Published in final edited form as:
Epidemiology. 2011 July ; 22(4): 582–585. doi:10.1097/EDE.0b013e31821db37e.

Causal mediation analysis with survival data

Tyler J. VanderWeele
Departments of Epidemiology and Biostatistics, Harvard School of Public Health

Abstract

Causal mediation analysis is considered for time-to-event outcomes and survival analysis models. Different possible effect decompositions are discussed for the survival function, hazard, mean survival time and median survival scales. Approaches to mediation analysis in the social sciences are related to counterfactual approaches using additive hazard, proportional hazard and accelerated failure time models. The product-coefficient method from the social sciences gives mediated effects on the hazard difference scale for additive hazard models, on the log mean survival time difference scale for accelerated failure time models, and on the log hazard scale for the proportional hazards model but only if the outcome is rare. With the proportional hazards model and a common outcome, the product-coefficient method can provide a valid test for the presence of a mediator effect but does not provide a measure. When additive hazard, accelerated failure time, or the rare-outcome proportional hazards models are employed and combined with the counterfactual approach, exposure-mediator interactions can be accommodated in a relatively straightforward manner.

Introduction

In the last few years there have been a number of papers developing methods for mediation analysis from a counterfactual perspective, building on some of the original insights of Robins and Greenland[1] and Pearl.[2] Until the paper by Lange and Hansen,[3] in this issue of Epidemiology, there has not, however, been any work addressing the survival-analysis setting from the perspective of causal inference. Using an additive hazard model, Lange and Hansen[3] have provided a useful flexible method to analyze direct and indirect effects for time-to-event data. Here, I would like to discuss different effect measures of interest when direct and indirect effects in survival analysis are in view, show how an approach similar to that of Lange and Hansen[3] is possible for a proportional hazards with a rare outcome or accelerated failure time models generally, and relate these ideas to previous work on mediation with survival data published in the social science literature.[4]

Concepts and Definitions

We will let A denote an exposure of interest, T a time-to-event outcome, M a mediator and C a set of covariates. We will let T_a denote the counterfactual event time if A had been set to a; likewise we let T_{am} denote the counterfactual event time if A had been set to a and M had been set to m. We let M_a be the counterfactual value of the mediator if A had been set to a. We restrict our attention here to the setting of a single event, rather than considering

Corresponding Author: Tyler J. VanderWeele Harvard School of Public Health Departments of Epidemiology and Biostatistics 677 Huntington Avenue Boston MA 02115 Phone: 617-432-7855 Fax: 617-432-1884 tvanderw@hsph.harvard.edu.

multiple events as in Lange and Hansen.[3] With these definitions we can also consider nested counterfactual event times. For example, $T_{aM_{a^*}}$ is an individual's event time if the exposure had been set to a and the mediator had been set to the level it would have been had exposure been a^*. We assume composition,[5] that $T_a = T_{aM_a}$. For an arbitrary time-to-event variable V we will let $S_V(t)$ denote the survival function at time t, that is $S_V(t) = P(V > t)$; the survival function conditional on covariates $C = c$ can likewise be defined as $S_V(t|c) = P(V > t|c)$. We will use $\lambda_V(t)$ and $\lambda_V(t|c)$ for the hazard or conditional hazard at time t, that is the instantaneous rate of the event conditional on $V \geq t$.

An interesting feature of survival data within the context of mediation analysis is that there are multiple ways or scales by which we might decompose a total effect comparing exposure levels a and a^* into direct and indirect effects. For example, if we were to consider the survival functions, we could decompose a comparison of the survival functions $S_{T_a}(t)$ and $S_{T_{a^*}}(t)$ as follows:

$$S_{T_a}(t) - S_{T_{a^*}}(t) = \left[S_{T_{aM_a}}(t) - S_{T_{aM_{a^*}}}(t)\right] + \left[S_{T_{aM_{a^*}}}(t) - S_{T_{a^*M_{a^*}}}(t)\right]$$

where the first expression in brackets is the natural indirect effect on the survival function scale and the second is the natural direct effect on the survival function scale. We could alternatively but similarly decompose the overall difference in hazards as the sum of natural indirect and direct effects on the hazard scale:

$$\lambda_{T_a}(t) - \lambda_{T_{a^*}}(t) = \left[\lambda_{T_{aM_a}}(t) - \lambda_{T_{aM_{a^*}}}(t)\right] + \left[\lambda_{T_{aM_{a^*}}}(t) - \lambda_{T_{a^*M_{a^*}}}(t)\right].$$

Both of these measures, along with a cumulative hazard effect decomposition, were considered by Lange and Hansen.[3] We could, however, also consider other effect decompositions. We could, for example, consider a decomposition in terms of mean survival times:

$$E(T_a) - E(T_{a^*}) = \left[E\left(T_{aM_a}\right) - E\left(T_{aM_{a^*}}\right)\right] + \left[E\left(T_{aM_{a^*}}\right) - E\left(T_{a^*M_{a^*}}\right)\right].$$

Or if we let let Q_a and Q_{am} denote the median counterfactual survival time if A had been set to a or if A had been set to a and M had been set to m, respectively, then we have the decomposition:

$$Q_a - Q_{a^*} = \left[Q_{aM_a} - Q_{aM_{a^*}}\right] + \left[Q_{aM_{a^*}} - Q_{a^*M_{a^*}}\right].$$

One could also consider using the difference in log-survival function, or log-hazards, or log-expected survival times, etc. For example, with log-hazard one has the decomposition:

$$\log\left\{\lambda_{T_a}(t)\right\} - \log\left\{\lambda_{T_{a^*}}(t)\right\} = \left[\log\left\{\lambda_{T_{aM_a}}(t)\right\} - \log\left\{\lambda_{T_{aM_{a^*}}}(t)\right\}\right] + \left[\log\left\{\lambda_{T_{aM_{a^*}}}(t)\right\} - \log\left\{\lambda_{T_{a^*M_{a^*}}}(t)\right\}\right]$$

which exponentiating can also be written:

$$\lambda_{T_a}(t)/\lambda_{T_{a^*}}(t)=\left[\lambda_{T_{aM_a}}(t)/\lambda_{T_{aM_{a^*}}}(t)\right]\times\left[\lambda_{T_{aM_{a^*}}}(t)/\lambda_{T_{a^*M_{a^*}}}(t)\right]$$

so that the hazard ratio is the product of the natural indirect and direct effect hazard ratios. All of the above measures could also be considered conditional on strata of covariates $C = c$. With each of these potential decompositions on the difference scale, one could calculate a ”proportion mediated” by taking a ratio of the natural indirect effect to the sum of the natural direct and indirect effects (i.e. the total effect). These measures of the proportion mediated may vary across scales. Also, depending on the specific survival model, the natural direct and indirect effects may be analytically tractable on certain scales but not on others.

Irrespective of the decomposition chosen, however, certain fairly strong no-unmeasured-confounding assumptions need to be made. Following an identification approach initiated by Pearl[2] and used by subsequent authors on mediation,[5−8] Lange and Hansen[3] make four assumptions about no confounding conditional on the covariates. These can essentially be stated as that, conditional on covariates, there is (i) no confounding for the exposure-outcome relationship, (ii) no confounding for the mediator-outcome relationship, (iii) no confounding for the exposure-mediator relationship, and (iv) no mediator-outcome confounder that is an effect of the exposure. These are assumptions (A.1)-(A.4) in Lange and Hansen, and we likewise assume that they hold here. Sensitivity analysis techniques for direct and indirect effects can be useful when these assumptions do not hold.[9,10]

Mediation with an Additive Hazard Model

Lange and Hansen[3] present an approach to mediation analysis with survival data using an additive hazard model. In the most basic form they consider, the model can be written as:

$$\lambda_T(t|a,m,c)=\lambda_0+\lambda_1 a+\lambda_2' c+\lambda_3 m. \tag{1}$$

They propose a linear regression model for the mediator, when it is continuous, with normally distributed error:

$$E[M|a,c]=\beta_0+\beta_1 a+\beta_2' c. \tag{2}$$

They proceed to show that on the hazard scale, natural direct and indirect effects are given by:

$$\begin{aligned}\lambda_{T_{aM_a}}(t)-\lambda_{T_{aM_{a^*}}}(t) &= \beta_1\lambda_3(a-a^*)\\ \lambda_{T_{aM_{a^*}}}(t)-\lambda_{T_{a^*M_{a^*}}}(t) &= \lambda_1(a-a^*)\end{aligned}$$

where the first expression is the indirect effect and the second the direct effect on the hazard scale.

The use of the coefficient λ_1 for the exposure in the model for the outcome as the direct effect, and the product of the coefficient for the exposure in the model for the mediator times the coefficient for the mediator in the model for the outcome ($\lambda_3\beta_1$) as a measure of the indirect effect, has a long history in the social sciences.[11,12] The causal inference literature has clarified the assumptions needed to interpret these measures as causal direct and indirect effects,[2,5,8] e.g. assumptions (i)-(iv) above. The causal inference literature has also given formal counterfactual definitions of these effects, and has extended the notions of direct and indirect effects to much more general settings. Lange and Hansen[3] have shown how these notions extend further to survival data and have provided a model - the additive hazards models - under which the traditional social science direct and indirect coefficient measures hold.

However, the paper of Lange and Hansen[3] goes much further than this. Their approach allows the hazard functions to vary over time, allows for the possibility of multiple types of events, and could be extended to incorporate exposure-mediator interactions as well. The generality of the approach proposed is impressive, and the methodology and software provided will certainly be of use for causal mediation analysis within a survival context. Additive hazard models are not employed with great frequency in the epidemiologic literature, but the paper by Lange and Hansen demonstrates their potential utility and perhaps should give epidemiologists reason to rethink their choice of survival analysis models.

Mediation with Accelerated Failure Time and Proportional Hazards Models

The survival analysis models most frequently employed in the epidemiologic and social science literatures are probably, first, the proportional hazards model, and, second, accelerated failure time models. The possibility of conducting mediation analysis with survival data under both models was in fact considered in a paper by Tein and MacKinnon[4] in the social science literature some years ago. There have traditionally been two methods for undertaking mediation analysis. The ”difference method”,[13] which is more common in epidemiology, considers an outcome model both with and without the mediator, and takes the difference in the coefficients for the exposure as the measure of the indirect or mediated effect. The ”product method”,[11] more common in the social sciences, takes as a measure of the indirect effect the product of the coefficient for the exposure in the model for the mediator (i.e. β_1 in model (2)) and the coefficient for the mediator in the model for the outcome. If the outcome and mediator are continuous and there are no interactions in the model for the outcome, then the two methods coincide.[8,14] However, with binary outcomes the two methods may diverge[8,15]; they will approximately coincide when the binary outcome is rare.[8]

Tein and MacKinnon[4] consider whether the two approaches coincide with proportional hazards and accelerated failure time models. They effectively use model (2) for the mediator and use

$$\lambda_T\,(t|a,m,c)=\lambda_T\,(t|0,0,0)\,e^{\gamma_1 a+\gamma_2 m+\gamma_4' c}. \tag{3}$$

for the proportional hazard model and

$$\log(T)=\theta_0+\theta_1 A+\theta_2 M+\theta_4' c+\nu\varepsilon. \tag{4}$$

for the accelerated failure time model where ε is a random variable following an extreme value distribution and v is a scale parameter so that T follows a Weibull distribution. Using simulations, Tein and MacKinnon find that the difference-method and product-method give different results for the proportional hazards model but the same results for the accelerated failure time model. Their results raise the question of whether either of these methods for either of the models has a clear causal interpretation. Lange and Hansen[3] have given a rigorous causal interpretation for the parameters of an additive hazard model. Do similar results hold for the proportional hazards or accelerated failure time models?

Let us first consider the accelerated failure time model. We note first that it is no coincidence that the product- and difference- methods coincide for the accelerated failure time model in (4). In the eAppendix (http://links.lww.com) we give an analytic proof that this is so, provided that the models are correctly specified and that there are no interactions in model (4); the result holds for arbitrary distributions of ε in model (4) i.e. not just Weibull models. We moreover show in the eAppendix that the measures of direct and indirect effects obtained by these methods are the natural direct and indirect effects on the mean survival time scale. That is, the natural direct effect $\log\{E(T_{aM_{a*}})\} - \log\{E(T_{a*M_{a*}})\}$ is equal to $\theta_1(a - a*)$ and the natural indirect effect $\log\{E(T_{aM_a})\} - \log\{E(T_{aM_{a*}})\}$ is equal to $\beta_1\theta_2(a - a*)$. In other words, we once again obtain the result that exposure-coefficient in model (4) for the outcome is a measure of the direct effect, and the product of the exposure-coefficient in model (2) for the mediator times the mediator-coefficient in model (4) for the outcome is a measure of the indirect effect. In fact, for the accelerated failure time model, these analytic expressions can be extended so as to allow for exposure-mediator interaction in model (4). Suppose we extended model (4) to allow for such interaction:

$$\log(T) = \theta_0 + \theta_1 A + \theta_2 M + \theta_3 AM + \theta_4' c + \nu\varepsilon. \tag{5}$$

If model (5) holds for the outcome and model (2) holds for the mediator, then natural direct and indirect effects on the log mean survival time scale conditional on $C = c$ are given by:

$$\begin{aligned}
\log\left\{E\left(T_{aM_a}|c\right)\right\} - \log\left\{E\left(T_{aM_{a^*}}|c\right)\right\} &= (\theta_2\beta_1 + \theta_3\beta_1 a)\,(a - a^*)\\
\log\left\{E\left(T_{aM_{a^*}}|c\right)\right\} - \log\left\{E\left(T_{a^*M_{a^*}}|c\right)\right\} &= \left\{\theta_1 + \theta_3\left(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2\sigma^2\right)\right\}(a - a^*) + 0.5\theta_3^2\sigma^2\left(a^2 - a^{*2}\right)
\end{aligned}$$

where the first expression is the natural indirect effect and the second expression is the natural direct effect, and where σ^2 is the variance of the error term in regression model (2) for the mediator. These results hold for arbitrary distributions for ε in model (5) but do require a normally distributed mediator in model (2). Note that when there is no interaction ($\theta_3 = 0$), the expressions reduce to those given above and considered by Tein and MacKinnon.[4] The expressions given here for the accelerated failure time model are analogous to those given by VanderWeele and Vansteelandt[8] for odds ratios for mediation analysis for a dichotomous outcome. Expressions for standard errors for these direct and indirect effects could likewise be adapted from VanderWeele and Vansteelandt.[8]

Let us now turn to the proportional hazards model in (3). With the proportional hazards model, somewhat analogous results can be obtained, but only when the outcome is rare. Specifically, consider an extension to model (3) which allows for exposure-mediator interaction:

$$\lambda_T(t|a, m, c) = \lambda_T(t|0, 0, 0)\, e^{\gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma_4' c}. \tag{6}$$

If model (6) holds for the outcome and model (2) holds for the mediator then we show in the eAppendix (http://links.lww.com), using arguments similar to those in Lin et al.,[16] that, provided the outcome is rare, natural direct and indirect effects on the log hazard ratio difference scale are given by:

$$\begin{aligned}\log\left\{\lambda_{T_{aM_a}}(t|c)\right\}-\log\left\{\lambda_{T_{aM_{a^*}}}(t|c)\right\} &= (\gamma_2\beta_1+\gamma_3\beta_1 a)(a-a^*)\\ \log\left\{\lambda_{T_{aM_{a^*}}}(t|c)\right\}-\log\left\{\lambda_{T_{a^*M_{a^*}}}(t|c)\right\} &= \left\{\gamma_1+\gamma_3\left(\beta_0+\beta_1 a^*+\beta_2' c+\gamma_2\sigma^2\right)\right\}(a-a^*)+0.5\gamma_3^2\sigma^2\left(a^2-a^{*2}\right)\end{aligned}$$

where σ^2 is again the variance of the error term in regression model (2) for the mediator. The expressions are likewise analogous to those obtained by VanderWeele and Vansteelandt[8] for a dichotomous outcome, but these expressions only apply for a rare outcome. Natural indirect and direct effect hazard ratios can be obtained by exponentiating the right hand side of the equalities. We moreover show in the eAppendix that when there is no exposure-mediator interaction as in model (3), and when the outcome is rare, then the product- and difference- methods will coincide approximately.

In the general setting (with non-rare outcome), unfortunately, neither the product-method or the difference-method for the proportional hazards model have any sort of clear causal interpretation as a measure of effect. Tein and MacKinnon[4] show that the product- and difference- methods can diverge, and that they may even be of opposite signs! Lange and Hansen[3] noted that in the general setting (i.e. common outcome) with the proportional hazards model, natural direct and indirect effects do not have any simple analytic expression. We do nevertheless show in the eAppendix that even if the outcome is common, the product method using models (2) and (3) at least provides a valid test for whether there is any mediated effect, provided the models are correctly specified and that assumptions (i)-(iv) hold. With the proportional hazards model and a common outcome, the product method can thus be useful at least in testing the hypothesis of any mediated effect. But neither the product- nor the difference- method should in general be used as a measure of an indirect effect. Indeed, Hafeman[17] has recently demonstrated the danger of using such measures in non-linear models; they can result in weighted averages of causal effects in which the weights do not in fact sum to one.

Conclusion

The discussion above has provided expressions for natural direct and indirect effects for the accelerated failure time model and for the proportional hazards model when the outcome is rare. A major contribution of the counterfactual approach to causal mediation analysis has been to clarify the no-confounding assumptions required for the identification of direct and indirect effects. Within the context of survival data, the counterfactual approach also clarifies when different methods for direct and indirect effects can be interpreted as measures of effect rather than simply as a test for a mediated effect. The causal inference approach clarifies further on what scale these measures apply when they can be so interpreted. The observations of Tein and MacKinnon[4] have been given a more rigorous formulation and the approach has been extended to allow for exposure-mediator interactions.

Because the proportional hazards model is commonly used in epidemiologic research, the development of methodology and causal direct and indirect effect measures that can be used in conjunction with the model when the outcome is common may be an important direction for future research. However, the additive hazard model employed by Lange and Hansen[3] constitutes an important and very general alternative to mediation analysis with survival data.

Acknowledgments

Funding: Supported by National Institutes of Health grant HD060696.

eAppendix for “Causal mediation analysis with survival data” by TJ VanderWeele

Equivalence of Product and Difference Method for the Accelerated Failure Time Model with No Exposure-Mediator Interaction

Suppose that the model (2) in the text for the mediator is correctly specified:

$$E\left[M|a,c\right]=\beta_0+\beta_1 a+\beta_2^{'}c. \tag{2}$$

along with model (4) for the outcome with the exposure and mediator so that:

$$\log\left(T\right)=\theta_0+\theta_1 A+\theta_2 M+\theta_4^{'}C+\nu\varepsilon. \tag{4}$$

Suppose also a model is fit for the outcome with just the exposure, not the mediator:

$$\log\left(T\right)=\phi_0+\phi_1 A+\phi_4^{'}c+\varkappa\varepsilon.$$

The difference method uses $\phi_1 - \theta_1$ as a measure of the indirect effect; the product method uses $\beta_1\theta_2$. We show that if all of the models are correctly specified these two are equal. This is because by the model for the outcome without the mediator we have:

$$E\left[\log\left(T\right)|a,c\right]=\phi_0+\phi_1 a+\phi_4^{'}c+\varkappa E\left[\varepsilon\right]$$

and by model (4) we have:

$$\begin{aligned}
E\left[\log\left(T\right)|a,c\right] &= E\left[E\left[\log\left(T\right)|a,M,c\right]\right]\\
&= \theta_0+\theta_1 a+\theta_2 E\left[M|a,c\right]+\theta_4^{'}c+\nu E\left[\varepsilon\right]\\
&= \theta_0+\theta_1 a+\theta_2\left\{\beta_0+\beta_1 a+\beta_2^{'}c\right\}+\theta_4^{'}c+\nu E\left[\varepsilon\right]\\
&= \left\{\theta_0+\theta_2\beta_0\right\}+\left\{\theta_1+\theta_2\beta_1\right\}a+\left\{\theta_4^{'}+\theta\beta_2^{'}\right\}c+\nu E\left[\varepsilon\right]
\end{aligned}$$

Because this holds for all *a*, we must have $\phi_1 = \{\theta_1 + \theta_2\beta_1\}$ and thus $\phi_1 - \theta_1 = \theta_2\beta_1$.

Formulas for Natural Direct and Indirect Effects for the Accelerated Failure Time Model with An Exposure-Mediator Interaction

Under model (2) for the mediator and model (5) for the outcome:

$$\log\left(T\right)=\theta_0+\theta_1 A+\theta_2 M+\theta_3 AM+\theta_4^{'}C+\nu\varepsilon \tag{5}$$

we have that

$$\begin{aligned}
E\left(T_{aM_{a^*}}|c\right) &= \int E\left[T_{am}|c, M_{a^*}{=}m\right] dP_{M_{a^*}}(m|c) \\
&= \int E\left[T_{am}|c\right] dP_{M_{a^*}}(m|c) \\
&= \int E\left[T|a, m, c\right] dP_M(m|a^*, c) \\
&= \int E\left[e^{\theta_0+\theta_1 a+\theta_2 m+\theta_3 am+\theta_4' c+\nu\varepsilon}\right] dP_M(m|a^*, c) \\
&= e^{\theta_0+\theta_1 a+\theta_4' c} E\left[e^{\nu\varepsilon}\right] E\left[e^{\theta_2 M+\theta_3 aM}\right] \\
&= e^{\theta_0+\theta_1 a+\theta_4' c} E\left[e^{\nu\varepsilon}\right] e^{(\theta_2+\theta_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2}
\end{aligned}$$

where the first equality follows by the law of iterated expectations, the second by assumption (iv), the third by assumptions (i)-(iii), the fourth by the acclerated failure time model, and the final one by the fact that M is normally distributed and has constant variance σ^2. Thus,

$$\log\left\{E\left(T_{aM_{a^*}}|c\right)\right\} = \log\left(E\left[e^{\nu\varepsilon}\right]\right) + \theta_0+\theta_1 a+\theta_4' c + (\theta_2+\theta_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right) + \frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2$$

and so

$$\begin{aligned}
\log\left\{E\left(T_{aM_a}|c\right)\right\} - \log\left\{E\left(T_{aM_{a^*}}|c\right)\right\} &= (\theta_2\beta_1+\theta_3\beta_1 a)(a-a^*) \\
\log\left\{E\left(T_{aM_{a^*}}|c\right)\right\} - \log\left\{E\left(T_{a^*M_{a^*}}|c\right)\right\} &= \left\{\theta_1+\theta_3\left(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2\sigma^2\right)\right\}(a-a^*) + 0.5\theta_3^2\sigma^2\left(a^2-a^{*2}\right).
\end{aligned}$$

Formulas for Natural Direct and Indirect Effects for the Proportional Hazards Model with Exposure-Mediator Interaction and a Rare Outcome

Under model (2) for the mediator and model (6) for the outcome:

$$\lambda_T(t|a, m, c) = \lambda_T(t|0, 0, 0)\, e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}. \tag{6}$$

we have that

$$\lambda_{T_{aM_{a^*}}}(t|c) = \frac{f_{T_{aM_{a^*}}}(t|c)}{S_{T_{aM_{a^*}}}(t|c)}$$

where $f_{T_{aM_{a^*}}}(t|c)$ and $S_{T_{aM_{a^*}}}(t|c)$ denote the conditional density and survival functions respectively for $T_{aM_{a^*}}$. We have that

$$\begin{aligned}
f_{T_{aM_{a^*}}}(t|c) &= \int f_{T_{am}}(t|c, M_{a^*}{=}m)\, dP_{M_{a^*}}(m|c) \\
&= \int f_{T_{am}}(t|c)\, dP_{M_{a^*}}(m|c) \quad \text{by assumption} \quad \text{(iv)} \\
&= \int f_T(t|a, m, c)\, dP_M(m|a^*, c) \quad \text{by assumptions} \quad \text{(i)} - \text{(iii)} \\
&= \int \lambda_T(t|0, 0, 0)\, e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c} \exp\left(-\Lambda_T(t|0, 0, 0)\, e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right) dP_M(m|a^*, c)
\end{aligned}$$

where $\Lambda_T(t|0, 0, 0) = \int_0^t \lambda_T(t|0, 0, 0)\, dt$. Likewise,

$$S_{T_{aM_{a^*}}}(t|c)=\int \exp\left(-\Lambda_T(t|0,0,0)\,e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right)dP_M(m|a^*,c).$$

Thus,

$$\lambda_{T_{aM_{a^*}}}(t|c)=\lambda_T(t|0,0,0)\exp\left(\gamma_1 a+\gamma_4' c\right)r(t;a,a^*,c)$$

where

$$r(t;a,a^*,c)=\frac{\int e^{(\gamma_2+\gamma_3 a)m}\exp\left(-\Lambda_T(t|0,0,0)\,e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right)dP_M(m|a^*,c)}{\int \exp\left(-\Lambda_T(t|0,0,0)\,e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right)dP_M(m|a^*,c)}.$$

Since M is normally distributed we have that (cf. Lin et al.[16]):

$$\begin{aligned} r(t;a,a^*,c) & \\ &= e^{(\gamma_2+\gamma_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right)+\frac{1}{2}(\gamma_2+\gamma_3 a)^2\sigma^2} \\ &\times\frac{\int \exp\left(-\Lambda_T(t|0,0,0)\,e^{(\gamma_2+\gamma_3 a)^2+\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\beta_2' c\right)\right)^2}{2}\right)dm}{\int \exp\left(-\Lambda_T(t|0,0,0)\,e^{\gamma_1 a+\gamma_2 m+\gamma_3 am+\gamma_4' c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a^*+\beta_2' c\right)\right)^2}{2}\right)dm}\end{aligned}$$

which can be approximated by $e^{(\gamma_2+\gamma_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right)+\frac{1}{2}(\gamma_2+\gamma_3 a)^2\sigma^2}$ if $\Lambda_T(t|0, 0, 0)$ is small (i.e. if the outcome is relatively rare). Thus

$$\lambda_{T_{aM_{a^*}}}(t|c)\approx\lambda_T(t|0,0,0)\,e^{\gamma_1 a+\gamma_4' c}e^{(\gamma_2+\gamma_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right)+\frac{1}{2}(\gamma_2+\gamma_3 a)^2\sigma^2}$$

and

$$\log\left\{\lambda_{T_{aM_{a^*}}}(t|c)\right\}=\log\left(\lambda_T(t|0,0,0)\right)+\gamma_1 a+\gamma_4' c+(\gamma_2+\gamma_3 a)\left(\beta_0+\beta_1 a^*+\beta_2' c\right)+\frac{1}{2}(\gamma_2+\gamma_3 a)^2\sigma^2.$$

From this it follows that,

$$\begin{aligned}\log\left\{\lambda_{T_{aM_a}}(t|c)\right\}-\log\left\{\lambda_{T_{aM_{a^*}}}(t|c)\right\} &= (\gamma_2\beta_1+\gamma_3\beta_1 a)(a-a^*) \\ \log\left\{\lambda_{T_{aM_{a^*}}}(t|c)\right\}-\log\left\{\lambda_{T_{a^*M_{a^*}}}(t|c)\right\} &= \left\{\gamma_1+\gamma_3\left(\beta_0+\beta_1 a^*+\beta_2' c+\gamma_2\sigma^2\right)\right\}(a-a^*)+0.5\gamma_3^2\sigma^2\left(a^2-a^{*2}\right).\end{aligned}$$

Equivalence of Product and Difference Method for the Proportional Hazards Model with No Exposure-Mediator Interaction and a Rare Outcome

Suppose that the model (2) in the text for the mediator is correctly specified:

$$E\left[M|a,c\right]=\beta_0+\beta_1 a+\beta_2^{'}c. \tag{2}$$

along with proportional hazards model (3):

$$\lambda_T\left(t|a,m,c\right)=\lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c} \tag{3}$$

Suppose also a proportional hazards model is also fit without the mediator:

$$\lambda_T\left(t|a,c\right)=\lambda_T\left(t|0,0\right)e^{\phi_1 a+\phi_4^{'}c}.$$

The difference method uses $\phi_1-\gamma_1$ as a measure of the indirect effect; the product method uses $\beta_1\gamma_2$. We show that if all of the models are correctly specified and the outcome is rare these two are approximately equal. This is because by the proportional hazards model without the mediator:

$$\lambda_T\left(t|a,c\right)=\lambda_T\left(t|0,0\right)e^{\phi_1 a+\phi_4^{'}c}$$

and by model (3)

$$\begin{aligned}\lambda_T\left(t|a,c\right) &= \frac{f_T\left(t|a,c\right)}{S_T\left(t|a,c\right)}\\ &= \frac{\int f_T\left(t|a,m,c\right)dP_M\left(m|a,c\right)}{\int S_T\left(t|a,m,c\right)dP_M\left(m|a,c\right)}\\ &= \frac{\int\lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)dP_M\left(m|a,c\right)}{\int\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)dP_M\left(m|a,c\right)}\\ &= \lambda_T\left(t|0,0,0\right)\ \exp\left(\gamma_1 a+\gamma_4^{'}c\right)r\left(t;a,c\right)\end{aligned}$$

where

$$r\left(t;a,c\right)=\frac{\int e^{\gamma_2 m}\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)dP_M\left(m|a,c\right)}{\int\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)dP_M\left(m|a,c\right)}.$$

As in the previous proof, since *M* is normally distributed we have that (cf. Lin et al.[16]):

$$r\left(t;a,c\right)\ =\ e^{\gamma_2\left(\beta_0+\beta_1 a+\beta_2^{'}c\right)+\frac{1}{2}\gamma_2^2\sigma^2}\times\frac{\int\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_2^2+\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a+\beta_2^{'}c\right)\right)^2}{2}\right)dm}{\int\exp\left(-\Lambda_T\left(t|0,0,0\right)e^{\gamma_1 a+\gamma_2 m+\gamma_4^{'}c}\right)\exp\left(-\frac{\left(m-\left(\beta_0+\beta_1 a+\beta_2^{'}c\right)\right)^2}{2}\right)dm}$$

which can be approximated by $e^{\gamma_2\left(\beta_0+\beta_1 a+\beta_2^{'}c\right)+\frac{1}{2}\gamma_2^2\sigma^2}$ if $\Lambda_T(t|0,0,0)$ is small (i.e. if the outcome is relatively rare). Thus

$$\lambda_T(t|a,c) \approx \left\{e^{\gamma_2\beta_0+\frac{1}{2}\gamma_2^2\sigma^2}\lambda_T(t|0,0,0)\right\}e^{(\gamma_1+\gamma_2\beta_1)a+(\gamma_2\beta_2+\gamma_4)'c}.$$

Because this holds for all a, we must have $\phi_1 \approx \{\gamma_1 + \gamma_2\beta_1\}$ and thus $\phi_1 - \gamma_1 \approx \gamma_2\beta_1$.

The Product Method for the Proportional Hazards Model with Common Outcome Yields a Valid Test of the Presence of Any Mediated Effect

We assume models (2) and (3) are correctly specified and that assumptions (i)-(iv) hold. In counterfactual notation, these are that for all a, a^*, m, (i) $T_{am} \perp\!\!\!\perp A|C$, (ii) $T_{am} \perp\!\!\!\perp M|C$, (iii) $M_a \perp\!\!\!\perp A|C$ and (iv) $T_{am} \perp\!\!\!\perp M_{a^*}|C$ where $X \perp\!\!\!\perp Y|Z$ denotes that X is independent of Y conditional on Z. On any causal diagram for which (iv) holds, it also follows that $(T_{am}, T_{am^*}) \perp\!\!\!\perp (M_a, M_{a^*})|C$. If in models (2) and (3) we have that $\gamma_2\beta_1 \neq 0$ then from this it follows that $\gamma_2 \neq 0$ and $\beta_1 \neq 0$. If $\beta_1 \neq 0$ then by assumption (iii) it follows that A has an effect on M in the sense that for some a and a^* there are individuals $\omega \in \Theta_1$ such that, $M_a(\omega) - M_{a^*}(\omega) \neq 0$. Let $m = M_a(\omega)$ and $m^* = M_{a^*}(\omega)$. If $\gamma_2 \neq 0$ then by assumptions (i) and (ii) it follows that M has an effect on Y with A fixed at a in the sense that there are individuals $\omega \in \Theta_2$ such that $T_{am}(\omega) - T_{am^*}(\omega) \neq 0$. Since $(T_{am}, T_{am^*}) \perp\!\!\!\perp (M_a, M_{a^*})|C$, it follows that there are individuals $\omega \in \Theta_1 \cap \Theta_2$ and thus for $\omega \in \Theta_1 \cap \Theta_2$, $0 \neq T_{am}(\omega) - T_{am^*}(\omega) = T_{aM_a}(\omega) - T_{aM_{a^*}}(\omega)$ i.e. $T_{aM_a}(\omega) \neq T_{aM_{a^*}}(\omega)$ so there are some individuals for whom the natural indirect effect is non-zero.

References

1. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992; 3(2):143–155. [PubMed: 1576220]
2. Pearl, J. Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence. Morgan Kaufmann; San Francisco: 2001. Direct and indirect effects.; p. 411-420.
3. Lange T, Hansen JV. Direct and indirect effects in a survival context. Epidemiology. 2011; 22:xxx–xxx.
4. Tein, J-Y.; MacKinnon, DP. Estimating mediated effects with survival data.. In: Yanai, H.; Rikkyo, AO.; Shigemasu, K.; Kano, Y.; Meulman, JJ., editors. New Developments on Psychometrics. Springer-Verlag Tokyo Inc; Tokyo, Japan: 2003. p. 405-412.
5. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Statist. Interface - Special Issue on Mental Health and Soc Behav Sci. 2009; 2(4):457–468.
6. Peterson ML, Sinisi SE, van der Laan MJ. Estimation of direct causal effects. Epidemiology. 2006; 17:276–84. [PubMed: 16617276]
7. VanderWeele TJ. Marginal structural models for the estimation of direct and indirect effects. Epidemiology. 2009; 20(1):18–26. [PubMed: 19234398]
8. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis with a dichotomous outcome. American Journal of Epidemiology. 2010; 172:1339–1348. [PubMed: 21036955]
9. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010; 21:540–551. [PubMed: 20479643]
10. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Pyschological Methods. 15:309–334.
11. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51(6):1173–1182. [PubMed: 3806354]
12. MacKinnon, DP. An Introduction to Statistical Mediation Analysis. Lawrence Erlbaum Associates; New York: 2008.

13. Judd CM, Kenny DA. Process analysis: estimating mediation in treatment evaluations. Eval Rev. 1981; 5(5):602–619.

14. MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivariate Behavioral Research. 1995; 30:41–62. [PubMed: 20157641]

15. MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. Evaluation Review. 17:144–158.

16. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeausred confounding in observational studies. Biometrics. 1998; 54:948–963. [PubMed: 9750244]

17. Hafeman DM. ”Proportion explained”: a causal interpretation for standard measures of indirect effect? Am J Epidemiol. 2009; 170(11):1443–1448. [PubMed: 19850625]