Published in final edited form as:

Genet Epidemiol. 2018 June ; 42(4): 394–404. doi:10.1002/gepi.22120.

A multiple mediator analysis approach to quantify the effects of the ADH1B and ALDH2 genes on hepatocellular carcinoma risk

Stephannie Shih¹, Yen-Tsung Huang^{1,2,3}, and Hwai-I Yang⁴

¹Department of Epidemiology, Brown University, Providence, RI 02912, United States

²Department of Biostatistics, Brown University, Providence, RI 02912, United States

³Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

⁴Genomics Research Center, Academia Sinica, Taipei 11529, Taiwan

Abstract

Previous work suggested a genetic component affecting the risk of Hepatocellular Carcinoma (HCC) and mediation analyses have elucidated potential indirect pathways of these genetic effects. Specifically, the effects of alcohol dehydrogenase (ADH1B) and aldehyde dehydrogenase (ALDH2) genes on HCC risk vary based on alcohol consumption habits. However, alcohol consumption may not be the only mediator in the identified pathway: factors related to alcohol consumption may contribute to the same indirect pathway. Thus, we developed a multi-mediator model to quantify the genetic effects on HCC risk through sequential dichotomous mediators under the counterfactual framework. Our method provided a closed form formula for the mediation effects through different indirect paths which requires no assumption for the rarity of outcome. In simulation studies of a finite sample, we presented the utility of the method with the variance of the effects estimated using the delta method and bootstrapping. We applied our method to data from participants in Taiwan (580 cases and 3,207 controls) and quantified the mediation effects of Single Nucleotide Polymorphisms (SNPs) in the ADH1B and ALDH2 genes on HCC through alcohol consumption (yes/no) and high alanine transaminase (ALT) levels (greater than or equal to 45 U/L or below 45 U/L). Assuming a dominant risk model, we identified that the SNPs' effects through alcohol consumption is more significant than through ALT levels on HCC risk. This new method provides insight to the magnitude of various casual mechanisms as a closed form solution and can be readily applied in other genomic studies.

Keywords

alanine transaminase; alcohol consumption; dichotomous mediators; dichotomous outcome

Corresponding Author: Yen-Tsung Huang, Institute of Statistical Science, Academia Sinica, 128 Academia Road, Taipei 11529, Taiwan; phone: +886-2-2783-5611 ext 315; ythuang@stat.sinica.edu.tw.

SUPPORTING INFORMATION

Supporting information can be found online for this article.

1. INTRODUCTION

Hepatocellular carcinoma (HCC) is a complex and common cancer worldwide with a high mortality rate (Ghouri, Mian, & Rowe, 2017; Zhu, Seto, Lai, & Yuen, 2016). Exposure to the hepatitis B virus (HBV) is a risk factor for HCC and in some countries such as Taiwan, exposure to HBV may have occurred early in life (perinatally or during childhood)(Chen & Yang, 2011). Another risk factor of HCC is alcohol consumption habits which may increase liver inflammation and cirrhosis risk (Seitz & Stickel, 2006; Zhu et al., 2016). Based on these associations, a previous study assessed genes encoding enzymes that help metabolize alcohol and may affect the risk of alcohol dependence. The study identified that the single nucleotides polymorphisms (SNPs) rs1229984 and rs671 for the alcohol dehydrogenase (ADH1B) gene and aldehyde dehydrogenase (ALDH2) gene respectively, may decrease HCC risk mediated by alcohol consumption (Edenberg et al., 2006; Eng, Luczak, & Wall, 2007; Liu et al., 2016). Incorporating biomarker information may further help understand the development of HCC. In particular, alanine transaminase (ALT) levels (reflective of liver damage) may be useful in understanding an individual's susceptibility to HCC and may be a better reflection of long term exposure to alcohol (Alatalo et al., 2008; Conigrave, Davies, Haber, & Whitfield, 2003; Kim, Flamm, Di Bisceglie, & Bodenheimer, 2008). Moreover, carrying deficient metabolic genes may increase the burden of detoxification by the liver and thus induce higher ALT levels through liver damage. The long-term liver damage due to the ineffective detoxification by ADH1B and ALDH2 enzymes and subsequent chronic inflammation may lead to the development of HCC. Thus, we will like to build upon the previous mediation analyses by identifying the direct effects of the ADH1B and ALDH2 genes on HCC risk and the indirect effects through alcohol consumption and ALT levels. Results can help inform how these intermediary factors may work together so interventions can be better developed to address this increased risk.

Traditional difference-based mediation methods quantify single mediator effects using two regression models: one estimating the outcome based on the exposure and another estimating the outcome based on both the exposure and mediator (Baron & Kenny, 1986). These methods can be further adapted to incorporate the counterfactual framework to address exposure-mediator interactions and other nonlinear effects that may bias estimates (Robins & Greenland, 1992). Additional work has been done to apply these methods to quantify the mediation effects for dichotomous outcomes and survival analyses (Imai, Keele, & Tingley, 2010; Lange & Hansen, 2011; Valeri & Vanderweele, 2013; VanderWeele, 2011; Vanderweele & Vansteelandt, 2010), but the methods estimating the mediation effects of a dichotomous mediator requires the rare outcome assumption, which may not be reasonable when studying increasingly prevalent outcomes (Valeri & Vanderweele, 2013). These methods have also been extended to incorporate multiple mediators that do not affect each other, but if it is known that one of these mediators affects the other, then analyzing the mediators sequentially may better quantify the shared effects of the mediators (VanderWeele & Vansteelandt, 2014; W. Wang, Nelson, & Albert, 2013). Moreover, the mechanism that hepatic inflammation measured by ALT as a consequence of metabolic genes affects the behavior of alcohol consumption and HCC risk violates the no recanting witness assumption (Valeri and Vanderweele 2013) and makes the mediation effect in the single-mediator model

non-identifiable. To address the issue, we resort to quantifying the path-specific effects to understand how ALT and alcohol consumption sequentially mediate the genetic effect on HCC risk.

In order to assess the effect of multiple sequential mediators of interest, the concept of path-specific effects can be applied to address potential confounders of the mediator-outcome relationship (Huang & Pan, 2016; VanderWeele & Vansteelandt, 2014; Vanderweele, Vansteelandt, & Robins, 2014). However, having two sequential mediators still introduces the so called “recanting witness assumption”, where the mediator’s effect may vary dependent on the specific pathway leading to the outcome of interest (Avin, Shpitser, & Pearl, 2005; Vanderweele et al., 2014). For example, under a two-mediator model, only three specific effects can be quantified without introducing additional sensitivity parameters to account for the relationships between the mediators (Daniel, De Stavola, Cousens, & Vansteelandt, 2015). Despite the availability of mediation analyses in the context of continuous outcomes and survival analyses (Daniel et al., 2015; Huang & Cai, 2016; Huang & Yang, 2017; Vanderweele et al., 2014), methodology decomposing these mediation effects into three path-specific effects have not been developed for the setting in our motivating example where mediators and the outcome are all dichotomous.

We would like to develop multi-mediator methods to quantify the path-specific effects mediated by a sequence of dichotomous mediators on a dichotomous outcome of interest. Moreover, we derive a computationally efficient closed form formula under both the risk difference and risk ratio scale. This is different strategy from available multi-mediator methods that rely on various weighting schemes that may not be able to easily incorporate exposure-mediator interactions in this context (VanderWeele & Vansteelandt, 2014) or Monte-Carlo simulations to estimate the combined distribution of the outcome and mediators (Daniel et al., 2015; Preacher & Hayes, 2008; VanderWeele & Vansteelandt, 2014; Vanderweele et al., 2014; Vansteelandt & Daniel, 2017). Wang et al (2012) proposed a structural estimating equation-based approach to investigate all possible paths, which is subject to the aforementioned non-identifiability issue with the two mediators (J. Wang et al., 2012). Herein we describe a new method incorporating the counterfactual framework that will allow for the identification of the three path-specific effects of the intermediary factors contributing to HCC risk: 1) the direct effect of these genetic SNPs, 2) the mediation effect through ALT levels but not explained by alcohol consumption, and 3) the mediation effect through alcohol consumption and possibly through ALT.

The rest of this paper is organized as follows. First, we discuss the definition of path-specific effects and the required identifiability assumption (Section 2.1), and present a closed form estimator to quantify the mediation effects of multiple dichotomous mediators on a dichotomous outcome (Section 2.2). Second, we present the simulation results implementing this model (Section 3.1). Lastly, we quantify the effects of genetic factors on HCC risk through an application of our method (Section 3.2) and conclude with discussion (Section 4). Although our specific example only focuses on a two-mediator model, the method can be readily adapted to incorporate additional mediators.

2. METHODS

2.1 Definitions and assumptions

Path-specific effects of interest are identified by deriving the probability of a outcome under the counterfactual framework (Rubin, 1974). This is represented by $Y(q, w, h)$ which is the counterfactual outcome of Y that would be observed were Q (SNPs), W (indicator of alcohol consumption) and H (ALT levels) set to q , w and h , respectively (Rubin, 1974). $H(q, w)$ is the potential outcome for H (ALT levels) were Q set to q , and W to w . $W(q)$ is the potential outcome for alcohol consumption where Q was set to q . The confounders are represented by the vector, X . The path-specific effects depicted in Figure 1 could be identified only under the assumptions that there was no unmeasured confounding for the following three relations. (Huang & Cai, 2016; Pearl, 2001; Vanderweele & Vansteelandt, 2010) The notation $A \perp B | C$ is used to describe that A is independent of B conditional on C .

1. There is no unmeasured confounding for the exposure or mediators and the outcome, represented by the following three scenarios:

Independence of W and $Y(q, w, h)$ conditional on Q and X , denoted as $W \perp Y(q, w, h) | Q, X$

Independence of H and $Y(q, w, h)$ conditional on Q , W and X , denoted as $H \perp Y(q, w, h) | Q, W, X$

Independence of Q and $Y(q, w, h)$ conditional on X , denoted as $Q \perp Y(q, w, h) | X$

2. There is no unmeasured confounding for the exposure and mediators:

Independence of Q and $H(q, w)$ conditional on X , denoted as $Q \perp H(q, w) | X$

Independence of W and $H(q, w)$ conditional on Q and X , denoted as $W \perp H(q, w) | Q, X$

Independence of Q and $W(q)$ conditional on X , denoted as $Q \perp W(q) | X$

3. There is no recanting witness assumptions:

No Q -affected confounder for W to Y or H to Y relationship conditional on X , denoted as $(W(q_c), H(q_b, w)) \perp Y(q_a, w, h) | X$ and

No confounder affected by Q for the W to H relationship conditional on X , denoted as $H(q_b, w) \perp W(q_c) | X$.

where q_a , q_b and q_c may be different intervention values.

With these assumptions, the counterfactual probability of having T2D as outcome of interest herein can be re-expressed as a probability of Y conditional on Q , W and H by summing over all possible combinations of H and W :

$$\begin{aligned}\Omega(P(Y = 1|q_a, q_b, q_c, X) = \\ P(Y(q_a, W(q_c), H(q_b, W(q_c))) = 1|X) = \\ \iint P(Y = 1|q_a, w, h, X) dF_{H|W}(h|q_b, w, X) dF_W(w|q_c, X)\end{aligned}\quad (1)$$

2.2 Multi-mediator model

We propose a series of logistic regression models to model the joint distribution of multiple dichotomous mediators and a dichotomous outcome.

For example, under a two-mediator model, the logistic regression equations are written as follows:

$$\text{logit}[P(Y = 1|Q, W, H, X)] = \beta_0 + \beta_q Q + \beta_w W + \beta_h H + \beta_x^T X \quad (2)$$

$$\text{logit}[P(H = 1|Q, W, X)] = \alpha_0 + \alpha_q Q + \alpha_w W + \alpha_x^T X \quad (3)$$

$$\text{logit}[P(W = 1|Q, X)] = \delta_0 + \delta_q Q + \delta_x^T X \quad (4)$$

To derive the probability model, we first let

$$\begin{aligned}\tau(w, h, q_a, q_b, q_c, x) = & \left(\frac{\exp(\beta_0 + \beta_q q_a + \beta_w w + \beta_h h + \beta_x^T x)}{1 + \exp(\beta_0 + \beta_q q_a + \beta_w w + \beta_h h + \beta_x^T x)} \right) \left(\frac{\exp(\alpha_0 + \alpha_q q_b + \alpha_w w + \alpha_x^T x) h}{1 + \exp(\alpha_0 + \alpha_q q_b + \alpha_w w + \alpha_x^T x)} \right) \\ & \left(\frac{\exp(\delta_0 + \delta_q q_c + \delta_x^T x) w}{1 + \exp(\delta_0 + \delta_q q_c + \delta_x^T x)} \right)\end{aligned}$$

As such, the marginal probability of the outcome of interest is the sum of the probabilities for all four possible combinations of values of H and W:

$$\begin{aligned}& \iint P(Y = 1|q_a, w, h, x) dF_{H|W}(h|q_b, w, x) dF_W(w|q_c, x) \\ &= \sum_{H=0}^1 \sum_{W=0}^1 \tau(w, h, q_a, q_b, q_c, x) \equiv \gamma(P(Y = 1|q_a, q_b, q_c, x))\end{aligned}$$

The path-specific effects (Path 1: $\lambda_{Q \rightarrow Y}$, Path 2: $\lambda_{Q \rightarrow H \rightarrow Y}$, Path 3: $\lambda_{Q \rightarrow W \rightarrow Y}$) under the risk difference scale are:

$$\lambda_{Q \rightarrow Y} = P(Y(q_1, W(q_0), H(q_0, W(q_0))) = 1 | x) - P(Y(q_0, W(q_0), H(q_0, W(q_0))) = 1 | x) = \gamma(P(Y = 1 | 1, 0, 0, x)) - \gamma(P(Y = 1 | 0, 0, 0, x))$$

$$\begin{aligned} \lambda_{Q \rightarrow H \rightarrow Y} &= P(Y(q_1, W(q_0), H(q_1, W(q_0))) = 1 | x) - P(Y(q_1, W(q_0), H(q_0, W(q_0))) = 1 | x) \\ &= \gamma(P(Y = 1 | 1, 1, 0, x)) - \gamma(P(Y = 1 | 1, 0, 0, x)) \end{aligned}$$

$$\begin{aligned} \lambda_{Q \rightarrow WY} &= P(Y(q_1, W(q_1), H(q_1, W(q_1))) = 1 | x) - P(Y(q_1, W(q_0), H(q_1, W(q_0))) = 1 | x) \\ &= \gamma(P(Y = 1 | 1, 1, 1, x)) - \gamma(P(Y = 1 | 1, 1, 0, x)) \end{aligned}$$

Based on Taylor series expansion, the delta method will be used to approximate asymptotic variances for the above estimators (Hosmer, Lemeshow, & May, 2008). Let $\hat{\theta} = (\hat{\beta}^T, \hat{\alpha}^T, \hat{\delta}^T)^T$

where $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_q, \hat{\beta}_w, \hat{\beta}_h, \hat{\beta}_x)^T$ is a vector of regression parameters in model (2),

$\hat{\alpha} = (\hat{\alpha}_0, \hat{\alpha}_q, \hat{\alpha}_w, \hat{\alpha}_x)^T$ is a vector of regression parameters in model (3), and $\hat{\delta} = (\hat{\delta}_0, \hat{\delta}_q, \hat{\delta}_x)^T$ is

a vector of regression parameters in model (4). We let $\lambda = (\lambda_{Q \rightarrow Y}, \lambda_{Q \rightarrow H \rightarrow Y}, \lambda_{Q \rightarrow WY})^T$ and note that λ is a function of θ , i.e., $\lambda \equiv \lambda(\theta)$. We plug in $\hat{\theta}$ in order to estimate λ and denote the plug-in estimator as $\hat{\lambda} \equiv \lambda(\hat{\theta})$. The covariance of $\hat{\lambda}$ would be approximated by the delta method as follows:

$$\widehat{Cov}(\hat{\lambda}) = \frac{d\lambda}{d\theta} \Big|_{\hat{\theta}} \widehat{Cov}(\hat{\theta}) \frac{d\lambda}{d\theta^T} \Big|_{\hat{\theta}},$$

where

$$\widehat{Cov}(\hat{\theta}) = \begin{bmatrix} \widehat{Cov}(\hat{\beta}) & 0 & 0 \\ 0 & \widehat{Cov}(\hat{\alpha}) & 0 \\ 0 & 0 & \widehat{Cov}(\hat{\delta}) \end{bmatrix}.$$

Calculations for each of the partial derivatives are shown in Supporting Information S1 with the coefficients substituted with the respective estimated values from the fitted outcome and mediator logistic regressions. The 95% confidence intervals of $\hat{\lambda}_{Q \rightarrow Y}$, for example, is approximated by: $\hat{\lambda}_{Q \rightarrow Y} \pm 1.96 * \sqrt{\widehat{Var}(\hat{\lambda}_{Q \rightarrow Y})}$, and similar confidence intervals can be calculated for $\hat{\lambda}_{Q \rightarrow H \rightarrow Y}$ and $\hat{\lambda}_{Q \rightarrow WY}$. Hypothesis tests for $\hat{\lambda}_{Q \rightarrow Y}$ are performed by comparing the test statistic, $\frac{|\hat{\lambda}_{Q \rightarrow Y}|}{\sqrt{\widehat{Var}(\hat{\lambda}_{Q \rightarrow Y})}}$, with the standard normal distribution to obtain the p-value by doubling the tail probability, and similarly for $\hat{\lambda}_{Q \rightarrow H \rightarrow Y}$ and $\hat{\lambda}_{Q \rightarrow WY}$.

In addition to the above analytic approach, one may also make statistical inference by bootstrap. Two sided p-values for each path-specific effect for the bootstrap method are calculated by doubling either the proportion of re-sampled (with replacement) path-specific effects from bootstrapping that were above 0 (the null hypothesis) or the proportion below 0, whichever provides the smaller value. Bootstrap 95% confidence intervals are calculated by choosing the 2.5 and 97.5 percentiles as the lower and upper limits for the intervals.

The path-specific effects under the risk ratio scale (RR) are:

$$\rho_{Q \rightarrow Y} = P(Y(q_1, W(q_0), H(q_0, W(q_0))) = 1 | x) / P(Y(q_0, W(q_0), H(q_0, W(q_0))) = 1 | x) = \gamma(P(Y = 1 | 1, 0, 0, x)) / \gamma(P(Y = 1 | 0, 0, 0, x))$$

$$\rho_{Q \rightarrow H \rightarrow Y} = P(Y(q_1, W(q_0), H(q_1, W(q_0))) = 1 | x) / P(Y(q_1, W(q_0), H(q_0, W(q_0))) = 1 | x) = \gamma(P(Y = 1 | 1, 1, 0, x)) / \gamma(P(Y = 1 | 1, 0, 0, x))$$

$$\rho_{Q \rightarrow WY} = P(Y(q_1, W(q_1), H(q_1, W(q_1))) = 1 | x) / P(Y(q_1, W(q_0), H(q_1, W(q_0))) = 1 | x) = \gamma(P(Y = 1 | 1, 1, 1, x)) / \gamma(P(Y = 1 | 1, 1, 0, x))$$

We let $\rho = (\rho_{Q \rightarrow Y}, \rho_{Q \rightarrow H \rightarrow Y}, \rho_{Q \rightarrow WY})^T$. Using the delta method, the variance of $\log \hat{\rho}$ is calculated by:

$$\widehat{Cov} \left(\log \left(\hat{\rho} \right) \right) = \frac{d \log \rho}{d \theta} \Big|_{\hat{\theta}} \widehat{Cov} \left(\hat{\theta} \right) \frac{d \log \rho}{d \theta^T} \Big|_{\hat{\theta}}.$$

For $\hat{\rho}_{Q \rightarrow Y}$, its 95% confidence interval is calculated by the expression

$$\exp \left(\log \left(\hat{\rho}_{Q \rightarrow Y} \right) \pm 1.96 * \sqrt{\widehat{Var} \left(\log \left(\hat{\rho}_{Q \rightarrow Y} \right) \right)} \right)$$

and the test statistic can be constructed as $\frac{\log \hat{\rho}_{Q \rightarrow Y}}{\sqrt{\widehat{Var}(\log \hat{\rho}_{Q \rightarrow Y})}}$; similarly for $\hat{\rho}_{Q \rightarrow H \rightarrow Y}$ and

$\hat{\rho}_{Q \rightarrow WY}$. P-values for each path-specific effect estimate are calculated from doubling the tail probability of the test statistic in the standard normal distribution. Again, statistical inference can also be made by bootstrap.

If the three path-specific effects quantified all follow the same direction, the proportion of the mediation effect can be estimated by calculating the proportion of each effect over the total effect which is equivalent to

$$P(Y(q_1, W(q_1), H(q_1, W(q_1))) = 1 | x) - P(Y(q_0, W(q_0), H(q_0, W(q_0))) = 1 | x) = \lambda_{Q \rightarrow Y} + \lambda_{Q \rightarrow H \rightarrow Y} + \lambda_{Q \rightarrow WY}$$

under the RD scale and

$$\log\{P(Y(q_1, W(q_1), H(q_1, W(q_1))) = 1|x)/P(Y(q_0, W(q_0), H(q_0, W(q_0))) = 1|x)\} \text{ under the log(RR)} \\ = \log(\rho_{Q \rightarrow Y} * \rho_{Q \rightarrow H \rightarrow Y} * \rho_{Q \rightarrow WY})$$

scale. As an example, the proportion of the mediated effect of path 2 $\lambda_{Q \rightarrow H \rightarrow Y}$ under the RD scale can be described using the following expression. $\lambda_{Q \rightarrow H \rightarrow Y}/(\lambda_{Q \rightarrow Y} + \lambda_{Q \rightarrow H \rightarrow Y} + \lambda_{Q \rightarrow WY})$. Under the log risk ratio scale, the proportion of the mediated effect of path 2 is $\log(\rho_{Q \rightarrow H \rightarrow Y})/\log(\rho_{Q \rightarrow Y} * \rho_{Q \rightarrow H \rightarrow Y} * \rho_{Q \rightarrow WY})$. These expressions can be similarly applied to path 3 under both the risk difference and risk ratio scales. The proportion of mediation helps quantify the relative contribution of path-specific effects but is not meaningful if the effects are in different directions.

Interaction terms can also be incorporated into these estimates by including these terms in the regression models and in the calculations using the delta method. In our analyses, we present results with exposure-mediator interactions, but mediator-mediator interactions can also be accommodated in a similar fashion (Supporting Information S1).

2.3 Simulation study

Numerical simulation is conducted to evaluate the performance of our estimators under the finite sample. Simulations studies of 1000 replicates are generated based on the regression equations 2-4 using R (core team, Austria) and Rstudio (Rstudio, USA). Values for the first mediator (W) are randomly sampled from a binomial distribution with the probability equal to the equation $P(W=1) = \exp(\delta_0 + \delta_q Q)/(1 + \exp(\delta_0 + \delta_q Q))$ where δ_0 , δ_q and genetic factors (Q) are pre-specified for each scenario listed in Supporting Information Table S2. Values for the second mediator (H) are also randomly sampled from a binomial distribution with the probability equal to the equation $P(H=1) = \exp(a_0 + a_q Q + a_w W)/(1 + \exp(a_0 + a_q Q + a_w W))$ where a_0 , a_q and a_w were pre-specified for each scenario and previously generated values of Q and W were used. Values for HCC status (Y) are also generated from a binomial distribution with the probability $P(Y=1) = \exp(\beta_0 + \beta_q Q + \beta_w W + \beta_h H)/(1 + \exp(\beta_0 + \beta_q Q + \beta_w W + \beta_h H))$. Once again, β_0 , β_q , β_w , and β_h are pre-specified for each scenario and values of Q , W , and H are generated from the two previous equations. Guidance on using the R codes implementing these methods are contained in the R help file of the program.

2.4 HCC cohort data

To understand the genetic effects on HCC risk, we apply our method to participants in the Risk Evaluation of Viral Load Elevation and Associated Liver Disease/Cancer-Hepatitis B Virus (REVEAL-HBV) cohort and the Taiwan Liver Cancer Network (TLCN) (Chen & Yang, 2011). Detailed information regarding the study design and participant characteristics have been described elsewhere (Chang et al., 2016; Chen & Yang, 2011; National Health Research Institutes, 2011). In short, the REVEAL-HBV cohort study followed participants aged 30 to 65 in Taiwan from 1991 to 2004 who were Hepatitis B virus surface antigen positive (indicating chronic infection with HBV (El-Serag, 2012)), and were anti-Hepatitis C Virus antibody (anti-HCV) negative at enrollment. Participants included in our analysis from TLCN were liver cancer patients from five medical centers in Taiwan who were over 30 years of age and anti-HCV seronegative. We only include participants who reported

information for all covariates and assume missing entries are missing at random. A total of 580 cases of HCC and 3207 individuals without HCC were included in our analyses and baseline characteristics are described in Supporting Information Table S2. The prevalence of the T allele in the ADH1B gene is approximately 92% in participants who developed HCC and 93% in those who did not. The prevalence of the A allele in the ALDH2 gene was found in approximately 50% of the participants in both cases and non-cases of HCC.

Following model 1 in Figure 2, we let Q represent the genotype for each SNP, W as whether the individual drinks four or more times per week for at least one year (yes vs. no), H as whether the individual has ALT levels equal to or higher than 45 U/L (yes vs. no), and Y as HCC status (Liu et al., 2016). Although the mediators (alcohol consumption and ALT) available to us in this application were dichotomous, we note that their underlying measures are supposed to be continuous. Caution should be taken when dichotomizing continuous mediator values, which may cause power losses or may even raise the issue of misclassification. In these scenarios, alternative Monte Carlo-based methods without closed form expression (Daniel et al., 2015) should be considered. Weighted logistic regressions are used to model the outcome and both mediators, adjusted for potential confounding factors: age, gender, and HBV DNA levels (El-Serag, 2012; Zhu et al., 2016). Cases are re-weighted in the REVEAL-HBV and TLCN cohort to account for unequal sampling and these weights are described in Supporting Information S3. For the reported mediation effects, we specified values for those who are at a higher risk of HCC; those of older age, men, having DNA levels over one million copies/mL, and the higher risk genotype in the other SNP (Zhu et al., 2016).

3 RESULTS

3.1 Simulation results

Overall, the calculated 95% confidence intervals from the simulations have proper coverages. With a dichotomous exposure, a higher prevalence of the outcome result in a lower CI coverage in path 3 ($\lambda_{Q \rightarrow WY}$) (for an exposure prevalence of 0.2, coverage is 89.5% in scenario 1 with an outcome prevalence of 0.618 compared to 95.4% in scenario 2 with an outcome prevalence of 0.288) and a lower prevalence of the outcome results in lower coverage of the true effects of path 2 ($\lambda_{Q \rightarrow H \rightarrow Y}$) (for an exposure prevalence of 0.2, 95% confidence interval coverage for risk difference estimates is 93.3% for scenario 1 and 91.3% in scenario 2) (Figures 3, 4 and Supporting Information Table S4 and 5). This can be observed similarly comparing all four simulation settings with the same exposure prevalence (Figures 3, 4 and Supporting Information Table S4 and 5). For the scenarios with less coverage for path 2, we find that the potential outcome probabilities to also be smaller and for higher potential outcome probabilities path 3 tends to be less well covered (Supporting Information Table S6). Furthermore, a lower prevalence of the exposure may generate more conservative confidence intervals (for simulation 2, 95% confidence interval: -0.12, 0.01 for risk difference estimates with an exposure prevalence of 0.2, and for an exposure prevalence of 0.6 95% CI: -0.14, -0.03) for the path-specific estimates due to a smaller proportion of individuals within each combination of the exposure, mediators, and outcome seen in simulations 2 and 4 (Supporting Information Table S7-9, eFigure 10 and 11). On the other

hand, a continuous exposure value (which may be of interest for other genomic applications) more consistently provided 95% confidence intervals capturing the true path-specific effect (Figures 3,4). For the scenarios tested, average bias and mean squared error were largest for path 1 under both the risk ratio and risk difference scale (Supporting Information Table S4 and 5).

The bootstrap method consistently estimates the true path-specific effects, and, unlike the delta method, estimates did not seem to be affected by parameter settings (Supporting Information Table S12). Using a simulated data set of size 1000, the 95% risk difference and risk ratio confidence intervals generated using the bootstrap method seems to be very similar to that using the delta method, although it may be slightly wider under certain conditions to estimate paths 2 and 3 and not path 1 (Supporting Information S8 and 9). Simulation 2 with a prevalence of 0.2 has a larger confidence interval using both the bootstrap and delta methods compared to a prevalence of 0.49, which may explain why there was better coverage of the true effects for the 95% confidence intervals. The 95% confidence intervals and P-values under the risk difference and risk ratio scales can be found in Supporting Information Tables 8 and 9. Estimates of path 3 are less accurate with higher exposure prevalence and we see that confidence intervals tended to be more conservative when using the delta method in these parameters settings (Supporting Information S10 and 11).

Analyzing data simulation 1, the delta method was faster in computing all three path-specific effects as it took 0.154 seconds while the bootstrap method resampling 1000 times took 22.478 seconds using a 2.2 GHz Intel Core i7 processor and 8GB 1600 MHz DDR3 memory.

3.2 HCC data analysis

A dominant model was used to assess both SNP effects as the presence of any of these deficient alleles affect enzyme activity: the T allele in the ADH1B gene causes more alcohol to be metabolized and the A allele in the ALDH2 gene causes an amino acid change which decreases acetaldehyde metabolism (Lewis & Smith, 2005; Liu et al., 2016). Furthermore, the deficient allele in the ADH1B gene decreases the likelihood of an individual to consume alcohol (Bierut et al., 2012). The individuals homozygous for the deficient allele in the ALDH2 gene show a more severe reaction to alcohol compared to individuals who are heterozygous for this gene, but in both individuals, there is a decrease in acetaldehyde metabolism (Lewis & Smith, 2005). Since both alcohol consumption and ALT levels were measured at baseline, there are various possibilities in which the mediators may affect each other indicated by model 1 and 2 in Figure 2. Model 1 was chosen over model 2 since the Akaike Information Criterion values were lower when more variables were included in the regression model (Supporting Information Table S13). Under the causal structure in model 1 (Figure 2) in which alcohol consumption affects ALT levels, having a T allele in the ADH1B gene (rs1229984) mediated only by alcohol consumption resulted in a decreased HCC risk of -0.016 (95% CI: -0.030, -0.002) or by a ratio of 0.970 (95% CI: 0.945, 0.996) (Table 1). The proportion of mediation for this effect is 0.593 under the RD scale and 0.591 under the log RR scale. The direct effect of the SNP and indirect effect through ALT levels did not have notable effects (proportion of mediation 25.9% RD scale).

Similar results were observed for the ALDH2 (rs671) SNP gene in Table 2 for those with the A allele in which the mediated effect through alcohol consumption decreased HCC risk by -0.47 (95% CI: -0.067 , -0.027) or by a factor of 0.917 (95% CI: 0.881 , 0.955). Since the direct effect of the gene and indirect effect only through ALT levels has an opposite effect on HCC risk, calculating the proportion of mediation is not meaningful. Including interaction terms between each SNP and alcohol consumption did not greatly change RD and RR estimates (Table 3 and 4).

Additional analyses were conducted to assess the robustness of our findings. We conducted a sensitivity analysis to quantify the magnitude in which an unmeasured confounder (family history of HCC) of alcohol consumption and HCC risk may affect our results. Since only the effect through alcohol consumption was significant in model 1, we used the sensitivity method to understand the biased effects through one mediator (VanderWeele, 2010). The sensitivity analyses suggested that our results are robust: the magnitude of the bias due to family history of HCC cannot fully explain the effects we observed (Supporting Information S14). The range of parameters used in our sensitivity analyses reflect what has been observed in the REVEAL cohort: prevalence of having a family history of HCC ranging from 2-3% and risk ratio of 1.3 for the effect of family history on HCC (Loomba et al., 2013). Family history was not collected in the TLCN population, and is the reason why it was not adjusted for in the analyses. Using a recessive and additive genetic model replicated most of the results, although the mediation effects through drinking were attenuated for the ADH1B gene (Supporting Information Table S15 and S16). If the true causal path is more accurately represented by model 2 of Figure 2, the mediated effect through drinking still remained significant (Supporting Information Table S17).

4 DISCUSSION

The method described in this paper can be used to assess the mediation effects for multiple dichotomous mediators on a dichotomous outcome with a closed form estimator for both a dichotomous and continuous exposure. Existing mediation methods are able to quantify the single mediation effects on a dichotomous outcome (Imai et al., 2010; Valeri & Vanderweele, 2013; Vanderweele & Vansteelandt, 2010), but may require a rare outcome in order to estimate the effects (Valeri & Vanderweele, 2013; Vanderweele & Vansteelandt, 2010). When multiple mediators are considered, the existing mediation methods may be computationally intensive or may require additional sensitivity parameters to estimate the relationship between the two mediators (Daniel et al., 2015; J. Wang et al., 2012). The proposed method differs from other methods in that it can estimate the probability of a dichotomous outcome, and therefore, any effect involving a contrast of the probability under two exposure status, such as the risk difference or risk ratio, without the rare outcome assumption and additional assumptions regarding the relationship between the mediators. We present the derivation and application for the two-mediator model, which can be easily extended to accommodate more than two mediators by following the similar development. We also present variance of the estimates by delta method and bootstrapping. Statistical inference based on the delta method is computationally efficient, compared with the bootstrap approach. If the prevalence of mediators and the outcome is too small or too large, the delta method-based inference may not be optimal since it requires large sample

approximation. Under such extreme scenarios, bootstrap method may be a plausible alternative provided that the computation cost is affordable. Our method can be adapted to analyze data from a case-control study using weighting methods as it was done previously in other mediation analyses (Vanderweele & Vansteelandt, 2010; Vanderweele et al., 2014; W. Wang et al., 2013). Due to the nature of this study, we only had coarsened measurements of our mediators. In other applications, it may be more favorable to conduct analyses using continuous measurements, and this motivates additional methods to accommodate these distributions.

We conclude that the genetic effects of ADH1B and ALDH2 on HCC are more strongly mediated by alcohol consumption and less by ALT levels. Through applying this mediation method, we are able to identify that effects of ALDH2 on HCC may vary depending on the intermediary factors involved and that the genetic effect through alcohol consumption habits actually decreased HCC risk. This information could provide additional evidence regarding the importance of addressing HCC risk by decreasing alcohol consumption in those with chronic HBV infection and with normal ADH1B and ALDH2 genotypes. The other mediator, ALT, did not have as strong of a mediation effect independent of alcohol consumption habits, which suggests either that ALT levels are not intermediate factors in the effects of these genes on HCC risk or the effects of ALT levels on HCC risk can be captured by the mediation effects through alcohol consumption. This result does not take away from the importance of ALT being a helpful screening test for liver disease compared to measuring alcohol consumption habits, but it may encourage additional research to understand whether additional ALT measurements could better identify the mediation effects independent of alcohol consumption (Kim et al., 2008). Our sensitivity analyses identified that our estimates are robust, and using different genetic models and causal relationships all supported the protective effects of the ALDH2 gene on HCC risk through drinking. The attenuated results for the ADH1B SNP may suggest a need for additional studies to further identify similar trends in other populations, but this may not be accessible as the allele frequencies may be very low (Bierut et al., 2012). Future development of this model would allow for more rigorous sensitivity analyses to quantifying the effects of other potentially unmeasured confounders (Alatalo et al., 2008; Kim et al., 2008; Liu et al., 2016). Most importantly, these results motivate additional mediation analyses to understand the other mechanisms that could describe the interplay between genetic and modifiable factors leading to disease. The R (core team, Austria) function to conduct the mediation analysis can be downloaded from <http://www.stat.sinica.edu.tw/personinfo/?sysid=5017880>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by US National Institutes of Health 5R03CA182937-02 (YTH), and Taiwan Ministry of Science and Technology 105-2118-M-001-014-MY3 (YTH).

We also like to acknowledge two anonymous reviewers for their helpful feedback on earlier drafts of the manuscript.

References

- Alatalo PI, Koivisto HM, Hietala JP, Puukka KS, Bloigu R, Niemela OJ. Effect of moderate alcohol consumption on liver enzymes increases with increasing body mass index. *Am J Clin Nutr*. 2008; 88(4):1097–1103. [PubMed: 18842799]
- Avin C, Shpitser I, Pearl J. Identifiability of path-specific effects. *Proceedings of the International Joint Conference on Artificial Intelligence*. 2005:357–363.
- 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):1173–1182. [PubMed: 3806354]
- Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Edenberg HJ. ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. *Mol Psychiatry*. 2012; 17(4):445–450. DOI: 10.1038/mp.2011.124 [PubMed: 21968928]
- Chang IC, Huang SF, Chen PJ, Chen CL, Chen CL, Wu CC, Liaw YF. The Hepatitis Viral Status in Patients With Hepatocellular Carcinoma: a Study of 3843 Patients From Taiwan Liver Cancer Network. *Medicine (Baltimore)*. 2016; 95(15):e3284.doi: 10.1097/md.0000000000003284 [PubMed: 27082566]
- Chen CJ, Yang HI. Natural history of chronic hepatitis B REVEALed. *J Gastroenterol Hepatol*. 2011; 26(4):628–638. DOI: 10.1111/j.1440-1746.2011.06695.x [PubMed: 21323729]
- Conigrave KM, Davies P, Haber P, Whitfield JB. Traditional markers of excessive alcohol use. *Addiction*. 2003; 98(Suppl 2):31–43. [PubMed: 14984240]
- Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S. Causal mediation analysis with multiple mediators. *Biometrics*. 2015; 71(1):1–14. DOI: 10.1111/biom.12248 [PubMed: 25351114]
- Edenberg HJ, Xuei X, Chen HJ, Tian H, Wetherill LF, Dick DM, Foroud T. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. *Hum Mol Genet*. 2006; 15(9):1539–1549. DOI: 10.1093/hmg/ddl073 [PubMed: 16571603]
- El-Serag HB. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology*. 2012; 142(6):1264–1273. e1261. DOI: 10.1053/j.gastro.2011.12.061 [PubMed: 22537432]
- Eng MY, Luczak SE, Wall TL. ALDH2, ADH1B, and ADH1C genotypes in Asians: a literature review. *Alcohol Res Health*. 2007; 30(1):22–27. [PubMed: 17718397]
- Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. *J Carcinog*. 2017; 16doi: 10.4103/jcar.JCar_9_16
- Hosmer, DW., Lemeshow, S., May, S. *Applied Survival Analysis*. John Wiley & Sons, Inc; 2008. Appendix 1: The Delta Method; p. 355–358.
- Huang YT, Cai T. Mediation analysis for survival data using semiparametric probit models. *Biometrics*. 2016; 72(2):563–574. DOI: 10.1111/biom.12445 [PubMed: 26618735]
- Huang YT, Pan WC. Hypothesis test of mediation effect in causal mediation model with high-dimensional continuous mediators. *Biometrics*. 2016; 72(2):402–413. DOI: 10.1111/biom.12421 [PubMed: 26414245]
- Huang YT, Yang HI. Causal mediation analysis of survival outcome with multiple mediators. *Epidemiology*. 2017; doi: 10.1097/ede.0000000000000651
- Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods*. 2010; 15(4):309–334. DOI: 10.1037/a0020761 [PubMed: 20954780]
- Kim WR, Flamm SL, Di Bisceglie AM, Bodenheimer HC. Serum activity of alanine aminotransferase (ALT) as an indicator of health and disease. *Hepatology*. 2008; 47(4):1363–1370. DOI: 10.1002/hep.22109 [PubMed: 18366115]
- Lange T, Hansen JV. Direct and indirect effects in a survival context. *Epidemiology*. 2011; 22(4):575–581. DOI: 10.1097/EDE.0b013e31821c680c [PubMed: 21552129]
- Lewis SJ, Smith GD. Alcohol, ALDH2, and esophageal cancer: a meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach. *Cancer Epidemiol Biomarkers Prev*. 2005; 14(8):1967–1971. DOI: 10.1158/1055-9965.epi-05-0196 [PubMed: 16103445]
- Liu J, Yang HI, Lee MH, Jen CL, Hu HH, Lu SN, Chen CJ. Alcohol Drinking Mediates the Association between Polymorphisms of ADH1B and ALDH2 and Hepatitis B-Related

- Hepatocellular Carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2016; 25(4):693–699. DOI: 10.1158/1055-9965.epi-15-0961 [PubMed: 26827895]
- Loomba R, Liu J, Yang H, Lee M, Lu S, Wang L, Chen C. Synergistic Effects of Family History of Hepatocellular Carcinoma and Hepatitis B Virus Infection on Risk for Incident Hepatocellular Carcinoma. *Clin Gastroenterol Hepatol.* 2013; 11(12):1636–1645.e1631-1633. DOI: 10.1016/j.cgh.2013.04.043 [PubMed: 23669307]
- National Health Research Institutes. Introduction to Taiwan Liver Cancer Network. 2011. Retrieved from http://tlcn.nhri.org.tw:8080/etlcn/webcont!Cont.action?news_id=201109221316678798919&lab_id=etlcn
- Pearl, J. Direct and indirect effects. San Francisco, CA: Morgan Kaufmann Publishers Inc; 2001.
- Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods.* 2008; 40(3):879–891. [PubMed: 18697684]
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology.* 1992; 3(2):143–155. [PubMed: 1576220]
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology.* 1974; 66(5):688–701.
- Seitz HK, Stickel F. Risk factors and mechanisms of hepatocarcinogenesis with special emphasis on alcohol and oxidative stress. *Biol Chem.* 2006; 387(4):349–360. DOI: 10.1515/bc.2006.047 [PubMed: 16606331]
- Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods.* 2013; 18(2):137–150. DOI: 10.1037/a0031034 [PubMed: 23379553]
- VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology.* 2010; 21(4):540–551. DOI: 10.1097/EDE.0b013e3181df191c [PubMed: 20479643]
- VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology.* 2011; 22(4):582–585. DOI: 10.1097/EDE.0b013e31821db37e [PubMed: 21642779]
- Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol.* 2010; 172(12):1339–1348. DOI: 10.1093/aje/kwq332 [PubMed: 21036955]
- VanderWeele TJ, Vansteelandt S. Mediation Analysis with Multiple Mediators. *Epidemiol Method.* 2014; 2(1):95–115. DOI: 10.1515/em-2012-0010 [PubMed: 25580377]
- Vanderweele TJ, Vansteelandt S, Robins JM. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology.* 2014; 25(2):300–306. DOI: 10.1097/ede.0000000000000034 [PubMed: 24487213]
- Vansteelandt S, Daniel RM. Interventional Effects for Mediation Analysis with Multiple Mediators. *Epidemiology.* 2017; 28(2):258–265. DOI: 10.1097/ede.0000000000000596 [PubMed: 27922534]
- Wang J, Spitz MR, Amos CI, Wu X, Wetter DW, Cinciripini PM, Shete S. Method for evaluating multiple mediators: mediating effects of smoking and COPD on the association between the CHRNA5-A3 variant and lung cancer risk. *PLoS One.* 2012; 7(10):e47705.doi: 10.1371/journal.pone.0047705 [PubMed: 23077662]
- Wang W, Nelson S, Albert JM. Estimation of causal mediation effects for a dichotomous outcome in multiple-mediator models using the mediation formula. *Stat Med.* 2013; 32(24):4211–4228. DOI: 10.1002/sim.5830 [PubMed: 23650048]
- Zhu RX, Seto WK, Lai CL, Yuen MF. Epidemiology of Hepatocellular Carcinoma in the Asia-Pacific Region. *Gut Liver.* 2016; 10(3):332–339. DOI: 10.5009/gnl15257 [PubMed: 27114433]

Figure 1. Causal Diagrams of the Path-Specific Effects. All possible effects are represented by diagram A, while diagram B represents the effect of Q on Y independent of W and H, diagram C represents the effect of Q on Y through H only, and diagram D represents the effect of Q on Y through W and potentially H.

Model 1

Model 2

Figure 2.
Two Candidate Causal Models of the Liver Cancer Study.

Figure 3. 95% Confidence Interval Coverage of the Path-Specific Effects for a Dichotomous or Continuous Exposure Under the Risk Difference Scale. Path 1 estimates are fully shaded, path 2 estimates are partially shaded, and path 3 estimates are not shaded. Setting parameters are listed in Supporting Information.

Figure 4. 95% Confidence Interval Coverage of the Path-Specific Effects for a Dichotomous or Continuous Exposure Under the Risk Ratio Scale. Path 1 estimates are fully shaded, path 2 estimates are partially shaded, and path 3 estimates are not shaded. Setting parameters are listed in Supporting Information.

Table 1

Risk Difference Estimates of Path-Specific Effects of SNPs on HCC Risk. The mediation effects reported reflect participants with characteristics associated with a higher risk: males, highest DNA level, highest age group, and have the CT/TT genotypes for ADH1B or GA/AA genotypes for ALDH2. Confidence Intervals (CI) were estimated using delta method and bootstrapping resampling size of 500.

Estimates	Risk Difference		
	Direct effect $\lambda_{Q \rightarrow Y}$	Indirect effect through ALT levels $\lambda_{Q \rightarrow H \rightarrow Y}$	Indirect effect through alcohol consumption $\lambda_{Q \rightarrow WY}$
ADH1B			
Point estimate	-0.004	-0.007	-0.016
Delta Method CI	(-0.097, 0.089)	(-0.047, 0.033)	(-0.030, -0.002)
Bootstrapping CI	(-0.067, 0.072)	(-0.058, 0.044)	(-0.045, -0.007)
ALDH2			
Point estimate	0.002	0.017	-0.047
Delta Method CI	(-0.051, 0.055)	(-0.005, 0.039)	(-0.067, -0.027)
Bootstrapping CI	(0.019, 0.089)	(-0.007, 0.050)	(-0.088, -0.047)

Table 2

Risk Ratio Estimates of Path-Specific Effects of SNPs on HCC Risk. The mediation effects reported reflect participants with characteristics associated with a higher risk: males, highest DNA level, highest age group, and have the CT/TT genotypes for ADH1B or GA/AA genotypes for ALDH2. Confidence Intervals (CI) were estimated using delta method and bootstrapping resampling size of 500.

Estimates	Risk Ratio		
	Direct effect $\rho_{Q \rightarrow Y}$	Indirect effect through ALT levels $\rho_{Q \rightarrow H \rightarrow Y}$	Indirect effect through alcohol consumption $\rho_{Q \rightarrow WY}$
ADH1B			
Point estimate	0.992	0.987	0.97
Delta Method CI	(0.838, 1.175)	(0.917, 1.063)	(0.945, 0.996)
Bootstrapping CI	(0.848, 1.225)	(0.859, 1.129)	(0.875, 0.980)
ALDH2			
Point estimate	1.004	1.031	0.917
Delta Method CI	(0.912, 1.105)	(0.991, 1.072)	(0.881, 0.955)
Bootstrapping CI	(1.052, 1.270)	(0.983, 1.134)	(0.784, 0.882)

Table 3

Risk Difference Estimates of Path-Specific Effects of SNPs on HCC risk with SNP and Alcohol Consumption Interaction. The mediation effects reported reflect participants with characteristics associated with a higher risk: males, highest DNA level, highest age group, and have CT/TT genotypes for ADH1B or GA/AA genotypes for ALDH2. Confidence Intervals (CI) were estimated using delta method and bootstrapping resampling size of 500.

Estimates	Risk Difference		
	Direct effect $\lambda_{Q \rightarrow Y}$	Indirect effect through ALT levels $\lambda_{Q \rightarrow H \rightarrow Y}$	Indirect effect through alcohol consumption $\lambda_{Q \rightarrow WY}$
ADH1B and alcohol consumption			
Point estimate	-0.004	-0.013	-0.017
Delta Method CI	(-0.099, 0.091)	(-0.055, 0.029)	(-0.031, -0.003)
Bootstrapping CI	(-0.067, 0.079)	(-0.067, 0.041)	(-0.049, -0.008)
ALDH2 and alcohol consumption			
Point estimate	0.011	0.015	-0.054
Delta Method CI	(-0.045, 0.068)	(-0.008, 0.038)	(-0.084, -0.024)
Bootstrapping CI	(0.023, 0.105)	(-0.008, 0.055)	(-0.110, -0.047)

Table 4

Risk Ratio Estimates of Path-Specific Effects of SNPs on HCC risk with SNP and Alcohol Consumption Interaction. The mediation effects reported reflect participants with characteristics associated with a higher risk: males, highest DNA level, highest age group, and have the CT/TT genotypes for ADH1B or GA/AA genotypes for ALDH2. Confidence Intervals (CI) were estimated using delta method and bootstrapping resampling size of 500.

Estimates	Risk Ratio		
	Direct effect $\rho_{Q \rightarrow Y}$	Indirect effect through ALT levels $\rho_{Q \rightarrow H \rightarrow Y}$	Indirect effect through alcohol consumption $\rho_{Q \rightarrow WY}$
ADH1B and alcohol consumption			
Point estimate	0.993	0.977	0.969
Delta Method CI	(0.837, 1.177)	(0.905, 1.054)	(0.943, 0.996)
Bootstrapping CI	(0.847, 1.233)	(0.846, 1.126)	(0.867, 0.980)
ALDH2 and alcohol consumption			
Point estimate	1.02	1.027	0.906
Delta Method CI	(0.922, 1.128)	(0.986, 1.069)	(0.856, 0.959)
Bootstrapping CI	(1.066, 1.334)	(0.982, 1.135)	(0.752, 0.882)