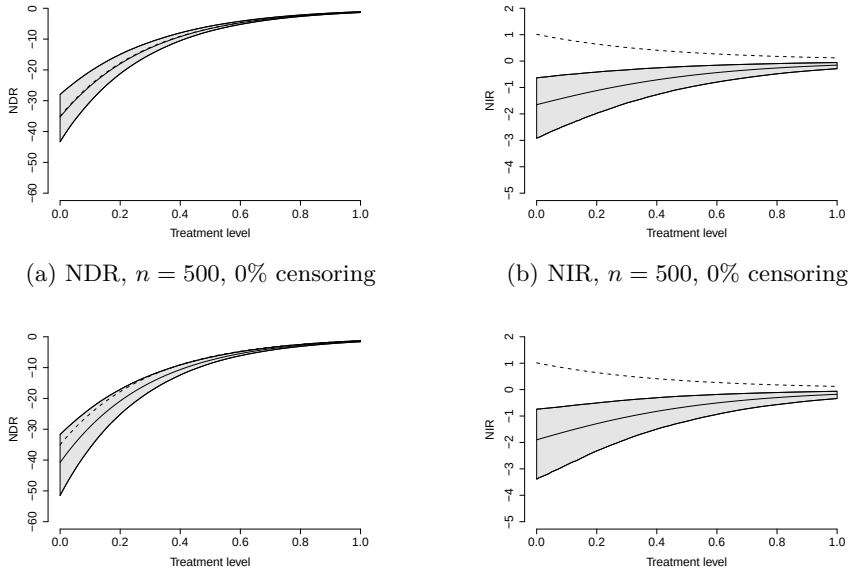


(d) NIR, $n = 200$, 25% censoring

FIGURE 1: NDR and NIR estimates based on the AFT model with Gamma mediators (log link). The dashed lines stand for the true (rescaled) effects, the middle solid lines stand for the median responsivenesses and the lower and upper solid lines for the 2.5% and 97.5% quantiles.

Figure 1 presents the estimates of the direct and indirect responsiveness by assuming the true underlying GLM Gamma model and samples of size 200. Figures 1a and 1b stand for estimates based on samples with no censoring, while Figures 1c and 1d are about the samples with 25% censoring. Simulations indicate that estimates under no censoring are unbiased, though estimates based on censored samples may present some bias. Indeed, by looking at Figures 1a and 1b, it is almost impossible to identify the true curve (dashed lines in all figures). However, in all cases, the true curve is between the lower and upper bands based on the 2.5% and 97.5% quantiles of the simulated estimates at each point of the grid. Similar results were found for sample size 500, with smaller variability, as expected.

We also considered the cases where estimation was based on the (misspecified) Gaussian distribution with its canonical link (identity). The results are in Figure 2 ($n = 500$). As noted in Figures 2a and 2c, the direct responsiveness does not seem to be very affected. In fact the estimates here are not very different as those obtained by using the true Gamma distribution. This is quite reasonable for the NDR, since it is less dependent on the mediator than the NIR. On the other hand, regardless of sample size, NIR estimates are highly biased (Figures 2b and 2d). Such estimates may even go towards to the wrong direction. In fact, at any treatment level, NIR estimates based on the Gaussian distribution are negative (indicating a decrease in the response variable as the treatment level



(a) NDR, $n = 500$, 0% censoring

(b) NIR, $n = 500$, 0% censoring

(c) NDR, $n = 500$, 25% censoring

(d) NIR, $n = 500$, 25% censoring

FIGURE 2: NDR and NIR estimates based on the AFT model with Gaussian mediators (identity link) and sample size 500. The dashed lines stand for the true (rescaled) effects, the middle solid lines stand for the median responsivenesses and the lower and upper solid lines for the 2.5% and 97.5% quantiles.

increases), while the true values are positive (which means an increase in response by increasing the level of treatment in that neighborhood).

5. Application

To illustrate the new approaches, we reexamine the mediation model for the incidence of liver cancer (Figure 3), which was proposed by Huang & Yang (2017). We use data from a subset of individuals ($n = 2,878$) who participated in a community-based prospective cohort study in Taiwan conducted from 1991 to 1992, see Huang et al. (2011), in which the viral load of hepatitis C (HCV) was measured at baseline, the viral load of hepatitis B (HBV) was measured during the follow-up, and the incidence of liver cancer (T) was recorded prospectively. We are interested in assessing the effect of HCV (A) directly on the liver incidence (hepatocellular carcinoma) and its effect mediated through HBV (M). The analyses were adjusted by measured confounders ($\mathbf{X}$): age group (30-39, 40-49, 50-59, ≥ 60 years), gender (female/male), alcohol consumption and cigarette smoking (no/yes). Viral loads of HBV and HCV were natural log transformed prior to analyses. For the initial analyses, we dichotomize HCV (detected if > 0/non detected otherwise).

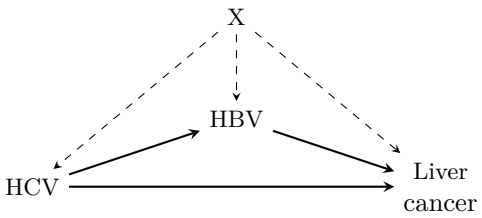


FIGURE 3: Causal mediation diagram for the relationship between HCV viral load and occurrence of liver cancer

We used total time scale for the survival analyses, such that the entry time was defined as the time of measuring HCV (at baseline) and the occurrence of liver cancer was assessed by data linkage of the national cancer registry and national death certification profiles from study entry to December 31, 2008 (end of the study). The event was also verified by medical records. Individuals who were not diagnosed of liver cancer until the end of the study or those who died from other causes were censored. The number of patients with liver cancer diagnosis was 188 (6.53%). Hence, the event is rare and suitable to be evaluated using Cox proportional hazards model. We compare the results for fitting AFT, Cox and Aalen models using GLMs, considering different link functions, for the mediator (HBV) at follow-up. To model the mediator HBV viral load, we considered the Normal distribution with identity link function and the Gamma distribution with the logarithm and reciprocal link functions. Confidence intervals for causal effects are obtained via bootstrap method for all survival models. We assumed that there were no unmeasured confounders of the relationship between HCV and time to liver cancer (A1), HBV and time to liver cancer (A2), and HCV and HBV (A3). The *composition* (A0) and identifiability assumptions (A4) are also required for any causal claims based on these analyses. All analyses were conducted in R (version 3.5.3).

5.1. Causal Effects for HCV

Results using the three mediation models revealed similar patterns under different effect scales. As indicated in Table 4, the NDE estimate by using the Cox model suggest that the detection of HCV viral load at baseline directly increases the risk of liver cancer ($\text{HR} = 3.26, 95\%CI = [1.75, 6.06]$)

TABLE 4: Estimated Natural Direct Effect (NDE) of hepatitis C detection on liver cancer incidence using different models. Taiwan.1991-2008.

Models	Estimate	95%CI
AFT $^{(\star)}$	-0.58	(-0.90; -0.27)
Cox model (HR)$^{(\star\star)}$	3.26	(1.75; 6.06)
Aalen model $^{(\star\star\star)}$	7.39	(1.42; 13.29)

$^{(\star)}$ Scale: Difference in mean survival time, $^{(\star\star)}$ Scale: HR=hazard ratio, $^{(\star\star\star)}$ Scale: Difference in hazard (per 1000 person-year)

On the other hand, NDE estimates by using the AFT and Aalen models indicate, respectively, a reduction of 56% (exp(−0.58)) in the expected time to diagnosis and a difference in the hazard of 7.39 (per 1000 person-years) of liver cancer for those with detected HBC viral load at baseline compared to non detects.

Table 5: Estimated Natural Indirect Effects (NIE) of hepatitis C detection on liver cancer incidence mediated through follow-up hepatitis B viral load using distinct models. Taiwan.1991-2008.

	GLM for Mediator (distribution / link function)		
	Normal	Gamma$^{(a),(b)}$	
Time-to-event	Identity	Log	Inverse
model	Estimate (95% CI)	Estimate (95%CI)	Estimate (95%CI)
AFT $^{(\star)}$	0.23 (0.12; 0.35)	0.21 (0.12; 0.29)	0.22 (0.13; 0.31)
Cox (HR)$^{(\star\star)}$	0.63 (0.49; 0.79)	0.47 (0.33; 0.64)	0.45 (0.32; 0.63)
Aalen$^{(\star\star\star)}$	-2.03 (-3.14;-1.02)	-2.07 (-2.97; -1.20)	-2.11 (-2.99; -1.19)

$^{(\star)}$ Scale: Difference in mean survival time, $^{(\star\star)}$ Scale: HR=hazard ratio
$^{(\star\star\star)}$ Scale: Difference in hazard (per 1000 person-year)
$^{(a)}$ AFT, Cox and Aalen estimates conditional on X, here computed for $\mathbf{X} = \mathbf{0}$,
$^{(b)}$ NIE in Aalen model is time-dependent, approximation is shown.

Table 5 presents estimates of NIE using different models for the mediator (varying both distribution of the mediator and link function). All models indicate a statistically significant mediation through follow-up HBV viral load, such that the NIE points out that the HCV viral load increases the mean time of liver cancer diagnosis, decreasing the hazard ratio and the difference in hazard, according to the distinct scales of the corresponding models. Residual analysis using the linear mediator model shows a poor fit (see Figure 1 in the Appendix D). Smallest AIC was obtained using the Gamma distribution with almost no difference between the link functions. It is important to highlight that the estimates for NIE under AFT, Cox and Aalen models are conditional on the covariates ($\mathbf{X}$) when using the Gamma distribution. To illustrate these results, we present the NIE estimates for 30-39 females, no smokers and no drinkers (reference groups for all covariates) in Table 5. We can compute these estimates for any covariate pattern of our interest. For instance, the NIE estimates using AFT model and the Gamma distribution are 0.18 (log link) and 0.17 (inverse link) for ≥ 60 years, males, smokers and drinkers. The estimates for the corresponding Cox models are 0.55 and 0.58.

The NIE for the Aalen model using the Gamma distribution, on the other hand, is time-dependent. However, for our current application $|\theta(\cdot)|$ is much larger than $B_2(t)$. We have that $\max(\hat{B}_2(t)) \approx 0.15$, with $\hat{\theta}(a, \mathbf{x} = \mathbf{0})$ varying between -0.28 and -0.22 and $\hat{\nu} = 6.11$, so that the NIE can be approximated by

$$\begin{aligned} \text{NIE} &= \beta_2 \mathrm{E}_t^x \left[\frac{1}{\theta(a^*, \boldsymbol{x}) + B_2(t)/\nu} - \frac{1}{\theta(a, \boldsymbol{x}) + B_2(t)/\nu} \right] \\ &\approx \beta_2 \mathrm{E}^x \left[\frac{1}{\theta(a^*, \boldsymbol{x})} - \frac{1}{\theta(a, \boldsymbol{x})} \right], \end{aligned}$$

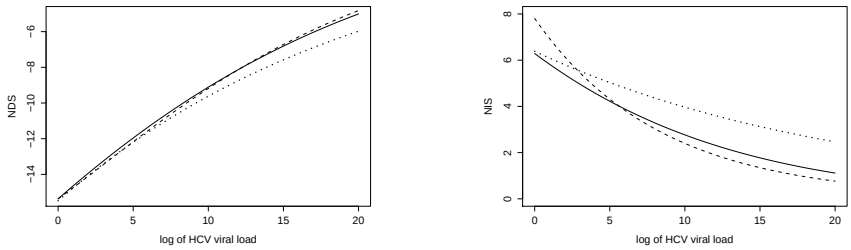
which does not depend on time. The NIE estimates using Aalen model and Gamma distribution indicate a difference in hazard (per 1000 person-year) of

-2.07 ($CI_{0.95} = [-2.97, -1.20]$) and -2.11 ($CI_{0.95} = [-2.99, -1.19]$) for 30-39 females, no smokers and no drinkers (reference groups for all covariates) using, respectively, log and inverse link functions. Results can be obtained for different covariate patterns.

Alternatively, we also considered the log HCV viral load in its original scale as described in Huang & Yang (2017), so that it is a continuous variable instead of a binary one as before. The estimated effects have the same direction as those shown in Tables 4 and 5. In short, by increasing the log HCV viral level in one unit, the estimated natural direct effect on the mean survival time is $\text{NDE}_{\text{AFT}} = -0.09$ ($95\%CI = [-0.14, -0.04]$); on the hazard ratio is $\exp(\text{NDE}_{\text{Cox}}) = 1.20$ ($95\%CI = [1.08, 1.33]$); and on the difference in the hazard is $\text{NDE}_{\text{Aalen}} = 1.20$ (per 1000 person-years) $95\%CI = [0.17, 1.33]$. On the other hand, NIE does not change substantially when considering different mediator models (varying distribution and link functions). For instance, the largest difference was verified for the Cox model, with NIE varying from $HR = 0.85, 95\%CI = [0.76, 0.92]$ to $HR = 0.93, 95\%CI = [0.89, 0.96]$ under, respectively, the Gamma (inverse link) and Gaussian distributions. Residual analysis for these mediator models were very similar as those presented in Appendix D.

5.2. Responsiveness Measures for HCV

To illustrate the responsiveness measure for continuous exposure, we consider the natural log transformed viral load of HCV. Figure 4 presents the natural direct and indirect responsiveness for the AFT model using different mediator models. Though the NDR_{aft} is not very sensitive to the mediation modeling, the

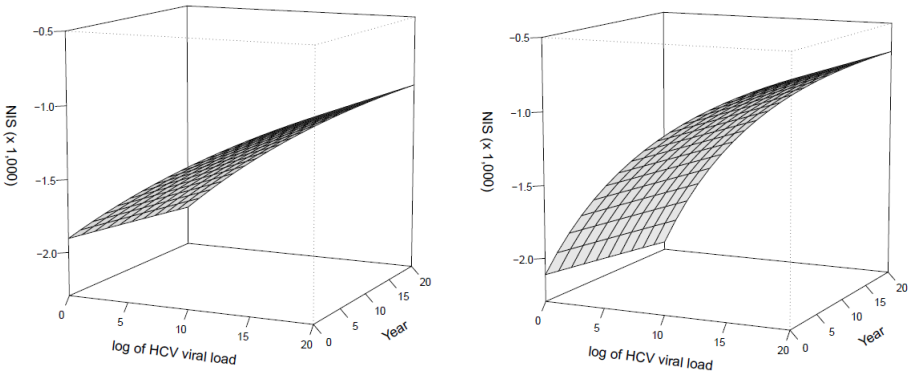


(a) Natural direct responsiveness (b) Natural indirect responsiveness

FIGURE 4: Responsiveness measures of the liver cancer in terms of the log HCV viral load using AFT model with Normal (dotted line) and Gamma distributions (logarithm link: solid line; inverse link: dashed line) for the mediator.

NIR_{aft} estimates indicate a larger responsiveness to treatment when considering the Gamma distribution. Small increases in the hepatitis C viral load directly imply a greater reduction in the expected time until the diagnosis of liver cancer when the viral load is relatively low. On the other hand, it is noted that, indirectly, the hepatitis B viral load (mediator) behaves in a protective way. However, such

a protective effect becomes less relevant as the viral load of hepatitis C increases. The responsiveness measures associated to the Aalen model depend both on HCV values and time (Figure 5). However, the effect of time on the hazard function is not as relevant as the effect of the HCV, as depicted in Figures 3(a) and 3(b). Regarding the log HCV viral load, the higher its value the faster the indirect effect on the hazard of diagnosis of liver cancer. Since we are assuming the coefficients in the Aaalen model are constant in time, $\text{NDR}_{\text{Aalen}}$ is invariant with respect to time and HCV viral load, so that $\text{NDR}_{\text{Aalen}} = \text{NDE}_{\text{Aalen}}$.



(a) Gamma with log link

(b) Gamma model with inverse link

FIGURE 5: Responsiveness measures of liver cancer hazard in terms of the log HCV viral load using Aalen model and a Gamma mediator with different link functions

6. Conclusions

This article presents formulas for GLM mediation when a time-to-event outcome is of interest and the AFT, Cox or Aalen models are used. Under the Aalen additive and Cox models, the effects are characterized, respectively, on the hazard difference and on the log hazard ratio scales. At the same time, AFT examine the effect on the mean survival time difference. Therefore, the effects estimated using AFT models are in the opposite direction compared to the other two models (Cox and Aalen), conveying complementary information when both are properly applied. We derive general expressions to the direct and indirect effects when the mediation process can be described by a GLM. Particularly, we allow interaction between exposure and mediator in all models. For concreteness, we focused on particular distributions for the mediator: Gaussian, Gamma, Binomial and Poisson, but related formulas can be readily extended to other distributions. Even though causal mediation estimators for survival models are available when the mediator is continuous (Normal distributed) or binary, to our knowledge, not much has been proposed for mediators of different nature to handle time-to-event data. For instance, Albert & Nelson (2011) consider

in another context, not including censored data, the use of GLM for mediation analysis. Tchetgen Tchetgen (2013), on the other hand, describes a inverse odds ratio-weighted approach for effect decomposition in GLMs with a nonlinear link function, including mediation analysis using the Cox proportional hazards model.

We successfully illustrate our approach by analyzing the incidence of liver cancer in a community-based prospective cohort study in Taiwan. The analysis showed that the HCV viral load directly affects liver cancer increasing its risk, but indirectly, via HBV viral load, has an opposite effect. The use of different effect scales to investigate the impact of HCV viral load (continuous or categorized) on liver cancer and mediator models for HBV viral load yielded consistent assessments. The data are fully anonymised and were made publicly available by Huang & Yang (2017), so that the interested reader may follow the link therein to download the data. This empirical example demonstrates the flexibility of our framework and its potential for studying mediation effects, linking theoretical knowledge and empirical evidence to refine the scientific theories for complex relationships. In fact, mediation models offer a more detailed view of the underlying mechanism and, hence, a more comprehensive understanding of the phenomena. Further refinements related to the application by considering different mediation arrangements are topics for subsequent research. Indeed, our data analysis using deviance residuals indicates the need for more sensitive models in order to capture asymmetry and heteroskedastic errors. For example, the modeling of the dispersion parameter, as in Smyth (1989), though not related to mediation analysis, could be considered as an alternative to be developed in this framework.

The proposed approach has a number of strenghts. Survival models provide a compelling framework for many research questions, and extensions via inclusion of GLMs for the mediator may encompass a broad field of applied research, allowing the often necessary control for confounding. In fact, a major limitation related to existing statistical theory and software tools for causal mediation analysis with survival outcomes is the confinement to continuous or binary mediators, see Huang & Yang (2017), Pratschke et al. (2016) and VanderWeele (2015). We provided a rigorous methodological justification of our approach, focusing on the identification of the direct and indirect natural effects and offering analytical expressions under the presence of GLM mediation. In particular, it is worth noticing that the additive hazard model naturally embodies the possible time dependence of the corresponding coefficients, yielding additional flexibility. We also have proposed responsiveness measures capable of assessing the local impact of small changes in the level of (continuous) treatment by both direct and indirect pathways. In particular, this allows us to assess the most critical levels of treatment (in terms of impact on the outcome) in the sense of identifying the regions in which small changes in exposure imply larger variations in the outcome. Estimation of the model's parameters for our approach may be performed using standard statistical packages. We use R (version 3.5.3) to conduct the simulation studies and the data analyses. The R code for our data analysis is provided at `https://github.com/lamorim-br/Mediation_survival-HBV.git`.

Nevertheless, it is important to mention some limitations to the proposed analytical framework. Firstly, assumptions regarding the survival part must

be carefully assessed according to the chosen survival model. For instance, the proportionality of the hazards under the Cox model should be evaluated. Besides that, the expressions for Cox models only apply under the rare-outcome assumption. Nevertheless, a weighting approach to handle common outcomes is described by VanderWeele (2015) and might be likewise extended to this general framework. Lange et al. (2012) also consider weighting each observation using a generalized marginal structural model, allowing the use of different link functions for the outcome (e.g., logistic model), and extended this framework for Cox and Aalen models. The assumptions for the GLM part should also be adressed accordingly. Secondly, the identification of the estimators are valid under stringent assumptions (A0)–(A4), which should be addressed to make causal claims. It is essential to highlight that we assumed the usual conditions of no unmeasured confoundness between exposure and outcome (A1), mediator and outcome (A2), and exposure and mediator (A3). The identification and measurement of confounders are required to draw defensible causal claims from non-experimental data, which depends on previous empirical evidence and solid knowledge of the theoretical mechanism. The theoretical and computational assessment of the efficiency and statistical consistency regarding the estimators of the natural direct and indirect effects is still a subject for future studies.

The development of approaches for causal mediation analysis for time-to-event outcomes is an active area of methodological research. We have discussed in this paper a particular case for mediation analysis for survival data when the mediator is measured somewhere between the exposure and outcome and may be described by means of a GLM model. Future extensions of our approach include mediation analysis for time-to-event data with multiple GLM mediators, which could allow estimation of path-specific effects through different mediators. Indeed, Miles et al. (2020) proposed a semiparametric procedure to assess path-specific effects in a different context, which points to further developments in the survival framework. We assumed no missing data and accurately measured variables, so that further developments may account for missing data and measurement errors. Additionaly, in many settings mediation analysis involves multiple mediators, time-varying causal effects, and time-varying exposures and mediators, see Didelez (2019), Fasanelli et al. (2019), Robins et al. (2000) and Vansteelandt et al. (2019). Extensions of our approach to circumstances involving multiple and time-varying mediators/confounders could be considered as well.

Acknowledgements

LDA was partly supported by a FIOTEC Fellowship (IGM-009-FEX-17).

[Received: March 2021 — Accepted: December 2021]

References

Aalen, O. O., Borgan, O. & Gjessing, H. K. (2008), *Survival and Event History Analysis: A Process Point of View*, Springer, New York.

Aalen, O., Røysland, K. & Gran, J. (2012), 'Causality, mediation and time: a dynamic viewpoint', *Journal of the Royal Statistical Society - Series A* **175**(4), 831–861.

Aalen, O., Stensrud, M., Didelez, V., Daniel, R., Røysland, K. & Strohmaier, S. (2020), 'Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model', *Biometrical Journal* **62**, 532–549.

Albert, J. & Nelson, S. (2011), 'Generalized Causal Mediation Analysis', *Biometrics* **67**, 1028–1038.

Baron, R. M. & Kenny, D. A. (1986), 'The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations', *Journal of Personality and Social Psychology* **51**, 1173–1182.

Bender, R., Augustin, T. & Blettner, M. (2005), 'Generating survival times to simulate Cox proportional hazards models', *Statistics in Medicine* **24**, 1713–1723.

Daniel, R., Stavola, B., Cousens, S. & Vansteelandt, S. (2015), 'Causal mediation analysis with multiple mediators', *Biometrics* **71**, 13–14.

Didelez, V. (2019), 'Defining causal mediation with a longitudinal mediator and a survival outcome', *Lifetime Data Analysis* **25**, 593–610.

Fasanelli, F., Giraudo, M., Ricceri, F., Valeri, L. & Zugna, D. (2019), 'Marginal Time-Dependent Causal Effects in Mediation Analysis With Survival Data', *American Journal of Epidemiology* **188**(5), 967–974.

Fulcher, I., Tchetgen Tchetgen, E. & Williams, P. (2017), 'Mediation Analysis for Censored Survival Data Under an Accelerated Failure Time Model', *Epidemiology* **28**(5), 660–666.

Huang, Y.-T., Jen, C.-L., Yang, H.-I., Lee, M.-H., Lu, S.-N., Iloeje, U. & Chen, C.-J. (2011), 'Lifetime risk and sex difference of hepatocellular carcinoma among patients with chronic hepatitis B and C', *Journal of Clinical Oncology* **29**(27), 3643–3650.

Huang, Y.-T. & Yang, H.-I. (2017), 'Causal mediation analysis of survival outcomes with multiple mediators', *Epidemiology* **28**(3), 370–378.

Iacobucci, D. (2012), 'Mediation analysis and categorical variables: the final frontier', *Journal of Consumer Psychology* **22**, 582–594.

Imai, K., Keele, L. & Tingley, D. (2010), 'A general approach to causal mediation analysis', *Psychological Methods* **15**, 309–334.

Kalbfleisch, J. D. & Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, 2 edn, Wiley, Hoboken, N.J.

Lange, T. & Hansen, J. (2011), 'Direct and indirect effects in a survival contex', *Epidemiology* **22**, 575–581.

Lange, T., Vansteelandt, S. & Bekaert, M. (2012), 'A simple unified approach for estimating natural direct and indirect effects', *American Journal of Epidemiology* **176**, 190–195.

Lin, S.-H., Young, J., Logan, R. & VanderWeele, T. (2017), 'Mediation analysis with a survival outcome with time-varying exposures, mediators and confounders', *Statistics in Medicine* **36**, 4153–4166.

Loyes, T., Moerkerke, B., De Smet, O., Byusse, A., S. & J. Vansteelandt, S. (2013), 'Flexible mediation analysis in the presence of nonlinear relations: beyond the mediation formula', *Multivariate Behavioral Research* **48**(6), 871–894.

MacKinnon, D. (2008), *Introduction to statistical mediation analysis*, Taylor and Francis, New York.

MacKinnon, D. P. & Dwyer, J. H. (1993), 'Estimating mediated effects in prevention studies', *Evaluation Review* **17**, 144–158.

MacKinnon, D., Warsi, G. & Dwyer, J. (1995), 'A simulation study of mediated effect measures', *Multivariate Behavioral Research* **30**, 41–62.

Miles, C., Shpitser, I., Kanki, P., Meloni, S. & Tchetgen Tchetgen, E. (2020), 'On semiparametric estimation of apath-specific effect in the presence of mediator-outcome confounding', *Biometrika* **107**(1), 159–172.

Pearl, J. (2001), 'Direct and indirect effects', *Proceedings of the 17th Annual Conference on Uncertainty in Artificial Intelligence (UAI-01)* pp. 411–442.

Pratschke, J., Haase, T., Comber, H., Sharp, L., Cancela, M. & Johnson, H. (2016), 'Mechanisms and mediation in survival analysis: towards an integrated analytical framework', *BMC Medical Research Methodology* pp. 16–27.

Robins, J., Hernán, M. & Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology', *Epidemiology* **11**, 550–560.

Smyth, G. (1989), 'Generalized linear models with varying dispersion', *Journal of the Royal Statistical Society: Series B* **51**(1), 47–60.

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

Tchetgen Tchetgen, E. & Shpitser, I. (2012), 'Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis', *Annals of statistics* **40**(3), 1816–1845.

VanderWeele, T. (2011), 'Causal mediation analysis with survival data', *Epidemiology* **22**(4), 582–585.

VanderWeele, T. (2015), *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press, New York.

VanderWeele, T. (2016), 'Mediation Analysis: a Practitioner's Guide', *Annual Review of Public Health* **37**, 17–32.

VanderWeele, T. & Vansteelandt, S. (2009), 'Conceptual issues concerning mediation, interventions and composition', *Statistics and its Interface* **2**, 457–468.

VanderWeele, T. & Vansteelandt, S. (2010), 'Odds ratios for mediation analysis with a dichotomous outcome', *American Journal of Epidemiology* **172**, 1339–1348.

Vansteelandt, S., Linder, M., Vandenberghe, S., Steen, J. & Madsen, J. (2019), 'Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding', *Statistics in Medicine* **38**, 4828–4840.

Appendix A.

Appendix A.1. Proof of formula (3) – Accelerated Failure Rate Model

Notice that

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = \mathrm{E}^{M_{a^*}}\mathrm{E}[T_{aM_{a^*}}|M_{a^*},\boldsymbol{x}] = \int \mathrm{E}[T_{aM_{a^*}}|M_{a^*}=m,\boldsymbol{x}]d\mathrm{P}^{M_{a^*}}(m|\boldsymbol{x}).$$

From (A3) and then (A1),

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = \int \mathrm{E}[T_{am}|m,\boldsymbol{x}]d\mathrm{P}^{M_{a^*}}(m|a^*,\boldsymbol{x}) = \int \mathrm{E}[T_{am}|a,m,\boldsymbol{x}]d\mathrm{P}^{M_{a^*}}(m|a^*,\boldsymbol{x}),$$

so that

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = \int \mathrm{E}[T|a,m,\boldsymbol{x}]d\mathrm{P}^{M}(m|a^*,\boldsymbol{x}).$$

Using the fact that the time to evento follows the accelerated failure rate model,

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = e^{\beta_0+\beta_1 a+\beta_4^\top \boldsymbol{x}}\mathrm{E}\left[e^{\gamma\varepsilon}\right]\int e^{(\beta_2+\beta_3 a)m}d\mathrm{P}^{M}(m|a^*,\boldsymbol{x}).$$

On the other hand, it is known that the moment generating function of M (given a^* and $\boldsymbol{x}$) is equal to

$$\mathrm{E}^{M}\left[e^{\tau M}|a^*,\boldsymbol{x}\right] = \exp\left\{\frac{b\left[\theta(a^*,\boldsymbol{x})+\tau\phi\right]-b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right\}, \tag{A1}$$

so that,

$$\mathrm{E}[T_{aM_{a^*}}|\boldsymbol{x}] = e^{\beta_0+\beta_1 a+\beta_4^\top \boldsymbol{x}}\mathrm{E}\left[e^{\gamma\varepsilon}\right]\exp\left\{\frac{b\left[\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a)\phi\right]-b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right\}.$$

Appendix A.2. Proof of formula (6 – Aalen Model)

From the proof of Proposition 4.7, Appendix A.4.4 (p. 502), VanderWeele (2015), if (A1)-(A4) hold, then, for $\delta > 0$,

$$\begin{aligned}&\mathrm{P}(T_{aM_{a^*}} \in (t, t+\delta]|T_{aM_{a^*}} \geq t, \boldsymbol{x})\\ &\quad = \int \mathrm{P}(T \in (t, t+\delta]|T \geq t, a, m, \boldsymbol{x})\frac{\mathrm{P}(T_{am} \geq t|\boldsymbol{x})}{I_1(t|a, a^*, \boldsymbol{x})}d\mathrm{P}^M(m|a^*, \boldsymbol{x}),\end{aligned}$$

where $I_1(t|a, a^*, \boldsymbol{x}) = \int \mathrm{P}(T_{am'} \geq t|\boldsymbol{x})d\mathrm{P}(m'|a^*, \boldsymbol{x})$. Now,

$$\begin{aligned}\mathrm{P}\{T_{am'} \geq t|\boldsymbol{x}\} &= \mathrm{P}\{T \geq t|a, m', \boldsymbol{x}\}\\ &= \exp\left\{B_0(t)+B_1(t)a+B_2(t)m'+B_3(t)am'+B_4(t)^\top \boldsymbol{x}\right\},\end{aligned}$$

where $B_j(t) = \int_0^t \beta_j(s)ds$, so that

$$\begin{aligned}&I_1(t|a, a^*, \boldsymbol{x})\\ &\quad = \int \exp\left\{B_0(t)+B_1(t)a+B_2(t)m'+B_3(t)am'+B_4(t)^\top \boldsymbol{x}\right\} d\mathrm{P}^M(m'|a^*, \boldsymbol{x})\end{aligned}$$

and hence

$$I_1(t|a, a^*, \boldsymbol{x}) = \exp\left\{B_0(t)+B_1(t)a+B_4(t)^\top \boldsymbol{x}\right\}\mathrm{E}\left[\exp\left\{(B_2(t)+B_3(t)a)M|a^*, \boldsymbol{x}\right\}\right].$$

Using the fact that $\lambda_{T_{aM_{a^*}}}(t) = \lim_{\delta\to 0}\mathrm{P}(t < T_{aM_{a^*}} \leq t+\delta|T_{aM_{a^*}} \geq t)/\delta$ can be written as

$$\lambda_{T_{aM_{a^*}}}(t) = \int \lim_{\delta\to 0}\frac{1}{\delta}\mathrm{P}(t < T_{aM_{a^*}} \leq t+\delta|T_{aM_{a^*}} \geq t, \boldsymbol{x})$$

and then applying the bounded convergence theorem, we have

$$\lambda_{T_{aM_{a^*}}}(t) = \int\int \frac{\lambda(t|a, m, \boldsymbol{x})\exp\left\{(B_2(t)+B_3(t)a)m\right\}}{\mathrm{E}\left[\exp\left\{(B_2(t)+B_3(t)a)M|a^*, \boldsymbol{x}\right\}\right]}d\mathrm{P}^M(m|a^*, \boldsymbol{x})d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}} \geq t).$$

Hence,

$$\lambda_{T_{aM_{a^*}}}(t) = \beta_0(t)+\beta_1(t)a+\int \beta_4(t)^\top \boldsymbol{x}d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}} \geq t)+(\beta_2(t)+\beta_3(t)a)I_2(t|a, a^*, \boldsymbol{x}),$$

where

$$I_2(t|a, a^*, \boldsymbol{x}) = \int \frac{\mathrm{E}\left[M\exp\left\{(B_2(t)+B_3(t)a)M\right\}|a^*, \boldsymbol{x}\right]}{\mathrm{E}\left[\exp\left\{(B_2(t)+B_3(t)a)M|a^*, \boldsymbol{x}\right\}\right]}d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}} \geq t).$$

Now, using the fact that the distribution of M is a member of the Exponential family (given a^* and $\boldsymbol{x}$),

$$\mathrm{E}\left[Me^{\tau M}|a^*,\boldsymbol{x}\right] = \exp\left\{\frac{b\left[\theta(a^*,\boldsymbol{x})+\tau\phi\right]-b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right\} b'\left[\theta(a^*,\boldsymbol{x})+\tau\phi\right],$$

so that

$$\frac{\mathrm{E}\left[Me^{\tau M}|a^*,\boldsymbol{x}\right]}{\mathrm{E}\left[e^{\tau M}|a^*,\boldsymbol{x}\right]} = b'\left[\theta(a^*,\boldsymbol{x})+\tau\phi\right],$$

and

$$I_2(t|a,a^*,\boldsymbol{x}) = \int b'\left[\theta(a^*,\boldsymbol{x})+(B_2(t)+B_3(t)a)\phi\right] d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}}\geq t).$$

On the one hand, using assumptions (A1)-(A4),

$$\mathrm{P}(T_{aM_{a^*}}\geq t|\boldsymbol{x}) = \int \mathrm{P}(T_{aM_{a^*}}\geq t|M_{a^*}=m,\boldsymbol{x})dP^{M_{a^*}}(m|\boldsymbol{x}),$$

so that

$$\mathrm{P}(T_{aM_{a^*}}\geq t|\boldsymbol{x}) = \int \mathrm{P}(T_{am}\geq t|m,\boldsymbol{x})dP^{M}(m|a^*,\boldsymbol{x}).$$

Therefore,

$$\mathrm{P}(T_{aM_{a^*}}\geq t|\boldsymbol{x}) = \mathrm{E}\left[S_T(t|a,M,\boldsymbol{x})|a^*,\boldsymbol{x}\right],$$

where $S_T(t|a,m,\boldsymbol{x})$ is the survival function of T. Besides that,

$$\mathrm{P}(T_{aM_{a^*}}\geq t) = \int\int \mathrm{P}(T_{aM_{a^*}}\geq t|M_{a^*}=m,\boldsymbol{x}')dP^{M_{a^*}}(m|\boldsymbol{x}')d\mathrm{P}^X(\boldsymbol{x}')$$

so that

$$\mathrm{P}(T_{aM_{a^*}}\geq t) = \int \mathrm{E}\left[S_T(t|a,M,\boldsymbol{x}')|a^*,\boldsymbol{x}'\right] d\mathrm{P}^X(\boldsymbol{x}').$$

Hence,

$$d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}}\geq t) = \frac{\mathrm{E}\left[S_T(t|a,M,\boldsymbol{x})|a^*,\boldsymbol{x}\right] d\mathrm{P}^X(\boldsymbol{x})}{\int \mathrm{E}\left[S_T(t|a,M,\boldsymbol{x}')|a^*,\boldsymbol{x}'\right] d\mathrm{P}^X(\boldsymbol{x}')}.$$

From (2),

$$\begin{aligned}&\mathrm{E}\left[S_T(t|a,M,\boldsymbol{x})|a^*,\boldsymbol{x}\right]\\ &\quad= \exp\left\{B_0(t)+B_1(t)a+B_4(t)^\top\boldsymbol{x}\right\}\mathrm{E}\left[\exp\left\{(B_2(t)+B_3(t)a)M\right\}|a^*,\boldsymbol{x}\right]\end{aligned}$$

and then using (A1),

$$\begin{aligned}&\mathrm{E}\left[S_T(t|a,M,\boldsymbol{x})|a^*,\boldsymbol{x}\right]\\ &= \exp\left\{B_0(t)+B_1(t)a+B_4(t)^\top\boldsymbol{x}+\frac{b\left[\theta^*(\boldsymbol{x})+(B_2(t)+B_3(t)a)\phi\right]-b\left[\theta^*(\boldsymbol{x})\right]}{\phi}\right\},\end{aligned}$$

so that

$$d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}}\geq t) = \frac{\exp\left\{B_4(t)^\top\boldsymbol{x}+\frac{b[\theta^*(\boldsymbol{x})+(B_2(t)+B_3(t)a)\phi]-b[\theta^*(\boldsymbol{x})]}{\phi}\right\} d\mathrm{P}^X(\boldsymbol{x})}{\int \exp\left\{B_4(t)^\top\boldsymbol{x}'+\frac{b[\theta^*(\boldsymbol{x}')+(B_2(t)+B_3(t)a)\phi]-b[\theta^*(\boldsymbol{x}')]}{\phi}\right\} d\mathrm{P}^X(\boldsymbol{x}')}.$$

Therefore, if we write for notational convenience $d\mathrm{P}^X(\boldsymbol{x}|T_{aM_{a^*}} \geq t) \equiv dQ_t(\boldsymbol{x}|a,a^*)$, then

$$I_2(t|a,a^*,\boldsymbol{x}) = \int b'\left[\theta(a^*,\boldsymbol{x}) + (B_2(t)+B_3(t)a)\phi\right] dQ_t(\boldsymbol{x}|a,a^*).$$

and

$$\lambda_{T_{aM_{a^*}}}(t) = \beta_0(t)+\beta_1(t)a+\int \beta_4(t)^\top \boldsymbol{x} dQ_t(\boldsymbol{x}|a,a^*)+(\beta_2(t)+\beta_3(t)a)I_2(t|a,a^*,\boldsymbol{x}),$$

or

$$\begin{aligned}\lambda_{T_{aM_{a^*}}}(t) =& \beta_0(t)+\beta_1(t)a \\ &+ \mathrm{E}_t^X\left[\beta_4(t)^\top \boldsymbol{X} + (\beta_2(t)+\beta_3(t)a)\mathrm{E}_{\theta(a^*,\boldsymbol{X})+(B_2(t)+B_3(t)a)\phi}M|a,a^*\right].\end{aligned}$$

Appendix A.3. Proof of formula (10 – Cox Model)

Assuming the Cox model, we know that

$$\lambda_{T_aM_{a^*}}(t|\boldsymbol{x}) = \lambda_o(t)\exp\left\{\beta_1 a + \beta_4^\top \boldsymbol{x}\right\} r(t|a,a^*,\boldsymbol{x}),$$

where

$$r(t|a,a^*,\boldsymbol{x}) = \frac{\int e^{(\beta_2+\beta_3 a)m}\exp\left\{-\Lambda_o(t)e^{\beta_1 a+\beta_4^\top \boldsymbol{x}+(\beta_2+\beta_3 a)m}\right\} dP(m|a^*,\boldsymbol{x})}{\int \exp\left\{-\Lambda_o(t)e^{\beta_1 a+\beta_4^\top \boldsymbol{x}+(\beta_2+\beta_3 a)m}\right\} dP(m|a^*,\boldsymbol{x})},$$

see VanderWeele (2015), Proposition 4.4, p. 496. Now, if the events are rare, i.e. $\Lambda_o(t) \approx 0$, and using (A1), then

$$r(t|a,a^*,\boldsymbol{x}) \approx \mathrm{E}\left[e^{(\beta_2+\beta_3 a)M}|a^*,\boldsymbol{x}\right] = \exp\left\{\frac{b\left[\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a)\phi\right]-b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right\}.$$

Hence,

$$\lambda_{T_aM_{a^*}}(t|\boldsymbol{x}) \approx \lambda_o(t)\exp\left\{\beta_1 a+\beta_4^\top \boldsymbol{x}+\frac{b\left[\theta(a^*,\boldsymbol{x})+(\beta_2+\beta_3 a)\phi\right]-b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right\}.$$

Appendix B. Binomial and Poisson Distributions for M

Binomial Mediator

Writing $g(\pi) = \log\{\pi/(1-\pi)\}$, $\phi = 1$ and $b(\theta) = m\log(1+e^\theta)$, we have the following cases:

- AFT and Cox Models: using (4) and (5) (or (11) and (12)), it follows that

$$\mathrm{NIE}_{\mathrm{Cox}} \approx \mathrm{NIE}_{\mathrm{AFT}} = m\log\frac{(1+e^{\theta(a,\boldsymbol{x})+\beta_2+\beta_3 a})/(1+e^{\theta(a,\boldsymbol{x})})}{(1+e^{\theta(a^*,\boldsymbol{x})+\beta_2+\beta_3 a})/(1+e^{\theta(a^*,\boldsymbol{x})})},$$

and

$$\text{NDE}_{\text{Cox}} \approx \text{NDE}_{\text{AFT}} = \beta_1(a-a^*) + m\log\frac{1+e^{\theta(a^*,\boldsymbol{x})+\beta_2+\beta_3 a}}{1+e^{\theta(a^*,\boldsymbol{x})+\beta_2+\beta_3 a^*}}.$$

- Aalen Model: using (8) and (9), we get

$$\text{NIE}_{\text{Aalen}} = \text{E}_t^X\left[\frac{m(\beta_2(t)+\beta_3(t)a)}{e^{-(B_2(t)+B_3(t)a)+\theta(a,\boldsymbol{X}))}+1} - \frac{m(\beta_2(t)+\beta_3(t)a)}{e^{-(B_2(t)+B_3(t)a)+\theta(a^*,\boldsymbol{X}))}+1}\middle| a, a^*\right]$$

and $\text{NDE}_{\text{Aalen}} = \beta_1(t)(a-a^*) + \mathcal{E}(t;a,a^*) - \mathcal{E}(t;a^*,a^*)$, with

$$\mathcal{E}(t;a,a^*) = \text{E}_t^X\left[\beta_4(t)^\top \boldsymbol{X} + \frac{m(\beta_2(t)+\beta_3(t)a)}{e^{-(B_2(t)+B_3(t)a+\theta(a^*,\boldsymbol{X}))}+1}\middle| a, a^*\right]$$

and $dQ_t(\boldsymbol{x}|a,a^*) \propto \exp\left\{B_4(t)^\top \boldsymbol{x}\right\}\left[\frac{1+e^{\theta(a^*,\boldsymbol{x})+B_2(t)+B_3(t)a}}{1+e^{\theta(a^*,\boldsymbol{x})}}\right]^m d\text{P}^X(\boldsymbol{x})$.

Poisson Mediator

Writing $g(\mu) = \log\mu$, $\phi = 1$ and $b(\theta) = e^\theta$, we have the following cases:

- AFT and Cox Models: using (4) and (5) (or (11) and (12)), it follows that

$$\text{NIE}_{\text{Cox}} \approx \text{NIE}_{\text{AFT}} = \left(e^{\beta_2+\beta_3 a}-1\right)\left(e^{\theta(a,\boldsymbol{x})} - e^{\theta(a^*,\boldsymbol{x})}\right),$$

and

$$\text{NDE}_{\text{Cox}} \approx \text{NDE}_{\text{AFT}} = \beta_1(a-a^*) + e^{\beta_2+\theta(a^*,\boldsymbol{x})}\left(e^{\beta_3 a} - e^{\beta_3 a^*}\right).$$

- Aalen Model: from (8) and (9), we have

$$\text{NIE}_{\text{Aalen}} = (\beta_2(t)+\beta_3(t)a)e^{B_2(t)+B_3(t)a}\text{E}_t^X\left[e^{\theta(a,\boldsymbol{X})} - e^{\theta(a^*,\boldsymbol{X})}|a,a^*\right]$$

and $\text{NDE}_{\text{Aalen}} = \beta_1(t)(a-a^*) + \mathcal{E}(t;a,a^*) - \mathcal{E}(t;a^*,a^*)$, with

$$\mathcal{E}(t;a,a^*) = \text{E}_t^X\left[\beta_4(t)^\top \boldsymbol{X} + (\beta_2(t)+\beta_3(t)a)e^{B_2(t)+B_3(t)a+\theta(a^*,\boldsymbol{X})}|a,a^*\right]$$

and $dQ_t(\boldsymbol{x}|a,a^*) \propto \exp\left\{B_4(t)^\top \boldsymbol{x} + e^{\theta(a^*,\boldsymbol{x})}\left(e^{B_2(t)+B_3(t)a}-1\right)\right\} d\text{P}^X(\boldsymbol{x})$.

Appendix C. Responsiveness Measures for the Aalen's Model

Write the density in (7) as $q_{t,a,a^*}(\boldsymbol{x}) = \varphi_{t,a,a^*}(\boldsymbol{x})/\Phi_{t,a,a^*}$, where

$$\varphi_{t,a,a^*}(\boldsymbol{x}) = \exp\left[B_4(t)^\top \boldsymbol{x} + \frac{b\left[\theta(a^*,\boldsymbol{x}) + (B_2(t)+B_3(t)a)\phi\right] - b\left[\theta(a^*,\boldsymbol{x})\right]}{\phi}\right]$$

and $\Phi_{t,a,a^*} = \int \varphi_{t,a,a^*}(\boldsymbol{x}')d\mathrm{P}^X(\boldsymbol{x})$. Then, $\partial_a\varphi_{t,a,a}(\boldsymbol{x}) = \phi^{-1}\tilde{\psi}_{1,t}(a,\boldsymbol{x})\varphi_{t,a,a}(\boldsymbol{x})$, where

$$\tilde{\psi}_{1,t}(a,\boldsymbol{x}) = (\theta'(a,\boldsymbol{x}) + B_3(t)\phi)\mathrm{E}_{\phi(B_2+B_3a)+\theta(a,\boldsymbol{x})}M - \theta'(a,\boldsymbol{x})\mathrm{E}_{\theta(a,\boldsymbol{x})}M$$

and $\partial_a\Phi_{t,a,a} = \phi^{-1}\int \psi_{0,t}(a,\boldsymbol{x}')\varphi_{t,a,a}(\boldsymbol{x}')d\mathrm{P}^X(\boldsymbol{x}')$, so that

$$\frac{\partial_a\Phi_{t,a,a}}{\Phi_{t,a,a}} = \frac{1}{\phi}\mathrm{E}^X_{t,a}\tilde{\psi}_{1,t}(a,\boldsymbol{X}),$$

with $\mathrm{E}^X_{t,a} = \mathrm{E}^X_{t,a,a}$. Hence, $\partial_a q_{t,a,a}(\boldsymbol{x}) = \phi^{-1}\psi_{1,t}(a,\boldsymbol{x})q_{t,a,a}(\boldsymbol{x})$, with

$$\psi_{1,t}(a,\boldsymbol{x}) = \tilde{\psi}_{1,t}(a,\boldsymbol{x}) - \mathrm{E}^X_{t,a}\tilde{\psi}_{1,t}(a,\boldsymbol{X}).$$

From (6), we have

$$\mathrm{TR}_{\mathrm{Aalen}} = \partial_a\lambda_{T_{aM_a}}(t)|_{a=a^*} = \beta_1(t) + \Xi_1 + \Xi_2 + \Xi_3, \tag{A2}$$

where $\Xi_1 = \beta_3(t)\mathrm{E}^X_{t,a^*}\mathrm{E}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M$,

$$\Xi_2 = \phi^{-1}(\beta_2(t) + \beta_3(t)a^*)\mathrm{E}^X_{t,a^*}(\phi B_3(t) + \theta'(a^*,\boldsymbol{X}))\mathrm{Var}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M,$$

and

$$\Xi_3 = \phi^{-1}\mathrm{E}^X_{t,a^*}\left[\beta_4(t)^\top\boldsymbol{X} + (\beta_2(t) + \beta_3(t)a^*)\mathrm{E}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M\right]\psi_{1,t}(a^*,\boldsymbol{X}).$$

Similarly, $\partial_a\varphi_{t,a,a^*}(\boldsymbol{x}) = \tilde{\psi}_{2,t}(a,\boldsymbol{x})\varphi_{t,a,a^*}(\boldsymbol{x})$, where

$$\tilde{\psi}_{2,t}(a,\boldsymbol{x}) = B_3(t)\mathrm{E}_{\theta(a^*,\boldsymbol{x})+(B_2+B_3a)\phi}M$$

and $\partial_a\Phi_{t,a,a^*} = \int \tilde{\psi}_{2,t}(a,\boldsymbol{x}')\varphi_{t,a,a^*}(\boldsymbol{x}')d\mathrm{P}^X(\boldsymbol{x}')$. Hence,

$$\frac{\partial_a\Phi_{t,a,a^*}}{\Phi_{t,a,a^*}} = \mathrm{E}^X_{t,a}\tilde{\psi}_{2,t}(a,\boldsymbol{X}),$$

so that $\partial_a q_{t,a,a^*}(\boldsymbol{x}) = \psi_{2,t}(a,\boldsymbol{x})q_{t,a,a}(\boldsymbol{x})$, where

$$\psi_{2,t}(a,\boldsymbol{x}) = \tilde{\psi}_{2,t}(a,\boldsymbol{x}) - \mathrm{E}^X_{t,a}\tilde{\psi}_{2,t}(a,\boldsymbol{X})$$

and, using (6) again,

$$\mathrm{NDR}_{\mathrm{Aalen}} = \partial_a\lambda_{T_{aM_{a^*}}}(t|\boldsymbol{x})|_{a=a^*} = \beta_1(t) + \Xi_1 + \Xi'_2 + \Xi'_3, \tag{A3}$$

where

- $\Xi'_2 = (\beta_2(t) + \beta_3(t)a^*)B_3(t)\mathrm{E}^X_{t,a}\mathrm{Var}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M$;
- $\Xi'_3 = \mathrm{E}^X_{t,a^*}\left[\beta_4(t)^\top\boldsymbol{X} + (\beta_2(t) + \beta_3(t)a^*)\mathrm{E}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M\right]\psi_{2,t}(a^*,\boldsymbol{X})$.

Finally,

$$\text{NIR}_{\text{Aalen}} = \text{TR}_{\text{Aalen}} - \text{NDR}_{\text{Aalen}} = (\Xi_2 - \Xi_2') + (\Xi_3 - \Xi_3'), \tag{A4}$$

where

$$\Xi_2 - \Xi_2' = \phi^{-1}(\beta_2(t) + \beta_3(t)a^*)\text{E}_{t,a}^X \theta'(a^*, \boldsymbol{X})\text{Var}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M$$

and

$$\Xi_3 - \Xi_3' = \text{E}_{t,a^*}^X \left[\beta_4(t)^\top \boldsymbol{X} + (\beta_2(t) + \beta_3(t)a^*)\text{E}_{\phi(B_2(t)+B_3(t)a^*)+\theta(a^*,\boldsymbol{X})}M\right] \psi_{3,t}(a^*, \boldsymbol{X})$$

with $\psi_{3,t}(a^*, \boldsymbol{x}) = \phi^{-1}\psi_{1,t}(a^*, \boldsymbol{x}) - \psi_{2,t}(a^*, \boldsymbol{x})$.

Appendix D. Additional results for causal mediation analysis

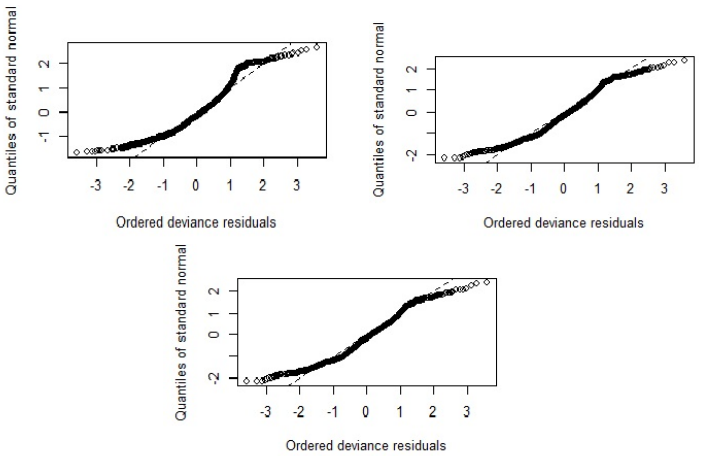


FIGURE 6: Deviance residual analysis when assuming (a) Normal, (b) Gamma with log link and (c) Gamma with inverse link distributions.