Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2017 August ; 66(4): 741–757. doi:10.1111/rssc.12188.

Causal Mediation Analysis for the Cox Proportional Hazards Model with a Smooth Baseline Hazard Estimator

Wei Wang^{*,†} and

Center of Biostatistics and Bioinformatics, New Guyton Research Building G562, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216

Jeffrey M. Albert

Department of Epidemiology and Biostatistics, School of Medicine WG-82S, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, Phone: (216) 368-1968, Fax: (216) 368-3970, jma13@case.edu

Abstract

An important problem within the social, behavioral, and health sciences is how to partition an exposure effect (e.g. treatment or risk factor) among specific pathway effects and to quantify the importance of each pathway. Mediation analysis based on the potential outcomes framework is an important tool to address this problem and we consider the estimation of mediation effects for the proportional hazards model in this paper. We give precise definitions of the total effect, natural indirect effect, and natural direct effect in terms of the survival probability, hazard function, and restricted mean survival time within the standard two-stage mediation framework. To estimate the mediation effects on different scales, we propose a mediation formula approach in which simple parametric models (fractional polynomials or restricted cubic splines) are utilized to approximate the baseline log cumulative hazard function. Simulation study results demonstrate low bias of the mediation effect estimators and close-to-nominal coverage probability of the confidence intervals for a wide range of complex hazard shapes. We apply this method to the Jackson Heart Study data and conduct sensitivity analysis to assess the impact on the mediation effects inference when the no unmeasured mediator-outcome confounding assumption is violated.

Keywords

mediation analysis; proportional hazards model; natural indirect effect; fractional polynomials; restricted cubic splines; mediation formula; sensitivity analysis

1. INTRODUCTION

Important questions within the social, behavioral, and health sciences are to determine the extent to which the effect of the exposure on the outcome is mediated through a causal intermediate variable, or mediator. The most commonly used method of mediation analysis

* Author for correspondence Phone: (601) 984-4361, Fax: (601) 984-1939, wwang@umc.edu.

† Present address: Division of Biostatistics, Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD, 20993, Phone: (301) 796-3060, wei.wang2@fda.hhs.gov

involves the fitting of a succession of linear regression models within the framework of linear structural equation modeling (LSEM) (Baron and Kenny, 1986), and the mediation effect can be assessed by difference in coefficients approach (MacKinnon and Dwyer, 1993) or product of coefficients approach (MacKinnon *et al.*, 2002). Recently, causal mediation analysis, namely methods based on a causal model (usually, potential outcomes) framework, has been developed to view mediation in terms of conceived manipulations (Robins and Greenland, 1992; Albert, 2008).

Some recent research has focused on mediation with the right censored survival outcomes. For example, the comorbidity variable (hypertension and different comorbidity scores) mediated the relation between age and survival among metastatic breast cancer patients (Jung *et al.*, 2012). A second example, which will be used in this paper, is the effect of cigarette smoking on the risk of stroke among the African-Americans in the Jackson Heart Study (JHS). It is well known that cigarette smoking strongly affects the risk of experiencing long-term ischemic and hemorrhage stroke at least partly through induced high blood pressure. Therefore, an important clinical and public health question is, to what extent can the adverse effects of cigarette smoking be mitigated by targeting its mediator blood pressure?

The survival analysis model most frequently employed in the literature is the Cox proportional hazards model which assumes that changes in a covariate have a multiplicative effect on the baseline hazard. A recent approach to mediation analysis in the survival context extends the LSEM method while assuming the Cox model. In this approach, coefficients representing hazard ratios of exposure with and without adjusting for the potential mediator are estimated, and the difference in the estimated coefficients is taken as a measure of the indirect or mediated effect (Jung *et al.*, 2012). The major limitations of this method are that the observed change in hazard ratios from two Cox models does not have a ‘causal’ interpretation and the important assumption of proportional hazards can never be satisfied for both models with and without the mediator. Lange and Hansen (2011) proposed a ‘product of coefficients’ approach to estimate the natural direct and indirect effects for the Aalen’s additive hazards model with normally distributed mediator. VanderWeele (2011) extended Lange and Hansen’s method, and proved that similar mediation effect estimates can be defined for the accelerated failure time model and for the Cox proportional hazards model; however, the method for the latter model depends on a rare outcome assumption. Tchetgen Tchetgen (2011) further developed some robust estimators of natural direct and indirect effects under the proportional hazards model and the additive hazards model, as well as a novel semiparametric sensitivity analysis technique to assess the impact of violation of the ignorability assumption for the mediator. Another approach used an inverse odds ratio-weighted estimation method to decompose total effects estimated in a Cox proportional hazards model into natural direct and indirect causal effects (Tchetgen Tchetgen, 2013). In the survival context, there are multiple scales by which the total effect might be defined and decomposed into the natural direct and indirect effects; examples include the survival probability, the hazard function, and mean survival time (VanderWeele, 2011). However, currently available mediation analysis approaches are all based on semiparametric survival models for which the natural direct and indirect effects may be analytically tractable on certain scales but not on others, e.g. the hazard decomposition for the Aalen’s additive

hazards model (Lange and Hansen, 2011), and the mean survival time decomposition for the accelerated failure time model (VanderWeele, 2011).

As a general causal model approach to mediation analysis, Pearl (2012) proposed a mediation formula for the assessment of natural direct and indirect effects, which is applicable to non-linear models for discrete, continuous and time-to-event outcome variables. This approach involves the estimation of relevant expected potential outcomes as integration/ summation of conditional mean of response variable over the probability density distribution of the mediators under the key ‘sequential ignorability’ assumption (Imai *and others*, 2010; VanderWeele, 2011; Lange and Hansen, 2011; Wang and Albert, 2012). An extension of the mediation formula approach to survival data under the proportional hazards assumption allowing for different mediation effect scales, requires a parametric estimation of the baseline hazard function, or more generally, the baseline distribution function, after fitting a Cox model. We consider using fractional polynomial (FP) functions (Royston and Altman, 1994; Sauerbrei and Royston, 1999) and restricted cubic regression splines (RCS) (Durrelman and Simon, 1989; Royston and Parmar, 2002; Rutherford *et al.*, 2015) to approximate the baseline log cumulative hazard function, both of which have relatively simple mathematical forms. In the present paper, focusing on a scenario in which the time-independent exposure and mediator are measured before the event/censoring process starts, we propose approaches to causal mediation analysis for a survival outcome under the Cox proportional hazards model that provide the mediation effect estimates on the restricted mean survival time (RMST), survival probability and hazard function scales. Simple parametric (FP and RCS) models are utilized to approximate baseline log cumulative hazard functions which are needed to estimate certain causal quantities through a mediation formula approach.

The structure of the paper is as follows. In Section 2, we define the natural indirect, natural direct, and total causal effects on mean survival time, survival probability and hazard function scales within a survival context. Section 3 presents a mediation formula approach to estimate the causal quantities under given identification assumptions. Then we describe FP functions, RCS, and other estimation issues in Section 4. Section 5 describes the method for sensitivity analysis that can be implemented by researchers to assess the impact on the mediation effects inference when the no unmeasured mediator-outcome confounding assumption is violated. A simulation study is reported in Section 6 to assess the mediation effects estimators using simulated complex baseline hazard functions under the proportional hazards model. Section 7 describes the application of the proposed method to the Jackson Heart Study data. Discussion and suggestions for further research are presented in Section 8.

2. CAUSAL QUANTITIES DEFINITION IN A SURVIVAL CONTEXT

Consider the standard two-stage mediation framework including a binary exposure or treatment indicator (X), a mediator (M), and a time-to-event outcome $Y = (T, \delta)$ ($T = \min(\tilde{T}, C)$ is the censored failure time, and $\tilde{T}, C$ are the underlying failure and censoring time, δ is a non-censoring indicator), where X may affect Y directly and/or X may affect M , which then affects Y . The relations among these variables are depicted in Figure 1. We assume that $\tilde{T}$ and C are independent given (X, M, W) , where W is baseline covariate that may affect time-

to-event T in the analysis model. To define the causal mediation effects, we use nested potential outcomes (Albert, 2008). Let $T(x, m)$ describe the time to event if the exposure was set to x and the mediator set to m , and $T(x, M(x'))$ denote the event time if the exposure was set to x , and the mediator is set to its potential outcome that would be observed if the exposure was set to x' . Extending this idea to probability distributions, we can define $h(t^*; x, M(x'))$ as the potential hazard rate (instantaneous probability of failure) at time t^* and $S(t^*; x, M(x'))$ as the probability of survival until time t^* , under the indicated manipulation of X and M , specifically exposure X was set to x , but the mediator, M , set to its potential outcome if X was set to x' (VanderWeele, 2011; Lange and Hansen, 2011). Under the standard two-stage mediation model, for time-to-event T , we could define the natural indirect effect and natural direct effect and also decompose the total exposure effect on mean survival time as follows,

$$\begin{aligned} IE_T(x) &= E\{T(x, M(1))\} - E\{T(x, M(0))\} \\ DE_T(1-x) &= E\{T(1, M(1-x))\} - E\{T(0, M(1-x))\} \\ TE_T &= E\{T(1, M(1))\} - E\{T(0, M(0))\} = IE_T(x) + DE_T(1-x) \end{aligned} \quad (1)$$

$IE_T(x)$ represents the difference for time to event between two mean potential outcomes that would result under exposure status x , but where the mediator takes values that would result under the two different exposure statuses (1 vs. 0). The proportion of exposure effect on the mean survival time due to the mediator is defined as,

$$P_T(x) \equiv IE_T(x) / TE_T \quad (2)$$

Similarly, as VanderWeele (2011) showed, we could also decompose the overall difference in survival functions and log-hazards at specific time t^* and define related mediation proportions,

$$\begin{aligned} TE_S(t^*) &= S(t^*; 1, M(1)) \\ &\quad - S(t^*; 0, M(0)) \\ &= IE_S(t^*; x) \\ &\quad + DE_S(t^*; 1-x) \\ &= [S(t^*; x, M(1)) - S(t^*; x, M(0))] \\ &\quad + [S(t^*; 1, M(1-x)) - S(t^*; 0, M(1-x))] \end{aligned} \quad (3)$$

$$\begin{aligned}
& \log TE_h(t^*) \\
&= \log \{h(t^*; 1, M(1))\} \\
&\quad - \log \{h(t^*; 0, M(0))\} \\
&= \log IE_h(t^*; x) \\
&\quad + \log DE_h(t^*; 1 - x) \\
&= [\log \{h(t^*; x, M(1))\} - \log \{h(t^*; x, M(0))\}] \\
&\quad + [\log \{h(t^*; 1, M(1 - x))\} - \log \{h(t^*; 0, M(1 - x))\}] \quad (4)
\end{aligned}$$

Of note, the defined causal quantities on survival probability and hazard scales are functions of time t , however, the causal quantities on mean survival times are independent of time t as t is integrated out in the computation of the expected value. In addition, while we only consider $x = 0$ and focus on the estimands $IE(0)$ and $DE(1)$ throughout the paper, the approaches for $IE(1)$ and $DE(0)$ estimation are completely analogous.

3. MEDIATION ANALYSIS WITH SURVIVAL DATA

We present the identification results for the natural direct and indirect effects using the nested potential outcome framework described above. Under a particular version of the sequential ignorability assumption, the mediation effect estimators are identified nonparametrically; specifically, we assume that the following two statements of conditional independence hold, which extends Imai *et al.* (2010).

Assumption 1 (Sequential Ignorability)

$$\{T(x', m), M(x)\} \perp\!\!\!\perp X | W=w \quad (5)$$

$$T(x', m) \perp\!\!\!\perp M(x) | W=w, X=x \quad (6)$$

Thus, the exposure (X) is first assumed to be ignorable, that is, statistically independent of potential outcomes and potential mediators given the baseline covariates (W), and then mediator (M) variable is assumed to be ignorable given the observed value of the exposure as well as baseline covariates. We also make the consistency assumption for the outcome and the mediator; for example, the potential outcome of $T(x, m)$ is equal to the observed survival outcome if the exposure of this individual is x and the mediator is m . (VanderWeele and Vansteelandt, 2009). Under Assumption 1, the functions of potential outcome for time-to-event $T\left(g\left\{T\left(x, M\left(x'\right)\right)\right\}\right)$ used to define the natural direct and indirect effects on different scales can be written as follows,

$$g\left\{T\left(x, M\left(x'\right)\right)\right\} = \iint g\{T|M=m, X=x, W=w\} dF_{M|X=x', W=w}(m) dF_W(w) \quad (7)$$

where $F_W(w)$ and $F_{M|X=x', W=w}(m)$ represent the distribution function of W for the reference population and the conditional distribution function of mediator M given X and W , respectively. Alternatively, we may sum over the empirical distribution of W in the reference population (e.g. exposed group) instead of integrating over the unknown distribution for W as indicated in (7). The term $g\left\{T\left(x, M\left(x'\right)\right)\right\}$ represents the function of potential outcome $T\left(x, M\left(x'\right)\right)$; for example, if g represents the expectation function, equation (7) provides potential outcome means that can be used to estimate the mediation effects on the mean survival time scale, or if g represents the survival function, equation (7) can be used to estimate the mediation effects on the survival function scale. Therefore, equation (7) provides the connection between the potential and observed outcomes and allows the estimation of the functions of potential outcome for time-to-event T using the observed data. The proof of equation (7) using Assumption 1 and the consistency assumption can be found in Appendix A of the online supporting material.

4. ASSOCIATION MODEL SPECIFICATION AND OTHER ESTIMATION ISSUES

Estimation of the mediation effects for survival data on different scales proceeds by estimating the association parameters in the mediation equation (7). In this paper, we consider semi-parametric Cox proportional hazards models and normally distributed continuous mediators.

4.1. Regression models

We assume the following general proportional hazards model for time-to-event T which is defined through the log cumulative hazard function as,

$$\log(H(t; X=x, M=m, W=w)) = \log(H_0(t)) + \beta_1 x + \beta_2 m + \beta' w \quad (8)$$

where $H_0(t)$ is the baseline cumulative hazard function at time t . We also consider appropriate regression models for the normally distributed continuous mediator M ,

$$M = \alpha_0 + \alpha_1 X + \alpha' W + \varepsilon, \text{ where } \varepsilon \sim N(0, \sigma^2) \quad (9)$$

In these two models, β_1 , β_2 , α_0 and α_1 are regression parameters, and W is baseline covariate vector with corresponding coefficient vectors β and α .

4.2. Baseline log cumulative hazard function approximation

For the defined mediation effects on different scales in equation (1), (3) and (4), identification of these causal quantities through equation (7) under the sequential ignorability assumption requires the baseline hazard function specification (specifically, $g\{T|M=m, X=x, W=w\}$ in equation (7)). An ideal approach for estimating the baseline

hazard function would be simple in form, suitable for calculating the mediation effects and allow this function to be approximated to a sufficient degree of accuracy. In the present paper, we consider the fractional polynomials and the restricted cubic regression splines, both of which have relatively simple mathematical forms and are applicable to a broad range of baseline hazard functions in biomedical research (Royston, 2011). The approach taken here is to model to logarithm of the baseline cumulative hazard function as a parametric function of survival time t in both families.

4.2.1 Fractional Polynomials—The logarithm of a baseline cumulative hazard function ($\ln H_0(t)$) is approximated with an FP function of degree $m > 0$ for an argument $t > 0$ which

is represented as $\theta_0 + \sum_{j=1}^m \theta_j t^{p_j}$. The powers $p_1 < p_2 < \dots < p_m$ are positive or negative integers or fractions selected from a restricted set, ' $P = \{-2, -1, -0.5, 0, 0.5, 1, 2, \max(3, m)\}$ ', where t^0 by convention denotes $\ln(t)$. In addition, the definition also includes possible 'repeated powers', for example, for an FP of degree $m = 3$ with powers $p = (-1, -1, 3)$, we have $\ln H_0(t) = \theta_0 + \theta_1 t^{-1} + \theta_2 t^{-1} \ln(t) + \theta_3 t^3$. An FP model of degree m is considered to have $2m$ degrees of freedom (d.f.) (excluding θ_0): 1 degree of freedom for coefficient θ_j and 1 degree of freedom for each power p_j . Sauerbrei and Royston (1999) argued that degree 2 is sufficient in most medical applications, and we only consider models in t of degree 1 or 2, with d.f. 2 or 4 respectively, therefore the number of total FP models we considered is $8(m = 1) + 36(m = 2) = 44$.

4.2.2 Restricted cubic splines—The logarithm of the baseline cumulative hazard function is modeled as a restricted cubic spline function $s(\ln(t); \gamma)$ of an argument $\ln(t)$ with $K - 1$ interior knots $k_1 < \dots < k_K$ and two boundary knots $k_{min} < k_1, k_{max} > k_K$, which is a weighted sum of series of basis functions, $\ln(t), v_1(\ln(t)), \dots, v_K(\ln(t))$. The basis functions $v_j(\ln(t))$ are derived from the cubic polynomial segments defined on the intervals between the knots and the restricted cubic splines are defined as cubic splines constrained to be linear beyond the boundary knots k_{min}, k_{max} . The spline is defined as,

$$s(\ln(t); \gamma) = \gamma_0 + \gamma_1 \ln(t) + \gamma_2 v_1(\ln(t)) + \dots + \gamma_{K+1} v_K(\ln(t)) \quad (10)$$

where the j th basis function is defined for $j = 1, \dots, K$ as,

$$\begin{aligned} v_j(\ln(t)) &= (\ln(t) - k_j)_+^3 \\ &\quad - \lambda_j (\ln(t) - k_{min})_+^3 \\ &\quad - (1 - \lambda_j) (\ln(t) - k_{max})_+^3, \lambda_j \\ &= \frac{k_{max} - k_j}{k_{max} - k_{min}} \quad \text{and} \quad (\ln(t) - a)_+ \\ &= \max(0, \ln(t) - a). \end{aligned}$$

The curve complexity and degrees of freedom (d.f.) is determined by the number of knots, which equals $K + 1$ ignoring γ_0 . The knot positions are often chosen according to the fixed

percentiles of the uncensored log survival times t , for example, 3 knots would be placed at the 25th, 50th and 75th of $\ln(t)$. The positions are recommended by Durrelman and Simon (1989) who suggested choosing knots close to the median uncensored log survival time to allow the data to be most closely modelled in the region of greatest density and hence usually of lowest variance. Models with d.f. > 4 were not recommended since the resulting curves are expected to be potentially unstable (Royston and Parmar, 2002). In this paper, we only considered restricted cubic splines with 1, 2 or 3 interior knots, corresponding to 2, 3 or 4 d.f. respectively.

4.2.3 Selecting a model for the baseline log cumulative hazard—The best-fitting model among FP functions and RCS models with different degrees of freedom ($47 = 44$ FPs + 3 RCSs potential models totally) can be assessed using the model selection criteria, such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC). For model (8), we selected the final baseline log cumulative hazard model that minimized the AIC since the BIC has been criticized for its tendency to give models that are too parsimonious (Weakliem, 1999).

4.3. Causal quantities on mean survival time

According to definition (1) and identifiability equation (7) from mediation formula, the estimation of the natural direct and indirect effects on mean survival time scale requires the distribution of time-to-event T over the whole range (time 0 to infinity). However, in the survival analysis, we usually have right censoring of event time and do not observe the right tail of survival distribution. Due to this limitation, it may not be applicable to estimate the natural direct and indirect effects on the (unrestricted) mean survival time scale. An alternative way is to consider causal quantities on restricted mean survival time (RMST). The RMST, $\mu(t^*)$ say, of a random variable T is the expected value of $\min(T, t^*)$ which may be evaluated as the area under the survival curve $S(t)$ up to t^* (Royston and Parmar, 2011),

$$\mu(t^*) = E(\min(T, t^*)) = \int_0^{t^*} S(t) dt$$

For our example, when T is time to stroke event (months), we may think of $\mu(t^*)$ as the ‘ t^* -month stroke free life expectancy’. We could decompose the total exposure effects on a RMST scale instead of an unrestricted mean survival time scale as,

$$\begin{aligned} TE_{RMST}(t^*) &= \mu(t^*; 1, M(1)) \\ &\quad - \mu(t^*; 0, M(0)) \\ &= IE_{RMST}(t^*; x) \\ &\quad + DE_{RMST}(t^*; 1 - x) \\ &= [\mu(t^*; x, M(1)) - \mu(t^*; x, M(0))] \\ &\quad + [\mu(t^*; 1, M(1 - x)) - \mu(t^*; 0, M(1 - x))] \quad (11) \end{aligned}$$

The natural direct and indirect effects on RMST can be estimated by applying the identifiability equation (7) to definition (11) with the AIC selected baseline log cumulative hazard model from either FP or RCS family. We chose the time point, t^* , for calculating the restricted mean survival time to be the last observed event time as recommended by Royston and Parmar (2011).

4.4 Mediation effect estimation on the hazard scale

The hazard function is not an unconditional probability, so the law of iterated expectations and the identifiability equation (7) cannot be used directly to estimate the hazard function of $T(x, M(x'))$. To solve this problem, we use,

$$h(t^*; x, M(x')) = \frac{f(t^*; x, M(x'))}{S(t^*; x, M(x'))} \quad (12)$$

where $f(t^*; x, M(x'))$ represents the probability density function for potential outcome $T(x, M(x'))$ at time point t^* (VanderWeele, 2011). With the formula (12), the necessary causal quantities for the hazard scale in (4) can be estimated using equation (7) and the estimated baseline parametric hazard function. In the mediation effects definition on the survival and hazard scales, we also set t^* equal to the last observed event time. Other specific time points with clinical significance may be chosen, but we recommend using a time point before the last observed event time in order to obtain 'accurate' baseline parametric distribution approximation for the mediation effects estimation.

5. SENSITIVITY ANALYSIS

In this section, we describe the method for sensitivity analysis to examine the impact on the mediation effects inference when the no unmeasured mediator-outcome confounding assumption (6) is violated (Albert and Wang, 2015). The quantities $g\{T(1, M(1))\}$ and $g\{T(0, M(0))\}$, but not $g\{T(0, M(1))\}$, are identifiable when unmeasured mediator-outcome confounders exist. In our approach, we consider a hybrid causal-observational model that extends the association Cox model for time-to-event T by incorporating causal as well as observational effects with the specified log cumulative hazard function,

$$\log(H(t; x, m | X=x', M(x')=m, W=w)) = \log(H_0(t)) + \beta_1(\phi x' + (1-\phi)x) + \beta_2 m + \beta' w \quad (13)$$

Here β_1 represents the association effect of X on T , and all β s in (13) are estimable from the corresponding association model if we set $x = x'$. When the sequential ignorability assumption (6) holds, ϕ is equal to 0 indicating that the functions of the potential outcome $T(x, m)$ do not depend on the observed group x' , and therefore the natural direct and indirect effects definition can be identified. Nonzero ϕ represents a departure from sequential

ignorability, and in our hybrid causal-observational model, ϕ can be interpreted as the nonidentifiable proportion of the association effect due to the cohort effect (or selection bias). Therefore, when ϕ is not equal to zero, estimation of the functions of the potential outcome $T(x, m)$ may depend on the observed group x' due to the unmeasured confounders between the mediator and the outcome (see equation (13)). Equivalently, $(1 - \phi)$ is the proportion of the association effect due to the causal effect of the exposure. It is possible for this proportion ϕ to be negative or greater than 1, as the causal effect $\beta_1(1 - \phi)$ and the cohort effect $\beta_1\phi$ need not be in the same direction. The natural direct and indirect effects can be estimated using equation (7) and (13) over varying ϕ to provide a sensitivity analysis.

6. SIMULATIONS

We conduct simulation studies to examine the natural indirect estimators on different scales from our mediation formula approach in a survival context. Event times were generated from eight sets of parametric distributions according to the approach set out by Bender *et al.* (2005). The parametric distributions were specified as: (1) FP with degree 1 and power $p = (0)$ (Weibull distribution); (2) FP with degree 2 and powers $p = (-0.5, -0.5)$; (3) RCS with 1 interior knot; (4) RCS with 3 interior knots; (5) Gompertz distribution with positive shape parameter leading to a hazard function that increases with time; (6) Gompertz distribution with negative shape parameter leading to a hazard function that decreases with time; (7) mixture Weibull distribution generating a bathtub shape hazard function; (8) mixture Weibull distribution generating an inverted bathtub shape hazard function. Plots of the baseline hazard function and the baseline cumulative hazard function from the eight simulated parametric distributions are presented in Appendix B of the online supporting material. These eight sets of parametric distributions include two FP models with degree 1 or 2, two RCS models with 1 or 3 internal knots, and another four common types of hazard function (i.e. increasing, decreasing, bathtub shape and inverted bathtub shape) that cannot be covered by our approximation models. For each specific parametric distribution, we considered two scenarios that are distinguished in the relative magnitude of the natural direct and indirect effects. The parameters β_2 and α_1 in model (8) and (9) and parameter β_1 in model (8) were adjusted to make the total exposure effects comparable through all sixteen simulation scenarios, and the mediation proportion is approximately 10% for the first scenario in each parametric distribution (scenario 1 for parametric distribution 1, scenario 3 for parametric 2 etc.) and 50% for the second scenario in each parametric distribution (scenario 2 for parametric distribution 1, scenario 4 for parametric 2 etc.). The log cumulative hazard function includes a binary exposure indicator X (1 if exposed with 50% frequency, 0, otherwise), a binary covariate W (constrained so that the exposed group has a 50% frequency of $W = 1$ and the non-exposed group has a 30% frequency of $W = 1$), and a normally distributed mediator variable M . For each given exposure, we generated the mediator variables using model (9). We then generated the time-to-event variate $\tilde{T}$ according to the eight parametric distribution listed above under the proportional hazards model (8) given the individual exposure and observed mediators and the censoring time C from the exponential distribution with mean $e^{5.5} = 244.7$. Administrative censoring was also employed to curtail the follow-up to 120 months. The censoring probabilities in the exposed and unexposed group for all 16 simulation scenarios are 30-50% and 10-25% respectively.

For each of above scenarios, 500 simulated datasets were generated with total sample sizes of 200 and 2000 respectively.

The true natural indirect effect on RMST, survival function and hazard function assessed at the last observed failure time for each data set is calculated by applying identifiability equation (7) to natural indirect effect definition (11), (3) and (4) using the true coefficients and simulated true parametric baseline functions. For each generated dataset, the proposed mediation formula approach was utilized to estimate the causal quantities, and 95% confidence intervals were constructed with percentile estimates from 200 bootstrap samples. AIC-selected baseline cumulative hazard functions among FP functions and RCS models with different degrees of freedom were used in the estimation of the relevant causal quantities. We considered the following simulation statistics, the average estimate of $IE(0)$; the average percent error of $IE(0)$ ($PE = 100 \times \{\text{Average Estimated } IE(0) - \text{True } IE(0)\} / \text{true } IE(0)$), a measure of relative bias; the SD of estimated $IE(0)$; the average estimated SE of estimated $IE(0)$; the coverage probability (CP, percent of simulated datasets for which the 95% confidence intervals for estimated causal quantities covered the true value). Similar summary statistics were calculated for mediation proportion estimates on the hazard scale.

The simulation results are given in Table 1 and Table 2 with total sample size 200. Table 1 shows simulation statistics for the estimated natural indirect effects on RMST and survival function and Table 2 shows simulation statistics for the estimated natural indirect effects on hazard function and related mediation proportion. From both tables, we see that the mediation formula approach generally produces a small relative bias of less than or equal to 9% in its estimation of $IE(0)$ on different scales for both scenarios (*Sce* 1 to 8) that can be covered by our 47 approximated parametric distribution, and scenarios (*Sce* 9 to 16) that cannot be covered by our approximated parametric distributions. The hazard function mediation effect estimators (Table 2) are usually less biased than those of the RMST and survival function (Table 1), and scenarios with balanced $IE(0)$ and $DE(1)$ usually have less relative bias than those with unbalanced $IE(0)$ and $DE(1)$ on RMST and survival function scales (scenario 5 vs. 6, 7 vs. 8, 9 vs. 10 etc. in Table 1). However, our approach shows a relative bias of up to 20% for estimators of the mediation proportion on the hazard scale (Table 2). The coverage probabilities of 95% CI from the bootstrap method are within 3.5% of the nominal level for all four estimated causal quantities under different scenarios. When the sample size increases to 1000 per group, the relative bias is less than 1.5% for the $IE(0)$ estimates on different scales and less than 2.5% for the mediation proportion estimates on the hazard scale (results not shown).

7. ANALYSIS OF JACKSON HEART STUDY DATA

To illustrate the new approach, we examined the relation between smoking, blood pressure, and incident stroke using the sample data from the Jackson Heart Study (JHS). The JHS is a prospective community-based cohort study initiated in 2000 to investigate the risk factors and causes of cardiovascular diseases in African Americans (Taylor *et al.*, 2005). The exposure for our example is the binary variable, current smoking status, coded as 1 for current smokers ('exposed') and 0 for non-current smokers ('unexposed') and the outcome considered is the adjudicated incident stroke event from JHS Visit 1 (2000-2004) to

administrative end date 12/31/2010. We also considered, in separate models, the potential mediator, continuous systolic blood pressure (SBP), mm Hg, measured using a random-zero sphygmomanometer. Smoking exposure and blood pressure were assessed at baseline JHS Visit 1. Our goal is to use the proposed methods to quantify the proportion of the smoking effect on stroke that is mediated through SBP.

The subset of JHS dataset we analyzed included 2,000 subjects with complete model variables. The crude incidence rate (per 1000 person-years) of total stroke was 6.27 (95% CI, [3.24, 10.96]) for current smokers ($n = 289$), and 3.85 (95% CI, [2.83, 5.12]) for non-current smokers ($n = 1,711$). The mean measured average SBP (SD) was 127.7 (20.0) for the exposed participants and 126.4 (18.2) for the unexposed participants. We consistently adjusted for age, gender and Body Mass Index (BMI) as the minimum set of confounders in both models (8) and (9). Figure 2 shows that AIC-selected FP with degree 1 and power $p = (0)$ provides a satisfactory fit comparing with the non-parametrical Cox model and was used as the parametric baseline log cumulative hazard function for the causal quantities estimation. Estimates of the natural direct, natural indirect, total effects and mediation proportion (%) for the incident stroke on the RMST, survival function and hazard function scales are provided in Table 3. The 95% asymptotic CIs are constructed using the bootstrap method. When the exposure changes from non-current smoker to current smoker, ‘stroke-free life expectancy (with follow up restricted to 102.28 months)’ decreases 1.06 months with an estimated 0.12 months (95% CI: 0.01, 0.30) attributable to the SBP and an estimated 0.94 months (−0.53, 2.17) because of the direct effect (or other unknown pathways). Likewise, the estimated stroke-free survival probability at the last observed event time (102.28 months) decreases by an estimated 2.05% (97.35% for $S(102.28; 0, M(0))$ vs. 95.30% for $S(102.28; 1, M(1))$), and this can be decomposed into an estimated decrease of 0.23% (0.01%, 0.61%) attributable to the SBP and an estimated decrease of 1.81% (−1.06%, 4.17%) attributable to other pathways. On the hazard function scale, the estimates indicate that a change in smoking status from non-current to current smoker produces a 1.79 times (0.82, 3.28) greater hazard of stroke at 102.28 months in which 1.09 (1.00, 1.22) or 14.8% (−62.2%, 110.0%) of the effect can be attributed to the pathway through SBP.

We also conducted sensitivity analyses using the hybrid observational-causal model approach and the elicitation of plausible sensitivity parameter value ϕ for the present data example is described in Appendix C of the online supporting material. We examined the natural direct and indirect effect estimates for varying proportion, ϕ , of the association effect due to the cohort effect (from −2 to 2 in increments of 0.02). The plot of $IE(0)$ estimates on the RMST and hazard scales over ϕ is shown in Figure 3A and Figure 3B respectively. $IE(0)$ estimates on the RMST scale decrease as ϕ increases and the elicited value $\phi = -0.06$ provides an estimate of −0.08 (95% CI: −0.28, 0.04). The plausible range for ϕ of [−0.3, 0] provides a range for estimates of $IE(0)$ on the RMST scale from −0.12 to 0.08. Figure 3A further shows that the 95% confidence intervals for $IE(0)$ fail to exclude 0 for value of $\phi < 0$; also, for $\phi < -0.18$, the estimate for $IE(0)$ is no longer negative. The $IE(0)$ estimates on the hazard scale increases as ϕ increases and the pattern is similar to that on the RMST scale in which the 95% confidence intervals for $IE(0)$ on the hazard scale fail to exclude 1 for value of $\phi < 0$, and the estimate for $IE(0)$ is less than 1 for $\phi < -0.18$. To conclude, 10-15% of the adverse effects of cigarette smoking on incident stroke in African Americans could be

mitigated by targeting its mediator, blood pressure; however the significance of the natural indirect effect may not hold if unmeasured confounders (inducing a cohort effect less than $-0.18\beta_1$) between the mediator and outcome exist.

8. DISCUSSION

In this paper, we provide an approach to causal mediation analysis using the mediation formula to estimate the mediation effects on different scales for survival data under the proportional hazards assumption. We proposed using parametric models, FP or RCS, to approximate the baseline hazard functions for the estimation of natural direct and indirect effects on the RMST, survival function and hazard function scales. Sensitivity analysis is performed to assess the effect of violation of no unmeasured mediator-outcome confounding assumption on mediation effects estimation using a hybrid causal-observational model approach. Comparing with VanderWeele's method (2011), our proposed approach is much more flexible by allowing mediation effects estimation on different scales for the Cox Proportional Hazards Model and does not rely on the rare outcome assumption. The method described has been implemented in a SAS Macro, which is available for downloading from the webpage, <https://github.com/souwwang/Cox-Survival-Mediation.git>.

Due to the right censoring of event times, we cannot observe the crucial upper tail of the survival distribution; thus our parametric approximation model (FP or RCS) can only capture the observed time-to-event T distribution through the last observed event time which makes the expected potential outcome estimation of T using the mediation formula approach inapplicable. For example, we tried to estimate the natural indirect and direct effects on the mean survival time scale defined as equation (1). The estimated natural indirect effects differed significantly ($-640.9 [-5982.5, -21.2]$ vs. $-817.7 [-14355.3, -8.4]$) using two parametric baseline log cumulative hazard functions (FP with degree 1 and power $p = (0)$ vs. FP with degree 2 and power $p = (-2, 0)$) with similar AIC (1010.4 vs. 1010.9) for our data example. In contrast, the estimated natural indirect effects on RMST, survival function, hazard function at the last observed event time are nearly the same for these two models. The result is not surprising considering the extrapolated survival distribution beyond the last observed event time used in both models for expected mean survival time calculation when estimating the mediation effects. To avoid this problem, we used the RMST, which has a straightforward interpretation, as the causal estimand. For the defined causal mediation effects on each of the RMST, survival function and hazard function scales, we need to choose a time point t^* at which each function is evaluated. In this paper, we used t^* (fixed for all the estimands) as the last observed event time, though an alternative, clinically motivated, t^* occurring before the last observed event time may be chosen. One can further plot the estimated value of the average causal mediation effects over the whole range from 0 to the last observed event time in order to investigate how the mediation effects change as a function of the chosen t^* .

In this paper, we used the AIC to select the final approximate parametric distribution used for mediation effects estimation. Alternatively, we may choose the BIC, which tends to give more parsimonious models (Rutherford *and others*, 2015). In our simulation studies, we designed eight sets of scenarios to cover a wide range of complex hazard shapes in

biomedical research. Although AIC-selected parametric model may deviate from true models especially when the sample size is small (data not shown), our simulation results showed that the baseline hazard function can be approximated to a sufficient degree of accuracy for the mediation effects estimation. More caution is needed for the inference regarding the mediation proportion, which tends to be unstable and is thus biased though the bias decreases as the sample size increases. This observation is consistent with previous work (Albert and Nelson, 2011).

The proposed sensitivity analysis in this paper mainly assesses the impact on the mediation effects inference when the no unmeasured mediator-outcome confounding assumption (6) is violated and we assumed the assumption (5) is met in the sensitivity analysis. The assumption (5) is easily satisfied in a randomized clinical trial, and in our observational JHS data example, we consistently adjusted for age, gender and BMI as the minimum set of confounders in the mediator model and the outcome model. However, it is possible that additional covariates other than age, gender and BMI could possibly confound the exposure-mediator or exposure-outcome relationship and invalidate the assumption (5).

In conclusion, we have proposed a mediation formula approach for causal mediation analysis of a survival outcome. This approach, which assumes a proportional hazards model and uses simple parametric models to approximate the baseline log cumulative hazard function, allows estimation of mediation effects defined on the RMST, survival probability, and hazard scales. We used a normally distributed continuous mediator in this paper, and the extension of our approach to binary or ordinal mediators using appropriate parametric models is straightforward. In addition, the causal quantities definition and estimation methods for survival data we proposed in this paper can also be extended to other semi-parametric survival models as the accelerated failure time models or the additive hazard models. Further work is needed to study multiple mediators or multiple stage mediators in the survival context.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Tyler J. VanderWeele for helpful discussion and Cuiyu Geng for assistance in conducting the simulation studies. Support for this research was provided in part by the National Heart, Lung, and Blood Institute and the National Institute on Minority Health and Health Disparities, with additional support from the National Institute on Biomedical Imaging and Bioengineering [the Jackson Heart Study contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C], and the National Institute of Dental and Craniofacial Research, National Institutes of Health [grant number R01DE022674 (J. Albert)].

REFERENCES

- Albert JM. Mediation analysis via potential outcomes models. *Stat. Med.* 2008; 27:1282–1304. [PubMed: 17691077]
- Albert JM, Nelson S. Generalized causal mediation analysis. *Biometrics.* 2011; 67:1028–38. [PubMed: 21306353]

- Albert JM, Wang W. Sensitivity analyses for parametric causal mediation effect estimation. *Biostatistics*. 2015; 16:339–351. [PubMed: 25395683]
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 1986; 51:1173–1182. [PubMed: 3806354]
- Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Stat. Med.* 2005; 24:1713–1723. [PubMed: 15724232]
- Durrelman S, Simon R. Flexible regression models with cubic splines. *Stat. Med.* 1989; 8:551–561. [PubMed: 2657958]
- Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol. Methods*. 2010; 15:309–334. [PubMed: 20954780]
- Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.*, 2010. 2010; 25:51–71.
- Jung SY, Rosenzweig M, Linkov F, Brufsky A, Weissfeld JL, Sereika SM. Comorbidity as a mediator of survival disparity between younger and older women diagnosed with metastatic breast cancer. *Hypertension*. 2012; 59:205–211. [PubMed: 22184319]
- Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011; 22:575–581. [PubMed: 21552129]
- MacKinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Evaluation Rev.* 1993; 17:144–158.
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods*. 2002; 7:83–104. [PubMed: 11928892]
- Pearl J. The causal mediation formula--a guide to the assessment of pathways and mechanisms. *Prev. Sci.* 2012; 13:426–436. [PubMed: 22419385]
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992; 3:143–155. [PubMed: 1576220]
- Royston P. Technical Report. University College; London: 2011. Estimating a smooth baseline hazard function for the Cox model..
- Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 1994; 43:429–467.
- Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat. Med.* 2002; 21:2175–2197. [PubMed: 12210632]
- Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat. Med.* 2011; 30:2409–2421. [PubMed: 21611958]
- Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *J. Statist. Comput. Simulation*. 2015; 85:777–793.
- Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors using fractional polynomials. *J. R. Stat. Soc. Ser. A. Stat. Soc.* 1999; 162:71–94.
- Taylor HA Jr, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, Nelson C, Wyatt SB. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn., Dis.* 2005; 15:S6–S17.
- Tchetgen Tchetgen EJ. On causal mediation analysis with a survival outcome. *Int. J. Biostat.* 2011; 7 Article 33.
- Tchetgen Tchetgen EJ. Inverse odds ratio-weighted estimation for causal mediation analysis. *Stat. Med.* 2013; 32:4567–4580. [PubMed: 23744517]
- VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2011; 22:582–585. [PubMed: 21642779]
- VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat. Interface*. 2009; 2:457–468.

- Wang W, Albert JM. Estimation of mediation effects for zero-inflated regression models. *Stat. Med.* 2012; 31:3118–3132. [PubMed: 22714572]
- Weakliem DL. A critique of the Bayesian information criterion for model selection. *Sociol. Method. Res.* 1999; 27:359–397.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Path diagram for a survival mediation model in the two-stage framework with exposure X , mediator M , time-to-event outcome T and covariates W .

Figure 2.

Jackson Heart Study stroke incidence data. Comparison of baseline (for a subject with systolic Blood Pressure 126.55 mmHg, BMI 30 kg/m², Age > 55.0 years old, Male, and non-Smoking) stroke-free survival probability (Panel A) and cumulative hazard (Panel B) for the AIC-selected FP model (dotted line), and the non-parametrical Cox model (solid line).

Figure 3.

Sensitivity analysis for the mediation effects of continuous mediator systolic blood pressure on the relation between smoking and incident stroke in the Jackson Heart Study. Panel A and B show estimated $IE(0)$ on the RMST and hazard scales at last observed event time (102.28 months). The solid line represents the estimated natural direct effects, and the gray areas represent the 95% CIs from the bootstrap method at each value of sensitivity parameter ϕ .

Table 1

Simulation statistics for the estimated natural indirect effects on RMST and survival function for survival data under the proportional hazards assumption, $n = 100$ per group.

Scenario	RMST [†]							Survival Function (%) [†]						
	True DE(1)	True IE(0)	Ave Est IE(0)	Ave PE (%)	SD of Est IE	AVE SE	CP (%)	True DE(1)	True IE(0)	Ave Est IE(0)	Ave PE (%)	SD of Est IE	AVE SE	CP (%)
1	-22.27	-1.75	-1.76	0.9	1.20	1.21	95.8	-25.44	-2.13	-2.14	0.5	1.42	1.47	96.2
2	-12.89	-10.15	-10.07	-0.5	3.17	3.05	93.2	-13.67	-11.55	-11.39	-1.4	3.31	3.27	93.0
3	-17.63	-1.34	-1.31	-2.6	1.02	0.97	92.2	-21.29	-1.67	-1.62	-2.6	1.18	1.25	93.4
4	-12.41	-9.99	-9.92	-0.3	3.16	3.04	94.0	-13.59	-11.41	-11.31	-0.9	3.32	3.30	94.8
5	-17.93	-1.37	-1.48	9.0	1.10	1.03	94.6	-21.45	-1.68	-1.83	8.7	1.32	1.31	95.4
6	-9.55	-7.06	-7.11	0.6	2.84	2.58	96.4	-11.41	-8.71	-8.76	0.6	2.91	3.00	95.8
7	-12.83	-0.94	-0.99	4.8	0.74	0.78	94.4	-22.16	-1.75	-1.82	3.8	1.32	1.40	95.0
8	-7.78	-5.60	-5.57	-0.4	2.03	2.07	93.4	-12.21	-9.58	-9.45	-1.4	3.07	3.18	93.6
9	-8.18	-0.62	-0.67	7.2	0.48	0.48	92.6	-28.27	-2.53	-2.63	3.9	1.80	1.78	93.8
10	-5.40	-4.06	-4.06	-0.1	1.30	1.35	96.4	-14.33	-13.22	-13.04	-1.4	3.48	3.55	95.6
11	-20.87	-1.65	-1.72	4.5	1.18	1.12	92.6	-25.35	-2.11	-2.19	3.6	1.46	1.46	92.2
12	-10.88	-8.32	-8.39	0.8	2.95	2.71	94.0	-12.92	-10.45	-10.48	0.2	3.22	3.19	94.6
13	-16.97	-1.27	-1.35	6.5	0.96	0.98	93.8	-24.1	-1.97	-2.08	5.9	1.44	1.47	93.8
14	-9.49	-6.96	-6.98	0.4	2.42	2.43	94.4	-12.7	-10.17	-10.12	-0.5	3.14	3.22	95.0
15	-13.69	-1.01	-1.05	3.4	0.81	0.81	92.0	-20.83	-1.62	-1.69	4.1	1.27	1.29	93.4
16	-10.42	-7.99	-7.8	-2.3	2.67	2.49	91.8	-13.57	-11.38	-11.00	-3.3	3.41	3.26	93.2

[†] assessed at last observed event time.

Table 2

Simulation statistics for the estimated natural indirect effects and mediation proportion on hazard scale for survival data under the proportional hazards assumption, $n = 100$ per group.

Scenario	Hazard Function [†]							Mediation Proportion on the Hazard Scale (%) [†]						
	True DE(1)	True IE(0)	Ave Est IE(0)	Ave PE (%)	SD of Est IE	AVE SE	CP (%)	True IE(0) Prop	Ave Est IE(0) Prop	Ave PE (%)	SD of Est IE Prop	AVE SE	CP (%)	
1	2.36	1.10	1.10	0.2	0.07	0.07	95.6	10.16	10.83	6.6	7.77	41.64	96.2	
2	1.45	1.49	1.49	-0.2	0.15	0.16	93.8	51.98	55.77	7.3	23.50	167.84	95.8	
3	2.39	1.10	1.10	-0.1	0.07	0.08	92.0	10.12	10.69	5.7	8.48	71.22	94.8	
4	1.45	1.50	1.51	0.8	0.16	0.16	94.0	51.96	57.44	10.5	48.28	126.97	95.2	
5	2.39	1.10	1.11	0.8	0.08	0.08	95.4	10.12	11.54	14.0	8.8	53.01	96.4	
6	1.50	1.54	1.56	1.2	0.18	0.20	96.4	51.58	61.88	20.0	79.12	610.84	97.4	
7	2.38	1.10	1.11	0.6	0.08	0.08	95.4	10.13	12.03	18.8	17.43	359.82	95.2	
8	1.48	1.52	1.54	0.7	0.18	0.19	95.2	51.71	58.17	12.5	38.62	358.83	96.0	
9	2.34	1.10	1.11	0.5	0.07	0.07	93.4	10.20	11.51	12.8	8.66	35.4	94.2	
10	1.41	1.46	1.47	0.3	0.14	0.15	94.4	52.24	54.08	3.5	38.74	322.86	96.6	
11	2.36	1.10	1.11	0.4	0.07	0.07	93.2	10.16	11.31	11.3	8.56	36.92	93.6	
12	1.47	1.51	1.52	0.7	0.16	0.17	94.6	51.83	56.21	8.5	23.22	371.91	95.6	
13	2.37	1.10	1.11	0.8	0.08	0.08	94.2	10.15	11.53	13.7	8.47	94.45	95.4	
14	1.47	1.52	1.53	0.7	0.17	0.18	95.2	51.79	56.57	9.2	29.78	295.14	95.8	
15	2.39	1.10	1.11	0.7	0.08	0.08	94.2	10.11	11.43	13.0	9.84	276.22	95.0	
16	1.45	1.50	1.49	-0.5	0.16	0.16	94.0	51.96	54.58	5.0	20.65	180.65	95.0	

[†] assessed at last observed event time.

Table 3

Estimated causal quantities and 95% CIs from bootstrap method of blood pressure on the relation between smoking and incident stroke in the Jackson Heart Study based on the mediation formula approach.

Causal Quantities	Mediation Formula Approach		
	RMST (Months) [†]	Survival Function [†]	Hazard Function [†]
Natural Indirect Effect $IE(0)$	-0.12 (-0.30, -0.01)	-0.23 (-0.61, -0.01)	1.09 (1.00, 1.22)
Natural Direct Effect $DE(1)$	-0.94 (-2.17, 0.53)	-1.81 (-4.17, 1.06)	1.64 (0.71, 3.03)
Mediation Proportion (%) $P(0)$	11.4 (-68.1, 110.4)	11.5 (-68.0, 110.4)	14.8 (-62.2, 110.0)
Total Exposure Effect TE	-1.06 (-2.30, 0.29)	-2.05 (-4.40, 0.56)	1.79 (0.82, 3.28)

[†] assessed at last observed event time (102.28 months).