Generalized interventional approach for causal mediation analysis with causally ordered multiple mediators

Sheng-Hsuan Lin*

*National Chiao Tung University, shenglin@nctu.edu.tw

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

https://biostats.bepress.com/harvardbiostat/paper217

Copyright ©2019 by the author.
Generalized interventional approach for causal mediation analysis with causally ordered multiple mediators

Sheng-Hsuan Lin

Abstract

Causal mediation analysis has demonstrated the advantage of mechanism investigation. In conditions with causally ordered mediators, path-specific effects (PSEs) are introduced for specifying the effect subject to a certain combination of mediators. However, most PSEs are unidentifiable. To address this, an alternative approach termed interventional analogue of PSE (iPSE), is widely applied to effect decomposition. Previous studies that have considered multiple mediators have mainly focused on two-mediator cases due to the complexity of the mediation formula. This study proposes a generalized interventional approach for the settings, with the arbitrary number of ordered multiple mediators to study the causal parameter identification as well as statistical estimation. It provides a general definition of iPSEs with a recursive formula, assumptions for nonparametric identification, a regression-based method, and a g-computation algorithm to estimate all iPSEs. We demonstrate that each iPSE reduces to the result of linear structural equation modeling subject to linear or log-linear models. This approach is applied to a Taiwanese cohort study for exploring the mechanism by which hepatitis C virus infection affects mortality through hepatitis B virus infection, liver function, and hepatocellular carcinoma. Software based on a g-computation algorithm allows users to easily apply this method for data analysis subject to various model choices according to the substantive knowledge for each variable. All methods and software proposed in this study contribute to comprehensively decompose a causal effect confirmed by data science and help disentangling causal mechanisms when the natural pathways are complicated.
Generalized interventional approach for causal mediation analysis with causally ordered multiple mediators

Sheng-Hsuan Lin

Institute of Statistics, National Chiao Tung University, 4F, Assembly Building I, 1001 University Road, Hsinchu 30010, TAIWAN(R.O.C)

shenglin@nctu.edu.tw

SUMMARY

Causal mediation analysis has demonstrated the advantage of mechanism investigation. In conditions with causally ordered mediators, path-specific effects (PSEs) are introduced for specifying the effect subject to a certain combination of mediators. However, most PSEs are unidentifiable. To address this, an alternative approach termed interventional analogue of PSE (iPSE), is widely applied to effect decomposition. Previous studies that have considered multiple mediators have mainly focused on two-mediator cases due to the complexity of the mediation formula. This study proposes a generalized interventional approach for the settings, with the arbitrary number of ordered multiple mediators to study the causal parameter identification as well as statistical estimation. It provides a general definition of iPSEs with a recursive formula, assumptions for nonparametric identification, a regression-based method, and a g-computation algorithm to estimate all iPSEs. We demonstrate that each iPSE reduces to the result of linear structural equation modeling subject to linear or log-linear models. This approach is applied to a Taiwanese cohort study for exploring the mechanism by which hepatitis C virus infection affects mortality through hepatitis B virus infection, liver function, and hepatocellular carcinoma. Software based on a g-computation algorithm allows users to easily apply this method for data analysis subject to various model choices according to the substantive knowledge for each variable. All methods and software proposed in this study contribute to comprehensively decompose a causal effect confirmed by data science and help disentangling causal mechanisms when the natural pathways are complicated.

Key words: Counterfactual; Pathway analysis; Causal mediation analysis; Ordered multiple mediator; Intervventional approach.
1. INTRODUCTION

Investigating the mechanism whereby a variable affects both experimental and observational study participants requires causal interpretation of the effects of an exposure or treatment on the outcome through some intermediate variable mediators. Recently, mediation analysis, such as linear structural equation modeling (SEM), has been widely used to study the causal mechanism among variables by decomposing the total effect of an exposure on an outcome into several pathways according to mediators of interest. Furthermore, causal mediation analysis, defining each path using counterfactual outcome models, expands the mediation analysis into more general settings, such as interaction and nonlinearities (Pearl, 2001; J. M. Robins & Greenland, 1992; T. VanderWeele & Vansteelandt, 2009). Numerous methodological techniques based on causal mediation analysis have been proposed, allowing various outcome scales, including additive and multiplicative scales, and other nonlinear models for time-to-event data (Didelez, Dawid, & Geneletti, 2012; Imai, Keele, & Tingley, 2010; Tchetgen & Shpitser, 2012; Valeri & VanderWeele, 2013; van der Laan & Petersen, 2008; T. VanderWeele & Vansteelandt, 2009; T. J. VanderWeele, 2009b, 2010; T. J. VanderWeele & Vansteelandt, 2010).

In the era of data science, the increasing accessibility of data collection provides the opportunity for investigation of complex causal mechanisms with multiple mediators. Hence, the development of corresponding methods allowing multiple mediators is critical for this setting with multiple mediators. Several approaches have been proposed. One method proposed by VanderWeele and Vansteelandt focused on the evaluation of the direct and indirect effects by treating all mediators as one unit (T. J. VanderWeele & Vansteelandt, 2014). In addition to the direct and indirect effects, the estimation of path-specific effects (PSEs), representing the causal effect through a specific pathway passing a subset of mediators, is required for completing the causal interpretation in the mediation analysis. To this end, Avin et al. and VanderWeele et al. have developed different approaches to identify the part of PSEs nonparametrically (Avin, Shpitser, & Pearl, 2005; T. J. VanderWeele & Vansteelandt, 2014). All types of PSEs cannot be identified nonparametrically (Albert & Nelson, 2011; Daniel, De Stavola, Cousens, & Vansteelandt, 2015). To address this, Lin and VanderWeele introduced randomly interventional definitions of direct and indirect effect to overcome the identification problem in settings with two ordered mediators. Daniel et al. and Lok also undertook similar research to obtain PSEs by changing the original definitions for PSE or by adding model assumptions (Daniel et al., 2015; Lok, 2016). However, because the number of PSEs increases exponentially as the number of mediators increases, the aforementioned research featured only two mediators and provided only approximate concepts for greater numbers of mediators due to notation complexity.
This study proposes a generalized framework in settings under which the mediators sequentially affect each other. We first introduce the generalized definition of the interventional version of PSE. Under the consistency and exchangeability assumptions, the interventional version of PSE is nonparametrically identified and formalized as a recursive formula. A regression-based algorithm and a g-computation algorithm were adopted for estimations.

The article is organized as follows. Section 2 introduces notation and definitions. Section 3 provides the assumptions for identification. Section 4 provides the formula for estimation of mediation effects using a regression-based model for the analytic form and the g-computation algorithm for numerical estimates. Section 5 provides the simulation study and application for the Taiwan cohort dataset. Section 6 discusses the limitations and potential relevant future research.
2. NOTATION AND DEFINITION FOR GENERALIZED INTERVENTIONAL
PATH-SPECIFIC EFFECTS

Let $A$ denote the exposure, $M = (M_1, M_2, ..., M_K)$ the causally ordered multiple mediators, $Y$ the outcome, $C_0$ the baseline confounders, and $C = (C_1, C_2, C_3, ..., C_K)$ the time-varying confounders. $C_i$ represents the confounders among $M_i$ and $Y$ for $i \in \{1,2, ..., K\}$. The causal relationships among these variables are demonstrated by a direct acyclic graph in Fig. 1.

For two nonnegative ordered integers $i_1$ and $i_2$, where $i_1 < i_2$, we denote $V_{(i_1,i_2)} = (v_{i_1}, v_{i_1+1}, ..., v_{i_2})$ as a subvector of a vector $V$; for $i_1 = i_2 = i$, we further define $V_{(i,i)} = v_i$, and for $i_1 > i_2$, $V_{(i_1,i_2)}$ is defined as a null vector. Let $Y(a, m_{(1,K)})$ be the counterfactual outcome given $(A, M_{(1,K)})$ is set to $(a, m_{(1,K)})$ (J. Robins, 1986). Let $M_i(a, m_{(1,i-1)})$ be the counterfactual value of $M_i$ given $(A, M_{(1,i-1)})$ is set to $(a, m_{(1,i-1)})$ for $i \in \{1,2,...,K\}$. Furthermore, we assume consistency (Pearl, 2009; T. VanderWeele & Vansteelandt, 2009; T. J. VanderWeele, 2009a), under which the observed $(A, M_{(1,K)})$ is equal to $(a, m_{(1,K)})$, and $Y(a, m_1, m_2, ..., m_K)$ is equal to the observed $Y$.

In the setting with $K$ causally ordered multiple mediators $M_{(1,K)}$, the number of PSEs increases exponentially ($= 2^K$) according to the involvement of $M_{(1,K)}$. Therefore, a definition system is required for a generalized setting. We first define the set of all paths by $K$ ordered mediators as $S = \{s | s = (I(M_1), ..., I(M_K))$ and $I(M_i) \in \{0,1\}\}$, where $I(M_i) = 1$ represents the path $s$ passing through the $i$-th mediator, $M_i$. For simplicity, each path $s = (I(M_1), ..., I(M_K))$ in $S$ is numbered as $d$, which is an integer converted by a one-to-one converted function: $f \left( (I(M_1), ..., I(M_K)) \right) = \sum_{i=1}^{K} I(M_i) \times 2^{i-1} + 1$. Use of a one-to-one function clearly ensures that each converted number is specifically mapped to one path. Consequently, on the basis of these converted numbers, the path-specific effect can be defined as a function of the converted number as follows:

**DEFINITION 1 (Path-Specific Effect, $PSE_K(d)$).**

$PSE_K(d)$ is the path-specific effect with respect to the path numbered $d$, where $d \in \{1,2,3, ..., 2^K\}$, and $K$ is the number of mediators.

Because most of the $PSE_K(d)$ elements that are based on standard definitions are not identified (Avin et al., 2005; T. J. Vanderweele, Vansteelandt, & Robins, 2014), an
alternative PSE definition (S.-H. Lin & VanderWeele, 2017; T. J. VanderWeele &
Tchetgen Tchetgen, 2017; T. J. VanderWeele & Vansteelandt, 2014) for generalized
settings with an arbitrary number of ordered mediators is termed “generalized
interventional path-specific effect” (giPSE). Before defining the giPSE, we must define
“iterative random draw of cross-world counterfactual mediators” and a “generalized
interventional mediation parameter” in advance, as Definition 2 and Definition 3.

DEFINITION 2 (Iterative random draw of cross-world counterfactual mediators,
\[
G_i \left( a_{(1,2^{i-1})} \right),
\]
\( G_1(a_1) \) is a random draw of \( M_1(a_1) \), which is the counterfactual value of \( M_1 \) given
\( A = a_1 \). \( G_2(a_1, a_2) \) is a random draw of \( M_2(a_1, G_1(a_2)) \), which is the counterfactual
value of \( M_2 \) given \( (A,M_1) \) is set to \( (a_1,G_1(a_2)) \). Consequently, for \( i \in \{3, ..., K\} \),
let \( G_i \left( a_{(1,2^{i-1})} \right) \) be a random draw of \( M_i(a_1,G_1(a_2), ..., G_{i-1} \left( a_{(2^{i-2}+1,2^{i-1})} \right)) \),
which is the counterfactual value of \( M_i \) given \( (A,M_{(1,i-1)}) \) is set to
\( (a_1,G_1(a_2), ..., G_{i-1} \left( a_{(2^{i-2}+1,2^{i-1})} \right)) \). For any \( i \in \{1, ..., K\} \), \( G_i \) is a function of
\( a_{(1,2^{i-1})} \).

On the basis of Definition 2, we can further define mediation parameters in a
generalized interventional form as Definition 3.

DEFINITION 3. The generalized interventional mediation parameter \( \varphi_K \left( a_{(1,2^K)} \right) \) is
given by
\[
\varphi_K \left( a_{(1,2^K)} \right) \equiv E \left[ Y \left( a_1, G_1(a_2), G_2(a_3,a_4), ..., G_K \left( a_{(2^{K-1}+1,2^K)} \right) \right) \right]
\]
The mediation parameter in Definition 3 is the expectation of a counterfactual outcome
given that \( (A,M_{(1,K)}) \) is set to \( (a_1,G_1(a_2), G_2(a_3,a_4), ..., G_K \left( a_{(2^{K-1}+1,2^K)} \right)) \).
Compared with the traditional definition, the traditional mediation parameter with the
settings for one mediator is \( E \left( Y(\{a_1,M_1(a_2)\}) \right) \) and with the settings for two mediators
is $E\left(Y\left(a_1, M_1(a_2), M_2(a_3, M_1(a_4))\right)\right)$. For two or more mediators, the PSE of each path cannot be based on the traditional definition, due to lack of identifiability. Definition 3 can be modified to fit the time-dependent outcome, and, more specifically, $\varphi_K(a_{1,2})$ is defined as the counterfactual survival function or hazard function in survival analysis. For simplicity, we mainly discuss the mediation parameter on the basis of the expectation. Next, we can use $\varphi$ to define giPSE.

**DEFINITION 4.** The generalized interventional path-specific effect (giPSE) is given by

$$\text{giPSE}(d) = Q(\varphi_K(a_1, a_2, ..., a_d = a^\ast_1, ..., a_{2^j}) \varphi_K(a_1, a_2, ..., a_d = a^\ast_0, ..., a_{2^j}))$$

In Definition 4, giPSE(d) is defined in terms of the change of $\varphi_K$ by changing the value of $a_d$ from $a^\ast_0$ to $a^\ast_1$ subject to fixing of all other variables’ values, and the definition of mediation parameters guarantees that the influence of changing $a_d$ reflects the effect of the exposure on the outcome through the d-th path. To clarify the relationship between the variable $a_d$ and the d-th path, we adopt a reverse interpretation. We assume that the d-th path consists of J mediators, $M_{t_1}, M_{t_2}, ..., M_{t_j}$ ($t_1 < t_2 < ... < t_j$). The variable structure of $\varphi_K(a_{1,2})$ reveals that $2^j-1$ variables are set before the $M_{t_j}$-related variables (that is, $a_{(2^j-1+1,2^j)}$), and among $G_{t_j}\left(a_{(2^j-1+1,2^j)}\right)$, $2^j-1$ variables are set before the $M_{t_{j-1}}$-related variables. Subsequently, in layer j, $a_{(2^j-1+1,2^j)}$ variables are set before the $M_{t_{j-1}}$-related variables. By counting the number of the prevariables of each layer, the order number of the variable related to the d-th path in $\varphi_K(a_{1,2})$ is $\sum_{j=1}^{l} 2^j-1 + 1$. The value of d is exactly equal to $\sum_{j=1}^{l} 2^j-1 + 1$ by the aforementioned definition. Accordingly, the difference in $\varphi_K$ caused by changes in $a_d$ is used to represent the effects of the d-th path.

The quantity of giPSE(d) is determined by a nonspecific comparative function $Q(x_1, x_2)$. For example, if Y is a binary variable, three types of $Q(x_1, x_2)$ are commonly used in medical research: (1) $Q(x_1, x_2) = (x_1 - x_2)$ for the risk difference scale, (2) $Q(x_1, x_2) = \frac{x_1}{x_2}$ for the risk ratio scale, and (3) $Q(x_1, x_2) =
\( \frac{x_1/(1-x_1)}{x_2/(1-x_2)} \) for the odds ratio scale. For simplicity, we use \( Q(x_1, x_2) = (x_1 - x_2) \) for the following article.

DEFINITION 5 (giPSE for decomposition of giTE).

Although \((a_1, a_2, \ldots, a_{d-1}, a_{d+1}, \ldots, a_{2K})\) can take any value in Definition 4, we specify giPSE using the following expression for convenience of decomposition:

\[
\text{giPSE}_K(d) = \varphi \left( \bar{a}^*_{(1)} d, \bar{a}^*_{(0)} 2^K - d \right) - \varphi \left( \bar{a}^*_{(1)} d-1, \bar{a}^*_{(0)} 2^K - d+1 \right)
\]

\[
\text{giTE}_K = \sum_{d=1}^{2^k} \text{giPSE}_K(d)
\]

where \( \bar{a}^*_{(1)} \) and \( \bar{a}^*_{(0)} \) represents a vector composed by \( a^*_{(1)} \) and \( a^*_{(0)} \) with length \( i \), respectively.

We provide examples of Definitions 1 to 5 with \( K = 1, 2, \) and \( 3 \) in Appendix 1.
3. IDENTIFICATION

To identify giPSE from the empirical data, a series of assumptions, for no unmeasured confounding among exposure, outcome, and mediators is required:

**Assumption 1.** No unmeasured confounding among exposure and outcome

\[ Y(a, m_{(1,K)}) \perp A | C_0. \]

**Assumption 2.** No unmeasured confounding among mediators and outcome

\[ Y(a, m_{(1,K)}) \perp M_i | C_{(0,i)}, A, M_{(1,i-1)} \text{ for } i \in \{1,2, ..., K\} \]

**Assumption 3.** No unmeasured confounding among mediators and exposure

\[ M_k(a, m_{(1,K-1)}) \perp A | C_0 \]

**Assumption 4.** No unmeasured confounding among mediators

\[ M_k(a, m_{(1,K-1)}) \perp M_i | C_{(0,i)}, A, M_{(1,i-1)} \text{ for } i \in \{1,2, ..., k-1\} \text{ and } k \in \{2, ..., K\} \]

Given these assumptions, two lemmas are introduced in this section for completing the identification of the generalized interventional mediation parameter (Definition 3). **Lemma 1** and **Lemma 2** demonstrate the identification of the conditional expectation for counterfactual outcome and the density functions of mediators, separately. **Lemma 2** further provides a recursive formula concerning mediators, and the development of the recursion significantly reduces the complexity of the identification result for giPSE with numerous mediators.

**Lemma 1 (g-formula for outcome).**

Under consistency assumption and Assumptions 1 and 2, the conditional expectation of a counterfactual outcome \( E[Y(a_1, m_{(1,K)}) | C_0] \), given \((A, M_{(1,K)})\) is set to \((a, m_{(1,K)})\), can be identified as

\[
\Gamma(c_0, a_1, m_{(1,K)}) = \int_{C_{(1,K)}} E[Y(a_1, c_{(0,K)}, m_{(1,K)})] \prod_{i=1}^{K} dF_{c_i | C_{(0,i-1)}, A, M_{(1,i-1)}}(c_i | C_{(0,i-1)}, a_1, m_{(1,i-1)}),
\]

for \( i \in \{1,2, ..., K\} \).

https://biostats.bepress.com/harvardbiostat/paper217
\[ \Gamma(c_0, a_1, m_{(1,K)}) \] is exactly equal to the g-formula, where \((A, M_{(1,K)})\) is treated as a time-varying exposure and \(C_{(1,k)}\) as a time-varying confounder conditional on baseline confounder \(C_0 = c_0\).

**Lemma 2** (Recursive g-formula for mediators).

Under consistency assumption and Assumptions 3 and 4, the conditional density of iterative random draw of the cross-world counterfactual mediator \(dF_{\mathcal{G}_i(a_{(2^i-1+1,2^i)})|c_0}(m_i|c_0)\) can be identified as \(H_i(m_i, a_{(2^i-1+1,2^i)}, c_0)\)

where

\[ H_i(m_1, a_1, c_0) = \int_{c_1} dF_{M_1|A, C_1, C_0}(m_1|a_1, c_1, c_0)dF_{C_1|A, C_0}(c_1|a_1, c_0). \]

and for \(i > 1\),

\[ H_i(m_i, a_{(2^i-1+1,2^i)}, c_0) = \]

\[ \int_{m_{(1, i-1)}} \int_{c_{(1,i)}} dF_{M_i|A, M_{(1,i-1)}, C_{(0,i)}, C_0}(m_i|a_{2^{i-1}+1}, m_{(1,i-1)}, c_{(1,i)}, c_0) \times \]

\[ \prod_{j=1}^{i} dF_{C_j|A, M_{(1,j-1)}, C_{(0,j-1)}, C_0}(c_j|a_{2^{j-1}+1}, m_{(1,j-1)}, c_{(1,j-1)}, c_0) \]

The first term

\[ \int_{c_{(1,i)}} dF_{M_i|A, M_{(1,i-1)}, C_{(0,i)}, C_0}(m_i|a_{2^{i-1}+1}, m_{(1,i-1)}, c_{(1,i)}, c_0) \times \]

\[ \prod_{j=1}^{i} dF_{C_j|A, M_{(1,j-1)}, C_{(0,j-1)}, C_0}(c_j|a_{2^{j-1}+1}, m_{(1,j-1)}, c_{(1,j-1)}, c_0) \]

is exactly equal to the g-formula where \(M_i\) is treated as the outcome and \((A, M_{(1,i-1)})\) as time-varying exposures conditional on the baseline confounder \(C_0 = c_0\). The proofs of Lemmas 1 and 2 feature in Appendix 2. As a result, we could identify the generalized interventional mediation parameter in Theorem 1.

**Theorem 1.** Under the consistency assumption, Lemma 1 and Lemma 2, the
generalized interventional mediator parameter in Definition 3 can be nonparametrically identified as follows:

\[
\varphi_K(a_{(1,K)}) = \int_{c_0} \int_{m_{(1,K)}} E[Y(a_1, m_{(1,K)})|c_0] \prod_{i=1}^{K} dF_{\hat{G}_i(a_{(2^{i-1},2^i)})}(m_1|c_0) dF_{c_0}(c_0)
\]

\[
= \int_{c_0} \int_{m_{(1,K)}} \Gamma(c_0, a_1, m_{(1,K)}) \prod_{i=1}^{K} H_i(m_i, a_{(2^{i-1},2^i)}) dF_{c_0}(c_0).
\]

Appendix 3 provides the proof in detail, and Appendix 4 provides numerous examples for the general form of \(H_K\) and for the identified formula of \(\varphi_K\) with one confounder or time-varying confounders.

Before addressing estimation, we here compare our definition as well as the identification process with the methods of Daniel et al. and Lok (Daniel et al., 2015; Lok, 2016). Daniel et al. retain the standard PSE definitions but specify a particular model for identification (Daniel et al., 2015), in which the counterfactual values of mediators are independent for different exposure values. Such counterfactual model assumptions allow the fourth assumption for mediation identification to remain valid in the presence of mediator-outcome confounders, which are affected by exposure. Although the theoretical proof was lacking, Daniel et al. used sensitivity analysis to demonstrate that their model did not feature severe bias. Lok introduced “organic” direct and indirect effects, which can be defined and estimated without relying on setting the mediator to specific values (Lok, 2016). Lok’s method provided an alternative definition of mediation parameter. “Organic” interventions (I) are introduced, which can affect the distribution of the mediator instead of the cross-world counterfactual model. Although Lok’s article included only one mediator, this method can potentially extend to multiple mediators. Despite the difference between definitions and counterfactual assumptions, the mediation formulas of two methods are identical to our method if no time-varying confounders exist. In addition, our method requires weaker assumptions about the counterfactual model or the existence of “organic intervention,” allows for time-varying confounders, and can be extended to the generalized condition, as demonstrated.
This section provides two methods, the regression-based approach and the g-computation approach, for estimating $g$PSE. The regression-based approach assumes that variables of interest follow linear additive models with a Gaussian distribution assumption, and the corresponding estimator can be calculated explicitly as the function of parameters for linear models. Although the analytic solution of the regression-based approach performs more efficiently, the requirements of linear and Gaussian assumptions restrict its flexibility. By contrast, the g-computation approach relaxes the assumption for distribution, and hence all variables of interest in analysis can be categorical as well as continuous. The g-computation approach adopts Monte Carlo simulation to approximate the causal effects of interest, which is widely used in all fields of causal inference, such as time-varying causal effects (S. H. Lin, Young, Logan, & VanderWeele, 2017; T. J. VanderWeele & Tchetgen Tchetgen, 2017), mediation analysis (S.-H. Lin & VanderWeele, 2017), and four-way decomposition (T. J. VanderWeele, 2014; T. J. VanderWeele & Shrier, 2016). The details of the two approaches are described in the following two paragraphs.

4.1. Regression-based approach

We calculate the estimators of $g$PSE for four scenarios that are common in data analysis: (Scenario 1) linear model with baseline confounders and main effect; (Scenario 2) linear model with baseline and time-varying confounders and main effect; (Scenario 3) linear model with baseline confounders, main effect, and interaction; (Scenario 4) dichotomous outcome with rare disease assumption. In Scenarios 1, 3, and 4, three mediators are demonstrated, the causal relationship of all variables is shown in Figure 2. In Scenario 2, two mediators with time-varying confounders are demonstrated and the caused relationship is shown in Figure 3. The case with less mediators and all proof details are shown in Appendix 5. We illustrated our method under three mediators.

Scenario 1. No time-varying confounding or interaction.

We assume that $Y$ and $M$ are based on linear regression with no time-varying confounders, giving the following regression assumptions:

\[ Y = \theta_0^Y + \theta_1^Y A + \theta_2^Y M_1 + \theta_3^Y M_2 + \theta_4^Y C + \epsilon_Y, \text{ where } \epsilon_Y \sim N(0, \sigma_Y^2) \] (Y-1)

\[ M_1 = \theta_0^1 + \theta_1^1 A + \theta_2^1 C + \epsilon_1, \text{ where } \epsilon_1 \sim N(0, \sigma_1^2) \] (M-1.1)

\[ M_2 = \theta_0^2 + \theta_1^2 A + \theta_2^2 C + \theta_1^1 M_1 + \epsilon_2, \text{ where } \epsilon_2 \sim N(0, \sigma_2^2) \] (M-1.2)
On the basis of the aforementioned regression assumptions:

\[ M_3 = \theta_0^3 + \theta_a^3 A + \theta_c^3 C + \theta_1^3 M_1 + \theta_2^3 M_2 + \epsilon_3, \text{ where } \epsilon_3 \sim N(0, \sigma_3^2) \] (M-1.3)

We assume \( Y \) and \( M \) to be based on linear regression with time-varying confounding, giving the following regression assumptions:

\[ Y = \theta_0^Y + \theta_a^Y A + \theta_1^Y M_1 + \theta_2^Y M_2 + \theta_{c_2}^Y C_2 + \theta_{c_1}^Y C_1 + \theta_{c_0}^Y C_0 + \epsilon_y, \text{ where } \epsilon_y \sim N(0, \sigma_y^2) \] (Y-2)

\[ M_1 = \theta_0^1 + \theta_a^1 A + \theta_{c_1}^1 C_1 + \theta_{c_0}^1 C_0 + \epsilon_1, \text{ where } \epsilon_1 \sim N(0, \sigma_1^2) \] (M-2.1)

\[ M_2 = \theta_0^2 + \theta_a^2 A + \theta_{c_2}^2 C_2 + \theta_{c_1}^2 C_1 + \theta_{c_0}^2 C_0 + \epsilon_2, \text{ where } \epsilon_2 \sim N(0, \sigma_2^2) \] (M-2.2)

\[ C_1 = \theta_{c_1}^0 + \theta_a^{c_1} A + \theta_{c_0}^{c_1} C_0 + \epsilon_{c_1}, \text{ where } \epsilon_{c_1} \sim N(0, \sigma_{c_1}^2) \] (C-2.1)

\[ C_2 = \theta_{c_2}^0 + \theta_a^{c_2} A + \theta_{c_1}^{c_2} C_1 + \theta_{c_0}^{c_2} C_0 + \epsilon_{c_2}, \text{ where } \epsilon_{c_2} \sim N(0, \sigma_{c_2}^2) \] (C-2.2)

On the basis of Theorem 1 and the aforementioned models, all giPSEs can be derived.
\( \text{giPSE}_2 = (\text{giPSE}_2(1), ..., \text{giPSE}_2(2^2)) \)

\[
\text{giPSE}_2 = (\theta^y_a + \theta^y_c \theta^c_a + \theta^y_c \theta^c_1 \theta^c_a + \theta^y_c \theta^c_1, \\
\theta^y_1 \theta^1_a + \theta^y_1 \theta^1_c \theta^c_a + \theta^y_1 \theta^c_2 \theta^1_1 + \theta^y_1 \theta^c_2 \theta^1_1 \theta^c_1, \\
\theta^y_2 \theta^2_a + \theta^y_2 \theta^2_c \theta^c_2 + \theta^y_2 \theta^c_2 \theta^1_1 + \theta^y_2 \theta^c_2 \theta^1_1 \theta^c_1, \\
\theta^y_3 \theta^3_a + \theta^y_3 \theta^3_c \theta^c_3 + \theta^y_3 \theta^c_3 \theta^3_1 + \theta^y_3 \theta^c_3 \theta^3_1 \theta^c_1)
\]

If we set all confounders to be empty, the result is identical to the SEM result. Moreover, each giPSE, in this case, is the sum of the SEM estimators with respect to the specific path. For example, \( \text{giPSE}_2(4) \) estimated by \( \theta^y_1 \theta^1_a + \theta^y_2 \theta^2_1 \theta^1_1 + \theta^y_2 \theta^2_2 \theta^1_1 + \theta^y_2 \theta^2_3 \theta^1_1 \theta^c_1 \)

\( \theta^y_2 \theta^2_1 \theta^2_1 \theta^c_1 \theta^c_1 \) corresponds to four paths: \( A \rightarrow M_1 \rightarrow M_2 \rightarrow Y, \ A \rightarrow C_1 \rightarrow M_1 \rightarrow M_2 \rightarrow Y, \ A \rightarrow C_1 \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow Y. \)

\( \theta^1_1 \rightarrow M_2 \rightarrow Y, \ A \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow Y, \text{ and } A \rightarrow C_1 \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow Y. \)

**Scenario 3.** No time-varying confounding but numerous interactions for three mediators.

On the basis of Fig. 2, we assume \( Y \) and \( M \) are based on linear regression with no time-varying confounder but allowing exposure-mediator interactions, giving the following regression assumptions:

\[
Y = \theta^y_0 + \theta^y_a A + \theta^y_1 M_1 + \theta^y_2 M_2 + \theta^y_3 M_3 + \theta^y_4 C_0 + \theta^y_5 A M_1 + \theta^y_6 A M_2 + \theta^y_7 A M_3 + \varepsilon_y, \text{ where } \varepsilon_y \sim N(0, \sigma^2_y) \quad (Y-3)
\]

\[
\theta^y_8 A M_1 + \varepsilon_y, \text{ where } \varepsilon_y \sim N(0, \sigma^2_y) \quad (Y-3)
\]

\[
M_1 = \theta^1_0 + \theta^1_a A + \theta^1_1 C_0 + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0, \sigma^2_1) \quad (M-3.1)
\]

\[
M_2 = \theta^2_0 + \theta^2_a A + \theta^2_1 M_1 + \theta^2_2 C_0 + \theta^2_3 A M_1 + \varepsilon_2, \text{ where } \varepsilon_2 \sim N(0, \sigma^2_2) \quad (M-3.2)
\]
\[ M_3 = \theta_0^3 + \theta_a^3 A + \theta_1^3 M_1 + \theta_2^3 M_2 + \theta_c^3 C_0 + \theta_{a_1}^3 A M_1 + \theta_{a_2}^3 A M_2 + \varepsilon_3, \]

where \( \varepsilon_3 \sim N(0, \sigma_3^2) \) (M-3.3)

On the basis of Theorem 1 and the aforementioned models, all giPSEs can be derived as

\[ \text{giPSE}_3(1) = \theta_1^y + \theta_{a_1}^y (\theta_0^1 + \theta_1^1 E(c_0)) + \theta_{a_2}^y (\theta_0^2 + \theta_1^2 \theta_0^1 + \theta_1^2 \theta_2^1 E(c_0) + \theta_2^2 E(c_0)) \]

\[ + \theta_{a_3}^y (\theta_0^3 + \theta_1^3 \theta_0^1 + \theta_1^3 \theta_2^1 E(c_0) + \theta_2^3 \theta_0^1 + \theta_2^3 \theta_2^1 \theta_2^1 E(c_0)) \]

\[ + \theta_3^2 \theta_0^2 E(c_0) + \theta_2^3 E(c_0) \]

\[ \text{giPSE}_3(2) = (\theta_1^y + \theta_{a_1}^y) \theta_a^1 \]

\[ \text{giPSE}_3(3) = (\theta_2^y + \theta_{a_2}^y) (\theta_0^2 + \theta_{a_1}^2 (\theta_0^1 + \theta_2^1 E(c_0))) \]

\[ \text{giPSE}_3(4) = (\theta_2^y + \theta_{a_2}^y) (\theta_1^2 \theta_0^1 + \theta_{a_1}^2 \theta_a^1) \]

\[ \text{giPSE}_3(5) = (\theta_3^y + \theta_{a_3}^y) (\theta_0^3 + \theta_{a_1}^3 (\theta_0^1 + \theta_2^1 E(c_0))) \]

\[ + \theta_{a_2}^3 (\theta_0^2 + \theta_{a_1}^2 (\theta_0^1 + \theta_2^1 E(c_0)) + \theta_3^2 E(c_0)) \]

\[ \text{giPSE}_3(6) = (\theta_3^y + \theta_{a_3}^y) (\theta_0^3 + \theta_{a_1}^3 \theta_a^1) \]

\[ \text{giPSE}_3(7) = (\theta_3^y + \theta_{a_3}^y) ((\theta_2^3 + \theta_{a_2}^3) (\theta_0^2 + \theta_{a_1}^2 (\theta_0^1 + \theta_2^1 E(c_0)))) \]

\[ \text{giPSE}_3(8) = (\theta_3^y + \theta_{a_3}^y) (\theta_1^2 \theta_0^1 + \theta_{a_1}^2 \theta_a^1) \]

Similarly, if we set all interactions to be empty, the result is identical to the result of SEM, and giPSEs subject to the interaction assumption can also be represented as the sum of a generalized SEM estimator (Bollen, 1995; Marcoulides & Schumacker, 2013).

\[ A \rightarrow M_1 \rightarrow Y, \text{ and } \theta_{a_1}^y \theta_a^1 \text{ is the product estimator of the path } A \theta_{a_1}^y \theta_a^1 \rightarrow M_1 \theta_{a_2}^1 \rightarrow Y. \]
Scenario 4. Log-link and logistic regression with rare disease assumption

The causal relationships of these variables are seen in Fig. 1. In this scenario 4, $M_1$, $M_2$, and $M_3$ follow the linear regression models (M-1.1), (M-1.2), and (M-1.3), respectively, and outcome $Y$ is binary according to logistic regression as follows:

$$\logit(Y) = \theta^Y_0 + \theta^Y_A + \theta^Y_1 M_1 + \theta^Y_2 M_2 + \theta^Y_3 M_3 + \theta^Y_C + \theta^Y_{a_1} AM_1 + \varepsilon_Y,$$

where

$$\varepsilon_Y \sim \mathcal{N}(0, \sigma^2_Y). \text{ (Y-4)}$$

The closed forms of giPSEs are absent. Therefore, a rare disease assumption must be made empirically when $\Pr(Y = 1) \leq 0.1$, and the logit($Y$) is approximately $\log(Y)$.

Because $Y$ is binary, we specify the comparative function $Q(x_1, x_2)$ in Definition 4 of giPSEs as multiplicative scale, namely, $Q(x_1, x_2) = x_1/x_2$; on the basis of Theorem 1 and (Y-4), (M-1.1), (M-1.2), all giPSEs can be derived as

$$\text{giPSE}_3 = \exp(\theta^Y_A), \exp\left(\frac{\theta^Y + \theta^Y_{a_1}}{\theta^Y_A}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y_A}\right), \exp\left(\frac{\theta^Y}{\theta^Y_{a_1}}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y_{a_1}}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y_{a_1}}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y_{a_1}}\right), \exp\left(\frac{\theta^Y_{a_1}}{\theta^Y_{a_1}}\right)$$

In the absence of interaction, this giPSE$_3$ can be reduced to the exponential form of SEM estimators. Therefore, to this form, the hypothesis testing can still be conducted using a Sobel test or joint test, but the SEM estimation is invalid.

This demonstrates that, when outcome $Y$ is neither identical link nor log link, no closed form of giPSE exists. Furthermore, when outcome $Y$ is linear with many interactions and time-varying confounders, the closed form of the giPSE is complicated. Therefore, we subsequently demonstrate use of the g-computation approach to address this problem.

4.2. G-computation approach

Although the regression-based approach provides the analytic estimator, it obviously has the limitation of the model specification to acquire the analytic form for estimators. To this end, various studies have been published with a general algorithm, Monte Carlo simulation, or g-computation, to infer causal parameter (Imai et al., 2010; King, Tomz, & Wittenberg, 2000). This algorithm is not tied to specific statistical models, which accommodate linear and nonlinear relationships in parametric models (e.g., probit or logit) and various types of covariates (for related methods, see Glynn, 2008 (Glynn, 2008); Huang, Sivaganesan, Succop, & Goodman, 2004 (B. Huang,
For this section, we extend the previous algorithm on one mediator (Imai et al., 2010) to the settings with causally ordered multiple mediators to estimate giPSE. The algorithm consists of four steps and assumes an arbitrary number for $K$ as the number of mediators. First, the parametric models for mediators and outcome are established and the model fitted using typical methods such as the least squares approach. Next, the sequential Monte Carlo draws from each mediator $M_1, ..., M_K$ are obtained followed by the Monte Carlo draws of the potential outcome. Subsequently, the relevant quantities of all giPSEs are computed on the basis of the Monte Carlo draws of potential outcome. Finally, the summary of giPSEs is computed. This algorithm is described as follows:

**Step 1.** Construct parametric models for mediators and outcomes. Let $\Theta_i$ be a vector of parameters in the $i$-th mediator regression model $f_{M_i}$ and $\Theta_Y$ be the parameter vector of outcome model $f_Y$. Estimate these parameter as follows:

1a) Fit the mediator models, $f_{M_1}(M_1|A, C, \Theta_1)$, ..., $f_{M_K}(M_K|A, M_{(1,K-1)}, C, \Theta_K)$ separately, and compute the MLEs of $\Theta_i$ and $\Sigma_{\Theta_i}$ as $\hat{\Theta}_i$ and $\hat{\Sigma}_{\Theta_i}$, respectively, where $\Sigma_{\Theta_i}$ is the variance-covariance matrix of $\hat{\Theta}_i$.

1b) Fit the outcome model $f_Y(Y|A, M_{(1,K)}, C, \Theta_Y)$, and compute the MLEs of $\Theta_Y$ and $\Sigma_{\Theta_Y}$ as $\hat{\Theta}_Y$ and $\hat{\Sigma}_{\Theta_Y}$, where $\Sigma_{\Theta_Y}$ is the variance-covariance matrix of $\hat{\Theta}_Y$.

**Step 2.** Simulate model parameters from their sampling distribution (usually multivariate normal (MN) distribution):

2a) Sample $J$ copies of $\Theta_i$ from $\text{MN}(\hat{\Theta}_i, \hat{\Sigma}_{\Theta_i})$, denoted as $\Theta_i^{(j)}$, where $i = 1, ..., K$, and $j = 1, ..., J$. $J$ is set as 10,000 in this study by the heuristic rule.

2b) Sample $J$ copies of $\Theta_Y$ from $\text{MN}(\hat{\Theta}_Y, \hat{\Sigma}_{\Theta_Y})$, denoted as $\Theta_Y^{(j)}$, where $j = 1, ..., J$.

**Step 3.** Repeat the following three steps for $j = 1, ..., J$:

3a) Compute the potential values of the mediators:
\[ M_1^{(j)}(a_1) = g \left( E( M_1 | \hat{\theta}_1, A = a_1, C) \right), \ldots, \]

\[ M_K^{(j)} \left( a_{(1:K-1)} \right) = g \left( E \left( M_K | \hat{\theta}_K, A = a_1, M_1 = M_1^{(j)}(a_2), \ldots, M_K = M_K^{(j)} \left( a_{(2:K-1:2,K)} \right), C \right) \right) \]

for \( a_{(1:K-1)} \) as all possible combinations, where \( g(\cdot) \) is the link function.

(3b) Compute the potential outcomes given the simulated mediators in (3a):

\[ Y^{(j)} \left( a_{(1:K)} \right) = \]

\[ g \left( E \left( M_K | \hat{\theta}, A = a_1, M_1 = M_1^{(j)}(a_2), \ldots, M_K = M_K^{(j)} \left( a_{(2:K-1:2,K)} \right), C \right) \right), \]

(3c) Compute the causal mediation effect as follows:

\[ \varphi^{(j)} \left( a_{(1:K)} \right) = \frac{1}{n} \sum Y^{(j)} \left( a_{(1:K)} \right), \]

where \( n \) is the sample size. Consequently, \( \text{gIPSE}^{(j)}(d) \) can be derived using

\[ \varphi^{(j)} \left( a_{(1:K)} \right). \]

Step 4. Use \( \text{gIPSE}^{(1)}(d), \ldots, \text{gIPSE}^{(J)}(d) \) to construct summary statistics such as point estimates, standard error, and confidence intervals.

The generality of this algorithm can be applied to any parametric statistical model. We have also developed a user-friendly R package which can be freely downloaded on the following website: http://shenglin.blog.nctu.edu.tw/methodology/.
5. NUMERICAL STUDIES

5.1. Simulation study

In this section, we report on a simulation study of the performance of g-computation, linear SEM, and analytic regression-based estimation for several settings with finite sample sizes of 1000 and 5000. Because no analytic solution exists to the estimation in the nonlinear model, we construct linear models with the interaction between exposure and mediators for four mediators and an outcome to illustrate the performance of our method. Consequently, one baseline continuous confounder $C_0$, one binary exposure $A$, four continuous mediators $M_1, M_2, M_3, M_4$, and a continuous outcome $Y$ are generated as follows.

$$C_0 \sim N(0,1), \ A \sim \text{ber}(0.5)$$

$$M_1 = 1 + A + C_0 + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0,1)$$

$$M_2 = 1 + A + M_1 + \theta_{a_1}^2 AM_1 + C_0 + \varepsilon_2, \text{ where } \varepsilon_2 \sim N(0,1)$$

$$M_3 = 1 + A + M_1 + M_2 + \theta_{a_1}^3 AM_1 + \theta_{a_2}^3 AM_2 + C_0 + \varepsilon_3, \text{ where } \varepsilon_3 \sim N(0,1)$$

$$M_4 = 1 + A + M_1 + M_2 + M_3 + \theta_{a_1}^4 AM_1 + \theta_{a_2}^4 AM_2 + \theta_{a_3}^4 AM_3 + C_0 + \varepsilon_4, \text{ where } \varepsilon_4 \sim N(0,1)$$

$$X = 1 + A + M_1 + M_2 + M_3 + M_4 + \theta_{b_1}^2 Y AM_1 + \theta_{b_2}^2 Y AM_2 + \theta_{b_3}^2 Y AM_3 + \theta_{a_4}^2 Y AM_4 + C_0 + \varepsilon_y, \text{ where } \varepsilon_y \sim N(0,1)$$

On the basis of these models, we discuss two situations to evaluate the characteristics of three methods under model misspecification, as follows:

Case I. Without any interaction effects: All coefficients for interactions $\theta_{a_1}^2, \theta_{a_2}^2, \theta_{a_1}^3, \theta_{a_2}^3, \theta_{a_1}^4, \theta_{a_2}^4, \theta_{a_1}^y, \theta_{a_2}^y, \theta_{a_3}^y, \theta_{a_4}^y, \theta_{a_1}^y, \theta_{a_2}^y, \theta_{a_3}^y, \theta_{a_4}^y$ were set to 0. In this case, linear SEM is equal to the regression-based approach.

Case II. With interaction effects: All coefficients for interactions $\theta_{a_1}^2, \theta_{a_2}^2, \theta_{a_1}^3, \theta_{a_2}^3, \theta_{a_1}^4, \theta_{a_2}^4, \theta_{a_1}^y, \theta_{a_2}^y, \theta_{a_3}^y, \theta_{a_4}^y, \theta_{a_1}^y, \theta_{a_2}^y, \theta_{a_3}^y, \theta_{a_4}^y$ were set to 2.

The g-computation approach is used with Monte Carlo size 1000 and bootstrap resampling size 1000, and the linear SEM model is fit without interaction for
comparison. We replicate the respective simulation procedure 1000 times and calculate the biases, empirical standard error (ESE), mean square error (MSE) estimated standard error (SSE), and coverage rate (COV) for the g-computation method and analytic regression-based estimation. The linear SEM adopts a delta method for the standard error estimation.

Table 1 demonstrates that in Case I, if no interaction occurs in models, then the three estimations are unbiased, and ESE is approximately equal to SSE for the three methods. But g-computation has larger variations for estimations of \( \text{giPSE}_4(1) \) and \( \text{giPSE}_4(9) \). Table 2 demonstrates that in Case II, the ESE is approximately equal to SSE for three estimations, but linear SEM is biased when interaction terms exist in models, the other two estimations are unbiased. Although g-computation has larger standard error and more computation time, it can be used when the PSE has no closed form or more complicated models. The closed form SEM method performs well in all respects, but the critical defect is that it requires a closed form. Both results are represented by boxplots in Figs. 4 and 5.

5.2. Application

From seven townships of Taiwan, 23,820 participants aged 30 to 65 years were recruited from 1991 to 2008. Hepatitis C virus (HCV) and hepatitis B virus (HBV) infection status and clinical data such as glutamate-pyruvate transaminase (GPT), for liver function, and ultrasound images were measured at the baseline and followed up every few years; hepatocellular carcinoma (HCC) status and death were confirmed through computerized data linkage with national cancer registry and death certification systems, respectively. The details of the participant enrollment have been described in other studies (C.-J. Chen et al., 2006; C. L. Chen et al., 2008; Iloeje et al., 2007).

We applied our method to a Taiwanese cohort dataset to investigate the mechanism of HCV infection on mortality. Three mediators (HBV infectious status, GPT and HCC) were considered in this mechanism. Age, gender, and smoking status were included as baseline confounders. The causal diagram (Y.-T. Huang et al., 2011) featuring in Fig. 6 is based on other articles from the literature. As a result, the effects can be decomposed into eight paths including: (1) not through the change of mediators (giPSE_3(1)); (2) through HBV infection change (giPSE_3(2)); (3) through GPT change only (giPSE_3(3)); (4) through HBV infection and then GPT change (giPSE_3(4)); (5) through HCC change only (giPSE_3(5)); (6) through HBV and then HCC change (giPSE_3(6)); (7) through GPT and then HCC change (giPSE_3(7)); (8) through HBV change that further influences the GPT and then HCC change (giPSE_3(8)). Decomposition of the overall effect into eight PSEs can facilitate understanding of the role of HBV infection, liver function (GPT),
and HCC, which allows prevention of the mechanism between hepatitis C and death.

The results in Table 4 provide the TE estimation and all eight pathways in scale of risk difference. TE demonstrated significant effects of 0.090 during 17 years’ follow-up (95% confidence interval [CI]: 0.070–0.109 ; P-value < 0.001). The effect of HCV through no mediator leading to death was 0.042 (95% CI: 0.023–0.061; P-value < 0.001) (giPSE_3(1)). The effect through GPT was 0.005 (95% CI: 0.002–0.008; P-value = 0.001) (giPSE_3(3)). The effect through HCC was 0.036 (95% CI: 0.029–0.044; P-value < 0.001) (giPSE_3(5)). The effect through both GPT and HCC was 0.006 (95% CI: 0.003–0.009; P-value < 0.001) (giPSE_3(7)). GPT and HCC can be seen to significantly mediate the causal mechanism of HCV on death, and their effects account for 5.7% and 40.8% of the overall effects, respectively. The path through GPT and HCC accounts for 6.8% of the total effect. However, the effect not explained by any mediators still accounts for 46.2%. This might be due to unmeasured adverse effects by anti-HCV drugs, acute and fatal inflammation, or the chronic consequences of inflammation such as cirrhosis. This merits further investigation. Although reported as a potential mediator by other studies, HBV did not contribute significantly to the mechanism in this study.
This work has three major contributions: First, we have built a general approach (including notation, definition, and estimation) for causal mediation analysis with arbitrary number of ordered multiple mediators, while previous literature restrict to only two or three mediators. Second, we demonstrate our method is a general form of previous mediation analysis. It is reduced to traditional structural equation model under linear or log link model, to causal mediation analysis with one mediator, and to the two mediation method proposed by Daniel et al at Biometrics under two mediator scenario. Third, a flexible algorithm built based on g-computation algorithm is proposed along with a user-friendly software online.

Several limitations merit attention, and some should be addressed in subsequent studies. First, unmeasured confounding assumptions are difficult to achieve when the covariates are not collected comprehensively. Sensitivity analysis is required when a set of confounders are known in other literature but not collected in a study. Second, the definitions of total effect and PSE in this study deviate from the traditional ones. Although two are identical in linear models, a bias formula should be developed for a general case when the link function of dependent variable is not an identity link. Third, both mediators and exposure in this method are restricted to one measurement at a fixed time point. However, most variables are measured repeatedly for time-to-event data such as liver function (GPT) and HCC status. Focusing on only one measurement and failing to use other measurements or time-to-event information is inefficient. Therefore, extending this method to time-varying settings and allowing time-to-event variables would be worthwhile. Finally, power decreases as the number of PSEs increases, and mediators do not contribute to mechanisms significantly. A criterion for path selection or mediator selection is required.

In conclusion, this method provides a generalized formula for multiple mediation analysis. It extends traditional linear SEM to settings with both continuous and categorical variables and also extends causal mediation analysis to settings when mediators mutually affect each other. Software based on a g-computation algorithm allows users to easily apply this method for data analysis subject to various model choices according to the substantive knowledge for each variable. All methods and software proposed in this study contribute to disentangling causal mechanisms when the natural pathways are complicated.
ACKNOWLEDGEMENT

We thank Professor Hwai-I Yang (Genomics Research Center, Academia Sinica, Taipei, Taiwan) for providing the HCC data set in this article. We also thank Shih-Wen Lin, Ying-Wen Liang, and Dr. An-Shun Tai for valuable comments on writing. This manuscript was edited by Wallace Academic Editing.
REFERENCES


Lok, J. J. (2016). Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. *Statistics in medicine*, 35(22), 4008-4020. doi:10.1002/sim.6990


VanderWeele, T. J. (2014). A unification of mediation and interaction: a four-way


Fig. 1. Causal relationships among these variables is demonstrated by a direct acyclic graph. $A, (M_1, M_2, ..., M_K), Y, C_0$ and $(C_1, C_2, C_3, ..., C_K)$ denote the exposure, the mediators, the outcome, the baseline confounders, and the time-varying confounders, respectively.

\[ C \rightarrow A \rightarrow C_1 \rightarrow M_1 \rightarrow \cdots \rightarrow C_K \rightarrow M_K \rightarrow Y \]

Fig. 2. Causal relationships among the exposure, the three causally ordered multiple mediators, the outcome variable, and the baseline confounder, which are denoted by $A$, $(M_1, M_2, M_3)$, $Y$, and $C_0$, respectively.

\[ C_0 \rightarrow A \rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow Y \]

Fig. 3. Causal relationships among the exposure, two causally ordered multiple mediators, the outcome variable, the baseline confounder, and the time-varying confounders, which are denoted by $A$, $(M_1, M_2)$, $Y$, $C_0$, and $(C_1, C_2)$, respectively.

\[ C_0 \rightarrow A \rightarrow C_1 \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow Y \]
Fig. 4. *Case I*: Boxplot of simulated values for g-computation, linear SEM, and analytic regression–based estimation with no interaction term. The red points present the true PSE values, and the white points present the means of the three methods’ simulated values.
Fig 5. Case II: Boxplot of simulated values for g-computation, linear SEM, and analytic regression–based estimation with interaction terms. The red points present the true value of PSE, and the white points present the mean of simulated values for the three methods.

Fig. 6. Causal diagram of HCC study.
Table 1. Case I: Simulation study without interaction term for three methods.

<table>
<thead>
<tr>
<th>True value</th>
<th>G-computation</th>
<th>Linear SEM</th>
<th>Analytic regression-based estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias</td>
<td>ESE</td>
<td>SSE</td>
</tr>
<tr>
<td>giPSE_4(1)</td>
<td>1</td>
<td>-0.002</td>
<td>0.093</td>
</tr>
<tr>
<td>giPSE_4(2)</td>
<td>1</td>
<td>-0.004</td>
<td>0.108</td>
</tr>
<tr>
<td>giPSE_4(3)</td>
<td>1</td>
<td>-0.007</td>
<td>0.102</td>
</tr>
<tr>
<td>giPSE_4(4)</td>
<td>1</td>
<td>-0.004</td>
<td>0.104</td>
</tr>
<tr>
<td>giPSE_4(5)</td>
<td>1</td>
<td>-0.001</td>
<td>0.106</td>
</tr>
<tr>
<td>giPSE_4(6)</td>
<td>1</td>
<td>-0.002</td>
<td>0.110</td>
</tr>
<tr>
<td>giPSE_4(7)</td>
<td>1</td>
<td>-0.004</td>
<td>0.102</td>
</tr>
<tr>
<td>giPSE_4(8)</td>
<td>1</td>
<td>-0.002</td>
<td>0.102</td>
</tr>
<tr>
<td>giPSE_4(9)</td>
<td>1</td>
<td>0.002</td>
<td>0.097</td>
</tr>
<tr>
<td>giPSE_4(10)</td>
<td>1</td>
<td>-0.003</td>
<td>0.107</td>
</tr>
<tr>
<td>giPSE_4(11)</td>
<td>1</td>
<td>-0.004</td>
<td>0.106</td>
</tr>
<tr>
<td>giPSE_4(12)</td>
<td>1</td>
<td>-0.002</td>
<td>0.105</td>
</tr>
<tr>
<td>giPSE_4(13)</td>
<td>1</td>
<td>0.004</td>
<td>0.098</td>
</tr>
<tr>
<td>giPSE_4(14)</td>
<td>1</td>
<td>0.003</td>
<td>0.107</td>
</tr>
<tr>
<td>giPSE_4(15)</td>
<td>1</td>
<td>0.002</td>
<td>0.104</td>
</tr>
<tr>
<td>giPSE_4(16)</td>
<td>1</td>
<td>0.003</td>
<td>0.100</td>
</tr>
</tbody>
</table>
Table 2. *Case II*: Simulation study for interaction term.

<table>
<thead>
<tr>
<th>True value</th>
<th>G-computation</th>
<th>Linear SEM</th>
<th>Analytic regression-based estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bias</td>
<td>ESE</td>
<td>SSE</td>
</tr>
<tr>
<td>giPSE4(1)</td>
<td>31</td>
<td>0.024</td>
<td>1.912</td>
</tr>
<tr>
<td>giPSE4(2)</td>
<td>3</td>
<td>-0.010</td>
<td>0.278</td>
</tr>
<tr>
<td>giPSE4(3)</td>
<td>9</td>
<td>0.014</td>
<td>0.711</td>
</tr>
<tr>
<td>giPSE4(4)</td>
<td>9</td>
<td>-0.022</td>
<td>0.777</td>
</tr>
<tr>
<td>giPSE4(5)</td>
<td>21</td>
<td>0.006</td>
<td>1.538</td>
</tr>
<tr>
<td>giPSE4(6)</td>
<td>9</td>
<td>-0.037</td>
<td>0.788</td>
</tr>
<tr>
<td>giPSE4(7)</td>
<td>27</td>
<td>-0.011</td>
<td>1.899</td>
</tr>
<tr>
<td>giPSE4(8)</td>
<td>27</td>
<td>-0.121</td>
<td>2.120</td>
</tr>
<tr>
<td>giPSE4(9)</td>
<td>45</td>
<td>0.062</td>
<td>2.767</td>
</tr>
<tr>
<td>giPSE4(10)</td>
<td>9</td>
<td>-0.015</td>
<td>0.783</td>
</tr>
<tr>
<td>giPSE4(11)</td>
<td>27</td>
<td>0.032</td>
<td>1.867</td>
</tr>
<tr>
<td>giPSE4(12)</td>
<td>27</td>
<td>-0.071</td>
<td>2.180</td>
</tr>
<tr>
<td>giPSE4(13)</td>
<td>63</td>
<td>0.036</td>
<td>3.778</td>
</tr>
<tr>
<td>giPSE4(14)</td>
<td>27</td>
<td>-0.097</td>
<td>2.166</td>
</tr>
<tr>
<td>giPSE4(15)</td>
<td>81</td>
<td>-0.011</td>
<td>4.588</td>
</tr>
<tr>
<td>giPSE4(16)</td>
<td>81</td>
<td>-0.319</td>
<td>5.697</td>
</tr>
</tbody>
</table>

- **giPSE**: Generalized Predictive Standard Errors
- **ESE**: Empirical Standard Error
- **SSE**: Standardized Standard Error
- **MSE**: Mean Square Error
- **COV**: Coefficient of Variation
Table 3. Estimation of all giPSEs and giTEs.

<table>
<thead>
<tr>
<th>Mediators involved</th>
<th>Estimate (proportion of mediator)</th>
<th>SE</th>
<th>95% CI lower</th>
<th>95% CI upper</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>giPSE₃(1) Null</td>
<td>0.042 (46.2%)</td>
<td>0.010</td>
<td>0.023</td>
<td>0.061</td>
<td>0.000*</td>
</tr>
<tr>
<td>giPSE₃(2) M₁</td>
<td>0.000 (0%)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.811</td>
</tr>
<tr>
<td>giPSE₃(3) M₂</td>
<td>0.005 (5.7%)</td>
<td>0.002</td>
<td>0.002</td>
<td>0.008</td>
<td>0.001*</td>
</tr>
<tr>
<td>giPSE₃(4) M₁, M₂</td>
<td>0.000 (0%)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.832</td>
</tr>
<tr>
<td>giPSE₃(5) M₃</td>
<td>0.036 (40.8%)</td>
<td>0.004</td>
<td>0.029</td>
<td>0.044</td>
<td>0.000*</td>
</tr>
<tr>
<td>giPSE₃(6) M₁, M₃</td>
<td>0.000 (0.1%)</td>
<td>0.002</td>
<td>-0.004</td>
<td>0.004</td>
<td>0.992</td>
</tr>
<tr>
<td>giPSE₃(7) M₂, M₃</td>
<td>0.006 (6.8%)</td>
<td>0.002</td>
<td>0.003</td>
<td>0.009</td>
<td>0.000*</td>
</tr>
<tr>
<td>giPSE₃(8) M₁, M₂, M₃</td>
<td>0.000 (0.4%)</td>
<td>0.001</td>
<td>-0.002</td>
<td>0.003</td>
<td>0.797</td>
</tr>
<tr>
<td>giTE₃</td>
<td>0.090</td>
<td>0.010</td>
<td>0.070</td>
<td>0.109</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

SE: standard error; CI: confidence interval
* significance level = 0.05
Appendix

Appendix 1: Examples of section 2

Figure S1. One mediator

\[ C_0 \rightarrow A \rightarrow C_1 \rightarrow M_1 \rightarrow Y \]

Figure S2. Two mediators

\[ C_0 \rightarrow A \rightarrow C_1 \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow Y \]

Figure S3. Three mediators

\[ C_0 \rightarrow A \rightarrow C_1 \rightarrow M_1 \rightarrow C_2 \rightarrow M_2 \rightarrow C_3 \rightarrow M_3 \rightarrow Y \]
1.1 For one mediator (K=1, Figure S1)

Definition 1
\[
d = I(M_1) \times 2^0 + 1 = \begin{cases} 
1, & \text{if } I(M_1) = 0 \\
2, & \text{if } I(M_1) = 1 
\end{cases}
\]

The notation \( I(M_1) \) represents whether through \( M_1 \) or not.

Definition 2
\( G_1(a_1) \) is a random draw of \( M_1(a_1) \), which is the counterfactual value of \( M_1 \) given \( A \) is set to \( a_1 \).

Definition 3
\[
\phi_1(a_{(1,2)}) = E[Y(a_1, G_1(a_2))]
\]

Define \( \phi_1 \) the expectation of a counterfactual outcome given \( (A, M_1) \) is set to \( (a_1, G_1(a_2)) \).

Definition 4
\[
giPSE(1) = G(\phi_1(a_1 = a^*_1, a_2), \phi_1(a_1 = a^*_0, a_2))
\]
\[
giPSE(2) = G(\phi_1(a_1, a_2 = a^*_1), \phi_1(a_1, a_2 = a^*_0))
\]

It is the change of \( \phi_1 \) by changing the value of \( a_d \) to \( a^*_1 \) and \( a^*_0 \) and keeping all other \( a \) as the same value \( (a_1, a_2, ..., a_{d-1}, a_{d+1}, ..., a_K) \).

Definition 5
\[
giPSE_1(1) = \phi_1(1,0) - \phi_1(0,0)
\]
\[
giPSE_1(2) = \phi_1(1,1) - \phi_1(1,0)
\]
\[
giTE_1 = giPSE_1(1) + giPSE_1(2)
\]

Define \( giPSE \) for one mediator (K=1).
1.2 For two mediators (K=2, Figure S2)

Definition 1

\[ d = I(M_1) \times 2^0 + I(M_2) \times 2 + 1 = \begin{cases} 
1 & \text{if} \ (I(M_2), I(M_1)) = (0,0) \\
2 & \text{if} \ (I(M_2), I(M_1)) = (0,1) \\
3 & \text{if} \ (I(M_2), I(M_1)) = (1,0) \\
4 & \text{if} \ (I(M_2), I(M_1)) = (1,1) 
\end{cases} \]

The notation \((I(M_2), I(M_1))\) represents whether through \(M_1\) and \(M_2\) or not.

Definition 2

\(G_2(a_1, a_2)\) is a random draw of \(M_2(a_1, G_1(a_2))\), which is the counterfactual value of \(M_2\) given \((A, M_1)\) is set to \((a_1, G_1(a_2))\).

Definition 3

\[ \varphi_2(a_{(1,4)}) = E[Y(a_1, G_1(a_2), G_2(a_3, a_4))] \]

Define \(\varphi_2\) the expectation of a counterfactual outcome given \((A, M_{(1,2)})\) is set to \((a_1, G_1(a_2), G_2(a_3, a_4))\).

Definition 4

\[ \text{gIPSE}(1) = G(\varphi_2(a_1 = a^*_1, a_2, a_3, a_4), \varphi_2(a_1 = a^*_0, a_2, a_3, a_4)) \]
\[ \text{gIPSE}(2) = G(\varphi_2(a_1, a_2 = a^*_1, a_3, a_4), \varphi_2(a_1, a_2 = a^*_0, a_3, a_4)) \]
\[ \text{gIPSE}(3) = G(\varphi_2(a_1, a_2, a_3 = a^*_1, a_4), \varphi_2(a_1, a_2, a_3 = a^*_0, a_4)) \]
\[ \text{gIPSE}(4) = G(\varphi_2(a_1, a_2, a_3, a_4 = a^*_1), \varphi_2(a_1, a_2, a_3, a_4 = a^*_0)) \]

It is the change of \(\varphi_2\) by changing the value of \(a_d\) to \(a^*_1\) and \(a^*_0\) and keeping all other \(a\) as the same value \((a_1, a_2, ..., a_{d-1}, a_{d+1}, ..., a_{2k})\).

Definition 5

\[ \text{gIPSE}_2(1) = \varphi_2(1,0,0,0) - \varphi_2(0,0,0,0) \]
\[ \text{giPSE}_2(2) = \varphi_2(1,1,0,0) - \varphi_2(1,0,0,0) \]
\[ \text{giPSE}_2(3) = \varphi_2(1,1,1,0) - \varphi_2(1,1,0,0) \]
\[ \text{giPSE}_2(4) = \varphi_2(1,1,1,1) - \varphi_2(1,1,1,0) \]

\[ \text{giTE}_2 = \text{giPSE}_2(1) + \text{giPSE}_2(2) + \text{giPSE}_2(3) + \text{giPSE}_2(4) \]

Define giPSE for two mediators (K=2).

### 1.3 For three mediators (K=3, Figure S3)

**Definition 1**

\[ d = I(M_1) \times 2^0 + I(M_2) \times 2 + I(M_3) \times 4 + 1 = \begin{cases} 
1 & \text{if } (I(M_3), I(M_2), I(M_1)) = (0,0,0) \\
2 & \text{if } (I(M_3), I(M_2), I(M_1)) = (0,0,1) \\
3 & \text{if } (I(M_3), I(M_2), I(M_1)) = (0,1,0) \\
4 & \text{if } (I(M_3), I(M_2), I(M_1)) = (0,1,1) \\
5 & \text{if } (I(M_3), I(M_2), I(M_1)) = (1,0,0) \\
6 & \text{if } (I(M_3), I(M_2), I(M_1)) = (1,0,1) \\
7 & \text{if } (I(M_3), I(M_2), I(M_1)) = (1,1,0) \\
8 & \text{if } (I(M_3), I(M_2), I(M_1)) = (1,1,1) 
\end{cases} \]

The notation \((I(M_3), I(M_2), I(M_1))\) represents whether through \(M_1, M_2\) and \(M_3\) or not.

**Definition 2**

\(G_3(a_1, a_2, a_3, a_4)\) is a random draw of \(M_3(a_1, G_1(a_2), G_2(a_3, a_4))\), which is the counterfactual value of \(M_3\) given \((A, M_1, M_2)\) is set to \((a_1, G_1(a_2), G_2(a_3, a_4))\).

**Definition 3**

\[ \varphi_3(a_{(1,8)}) = E[Y(a_1, G_1(a_2), G_2(a_3, a_4), G_3(a_5, a_6, a_7, a_8))] \]

Define \(\varphi_3\) the expectation of a counterfactual outcome given \((A, M_{(1,3)})\) is set to \((a_1, G_1(a_2), G_2(a_3, a_4), G_3(a_5, a_6, a_7, a_8))\).

**Definition 4**
giPSE(1) = G(\varphi_3(a_1 = a^{*}_{(1)}, a_2, ..., a_8), \varphi_3(a_1 = a^{*}_{(0)}, a_2, ..., a_8))

giPSE(2) = G(\varphi_3(a_1, a_2 = a^{*}_{(1)}, a_3, ..., a_8), \varphi_3(a_1, a_2 = a^{*}_{(0)}, a_3, ..., a_8))

giPSE(3) = G(\varphi_3(a_2, a_3 = a^{*}_{(1)}, a_4, ..., a_8), \varphi_3(a_1, a_2, a_3 = a^{*}_{(0)}, a_4, ..., a_8))

giPSE(4) = G(\varphi_3(a_1, a_2, a_3, a_4 = a^{*}_{(1)}, a_5, ..., a_8), \varphi_3(a_1, a_2, a_3, a_4 = a^{*}_{(0)}, a_5, ..., a_8))

giPSE(5) = G(\varphi_3(a_1, ..., a_4, a_5 = a^{*}_{(1)}, a_6, a_7, a_8), \varphi_3(a_1, ..., a_4, a_5 = a^{*}_{(0)}, a_6, a_7, a_8))

giPSE(6) = G(\varphi_3(a_1, ..., a_5, a_6 = a^{*}_{(1)}, a_7, a_8), \varphi_3(a_1, ..., a_5, a_6 = a^{*}_{(0)}, a_7, a_8))

giPSE(7) = G(\varphi_3(a_1, ..., a_6, a_7 = a^{*}_{(1)}, a_8), \varphi_3(a_1, ..., a_6, a_7 = a^{*}_{(0)}, a_8))

giPSE(8) = G(\varphi_3(a_1, ..., a_7, a_8 = a^{*}_{(1)}), \varphi_3(a_1, ..., a_7, a_8 = a^{*}_{(0)}))

It is the change of \varphi_3 by changing the value of \( a_d \) to \( a^{*}_{(0)} \) and \( a^{*}_{(1)} \) and keeping all other a as the same value \( (a_1, a_2, ..., a_{d-1}, a_{d+1}, ..., a_k) \).

**Definition 5**

\[
\text{giPSE}_3(1) = \varphi_3(1,0,0,0,0,0,0,0) - \varphi_3(0,0,0,0,0,0,0,0) \\
\text{giPSE}_3(2) = \varphi_3(1,1,0,0,0,0,0,0) - \varphi_3(1,0,0,0,0,0,0,0) \\
\text{giPSE}_3(3) = \varphi_3(1,1,1,0,0,0,0,0) - \varphi_3(1,1,0,0,0,0,0,0) \\
\text{giPSE}_3(4) = \varphi_3(1,1,1,1,0,0,0,0) - \varphi_3(1,1,1,0,0,0,0,0) \\
\text{giPSE}_3(5) = \varphi_3(1,1,1,1,1,0,0,0) - \varphi_3(1,1,1,1,0,0,0,0) \\
\text{giPSE}_3(6) = \varphi_3(1,1,1,1,1,1,0,0) - \varphi_3(1,1,1,1,1,0,0,0) \\
\text{giPSE}_3(7) = \varphi_3(1,1,1,1,1,1,1,0) - \varphi_3(1,1,1,1,1,1,0,0) \\
\text{giPSE}_3(8) = \varphi_3(1,1,1,1,1,1,1,1) - \varphi_3(1,1,1,1,1,1,1,0)
\]

\( \text{giTE}_3 = \text{giPSE}_3(1) + \text{giPSE}_3(2) + \text{giPSE}_3(3) + \text{giPSE}_3(4) + \text{giPSE}_3(5) + \text{giPSE}_3(6) + \text{giPSE}_3(7) + \text{giPSE}_3(8) \)

Define giPSE for three mediators (K=3).
Appendix 2: Proof of Lemmas 1 and 2

The proof of Lemma 1 is conducted as follows:

First, we have

\[
E[Y(a_1, m_{(1,K)}) | c_0] = E[Y(a_1, m_{(1,K)}) | a_1, c_0]
\]

\[
= \int_{c_1} E[Y(a_1, m_{(1,K)}) | a_1, c_{(0,1)}] dF_{c_1 | A, C_0}(c_1 | a_1, c_0)
\]

\[
= \int_{c_1} E[Y(a_1, m_{(1,K)}) | a_1, c_{(0,1)}, M_1 = m_1] dF_{c_1 | A, C_0}(c_1 | a_1, c_0) \text{ ... (s2.1)}
\]

where the first equality follows by Assumption 1 \( (Y(a_1, m_{(1,K)}) \perp A | C_0) \), the second by the law of total expectation, and the last by Assumption 2 \( (Y(a, m_{(1,K)}) \perp M_1 | C_{(0,1)}, A) \).

Next, we identify the general form of the expectation in (s2.1) by \( E[Y(a_1, m_{(1,K)}) | a_1, c_{(0,i)}, M_{(1,i)} = m_{(1,i)}] \)

\[
= \int_{c_{i+1}} E[Y(a_1, m_{(1,K)}) | a_1, c_{(0,i+1)}, M_{(1,i)} = m_{(1,i)}] \times
\]

\[
dF_{c_{i+1} | A, C_{(0,i)}, M_{(1,i)}}(c_{i+1} | a_1, c_{(0,i)}, M_{(1,i)} = m_{(1,i)})
\]

\[
= \int_{c_{i+1}} E[Y(a_1, m_{(1,K)}) | A, C_{(0,i+1)}, M_{(1,i+1)} = m_{(1,i+1)}] \times
\]

\[
dF_{c_{i+1} | A, C_{(0,i)}, M_{(1,i)}}(c_{i+1} | a_1, c_{(0,i)}, M_{(1,i)} = m_{(1,i)}) \text{ ... (s2.2)}
\]

where the first equality follows by the law of total expectation with respect to \( c_{i+1} \), and the last by Assumption 2 \( (Y(a, m_{(1,K)}) \perp M_{i+1} | C_{(0,1)}, A, M_{(1,i)}) \). Equation (s2.2) shows that the expectation in (s2.1) can be iteratively identify. Therefore, by the consistency assumption, the conditional expectation in Lemma 1 subsequently be identified as

\[
E[Y(a_1, m_{(1,K)}) | c_0] = \int_{c_{(1:K)}} E[Y(a_1, m_{(1,K)}) | a_1, c_{(0,K)}, M_{(1,K)} = m_{(1,K)}] \times
\]

\[
\prod_{i=1}^{K} dF_{c_i | A, C_{(0,i-1)}, M_{(1,i-1)}}(c_i | a_1, c_{(0,i-1)}, M_{(1,i-1)} = m_{(1,i-1)})
\]

\[
= \int_{c_{(1:K)}} E[Y(a_1, c_{(0,K)}, M_{(1,K)} = m_{(1,K)})] \prod_{i=1}^{K} dF_{c_i | A, C_{(0,i-1)}, M_{(1,i-1)}}(c_i | a_1, c_{(0,i-1)}, M_{(1,i-1)} = m_{(1,i-1)}).
\]

37
The proof of Lemma 2 (Recursive formula) is shown by induction as follows:

For \( i = 1 \), it is
\[
dF_{G_i(a_1)}|_{c_0}(m_1 | c_0) = dF_{G_i(a_1)}|_{A, c_0}(m_1 | a_1, c_0)
\]
\[
= \int_{c_1} dF_{G_i(a_1)}|_{A, (c_0, 1)}(m_1 | a_1, c_{(0, 1)}) dF_{C_i}|_{A, c_0}(c_1 | a_1, c_0)
\]
\[
= \int_{c_1} dF_{M_i(a_1)}|_{A, (c_0, 1)}(m_1 | a_1, c_{(0, 1)}) dF_{C_i}|_{A, c_0}(c_1 | a_1, c_0)
\]
\[
= \int_{c_1} dF_{M_i(a_1)}|_{A, c_0}(m_1 | a_1, c_{(0, 1)}) dF_{C_i}|_{A, c_0}(c_1 | a_1, c_0)
\]
where the first equality follows by the independency of \( G_1(a_1) \), the second by the law of total expectation, the third by the identical distribution of \( M_1 \) and \( G_1 \), and the fourth by the consistency assumption.

Next, by induction, we assume \( dF_{G_i(a_1)}|_{c_0}(m_1 | c_0) \), \( \ldots, dF_{G_{i-1}(a_{(2i-2,1,2i-1)})}|_{c_0}(m_{i-1} | c_0) \)

have been identified as \( H_1(m_1, a_1, c_0), \ldots, H_{i-1}(m_{i-1}, a_{(2i-2,1,2i-1)}, c_0) \), and then it now need to identify \( dF_{G_i(a_{(2i-1,1,2i)})}|_{c_0}(m_i | c_0) \). The proof is shown as below:

\[
dF_{G_i(a_{(2i-1,1,2i)})}|_{c_0}(m_i | c_0)
\]
\[
= dF_{G_i(a_{(2i-1,1,2i)})}|_{A, c_0}(m_i | a_{2^{-i}+1}, c_0)
\]
\[\because G_i \text{ depends to all variables}\]
\[
= dF_{M_i\left(a_{2^{-i}+1}, G_1(a_{2^{-i}+2}), G_2(a_{2^{-i}+3}a_{2^{-i}+4}) \ldots G_{i-1}\left(a_{(2i-2,1,2i-1)}\right)\right)|_{A, c_0}(m_i | a_{2^{-i}+1}, c_0)
\]
\[\because G_i \text{ and } M_i \text{ have identical distribution}\]
\[
= \int_{m_1} dF_{M_i\left(a_{2^{-i}+1}, m_1, G_2(a_{2^{-i}+3}a_{2^{-i}+4}) \ldots G_{i-1}\left(a_{(2i-2,1,2i-1)}\right)\right)|_{A, c_0}(m_i | a_{2^{-i}+1}, c_0) \times
\]
\[dF_{G_i|_{A, c_0}}(G_1(a_{2^{-i}+2}) = m_1 | a_{2^{-i}+1}, c_0) \]
\[
= \int_{m_1} dF_{M_i\left(a_{2^{-i}+1}, m_1, G_2(a_{2^{-i}+3}a_{2^{-i}+4}) \ldots G_{i-1}\left(a_{(2i-2,1,2i-1)}\right)\right)|_{A, c_0}(m_i | a_{2^{-i}+1}, c_0) \times
\]
\[ dF_{G_1|A,C_0}(G_1(a_{2i-1+2}) = m_1|a_{2i-1+1}, c_0) \]

(\(\because\) \(G_1\) independs to all variables)

\[ = \int_{i=1}^{M1} dF_M(a_{2i-1+1}, m_1, a_2(a_{2i-1+1}), \ldots, a_{2i-1}(a_{2i-2+1}, a_{2i+1})) |_{A,C_0,M1} (m_1|a_{2i-1+1}, c_0, m_1) \times \]

\[ dF_{G_1|A,C_0}(G_1(a_{2i-1+2}) = m_1|a_{2i-1+1}, c_0) \]

(\(\because\) Assumption 4)

In the last equation, the second component has assumed be identified, and hence we only calculate the first component.

\[ dF_M(a_{2i-1+1}, m_1, a_2(a_{2i-1+1}), \ldots, a_{2i-1}(a_{2i-2+1}, a_{2i+1})) |_{A,C_0,M1} (m_1|a_{2i-1+1}, c_0, m_1) \]

\[ = \int_{i=1}^{M1} dF_M(a_{2i-1+1}, m_1, a_2(a_{2i-1+1}), \ldots, a_{2i-1}(a_{2i-2+1}, a_{2i+1})) |_{A,C_0,M1} (m_1|a_{2i-1+1}, c_0, m_1) \times \]

\[ dF_{c_i|A,C_0,M1}(c_i|a_{2i-1+1}, c_0) \]

Keeping the second component in the equation above, the first component can be identified with respect to \(m_2\). Iteratively, we have

\[ dF_{G_1(a_{2i-1+1}, a_{2i+1})}(m_1|c_0) \]

\[ = \int_{m(i-1)}^{M1} \int_{c(i)}^{M1} dF_M(a_{2i-1+1}, m_1, m_2, \ldots, m_i) |_{A,M, c(i)} (m_1|a_{2i-1+1}, m_1, c(i), c_0) \times \]

\[ \Pi_{j=1}^{i} dF_{c,j|A,M, c, c_0}(c_j|a_{2i-1+1}, m_1, c(i), c_0) \Pi_{j=1}^{i} H_j \left( m_j, a_{2j-1+2}a_{2j+2}, c_0 \right) \]

(\(\because\) consistency assumption)

\[ = \int_{m(i-1)}^{M1} \int_{c(i)}^{M1} dF_M(a_{2i-1+1}, m_1, m_1, c(i), c_0) \times \]

\[ \Pi_{j=1}^{i} dF_{c,j|A,M, c, c_0}(c_j|a_{2i-1+1}, m_1, c(i), c_0) \Pi_{j=1}^{i} H_j \left( m_j, a_{2j-1+2}a_{2j+2}, c_0 \right) \]
Appendix 3: Proof of Theorem 1

\[ E \left[ Y \left( a_1, G_1(a_2), G_2(a_3, a_4), \ldots, G_K \left( a_{(2^K-1+1,2^K)} \right) \right) \right] \]

\[ = \int_{c_0} E \left[ Y \left( a_1, G_1(a_2), G_2(a_3, a_4), \ldots, G_K \left( a_{(2^K-1+1,2^K)} \right) \right) \right] \text{d}F_{c_0}(c_0) \]

\[ = \int_{c_0} \int_{m_1} E \left[ Y \left( a_1, m_1, G_2(a_3, a_4), \ldots, G_K \left( a_{(2^K-1+1,2^K)} \right) \right) \right] \text{d}F_{G_1(a_1)}(m_1) \text{d}F_{c_0}(c_0) \]

\[ = \int_{c_0} \int_{m_1} E \left[ Y \left( a_1, m_1, m_2, \ldots, G_K \left( a_{(2^K-1+1,2^K)} \right) \right) \right] \text{d}F_{G_2(a_3,a_4)}(m_2) \text{d}F_{G_1(a_1)}(m_1) \text{d}F_{c_0}(c_0) \]

(∵ \( G_1 \) independs to all variables)

Iteratively, the finial equation can be done as

\[ \int_{c_0} \int_{m_{(1,K)}} E \left[ Y(a_1, m_{(1,K)}) \right] \prod_{i=1}^{K} \text{d}F_{G_i} \left( a_{(2^{i-1}+1,2^i)} \right) \text{d}F_{c_0}(c_0) \]

\[ = \int_{c_0} \int_{m_{(1,K)}} F \int_{c_0, a_1, m_{(1,K)}} \prod_{i=1}^{K} H_i \left( m_i, a_{(2^{i-1}+1,2^i)} \right) \text{d}F_{c_0}(c_0). \]
Appendix 4: Examples of Lemma 2 and Theorem 1 with one confounders or time-varying confounders

4.1 No time-varying confounding or interaction

4.1.1 For one mediator (K=1)

Theorem 1

\[
\varphi_1(a_1, a_2) = \int_{c_0} \int_{m_1} E[Y(a_1, m_1)|c_0] dF_{G_1(a_2)|c_0}(m_1|c_0)dF_{C_0}(c_0)
\]

\[
= \int_{c_0} \int_{m_1} E[Y(a_1, m_1, c_0)|a_2] dF_{G_1(a_2)|c_0}(m_1|c_0)dF_{C_0}(c_0)
\]

Lemma 2

\[
dF_{G_1(a_2)|c_0}(m_1|c_0) = H_1(m_1, a_2) = dF_{M_1|A,C_0}(m_1|a_2, c_0)
\]

4.1.2 For two mediators (K=2)

Theorem 1

\[
\varphi_2(a_1, a_2, a_3, a_4) = \int_{c_0} \int_{m_1,m_2} E[Y(a_1, m_1, m_2)|c_0] dF_{G_1(a_2)|c_0}(m_1|c_0) dF_{G_2(a_3,a_4)|c_0}(m_2|c_0)dF_{C_0}(c_0)
\]

\[
= \int_{c_0} \int_{m_1,m_2} E[Y(a_1, m_1, m_2, c_0)|a_3,a_4] dF_{G_1(a_2)|c_0}(m_1|c_0) dF_{G_2(a_3,a_4)|c_0}(m_2|c_0)dF_{C_0}(c_0)
\]

Lemma 2

\[
dF_{G_1(a_2)|c_0}(m_1) = H_1(m_1, a_2) = dF_{M_1|A,C_0}(m_1|a_2, c_0)
\]

\[
dF_{G_2(a_3,a_4)|c_0}(m_2) = H_2(m_2, a_3, a_4)
\]

\[
= \int_{m_1} dF_{M_2|A,C_0,M_1}(m_2|m_3,m_4,c_0,m_1) dF_{M_1|A,C_0}(m_1|a_3,c_0)
\]

4.1.3 For three mediators (K=3)

Theorem 1
Theorem 1

\[ \varphi_3(a_1, a_2, ..., a_7, a_8) = \int_{c_0} \int_{m_1, m_2, m_3} E[Y(a_1, m_1, m_2, m_3 | c_0)] dF_{G_1(a_2)}(m_1) dF_{G_2(a_3, a_4)}(m_2) dF_{G_3(a_5, a_6, a_7, a_8)}(m_3) dF_0(c_0) \]

\[ = \int_{c_0} \int_{m_1, m_2, m_3} E[Y(a_1, m_1, c_0, m_2, m_3)] dF_{G_1(a_2)}(m_1) dF_{G_2(a_3, a_4)}(m_2) dF_{G_3(a_5, a_6, a_7, a_8)}(m_3) dF_0(c_0) \]

Lemma 2

\[ dF_{G_1(a_2)} | c_0(m_1) = H_1(m_1, a_2) = dF_{M_1} | A, C_0(m_1 | a_2, c_0) \]

\[ dF_{G_2(a_3, a_4)} | c_0(m_2) = H_2(m_2, a_3, a_4) \]

\[ = \int_{m_1} dF_{M_2} | A, C_0, M_1(m_2 | a_3, c_0, m_1) dF_{M_1} | A, C_0(m_1 | a_4, c_0) \]

\[ dF_{G_3(a_5, a_6, a_7, a_8)} | c_0(m_3) = H_3(m_2, a_5, a_6, a_7, a_8) \]

\[ = \int_{m_1, m_2} dF_{M_3} | A, C_0, M_1, M_2(m_3 | a_5, c_0, m_1, m_2) \int_{m_1} dF_{M_2} | A, C_0, M_1(m_2 | a_6, c_0, m_1) dF_{M_1} | A, C_0(m_1 | a_7, c_0) \]

\[ dF_{M_1} | A, C_0(m_1 | a_8, c_0) \]

4.2 Time-varying confounding or interaction

4.2.1 For one mediator (K=1)

Theorem 1

\[ \varphi_1(a_1, a_2) = \int_{c_0} \int_{m_1} E[Y(a_1, m_1 | c_0)] dF_{G_1(a_2)}(m_1) dF_0(c_0) \]

\[ E[Y(a_1, m_1 | c_0)] = \int_{c_1} E[Y | a_1, m_1, c_1, c_0] dF_{C_1 | C_0, A}(c_1 | c_0, a_1) \]

Lemma 2

\[ dF_{G_1(a_2)} | c_0(m_1) = H_1(m_1, a_2) = \int_{c_1} dF_{M_1} | A, C_1, C_0(m_1 | a_2, c_1, c_0) dF_{C_1 | A, C_0}(c_1 | a_2, c_0) \]

4.2.2 For two mediators (K=2)

Theorem 1

\[ \varphi_2(a_1, a_2, a_3, a_4) = \]
\[
\int_{c_0} \int_{m_1, m_2} E[Y(a_1, m_1, m_2 | c_0)] \, dF_{G_1(a_2)}(m_1) \, dF_{G_2(a_3, a_4)}(m_2) \, dF_{C_0}(c_0)
\]

\[
E[Y(a_1, m_1, m_2 | c_0)] = \int_{c_1, c_2} E[Y|a_1, m_1, m_2, c_2, c_1, c_0] \, dF_{C_1|c_0,A}(c_1|c_0, a_1) \, dF_{C_2|c_1,c_0,A,M_1}(c_2|c_1, c_0, a_1, m_1)
\]

**Lemma 2**

\[
dF_{G_1(a_2)|c_0}(m_1) = H_1(m_1, a_2) = \int_{c_1} dF_{M_1|A,C_1,C_0}(m_1|a_2, c_1, c_0) \, dF_{C_1|A,C_0}(c_1|a_2, c_0)
\]

\[
dF_{G_2(a_3, a_4)|c_0}(m_2) = H_2(m_2, a_3, a_4)
\]

\[
= \int_{c_1, c_2} \int_{m_1} dF_{M_1|A,C_2,C_1,C_0,M_1}(m_2|a_3, c_2, c_1, c_0, m_1) \, dF_{C_2|c_1,c_0,A,M_1}(c_2|c_1, c_0, a_3, m_1)
\]

\[
dF_{M_1|A,C_1,C_0}(m_1|a_4, c_1, c_0) \int_{c_1} dF_{C_1|A,C_0}(c_1|a_4, c_0)
\]
Appendix 5: Examples of Section 4

This part illustrates the analytic regression-based estimator of PSE under the different scenarios.

Scenario 1. No time-varying confounding or interaction

5.1.1 For one mediator (K=1)

Assume Y and M are linear regression with no time-varying confounding.

\[ Y = \theta_0^Y + \theta_a^Y A + \theta_1^Y M_1 + \theta_c^Y C + \varepsilon_y, \text{ where } \varepsilon_y \sim N(0, \sigma_{\varepsilon_y}^2) \]

\[ M_1 = \theta_0^1 + \theta_a^1 A + \theta_1^1 C + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0, \sigma_{\varepsilon_1}^2) \]

\[ \varphi_1(a_1, a_2) = \int_{c_0} \int_{m_1} E[Y|a_1, m_1, c_0] \, dF_{M_1|A,C_0}(m_1|a_2, c_0) \, dF_{C_0}(c_0) \]

\[ = \theta_0^Y + \theta_a^Y a_1 + \theta_1^Y (\theta_0^1 + \theta_a^1 a_2 + \theta_c^1 E(c_0)) + \theta_c^Y E(c_0) \]

\[ = \theta_0^Y + \theta_a^Y a_1 + \theta_1^Y (\theta_0^1 + \theta_a^1 a_2) + \text{constant} \]

Calculate giPSE,

\[ \text{giPSE}_1(1) = \varphi_1(1,0) - \varphi_1(0,0) = \theta_a^Y \]

\[ \text{giPSE}_1(2) = \varphi_1(1,1) - \varphi_1(1,0) = \theta_1^Y \theta_a^1. \]

5.1.2 For two mediators (K=2)

Assume Y and M are linear regression with no time-varying confounding.

\[ Y = \theta_0^Y + \theta_a^Y A + \theta_1^Y M_1 + \theta_2^Y M_2 + \theta_c^Y C + \varepsilon_y, \text{ where } \varepsilon_y \sim N(0, \sigma_{\varepsilon_y}^2) \]

\[ M_1 = \theta_0^1 + \theta_a^1 A + \theta_1^1 C + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0, \sigma_{\varepsilon_1}^2) \]

\[ M_2 = \theta_0^2 + \theta_a^2 A + \theta_1^2 C + \theta_2^1 M_1 + \varepsilon_2, \text{ where } \varepsilon_2 \sim N(0, \sigma_{\varepsilon_2}^2) \]

\[ \varphi_2(a_1, a_2, a_3, a_4) \]

\[ = \int_{c_0} \int_{m_1} \int_{m_2} E[Y|a_1, m_1, c_0, m_2] \, dF_{M_1|A,C_0}(m_1|a_2, c_0) \, dF_{M_2|A,C_0,M_1}(m_2|a_3, c_0, m_1) \, dF_{M_1|A,C_0}(m_1|a_4, c_0) \, dF_{C_0}(c_0) \]

\[ = \theta_0^Y + \theta_a^Y a_1 + \theta_1^Y (\theta_0^1 + \theta_a^1 a_2 + \theta_c^1 E(c_0)) + \theta_2^Y (\theta_0^2 + \theta_a^2 a_3 + \theta_1^2 E(c_0)) + \theta_c^Y E(c_0) \]

\[ = \theta_0^Y + \theta_a^Y a_1 + \theta_1^Y (\theta_0^1 + \theta_a^1 a_2) + \theta_2^Y (\theta_0^2 + \theta_a^2 a_3) + \text{constant} \]

Calculate giPSE,

\[ \text{giPSE}_2(1) = \varphi_2(1,0,0,0) - \varphi_2(0,0,0,0) = \theta_a^Y \]

\[ \text{giPSE}_2(2) = \varphi_2(1,1,0,0) - \varphi_2(1,0,0,0) = \theta_1^Y \theta_a^1 + \theta_2^Y \theta_1^2. \]
\[
\begin{align*}
&= \theta_0^Y + \theta_1^Y E(c_0) + \theta_2^Y a_1 + \theta_3^Y (\theta_0^1 + \theta_1^1 a_2 + \theta_2^1 E(c_0)) + \theta_3^Y (\theta_0^2 + \theta_1^2 a_3 + \theta_2^2 E(c_0)) \\
&\quad + \theta_3^2 (\theta_0^3 + \theta_1^3 a_4 + \theta_2^3 E(c_0))) \\
&= \theta_0^Y a_1 + \theta_1^Y \theta_0^1 a_2 + \theta_2^Y \theta_0^1 a_3 + \theta_3^Y \theta_0^1 \theta_1^1 a_4 + \text{constant}
\end{align*}
\]

Calculate giPSE,

\[
\text{giPSE}_2(1) = \varphi_2(1,0,0,0) - \varphi_2(0,0,0,0) = \theta_0^Y
\]

\[
\text{giPSE}_2(2) = \varphi_2(1,1,0,0) - \varphi_2(1,0,0,0) = \theta_1^Y \theta_0^1
\]

\[
\text{giPSE}_2(3) = \varphi_2(1,1,1,0) - \varphi_2(1,1,0,0) = \theta_2^Y \theta_0^2
\]

\[
\text{giPSE}_2(4) = \varphi_2(1,1,1,1) - \varphi_2(1,1,1,0) = \theta_2^Y \theta_1^2 \theta_0^1.
\]

### 5.1.3 For three mediators (K=3)

Assume Y and M are linear regression with no time-varying confounding.

\[
Y = \theta_0^Y + \theta_1^Y A + \theta_2^Y M_1 + \theta_3^Y M_2 + \theta_3^Y M_3 + \theta_4^Y C + \varepsilon_Y, \text{ where } \varepsilon_Y \sim N(0, \sigma_Y^2)
\]

\[
M_1 = \theta_0^1 + \theta_1^1 A + \theta_2^1 C + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0, \sigma_1^2)
\]

\[
M_2 = \theta_0^2 + \theta_1^2 A + \theta_2^2 C + \theta_1^2 M_1 + \varepsilon_2, \text{ where } \varepsilon_2 \sim N(0, \sigma_2^2)
\]

\[
M_3 = \theta_0^3 + \theta_1^3 A + \theta_2^3 C + \theta_1^3 M_1 + \theta_2^3 M_2 + \varepsilon_3, \text{ where } \varepsilon_3 \sim N(0, \sigma_3^2)
\]

\[
\varphi_3(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)
\]

\[
= \int_{\text{c_0}} \int_{m_1, m_2, m_3} E(Y|a_1, m_1, c_0, m_2, m_3) dF_{M_1|A,C_0}(m_1 | a_2, c_0)
\]

\[
\int_{m_1} dF_{M_2|A,C_0,M_1}(m_2|a_3, c_0, m_1) dF_{M_1|A,C_0}(m_1 | a_4, c_0)
\]

\[
\int_{m_2} dF_{M_3|A,C_0,M_1,M_2}(m_3|a_5, c_0, m_1, m_2) \int_{m_1} dF_{M_1|A,C_0}(m_4|a_8, c_0) \int_{m_1} dF_{M_2|A,C_0,M_1}(m_2|a_7, c_0, m_1)
\]

45
\[ dF_{M_1|A,c_0}(m_1|a_8,c_0)df_{c_0}(c_0) \]
\[ = \theta^\gamma_\alpha a_1 + \theta^\gamma_1 a_2 + \theta^\gamma_2 a_3 + \theta^\gamma_3 a_4 + \theta^\gamma_4 a_5 + \theta^\gamma_5 a_6 + \theta^\gamma_6 a_7 + \theta^\gamma_7 a_8 + \text{constant} \]

Calculate giPSE,
\[ \text{giPSE}_3(1) = \varphi_3(1,0,0,0,0,0,0,0) - \varphi_3(0,0,0,0,0,0,0,0) = \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(2) = \varphi_3(1,1,0,0,0,0,0,0) - \varphi_3(1,0,0,0,0,0,0,0) = \theta^\gamma_1 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(3) = \varphi_3(1,1,1,0,0,0,0,0) - \varphi_3(1,1,0,0,0,0,0,0) = \theta^\gamma_2 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(4) = \varphi_3(1,1,1,1,0,0,0,0) - \varphi_3(1,1,1,0,0,0,0,0) = \theta^\gamma_3 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(5) = \varphi_3(1,1,1,1,1,0,0,0) - \varphi_3(1,1,1,1,0,0,0,0) = \theta^\gamma_3 \theta^\gamma_1 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(6) = \varphi_3(1,1,1,1,1,1,0,0) - \varphi_3(1,1,1,1,1,0,0,0) = \theta^\gamma_3 \theta^\gamma_3 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(7) = \varphi_3(1,1,1,1,1,1,1,0) - \varphi_3(1,1,1,1,1,1,0,0) = \theta^\gamma_3 \theta^\gamma_3 \theta^\gamma_1 \theta^\gamma_\alpha \]
\[ \text{giPSE}_3(8) = \varphi_3(1,1,1,1,1,1,1,1) - \varphi_3(1,1,1,1,1,1,1,0) = \theta^\gamma_3 \theta^\gamma_3 \theta^\gamma_3 \theta^\gamma_1 \theta^\gamma_\alpha . \]

**Scenario 2. With time-varying confounding but no interaction**

**5.2.1 For one mediator (K=1)**

Assume Y and M are linear regression with no time-varying confounding.

\[ Y = \theta^\gamma_0 + \theta^\gamma_\alpha A + \theta^\gamma_1 M_1 + \theta^\gamma_2 C_1 + \theta^\gamma_3 \epsilon_\gamma , \text{ where } \epsilon_\gamma \sim N(0,\sigma^2_\gamma) \]
\[ M_1 = \theta^\gamma_0 + \theta^\gamma_1 A + \theta^\gamma_2 C_1 + \epsilon_1 , \text{ where } \epsilon_1 \sim N(0,\sigma^2_1) \]
\[ C_1 = \theta^\gamma_0 + \theta^\gamma_1 A + \epsilon_1 , \text{ where } \epsilon_1 \sim N(0,\sigma^2_1) \]

https://biostats.bepress.com/harvardbiostat/paper217
\[ \varphi_1(a_1, a_2) \]
\[ = \int_{c_0} \int_{m_1} \int_{c_1} E[Y|a_1, m_1, c_1, c_0] dF_{C_1|c_0,A}(c_1|c_0, a_1) \int_{c_1} dF_{M_1|A,c_1,c_0}(m_1|a_2, c_1, c_0) dF_{C_1|A,c_0}(c_1|a_2, c_0) \]
\[ dF_{C_0}(c_0) \]
\[ = (\theta_1^Y + \theta_1^Y \theta_1^C) a_1 + (\theta_1^Y + \theta_1^Y \theta_1^C) a_2 + \text{constant} \]

Calculate giPSE,
\[ \text{giPSE}_1(1) = \varphi_1(1,0) - \varphi_1(0,0) = \theta_1^Y + \theta_1^Y \theta_1^C \]
\[ \text{giPSE}_1(2) = \varphi_1(1,1) - \varphi_1(1,0) = \theta_1^Y + \theta_1^Y \theta_1^C \]

Same as SEM.

5.2.2 For two mediators (K=2)
Assume Y and M are linear regression with no time-varying confounding.
\[ Y = \theta_1^Y + \theta_1^Y A + \theta_1^Y M_1 + \theta_2^Y M_2 + \theta_2^Y C_2 + \theta_1^Y C_1 + \theta_0^Y C_0 + \varepsilon_Y \text{, where} \]
\[ \varepsilon_Y \sim N(0, \sigma_Y^2) \]
\[ M_2 = \theta_2^A + \theta_1^Y A + \theta_2^Y M_1 + \theta_2^Y C_2 + \theta_1^Y C_1 + \theta_0^C C_0 + \varepsilon_2 \text{, where} \]
\[ \varepsilon_2 \sim N(0, \sigma_2^2) \]
\[ M_1 = \theta_1^A + \theta_1^A A + \theta_2^C C_1 + \theta_0^C C_0 + \varepsilon_1 \text{, where} \]
\[ \varepsilon_1 \sim N(0, \sigma_1^2) \]
\[ C_2 = \theta_0^C + \theta_2^C A + \theta_2^C M_1 + \theta_2^C C_1 + \theta_2^C C_0 + \varepsilon_2 \text{, where} \]
\[ \varepsilon_2 \sim N(0, \sigma_2^2) \]
\[ C_1 = \theta_0^C + \theta_1^C A + \theta_0^C C_0 + \varepsilon_1 \text{, where} \]
\[ \varepsilon_1 \sim N(0, \sigma_1^2) \]

\[ \varphi_1(a_1, a_2, a_3, a_4) \]
\[ = \int_{c_0} \int_{m_1,m_2} \int_{c_1,c_2} E[Y|a_1, m_1, c_1, c_0] dF_{C_1|c_0,A}(c_1|c_0, a_1) dF_{C_2|c_1,c_0,A,M_1}(c_2|c_1, c_0, a_1, m_1) \]
\[ \int_{c_1} dF_{M_1|A,c_1,c_0}(m_1|a_2, c_1, c_0) dF_{C_1|A,c_0}(c_1|a_2, c_0) \int_{c_1,c_2} dF_{M_2|A,c_2,c_1,c_0}(m_2|a_3, c_2, c_1, c_0, m_1) \]
\[ \begin{align*}
\text{Scenario 3.} & \quad \text{giPSE} \\
\text{giPSE}_2(1) & = \varphi_2(1,0,0,0) - \varphi_2(0,0,0,0) = \theta_2^{y_1} + \theta_2^{y_2} \theta_1^{c_2} + \theta_2^{y_2} \theta_1^{c_2} \theta_1^{c_1} + \theta_2^{y_2} \theta_1^{c_1} \\
\text{giPSE}_2(2) & = \varphi_2(1,1,0,0) - \varphi_2(1,0,0,0) \\
& = \theta_1^{y_1} \theta_1^{c_1} + \theta_1^{y_2} \theta_1^{c_1} + \theta_1^{y_2} \theta_1^{c_2} \theta_1^{c_1} + \theta_1^{y_2} \theta_1^{c_1} \\
\text{giPSE}_2(3) & = \varphi_2(1,1,1,0) - \varphi_2(1,1,0,0) \\
& = \theta_2^{y_1} \theta_2^{c_1} + \theta_2^{y_2} \theta_2^{c_1} + \theta_2^{y_2} \theta_2^{c_2} \theta_1^{c_1} + \theta_2^{y_2} \theta_2^{c_1} \\
\text{giPSE}_2(4) & = \varphi_2(1,1,1,1) - \varphi_2(1,1,1,0) \\
& = \theta_2^{y_1} \theta_2^{c_1} + \theta_2^{y_2} \theta_2^{c_1} \theta_1^{c_1} + \theta_2^{y_2} \theta_2^{c_2} \theta_1^{c_1} + \theta_2^{y_2} \theta_2^{c_1} \theta_1^{c_1}.
\end{align*} \]

5.3.1 For one mediator (K=1)

\[ \begin{align*}
\varphi_1(a_1, a_2) & = \int_{c_0} \int_{m_1} E[Y|a_1, c_0, m_1] dF_{M_1|A,C_0}(m_1|a_2, c_0) dF_C(c_0) \\
& = \theta_1^{y_1} a_1 + \theta_1^{y_2} a_2 + \theta_1^{y_1} \theta_1^{c_1} + \theta_1^{y_2} \theta_1^{c_2} a_2 + \theta_1^{y_2} \theta_1^{c_2} \theta_1^{c_1} \theta_1^{c_1} \theta_1^{c_1}.
\end{align*} \]

Calculate giPSE.
\[ \text{giPSE}_1(1) = \varphi_1(1,0) - \varphi_1(0,0) = \theta_a^\gamma + \theta_{a_1}^\gamma \theta_0^1 + \theta_{a_2}^\gamma \theta_c^1 E(c_0) \]
\[ \text{giPSE}_1(2) = \varphi_1(1,1) - \varphi_1(1,0) = \theta_1^\gamma \theta_{a_1}^1 + \theta_1^\gamma \theta_a^1 \]

5.3.2 For two mediators (K=2)

\[ Y = \theta_0^0 + \theta_2^0 A + \theta_1^0 M_1 + \theta_2^0 M_2 + \theta_2^0 C_0 + \theta_{a_1}^0 A M_1 + \theta_{a_2}^0 A M_2 + \theta_{12}^0 M_1 M_2 + \epsilon_Y, \]
\[ \text{where } \epsilon_Y \sim N(0, \sigma_Y^2) \]
\[ M_1 = \theta_0^1 + \theta_{a_2}^1 A + \theta_{c_0}^1 C_0 + \epsilon_1, \quad \text{where } \epsilon_1 \sim N(0, \sigma_1^2) \]
\[ M_2 = \theta_0^2 + \theta_{a_2}^2 A + \theta_{c_0}^2 C_0 + \theta_{12}^2 M_1 + \theta_{a_2}^2 A M_1, \quad \text{where } \epsilon_2 \sim N(0, \sigma_2^2) \]
\[ \varphi_2(a_1, a_2, a_3, a_4) \]
\[ = \int_{c_0} \int_{m_1, m_2} E[Y|a_1, c_0, m_1, m_2] dF_{M_1|A,C_0}(m_1|a_2, c_0)dF_{M_2|A,C_0,m_1}(m_2|a_3, c_0, m_1)dF_{M_1|A,C_0}(m_1|c_0, a_4) \]

\[ dF_{c_0}(c_0) = \theta_0^y + \theta_{a_2}^y a_1 + (\theta_{a_1}^y + \theta_{a_2}^y a_1)(\theta_0^0 + \theta_{a_2}^0 a_2 + \theta_c^0 E(c_0)) \]
\[ + (\theta_{a_2}^y + \theta_{a_2}^y a_1)(\theta_0^0 + \theta_{a_2}^y a_3 + \theta_c^0 E(c_0) + \theta_2^0(\theta_0^0 + \theta_{a_2}^0 a_4 + \theta_c^0 E(c_0))) \]
\[ + \theta_{12}^y (\theta_0^0 + \theta_{a_2}^0 a_2 + \theta_c^0 E(c_0))(\theta_0^0 + \theta_{a_2}^0 a_3 + \theta_2^0 E(c_0) + \theta_1^0(\theta_0^0 + \theta_{a_2}^0 a_4 + \theta_c^0 E(c_0))) \]
\[ + \theta_{a_2}^0 a_3 (\theta_0^0 + \theta_{a_2}^0 a_4 + \theta_c^0 E(c_0))) \]

Calculate giPSE

\[ \text{giPSE}_2(1) = \varphi_2(1,0,0,0) - \varphi_2(0,0,0,0) \]
\[ = \theta_a^\gamma + \theta_{a_1}^\gamma (\theta_0^1 + \theta_c^1 E(c_0)) + \theta_{a_2}^\gamma (\theta_0^1 + \theta_c^1 E(c_0)) + \theta_1^\gamma \theta_0^1 \]
\[ + \theta_1^\gamma \theta_c^1 E(c_0)) \]
\[ \text{giPSE}_2(2) = \varphi_2(1,1,0,0) - \varphi_2(1,0,0,0) = (\theta_1^\gamma + \theta_{a_2}^\gamma) \theta_a^1 \]
\[ \text{giPSE}_2(3) = \varphi_2(1,1,1,0) - \varphi_2(1,1,0,0) \]
\[ = (\theta_2^\gamma + \theta_{a_2}^\gamma) (\theta_2^0 + \theta_{a_1}^0 + \theta_c^0 E(c_0)) + \theta_{12}^0 (\theta_0^1 + \theta_a^1 \]
\[ + \theta_c^1 E(c_0)) (\theta_0^1 + \theta_{a_1}^0 + \theta_c^1 E(c_0)) \]
\[ \text{giPSE}_2(4) = \varphi_2(1,1,1,1) - \varphi_2(1,1,1,0) \]
\[ = (\theta_2^\gamma + \theta_{a_2}^\gamma) (\theta_2^0 + \theta_{a_1}^0 + \theta_a^0) + \theta_{12}^0 (\theta_0^1 + \theta_a^1 a_2 + \theta_c^1 E(c_0)) + \theta_{a_2}^0 \theta_a^1 \]

5.3.3 For three mediators (K=3)
\[ Y = \theta_0^\psi + \theta_2^\psi A + \theta_1^\psi M_1 + \theta_3^\psi M_2 + \theta_3^\psi C_0 + \theta_1^\psi A_M + \theta_2^\psi A_M + \theta_3^\psi A_M + \varepsilon_y, \quad \text{where } \varepsilon_y \sim N(0, \sigma_y^2) \]

\[ M_1 = \theta_0^1 + \theta_1^1 A + \theta_2^1 C_0 + \varepsilon_1, \quad \text{where } \varepsilon_1 \sim N(0, \sigma_1^2) \]

\[ M_2 = \theta_0^2 + \theta_1^2 A + \theta_2^2 M_1 + \theta_3^2 C_0 + \theta_2^2 A_M + \varepsilon_2, \quad \text{where } \varepsilon_2 \sim N(0, \sigma_2^2) \]

\[ M_3 = \theta_0^3 + \theta_1^3 A + \theta_2^3 M_1 + \theta_2^3 M_2 + \theta_3^3 C_0 + \theta_3^3 A_M + \varepsilon_3, \quad \text{where } \varepsilon_3 \sim N(0, \sigma_3^2) \]

\[ \varphi_3(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) = \int_{c_0} \int_{m_1, m_2, m_3} E[Y|a_1, c_0, m_1, m_2, m_3] \, dF_{M_1|A,C_0}(m_1|a_2, c_0) \, dF_{M_2|A,C_0,M_1}(m_2|a_3, c_0, m_1) \]

\[ = \theta_1^\psi a_1 + (\theta_1^\psi + \theta_{a_1}^\psi)(\theta_0^1 + \theta_1^1 a_2 + \theta_1^3 E(c_0)) \]

\[ + (\theta_2^\psi + \theta_{a_2}^\psi)(\theta_0^2 + \theta_2^2 a_3 + \theta_2^2 \theta_0^1 a_4 + \theta_2^3 E(c_0)) + \theta_2^2 E(c_0) \]

\[ + \theta_{a_2}^3 a_3(\theta_0^1 + \theta_1^1 a_4 + \theta_1^3 E(c_0)) \]

\[ + (\theta_3^\psi + \theta_{a_3}^\psi)(\theta_0^3 + \theta_3^3 a_5 + (\theta_3^3 + \theta_{a_1}^3)(\theta_0^1 + \theta_1^1 a_6 + \theta_1^3 E(c_0))) \]

\[ + (\theta_2^3 + \theta_{a_2}^3)(\theta_0^2 + \theta_2^2 a_7 + \theta_1^2(\theta_0^1 + \theta_1^1 a_8 + \theta_1^3 E(c_0)) + \theta_2^2 E(c_0) \]

\[ + \theta_{a_2}^2 a_7(\theta_0^1 + \theta_1^1 a_6 + \theta_1^3 E(c_0))) + \theta_2^2 E(c_0) \]

Calculate giPSE

\[ \text{giPSE}_3(1) = \theta_1^\psi + \theta_{a_1}^\psi(\theta_0^1 + \theta_1^3 E(c_0)) + \theta_{a_2}^\psi(\theta_0^2 + \theta_2^2 \theta_0^1 + \theta_2^3 \theta_1^1 E(c_0)) + \theta_2^2 E(c_0) \]

\[ + \theta_{a_3}^\psi \theta_3^0 + \theta_3^3 \theta_0^1 + \theta_1^3 \theta_1^3 E(c_0) + \theta_2^3 \theta_0^2 + \theta_2^2 \theta_1^0 + \theta_2^3 \theta_1^1 \theta_1^2 E(c_0) \]

\[ \text{giPSE}_3(2) = (\theta_1^\psi + \theta_{a_1}^\psi) \theta_0^1 \]

\[ \text{giPSE}_3(3) = (\theta_2^\psi + \theta_{a_2}^\psi)(\theta_0^2 + \theta_{a_1}^\psi(\theta_0^1 + \theta_1^1 E(c_0))) \]
\begin{align*}
giPSE_3(4) &= (\theta_2^\gamma + \theta_2^\gamma)(\theta_1^\alpha \theta_1^\alpha + \theta_1^\gamma) \\
giPSE_3(5) &= (\theta_3^\gamma + \theta_3^\gamma)(\theta_2^\alpha + \theta_2^\alpha_1(\theta_0^0 + \theta_0^1 E(c_0))) \\
&\quad + \theta_3^\gamma(\theta_0^0 + \theta_0^1(\theta_0^0 + \theta_0^1 E(c_0)) + \theta_0^1 E(c_0)) \\
giPSE_3(6) &= (\theta_3^\gamma + \theta_3^\gamma)((\theta_2^\alpha + \theta_2^\alpha_1)\theta_2^\alpha) \\
giPSE_3(7) &= (\theta_3^\gamma + \theta_3^\gamma)((\theta_2^\alpha + \theta_2^\alpha_1)(\theta_0^0 + \theta_0^1 E(c_0))) \\
giPSE_3(8) &= (\theta_3^\gamma + \theta_3^\gamma)(\theta_2^\alpha + \theta_2^\alpha_1(\theta_0^2 \theta_0^1 + \theta_0^2 \theta_1^\alpha)) \\
\end{align*}

\section*{Scenario 4. Log-link and logistic regression with rare disease assumption}

\subsection*{5.4.1 For one mediator (K=1)}

\text{logit}(Y) = \theta_0^\gamma + \theta_0^\gamma A + \theta_1^\gamma M_1 + \theta_0^\gamma C_0 + \varepsilon_Y, \text{ where } \varepsilon_Y \sim N(0, \sigma_Y^2)

M_1 = \theta_0^0 + \theta_0^1 A + \theta_0^1 C_0 + \varepsilon_1, \text{ where } \varepsilon_1 \sim N(0, \sigma_1^2)

For y \ll 1, \text{ logit}(Y) \approx \text{ log}(Y).

Hence \ Y = \exp(\theta_0^\gamma + \theta_0^\gamma A + \theta_1^\gamma M_1 + \theta_0^\gamma C_0 + \varepsilon_Y)

\varphi_1(a_1, a_2) = \int_{c_0} \int_{m_1} E[Y|a_1, m_1, c_0] dF_{M_1|A,C_0}(m_1|a_2, c_0) dF_{C_0}(c_0)

= \int_{c_0} \int_{m_1} \exp(\theta_0^\gamma + \theta_0^\gamma A + \theta_1^\gamma M_1 + \theta_0^\gamma C_0) dF_{M_1|A,C_0}(m_1|a_2, c_0) dF_{C_0}(c_0)

= \exp(\theta_0^\gamma + \theta_0^\gamma A_1)E(\exp(\theta_0^\gamma C_0)) \int_{c_0} \exp(\mu_m \theta_1^\gamma + \frac{1}{2} (\theta_1^\gamma \sigma_1^2)) dF_{C_0}(c_0)

, where \ M_1 \sim N(\mu_m, \sigma_1^2)

= \exp(\theta_0^\gamma A_1 + \theta_1^\gamma \theta_1^\gamma A_2) \cdot \text{ Constant}

Calculate giPSE by RR scale

\text{giPSE}_1(1) = \varphi_1(1,0)/\varphi_1(0,0) = \theta_0^\gamma
giPSE₁(2) = \varphi₁(1,1)/\varphi₁(1,0) = \theta₁^\gamma\theta₁^α

5.4.2 For two mediators (K=2)

\text{logit}(Y) = \theta₀^\gamma + \theta₁^\gamma A + \theta₂^\gamma M₁ + \theta₂^\gamma M₂ + \theta₃^\gamma C₀ + \varepsilon_y , \text{ where } \varepsilon_y \sim N(0, \sigma₂^\gamma)

M₁ = \theta₁^α A + \theta₁^\gamma C₀ + \varepsilon₁ , \text{ where } \varepsilon₁ \sim N(0, \sigma₁^α)

M₂ = \theta₂^α A + \theta₂^\gamma C₀ + \theta₂^\gamma M₁ + \varepsilon₂ , \text{ where } \varepsilon₂ \sim N(0, \sigma₂^α)

For \ y \ll 1, \text{ logit}(Y) \approx \log(Y).

Hence \ Y = \exp(\theta₀^\gamma + \theta₁^\gamma A + \theta₁^\gamma M₁ + \theta₂^\gamma M₂ + \theta₃^\gamma C₀ + \varepsilon_y)

\varphi₂(a₁, a₂, a₃, a₄)

= \int_{c₀} \int_{m₁,m₂} \exp(\theta₀^\gamma + \theta₁^\gamma A + \theta₁^\gamma M₁ + \theta₂^\gamma M₂ + \theta₃^\gamma C₀ +)dF_{M₁\mid A,C₀}(m₁ \mid a₂, c₀)

\int_{m₁} dF_{M₂\mid A,C₀,m₁}(m₂ \mid a₃, c₀, m₁) dF_{M₁\mid A,C₀}(m₁ \mid a₄, c₀) dF_{C₀}(c₀)

= \exp(\theta₀^\gamma + \theta₁^\gamma a₁)E(\exp(\theta_c^\gamma c₀)) \int_{c₀} \exp(\mu_{m₁} \theta₁^\gamma + \frac{1}{2}(\theta₁^\gamma σ₁)^2) \exp(\mu_{m₂} \theta₂^\gamma + \frac{1}{2}(\theta₂^\gamma σ₂)^2) dF_{C₀}(c₀)

, \text{ where } M₁ \sim N(\mu_{m₁}, \sigma₁^2), M₂ \sim N(\mu_{m₂}, \sigma₂^2)

= \exp(\theta₁^\gamma a₁ + \theta₁^\gamma \theta₁^α a₂ + \theta₂^\gamma \theta₁^α a₃ + \theta₂^\gamma \theta₁^\gamma \theta₁^α a₄) \cdot \text{constant}

Calculate giPSE by RR scale

giPSE₂(1) = \varphi₂(1,0,0,0)/\varphi₂(0,0,0,0) = \theta₁^\gamma

giPSE₂(2) = \varphi₂(1,1,0,0)/\varphi₂(1,0,0,0) = \theta₁^\gamma\theta₁^α

giPSE₂(3) = \varphi₂(1,1,1,0)/\varphi₂(1,1,0,0) = \theta₂^\gamma\theta₂^α

giPSE₂(4) = \varphi₂(1,1,1,1)/\varphi₂(1,1,1,0) = \theta₂^\gamma\theta₂^α\theta₂^α

52
5.4.3 For three mediators (K=3)

\[ \text{logit}(Y) = \theta_0^Y + \theta_a^Y M_1 + \theta_2^M M_2 + \theta_3^M M_3 + \theta_C^Y + \varepsilon_Y , \] where \( \varepsilon_Y \sim N(0, \sigma_Y^2) \)

\[ M_1 = \theta_0^1 + \theta_1^A M_1 + \theta_C^C M_1 + \varepsilon_1 , \] where \( \varepsilon_1 \sim N(0, \sigma_1^2) \)

\[ M_2 = \theta_0^2 + \theta_1^A M_2 + \theta_C^C M_2 + \varepsilon_2 , \] where \( \varepsilon_2 \sim N(0, \sigma_2^2) \)

\[ M_3 = \theta_0^3 + \theta_1^A M_2 + \theta_2^M M_2 + \theta_3^M M_2 + \varepsilon_3 , \] where \( \varepsilon_3 \sim N(0, \sigma_3^2) \)

For \( y < 1 \), \( \text{logit}(Y) \approx \text{log}(Y) \).

Hence \( Y = \exp(\theta_0^Y + \theta_a^Y M_1 + \theta_2^M M_2 + \theta_3^M M_3 + \theta_C^Y + \varepsilon_Y) \)

\[ \varphi_3(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) \]

\[ = \int c_0 \int m_1, m_2, m_3 E[Y | a_1, m_1, c_0, m_2, m_3] dF_{M_1|A,C}(m_1 | a_2, c_0) \]

\[ \int m_1 dF_{M_2|A,C,M_1}(m_2 | a_3, c_0, m_1) dF_{M_1|A,C}(m_1 | a_2, c_0) \]

\[ \int m_1 dF_{M_3|A,C,M_1,M_2}(m_3 | a_4, c_0, m_2) \int m_1 dF_{M_1|A,C}(m_2 | a_8, c_0) \int m_1 dF_{M_2|A,C,M_1}(m_2 | a_7, c_0, m_1) \]

\[ dF_{M_1|A,C}(m_1 | a_8, c_0) dF_{C_0}(c_0) \]

\[ = \exp(\theta_0^Y + \theta_a^Y a_1) E(\exp(\theta_C^a c_0)) \int c_0 \exp(\mu_{m_1} \theta_1^Y + \frac{1}{2}(\theta_1^Y \sigma_1)^2) \exp(\mu_{m_2} \theta_2^Y + \frac{1}{2}(\theta_2^Y \sigma_2)^2) dF_{C_0}(c_0) \]

, where \( M_1 \sim N(\mu_{m_1}, \sigma_1^2), M_2 \sim N(\mu_{m_2}, \sigma_2^2), M_3 \sim N(\mu_{m_3}, \sigma_3^2) \)

\[ = \exp(\theta_a^Y a_1 + \theta_a^Y \theta_1^1 a_2 + \theta_a^Y \theta_2^1 a_3 + \theta_a^Y \theta_2^1 \theta_1^1 a_4 + \theta_a^Y (\theta_3^1 a_5 + \theta_3^1 \theta_1^1 a_6 + \theta_3^1 \theta_1^1 \theta_1^1 a_7 + \theta_3^1 \theta_1^1 \theta_1^1 a_8)) + \text{constant} \]

Calculate giPSE by RR scale
giPSE_3(1) = \varphi_3(1,0,0,0,0,0,0)/\varphi_3(0,0,0,0,0,0,0) = \theta_1^3 \\

giPSE_3(2) = \varphi_3(1,1,0,0,0,0,0)/\varphi_3(1,0,0,0,0,0,0) = \theta_1^1 \theta_1^2 \\

giPSE_3(3) = \varphi_3(1,1,1,0,0,0,0)/\varphi_3(1,1,0,0,0,0,0) = \theta_2^1 \theta_1^2 \\

giPSE_3(4) = \varphi_3(1,1,1,1,0,0,0)/\varphi_3(1,1,1,1,0,0,0) = \theta_2^1 \theta_1^3 \\

giPSE_3(5) = \varphi_3(1,1,1,1,1,0,0)/\varphi_3(1,1,1,1,1,0,0) = \theta_3^1 \theta_1^3 \\

giPSE_3(6) = \varphi_3(1,1,1,1,1,1,0)/\varphi_3(1,1,1,1,1,1,0) = \theta_3^1 \theta_1^3 \theta_1^2 \\

giPSE_3(7) = \varphi_3(1,1,1,1,1,1,1)/\varphi_3(1,1,1,1,1,1,0) = \theta_3^1 \theta_1^3 \theta_1^2 \\

giPSE_3(8) = \varphi_3(1,1,1,1,1,1,1)/\varphi_3(1,1,1,1,1,1,0) = \theta_3^1 \theta_1^3 \theta_1^2 \theta_1^1