



Published in final edited form as:

*Epidemiology*. 2017 May ; 28(3): 370–378. doi:10.1097/EDE.0000000000000651.

## Causal mediation analysis of survival outcome with multiple mediators

Yen-Tsung Huang<sup>1,2</sup> and Hwai-I Yang<sup>3</sup>

<sup>1</sup>Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

<sup>2</sup>Departments of Epidemiology and Biostatistics, Brown University, Providence, RI 02912, United States

<sup>3</sup>Genomics Research Center, Academia Sinica, Taipei, Taiwan

### Abstract

**Background**—Mediation analyses have been a popular approach to investigate the effect of an exposure on an outcome through a mediator. Mediation models with multiple mediators have been proposed for continuous and dichotomous outcomes. However, development of multi-mediator models for survival outcomes is still limited.

**Methods**—We present methods for multi-mediator analyses using three survival models: Aalen additive hazard models, Cox proportional hazard models, and semiparametric probit models. Effects through mediators can be characterized by path-specific effects, for which definitions and identifiability assumptions are provided. We derive closed form expressions for path-specific effects for the three models, which are intuitively interpreted using a causal diagram.

**Results**—Mediation analyses using Cox models under the rare outcome assumption and Aalen additive hazard models consider effects on log hazard ratio and hazard difference, respectively; analyses using semiparametric probit models consider effects on difference in transformed survival time as well as survival probability. The three models were applied to a hepatitis study where we investigated effects of hepatitis C on liver cancer incidence mediated through baseline and/or follow-up hepatitis B viral load. The three methods show consistent results on respective effect scales, which suggest an adverse estimated effect of hepatitis C on liver cancer not mediated through hepatitis B, and a protective estimated effect mediated through the baseline (and possibly follow-up) of hepatitis B viral load.

**Conclusions**—Causal mediation analyses of survival outcome with multiple mediators are developed for additive hazard and proportional hazard and probit models with utility demonstrated in a hepatitis study.

---

**Corresponding Author:** Yen-Tsung Huang, Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, TAIWAN; Departments of Epidemiology and Biostatistics, Brown University, 121 South Main Street, Providence, RI 02912, USA; phone: 886-2-2783-5611 ext 315; ythuang@stat.sinica.edu.tw.

**Availability of Data and Code:** Computation code and data are available at <http://www.stat.sinica.edu.tw/ythuang/M3S-Cox%20and%20M3S-Aalen.zip> and <http://www.stat.sinica.edu.tw/ythuang/M3S-Probit.zip>.

**Conflict of Interest:** NA

## Keywords

Additive hazard model; causal mediation model; Cox proportional hazard model; multiple mediators; semiparametric probit model; survival analysis

---

## INTRODUCTION

Mediation analyses first proposed in psychology literature<sup>1,2</sup> have become a very popular approach and been applied to a wide range of epidemiologic studies<sup>3-7</sup>. Mediation models characterize relationships of an exposure, a mediator, and an outcome. Specifically, in mediation analyses, the effect of the exposure on the outcome can be decomposed into natural direct effect, the effect of the exposure on the outcome not mediated by the mediator, and natural indirect effect, the effect of the exposure on the outcome mediated through the mediator. By employing the counterfactual framework<sup>8</sup>, causal mediation models can be formulated as a graphic model illustrated using a directed acyclic graph (DAG)<sup>9</sup>, and causal assumptions for effect identifiability have been carefully studied<sup>10</sup>. Built upon the framework of causal inference, the methodology of mediation analyses has been generalized from continuous outcomes to dichotomous outcomes<sup>11,12</sup> and time-to-event survival outcomes<sup>13,14</sup>. The identification of natural direct and indirect effect in mediation analyses with an exposure  $S$ , a mediator  $M$ , and a survival outcome  $Y$  requires a set of assumptions<sup>14</sup>. The assumptions can be stated as that, conditional on measured confounders  $X$ , there is 1) no confounding for the relationship of  $S$  and  $Y$ , 2) no confounding for the relationship of  $M$  and  $Y$ , conditional on  $S$ , 3) no confounding for the  $S$ - $M$  association, and 4) no confounder for the  $M$ - $Y$  relationship caused by  $S$ ; additional standard assumptions include positivity and consistency discussed in eAppendix A1.

This paper is motivated by a hepatitis study where mediation by hepatitis B and C in relation to liver cancer was investigated<sup>15</sup>. The study was conducted in a community-based prospective cohort study in Taiwan, in which viral load of hepatitis B (HBV) and C (HCV) viruses was measured at the baseline, and the incidence of liver cancer was recorded prospectively<sup>16</sup>. Based on the existing scientific evidence<sup>17-23</sup>, a mediation model was proposed with the exposure being hepatitis C viral load, the mediator being hepatitis B viral load and the outcome being the incidence of liver cancer. Additionally, hepatitis B viral load was also measured during the follow-up. We extend the above mediation model to include exposure, HCV viral load ( $S$ ), the outcome, the liver cancer incidence ( $Y$ ) and two mediators, one being the baseline HBV viral load ( $M_1$ ) and the other being the follow-up HBV viral load ( $M_2$ ) (Figure 1). Note that if one is interested in the mediator model with only one mediator  $M_2$ , then the existence of  $M_1$  violates the aforementioned assumption 4). To address the issue, we study three effects: 1) the effect of HCV on the liver cancer incidence mediated through the baseline HBV viral load and possibly through the follow-up HBV viral load (the black path), 2) the effect of HCV on the liver cancer incidence mediated through the follow-up HBV viral load, but not through the baseline HBV viral load (the dark gray path), and 3) the effect of HCV on the liver cancer incidence not mediated by HBV regardless through the baseline or follow-up viral load (the light gray path).

To identify the above effects, one has to conduct mediation analyses for a model with a survival outcome and multiple mediators. The three different effects have been termed as path-specific effects, which characterize mediation effects for various pathways through different mediators<sup>24</sup>. Regression-based and weighting-based approaches have been proposed to estimate path-specific effects<sup>25</sup>. The path-specific effect approach has also been proposed as a method to adjust for exposure-induced confounding for the mediator-outcome association<sup>26</sup>, similar to our problem in the hepatitis study. Causal mediation models have been generalized to incorporate mixed variable types such as a combination of continuous and dichotomous mediators<sup>27</sup> or a set of high-dimensional continuous mediators<sup>28</sup>. However, these methods focus on non-censored outcomes. There has been some work on mediation analyses for survival outcomes. Lange and Hansen (2011)<sup>13</sup> introduced mediation analyses using Aalen additive hazard model<sup>29</sup>; VanderWeele (2011)<sup>14</sup> presented mediation analyses using Cox proportional hazard models<sup>30</sup> and accelerated failure time models<sup>31</sup>; Tchetgen Tchetgen (2011) proposed studying mediation in Cox models by a doubly robust estimator<sup>32</sup>. Except for a resampling-based method presented by Lange et al. (2013)<sup>33</sup>, literature on survival models that incorporate multiple mediators is still very limited.

Because the numerical approach may have its limitation in computation costs, this paper aims to present analytic methods, which also provide better mechanistic understanding of the parameters involved in mediation effects. Huang and Cai (2016) have recently developed a new mediation analysis for survival outcome using semiparametric probit models<sup>34</sup>. Instead of focusing on a specific effect scale, e.g., hazard difference, hazard ratio, or ratio of survival time, the probit model is able to quantify the effect on the entire survival probability. Due to the conjugacy property of normal distributions, this model can be easily used to study multiple mediators. Moreover, we develop two additional multi-mediator models that use Aalen additive hazard and Cox proportional hazard models. We illustrate the utility of the three multi-mediator models in the motivating hepatitis study where we examine path-specific effects on different effect scales, including the difference in transformed survival time, the difference in survival probability across follow-up time, the hazard difference, and the hazard ratio.

## DEFINITIONS AND ASSUMPTIONS

We first define a general outcome  $Y$  to be a function of time to liver cancer development  $T$ ,  $Y = \mathcal{H}(T)$ . Let  $Y(s, m_1, m_2)$  be the counterfactual outcome that would have been observed had  $S$  (HCV baseline viral load),  $M_1$  (HBV baseline viral load) and  $M_2$  (HBV follow-up viral load) been set to  $s$ ,  $m_1$  and  $m_2$ , respectively;  $M_2(s, m_1)$  be the counterfactual value of  $M_2$  (HBV follow-up viral load) had  $S$  and  $M_1$  been set to  $s$  and  $m_1$ , respectively; and  $M_1(s)$  be the counterfactual value of  $M_1$  (HBV baseline viral load) had  $S$  been set to  $s$ . By extension of natural direct and indirect to two-mediator models, we define three path-specific effects on  $Y$ :

$$\begin{aligned} \Delta_{S \rightarrow Y} &= Y(s_1, M_1(s_0), M_2(s_0, M_1(s_0))) - Y(s_0, M_1(s_0), M_2(s_0, M_1(s_0))) \\ \Delta_{S \rightarrow M_2 \rightarrow Y} &= Y(s_1, M_1(s_0), M_2(s_1, M_1(s_0))) - Y(s_1, M_1(s_0), M_2(s_0, M_1(s_0))) \\ \Delta_{S \rightarrow M_1 Y} &= Y(s_1, M_1(s_1), M_2(s_1, M_1(s_1))) - Y(s_1, M_1(s_0), M_2(s_1, M_1(s_0))). \end{aligned} \quad (1)$$

A series of no unmeasured confounding assumptions needs to be made to identify the three path-specific effects<sup>26,34</sup>. With  $A \perp\!\!\!\perp B|C$  for the condition that  $A$  is independent of  $B$  conditional on  $C$ , we list six sufficient conditions for identifying the PSE in the context of the motivating hepatitis study. To simplify notation, the following assumptions are presented as marginal exchangeability, and they can be generalized to be conditionally exchangeable on covariates  $X$ , in which known confounders such as age and gender can be adjusted to achieve exchangeability.

1.  $T(s, m_1, m_2) \perp\!\!\!\perp (M_1, M_2) | S$ : there is no confounding for the joint effect of the baseline and follow-up hepatitis B viral load ( $M_1, M_2$ ) on the time to liver cancer development, conditional on the baseline hepatitis C viral load ( $S$ ).
2.  $T(s, m_1, m_2) \perp\!\!\!\perp S$ : there is no confounding for the effect of the baseline hepatitis C viral load ( $S$ ) on the time to liver cancer incidence.
3.  $M_2(s, m_1) \perp\!\!\!\perp (S, M_1)$ : there is no confounding for the joint effect of ( $S, M_1$ ), the baseline HBV and HCV viral load on the follow-up HBV viral load ( $M_2$ ).
4.  $M_1(s) \perp\!\!\!\perp S$ : there is no confounding for the effect of the baseline HCV viral load ( $S$ ) on the baseline HBV viral load ( $M_1$ ).
5.  $T(s, m_1, m_2) \perp\!\!\!\perp (M_1(s^*), M_2(s^{**}, m_1))$ : there is no baseline HCV viral load ( $S$ )-induced factor that can confound the baseline HBV viral load-survival time ( $M_1-T$ ) and the follow-up HBV viral load-survival time ( $M_2-T$ ) joint relation, where  $s^*$  and  $s^{**}$  are interventions for the baseline HCV viral load with different values than  $s$  and each other.
6.  $M_2(s, m_1) \perp\!\!\!\perp M_1(s^*)$ : there is no baseline HCV viral load-induced factor that confounds the  $M_1$ - $M_2$  (baseline vs. follow-up HBV viral load) association.

Besides no unmeasured confounding assumptions, standard assumptions of positivity and consistency are also required (see eAppendix A1).

## MULTI-MEDIATOR MODELS OF SURVIVAL OUTCOME

In this section, we present three methods for mediation analyses with one exposure  $S$ , two mediators  $M_1$  and  $M_2$ , and a survival outcome  $Y$ , as shown in Figure 1. The methods can easily be extended to more than two mediators.

### Aalen additive hazard model

We propose two linear regression models for  $M_1$  and  $M_2$ , respectively:

$$M_{1i} = \delta_X^T \mathbf{X}_i + \delta_S S_i + \varepsilon_{M_{1i}} \quad (2)$$

$$M_{2i} = \alpha_X^T \mathbf{X}_i + \alpha_S S_i + \alpha_M M_{1i} + \varepsilon_{M_{2i}}, \quad (3)$$

where  $\mathbf{X}$  is covariates with the first element being 1 for the intercept; the error terms  $\varepsilon_{M_{1i}}$  and  $\varepsilon_{M_{2i}}$  are independent and normally distributed with mean zero and respective variances,  $\sigma_{M_1}^2$  and  $\sigma_{M_2}^2$ . We propose the following additive hazard model for the outcome  $Y \equiv \lambda(t)$ :

$$\lambda(t | \mathbf{X}_i, S_i, M_{1i}, M_{2i}) = \lambda_i = \lambda_0(t) + \lambda_X^T \mathbf{X}_i^* + \lambda_S S_i + \lambda_{M_1} M_{1i} + \lambda_{M_2} M_{2i}, \quad (4)$$

where  $\lambda_i$  is the hazard of developing liver cancer for subject  $i$ ;  $\lambda_0(t)$  is the baseline hazard; and  $\lambda_X, \lambda_S, \lambda_{M_1}$  and  $\lambda_{M_2}$  are regression coefficients for the covariates  $\mathbf{X}^*$  ( $\mathbf{X}$  without the first element), the baseline HCV viral load  $S$ , the baseline HBV viral load  $M_1$  and the follow-up HBV viral load  $M_2$ , respectively. Based on the six assumptions in the Definitions and Assumptions section one can express the counterfactual hazard as:

$$\lambda(T(s_a, M_1(s_c), M_2(s_b, M_1(s_c))); t | \mathbf{X}) = \left\{ \lambda_0(t) + \lambda_X^T \mathbf{X}^* + \lambda_{M_1} \delta_X^T \mathbf{X} + \lambda_{M_2} \alpha_X^T \mathbf{X} + \lambda_{M_2} \alpha_M \delta_X^T \mathbf{X} - \sigma_{W_\lambda}^2 \right\} + \lambda_S s_a + \lambda_{M_2} \alpha_S s_b + (\lambda_{M_1} + \lambda_{M_2} \alpha_M) \delta_S s_c,$$

where  $\sigma_{W_\lambda}^2 = \lambda_{M_1}^2 \sigma_{M_1}^2 + \lambda_{M_2}^2 \sigma_{M_2}^2 + \lambda_{M_2}^2 \alpha_M^2 \sigma_{M_1}^2 + 2\lambda_{M_1} \lambda_{M_2} \alpha_M \sigma_{M_1}^2$ . Derivation of the above expression is in eAppendix. The model for the second mediator (3) and the survival model (4) focus on the main effects, and they can incorporate  $S$ -by- $M_1$  and  $S$ -by- $M_2$  cross-product interaction terms by replacing  $\alpha_M, \lambda_{M_1}$  and  $\lambda_{M_2}$  with  $\alpha_M + \alpha_{SM} s_b, \lambda_{M_1} + \lambda_{SM1} s_a$  and  $\lambda_{M_2} + \lambda_{SM2} s_a$ , respectively. We then use the definition in (1) to re-expressed path-specific effects on the scale of hazard difference by using the above result:

$$\begin{aligned} \Delta_{S \rightarrow Y}^{Aalen} &= \lambda(T(s_1, M_1(s_0), M_2(s_0, M_1(s_0))); t | \mathbf{X}) - \lambda(T(s_0, M_1(s_0), M_2(s_0, M_1(s_0))); t | \mathbf{X}) \\ &= \lambda_S (s_1 - s_0) \\ \Delta_{S \rightarrow M_2 \rightarrow Y}^{Aalen} &= \lambda(T(s_1, M_1(s_0), M_2(s_1, M_1(s_0))); t | \mathbf{X}) - \lambda(T(s_1, M_1(s_0), M_2(s_0, M_1(s_0))); t | \mathbf{X}) \\ &= \lambda_{M_2} \alpha_S (s_1 - s_0) \\ \Delta_{S \rightarrow M_1 \rightarrow Y}^{Aalen} &= \lambda(T(s_1, M_1(s_1), M_2(s_1, M_1(s_1))); t | \mathbf{X}) - \lambda(T(s_1, M_1(s_0), M_2(s_1, M_1(s_0))); t | \mathbf{X}) \\ &= (\lambda_{M_1} + \lambda_{M_2} \alpha_M) \delta_S (s_1 - s_0). \end{aligned}$$

(5)

The results in (5) have intuitive interpretations and can be easily visualized using the DAG in Figure 1. Each arrow in Figure 1 has its corresponding effect parameter in models (2)–(4):  $\delta_S$  represents the effect of the baseline HCV viral load  $S$  on the baseline HBV viral load  $M_1$ ;  $\alpha_S$  and  $\alpha_M$  respectively represent the effects of the baseline HCV viral load  $S$  and the

baseline HBV viral load  $M_1$  on the follow-up HBV viral load  $M_2$ ;  $\lambda_S$ ,  $\lambda_{M_1}$ , and  $\lambda_{M_2}$  respectively represent the effects of  $S$ ,  $M_1$  and  $M_2$  on the hazard of developing liver cancer. The path of  $\Delta_{S \rightarrow Y}$  has only one arrow with effect parameter  $\delta_S$ .  $\Delta_{S \rightarrow M_2 \rightarrow Y}$  has two arrows:  $S \rightarrow M_2$  and  $M_2 \rightarrow Y$  with respective effect parameters being  $\alpha_S$  and  $\lambda_{M_2}$ , and the path-specific effect is the product of the two parameters  $\alpha_S \lambda_{M_2}$ .  $\Delta_{S \rightarrow M_1 Y}$  contains two paths:  $S \rightarrow M_1 \rightarrow M_2 \rightarrow Y$  and  $S \rightarrow M_1 \rightarrow Y$ . The path  $S \rightarrow M_1 \rightarrow M_2 \rightarrow Y$  consists of three arrows:  $S \rightarrow M_1$ ,  $M_1 \rightarrow M_2$  and  $M_2 \rightarrow Y$  with effect parameters being  $\delta_S$ ,  $\alpha_M$  and  $\lambda_{M_2}$ , respectively; the path  $S \rightarrow M_1 \rightarrow Y$  consists of two arrows:  $S \rightarrow M_1$  and  $M_1 \rightarrow Y$  with effect parameters being  $\delta_S$  and  $\lambda_{M_1}$ , respectively. The effect of  $\Delta_{S \rightarrow M_1 Y}$  is the sum of  $\delta_S \alpha_M \lambda_{M_2}$  and  $\delta_S \lambda_{M_1}$ , which are the products of effect parameters along the two paths.

$$\begin{aligned} \Delta_{TE}^{Aalen} &\equiv \lambda(T(s_1); t | \mathbf{X}) \\ &\quad - \lambda(T(s_0); t | \mathbf{X}) \\ &= \lambda(T(s_1, M_1(s_1), M_2(s_1, M_1(s_1))); t | \mathbf{X}) \\ &\quad - \lambda(T(s_0, M_1(s_0), M_2(s_0, M_1(s_0))); t | \mathbf{X}), \end{aligned}$$

Total effect, defined as  $\Delta_{TE}^{Aalen} = \Delta_{S \rightarrow Y}^{Aalen} + \Delta_{S \rightarrow M_2 \rightarrow Y}^{Aalen} + \Delta_{S \rightarrow M_1 Y}^{Aalen}$ . The additive hazard model (4) can incorporate time-dependent effects by generalizing  $\lambda_S$ ,  $\lambda_{M_1}$ ,  $\lambda_{M_2}$ , and  $\lambda_X$  to  $\lambda_S(t)$ ,  $\lambda_{M_1}(t)$ ,  $\lambda_{M_2}(t)$ , and  $\lambda_X(t)$ , respectively.

### Cox proportional hazard model

Unlike additive hazard models, which assume hazard linearly determined by the predictors, Cox models assume that the hazard is determined by the predictors exponentially (or linearly on the log hazard scale):

$$\log \lambda_i = \log \lambda_0(t) + \gamma_X^T \mathbf{X}_i^* + \gamma_S S_i + \gamma_{M_1} M_{1i} + \gamma_{M_2} M_{2i}, \quad (6)$$

where  $\gamma_X$ ,  $\gamma_S$ ,  $\gamma_{M_1}$ , and  $\gamma_{M_2}$  are regression coefficients for  $\mathbf{X}^*$ ,  $S$ ,  $M_1$ , and  $M_2$ , respectively. Provided that the assumptions in the Definitions and Assumptions section hold and the outcome event is rare, one can approximate the counterfactual log hazard as:

$$\begin{aligned} &\log \lambda(T(s_a, M_1(s_c), M_2(s_b, M_1(s_c))); t | \mathbf{X}) \\ \approx &\left\{ \log \lambda_0(t) + \gamma_X^T \mathbf{X}^* + \gamma_{M_1} \delta_X^T \mathbf{X} + \gamma_{M_2} \alpha_X^T \mathbf{X} + \gamma_{M_2} \alpha_M \delta_X^T \mathbf{X} + \frac{1}{2} \sigma_{W_\gamma}^2 \right\} + \gamma_S s_a + \gamma_{M_2} \alpha_S s_b + (\gamma_{M_1} + \gamma_{M_2} \alpha_M) \delta_S s_c, \end{aligned}$$

where  $\sigma_{W_\gamma}^2 = \gamma_{M_1}^2 \sigma_{M_1}^2 + \gamma_{M_2}^2 \sigma_{M_2}^2 + \gamma_{M_2}^2 \alpha_M^2 \sigma_{M_1}^2 + 2\gamma_{M_1} \gamma_{M_2} \alpha_M \sigma_{M_1}^2$ . Derivation of the above expression is provided in the eAppendix (Section A5). Again, it can be extended to incorporate  $S$ -by- $M_1$  and  $S$ -by- $M_2$  cross-product interaction terms by replacing  $\alpha_M$ ,  $\gamma_{M_1}$ , and  $\gamma_{M_2}$  in (3) and (6) with  $\alpha_M + \alpha_{SM} s_b$ ,  $\gamma_{M_1} + \gamma_{SM_1} s_a$ , and  $\gamma_{M_2} + \gamma_{SM_2} s_a$ , respectively. We let  $Y \equiv \log \lambda(t | \mathbf{X})$ , and re-express the three path-specific effects in (1) as effects on log hazard ratio by using the above expression:

$$\begin{aligned}
 \Delta_{S \rightarrow Y}^{Cox} &= \log \lambda (T(s_1, M_1(s_0), M_2(s_0, M_1(s_0))) ; t | \mathbf{X}) - \log \lambda (T(s_0, M_1(s_0), M_2(s_0, M_1(s_0))) ; t | \mathbf{X}) \\
 &\approx \gamma_S (s_1 - s_0) \\
 \Delta_{S \rightarrow M_2 \rightarrow Y}^{Cox} &= \log \lambda (T(s_1, M_1(s_0), M_2(s_1, M_1(s_0))) ; t | \mathbf{X}) - \log \lambda (T(s_1, M_1(s_0), M_2(s_0, M_1(s_0))) ; t | \mathbf{X}) \\
 &\approx \gamma_{M_2} \alpha_S (s_1 - s_0) \\
 \Delta_{S \rightarrow M_1 Y}^{Cox} &= \log \lambda (T(s_1, M_1(s_1), M_2(s_1, M_1(s_1))) ; t | \mathbf{X}) - \log \lambda (T(s_1, M_1(s_0), M_2(s_1, M_1(s_0))) ; t | \mathbf{X}) \\
 &\approx (\gamma_{M_1} + \gamma_{M_2} \alpha_M) \delta_S (s_1 - s_0).
 \end{aligned}$$

(7)

Similar to mediation analyses using Aalen additive hazard models, one can visualize the three path-specific effects  $\Delta^{Cox} = (\Delta_{S \rightarrow Y}^{Cox}, \Delta_{S \rightarrow M_2 \rightarrow Y}^{Cox}, \Delta_{S \rightarrow M_1 Y}^{Cox})^T$  as products of parameters in Figure 1. Total effect is sum of the three path-specific effects:

$$\begin{aligned}
 \Delta_{TE}^{Cox} &\equiv \log \lambda (T(s_1) ; t | \mathbf{X}) - \log \lambda (T(s_0) ; t | \mathbf{X}) = \\
 &\log \lambda (T(s_1, M_1(s_1), M_2(s_1, M_1(s_1))) ; t | \mathbf{X}) - \\
 &\log \lambda (T(s_0, M_1(s_0), M_2(s_0, M_1(s_0))) ; t | \mathbf{X}) = \Delta_{S \rightarrow Y}^{Cox} + \Delta_{S \rightarrow M_2 \rightarrow Y}^{Cox} + \Delta_{S \rightarrow M_1 Y}^{Cox}.
 \end{aligned}$$

Variances and confidence intervals for  $\Delta^{Aalen}$  and  $\Delta^{Cox}$  can be calculated using a resampling method that takes random draws from multivariate normal distribution of estimates for  $(\delta_S, \alpha_S, \alpha_M, \lambda_S, \lambda_{M_1}, \lambda_{M_2})$  or  $(\delta_S, \alpha_S, \alpha_M, \gamma_S, \gamma_{M_1}, \gamma_{M_2})$ <sup>13,28</sup> with detail provided in the eAppendix (Sections A4 and A5).

### Probit model

For completeness as well as comparison with the above two models, we present another multi-mediator model that has been published<sup>34</sup>. This model proposes a semiparametric probit model for the survival time:

$$H(T_i) = - \left( \beta_X^T \mathbf{X}_i + \beta_S S_i + \beta_{M_1} M_{1i} + \beta_{M_2} M_{2i} \right) + \varepsilon_{T_i}, \quad (8)$$

where the outcome  $H(T)$  is a nonparametric transformation  $H(\cdot)$  of the survival time  $T$ ;  $\varepsilon_T$  is a standard normal random variable, independent of  $\varepsilon_{M_1}$  and  $\varepsilon_{M_2}$ ; and  $\beta_X, \beta_S, \beta_{M_1}$  and  $\beta_{M_2}$  are regression coefficients for the covariates  $\mathbf{X}$ , the baseline HCV viral load  $S$ , the baseline HBV viral load  $M_1$  and the follow-up HBV viral load  $M_2$ , respectively. By the assumptions in Section 2, one can show that

$$\begin{aligned}
 &E [ H(T; s_a, M_1(s_c), M_2(s_b, M_1(s_c))) | \mathbf{X} ] \\
 &= - \left[ \left\{ \beta_X^T \mathbf{X} + \beta_{M_1} \delta_X^T \mathbf{X} + \beta_{M_2} \alpha_X^T \mathbf{X} + \beta_{M_2} \alpha_M \delta_X^T \mathbf{X} \right\} + \beta_S s_a + \beta_{M_2} \alpha_S s_b + (\beta_{M_1} + \beta_{M_2} \alpha_M) \delta_S s_c \right].
 \end{aligned}$$

The derivation is in the eAppendix (Section A3). Similar to the additive hazard and proportional hazard models, it can be extended to incorporate  $S$ -by- $M_1$  and  $S$ -by- $M_2$  cross-product interaction terms by replacing  $\alpha_M, \beta_{M_1}$ , and  $\beta_{M_2}$  in (3) and (8) with  $\alpha_M + \alpha_{SM} s_b, \beta_{M_1} + \beta_{SM_1} s_a$ , and  $\beta_{M_2} + \beta_{SM_2} s_a$ , respectively. By letting  $Y \equiv E[H(T) | X]$  and using the above result, the path-specific effects defined in (1) can be expressed as:

$$\begin{aligned} \Delta_{S \rightarrow Y}^{Probit} &= E[H(T; s_1, M_1(s_0), M_2(s_0, M_1(s_0))) | \mathbf{X}] - E[H(T; s_0, M_1(s_0), M_2(s_0, M_1(s_0))) | \mathbf{X}] \\ &= -\beta_S (s_1 - s_0) \\ \Delta_{S \rightarrow M_2 \rightarrow Y}^{Probit} &= E[H(T; s_1, M_1(s_0), M_2(s_1, M_1(s_0))) | \mathbf{X}] - E[H(T; s_1, M_1(s_0), M_2(s_0, M_1(s_0))) | \mathbf{X}] \\ &= -\beta_{M_2} \alpha_S (s_1 - s_0) \\ \Delta_{S \rightarrow M_1 Y}^{Probit} &= E[H(T; s_1, M_1(s_1), M_2(s_1, M_1(s_1))) | \mathbf{X}] - E[H(T; s_1, M_1(s_0), M_2(s_1, M_1(s_0))) | \mathbf{X}] \\ &= -(\beta_{M_1} + \beta_{M_2} \alpha_M) \delta_S (s_1 - s_0). \end{aligned}$$

(9)

Note that the three path-specific effects  $\Delta^{Probit} = (\Delta_{S \rightarrow Y}^{Probit}, \Delta_{S \rightarrow M_2 \rightarrow Y}^{Probit}, \Delta_{S \rightarrow M_1 Y}^{Probit})^T$  are effects on the difference in the transformed survival time. Similar to the other two models,

the three path-specific effects  $\Delta^{Probit} = (\Delta_{S \rightarrow Y}^{Probit}, \Delta_{S \rightarrow M_2 \rightarrow Y}^{Probit}, \Delta_{S \rightarrow M_1 Y}^{Probit})^T$  can also be visualized as products of parameters in Figure 1.

Next we present another set of path-specific effects, the difference in survival probability.

Note that the survival probability in the motivating hepatitis study represents the probability that a subject has not yet developed liver cancer. Since  $\varepsilon_{M_1}, \varepsilon_{M_2}$ , and  $\varepsilon_T$  are all normally distributed, one can further show that the distribution function for the survival time  $T, F_T(t)$  is conjugate with the distributions for  $M_2$  and  $M_1$ . The counterfactual outcome defined as a cumulative distribution function of counterfactual survival time

$T(s_a, M_1(s_c), M_2(s_b, M_1(s_c)))$  can be expressed as:

$$\begin{aligned} &F_{T(s_a, M_1(s_c), M_2(s_b, M_1(s_c)))}(t | \mathbf{X}) \\ = \Phi &\left( \left[ \log \Lambda(t) + \beta_X^T \mathbf{X} + \beta_{M_1} \delta_X^T \mathbf{X} + \beta_{M_2} \alpha_X^T \mathbf{X} + \beta_{M_2} \alpha_M \delta_X^T \mathbf{X} \right] + \beta_S s_a + \beta_{M_2} \alpha_S s_b + (\beta_{M_1} + \beta_{M_2} \alpha_M) \delta_S s_c \right] \sigma^{*-1} \end{aligned}$$

$$\equiv 1 - \Omega(t | s_a, s_b, s_c),$$

where  $\sigma^{*-1} = 1 + \beta_{M_2}^2 \sigma_{M_2}^2 + (\beta_{M_1} + \beta_{M_2} \alpha_M)^2 \sigma_{M_1}^2$  and  $\Lambda(t)$  is the baseline cumulative hazard (see eAppendix for the derivation). Therefore, by letting  $Y_t \equiv F_T(t | \mathbf{X})$ , we obtain another set of path-specific effects  $\Omega = (\Omega_{S \rightarrow Y_t}, \Omega_{S \rightarrow M_2 \rightarrow Y_t}, \Omega_{S \rightarrow M_1 Y_t})^T$  that characterizes the effects on entire survival probability as a function of the follow-up time:



$$\begin{aligned}
 \Omega_{S \rightarrow Y_t} &= F_{T(s_0, M_1(s_0), M_2(s_0), M_1(s_0))}(t|\mathbf{X}) - F_{T(s_1, M_1(s_0), M_2(s_0), M_1(s_0))}(t|\mathbf{X}) \\
 &= \Omega(t|s_a=s_1, s_b=s_0, s_c=s_0) - \Omega(t|s_a=s_0, s_b=s_0, s_c=s_0) \\
 \Omega_{S \rightarrow M_2 \rightarrow Y_t} &= F_{T(s_1, M_1(s_0), M_2(s_0), M_1(s_0))}(t|\mathbf{X}) - F_{T(s_1, M_1(s_0), M_2(s_1), M_1(s_0))}(t|\mathbf{X}) \\
 &= \Omega(t|s_a=s_1, s_b=s_1, s_c=s_0) - \Omega(t|s_a=s_1, s_b=s_0, s_c=s_0) \\
 \Omega_{S \rightarrow M_1 Y_t} &= F_{T(s_1, M_1(s_0), M_2(s_1), M_1(s_0))}(t|\mathbf{X}) - F_{T(s_1, M_1(s_1), M_2(s_1), M_1(s_1))}(t|\mathbf{X}) \\
 &= \Omega(t|s_a=s_1, s_b=s_1, s_c=s_1) - \Omega(t|s_a=s_1, s_b=s_1, s_c=s_0).
 \end{aligned} \tag{10}$$

The difference in survival probability is a very general effect because other effect scales such as the average survival time and the hazard are functions of the survival probability.  $\Lambda(t), \sigma^*, \delta's, \alpha's, \beta's$  all can be estimated consistently with least square estimators and a nonparametric maximum likelihood estimator. It follows that path-specific effects can be estimated with the closed form formula in (9) and (10) by plugging in the estimates. Total

$$\begin{aligned}
 \Delta_{TE}^{Probit} &= E[H(T; s_1) | \mathbf{X}] \\
 &\quad - E[H(T; s_0) | \mathbf{X}] \\
 &= E[H(T; s_1, M_1(s_1), M_2(s_1), M_1(s_1)) | \mathbf{X}] \\
 &\quad - E[H(T; s_0, M_1(s_0), M_2(s_0), M_1(s_0)) | \mathbf{X}] \text{ and}
 \end{aligned}$$

effect is defined as

$$\Omega_{TE} = F_{T(s_0)}(t|\mathbf{X}) - F_{T(s_1)}(t|\mathbf{X}) = F_{T(s_0, M_1(s_0), M_2(s_0), M_1(s_0))}(t|\mathbf{X}) - F_{T(s_1, M_1(s_1), M_2(s_1), M_1(s_1))}(t|\mathbf{X})$$

can be expressed as  $\Delta_{TE}^{Probit} = \Delta_{S \rightarrow Y}^{Probit} + \Delta_{S \rightarrow M_2 \rightarrow Y}^{Probit} + \Delta_{S \rightarrow M_1 Y}^{Probit}$  and

$\Omega_{TE} = \Omega_{S \rightarrow Y_t} + \Omega_{S \rightarrow M_2 \rightarrow Y_t} + \Omega_{S \rightarrow M_1 Y_t}$  respectively. Variances and confidence intervals for  $\Delta^{Probit}$  and  $\Omega$  can be calculated using the non-parametric maximum likelihood estimator and the functional delta method<sup>34,35</sup>.

The proposed method for mediation analyses using semiparametric probit models has been implemented in Matlab code, and those using Aalen additive hazard models and Cox proportional hazard models have been implemented in R code. We term these methods M3S (Multi-Mediation Models for Survival data). Relevant code is available on our website (<http://www.stat.sinica.edu.tw/ythuang/M3S-Probit.zip>; <http://www.stat.sinica.edu.tw/ythuang/M3S-Cox%20and%20M3S-Aalen.zip>). R scripts for the Aalen and Cox models are also in the eAppendix (Section A7).

## DATA APPLICATION

Details of the study design are provided in eAppendix A6. The ethical review is not required for the secondary analyses of the existing and de-identified data. We first conducted one-mediator analyses, similar to Huang et al (2016)<sup>15</sup>, to examine the effect of the HCV viral load ( $S$ ) on the incidence of liver cancer ( $Y$ ) mediated through the follow-up HBV viral load ( $M$ ). Natural direct effect, natural indirect effect, and total effect for the HCV viral load increasing from the minimum to the first quartile of the detectable log-transformed viral level are presented in Table 1. Mediation analyses using probit models estimated that an increase in the HCV viral load decreased the transformed survival time  $H(T)$  using the natural direct effect ( $-0.32$  (95% confidence interval [CI]:  $-0.57, -0.076$ )), which is not through HBV, but increased  $H(T)$  by the natural indirect effect ( $0.19$  (95% CI:  $0.12, 0.26$ ));

the opposite natural direct and natural indirect effects masked each other in the total effect ( $-0.13$  (95% CI:  $-0.38, 0.12$ )). We noted that the non-parametric transformation  $H(\cdot)$  made it less intuitive to interpret the effects in absolute units such as months. Results using Aalen additive hazard models and Cox proportional hazard models revealed a similar pattern (Table 1). While the three models provided similar results under different effect scales, probit models can further characterize effects on the survival probability as a function of the follow-up time. The natural direct and indirect effects, respectively, decreased and increased the survival probability from getting liver cancer monotonically across the follow-up time: for an increase from the minimum to the first quartile of the detectable HCV viral level, the effect independent of the follow-up HBV DNA (natural direct effect) decreased the 15-year probability of not having an HCC event by 0.3%, and the effect mediated by the follow-up HBV DNA (natural indirect effect) increased the probability by 0.18% (Figure 2a). Again, the opposite effects masked each other in total and marginal effects (Figure 2b) where the marginal effect was estimated from the probit model that simply regressed  $H(T)$  on HCV viral load  $S$ , adjusting for covariates  $X$ . Besides the total effect, the marginal effect was another estimate for the overall effect, which did not decompose the effect into the natural direct and indirect effects; but similar to the total effect, the marginal effect was also conditional on  $X$ .

According to the epidemiology of hepatitis in Taiwan, most dual infected participants were chronic hepatitis B carriers superinfected by HCV<sup>17,20</sup>. The literature supports the assumption that HBV is suppressed by the superimposed HCV<sup>18,19,21–23</sup>. Therefore, the HCV ( $S$ )-induced baseline HBV viral load ( $M_1$ ) may be a common cause of the follow-up HBV viral load ( $M_2$ ) and the outcome ( $Y$ ) (Figure 1), which violates the identifiability assumption  $Y(s, m) \prod M(s^*)$  in the single-mediator model. To address the issue, we resorted to the two-mediator model investigating path-specific effects<sup>26</sup> (Table 2). In analyses using Cox models, an increase in the HCV viral load increased the risk of liver cancer through the mechanism of  $\Delta_{S \rightarrow Y}$  with a hazard ratio of 1.17 (95% CI: 1.06, 1.29), but decreased the risk through the mechanism of  $\Delta_{S \rightarrow M_1 Y}$  with a hazard ratio of 0.91 (95% CI: 0.88, 0.94); the HCV viral load did not change the risk through the mechanism only mediated by the follow-up HBV viral load ( $\Delta_{S \rightarrow M_2 \rightarrow Y}$ : 1.00 (95% CI: 0.99, 1.01)); and the total effect was 1.06 (95% CI: 0.96, 1.18). Probit models characterized path-specific effects on the entire survival function, which suggests that  $\Omega_{S \rightarrow Y_t}$  and  $\Omega_{S \rightarrow M_1 Y_t}$ , respectively decreased and increased the survival probability (from getting liver cancer) monotonically across the follow-up time;  $\Omega_{S \rightarrow M_2 \rightarrow Y_t}$  covered the probability difference of zero; and the three effects masked each other in total and marginal effects (Figure 3a and b).

The nonparametric transformation function  $H(\cdot)$  depicted in Figure 4 is a monotonically increasing function with slight non-linearity and was estimated as a step function using the nonparametric maximum likelihood estimator<sup>34</sup>. The normality assumption was also evaluated for probit models using a bootstrap procedure<sup>34</sup>, which revealed no departure from the normal distribution.

## DISCUSSION

We present multi-mediator analyses under semiparametric probit models, additive hazard models, and Cox proportional hazard models and provide closed-form expressions for path-specific effects. The closed-form expression for Cox models can only be obtained under the rare-outcome assumption, which may be a limitation for its application. Mediation analyses under additive hazard models and Cox models characterize effects on specific scales; additive hazard models focus on the hazard difference and Cox models focus on the log hazard ratio. Probit models examine the effect on the transformed survival time difference as well as the difference in survival probability across the follow-up time. The closed-form expressions under the three models have intuitive interpretations using a causal diagram labeled with effect parameters on each arrow. Effects estimated from the three models convey complementary information. The effect on survival probability by probit models is very general and may be used to estimate any summary outcome such as average survival time or 5-year survival risk. The effects estimated using probit models is in the opposite direction compared to the other two models because probit models characterize the difference in survival time and survival probability, which is qualitatively the opposite of the hazard difference or the hazard ratio. Because of the lack of conjugacy between product of normal and normal distributions, we may not be able to provide closed-form expressions for a mediator-by-mediator interaction effect (e.g.,  $\beta_{M_1M_2} M_1 M_2$ ) or a non-linear mediator effect (e.g.,  $\beta_{M_1^2} M_1^2$ ). One may have to utilize the simulation-based approach to approximate these effects<sup>33</sup>. In the hepatitis study, there was no evidence for such effects.

Causal multi-mediator models have several advantages. First, models with multiple mediators are able to characterize more comprehensive mechanistic structure. It is often not plausible to assume that the effect of an exposure on an outcome is mediated only by a single mediator. Secondly, path-specific effects in multi-mediator models constitute an approach to address identifiability issue due to the undue influence by the exposure-induced mediator-outcome confounder<sup>26</sup>, e.g., the baseline HBV viral load in the hepatitis study. Finally, being able to incorporate more potential mediators renders the identifiability assumptions (5) and (6) more plausible<sup>25</sup> because the inclusion of additional mediators makes it less likely to miss alternative pathways from the exposure to the outcome.

The analyses of the hepatitis study were limited to participants with no missing data on the hepatitis B and C viral load and covariates, and assumed that the viral load was accurately measured. Mediation analyses accounting for missing data and measurement error have been developed<sup>36,37</sup>, but how to incorporate in the proposed multi-mediator analyses for survival outcomes may require additional development.

The finding that  $\Delta_{S \rightarrow M_1 Y}$  had stronger effects than  $\Delta_{S \rightarrow M_2 \rightarrow Y}$  may not exclude the possibility that the follow-up HBV DNA is a mediator since  $\Delta_{S \rightarrow M_1 Y}$  contains the effect mediated by both baseline and follow-up HBV DNA levels. However, our analyses does not support that there existed an effect mediated only through the follow-up HBV. The finding has an important public health implication: Taiwanese patients with dual infection of hepatitis B and C may not require more intensive follow-up for HBV activity than those with HBV

single infection if their baseline HBV viral load is low. Analytically, one may propose a mediation model with the baseline and follow-up HBV viral load as a single set of mediators, and interrogates the longitudinal mediation effect through HBV<sup>38</sup>. Such analyses evaluate an overall mediation effect but would not be able to provide the respective effect measures of different mechanisms/paths.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

**Sources of Financial Support:** NIH/NCI 5R03CA 182937-02, NIH/NIA 1R01AG048825-01

## References

1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986; 51(6):1173–82. [PubMed: 3806354]
2. MacKinnon, D. Introduction to statistical mediation analysis. New York: Taylor and Francis; 2008.
3. Huang YT, Freeman JR, Yang HI, Liu J, Lee MH, Chen CJ. Mediation effect of hepatitis B and C on mortality. *Eur J Epidemiol.* 2016
4. Pan WC, Wu CD, Chen MJ, Huang YT, Chen CJ, Su HJ, Yang HI. Fine Particle Pollution, Alanine Transaminase, and Liver Cancer: A Taiwanese Prospective Cohort Study (REVEAL-HBV). *J Natl Cancer Inst.* 2016; 108(3)
5. Song Y, Huang YT, Song Y, Hevener AL, Ryckman KK, Qi L, LeBlanc ES, Kazlauskaitė R, Brennan KM, Liu S. Birthweight, mediating biomarkers and the development of type 2 diabetes later in life: a prospective study of multi-ethnic women. *Diabetologia.* 2015; 58(6):1220–30. [PubMed: 25567102]
6. VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, Han Y, Spitz MR, Shete S, Wu X, Gaborieau V, Wang Y, McLaughlin J, Hung RJ, Brennan P, Amos CI, Christiani DC, Lin X. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol.* 2012; 175(10):1013–20. [PubMed: 22306564]
7. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol.* 2013; 31(2):142–7. [PubMed: 23334450]
8. Rubin D. Bayesian inference of causal effects. *Annals of Statistics.* 1978; 6:34–58.
9. Pearl, J. Direct and indirect effects Seventeenth Conference on Uncertainty and Artificial Intelligence. San Francisco, CA: Morgan Kaufmann; 2001. p. 411-420.
10. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992; 3(2):143–55. [PubMed: 1576220]
11. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods.* 2010; 15(4):309–34. [PubMed: 20954780]
12. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol.* 2010; 172(12):1339–48. [PubMed: 21036955]
13. Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology.* 2011; 22(4): 575–81. [PubMed: 21552129]
14. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology.* 2011; 22(4):582–5. [PubMed: 21642779]

15. Huang YT, Yang HI, Liu J, Lee MH, Freeman JR, Chen CJ. Mediation Analysis of Hepatitis B and C in Relation to Hepatocellular Carcinoma Risk. *Epidemiology*. 2016; 27(1):14–20. [PubMed: 26443934]
16. Huang YT, Jen CL, Yang HI, Lee MH, Su J, Lu SN, Iloeje UH, Chen CJ. Lifetime risk and sex difference of hepatocellular carcinoma among patients with chronic hepatitis B and C. *J Clin Oncol*. 2011; 29(27):3643–50. [PubMed: 21859997]
17. Chen CJ, Wang LY, Yu MW. Epidemiology of hepatitis B virus infection in the Asia-Pacific region. *J Gastroenterol Hepatol*. 2000; 15(Suppl):E3–6. [PubMed: 10921373]
18. Koike K, Yasuda K, Yotsuyanagi H, Moriya K, Hino K, Kurokawa K, Iino S. Dominant replication of either virus in dual infection with hepatitis viruses B and C. *J Med Virol*. 1995; 45(2):236–9. [PubMed: 7539830]
19. Liaw YF. Role of hepatitis C virus in dual and triple hepatitis virus infection. *Hepatology*. 1995; 22(4 Pt 1):1101–8. [PubMed: 7557857]
20. Lu SN, Chen HC, Tang CM, Wu MH, Yu ML, Chuang WL, Lu CF, Chang WY, Chen CJ. Prevalence and manifestations of hepatitis C seropositivity in children in an endemic area. *Pediatr Infect Dis J*. 1998; 17(2):142–5. [PubMed: 9493811]
21. Schuttler CG, Fiedler N, Schmidt K, Repp R, Gerlich WH, Schaefer S. Suppression of hepatitis B virus enhancer 1 and 2 by hepatitis C virus core protein. *J Hepatol*. 2002; 37(6):855–62. [PubMed: 12445429]
22. Shih CM, Lo SJ, Miyamura T, Chen SY, Lee YH. Suppression of hepatitis B virus expression and replication by hepatitis C virus core protein in HuH-7 cells. *J Virol*. 1993; 67(10):5823–32. [PubMed: 8396658]
23. Tsiquaye KN, Portmann B, Tovey G, Kessler H, Hu S, Lu XZ, Zuckerman AJ, Craske J, Williams R. Non-A, non-B hepatitis in persistent carriers of hepatitis B virus. *J Med Virol*. 1983; 11(3):179–89. [PubMed: 6408222]
24. Avin, C., Spitzer, I., Pearl, J. Identifiability of path-specific effects. International Joint Conference on Artificial Intelligence; Edinburgh, Scotland. 2005. p. 357-363.
25. VanderWeele TJ, Vansteelandt S. Mediation Analysis with Multiple Mediators. *Epidemiol Method*. 2014; 2(1):95–115. [PubMed: 25580377]
26. Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*. 2014; 25(2):300–6. [PubMed: 24487213]
27. Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics*. 2011; 67(3):1028–38. [PubMed: 21306353]
28. Huang YT, Pan WC. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*. 2016; 72(2):402–413. [PubMed: 26414245]
29. Aalen, O. A model for non-parametric regression analysis of counting process. In: Klonecki, W.Kozek, A., Rosinski, J., editors. *Lecture Notes in Statistics, Mathematical Statistics and Probability Theory*. New York: Springer-Verlag; 1980. p. 1-25.
30. Cox D. Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
31. Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med*. 1992; 11(14–15):1871–9. [PubMed: 1480879]
32. Tchetgen Tchetgen EJ. On causal mediation analysis with a survival outcome. *Int J Biostat*. 2011; 7(1) Article 33.
33. Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *Am J Epidemiol*. 2014; 179(4):513–8. [PubMed: 24264291]
34. Huang YT, Cai T. Mediation analysis for survival data using semiparametric probit models. *Biometrics*. 2016; 72(2):563–574. [PubMed: 26618735]
35. Zeng D, Lin D. Maximum Likelihood Estimation in Semiparametric Models with Censored Data. *Journal of the Royal Statistical Society, Series B*. 2007; 69(4):507–564.
36. Zhang Z, Wang L. Methods for mediation analysis with missing data. *Psychometrika*. 2013; 78(1): 154–84. [PubMed: 25107523]

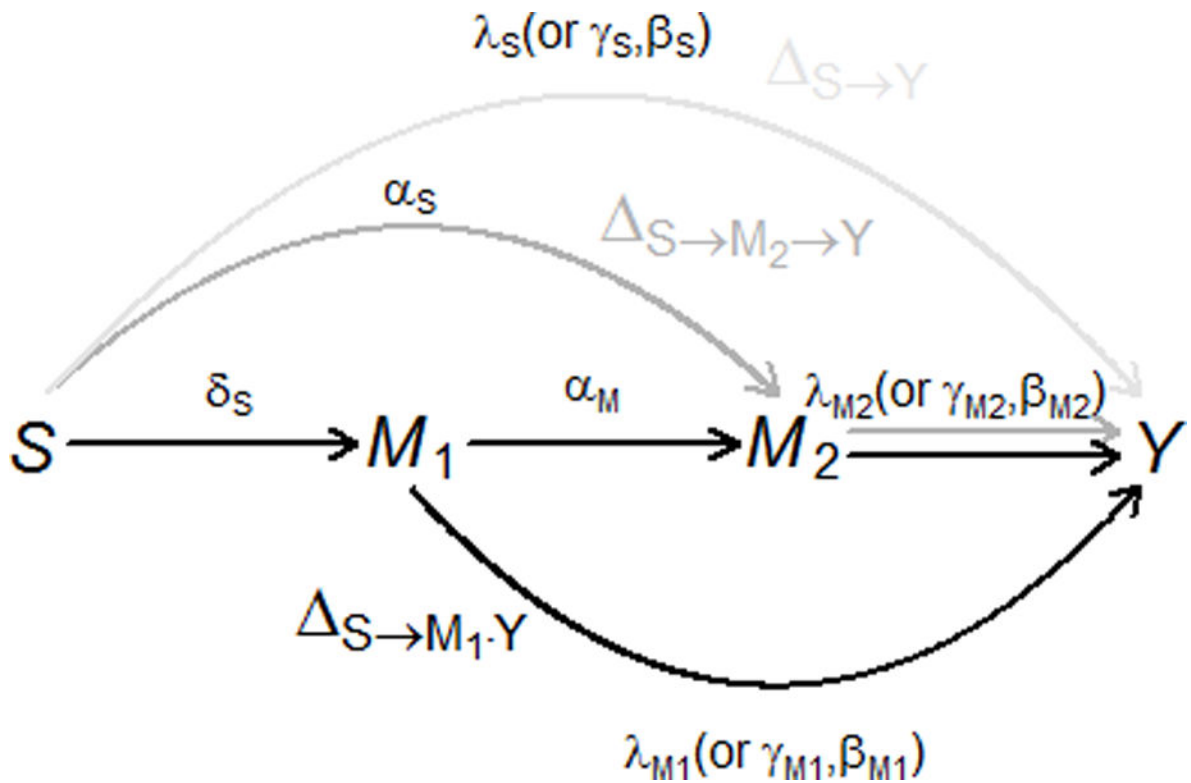
37. VanderWeele TJ, Valeri L, Ogburn EL. The role of measurement error and misclassification in mediation analysis: mediation and measurement error. *Epidemiology*. 2012; 23(4):561–4. [PubMed: 22659547]
38. Bind MA, Vanderweele TJ, Coull BA, Schwartz JD. Causal mediation analysis for longitudinal data with exogenous exposure. *Biostatistics*. 2016; 17(1):122–34. [PubMed: 26272993]

Author Manuscript

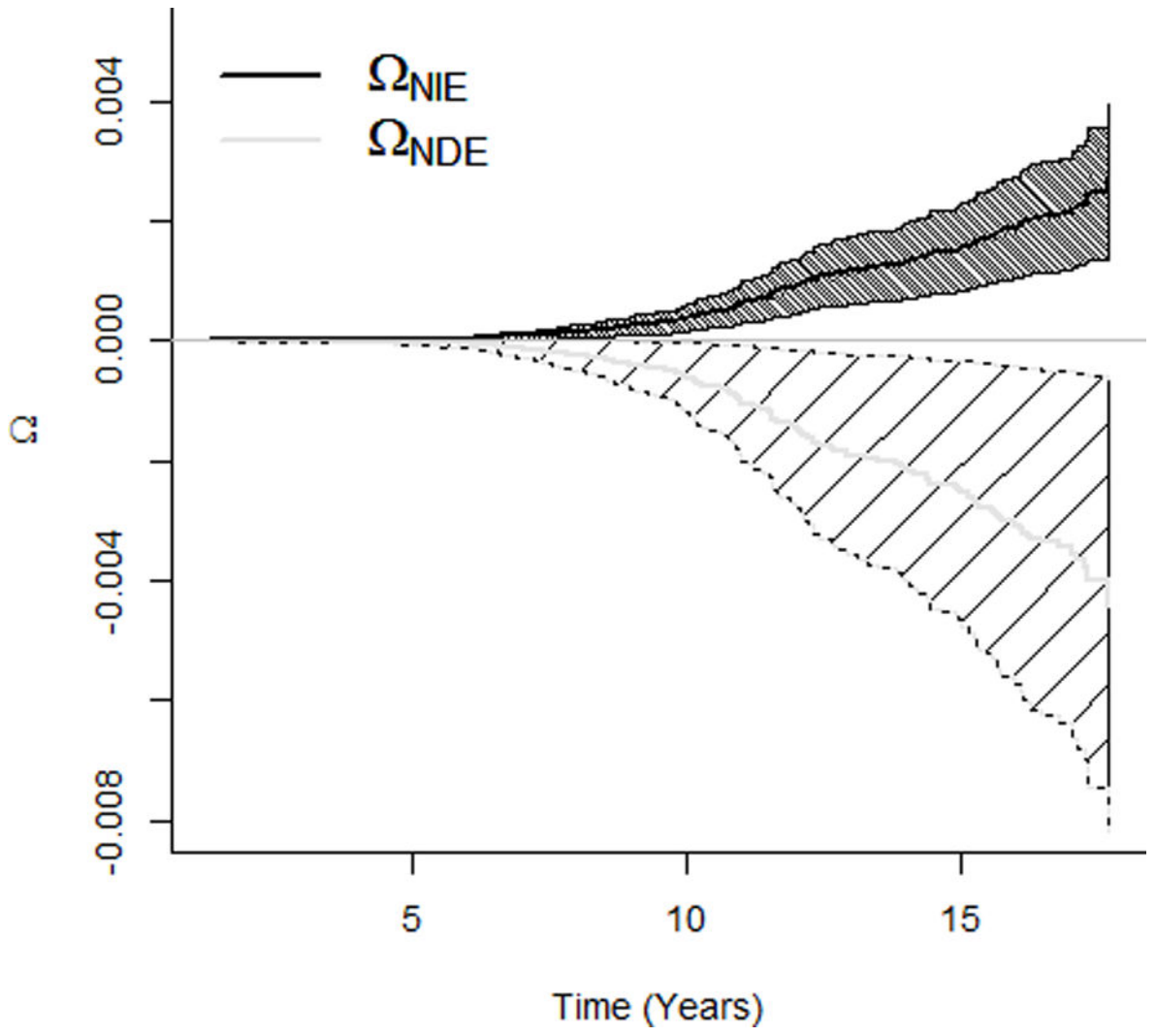
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 1.** The causal diagram of the baseline HCV viral load  $S$ , the baseline HBV viral load  $M_1$ , the follow-up HBV viral load  $M_2$ , and the outcome of interest  $Y$ . The outcome  $Y$  is  $\lambda(t|X)$ ,  $\log\lambda(t|X)$  and  $H(T)$  in additive hazard model, Cox proportional hazard model, and probit model, respectively. Three path-specific effects are in different gray scales:  $\Delta_{S \rightarrow Y}$ , the effect of the baseline HCV viral load  $S$  on the outcome independent of the HBV viral load  $M_1$  and  $M_2$  is in light gray;  $\Delta_{S \rightarrow M_2 \rightarrow Y}$ , the effect of the HCV viral load  $S$  on the outcome mediated only through the follow-up HBV viral load  $M_2$  (i.e., not through the baseline HBV viral load  $M_1$ ) is in dark gray;  $\Delta_{S \rightarrow M_1 \cdot Y}$ , the effect mediated through the baseline HBV viral load  $M_1$ , and possibly through the follow-up HBV viral load  $M_2$  is in black. Greek letters are the regression coefficients in models (2)–(4) (or (6), (8)) which each arrow corresponds to. HCV indicates hepatitis C virus, HBV hepatitis B virus.



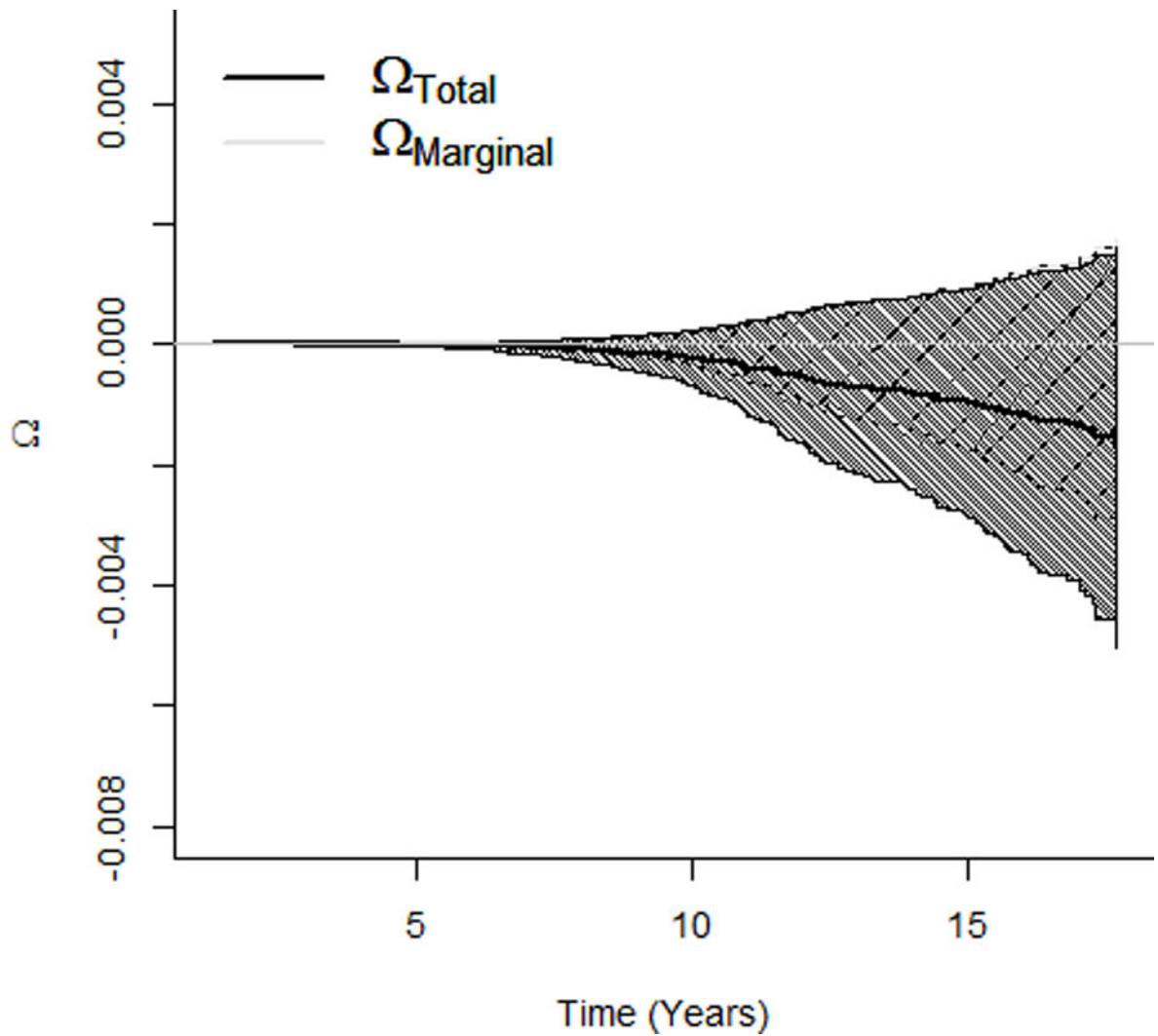
Author Manuscript

Author Manuscript

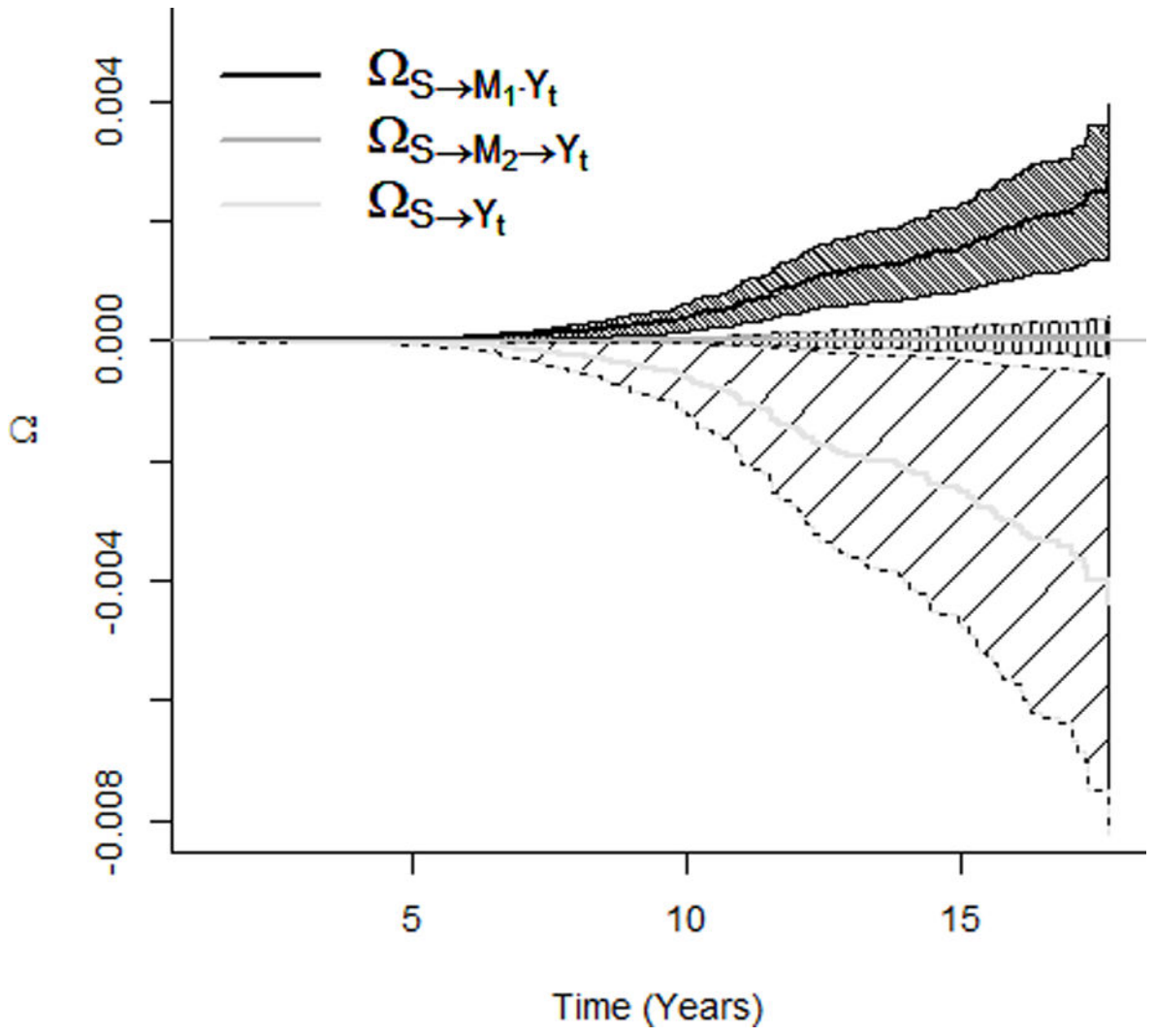
Author Manuscript

Author Manuscript





**Figure 2.** Effects of the HCV viral load (increasing from the minimum to the first quartile of the detectable HCV viral level) on survival probability of the liver cancer incidence. a. NDE and NIE on survival probability of getting liver cancer. b. Total and marginal effects on survival probability. HCV indicates hepatitis C virus, NDE natural direct effect, NIE natural indirect effect.

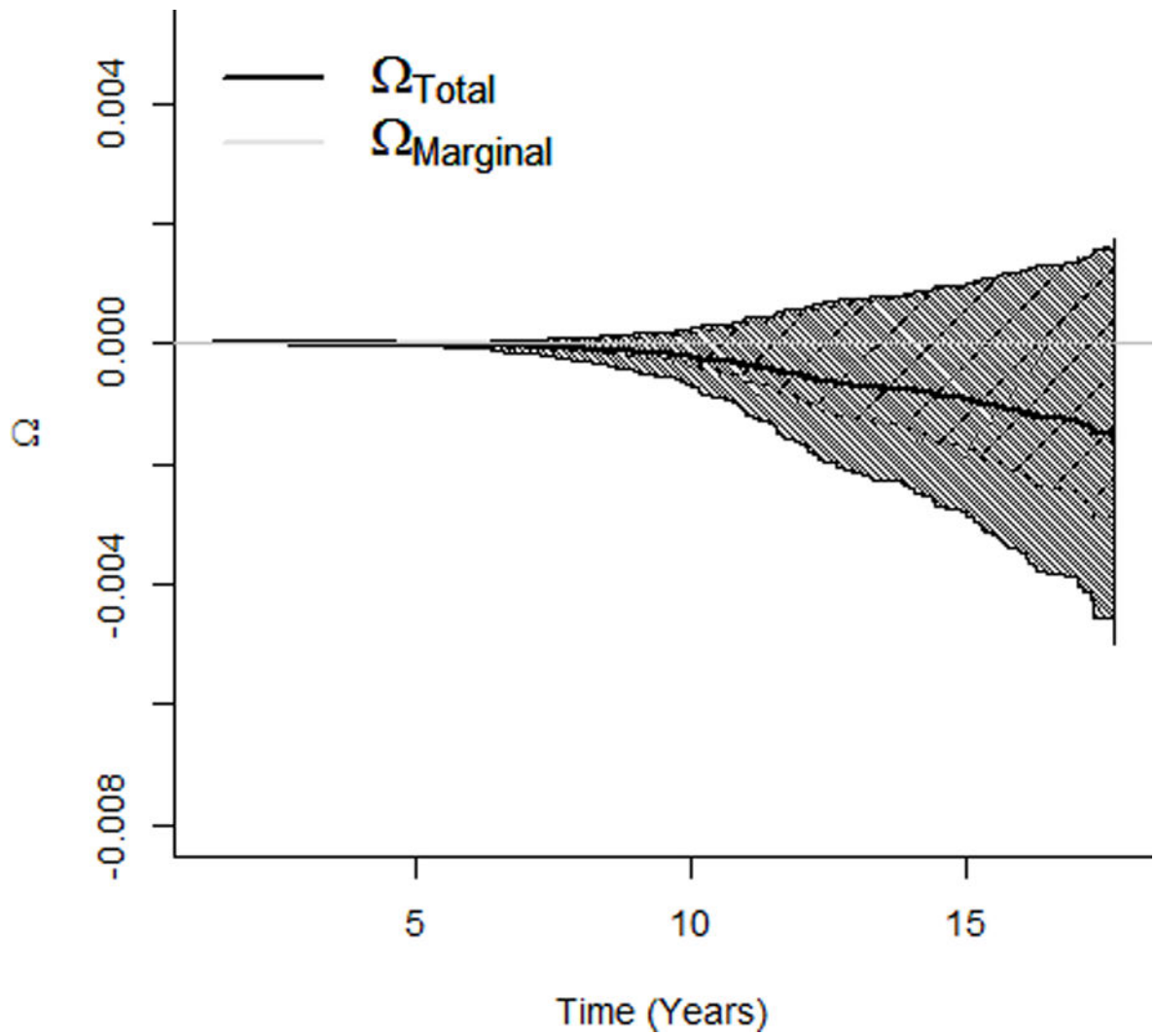


Author Manuscript

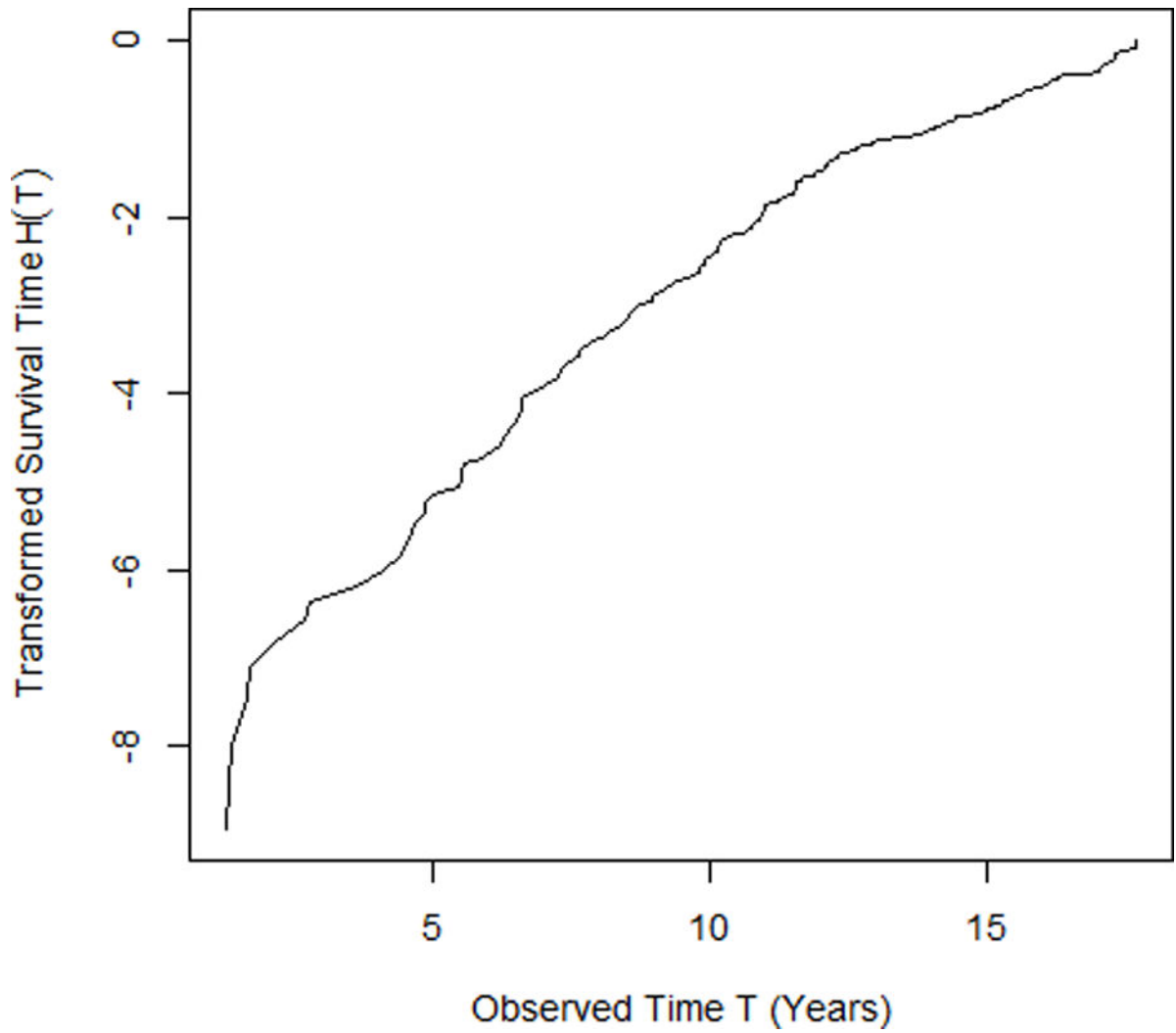
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3.** Effects of the HCV viral load (increasing from the minimum to the first quartile of the detectable HCV viral level) on survival probability of the liver cancer incidence. a. Path-specific effects on survival probability of getting liver cancer. b. Total and marginal effects on survival probability of getting liver cancer. HCV indicates hepatitis C virus.



**Figure 4.** The transformed survival time  $H(T)$  versus the observed survival time  $T$  under the semiparametric probit model in (8).

**Table 1**

Natural direct, natural indirect, and total effects of hepatitis C viral load (increase from the minimum to the first quartile of the detectable HCV viral level) on liver cancer incidence mediated through the follow-up HBV viral load.

	<b>Effect Estimate</b>	<b>95% Confidence Interval</b>
Probit; Effect scale: Difference in transformed survival time		
Natural Direct Effect	-0.32	(-0.57, -0.076)
Natural Indirect Effect	0.19	(0.12, 0.26)
Total Effect	-0.13	(-0.38, 0.12)
Aalen; Effect scale: Difference in hazard (per 1000 person-year)		
Natural Direct Effect	0.84	(-0.066, 1.75)
Natural Indirect Effect	-0.36	(-0.53, -0.21)
Total Effect	0.49	(-0.43, 1.40)
Cox; Effect scale: Hazard ratio		
Natural Direct Effect	1.14	(1.03, 1.26)
Natural Indirect Effect	0.92	(0.89, 0.95)
Total Effect	1.05	(0.95, 1.16)

HCV indicates hepatitis C virus, HBV hepatitis B virus.

**Table 2**

Path-specific effects and total effects of HCV viral load ( $S$ ) increase from the minimum to the first quartile of the detectable HCV viral level) on liver cancer incidence ( $Y$ ) mediated through the baseline HBV viral load ( $M_1$ ) and the follow-up HBV viral load ( $M_2$ ).

	Effect Estimate	95% Confidence Interval
Probit model; Effect scale: Difference in transformed survival time		
$\Delta_{S \rightarrow Y}$	-0.32	(-0.57, -0.073)
$\Delta_{S \rightarrow M_2 \rightarrow Y}$	0.0042	(-0.018, 0.026)
$\Delta_{S \rightarrow M_1 Y}$	0.19	(0.12, 0.26)
$\Delta_{TE}$	-0.13	(-0.38, 0.13)
Aalen model; Effect scale: Difference in hazard (per 1000 person-year)		
$\Delta_{S \rightarrow Y}$	0.92	(0.021, 1.82)
$\Delta_{S \rightarrow M_2 \rightarrow Y}$	-0.010	(-0.067, 0.044)
$\Delta_{S \rightarrow M_1 Y}$	-0.42	(-0.60, -0.27)
$\Delta_{TE}$	0.49	(-0.43, 1.40)
Cox model; Effect scale: Hazard ratio		
$\Delta_{S \rightarrow Y}$	1.17	(1.06, 1.29)
$\Delta_{S \rightarrow M_2 \rightarrow Y}$	1.00	(0.99, 1.01)
$\Delta_{S \rightarrow M_1 Y}$	0.91	(0.88, 0.94)
$\Delta_{TE}$	1.06	(0.96, 1.18)

HCV indicates hepatitis C virus, HBV hepatitis B virus.