Decomposition of Probabilities of Causation with Two Mediators

Yuta Kawakami[1] **Jin Tian**[1]

[1]Mohamed bin Zayed University of Artificial Intelligence, UAE

Abstract

Mediation analysis for probabilities of causation (PoC) provides a fundamental framework for evaluating the necessity and sufficiency of treatment in provoking an event through different causal pathways. One of the primary objectives of causal mediation analysis is to decompose the total effect into path-specific components. In this study, we investigate the path-specific probability of necessity and sufficiency (PNS) to decompose the total PNS into path-specific components along distinct causal pathways between treatment and outcome, incorporating two mediators. We define the path-specific PNS for decomposition and provide an identification theorem. Furthermore, we conduct numerical experiments to assess the properties of the proposed estimators from finite samples and demonstrate their practical application using a real-world educational dataset.

1 INTRODUCTION

Probabilities of causation (PoC) [Robins and Greenland, 1989, Tian and Pearl, 2000, Pearl, 2009, Kuroki and Cai, 2011, Dawid et al., 2014, Murtas et al., 2017, Shingaki and Kuroki, 2021, Kawakami et al., 2023, Rubinstein et al., 2025] and causal mediation analysis [Wright, 1921, 1934, Baron and Kenny, 1986, Robins and Greenland, 1992, Imai et al., 2010a,b, Tchetgen and Shpitser, 2012, Rubinstein et al., 2025] are valuable tools for decision-making and enhancing explainability in AI (XAI) [Shin, 2021]. One of the primary objectives of causal mediation analysis is to decompose the total effect into path-specific components. Recently, [Kawakami and Tian, 2025] provided mediation analysis for the probability of necessity and sufficiency (PNS) of the treatment (X) in causing the outcome (Y) via a mediator (M). They introduced the natural direct PNS (ND-PNS) and natural indirect PNS (NI-PNS) such that PNS can be decomposed into its direct component ND-PNS and indirect component NI-PNS. ND-PNS and NI-PNS can be used to answer the following causal questions:

(**Q-a**). *Would the treatment still be necessary and sufficient had there been no influence via the mediator M?*
(**Q-b**). *Would the treatment still be necessary and sufficient had the influence only existed via the mediator M?*

In many studies, researchers often face multiple mediators and study the causal effects along a specific pathway. Causal mediation analysis of the total effect $\mathbb{E}[Y_x - Y_{x'}]$ with two or more mediators has been studied across various fields, including statistics [Lin and VanderWeele, 2017, Miles et al., 2017, 2019, Zhou, 2022], AI [Avin et al., 2005, Shpitser and Pearl, 2008], medicine [Albert and Nelson, 2011, VanderWeele and Vansteelandt, 2014, Daniel et al., 2015, Vansteelandt and Daniel, 2017], political science [Zhou and Yamamoto, 2023], and cognitive science [Shpitser, 2013]. For example, Daniel et al. [2015] studied the effect of heavy drinking in the previous year on systolic blood pressure, mediated through body mass index (BMI) and gamma-glutamyl transpeptidase (GGT).

In this paper, we study the decomposition of PNS into path-specific components with two mediators and aim to answer the following four additional path-specific causal questions, incorporating an additional mediator N downstream of M:

(**Q-a1**). *Would the treatment still be necessary and sufficient had the influence via neither of the two mediators M and N existed?*
(**Q-a2**). *Would the treatment still be necessary and sufficient had the influence only existed via the second mediator N but not via the first mediator M?*
(**Q-b1**). *Would the treatment still be necessary and sufficient had the influence only existed via both the first mediator M and the second mediator N?*
(**Q-b2**). *Would the treatment still be necessary and*

sufficient had the influence only existed via the first mediator M but not via the second mediator N?

We provide definitions of path-specific PNS that enable the decomposition of the PNS into path-specific components. We then present an identification theorem for the path-specific PNS. Finally, we conduct numerical experiments to evaluate the finite-sample properties of the estimators and demonstrate their application using a real-world educational dataset.

2 BACKGROUNDS AND NOTATIONS

We represent a single or vector variable with a capital letter (X) and its realized value with a small letter (x). Let $\mathbb{I}(\cdot)$ be an indicator function that takes 1 if the statement in $(\cdot)$ is true and 0 otherwise, and $\mathbb{1}(\cdot)$ be a delta function. Denote Ω_Y be the domain of variable Y, $\mathbb{E}[Y]$ be the expectation of Y, $\mathbb{P}(Y \prec y)$ be the cumulative distribution function (CDF) of continuous variable Y, and $\mathfrak{p}_Y(y)$ be the probability density function (PDF) of continuous variable Y. We use $X \perp\!\!\!\perp Y|C$ to denote that X and Y are conditionally independent given C. We use $\preceq$ to denote a total order. In the univariate case, the total order $\preceq$ reduces to the standard order $\leq$. A formal definition of total order is given in Appendix A.

Structural causal models (SCM). We use the language of SCMs as our basic framework and follow the standard definition in the following [Pearl, 2009]. An SCM $\mathcal{M}$ is a tuple $\langle \boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, \mathbb{P}_{\boldsymbol{U}} \rangle$, where $\boldsymbol{U}$ is a set of exogenous (unobserved) variables following a distribution $\mathbb{P}_{\boldsymbol{U}}$, and $\boldsymbol{V}$ is a set of endogenous (observable) variables whose values are determined by structural functions $\mathcal{F} = \{f_{V_i}\}_{V_i \in \boldsymbol{V}}$ such that $v_i := f_{V_i}(\mathbf{pa}_{V_i}, \boldsymbol{u}_{V_i})$ where $\mathbf{PA}_{V_i} \subseteq \boldsymbol{V}$ and $\boldsymbol{U}_{V_i} \subseteq \boldsymbol{U}$. Each SCM $\mathcal{M}$ induces an observational distribution $\mathbb{P}_{\boldsymbol{V}}$ over $\boldsymbol{V}$, and a causal graph $G(\mathcal{M})$ in which there exists a directed edge from every variable in $\mathbf{PA}_{V_i}$ and $\boldsymbol{U}_{V_i}$ to V_i. An intervention of setting a set of endogenous variables $\boldsymbol{X}$ to constants $\boldsymbol{x}$, denoted by $do(\boldsymbol{x})$, replaces the original equations of $\boldsymbol{X}$ by the constants $\boldsymbol{x}$ and induces a *sub-model* $\mathcal{M}_{\boldsymbol{x}}$. We denote the potential outcome Y under intervention $do(\boldsymbol{x})$ by $Y_{\boldsymbol{x}}(\boldsymbol{u})$, which is the solution of Y in the sub-model $\mathcal{M}_{\boldsymbol{x}}$ given $\boldsymbol{U} = \boldsymbol{u}$.

Probabilities of causation (PoC) and mediation analysis for PoC. Kawakami et al. [2024] consider the following SCM:

$$Y := f_Y(X, C, U^Y), X := f_X(C, U^X), C := f_C(U^C), \tag{1}$$

where all variables can be vectors, and U^X, U^C, and U^Y are latent exogenous variables, and defined the (multivariate conditional) PoC for vectors of continuous or discrete variables as follows:

Definition 2.1 (PNS with Evidence)**.** *[Kawakami et al., 2024] The PNS with evidence is defined as*

$$\text{PNS}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x | \mathcal{E}, C = c).$$

In the above definition, C and $\mathcal{E}$ represent the information used to characterize a specific targeted subpopulation. C consists of subjects' pre-treatment covariates, and $\mathcal{E}$ contains information about post-treatment variables, commonly referred to as evidence. $\text{PNS}(y; x', x, \mathcal{E}, c)$ provides a measure of the necessity and sufficiency of x w.r.t. x' to produce $Y \succeq y$ given $C = c$ and evidence $\mathcal{E}$, that is, when X is set to $X = x$, the event $Y \succeq y$ occurs; when X is set to $X = x'$, the event $Y \succeq y$ does not occur. Note that PNS with $\mathcal{E} = (y \leq Y, X = x)$ coincides with the probability of necessity (PN), and PNS with $\mathcal{E} = (Y < y, X = x')$ coincides with the probability of sufficiency (PS) [Kawakami and Tian, 2025].

We will often call PNS *total PNS (T-PNS)* and denote it by $\text{T-PNS}(y; x', x, \mathcal{E}, c)$ for convenience.

Recently, Kawakami and Tian [2025] considered the following SCM $\mathcal{M}_1$, corresponding to the causal graph in Figure 1:

$$\begin{aligned} Y := f_Y(X, M, C, U^Y), M := f_M(X, C, U^M), \\ X := f_X(C, U^X), C := f_C(U^C), \end{aligned} \tag{2}$$

where all variables can be vectors, and U^X, U^C, U^Y, and U^M are latent exogenous variables.

Kawakami and Tian [2025] defined the (conditional) controlled direct, natural direct, and natural indirect probabilities of necessity and sufficiency with evidence.

Definition 2.2 (CD-PNS, ND-PNS, and NI-PNS with Evidence)**.** *The controlled direct, natural direct, and natural indirect PNS (CD-PNS, ND-PNS, and NI-PNS) with evidence w.r.t. M are defined by* $\text{CD-PNS}(y; x', x, m, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x',m} \prec y \preceq Y_{x,m} | \mathcal{E}, C = c)$*,* $\text{ND-PNS}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y | \mathcal{E}, C = c)$*, and* $\text{NI-PNS}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x} | \mathcal{E}, C = c)$*.*

ND-PNS and NI-PNS can answer the causal questions (Q-a) and (Q-b), respectively. T-PNS is decomposed as ND-PNS + NI-PNS. However, their applicability is restricted to the single mediator case, limiting their ability to capture more complex mediation pathways involving additional mediators.

Causal mediation analysis for two mediators. Researchers often consider the following SCM $\mathcal{M}_2$ with two mediators, corresponding to the causal graph in Figure 2:

$$\begin{aligned} Y := f_Y(X, M, N, C, U^Y), N := f_N(X, M, C, U^N), \\ M := f_M(X, C, U^M), X := f_X(C, U^X), C := f_C(U^C), \end{aligned} \tag{3}$$

where all variables can be vectors, and U^X, U^C, U^Y, U^M, and U^N are latent exogenous variables. We assume that the

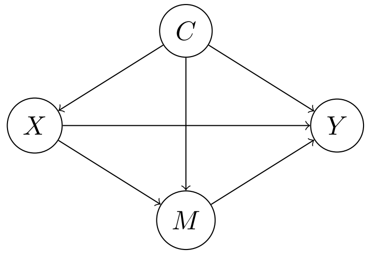


Figure 1: A causal graph representing SCM $\mathcal{M}_1$.

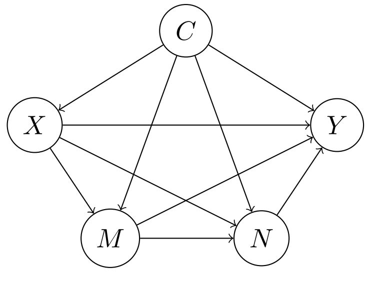


Figure 2: A causal graph representing SCM $\mathcal{M}_2$.

domains Ω_Y and $\Omega_{U^Y} \times \Omega_{U^M} \times \Omega_{U^N}$ are totally ordered sets with $\preceq$. SCM $\mathcal{M}_2$ means that two mediators are causally ordered, or M is the cause of N.

One of the most widely used models in mediation analysis with multiple mediators is a linear SCM $\mathcal{M}^{L2}$ consisting of $Y := \alpha_0 + \alpha_1 X + \alpha_2 M + \alpha_3 N + \alpha_4 C + U^Y$, $N := \beta_0 + \beta_1 X + \beta_2 M + \beta_3 C + U^N$, $M := \gamma_0 + \gamma_1 X + \beta_3 C + U^M$, where $U^C \sim \mathcal{N}(0, \sigma_C)$, $U^X \sim \mathcal{N}(0, \sigma_X)$, $U^Y \sim \mathcal{N}(0, \sigma_Y)$, $U^M \sim \mathcal{N}(0, \sigma_M)$, $U^N \sim \mathcal{N}(0, \sigma_N)$, and they are mutually independent normal distributions. $\mathcal{N}(\mu, \sigma)$ means a normal distribution whose mean is μ and standard deviation is σ.

Then, Daniel et al. [2015] defined the natural path-specific causal effects for binary treatment using the expectation of the counterfactuals, e.g., $\mathbb{E}[Y_{1,M_1,N_{1,M_0}}] - \mathbb{E}[Y_{1,M_1,N_{0,M_0}}]$. They impose the following assumption to identify the path-specific causal effects.

Assumption 2.1. *The following conditional independence statements hold:* (1)$\{Y_{x,m,n}, M_x, N_{x,m}\} \perp\!\!\!\perp X | C = c$, (2)$\{Y_{x,m,n}, N_{x,m}\} \perp\!\!\!\perp M | C = c, X = x$, *and* (3)$Y_{x,m,n} \perp\!\!\!\perp N | C = c, X = x$, *for any* $m \in \Omega_M$, $n \in \Omega_N$, $x \in \Omega_X$, *and* $c \in \Omega_C$, *where* $\mathfrak{p}_{X|C}(x|c) > 0$, $\mathfrak{p}_{M|C,X}(m|c,x) > 0$, $\mathfrak{p}_{N|C,X,M}(n|c,x'',m') > 0$, *and* $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) > 0$ *for any* $m, m' \in \Omega_M$, $n \in \Omega_N$, $x \in \Omega_X$, *and* $c \in \Omega_C$.

These independence conditions hold when there are no unmeasured confounders (or bidirected edges) between $\{X, M, N\} \to Y$, $\{X, M\} \to N$, and $X \to M$. The consistency conditions $\{X, M, N\}$ on Y, X on M, and $\{X, M\}$ on N assumed in [Daniel et al., 2015] hold under SCM $\mathcal{M}_2$.

Lemma 2.1. *[Daniel et al., 2015] Under SCM* $\mathcal{M}_2$ *and Assumption 2.1, the conditional CDF of potential outcome* $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y | C = c)$ *is given by*

$$\begin{aligned}\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y | C = c) = \int_{\Omega_M} \int_{\Omega_M} \int_{\Omega_N} \\ \mathbb{P}(Y \prec y | X = x, M = m, N = n, C = c) \\ \times\, \mathfrak{p}_{N|C,X,M}(n|c,x'',m')\, \mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) \\ \times\, \mathfrak{p}_{M|C,X}(m|c,x')\, dndmdm' \qquad (4)\end{aligned}$$

for any $x, x', x'', x''' \in \Omega_X$, $y \in \Omega_Y$, *and* $c \in \Omega_C$.

This lemma does not imply the identification of $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y | C = c)$. Instead, it states that $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y | C = c)$ is identifiable if $\mathfrak{p}_{M_{x''}|C,M_{x'}}(m'|c,m)$ is known or identifiable. Appendix B presents the explicit form of the distribution of $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ derived under two simple SCMs.

Furthermore, [Daniel et al., 2015] showed three special cases in which $\mathfrak{p}_{M_{x''}|C,M_{x'}}(m'|c,m)$ is identifiable. First, if $x' = x'''$, then $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) = \mathbb{1}(m = m')$ holds. Second, if there exists no effect of M on N, then $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y | C = c) = \int_{\Omega_M} \int_{\Omega_M} \int_{\Omega_N} \mathbb{P}(Y \prec y | X = x, M = m, N = n, C = c) \mathfrak{p}_{N|C,X}(n|c,x'') \mathfrak{p}_{M|C,X}(m|c,x') dndm$ holds. Third, if we assume a specific model with Gaussian noise, i.e., $M | X, C \sim \mathcal{N}(f(X, C; \alpha), \sigma^2)$, then we can identify $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m)$, where $f(X, C; \alpha)$ represents a parametric model.

3 PATH-SPECIFIC PNS WITH TWO MEDIATORS

We can extend the definition proposed by Kawakami and Tian [2025] to a two-mediator setting by replacing the mediator M in Kawakami and Tian [2025] with the set $\{M, N\}$. However, this extension accounts for only two aggregated pathways: $X \to Y$ and $X \to \{M, N\} \to Y$. Alternatively, one may apply the definition proposed by Kawakami and Tian [2025] while ignoring N, treating it as an unobserved exogenous variable of Y. This approach accounts for only two marginalized pathways: $X \to Y$ and $X \to M \to Y$. In Figure 2, conditioning $C = c$, there exist four pathways between X and Y, $X \to Y$, $X \to N \to Y$, $X \to M \to N \to Y$, and $X \to M \to Y$.

In this paper, we propose a definition of path-specific PNSs that satisfies the key decomposition relationships, ensuring that the sum of all path-specific components equals T-PNS.

Definition 3.1 (Path-specific PNS for decomposition)**.** *We define four types of path-specific PNS with two mediators as follows:*

$$\text{PNS}^{X \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y,$$

$$Y_{x',M_x,N_{x,M_x}} \prec y|\mathcal{E}, C = c), \quad (5)$$

$$\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y, y \preceq Y_{x',M_x,N_{x,M_x}}|\mathcal{E}, C = c), \quad (6)$$

$$\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}, Y_{x',M_x,N_{x',M_{x'}}} \prec y|\mathcal{E}, C = c), \quad (7)$$

$$\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}, y \preceq Y_{x',M_x,N_{x',M_{x'}}}|\mathcal{E}, C = c), \quad (8)$$

where $\mathcal{E} \triangleq (X = x^e, Y \in \mathcal{I}_Y)$ *and* $\mathcal{I}_Y$ *is a half-open interval* $[y^l, y^u)$ *or a closed interval* $[y^l, y^u]$ *w.r.t.* $\prec$.

If $C = \emptyset$ and $\mathcal{E} = \emptyset$, they reduce to the path-specific PNS for the entire population. We provide pathway representations of $\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c)$, $\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c)$, $\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c)$, and $\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c)$ in Appendix C.

Remark. For binary outcome $Y \in \{0, 1\}$, the definitions reduce to $\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} = 0, Y_x = 1, Y_{x',M_x} = 0, Y_{x',M_x,N_{x,M_x}} = 0|\mathcal{E}, C = c)$, $\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} = 0, Y_x = 1, Y_{x',M_x} = 0, Y_{x',M_x,N_{x,M_x}} = 1|\mathcal{E}, C = c)$, $\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} = 0, Y_x = 1, Y_{x',M_x} = 1, Y_{x',M_x,N_{x',M_{x'}}} = 0|\mathcal{E}, C = c)$, and $\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} = 0, Y_x = 1, Y_{x',M_x} = 1, Y_{x',M_x,N_{x',M_{x'}}} = 1|\mathcal{E}, C = c)$.

The following decomposition relationships hold:

Proposition 3.1. *T-PNS can be decomposed as follows:*

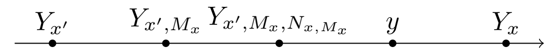
$$\begin{aligned}&\text{T-PNS}(y; x', x, \mathcal{E}, c)\\ &= \text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c)\\ &\quad + \text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c)\\ &\quad + \text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c). \quad (9)\end{aligned}$$

This decomposition property is essential for causal mediation analysis.

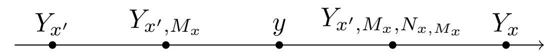
Proposition 3.2. *We have*

$$\begin{aligned}&\mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y|\mathcal{E}, C = c)\\ &= \text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c), \quad (10)\end{aligned}$$

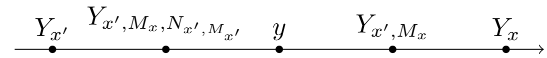
$$\begin{aligned}&\mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}|\mathcal{E}, C = c)\\ &= \text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c)\\ &\quad + \text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c). \quad (11)\end{aligned}$$



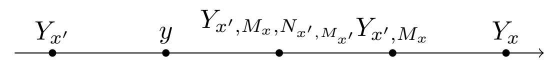
(a1) Order of potential outcomes in $\text{PNS}^{X\to Y}$.



(a2) Order of potential outcomes in $\text{PNS}^{X\to N\to Y}$.



(b1) Order of potential outcomes in $\text{PNS}^{X\to M\to N\to Y}$.



(b2) Order of potential outcomes in $\text{PNS}^{X\to M\to Y}$.

Figure 3: Order of potential outcomes.

The expressions $\mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y|\mathcal{E}, C = c)$ and $\mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}|\mathcal{E}, C = c)$ coincide with the definitions proposed by Kawakami and Tian [2025] by simply ignoring N, considering it as one of the unobserved exogenous variables of Y. We denote these quantities as follows: $\text{ND-PNS}^M(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y|\mathcal{E}, C = c)$ and $\text{NI-PNS}^M(y; x', x, \mathcal{E}, c) = \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}|\mathcal{E}, C = c)$.

First, $\text{PNS}^{X\to Y}$ has the following four counterfactuals:

(**A1**). *"had the treatment been x', the outcome would be $Y \prec y$"* ($Y_{x'} = Y_{x',M_{x'},N_{x',M_{x'}}} \prec y$),
(**A2**). *"had the treatment been x' but the first mediator was kept at the same value M_x, the outcome would be $Y \prec y$"* ($Y_{x',M_x} = Y_{x',M_x,N_{x',M_x}} \prec y$),
(**A3**). *"had the treatment been x' but the first mediator was kept at the same value M_x and the second mediator was kept at the same value N_{x,M_x}, the outcome would be $Y \prec y$"* ($Y_{x',M_x,N_{x,M_x}} \prec y$), and
(**A4**). *"had the treatment been x, the outcome would be $y \preceq Y$"* ($y \preceq Y_x = Y_{x,M_x,N_{x,M_x}}$).

The relative values of the potential outcomes are presented in Figure 3 (a1). Conditions (A1) and (A4) imply that $Y_{x'} \prec y \preceq Y_x$, which corresponds to the same condition in T-PNS and represents that the treatment x is necessary and sufficient w.r.t. x' to provoke the event $y \preceq Y$ given $C = c$. Conditions (A2) and (A4) imply that $Y_{x',M_x} \prec y \preceq Y_{x,M_x}$, which represents the necessity and sufficiency of x w.r.t. x' to produce $Y \succeq y$ given $C = c$, while keeping the values of the first mediator by the same as M_x. Conditions (A3) and (A4) imply that $Y_{x',M_x,N_{x,M_x}} \prec y \preceq Y_{x,M_x}$, which represents the necessity and sufficiency of x w.r.t. x' to produce $Y \succeq y$ given $C = c$, while keeping the values of the first mediator by the same as M_x and the second mediator by the same as N_{x,M_x}. Then, these conditions indicate that the treatment would be necessary and sufficient even if the influence only existed neither via the first or second mediators. $\text{PNS}^{X\rightarrow Y}$ answers the question (Q-a1).

Second, $\text{PNS}^{X\rightarrow N\rightarrow Y}$ has four counterfactuals:

(**B1**). *"had the treatment been x', the outcome would be $Y \prec y$"* ($Y_{x'} = Y_{x',M_{x'},N_{x',M_{x'}}} \prec y$),
(**B2**). *"had the treatment been x' but the first mediator was kept at the same value M_x, the outcome would be $Y \prec y$"* ($Y_{x',M_x} = Y_{x',M_x,N_{x',M_x}} \prec y$),
(**B3**). *"had the treatment been x' but the first mediator was kept at the same value M_x and the second mediator was kept at the same value N_{x,M_x}, the outcome would be $y \preceq Y$"* ($y \preceq Y_{x',M_x,N_{x,M_x}}$), and
(**B4**). *"had the treatment been x, the outcome would be $y \preceq Y$"* ($y \preceq Y_x = Y_{x,M_x,N_{x,M_x}}$).

The relative values of the potential outcomes are shown in Figure 3 (a2). Conditions (B1) and (B4) imply that $Y_{x'} \prec y \preceq Y_x$, which corresponds to the same condition in T-PNS and represents that the treatment x is necessary and sufficient w.r.t. x' to provoke the event $y \preceq Y$ given $C = c$. Conditions (B2) and (B4) imply that $Y_{x',M_x} \prec y \preceq Y_{x,M_x}$, which represents the necessity and sufficiency of x w.r.t. x' to produce $Y \succeq y$ given $C = c$, while keeping the values of the first mediator the same as M_x. Conditions (B2) and (B3) imply that $Y_{x',M_x} \prec y \preceq Y_{x',M_x,N_{x,M_x}}$, which represents the necessity and sufficiency of N_{x,M_x} w.r.t. N_{x',M_x} to produce $Y \succeq y$ given $C = c$, while keeping the values of the first mediator the same as M_x. Then, these conditions indicate that the treatment would be necessary and sufficient even if there existed only the influence via the second mediator, not via the first mediator. $\text{PNS}^{X\rightarrow N\rightarrow Y}$ answers the question (Q-a2).

Third, $\text{PNS}^{X\rightarrow M\rightarrow N\rightarrow Y}$ has four counterfactuals:

(**C1**). *"had the treatment been x', the outcome would be $Y \prec y$"* ($Y_{x'} = Y_{x',M_{x'},N_{x',M_{x'}}} \prec y$),
(**C2**). *"had the treatment been x' but the first mediator was kept at the same value M_x and the second mediator was kept at the same value $N_{x',M_{x'}}$, the outcome would be $Y \prec y$"* ($Y_{x',M_x,N_{x',M_{x'}}} \prec y$),
(**C3**). *"had the treatment been x' but the first mediator was kept at the same value M_x, the outcome would be $y \preceq Y$"* ($y \preceq Y_{x',M_x} = Y_{x',M_x,N_{x',M_x}}$), and
(**C4**). *"had the treatment been x, the outcome would be $y \preceq Y$"* ($y \preceq Y_x = Y_{x,M_x,N_{x,M_x}}$).

The relative values of the potential outcomes are shown in Figure 3 (b1). Conditions (C1) and (C4) imply that $Y_{x'} \prec y \preceq Y_x$, which corresponds to the same condition in T-PNS and states that the treatment x is necessary and sufficient w.r.t. x' to provoke the event $y \preceq Y$ given $C = c$. Conditions (C1) and (C3) imply that $Y_{x',M_{x'}} \prec y \preceq Y_{x',M_x}$, which represents the necessity and sufficiency of M_x w.r.t. $M_{x'}$ to produce $Y \succeq y$ given $C = c$, when the treatment is set to x'. Conditions (C2) and (C3) imply that $Y_{x',M_x,N_{x',M_{x'}}} \prec y \preceq Y_{x',M_x,N_{x',M_x}}$, which represents the necessity and sufficiency of N_{x',M_x} w.r.t. $N_{x',M_{x'}}$ to produce $Y \succeq y$ given $C = c$, while keeping the values of the first mediator fixed at M_x and setting the treatment to x'. Then, these conditions indicate that the treatment would be necessary and sufficient if there existed only the influence both via the first and the second mediators. $\text{PNS}^{X\rightarrow M\rightarrow N\rightarrow Y}$ answers the question (Q-b1).

Fourth, $\text{PNS}^{X\rightarrow M\rightarrow Y}$ has four counterfactuals:

(**D1**). *"had the treatment been x', the outcome would be $Y \prec y$"* ($Y_{x'} = Y_{x',M_{x'},N_{x',M_{x'}}} \prec y$),
(**D2**). *"had the treatment been x' but the first mediator was kept at the same value M_x and the second mediator was kept at the same value $N_{x',M_{x'}}$, the outcome would be $y \preceq Y$"* ($y \preceq Y_{x',M_x,N_{x',M_{x'}}}$),
(**D3**). *"had the treatment been x' but the first mediator was kept at the same value M_x, the outcome would be $y \preceq Y$"* ($y \preceq Y_{x',M_x} = Y_{x',M_x,N_{x',M_x}}$), and
(**D4**). *"had the treatment been x, the outcome would be $y \preceq Y$"* ($y \preceq Y_x = Y_{x',M_{x'},N_{x',M_{x'}}}$).

The relative values of the potential outcomes are shown in Figure 3 (b2). Conditions (D1) and (D4) imply that $Y_{x'} \prec y \preceq Y_x$, which corresponds to the same condition in T-PNS and states that the treatment x is necessary and sufficient w.r.t. x' to provoke the event $y \preceq Y$ given $C = c$. Conditions (D1) and (D3) imply that $Y_{x',M_{x'}} \prec y \preceq Y_{x',M_x}$, which represents the necessity and sufficiency of M_x w.r.t. $M_{x'}$ to produce $Y \succeq y$ given $C = c$, when the treatment is set to x'. Conditions (D1) and (D2) imply that $Y_{x',M_{x'},N_{x',M_{x'}}} \prec y \preceq Y_{x',M_x,N_{x',M_{x'}}}$, which represents the necessity and sufficiency of M_x w.r.t. $M_{x'}$ to produce $Y \succeq y$ given $C = c$, while keeping the values of the second mediator fixed at $N_{x',M_{x'}}$ and setting the treatment to x'. Then, these conditions indicate that the treatment would be necessary and sufficient if there existed only the influence via the first mediator, not via the second mediator. $\text{PNS}^{X\rightarrow M\rightarrow Y}$ answers the question (Q-b2).

Remark. Pearl [2001] defined the path-specific effects from the perspective of a *path-deactivation process*. In the two-mediator setting, each path-specific effect, from the perspective of a path-deactivation process, is expressed as: $Y_{x,M_{x'},N_{x',M_{x'}}} - Y_{x'}$, $Y_{x',M_{x'},N_{x,M_{x'}}} - Y_{x'}$, $Y_{x',M_{x'},N_{x',M_x}} - Y_{x'}$, and $Y_{x',M_x,N_{x',M_{x'}}} - Y_{x'}$, respectively. Similarly, one may define the four types of path-specific PNS with two mediators as follows: $\text{PNS'}^{X\rightarrow Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_{x,M_{x'},N_{x',M_{x'}}} | \mathcal{E}, C = c)$, $\text{PNS'}^{X\rightarrow N\rightarrow Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_{x',M_{x'},N_{x,M_{x'}}} | \mathcal{E}, C = c)$, $\text{PNS'}^{X\rightarrow M\rightarrow N\rightarrow Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec$

$y \preceq Y_{x',M_{x'},N_{x',M_x}}|\mathcal{E}, C = c)$, and $\text{PNS}'^{X\to M\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_{x',M_x,N_{x',M_{x'}}}|\mathcal{E}, C = c)$. However, these definitions do not satisfy the desired decomposition relationships. We adopt a different approach from his path-deactivation process in the context of T-PNS decomposition.

4 IDENTIFICATION OF PATH-SPECIFIC PNS WITH TWO MEDIATORS

In this section, we provide an identification theorem for the path-specific PNS for decomposition.

New identification assumption of the conditional PDF of potential outcomes. Identifying the conditional PDF of potential outcomes $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m)$ remains challenging in Lemma 2.1, and Daniel et al. [2015] addressed this issue by employing a specific parametric model.

We introduce a new nonparametric identification assumption for the counterfactual conditional PDF $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m)$ as follows.

Assumption 4.1 (Strictly monotonicity over f_M)**.** *The function f_M is either strictly monotonic increasing on U^M for all $x \in \Omega_X$ and $c \in \Omega_C$, or strictly monotonic decreasing on U^M for all $x \in \Omega_X$ and $c \in \Omega_C$, almost surely w.r.t. $\mathbb{P}_{U^M}$ with $\mathfrak{p}(u_M) > 0$.*

This assumption is similar to Assumption 3.5 (strict monotonicity over f_Y) in [Kawakami et al., 2024]. The parametric model used in [Daniel et al., 2015] satisfies this assumption. Then $\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m)$ is identifiable as follows.

Lemma 4.1. *Under SCM $\mathcal{M}_2$ and Assumption 4.1, we have*

$$\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) = \mathbb{1}(F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c)) = m'), \tag{12}$$

where $F^{-1}_{M|X=x''',C=c}$ is the inverse function of conditional CDF $\mathbb{P}(M \preceq m|X=x''',C=c)$ on m given $X = x'''$ and $C = c$.

Note that, under Assumption 4.1, $\mathbb{P}(M \preceq m|X = x''', C = c)$ is monotonically increasing and continuous in m. Consequently, an inverse function $F^{-1}_{M|X=x''',C=c}$ always exists. When $x''' = x'$, the expression $\mathbb{1}(F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X = x', C = c)) = m')$ reduces to $\mathbb{1}(m = m')$, which aligns with the result presented in Daniel et al. [2015].

We then have the following theorem:

Theorem 4.1. *Under SCM $\mathcal{M}_2$ and Assumption 4.1, for any $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y|C = c)$ is identifiable by $\theta(y; x, x', x'', x''', c)$, where*

$$\begin{aligned}\theta(y;x,x',x'',x''',c) = &\int_{\Omega_M}\int_{\Omega_N} \mathbb{P}(Y \prec y|X=x,M=m,N=n,C=c)\\ &\times \mathfrak{p}_{N|C,X,M}\big(n|c,x'',F^{-1}_{M|X=x''',C=c}\\ &\qquad(\mathbb{P}(M \preceq m|X=x',C=c))\big)\\ &\times \mathfrak{p}_{M|C,X}(m|c,x')\,dndm.\end{aligned} \tag{13}$$

Identification of path-specific PNS with two mediators. The identification of PoC relies on monotonicity assumptions, as discussed in the literature [Tian and Pearl, 2000, Kawakami et al., 2024, Kawakami and Tian, 2025]. We adopt similar assumptions to those in [Kawakami and Tian, 2025].

Assumption 4.2. *Potential outcome $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ has conditional PDF $\mathfrak{p}_{Y_{x,M_{x'},N_{x'',M_{x'''}}}|C=c}$ for each $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, and its support $\{y \in \Omega_Y : \mathfrak{p}_{Y_{x,M_{x'},N_{x'',M_{x'''}}}|C=c}(y) \neq 0\}$ is the same for each $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$.*

Assumption 4.2 is reasonable for continuous variables. For example, potential outcomes $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ often have $[-\infty, \infty]$ support, such as in linear SCM $\mathcal{M}^{L2}$.

Let a compounded function g be

$$\begin{aligned}g(x,x',x'',x''',c,\tilde{U}) \triangleq\ & f_Y(x, f_M(x',c,U^M),\\ & f_N(x'', f_M(x''',c,U^M),c,U^N),c,U^Y)\end{aligned} \tag{14}$$

for all $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, where $\tilde{U} = (U^Y, U^M, U^N)$. We assume the following for identifying path-specific PNS:

Assumption 4.3 (Monotonicity over g)**.** *The function $g(x, x', x'', x''', c, \tilde{U})$ is either monotonic increasing on $\tilde{U}$ for all $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, or monotonic decreasing on $\tilde{U}$ for all $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, almost surely w.r.t. $\mathbb{P}_{\tilde{U}}$.*

Or, alternatively, we assume the following:

Assumption 4.3' (Conditional monotonicity over $Y_{x,M_{x'},N_{x'',M_{x'''}}}$) *The potential outcomes $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ satisfy: for any $x, x', x'', x''', x^*, x^{**}, x^{***}, x^{****} \in \Omega_X$, $y \in \Omega_Y$, and $c \in \Omega_C$, either $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y \preceq Y_{x^*,M_{x^{**}},N_{x^{***},M_{x^{****}}}}|C = c) = 0$ or $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y \preceq Y_{x^*,M_{x^{**}},N_{x^{***},M_{x^{****}}}}|C = c) = 0$.*

Similarly to the results in Kawakami and Tian [2025], Assumptions 4.3 and 4.3' are equivalent under Assumption 4.2. We provide the proof of this statement in Appendix D. We note that both the linear SCM $\mathcal{M}^{L2}$ and the nonlinear SCM with normal distribution $\mathcal{M}^{N}$ satisfy Assumption 4.3 with $\tilde{U} = U^Y + U^M + U^N$.

Then, the path-specific PNSs are identifiable as follows.

Theorem 4.2 (Identification of path-specific PNS). *Let $\mathcal{I}_Y$ be a half-open interval in evidence $\mathcal{E}$. We have the following two statements.*

(1). *Under SCM $\mathcal{M}_2$ and Assumptions 2.1, 4.2 and 4.3, for any $x', x \in \Omega_X$, $y \in \Omega_Y$, and $c \in \Omega_C$, we have*

(1A). *If $\mathbb{P}(Y \prec y^l|X = x^e, C = c) \neq \mathbb{P}(Y \prec y^u|X = x^e, C = c)$, then*

$$\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^1/\delta, 0\}, \quad (15)$$

$$\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^2/\delta, 0\}, \quad (16)$$

where

$$\gamma^1 = \min\big\{\theta(y; x', x', x', x', c), \theta(y; x', x, x', x, c), \theta(y; x', x, x, x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c)\big\} - \max\big\{\theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c)\big\}, \quad (17)$$

$$\gamma^2 = \min\big\{\theta(y; x', x', x', x', c), \theta(y; x', x, x', x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c)\big\} - \max\big\{\theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c), \theta(y; x', x, x, x, c)\big\}, \quad (18)$$

$$\delta = \mathbb{P}(Y \prec y^u|X = x^e, C = c) - \mathbb{P}(Y \prec y^l|X = x^e, C = c). \quad (19)$$

(1B). *If $\mathbb{P}(Y \prec y^l|X = x^e, C = c) = \mathbb{P}(Y \prec y^u|X = x^e, C = c)$, then*

$$\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\big(\theta(y; x', x', x', x', c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c), \theta(y; x', x, x', x, c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c), \theta(y; x', x, x, x, c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c)\big), \quad (20)$$

$$\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\big(\theta(y; x', x', x', x', c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c), \theta(y; x', x, x', x, c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x, x, c)\big). \quad (21)$$

(2). *Under SCM $\mathcal{M}_2$ and Assumptions 2.1, 4.1, 4.2 and 4.3, for any $x', x \in \Omega_X$, $y \in \Omega_Y$, and $c \in \Omega_C$, we have*

(2A). *If $\mathbb{P}(Y \prec y^l|X = x^e, C = c) \neq \mathbb{P}(Y \prec y^u|X = x^e, C = c)$, then*

$$\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^3/\delta, 0\}, \quad (22)$$

$$\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^4/\delta, 0\}, \quad (23)$$

where

$$\gamma^3 = \min\big\{\theta(y; x', x', x', x', c), \mathbb{P}(Y \prec y^u|X = x^e, C = c), \theta(y; x', x, x', x', c)\big\} - \max\big\{\theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c), \theta(y; x', x, x', x, c)\big\}, \quad (24)$$

$$\gamma^4 = \min\big\{\theta(y; x', x', x', x', c), \mathbb{P}(Y \prec y^u|X = x^e, C = c)\big\} - \max\big\{\theta(y; x, x, x, x, c), \theta(y; x', x, x', x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c), \theta(y; x', x, x', x', c)\big\}. \quad (25)$$

(2B). *If $\mathbb{P}(Y \prec y^l|X = x^e, C = c) = \mathbb{P}(Y \prec y^u|X = x^e, C = c)$, then*

$$\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\big(\theta(y; x', x', x', x', c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x, c), \theta(y; x', x, x', x', c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c)\big), \quad (26)$$

$$\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\big(\theta(y; x', x', x', x', c) \leq \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x', c)\big). \quad (27)$$

Only the identifications of $\text{PNS}^{X\to M\to N\to Y}$ and $\text{PNS}^{X\to M\to Y}$ require Assumption 4.1 since all counterfactuals in $\text{PNS}^{X\to N\to Y}$ and $\text{PNS}^{X\to Y}$ satisfy $x' = x'''$, which corresponds to a special case in Daniel et al. [2015].

Remark. When $\mathcal{I}_Y$ is a closed interval $[y^l, y^u]$ in evidence $\mathcal{E}$, the identification results are obtained by replacing "$Y \prec y^u$" with "$Y \preceq y^u$" in Theorem 4.2. By setting $y^u = \infty$ and $y^l = -\infty$, the identification theorem for the path-specific PNS can be derived for the entire population in the absence of evidence ($\mathcal{E} = \emptyset$).

5 NUMERICAL EXPERIMENTS

We conduct numerical experiments to illustrate the finite-sample properties of estimators for the path-specific PNS.

Estimation methods. The path-specific PNSs for decomposition under SCM $\mathcal{M}^{L2}$ are estimable using simple linear regressions. $\theta(y; x, x', x'', x''', c)$ in Theorem 4.1 is estimated by $\mathbb{P}(Z < y)$, where $Z \sim \mathcal{N}(\hat{\alpha}_0 + \hat{\alpha}_1 x + \hat{\alpha}_2(\hat{\gamma}_0 + \hat{\gamma}_1 x' + \hat{\gamma}_2 c) + \hat{\alpha}_3(\hat{\beta}_0 + \hat{\beta}_1 x'' + \hat{\beta}_2(\hat{\gamma}_0 + \hat{\gamma}_1 x''' + \hat{\gamma}_2 c) + \hat{\beta}_3 c) + \hat{\alpha}_4 c, \{(\hat{\alpha}_2 + \hat{\alpha}_3\hat{\beta}_2)^2\hat{\sigma}_M^2 + \hat{\alpha}_3^2\hat{\sigma}_N^2 + \hat{\sigma}_Y^2\}^{1/2})$ and $\{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\sigma}_Y, \hat{\sigma}_M, \hat{\sigma}_N\}$ are the estimated parameters of the three linear regressions, $Y \sim X + M + N$, $N \sim X + M$, and $M \sim X$.

Setting. We consider the following SCM:

$$Y := X + M + N + C + U^Y, N := X + M + C + U^N, \quad M := X + C + U^M, X := C + U^X, C := U^C, \quad (28)$$

where $U^C \sim \mathcal{N}(0, 1)$, $U^X \sim \mathcal{N}(0, 1)$, $U^Y \sim \mathcal{N}(0, 1)$, $U^M \sim \mathcal{N}(0, 1)$, $U^N \sim \mathcal{N}(0, 1)$, which are mutually independent normal distributions. This SCM satisfies Assumptions 2.1, 4.1, 4.2, 4.3, and 4.3'. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$.

Table 1: Results of numerical experiments. We present the estimates of T-PNS, ND-PNS^M, NI-PNS^M, PNS$^{X\to Y}$, PNS$^{X\to N\to Y}$, PNS$^{X\to M\to N\to Y}$, and PNS$^{X\to M\to Y}$. Additionally, we report the mean of each estimator accompanied by their 95% confidence interval.

Estimators	$N=20$	$N=100$	$N=10000$	Ground Truth
T-PNS	0.447 ($[0.286, 0.625]$)	0.448 ($[0.385, 0.519]$)	0.449 ($[0.443, 0.455]$)	0.449
ND-PNS^M	0.153 ($[0.040, 0.343]$)	0.155 ($[0.103, 0.214]$)	0.156 ($[0.150, 0.161]$)	0.156
NI-PNS^M	0.296 ($[0.154, 0.443]$)	0.293 ($[0.231, 0.355]$)	0.293 ($[0.282, 0.299]$)	0.293
PNS$^{X\to Y}$	0.057 ($[0.003, 0.153]$)	0.059 ($[0.032, 0.093]$)	0.059 ($[0.056, 0.062]$)	0.059
PNS$^{X\to N\to Y}$	0.095 ($[0.015, 0.251]$)	0.097 ($[0.059, 0.144]$)	0.097 ($[0.093, 0.101]$)	0.097
PNS$^{X\to M\to N\to Y}$	0.133 ($[0.037, 0.287]$)	0.134 ($[0.096, 0.180]$)	0.134 ($[0.130, 0.139]$)	0.135
PNS$^{X\to M\to Y}$	0.163 ($[0.031, 0.319]$)	0.160 ($[0.108, 0.215]$)	0.158 ($[0.153, 0.164]$)	0.158

Results. We present the results of each estimator in Table 1. All means of the estimates are close to the ground truth. However, for small sample sizes, the estimators exhibit large 95 % CIs, indicating higher variability in estimation.

We provide three additional experiments in Appendix F.1 under the following conditions: (1) no effect between M and N, (2) no effect between $\{M, N\}$ and Y, and (3) only effect through $X\to M\to N\to Y$. In the setting (1), PNS$^{X\to M\to N\to Y}$ is equal to 0. In the setting (2), PNS$^{X\to N\to Y}$, PNS$^{X\to M\to N\to Y}$, and PNS$^{X\to M\to Y}$ are all equal to 0. In the setting (3), PNS$^{X\to Y}$, PNS$^{X\to N\to Y}$, and PNS$^{X\to M\to Y}$ are all equal to 0. These results offer intuitive decompositions of T-PNS.

We provide a sensitivity analysis on the monotonicity assumption in Appendix F.2 by introducing a non-monotonic term in SCM, i.e., $Y := X + M + N + C + \alpha U^Y + (1-\alpha)(U^Y)^4$, where $\alpha \in [0, 1]$ controls the degree of the violation of the monotonicity. We observe that the magnitude of bias increases with greater violations of the monotonicity. We additionally report experimental results using logistic regression for binary outcomes in Appendix F.3. The estimates obtained from logistic regression are reliable when the sample size is large.

6 APPLICATION TO REAL-WORLD

We present an application using a real-world dataset.

Dataset. We use an open dataset from (`https://archive.ics.uci.edu/dataset/320/student+performance`) on student performance in mathematics from secondary education in two Portuguese schools. Secondary education lasts for three years, and students are tested once per year, resulting in a total of three tests. The dataset includes attributes related to demographics, social factors, school-related features, and student grades. The sample size is 649, with no missing values. Prior research using this dataset aimed to predict students' performance based on their attributes [Cortez and Silva, 2008, Helwig, 2017]. Kawakami et al. [2024] assess the causal relationship between the students' performance, study time, and extra paid classes via estimating PoC. In this paper, we analyze the causal relationship between students' performance in the final period, study time, and extra paid classes, considering their performance in the first and second periods as mediators.

Variables. We consider the mathematics scores in the first period (M) and the second period (N) as mediators, while the mathematics score in the final period (Y) serves as the outcome variable. They take values from 0 to 20, respectively. We note that Kawakami et al. [2024] considered all mathematics scores from the first period, the second period, and the final period as the vector of outcome variables. We consider "study time in a week" (X^1) and "extra paid classes within the course subject" (X^2) (where yes: $X^2 = 2$, no: $X^2 = 1$) as the treatment variables, denoted as $X = (X^1, X^2)$. We select "sex," "failures," "schoolsup," "famsup," and "goout" as the covariates (C), following their selection in [Helwig, 2017] and note that they are also used in [Kawakami et al., 2024]. We estimate the path-specific PNS using linear regression models, as described in Section 5. We conduct 1,000 bootstrap resampling iterations [Efron, 1979] to examine the distribution of the estimators.

In this dataset, Assumption 4.3', for instance, $\mathbb{P}(Y_{x,M_{x'},N_{x,M_x}} \prec y \preceq Y_{x,M_{x'},N_{x,M_{x'}}} | C = c) = 0$ is reasonable. This is because the scores in the first period, had she studied four hours a week and taken extra classes, appear to be greater than those in the first period had she studied only one hour a week and taken no extra classes. Moreover, if her scores in the first had been higher, the scores in the second periods would also have been higher. Furthermore, if her scores in the second had been higher, the scores in the third periods would also have been higher. Assumption 4.1 is also reasonable, as, for example, if U^M represents a genetic factor influencing mathematics performance, it can exert a strictly monotonic increasing effect on the scores in the first period.

Results. We consider the subject whose ID number is 1 and set the values of her covariates as c_1. We define the treatment values as $x' = (1, 1)$, $x = (4, 2)$, set the outcome threshold to $y = 10$, and let the evidence be $\mathcal{E} = \emptyset$. The estimates of T-PNS, ND-PNS^M, and NI-PNS^M at $(y; x', x, \mathcal{E}, c_1)$ are 15.259% (CI : $[0.000\%, 33.022\%]$), 1.032 %(CI : $[0.000\%, 7.452\%]$), and 14.226 %(CI : $[0.000\%, 2.304\%]$), respectively. The results indicate that studying 4 hours a week and taking extra classes would be necessary and sufficient to achieve a score above 10 in the final period for 15.259 % of subjects. Moreover, when ignoring the scores in the second period, studying 4 hours a week and taking extra classes would remain necessary and sufficient at the same level of T-PNS if the influence existed solely through the scores in the first period.

We ask four further causal questions about this dataset, considering the scores in the first and second periods:

(**Q-a1'**). *Would studying 4 hours a week and taking extra classes still be necessary and sufficient to achieve a score above 10 in the final period if the influence through the scores in the first and second periods had not existed?*
(**Q-a2'**). *Would studying 4 hours a week and taking extra classes still be necessary and sufficient to achieve a score above 10 in the final period if the influence occurred only through the scores in the second period, and not through the scores in the first period?*
(**Q-b1'**). *Would studying 4 hours a week and taking extra classes still be necessary and sufficient to achieve a score above 10 in the final period if the influence occurred only through both the scores in the first and second periods?*
(**Q-b2'**). *Would studying 4 hours a week and taking extra classes still be necessary and sufficient to achieve a score above 10 in the final period if the influence occurred only through the scores in the first period, and not through the scores in the second period?*

Then, the estimates of PNS$^{X\to Y}$, PNS$^{X\to N\to Y}$, PNS$^{X\to M\to N\to Y}$, and PNS$^{X\to M\to Y}$ at $(y; x', x, \mathcal{E}, c_1)$ are

$$\begin{aligned}
&\text{PNS}^{X\to Y}: && 0.149\%(\text{CI} : [0.000\%, 2.304\%]),\\
&\text{PNS}^{X\to N\to Y}: && 0.883\%(\text{CI} : [0.000\%, 6.239\%]),\\
&\text{PNS}^{X\to M\to N\to Y}: && 0.000\%(\text{CI} : [0.000\%, 0.000\%]),\\
&\text{PNS}^{X\to M\to Y}: && 14.226\%(\text{CI} : [0.000\%, 29.405\%]).
\end{aligned}$$

The results suggest that the necessity and sufficiency of the treatment are almost entirely attributed to the indirect influence occurring solely through the first mediator, while accounting for the path through the score in the second period N.

Additionally, we provide the estimates under the evidence condition $\mathcal{E} = (X = 0, 10 \leq Y < 15)$, while maintaining the same settings as in Appendix G. Similar results are observed for the subpopulations defined by the specified evidence.

7 CONCLUSION

We study path-specific PNSs with two mediators to answer the causal questions (Q-a1), (Q-a2), (Q-b1), and (Q-b2) by decomposing T-PNS into its path-specific components. Researchers sometimes analyze causal models with more than three mediators. We provide the definitions of the path-specific PNS with three mediators in Appendix E. The definitions of path-specific PNS along an arbitrary path in an arbitrary causal graph that satisfy the decomposition relationships will be left for future research.

In the settings where the monotonicity assumption does not hold, we can aim to derive bounds for the path-specific PNS [Tian and Pearl, 2000, Li and Pearl, 2024]. One approach is to use Fréchet inequalities [Fréchet, 1960]. Deriving bounds for the path-specific PNS will be a future work.

ACKNOWLEDGEMENTS

The authors thank the anonymous reviewers for their time and thoughtful comments.

References

Jeffrey M Albert and Suchitra Nelson. Generalized causal mediation analysis. *Biometrics*, 67(3):1028–1038, Sep 2011.

Chen Avin, Ilya Shpitser, and Judea Pearl. Identifiability of path-specific effects. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, IJCAI'05, page 357–363, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51 (6):1173, 1986.

P. Cortez and A. M. Gonçalves Silva. Using data mining to predict secondary school student performance. In *Proceedings of 5th Annual Future Business Technology Conference*, pages 5–12, 2008.

R M Daniel, B L De Stavola, S N Cousens, and S Vansteelandt. Causal mediation analysis with multiple mediators. *Biometrics*, 71(1):1–14, Mar 2015.

A. Philip Dawid, Rossella Murtas, and Monica Musio. Bounding the probability of causation in mediation analysis. *ArXiv*, abs/1411.2636, 2014.

B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26, 1979.

Maurice Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut international de statistique*, pages 10–32, 1960.

E. Harzheim. *Ordered Sets*. Advances in Mathematics - Kluwer Academic Publishers. Springer, 2005.

Nathaniel E Helwig. Adding bias to reduce variance in psychological results: A tutorial on penalized regression. *The Quantitative Methods for Psychology*, 13(1):1–19, 2017.

Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychol Methods*, 15(4):309–334, Dec 2010a.

Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*, 25(1):51 – 71, 2010b.

Yuta Kawakami and Jin Tian. Mediation analysis for probabilities of causation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 26823–26832, 2025.

Yuta Kawakami, Ryusei Shingaki, and Manabu Kuroki. Identification and estimation of the probabilities of potential outcome types using covariate information in studies with non-compliance. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(10):12234–12242, Jun. 2023.

Yuta Kawakami, Manabu Kuroki, and Jin Tian. Probabilities of causation for continuous and vector variables. *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI-2024)*, 2024.

Manabu Kuroki and Zhihong Cai. Statistical analysis of 'probabilities of causation' using co-variate information. *Scandinavian Journal of Statistics*, 38(3):564–577, 2011.

Ang Li and Judea Pearl. Probabilities of causation with nonbinary treatment and effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20465–20472, 2024.

Sheng-Hsuan Lin and Tyler VanderWeele. Interventional approach for path-specific effects. *Journal of Causal Inference*, 5(1):20150027, 2017.

C H Miles, I Shpitser, P Kanki, S Meloni, and E J Tchetgen Tchetgen. On semiparametric estimation of a path-specific effect in the presence of mediator-outcome confounding. *Biometrika*, 107(1):159–172, 11 2019.

Caleb H Miles, Ilya Shpitser, Phyllis Kanki, Seema Meloni, and Eric J Tchetgen Tchetgen. Quantifying an adherence path-specific effect of antiretroviral therapy in the nigeria pepfar program. *J Am Stat Assoc*, 112(520):1443–1452, 2017.

Rossella Murtas, Alexander Philip Dawid, and Monica Musio. New bounds for the probability of causation in mediation analysis. *arXiv: Statistics Theory*, 2017.

Judea Pearl. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, page 411–420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.

J M Robins and S Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2): 143–155, Mar 1992.

James Robins and Sander Greenland. The probability of causation under a stochastic model for individual risk. *Biometrics*, 45(4):1125–1138, 1989.

Max Rubinstein, Maria Cuellar, and Daniel Malinsky. Mediated probabilities of causation. *Journal of Causal Inference*, 13(1):20240019, 2025. doi: 10.1515/jci-2024-0019.

Donghee Shin. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human-Computer Studies*, 146:102551, 2021.

Ryusei Shingaki and Manabu Kuroki. Identification and estimation of joint probabilities of potential outcomes in observational studies with covariate information. In *Advances in Neural Information Processing Systems*, volume 34, pages 26475–26486. Curran Associates, Inc., 2021.

Ilya Shpitser. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science*, 37(6):1011–1035, 2013.

Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. *J. Mach. Learn. Res.*, 9: 1941–1979, jun 2008.

Eric J. Tchetgen Tchetgen and Ilya Shpitser. Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness and sensitivity analysis. *The Annals of Statistics*, 40(3):1816 – 1845, 2012.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1):287–313, 2000.

T J VanderWeele and S Vansteelandt. Mediation analysis with multiple mediators. *Epidemiol Methods*, 2(1):95–115, Jan 2014.

Stijn Vansteelandt and Rhian M Daniel. Interventional effects for mediation analysis with multiple mediators. *Epidemiology*, 28(2):258–265, Mar 2017.

Sewall Wright. Correlation and causation. *Journal of agricultural research*, 20(7):557–585, 1921.

Sewall Wright. The method of path coefficients. *The annals of mathematical statistics*, 5(3):161–215, 1934.

Xiang Zhou. Semiparametric estimation for causal mediation analysis with multiple causally ordered mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(3):794–821, 2022.

Xiang Zhou and Teppei Yamamoto. Tracing causal paths from experimental and observational data. *The Journal of Politics*, 85(1):250–265, 2023.

APPENDIX TO "DECOMPOSITION OF PROBABILITIES OF CAUSATION WITH TWO MEDIATORS"

A TOTAL ORDERS

We show the definition of the total order used in the body of the paper. The definition of the total order is as follows [Harzheim, 2005]:

Definition A.1 (Total order). *A total order on a set Ω is a relation "$\preceq$" on Ω satisfying the following four conditions for all $a_1, a_2, a_3 \in \Omega$:*

1. $a_1 \preceq a_1$*;*
2. *if* $a_1 \preceq a_2$ *and* $a_2 \preceq a_3$ *then* $a_1 \preceq a_3$*;*
3. *if* $a_1 \preceq a_2$ *and* $a_2 \preceq a_1$ *then* $a_1 = a_2$*;*
4. *at least one of* $a_1 \preceq a_2$ *and* $a_2 \preceq a_1$ *holds.*

In this case we say that the ordered pair $(\Omega, \preceq)$ is a totally ordered set. The inequality $a \preceq b$ of total order means $a \prec b$ or $a = b$, and the relationship $\neg(a \preceq b) \Leftrightarrow a \succ b$ holds for a totally ordered set, where $\neg$ means the negation.

B MARGINAL DISTRIBUTION OF COUNTERFACTUAL UNDER SPECIAL CASES

Linear SCM. We provide explicit forms of path-specific PNS under linear SCM $\mathcal{M}^{L2}$, i.e.,

$$\begin{aligned} Y &:= \alpha_0 + \alpha_1 X + \alpha_2 M + \alpha_3 N + \alpha_4 C + U^Y, \\ N &:= \beta_0 + \beta_1 X + \beta_2 M + \beta_3 C + U^N, M := \gamma_0 + \gamma_1 X + \beta_3 C + U^M, \end{aligned} \tag{29}$$

where $U^C \sim \mathcal{N}(0, \sigma_C)$, $U^X \sim \mathcal{N}(0, \sigma_X)$, $U^Y \sim \mathcal{N}(0, \sigma_Y)$, $U^M \sim \mathcal{N}(0, \sigma_M)$, $U^N \sim \mathcal{N}(0, \sigma_N)$, and they are mutually independent normal distributions. $\mathcal{N}(\mu, \sigma)$ means a normal distribution whose mean is μ and standard deviation is σ. The potential outcome $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ given $C = c$ is explicitly expressed as:

$$\begin{aligned} Y_{x,M_{x'},N_{x'',M_{x'''}}} &= \alpha_0 + \alpha_1 x + \alpha_2(\gamma_0 + \gamma_1 x' + \beta_3 c + U^M) \\ &\quad + \alpha_3(\beta_0 + \beta_1 x'' + \beta_2(\gamma_0 + \gamma_1 x''' + \beta_3 c + U^M) + \beta_3 c + U^N) + \alpha_4 c + U^Y. \end{aligned} \tag{30}$$

Then, using a Gaussian distribution $Z \sim \mathcal{N}(0, 1)$, the counterfactual $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ given $C = c$ is expressed as

$$\begin{aligned} &(\alpha_0 + \alpha_1 x + \alpha_2(\gamma_0 + \gamma_1 x' + \gamma_2 c) + \alpha_3(\beta_0 + \beta_1 x'' + \beta_2(\gamma_0 + \gamma_1 x''' + \gamma_2 c) + \beta_3 c) + \alpha_4 c) \\ &\quad + \{(\alpha_2 + \alpha_3\beta_2)^2\sigma_M^2 + \alpha_3^2\sigma_N^2 + \sigma_Y^2\}^{1/2} Z. \end{aligned} \tag{31}$$

Then, we have

$$\begin{aligned} &\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} < y | C = c) \\ &= \mathbb{P}((\alpha_0 + \alpha_1 x + \alpha_2(\gamma_0 + \gamma_1 x' + \gamma_2 c) + \alpha_3(\beta_0 + \beta_1 x'' + \beta_2(\gamma_0 + \gamma_1 x''' + \gamma_2 c) + \beta_3 c) + \alpha_4 c) \\ &\quad + \{(\alpha_2 + \alpha_3\beta_2)^2\sigma_M^2 + \alpha_3^2\sigma_N^2 + \sigma_Y^2\}^{1/2} Z < y). \end{aligned} \tag{32}$$

The parameters $\{\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_1, \beta_2, \beta_3, \gamma_0, \gamma_1, \sigma_M, \sigma_N, \sigma_Y\}$ are identified through three regressions $M \sim X + C$, $N \sim X + M + C$, and $Y \sim X + M + N + C$. Since SCM $\mathcal{M}^{L2}$ satisfies 2.1, 4.1, 4.2 and 4.3, the path-specific PNS are identified via Theorem 4.2, based on the marginal distribution of $Y_{x,M_{x'},N_{x'',M_{x'''}}}$.

Binary Outcome Case with Logistic Model. When the outcome is binary and modeled using logistic regression by

$$\mathbb{P}(Y = 1) = \frac{1}{1 + \exp(-(\alpha_0 + \alpha_1 X + \alpha_2 M + \alpha_3 N + \alpha_4 C))}, \tag{33}$$

$$N := \beta_0 + \beta_1 X + \beta_2 M + \beta_3 C + U^N, M := \gamma_0 + \gamma_1 X + \beta_3 C + U^M, \tag{34}$$

then, the distribution of potential outcomes $Y_{x,M_{x'},N_{x'',M_{x'''}}}$ given $C=c$, $U^M=u^M$ and $U^N=u^N$ is explicitly expressed as:

$$\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}}=1|C=c,U^M=u^M,U^N=u^N) \tag{35}$$
$$=\Big\{1+\exp\Big(-(\alpha_0+\alpha_1 x+\alpha_2(\gamma_0+\gamma_1 x'+\beta_3 c+u_M) \tag{36}$$
$$+\alpha_3(\beta_0+\beta_1 x''+\beta_2(\gamma_0+\gamma_1 x'''+\beta_3 c+u_M)+\beta_3 c+u_N)+\alpha_4 c)\Big)\Big\}^{-1}. \tag{37}$$

The parameters $\{\alpha_0,\alpha_1,\alpha_2,\alpha_3,\alpha_4,\beta_0,\beta_1,\beta_2,\beta_3,\gamma_0,\gamma_1\}$ are identified through regressions $M\sim X+C$ and $N\sim M+X+C$. Since U^N and U^M are identified as residuals of two regressions $M\sim X+C$ and $N\sim M+X+C$, $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}}=1|C=c)$ is given by $\mathbb{E}_{(U^M,U^N)}[\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}}=1|C=c,U^M,U^N)]$. Since this SCM satisfies 2.1, 4.1, 4.2 and 4.3, the path-specific PNS are identified via Theorem 4.2, based on the marginal distribution of $Y_{x,M_{x'},N_{x'',M_{x'''}}}$.

C PATHWAY REPRESENTATIONS

We provide pathway representations of $\text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c)$, $\text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c)$, $\text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c)$, and $\text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c)$ in Figures 4, 5, 6, and 7, respectively.

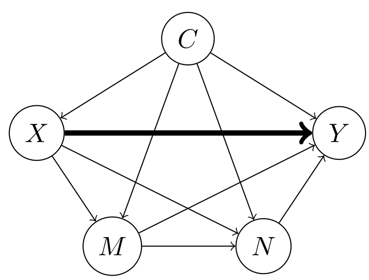


Figure 4: Pathway representation of $\text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c)$.

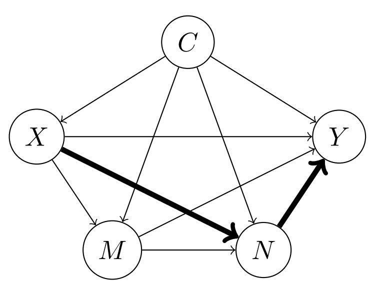


Figure 5: Pathway representation of $\text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c)$.

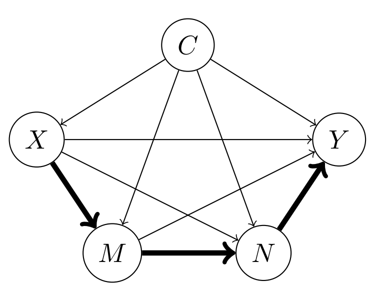


Figure 6: Pathway representation of $\text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c)$.

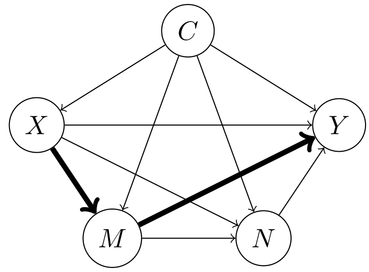


Figure 7: Pathway representation of $\text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c)$.

D PROOFS

In this section, we provide the proofs of propositions and theorems in the body of the paper.

Proposition 3.1. *T-PNS can be decomposed as follows:*

$$\begin{aligned}\text{T-PNS}(y;x',x,\mathcal{E},c) &= \text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c)\\ &\quad + \text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c).\end{aligned} \tag{38}$$

Proof. For any $x',x\in\Omega_X$, $y\in\Omega_Y$, and $c\in\Omega_C$, T-PNS, ND-PNS, and NI-PNS are decomposed by

$$\begin{aligned}\text{T-PNS}(y;x',x,\mathcal{E},c) &= \mathbb{P}(Y_{x'}\prec y\preceq Y_x|\mathcal{E},C=c)\\ &= \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y, Y_{x',M_x,N_{x,M_x}}\prec y\Big|\mathcal{E},C=c\Big)\\ &+ \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y, y\preceq Y_{x',M_x,N_{x,M_x}}\Big|\mathcal{E},C=c\Big)\\ &+ \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, Y_{x',M_x,N_{x',M_{x'}}}\prec y\Big|\mathcal{E},C=c\Big)\\ &+ \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, y\preceq Y_{x',M_x,N_{x',M_{x'}}}\Big|\mathcal{E},C=c\Big)\\ &= \text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c)\\ &\qquad + \text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c).\end{aligned} \tag{39}$$

□

Proposition 3.2. *We have*

$$\mathbb{P}(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y|\mathcal{E},C=c) = \text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c), \tag{40}$$

$$\mathbb{P}(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}|\mathcal{E},C=c) = \text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c). \tag{41}$$

Proof. We have

$$\begin{aligned}\mathbb{P}(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y|\mathcal{E},C=c) &= \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y, Y_{x',M_x,N_{x,M_x}}\prec y\Big|\mathcal{E},C=c\Big)\\ &+ \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y, y\preceq Y_{x',M_x,N_{x,M_x}}\Big|\mathcal{E},C=c\Big)\\ &= \text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c),\end{aligned} \tag{42}$$

$$\begin{aligned}\mathbb{P}(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}|\mathcal{E},C=c) &= \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, Y_{x',M_x,N_{x',M_{x'}}}\prec y\Big|\mathcal{E},C=c\Big)\\ &+ \mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, y\preceq Y_{x',M_x,N_{x',M_{x'}}}\Big|\mathcal{E},C=c\Big)\\ &= \text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c) + \text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c).\end{aligned} \tag{43}$$

□

Lemma 4.1. *Under SCM $\mathcal{M}_2$ and Assumption 4.1, we have*

$$\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) = \mathbb{1}(F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c)) = m') \tag{44}$$

where $F^{-1}_{M|X=x''',C=c}$ is the inverse function of conditional CDF $\mathbb{P}(M \preceq m|X=x''',C=c)$ on m given $X=x'''$ and $C=c$.

Proof. We have

$$\mathfrak{p}_{M_{x'''}|C,M_{x'}}(m'|c,m) = \frac{d}{dm'}\mathbb{P}(M_{x'''} \prec m'|C=c, M_{x'}=m). \tag{45}$$

When f_M is strictly monotonic increasing, we denote $\overline{u}_{M_{x'}|C=c} := \sup\{u_M \in \Omega_{U^M}; M_{x'} \preceq m\}$ given $C=c$, and we have $\mathbb{P}(U^M < \overline{u}_{M_{x'}}|C=c) = \mathbb{P}(M_{x'} \preceq m|C=c) = \mathbb{P}(M \preceq m|X=x',C=c)$. When f_M is strictly monotonic decreasing, we denote $\overline{u}_{M_{x'}|C=c} := \inf\{u_{M_{x'}} \in \Omega_{U^M}; M_{x'} \preceq m\}$ given $C=c$, and we have $\mathbb{P}(U^M > \overline{u}_{M_{x'}}|C=c) = \mathbb{P}(M_{x'} \preceq m|C=c) = \mathbb{P}(M \preceq m|X=x',C=c)$. In the both case, we have

$$\begin{aligned}\frac{d}{dm'}\mathbb{P}_M(M_{x'''} \prec m'|C=c, M_{x'}=m) &= \frac{d}{dm'}\mathbb{P}_M(M_{x'''} \prec m'|C=c, U^M=\overline{u}_{M_{x'}}) \\ &= \frac{d}{dm'}\mathbb{P}_M(M_{x'''}(\overline{u}_{M_{x'}}) \prec m'|C=c).\end{aligned} \tag{46}$$

The value of $M_{x'''}(\overline{u}_{M_{x'}})$, which is a potential outcome $M_{x'''}$ for the subject $U^M = \overline{u}_{M_{x'}}$ and is a constant, is equal to $F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c))$ since we have $\mathbb{P}(M_{x'} \preceq m|C=c) = \mathbb{P}(M_{x'''} \preceq M_{x'''}(\overline{u}_{M_{x'}})|C=c)$ from the strict monotonicity. Then, we have

$$\begin{aligned}\frac{d}{dm'}\mathbb{P}_M(M_{x'''} \prec m'|C=c, M_{x'}=m) &= \frac{d}{dm'}\mathbb{I}(F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c)) \prec m') \\ &= \mathbb{1}(F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c)) = m').\end{aligned} \tag{47}$$

□

Theorem 4.1. *Under SCM $\mathcal{M}_2$ and Assumption 4.1, for any $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, $\mathbb{P}(Y_{x,M_{x'},N_{x'',M_{x'''}}} \prec y|C=c)$ is identifiable by $\theta(y;x,x',x'',x''',c)$, where*

$$\begin{aligned}\theta(y;x,x',x'',x''',c) = \int_{\Omega_M}\int_{\Omega_N} &\mathbb{P}(Y \prec y|X=x, M=m, N=n, C=c) \\ \times\, \mathfrak{p}_{N|C,X,M}(n|c,x'',&F^{-1}_{M|X=x''',C=c}(\mathbb{P}(M \preceq m|X=x',C=c)))\mathfrak{p}_{M|C,X}(m|c,x')\, dndm.\end{aligned} \tag{48}$$

Proof. This is because Eq. 4 holds and $\mathfrak{p}_{M_{x''}|C,M_{x'}}(m'|c,m)$ is identifiable from Lemma 4.1. □

Theorem 4.2. *Let $\mathcal{I}_Y$ be a half-open interval in evidence $\mathcal{E}$. We have the following two statements.*

(1) Under SCM $\mathcal{M}_2$ and Assumptions 2.1, 4.2 and 4.3, for any $x', x \in \Omega_X$, $y \in \Omega_Y$, and $c \in \Omega_C$, we have

(1A). *If $\mathbb{P}(Y \prec y^l|X=x^e, C=c) \neq \mathbb{P}(Y \prec y^u|X=x^e, C=c)$, then*

$$\text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) = \max\{\gamma^1/\delta, 0\}, \text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c) = \max\{\gamma^2/\delta, 0\}, \tag{49}$$

where

$$\begin{aligned}\gamma^1 = \min\Big\{&\theta(y;x',x',x',x',c), \theta(y;x',x,x',x,c), \theta(y;x',x,x,x,c), \mathbb{P}(Y \prec y^u|X=x^e, C=c)\Big\} \\ - \max\Big\{&\theta(y;x,x,x,x,c), \mathbb{P}(Y \prec y^l|X=x^e, C=c)\Big\},\end{aligned} \tag{50}$$

$$\begin{aligned}\gamma^2 = \min\Big\{&\theta(y;x',x',x',x',c), \theta(y;x',x,x',x,c), \mathbb{P}(Y \prec y^u|X=x^e, C=c)\Big\} \\ - \max\Big\{&\theta(y;x,x,x,x,c), \mathbb{P}(Y \prec y^l|X=x^e, C=c), \theta(y;x',x,x,x,c)\Big\},\end{aligned} \tag{51}$$

$$\delta = \mathbb{P}(Y \prec y^u|X = x^e, C = c) - \mathbb{P}(Y \prec y^l|X = x^e, C = c). \quad (52)$$

(1B). *If* $\mathbb{P}(Y \prec y^l|X = x^e, C = c) = \mathbb{P}(Y \prec y^u|X = x^e, C = c)$*, then*

$$\begin{aligned}
&\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c)\mathbb{I}\Big(\theta(y; x', x', x', x', c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c),\\
&\theta(y; x', x, x', x, c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c), \theta(y; x', x, x, x, c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c)\Big),\\
&\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\Big(\theta(y; x', x', x', x', c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c),\\
&\theta(y; x', x, x', x, c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x, x, c)\Big), \quad (53)
\end{aligned}$$

(2) Under SCM $\mathcal{M}_2$ *and Assumptions 2.1, 4.1, 4.2 and 4.3, for any* $x', x \in \Omega_X$*,* $y \in \Omega_Y$*, and* $c \in \Omega_C$*, we have*

(2A). *If* $\mathbb{P}(Y \prec y^l|X = x^e, C = c) \ne \mathbb{P}(Y \prec y^u|X = x^e, C = c)$*, then*

$$\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^3/\delta, 0\}, \quad (54)$$

$$\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) = \max\{\gamma^4/\delta, 0\}, \quad (55)$$

where

$$\begin{aligned}
\gamma^3 &= \min\Big\{\theta(y; x', x', x', x', c), \mathbb{P}(Y \prec y^u|X = x^e, C = c), \theta(y; x', x, x', x', c)\Big\}\\
&- \max\Big\{\theta(y; x, x, x, x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c), \theta(y; x', x, x', x, c)\Big\}, \quad (56)\\
\gamma^4 &= \min\Big\{\theta(y; x', x', x', x', c), \mathbb{P}(Y \prec y^u|X = x^e, C = c)\Big\}\\
&- \max\Big\{\theta(y; x, x, x, x, c), \theta(y; x', x, x', x, c), \mathbb{P}(Y \prec y^l|X = x^e, C = c), \theta(y; x', x, x', x', c)\Big\}, \quad (57)\\
\delta &= \mathbb{P}(Y \prec y^u|X = x^e, C = c) - \mathbb{P}(Y \prec y^l|X = x^e, C = c). \quad (58)
\end{aligned}$$

(2B). *If* $\mathbb{P}(Y \prec y^l|X = x^e, C = c) = \mathbb{P}(Y \prec y^u|X = x^e, C = c)$*, then*

$$\begin{aligned}
&\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\Big(\theta(y; x', x', x', x', c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c),\\
&\mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x, c), \theta(y; x', x, x', x', c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c)\Big), \quad (59)\\
&\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) = \mathbb{I}\Big(\theta(y; x', x', x', x', c) \le \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x, x, x, x, c),\\
&\mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x, c), \mathbb{P}(Y \prec y^u|X = x^e, C = c) < \theta(y; x', x, x', x', c)\Big). \quad (60)
\end{aligned}$$

Proof. Let $\mathcal{I}_Y$ be a half-open interval in the evidence $\mathcal{E}$. Under SCM and Assumptions 2.1, 4.2 and 4.3, for any $x', x \in \Omega_X$, $y \in \Omega_Y$, and $c \in \Omega_C$, we have $\mathbb{P}(Y \prec y^u|X = x^e, C = c)$ and $\mathbb{P}(Y \prec y^l|X = x^e, C = c)$. Without loss of generality, the function $g(x, x', x'', x''', c, \tilde{U})$ is monotonic increasing on $\tilde{U}$ for all $x, x', x'', x''' \in \Omega_X$ and $c \in \Omega_C$, almost surely w.r.t. $\mathbb{P}_{\tilde{U}}$. Let $u_{x,y} = \{u \in \Omega_{\tilde{U}}; Y_x(u) \prec y\}$, $u_{x,x',y} = \{u \in \Omega_{\tilde{U}}; Y_{x,M_{x'}} \prec y\}$, and $u_{x,x',x'',x'''y} = \{u \in \Omega_{\tilde{U}}; Y_{x,M_{x'},N_{x''},M_{x'''}} \prec y\}$.

(A). If $\mathbb{P}(Y \prec y^l|X = x^e, C = c) \ne \mathbb{P}(Y \prec y^l|X = x^e, C = c)$, then we have

$$\begin{aligned}
&\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c)\\
&= \mathbb{P}\Big(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y, Y_{x',M_x,N_{x,M_x}} \prec y\Big|\mathcal{E}, C = c\Big)\\
&= \frac{\mathbb{P}\Big(u_{x',y} \le \tilde{U} < u_{x,y}, u_{x',x,y} \le \tilde{U}, u_{x',x,x,x,y} \le \tilde{U}, u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u}\Big|C = c\Big)}{\mathbb{P}\Big(u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u}\Big|C = c\Big)}\\
&= \Bigg[\min\Big\{\mathbb{P}\left(Y_{x'} \prec y\Big|C = c\right), \mathbb{P}\left(Y_{x',M_x} \prec y\Big|C = c\right), \mathbb{P}\left(Y_{x',M_x,N_{x,M_x}} \prec y\Big|C = c\right), \mathbb{P}(Y_{x^e} \prec y^u|C = c)\Big\}
\end{aligned}$$

$$-\max\Big\{\mathbb{P}\left(Y_x \prec y \Big| C=c\right), \mathbb{P}(Y_{x^e} \prec y^l | C=c)\Big\}\Bigg] \Bigg/ \Big\{\mathbb{P}(Y_{x^e} \prec y^u | C=c) - \mathbb{P}(Y_{x^e} \prec y^l | C=c)\Big\}, \quad (61)$$

$$\begin{aligned}
&\text{PNS}^{X\to N\to Y}(y; x', x, \mathcal{E}, c) \\
&= \mathbb{P}\Big(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y, y \preceq Y_{x',M_x,N_{x,M_x}} \Big| \mathcal{E}, C=c\Big) \\
&= \frac{\mathbb{P}\Big(u_{x',y} \le \tilde{U} < u_{x,y}, u_{x',x,y} \le \tilde{U}, \tilde{U} < u_{x',x,x,x,y}, u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)}{\mathbb{P}\Big(u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)} \\
&= \Bigg[\min\Big\{\mathbb{P}\left(Y_{x'} \prec y \Big| C=c\right), \mathbb{P}\left(Y_{x',M_x} \prec y \Big| C=c\right), \mathbb{P}(Y_{x^e} \prec y^u | C=c)\Big\} - \max\Big\{\mathbb{P}\left(Y_x \prec y \Big| C=c\right), \\
&\mathbb{P}(Y_{x^e} \prec y^l | C=c), \mathbb{P}\left(Y_{x',M_x,N_{x,M_x}} \prec y \Big| C=c\right)\Big\}\Bigg] \Bigg/ \Big\{\mathbb{P}(Y_{x^e} \prec y^u | C=c) - \mathbb{P}(Y_{x^e} \prec y^l | C=c)\Big\}, \quad (62)
\end{aligned}$$

$$\begin{aligned}
&\text{PNS}^{X\to M\to N\to Y}(y; x', x, \mathcal{E}, c) \\
&= \mathbb{P}\Big(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}, Y_{x',M_x,N_{x',M_{x'}}} \prec y \Big| \mathcal{E}, C=c\Big) \\
&= \frac{\mathbb{P}\Big(u_{x',y} \le \tilde{U} < u_{x,y}, \tilde{U} < u_{x',x,y}, u_{x',x,x',x',y} \le \tilde{U}, u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)}{\mathbb{P}\Big(u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)} \\
&= \Bigg[\min\Big\{\mathbb{P}\left(Y_{x'} \prec y \Big| C=c\right), \mathbb{P}(Y_{x^e} \prec y^u | C=c), \mathbb{P}\left(Y_{x',M_x,N_{x',M_{x'}}} \prec y \Big| C=c\right)\Big\} - \max\Big\{\mathbb{P}\left(Y_x \prec y \Big| C=c\right), \\
&\mathbb{P}(Y_{x^e} \prec y^l | C=c), \mathbb{P}\left(Y_{x',M_x} \prec y \Big| C=c\right)\Big\}\Bigg] \Bigg/ \Big\{\mathbb{P}(Y_{x^e} \prec y^u | C=c) - \mathbb{P}(Y_{x^e} \prec y^l | C=c)\Big\}, \quad (63)
\end{aligned}$$

$$\begin{aligned}
&\text{PNS}^{X\to M\to Y}(y; x', x, \mathcal{E}, c) \\
&= \mathbb{P}\Big(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M_x}, y \preceq Y_{x',M_x,N_{x',M_{x'}}} \Big| \mathcal{E}, C=c\Big) \\
&= \frac{\mathbb{P}\Big(u_{x',y} \le \tilde{U} < u_{x,y}, \tilde{U} < u_{x',x,y}, \tilde{U} < u_{x',x,x',x',y}, u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)}{\mathbb{P}\Big(u_{x^e,y^l} \le \tilde{U} < u_{x^e,y^u} \Big| C=c\Big)} \\
&= \Bigg[\min\Big\{\mathbb{P}\left(Y_{x'} \prec y \Big| C=c\right), \mathbb{P}(Y_{x^e} \prec y^u | C=c)\Big\} - \max\Big\{\mathbb{P}\left(Y_x \prec y \Big| C=c\right), \mathbb{P}\left(Y_{x',M_x} \prec y \Big| C=c\right), \\
&\mathbb{P}(Y_{x^e} \prec y^l | C=c), \mathbb{P}\left(Y_{x',M_x,N_{x',M_{x'}}} \prec y \Big| C=c\right)\Big\}\Bigg] \Bigg/ \Big\{\mathbb{P}(Y_{x^e} \prec y^u | C=c) - \mathbb{P}(Y_{x^e} \prec y^l | C=c)\Big\}. \quad (64)
\end{aligned}$$

(B). If $\mathbb{P}(Y \prec y^l | X = x^e, C = c) = \mathbb{P}(Y \prec y^l | X = x^e, C = c)$, then we have

$$\begin{aligned}
&\text{PNS}^{X\to Y}(y; x', x, \mathcal{E}, c) \\
&= \mathbb{P}\Big(Y_{x'} \prec y \preceq Y_x, Y_{x',M_x} \prec y, Y_{x',M_x,N_{x,M_x}} \prec y \Big| \mathcal{E}, C=c\Big) \\
&= \mathbb{P}\Big(u_{x',y} \le \tilde{U} < u_{x,y}, u_{x',x,y} \le \tilde{U}, u_{x',x,x,x,y} \le \tilde{U} \Big| \tilde{U} = u_{x^e,y^u}, C=c\Big) \\
&= \mathbb{I}\Big(\theta(y; x', x', x', x', c) \le \mathbb{P}(Y_{x^e} \prec y^u | C=c) < \theta(y; x, x, x, x, c), \\
&\qquad \theta(y; x', x, x', x, c) \le \mathbb{P}(Y_{x^e} \prec y^u |_{x^e} C=c), \theta(y; x', x, x, x, c) \le \mathbb{P}(Y_{x^e} \prec y^u | C=c)\Big), \quad (65)
\end{aligned}$$

$$\begin{aligned}
&\text{PNS}^{X\to N\to Y}(y;x',x,\mathcal{E},c)\\
&=\mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, Y_{x',M_x}\prec y, y\preceq Y_{x',M_x,N_{x,M_x}}\Big|\mathcal{E},C=c\Big)\\
&=\mathbb{P}\Big(u_{x',y}\le\tilde{U}<u_{x,y}, u_{x',x,y}\le\tilde{U},\tilde{U}<u_{x',x,x,x,y}\Big|\tilde{U}=u_{x^e,y^u},C=c\Big)\\
&=\mathbb{I}\Big(\theta(y;x',x',x',x',c)\le\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x,x,x,x,c),\\
&\qquad\theta(y;x',x,x',x,c)\le\mathbb{P}(Y_{x^e}\prec y^u|C=c),\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x',x,x,x,c)\Big),
\end{aligned}\tag{66}$$

$$\begin{aligned}
&\text{PNS}^{X\to M\to N\to Y}(y;x',x,\mathcal{E},c)\\
&=\mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, Y_{x',M_x,N_{x',M_{x'}}}\prec y\Big|\mathcal{E},C=c\Big)\\
&=\mathbb{P}\Big(u_{x',y}\le\tilde{U}<u_{x,y},\tilde{U}<u_{x',x,y}, u_{x',x,x',x',y}\le\tilde{U}\Big|\tilde{U}=u_{x^e,y^u},C=c\Big)\\
&=\mathbb{I}\Big(\theta(y;x',x',x',x',c)\le\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x,x,x,x,c),\\
&\qquad\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x',x,x',x,c),\theta(y;x',x,x',x',c)\le\mathbb{P}(Y_{x^e}\prec y^u|C=c)\Big),
\end{aligned}\tag{67}$$

$$\begin{aligned}
&\text{PNS}^{X\to M\to Y}(y;x',x,\mathcal{E},c)\\
&=\mathbb{P}\Big(Y_{x'}\prec y\preceq Y_x, y\preceq Y_{x',M_x}, y\preceq Y_{x',M_x,N_{x',M_{x'}}}\Big|\mathcal{E},C=c\Big)\\
&=\mathbb{P}\Big(u_{x',y}\le\tilde{U}<u_{x,y},\tilde{U}<u_{x',x,y},\tilde{U}<u_{x',x,x',x',y}\Big|\tilde{U}=u_{x^e,y^u},C=c\Big)\\
&=\mathbb{I}\Big(\theta(y;x',x',x',x',c)\le\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x,x,x,x,c),\\
&\qquad\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x',x,x',x,c),\mathbb{P}(Y_{x^e}\prec y^u|C=c)<\theta(y;x',x,x',x',c)\Big).
\end{aligned}\tag{68}$$

Furthermore, under Assumption 4.1, $\theta(y;x',x,x',x',c)$ is identifiable. □

Proof of Statement. We prove the statement "Assumptions 4.3 and 4.3' are equivalent under Assumption 4.2" in the body of the paper.

Proof. We show the proof of equivalence of Assumptions 4.3 and 4.3' under Assumption 4.2.

(Assumption 4.3 ⇒ Assumption 4.3'.) For any $c\in\Omega_C$, from Assumption 4.2, if we have the negation of Assumption 4.3'

there exists a set $\mathcal{U}\subset\Omega_{\tilde{U}}$ such that $0<\mathbb{P}(\mathcal{U})<1$, and

$$\begin{aligned}
&g(x,x',x'',x''',c,\tilde{u}_0)\succeq y\succ g(x^*,x^{**},x^{***},x^{****},c,\tilde{u}_1)\\
&\wedge\, g(x,x',x'',x''',c,\tilde{u}_0)\prec y\preceq g(x^*,x^{**},x^{***},x^{****},c,\tilde{u}_1)
\end{aligned}\tag{69}$$

for some $x,x',x'',x''',x^*,x^{**},x^{***},x^{****}\in\Omega_X$ and $y\in\Omega_Y$ and for any $\tilde{u}_0,\tilde{u}_1\in\mathcal{U}$ such that $\tilde{u}_0\preceq\tilde{u}_1$,

then we have

$g(x,x',x'',x''',c,\tilde{u}_0)\succeq y\succ g(x^*,x^{**},x^{***},x^{****},c,\tilde{u}_0)$ and
$g(x,x',x'',x''',c,\tilde{u}_1)\prec y\preceq g(x^*,x^{**},x^{***},x^{****},c,\tilde{u}_1)$ for some $x,x',x'',x''',x^*,x^{**},x^{***},x^{****}\in\Omega_X$ and $y\in\Omega_Y$
and
for any $\tilde{u}_0,\tilde{u}_1\in\mathcal{U}$ such that $\tilde{u}_0\preceq\tilde{u}_1$,

and we also have

$g(x,x',x'',x''',c,\tilde{u})\succeq y\succ g(x^*,x^{**},x^{***},x^{****},c,\tilde{u})$ and $g(x,x',x'',x''',c,\tilde{u})\prec y\preceq g(x^*,x^{**},x^{***},x^{****},c,\tilde{u})$ for some $x,x',x'',x''',x^*,x^{**},x^{***},x^{****}\in\Omega_X$ and $y\in\Omega_Y$ and for any $\tilde{u}\in\mathcal{U}$.

This implies the negation of Assumption 5 $\mathbb{P}((Y_{x,M_{x'},N_{x''},M_{x'''}} \prec y \preceq Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}}|C = c) \neq 0$ and $\mathbb{P}(Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}} \prec y \preceq (Y_{x,M_{x'},N_{x''},M_{x'''}}|C = c) \neq 0$ for some $x, x', x'', x''', x^*, x^{**}, x^{***}, x^{****} \in \Omega_Y$ and $y \in \Omega_Y$ since $g(x, x', x'', x''', c, \tilde{u}) \succeq y \succ g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \Leftrightarrow Y_{x',M_x}(c, \tilde{u}) \succeq y \succ Y_{x''',M_{x''}}(c, \tilde{u})$ and $g(x, x', x'', x''', c, \tilde{u}) \prec y \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \Leftrightarrow Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}}(c, \tilde{u}) \succeq y \succ Y_{x,M_{x'},N_{x''},M_{x'''}}(c, \tilde{u})$.

(Assumption 4.3' $\Rightarrow$ Assumption 4.3.) For any $c \in \Omega_C$, we denote $\tilde{u}_{sup} = \sup\{\tilde{u} : g(x, x', x'', x''', c, \tilde{u}) \preceq y\}$. We consider the situations "the function $g(x, x', x'', x''', c, \tilde{U})$ is monotonic increasing on $\tilde{U}$" and "the function $g(x, x', x'', x''', c, \tilde{U})$ is monotonic decreasing on $\tilde{U}$", separately.

(1). If the function $g(x, x', x'', x''', c, \tilde{U})$ is **monotonic increasing** on $\tilde{U}$ for all $x \in \Omega_X$ almost surely w.r.t. $\mathbb{P}_{\tilde{U}}$, we have

$$\begin{aligned} &g(x, x', x'', x''', c, \tilde{u}_{sup}) \preceq g(x, x', x'', x''', c, \tilde{u}) \\ &\text{and } g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \end{aligned} \tag{70}$$

for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $\tilde{u} \succeq \tilde{u}_{sup}$. We have the following statements:

1. Supposing $g(x, x', x'', x''', c, \tilde{u}_{sup}) \succ g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup})$, we have $y = g(x, x', x'', x''', c, \tilde{u}_{sup}) \succ g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \succeq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) = Y_{x''',M_{x''}}(c, \tilde{u})$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $g(x, x', x'', x''', c, \tilde{u}) \prec y$. It means $Y_{x,M_{x'},N_{x''},M_{x'''}}(c, \tilde{u}) \prec y \Rightarrow Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}}(c, \tilde{u}) \prec y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ and $\mathbb{P}(Y_{x,M_{x'},N_{x''},M_{x'''}} \prec y \preceq Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}}|C = c) = 0$.
2. Supposing $g(x, x', x'', x''', c, \tilde{u}_{sup}) \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup})$, we have $g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \succeq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \succeq g(x, x', x'', x''', c, \tilde{u}_{sup}) = y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $g(x, x', x'', x''', c, \tilde{u}) \succeq y$. It means $Y_{x,M_{x'},N_{x''},M_{x'''}}(c, \tilde{u}) \succeq y \Rightarrow Y_{x''',M_{x''}}(c, \tilde{u}) \succeq y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ and $\mathbb{P}(Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}} \prec y \preceq Y_{x,M_{x'},N_{x''},M_{x'''}}|C = c) = 0$.

Then, these results imply Assumption 4.3'.

(2). If the function $g(x, x', x'', x''', c, \tilde{U})$ is **monotonic decreasing** on $\tilde{U}$ for all $x \in \Omega_X$ almost surely w.r.t. $\mathbb{P}_{\tilde{U}}$, we have

$$\begin{aligned} &g(x, x', x'', x''', c, \tilde{u}_{sup}) \succeq g(x, x', x'', x''', c, \tilde{u}) \\ &\text{and } g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \succeq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \end{aligned} \tag{71}$$

for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $\tilde{u} \succeq \tilde{u}_{sup}$. We have the following statements:

1. Supposing $g(x, x', x'', x''', c, \tilde{u}_{sup}) \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup})$, we have $y = g(x, x', x'', x''', c, \tilde{u}_{sup}) \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \preceq g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) = Y_{x''',M_{x''}}(c, \tilde{u})$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $g(x, x', x'', x''', c, \tilde{u}) \succeq y$. It means $Y_{x',M_x}(c, \tilde{u}) \succeq y \Rightarrow Y_{x''',M_{x''}}(\tilde{u}) \succeq y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ and $\mathbb{P}(Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}} \prec y \preceq Y_{x,M_{x'},N_{x''},M_{x'''}}|C = c) = 0$.
2. Supposing $g(x, x', x'', x''', c, \tilde{u}_{sup}) \succ g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup})$, we have $g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}) \prec g(x^*, x^{**}, x^{***}, x^{****}, c, \tilde{u}_{sup}) \preceq g(x, x', x'', x''', c, \tilde{u}_{sup}) = y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ such that $g(x, x', x'', x''', c, \tilde{u}) \prec y$. It means $Y_{x',M_x}(c, \tilde{u}) \prec y \Rightarrow Y_{x''',M_{x''}}(c, \tilde{u}) \prec y$ for $\mathbb{P}_{\tilde{U}}$-almost every $\tilde{u} \in \Omega_{\tilde{U}}$ and $\mathbb{P}(Y_{x,M_{x'},N_{x''},M_{x'''}} \prec y \preceq Y_{x^*,M_{x^{**}},N_{x^{***}},M_{x^{****}}}|C = c) = 0$.

Then, these resluts imply Assumption 4.3'. In conclusion, Assumption 4.3' implies Assumption 4.3 under Assumption 4.2.

□

E PATH-SPECIFIC PNS WITH THREE MEDIATORS

We consider the following SCM $\mathcal{M}_3$:

$$\begin{aligned} Y := f_Y(X, M^1, M^2, M^3, C, U^Y), M_3 := f_{M^3}(X, M^1, M^2, C, U^{M^3}), \\ M_2 := f_{M^2}(X, M^1, C, U^{M^2}), M^1 := f_{M^1}(X, C, U^{M^1}), X := f_X(C, U^X), C := f_C(U^C), \end{aligned} \tag{72}$$

where all variables can be vectors, and U^X, U^C, U^Y, U^{M^1}, U^{M^2}, and U^{M^3} are latent exogenous variables. Assume that the domains Ω_Y and $\Omega_{U^Y} \times \Omega_{U^{M^1}} \times \Omega_{U^{M^2}} \times \Omega_{U^{M^3}}$ are totally ordered sets with $\preceq$. Three mediators are causally ordered, or M^1 is the cause of M^2 and M^2 is the cause of M^3. We give the definitions of the path-specific PNS with three mediators, and they have five counterfactual conditions, respectively.

Definition E.1 (Path-Specific PNS with Three Mediators)**.** *We define eight types of path-specific PNS with three mediators as follows:*

$$\text{PNS}^{X\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, Y_{x',M^1_x} \prec y, Y_{x',M^1_x,M^2_{x,M^1_x}} \prec y, \\ Y_{x',M^1_x,M^2_{x,M^1_x},M^3_{x,M^1_x,M^2_{x,M^1_x}}} \prec y \Bigg| \mathcal{E}, C=c \Bigg), \tag{73}$$

$$\text{PNS}^{X\to M^3\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, Y_{x',M^1_x} \prec y, Y_{x',M^1_x,M^2_{x,M^1_x}} \prec y, \\ y \preceq Y_{x',M^1_x,M^2_{x,M^1_x},M^3_{x,M^1_x,M^2_{x,M^1_x}}} \Bigg| \mathcal{E}, C=c \Bigg), \tag{74}$$

$$\text{PNS}^{X\to M^2\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, Y_{x',M^1_x} \prec y, y \preceq Y_{x',M^1_x,M^2_{x,M^1_x}}, \\ Y_{x',M^1_x,M^2_{x',M^1_x},M^3_{x',M^1_x,M^2_{x,M^1_x}}} \prec y \Bigg| \mathcal{E}, C=c \Bigg), \tag{75}$$

$$\text{PNS}^{X\to M^2\to M^3\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, Y_{x',M^1_x} \prec y, y \preceq Y_{x',M^1_x,M^2_{x,M^1_x}}, \\ y \preceq Y_{x',M^1_x,M^2_{x',M^1_x},M^3_{x',M^1_x,M^2_{x,M^1_x}}} \Bigg| \mathcal{E}, C=c \Bigg), \tag{76}$$

$$\text{PNS}^{X\to M^1\to M^2\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M^1_x}, Y_{x',M^1_x,M^2_{x',M^1_{x'}}} \prec y, \\ Y_{x',M^1_x,M^2_{x',M^1_{x'}},M^3_{x',M^1_x,M^2_{x',M^1_x}}} \prec y \Bigg| \mathcal{E}, C=c \Bigg), \tag{77}$$

$$\text{PNS}^{X\to M^1\to M^2\to M^3\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M^1_x}, Y_{x',M^1_x,M^2_{x',M^1_{x'}}} \prec y, \\ y \preceq Y_{x',M^1_x,M^2_{x',M^1_{x'}},M^3_{x',M^1_x,M^2_{x',M^1_x}}} \Bigg| \mathcal{E}, C=c \Bigg), \tag{78}$$

$$\text{PNS}^{X\to M^1\to Y}(y;x',x,\mathcal{E},c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x',M^1_x}, y \preceq Y_{x',M^1_x,M^2_{x',M^1_{x'}}}, \\ Y_{x',M^1_{x'},M^2_{x',M^1_{x'}},M^3_{x',M^1_x,M^2_{x',M^1_{x'}}}} \prec y \Bigg| \mathcal{E}, C=c \Bigg), \tag{79}$$

$$\text{PNS}^{X \to M^1 \to M^3 \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}\Bigg(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x', M^1_x}, y \preceq Y_{x', M^1_x, M^2_{x', M^1_{x'}}}, \\ y \preceq Y_{x', M^1_{x'}, M^2_{x', M^1_{x'}}, M^3_{x', M^1_x, M^2_{x', M^1_{x'}}}} \Bigg| \mathcal{E}, C = c\Bigg), \tag{80}$$

where $\mathcal{E} \triangleq (X = x^*, Y \in \mathcal{I}_Y)$, and $\mathcal{I}_Y$ is a half-open interval $[y^l, y^u)$ or a closed interval $[y^l, y^u]$ w.r.t. $\prec$.

We have the following seven decompositions:

$$\text{PNS}^{(M^1,M^2);X \to Y}(y; x', x, \mathcal{E}, c) = \text{PNS}^{X \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X \to M^3 \to Y}(y; x', x, \mathcal{E}, c) \tag{81}$$

$$\text{PNS}^{(M^1,M^2);X \to M^2 \to Y}(y; x', x, \mathcal{E}, c) = \text{PNS}^{X \to M^2 \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X \to M^2 \to M^3 \to Y}(y; x', x, \mathcal{E}, c) \tag{82}$$

$$\text{PNS}^{(M^1,M^2);X \to M^1 \to M^1 \to Y}(y; x', x, \mathcal{E}, c) = \text{PNS}^{X \to M^1 \to M^2 \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X \to M^1 \to M^2 \to M^3 \to Y}(y; x', x, \mathcal{E}, c) \tag{83}$$

$$\text{PNS}^{(M^1,M^2);X \to M^1 \to Y}(y; x', x, \mathcal{E}, c) = \text{PNS}^{X \to M^1 \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{X \to M^1 \to M^3 \to Y}(y; x', x, \mathcal{E}, c), \tag{84}$$

$$\text{ND-PNS}^{M^1}(y; x', x, \mathcal{E}, c) = \text{PNS}^{(M^1,M^2);X \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{(M^1,M^2);X \to M^2 \to Y}(y; x', x, \mathcal{E}, c) \tag{85}$$

$$\text{NI-PNS}^{M^1}(y; x', x, \mathcal{E}, c) = \text{PNS}^{(M^1,M^2);X \to M^1 \to M^1 \to Y}(y; x', x, \mathcal{E}, c) + \text{PNS}^{(M^1,M^2);X \to M^1 \to Y}(y; x', x, \mathcal{E}, c) \tag{86}$$

$$\text{T-PNS}(y; x', x, \mathcal{E}, c) = \text{ND-PNS}^{M^1}(y; x', x, \mathcal{E}, c) + \text{NI-PNS}^{M^1}(y; x', x, \mathcal{E}, c), \tag{87}$$

where

$$\text{PNS}^{(M^1,M^2);X \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x', {M^1}_x} \prec y, Y_{x', {M^1}_x, {M^2}_{x, {M^1}_x}} \prec y | \mathcal{E}, C = c), \tag{88}$$

$$\text{PNS}^{(M^1,M^2);X \to M^2 \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x', {M^1}_x} \prec y, y \preceq Y_{x', {M^1}_x, {M^2}_{x, {M^1}_x}} | \mathcal{E}, C = c), \tag{89}$$

$$\text{PNS}^{(M^1,M^2);X \to M^1 \to M^1 \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x', {M^1}_x}, Y_{x', {M^1}_x, {M^2}_{x', {M^1}_{x'}}} \prec y | \mathcal{E}, C = c), \tag{90}$$

$$\text{PNS}^{(M^1,M^2);X \to M^1 \to Y}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x', {M^1}_x}, y \preceq Y_{x', {M^1}_x, {M^2}_{x', {M^1}_{x'}}} | \mathcal{E}, C = c), \tag{91}$$

$$\text{ND-PNS}^{M^1}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, Y_{x', {M^1}_x} \prec y | \mathcal{E}, C = c), \tag{92}$$

$$\text{NI-PNS}^{M^1}(y; x', x, \mathcal{E}, c) \triangleq \mathbb{P}(Y_{x'} \prec y \preceq Y_x, y \preceq Y_{x', {M^1}_x} | \mathcal{E}, C = c). \tag{93}$$

F ADDITIONAL NUMERICAL EXPERIMENTS

F.1 SPECIAL CASES

We provide three additional experiments under (1) no effect between M and N, (2) no effect between $\{M, N\}$ and Y, (3) only effect through $X \to M \to N \to Y$.

(1). No effect between M and N. We consider the situation where there is no effect between M and N.

Setting. We consider the following linear SCM:

$$Y := X + M + N + C + U^Y, N := X + C + U^N, M := X + C + U^M, X := C + U^X, C := U^C, \tag{94}$$

where $U^C \sim \mathcal{N}(0,1)$, $U^X \sim \mathcal{N}(0,1)$, $U^Y \sim \mathcal{N}(0,1)$, $U^M \sim \mathcal{N}(0,1)$, $U^N \sim \mathcal{N}(0,1)$ and they are mutually independent normal distributions. This SCM satisfies Assumptions 2.1, 4.1, 4.2, 4.3, and 4.3'. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$.

Results. The ground truth of T-PNS is 0.458. The ground truth of $\text{PNS}^{X \to Y}$ is 0.082 and the estimates are

$N = 20$: 0.083 (95%CI: $[0.003, 0.213]$),
$N = 100$: 0.081 (95%CI: $[0.044, 0.127]$),
$N = 10000$: 0.082 (95%CI: $[0.078, 0.086]$).

The ground truth of PNS$^{X \to N \to Y}$ is 0.158 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.158 \text{ (95\%CI: } [0.030, 0.350]), \\ N = 100&: \quad 0.158 \text{ (95\%CI: } [0.102, 0.221]), \\ N = 10000&: 0.158 \text{ (95\%CI: } [0.151, 0.164]). \end{aligned}$$

The ground truth of PNS$^{X \to M \to N \to Y}$ is 0.000 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.018 \text{ (95\%CI: } [0.000, 0.090]), \\ N = 100&: \quad 0.008 \text{ (95\%CI: } [0.000, 0.038]), \\ N = 10000&: 0.001 \text{ (95\%CI: } [0.000, 0.004]). \end{aligned}$$

The ground truth of PNS$^{X \to M \to Y}$ is 0.218 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.199 \text{ (95\%CI: } [0.057, 0.353]), \\ N = 100&: \quad 0.211 \text{ (95\%CI: } [0.154, 0.265]), \\ N = 10000&: 0.217 \text{ (95\%CI: } [0.212, 0.223]). \end{aligned}$$

All means of the estimators are close to the ground truth. However, estimators for small sample sizes have large 95 $\%$ CIs.

(2). No effect between $\{M, N\}$ and Y. We consider the situation where there is no effect between $\{M, N\}$ and Y.

Setting. We consider the following linear SCM:

$$Y := X + C + U^Y, N := X + M + C + U^N, M := X + C + U^M, X := C + U^X, C := U^C, \tag{95}$$

where $U^C \sim \mathcal{N}(0,1)$, $U^X \sim \mathcal{N}(0,1)$, $U^Y \sim \mathcal{N}(0,1)$, $U^M \sim \mathcal{N}(0,1)$, $U^N \sim \mathcal{N}(0,1)$ and they are mutually independent normal distributions. This SCM satisfies Assumptions 2.1, 4.1, 4.2, 4.3, and 4.3'. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$.

Results. The ground truth of T-PNS is 0.346. The ground truth of PNS$^{X \to Y}$ is 0.346 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.282 \text{ (95\%CI: } [0.037, 0.481]), \\ N = 100&: \quad 0.314 \text{ (95\%CI: } [0.207, 0.400]), \\ N = 10000&: 0.339 \text{ (95\%CI: } [0.328, 0.347]). \end{aligned}$$

The ground truth of PNS$^{X \to N \to Y}$ is 0.000 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.027 \text{ (95\%CI: } [0.000, 0.185]), \\ N = 100&: \quad 0.011 \text{ (95\%CI: } [0.000, 0.074]), \\ N = 10000&: 0.001 \text{ (95\%CI: } [0.000, 0.007]). \end{aligned}$$

The ground truth of PNS$^{X \to M \to N \to Y}$ is 0.000 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.011 \text{ (95\%CI: } [0.000, 0.103]), \\ N = 100&: \quad 0.005 \text{ (95\%CI: } [0.000, 0.042]), \\ N = 10000&: 0.000 \text{ (95\%CI: } [0.000, 0.004]). \end{aligned}$$

The ground truth of PNS$^{X \to M \to Y}$ is 0.000 and the estimates are

$$\begin{aligned} N = 20&: \quad 0.025 \text{ (95\%CI: } [0.000, 0.168]), \\ N = 100&: \quad 0.012 \text{ (95\%CI: } [0.000, 0.071]), \\ N = 10000&: 0.001 \text{ (95\%CI: } [0.000, 0.007]). \end{aligned}$$

All means of the estimators are close to the ground truth. However, estimators for small sample sizes have large 95 $\%$ CIs.

(3). Only effect through $X \to M \to N \to Y$. We consider the situation where there is only effect through $X \to M \to N \to Y$.

Setting. We consider the following linear SCM:

$$Y := N + C + U^Y, N := M + C + U^N, M := X + C + U^M, X := C + U^X, C := U^C, \tag{96}$$

where $U^C \sim \mathcal{N}(0,1)$, $U^X \sim \mathcal{N}(0,1)$, $U^Y \sim \mathcal{N}(0,1)$, $U^M \sim \mathcal{N}(0,1)$, $U^N \sim \mathcal{N}(0,1)$ and they are mutually independent normal distributions. This SCM satisfies Assumptions 2.1, 4.1, 4.2, 4.3, and 4.3'. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$.

Results. The ground truth of T-PNS is 0.219. The ground truth of PNS$^{X \to Y}$ is 0.000 and the estimates are

$N = 20$: 0.017 (95%CI: $[0.000, 0.124]$),
$N = 100$: 0.008 (95%CI: $[0.000, 0.049]$),
$N = 10000$: 0.001 (95%CI: $[0.000, 0.004]$).

The ground truth of PNS$^{X \to N \to Y}$ is 0.000 and the estimates are

$N = 20$: 0.021 (95%CI: $[0.000, 0.144]$),
$N = 100$: 0.008 (95%CI: $[0.000, 0.050]$),
$N = 10000$: 0.001 (95%CI: $[0.000, 0.005]$).

The ground truth of PNS$^{X \to M \to N \to Y}$ is 0.219 and the estimates are

$N = 20$: 0.146 (95%CI: $[0.000, 0.293]$),
$N = 100$: 0.189 (95%CI: $[0.123, 0.247]$),
$N = 10000$: 0.215 (95%CI: $[0.208, 0.222]$).

The ground truth of PNS$^{X \to M \to Y}$ is 0.000 and the estimates are

$N = 20$: 0.034 (95%CI: $[0.000, 0.189]$),
$N = 100$: 0.014 (95%CI: $[0.000, 0.068]$),
$N = 10000$: 0.001 (95%CI: $[0.000, 0.007]$).

All means of the estimators are close to the ground truth. However, estimators for small sample sizes have large 95 $\%$ CIs.

F.2 SENSITIVITY ANALYSIS FOR VIOLATION OF MONOTONICITY

We conduct a sensitivity analysis to assess the impact of violations of the monotonicity assumption.

Setting. We consider the following SCM:

$$Y := X + M + N + C + \alpha U^Y + (1-\alpha)(U^Y)^4, N := X + M + C + U^N, M := X + C + U^M, X := C + U^X, C := U^C, \tag{97}$$

where $U^C \sim \mathcal{N}(0,1)$, $U^X \sim \mathcal{N}(0,1)$, $U^Y \sim \mathcal{N}(0,1)$, $U^M \sim \mathcal{N}(0,1)$, $U^N \sim \mathcal{N}(0,1)$, which are mutually independent normal distributions. This SCM violates the monotonicities. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$. We examine the cases $\alpha = 0.5$ and $\alpha = 0$, which correspond to a moderate violation and a strong violation of monotonicity, respectively. The case where $\alpha = 1$ corresponds to the setting described in Section 5.

Results ($\alpha = 0.5$; moderate violation). The ground truth of T-PNS is 0.365. The ground truth of PNS$^{X \to Y}$ is 0.039 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.048 \text{ (95\%CI: } [0.000, 0.188]),\\
N = 100&: \quad 0.048 \text{ (95\%CI: } [0.000, 0.115]),\\
N = 10000&: 0.048 \text{ (95\%CI: } [0.040, 0.057]).
\end{aligned}
$$

The ground truth of PNS$^{X\to N\to Y}$ is 0.073 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.053 \text{ (95\%CI: } [0.000, 0.176]),\\
N = 100&: \quad 0.056 \text{ (95\%CI: } [0.000, 0.104]),\\
N = 10000&: 0.056 \text{ (95\%CI: } [0.050, 0.062]).
\end{aligned}
$$

The ground truth of PNS$^{X\to M\to N\to Y}$ is 0.111, and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.078 \text{ (95\%CI: } [0.000, 0.214]),\\
N = 100&: \quad 0.070 \text{ (95\%CI: } [0.000, 0.135]),\\
N = 10000&: 0.063 \text{ (95\%CI: } [0.055, 0.071]).
\end{aligned}
$$

The ground truth of PNS$^{X\to M\to Y}$ is 0.142 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.101 \text{ (95\%CI: } [0.000, 0.253]),\\
N = 100&: \quad 0.080 \text{ (95\%CI: } [0.000, 0.163]),\\
N = 10000&: 0.069 \text{ (95\%CI: } [0.057, 0.081]).
\end{aligned}
$$

Results ($\alpha = 0$; strong violation). The ground truth of T-PNS is 0.330. The ground truth of PNS$^{X\to Y}$ is 0.035 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.048 \text{ (95\%CI: } [0.000, 0.201]),\\
N = 100&: \quad 0.041 \text{ (95\%CI: } [0.000, 0.125]),\\
N = 10000&: 0.032 \text{ (95\%CI: } [0.022, 0.043]).
\end{aligned}
$$

The ground truth of PNS$^{X\to N\to Y}$ is 0.064 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.038 \text{ (95\%CI: } [0.000, 0.149]),\\
N = 100&: \quad 0.034 \text{ (95\%CI: } [0.000, 0.091]),\\
N = 10000&: 0.034 \text{ (95\%CI: } [0.027, 0.042]).
\end{aligned}
$$

The ground truth of PNS$^{X\to M\to N\to Y}$ is 0.101, and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.048 \text{ (95\%CI: } [0.000, 0.184]),\\
N = 100&: \quad 0.040 \text{ (95\%CI: } [0.000, 0.106]),\\
N = 10000&: 0.044 \text{ (95\%CI: } [0.028, 0.044]).
\end{aligned}
$$

The ground truth of PNS$^{X\to M\to Y}$ is 0.130 and the estimates are

$$
\begin{aligned}
N = 20&: \quad 0.069 \text{ (95\%CI: } [0.000, 0.232]),\\
N = 100&: \quad 0.049 \text{ (95\%CI: } [0.000, 0.137]),\\
N = 10000&: 0.038 \text{ (95\%CI: } [0.026, 0.048]).
\end{aligned}
$$

The bias becomes large under strong violations of monotonicity.

F.3 BINARY OUTCOME

We conduct additional experiments using a logistic model for binary outcomes.

Setting. We consider the following SCM: Y is randomly chosen from $\{0, 1\}$ with the probability

$$\mathbb{P}(Y = 1) = \frac{1}{1 + \exp(-10(X + M + N + C))}, \tag{98}$$

and

$$N := X + M + C + U^N, M := X + C + U^M, X := C + U^X, C := U^C, \tag{99}$$

where $U^C \sim \mathcal{N}(0, 1)$, $U^X \sim \mathcal{N}(0, 1)$, $U^M \sim \mathcal{N}(0, 1)$, and $U^N \sim \mathcal{N}(0, 1)$, which are mutually independent normal distributions. This SCM satisfies Assumptions 2.1, 4.1, 4.2, 4.3, and 4.3'. We estimate the model parameters using logistic regression. We let $x' = 0$, $x = 1$, $y = 0$, $c = 0$, and $\mathcal{E} = \emptyset$. We simulate 1000 times with the sample size $N = 20$, $N = 100$, and $N = 10000$.

Results. The ground truth of T-PNS is 0.466. The ground truth of PNS$^{X\to Y}$ is 0.054 and the estimates are

$$\begin{aligned} N = 20: &\quad 0.092 \text{ (95\%CI: } [0.000, 0.950]), \\ N = 100: &\quad 0.054 \text{ (95\%CI: } [0.000, 0.159]), \\ N = 10000: &\ 0.053 \text{ (95\%CI: } [0.048, 0.058]). \end{aligned}$$

The ground truth of PNS$^{X\to N\to Y}$ is 0.098 and the estimates are

$$\begin{aligned} N = 20: &\quad 0.080 \text{ (95\%CI: } [0.000, 0.359]), \\ N = 100: &\quad 0.096 \text{ (95\%CI: } [0.013, 0.202]), \\ N = 10000: &\ 0.096 \text{ (95\%CI: } [0.089, 0.103]). \end{aligned}$$

The ground truth of PNS$^{X\to M\to N\to Y}$ is 0.141, and the estimates are

$$\begin{aligned} N = 20: &\quad 0.110 \text{ (95\%CI: } [0.000, 0.306]), \\ N = 100: &\quad 0.130 \text{ (95\%CI: } [0.000, 0.232]), \\ N = 10000: &\ 0.141 \text{ (95\%CI: } [0.137, 0.146]). \end{aligned}$$

The ground truth of PNS$^{X\to M\to Y}$ is 0.173 and the estimates are

$$\begin{aligned} N = 20: &\quad 0.149 \text{ (95\%CI: } [0.000, 0.439]), \\ N = 100: &\quad 0.182 \text{ (95\%CI: } [0.031, 0.329]), \\ N = 10000: &\ 0.178 \text{ (95\%CI: } [0.167, 0.190]). \end{aligned}$$

The estimates obtained from logistic regression are reliable when the sample size is large.

G ADDITIONAL INFORMATION ABOUT THE APPLICATION TO REAL-WORLD

Let the evidence be $\mathcal{E} = (X = 0, 10 < Y \leq 15)$, and the other settings are the same as in the body of the paper. The estimates at $(y; x', x, \mathcal{E}, c)$ are

$$\begin{aligned}
&\text{T-PNS}: && 23.950\%(\text{CI}: [0.000\%, 62.123\%]), \\
&\text{ND-PNS}^M: && 2.430\%(\text{CI}: [0.000\%, 18.587\%]), \\
&\text{NI-PNS}^M: && 21.520\%(\text{CI}: [0.000\%, 52.851\%]), \\
&\text{PNS}^{X\to Y}: && 0.354\%(\text{CI}: [0.000\%, 4.998\%]), \\
&\text{PNS}^{X\to N\to Y}: && 2.075\%(\text{CI}: [0.000\%, 16.048\%]), \\
&\text{PNS}^{X\to M\to N\to Y}: && 0.000\%(\text{CI}: [0.000\%, 0.000\%]), \\
&\text{PNS}^{X\to M\to Y}: && 21.520\%(\text{CI}: [0.000\%, 52.851\%]).
\end{aligned}$$