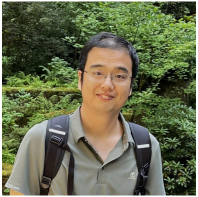
Mediation Analysis with Separately Observed Mediator and Outcome

Sijia (Lucy) Li

Assistant Professor
Department of Biostatistics, UCLA

Joint work with



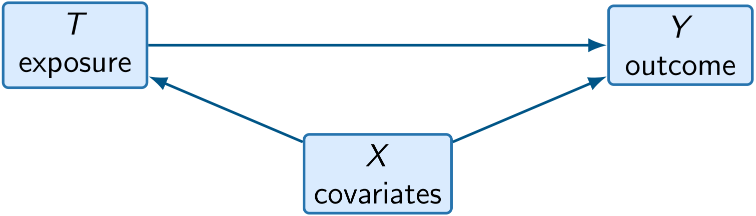
Ruoyu Wang, Ph.D.

Department of Biostatistics
Harvard University

Roadmap for today

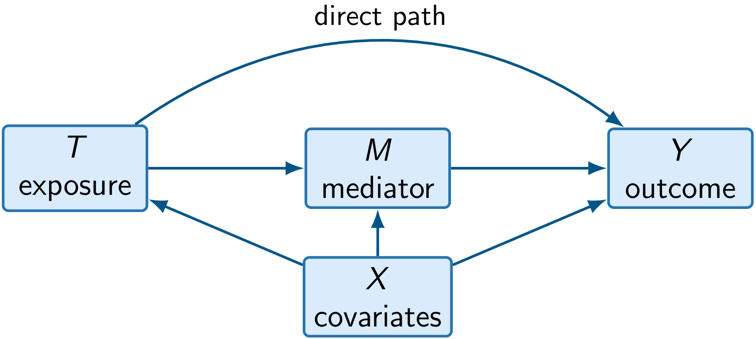
1. Background on mediation analysis
2. Two layers of challenges
3. Relationship to existing literature
4. Main results on identification and estimation
5. Simulation
6. Data Application

Mediation analysis: the usual goal



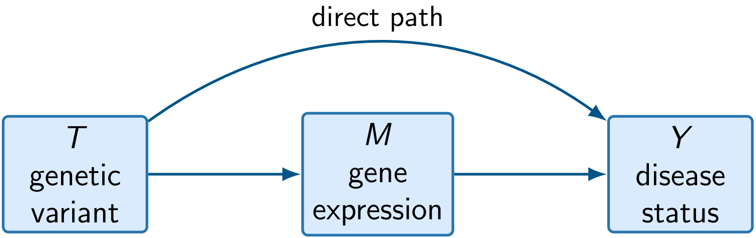
Mediation analysis starts with the total effect of T on Y.

Mediation analysis: the usual goal



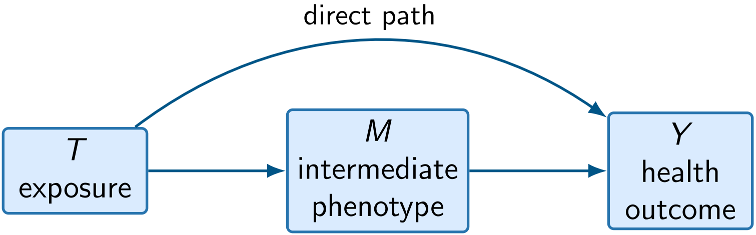
Mediation analysis asks how much of the effect of T on Y is transmitted through M.

Example I: integrative genomics



Does the genetic effect on disease operate through gene expression?

Example II: population health



How much of the exposure effect operates through biomarkers, medication use, or care processes?

Ideal setting for mediation analysis

If we observe X, T, M and Y in a single data source, under certain assumptions, we can identify[1]

- Natural direct effect (NDE):

$$\mathrm{NDE} = \mathbb{E}\{Y(1, M(0))\} - \mathbb{E}\{Y(0, M(0))\}$$

- Natural indirect effect (NIE):

$$\mathrm{NIE} = \mathbb{E}\{Y(1, M(1))\} - \mathbb{E}\{Y(1, M(0))\}$$

[1]We use the nested potential outcome $Y(t, M(t'))$ to denote the hypothetical outcome if the exposure were set at $T = t$ and the mediator were set at what would happen under $T = t'$.

Ideal setting for mediation analysis

If we observe X, T, M and Y in a single data source, under certain assumptions, we can identify[1]

- Natural direct effect (NDE):

$$\mathrm{NDE} = \mathbb{E}\{Y(1, M(0))\} - \mathbb{E}\{Y(0, M(0))\}$$

- Natural indirect effect (NIE):

$$\mathrm{NIE} = \mathbb{E}\{Y(1, M(1))\} - \mathbb{E}\{Y(1, M(0))\}$$

[1] We use the nested potential outcome $Y(t, M(t'))$ to denote the hypothetical outcome if the exposure were set at $T = t$ and the mediator were set at what would happen under $T = t'$.

Ideal setting: identification assumptions

Assumption 1. (Sequential ignorability Imai et al. (2010b))

$$\{Y(t', m), M(t)\} \perp\!\!\!\perp T \mid X$$
$$Y(t', m) \perp\!\!\!\perp M(t) \mid T = t, X$$

for $t, t' = \{0, 1\}$ and all $x \in \mathcal{X}$ while also assuming $0 < \Pr(T = t \mid X = x)$ and $0 < p(M(t) = m \mid T = t, X = x)$ for all $m \in \mathcal{M}$.

Ideal setting: identification assumptions

Assumption 1. (Sequential ignorability Imai et al. (2010b))

$$\{Y(t', m), M(t)\} \perp\!\!\!\perp T \mid X$$
$$Y(t', m) \perp\!\!\!\perp M(t) \mid T = t, X$$

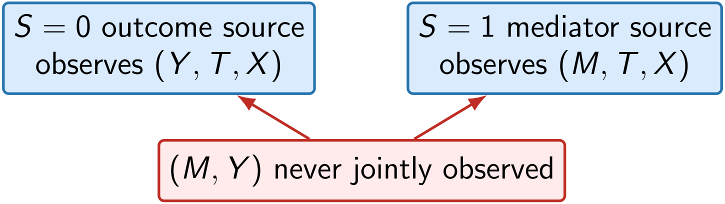
for $t, t' = \{0, 1\}$ and all $x \in \mathcal{X}$ while also assuming $0 < \Pr(T = t \mid X = x)$ and $0 < p(M(t) = m \mid T = t, X = x)$ for all $m \in \mathcal{M}$.

Key idea: Sequential ignorability turns the cross-world quantity into observed-data functionals:

$$\mathbb{E}\{Y(t, M(t'))\} = \int E[Y \mid T = t, M = m, X = x] dF_{M \mid T = t', X = x}(m) dF_X(x)$$

Practical barrier #1

In many settings, the mediator and outcome are measured in **different data sources**.



Practical barrier #1

In many settings, the mediator and outcome are measured in **different data sources**.

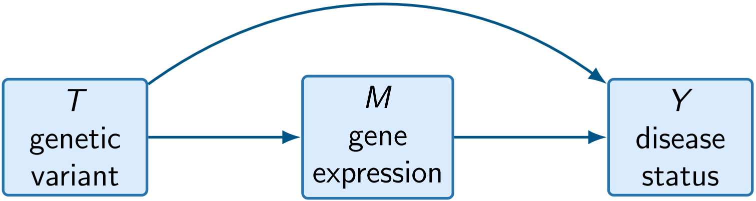
$S = 0$ outcome source observes (Y, T, X)

$S = 1$ mediator source observes (M, T, X)

(M, Y) never jointly observed

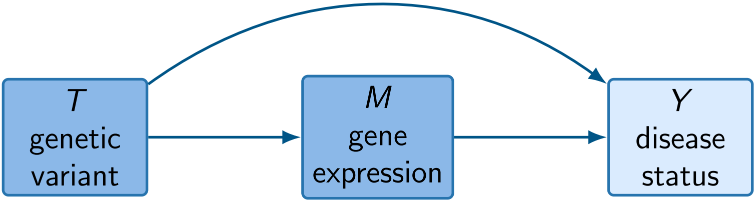
As a result, we can't estimate $E[Y \mid M, T, X]$.

Example I: integrative genomics



Does the genetic effect on disease operate through gene expression?

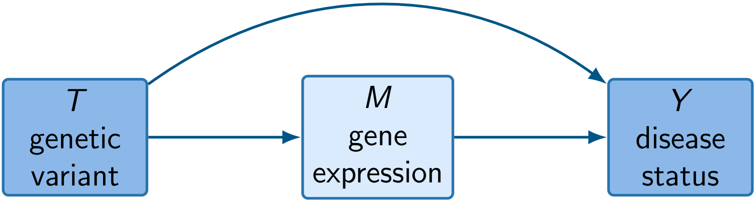
Example I: integrative genomics



$S = 1$
eQTL studies

Does the genetic effect on disease operate through gene expression?

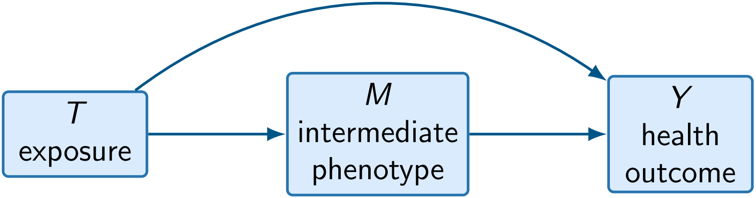
Example I: integrative genomics



$S = 0$
GWAS studies

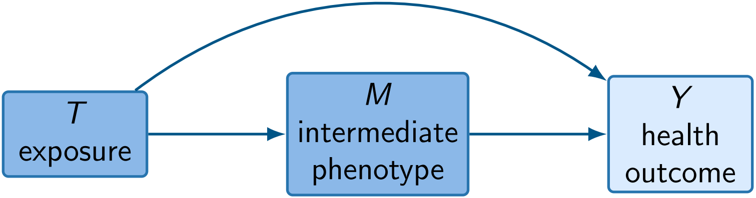
Does the genetic effect on disease operate through gene expression?

Example II: population health



How much of the exposure effect operates through biomarkers, medication use, or care processes?

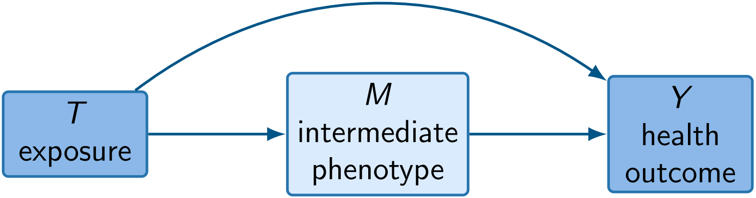
Example II: population health



$S = 1$
Dedicated studies

How much of the exposure effect operates through biomarkers, medication use, or care processes?

Example II: population health

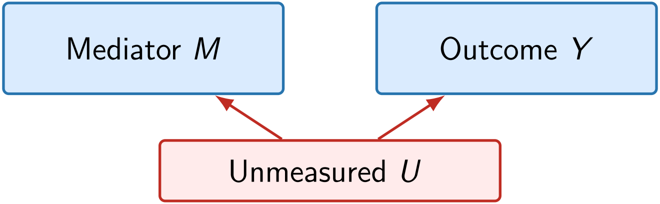


$S = 0$
observational cohorts

How much of the exposure effect operates through biomarkers, medication use, or care processes?

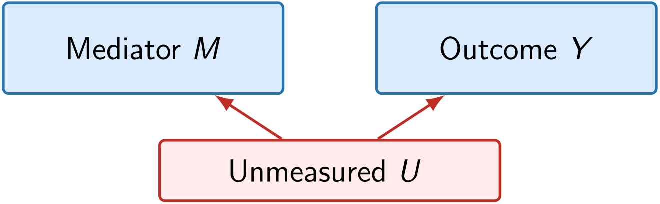
Practical barrier #2

The presence of unmeasured mediator–outcome confounding.



Practical barrier #2

The presence of unmeasured mediator–outcome confounding.



Even if (M, Y) are jointly observed,

$$\mathbb{E}\{Y(t, M(t'))\} \neq \int \mathbb{E}[Y \mid T = t, M = m, X = x] dF_{M \mid T = t', X = x}(m) dF_X(x)$$

This is a data fusion problem

We observe two complementary samples:

$$S = 0 : \quad (Y, T, X), \qquad S = 1 : \quad (M, T, X).$$

The challenge: No jointly observed (M, Y) and unmeasured $M - Y$ confounding

Goal: identify natural direct and indirect effects for a target population (for instance, $S = 0$).

How?

This is a data fusion problem

We observe two complementary samples:

$$S = 0 : \quad (Y, T, X), \qquad S = 1 : \quad (M, T, X).$$

The challenge: No jointly observed (M, Y) and unmeasured $M - Y$ confounding

Goal: identify natural direct and indirect effects for a target population (for instance, $S = 0$).

How? somehow **bridge** the two data sources...

Prior works on data fusion

Common theme: Express the target estimand as a functional of several components of the target distribution Q, i.e., $Q = P(\cdot \mid S = 0)$.

Prior works on data fusion

Common theme: Express the target estimand as a functional of several components of the target distribution Q, i.e., $Q = P(\cdot \mid S = 0)$.

Target estimand (assume $U = \emptyset$):

$$\text{NIE} := \phi\Big(Q_{Y|M,T,X}, Q_{M|T,X}, Q_X \Big) .$$

Prior works on data fusion

Common theme: Express the target estimand as a functional of several components of the target distribution Q, i.e., $Q = P(\cdot \mid S = 0)$.

Target estimand (assume $U = \emptyset$):

$$\mathrm{NIE} := \phi\Big(Q_{Y|M,T,X}, Q_{M|T,X}, Q_X \Big).$$

Then each component must be identified via at least one data source.

Prior works on data fusion

Common theme: Express the target estimand as a functional of several components of the target distribution Q, i.e., $Q = P(\cdot \mid S = 0)$.

Target estimand (assume $U = \emptyset$):

$$\mathrm{NIE} := \phi\Big(Q_{Y|M,T,X}, Q_{M|T,X}, Q_X \Big) .$$

Then each component must be identified via at least one data source. We need to find data sources *aligning* with Q in terms of

- $Q_{Y|M,T,X}$: data source $S = \{2, 4, 5\}$
- $Q_{M|T,X}$, Q_X: data source $S = \{1, 2\}$

Prior works on data fusion

Many existing data fusion work falls under this common theme:

- Intermediate variables (surrogates, short-term outcome): Huang et al. (2023); Athey et al. (2025); Kallus and Mao (2025)

Prior works on data fusion

Many existing data fusion work falls under this common theme:

- Intermediate variables (surrogates, short-term outcome): Huang et al. (2023); Athey et al. (2025); Kallus and Mao (2025)
- General recipes applicable to mediation effects: Li and Luedtke (2023); Graham et al. (2024)

Prior works on data fusion

Many existing data fusion work falls under this common theme:

- Intermediate variables (surrogates, short-term outcome): Huang et al. (2023); Athey et al. (2025); Kallus and Mao (2025)
- General recipes applicable to mediation effects: Li and Luedtke (2023); Graham et al. (2024)

These methods fall short, since

- $Q_{Y|M,T,X}$: data source $S = ?$

Prior works on data fusion

Many existing data fusion work falls under this common theme:

- Intermediate variables (surrogates, short-term outcome): Huang et al. (2023); Athey et al. (2025); Kallus and Mao (2025)
- General recipes applicable to mediation effects: Li and Luedtke (2023); Graham et al. (2024)

These methods fall short, since

- $Q_{Y|M,T,X}$: data source $S =?$

In other words, the alignment is **latent** in our setting.

Relationship with missing data literature

A missing-at-random view would require positivity.

Here, we violate that positivity condition as the probability of observing (M, Y) is zero.

Source	M	Y
Outcome source $S = 0$	–	observed
Mediator source $S = 1$	observed	–

Relationship with missing data literature

A missing-at-random view would require positivity.

Here, we violate that positivity condition as the probability of observing (M, Y) is zero.

Source	M	Y
Outcome source $S = 0$	–	observed
Mediator source $S = 1$	observed	–

This is a different setting from missing data (Robins et al., 1994; Tsiatis, 2006; Wang et al., 2009), and semi-supervised learning under blockwise missingness (Xu et al., 2025b,a).

Prior works on "extreme form of missing data"

These work consider settings closer to ours.

- Derkach et al. (2024) considered fusing sources with each containing only two of the three key variables, and relied on a parametric form of the distribution of $Y \mid M, T, X$.

Prior works on “extreme form of missing data”

These work consider settings closer to ours.

- Derkach et al. (2024) considered fusing sources with each containing only two of the three key variables, and relied on a parametric form of the distribution of $Y \mid M, T, X$.
- Evans et al. (2021) proposed estimators under a regression setting when Y and two covariates of interest (V, L) are never jointly observed, and relied on parametric form of the distribution of $Y \mid V, L$.

Our contributions

Existing approaches for mediation analysis often require at least one of the following:

- **Jointly observed (M,Y)**
- **No unmeasured M-Y confounding**

Our contributions

Existing approaches for mediation analysis often require at least one of the following:

- **Jointly observed (M,Y)**
- **No unmeasured M-Y confounding**

In this work, we develop a **semiparametric** approach that tackles the **two aforementioned challenges** simultaneously:

- **New identification strategies**
- **Efficient & robust estimators with theoretical guarantees**

Our proposal

The use of shared instrumental variables (IVs)

$$S = 0: \quad (Y, T, \mathbf{Z}, X), \qquad S = 1: \quad (M, T, \mathbf{Z}, X).$$

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

- **Average treatment effect** under unmeasured confounding: (Wang and Tchetgen Tchetgen, 2018; Dong et al., 2025; Chen et al., 2025)

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

- **Average treatment effect** under unmeasured confounding: (Wang and Tchetgen Tchetgen, 2018; Dong et al., 2025; Chen et al., 2025)
- **Mediation effects** for specific compliance groups: (Imai et al., 2013; Frölich and Huber, 2017; Rudolph et al., 2021, 2024)

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

Data-fusion role: bridge complementary samples

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

Data-fusion role: bridge complementary samples

Two-sample IV settings use a shared Z to bridge sources:

$$\text{Sample 1: } (Z, T), \qquad \text{Sample 2: } (Z, Y).$$

- Two-sample IVs: (Angrist and Krueger, 1992; Inoue and Solon, 2010; Zhao et al., 2019; Sun and Miao, 2022)

Roles of instrumental variables

$$Z \not\perp T, \qquad Z \nrightarrow Y \text{ except through } T, \qquad Z \perp U \mid X.$$

Classical role: address unmeasured confounding

Data-fusion role: bridge complementary samples

Our identification strategy in a nutshell

Use instrumental variables Z to align

$$M \mid T = t, X, S = 0 \quad \rightsquigarrow \quad M \mid T = t', X, S = 0$$

Applying the reweighting to source $S = 0$ identifies $\mathbb{E}\{Y(t, M(t')) \mid S = 0\}$ under a no-interaction condition and a latent exchangeability condition.

Our identification strategy in a nutshell

Use instrumental variables Z to align

$$M \mid T = t, X, S = 0 \quad \rightsquigarrow \quad M \mid T = t', X, S = 0$$

Applying the reweighting to source $S = 0$ identifies $\mathbb{E}\{Y(t, M(t')) \mid S = 0\}$ under a no-interaction condition and a latent exchangeability condition.

Observed data structure

Let $O = (S, X, T, Z, (1 - S)Y, SM) \sim P \in \mathcal{P}$.

Equivalently:

$$S = 0 : (Y, T, Z, X), \qquad S = 1 : (M, T, Z, X).$$

- $S \in \{0, 1\}$: data source indicator
- $X \in \mathbb{R}^d$: covariates
- $T \in \{0, 1\}$: binary exposure; $Z \in \mathbb{R}^{|\mathcal{Z}|}$: shared instruments
- $M \in \mathbb{R}$: mediator; $Y \in \mathbb{R}$: outcome
- U: unmeasured mediator-outcome confounder

Target estimands

We treat $S = 0$ as our target population. Let

$$\phi^{tt'} = \mathbb{E}\{Y(t, M(t')) \mid S = 0\}.$$

Then

$$\text{NDE} = \phi^{10} - \phi^{00}, \qquad \text{NIE} = \phi^{11} - \phi^{10}.$$

Key difficulty: identifying cross-world term

$$\phi^{10} = \mathbb{E}\{Y(1, M(0)) \mid S = 0\}.$$

Structural Causal Model

We assume that there exists deterministic functions f_M and f_Y such that

$$M = f_M(T, Z, X, U, \varepsilon_M),$$
$$Y = f_Y(M, T, X, U, \varepsilon_Y),$$

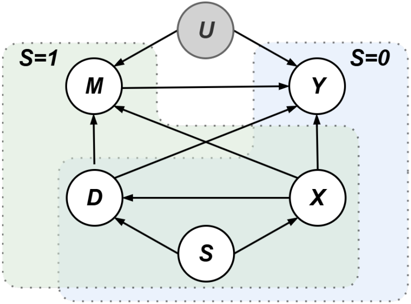


Figure: An illustration of the data generating process. $D = (T, Z)$

Identification ingredients

Our identification strategies rest on the following assumptions.

1. $(T, Z) \perp\!\!\!\perp (\varepsilon_Y, \varepsilon_M, U) \mid X, S$.
2. $\varepsilon_Y \perp\!\!\!\perp \varepsilon_M \mid X, U, S$.
3. $\epsilon < P(T = 1 \mid z, a, x, s) < 1 - \epsilon$ for any z, a, x, s P-everywhere.

Identification ingredients

Our identification strategies rest on the following assumptions.

1. $(T, Z) \perp\!\!\!\perp (\varepsilon_Y, \varepsilon_M, U) \mid X, S$.
2. $\varepsilon_Y \perp\!\!\!\perp \varepsilon_M \mid X, U, S$.
3. $\epsilon < P(T = 1 \mid z, a, x, s) < 1 - \epsilon$ for any z, a, x, s P-everywhere.

Under the proposed SCM with $U = \emptyset$, these assumptions reduce to classical sequential ignorability (Imai et al., 2010a).

Identification ingredients

Our identification strategies rest on the following assumptions.

1. $(T, Z) \perp\!\!\!\perp (\varepsilon_Y, \varepsilon_M, U) \mid X, S$.
2. $\varepsilon_Y \perp\!\!\!\perp \varepsilon_M \mid X, U, S$.
3. $\epsilon < P(T = 1 \mid z, a, x, s) < 1 - \epsilon$ for any z, a, x, s P-everywhere.
4. $E[Y \mid M = m, T, Z, X, U, S] - E[Y \mid M = m', T, Z, X, U, S]$ does not depend on U.

To accommodate U, we assume conditional on (T, Z, X, S), U is not an effect modifier.

Identification ingredients

Our identification strategies rest on the following assumptions.

1. $(T, Z) \perp\!\!\!\perp (\varepsilon_Y, \varepsilon_M, U) \mid X, S$.
2. $\varepsilon_Y \perp\!\!\!\perp \varepsilon_M \mid X, U, S$.
3. $\epsilon < P(T = 1 \mid z, a, x, s) < 1 - \epsilon$ for any z, a, x, s P-everywhere.
4. $E[Y \mid M = m, T, Z, X, U, S] - E[Y \mid M = m', T, Z, X, U, S]$ does not depend on U.
5. $c < P(S = 1 \mid T, Z, X) < 1 - c \quad P$-a.s. In addition, $S \perp\!\!\!\perp (U, M, Y) \mid T, Z, X$.

We additionally need cross-source exchangeability conditions linking the mediator and outcome sources.

What does Z do?

Condition 1 (Completeness). For any s, t, x, and any square-integrable function h,

$$E[h(M) \mid T = t, Z, X = x, S = s] = 0$$

almost surely implies $h = 0$ almost surely.

What does Z do?

Condition 1 (Completeness). For any s, t, x, and any square-integrable function h,

$$E[h(M) \mid T = t, Z, X = x, S = s] = 0$$

almost surely implies $h = 0$ almost surely.

This means: any variability in M is captured by variability in Z.

What does Z do?

Condition 1 (Completeness). For any s, t, x, and any square-integrable function h,

$$E[h(M) \mid T = t, Z, X = x, S = s] = 0$$

almost surely implies $h = 0$ almost surely.

This means: any variability in M is captured by variability in Z.

Great discussions on the role of completeness condition: Miao et al. (2023); Guo et al. (2025), and on alternative conditions: Bennett et al. (2023, 2025)

Bridge functions

Lemma 1. (Existence of bridge functions w). Under Condition 1 and additional regularity conditions, there exists some function vector $w = (w^{11}, w^{10}, w^{00})$ that satisfies

$$\int w^{tt'}(z,x) p_{M|T,Z,X,S}(m \mid t,z,x,1) p_{T,Z|X,S}(t,z \mid x,0) dz$$
$$= \int p_{M|T,Z,X,S}(m \mid t',z,x,1) p_{Z|X,S}(z \mid x,0) dz$$

for $t, t' \in \{0,1\}$ and x P-everywhere.

Bridge functions

Lemma 1. (Existence of bridge functions w). Under Condition 1 and additional regularity conditions, there exists some function vector $w = (w^{11}, w^{10}, w^{00})$ that satisfies

$$\int w^{tt'}(z,x)\, p_{M|T,Z,X,S}(m \mid t, z, x, 1)\, p_{T,Z|X,S}(t, z \mid x, 0) dz$$
$$= \int p_{M|T,Z,X,S}(m \mid t', z, x, 1)\, p_{Z|X,S}(z \mid x, 0) dz$$

for $t, t' \in \{0, 1\}$ and x P-everywhere.

Bridge functions

Lemma 1. (Existence of bridge functions w). Under Condition 1 and additional regularity conditions, there exists some function vector $w = (w^{11}, w^{10}, w^{00})$ that satisfies

$$\int w^{tt'}(z,x) p_{M|T,Z,X,S}(m \mid t,z,x,1)\, p_{T,Z|X,S}(t,z \mid x,0)\, dz$$
$$= \int p_{M|T,Z,X,S}(m \mid t',z,x,1)\, p_{Z|X,S}(z \mid x,0)\, dz$$

for $t, t' \in \{0,1\}$ and x P-everywhere.

Bridge functions

Lemma 1. (Existence of bridge functions w). Under Condition 1 and additional regularity conditions, there exists some function vector $w = (w^{11}, w^{10}, w^{00})$ that satisfies

$$\int w^{tt'}(z,x) p_{M|T,Z,X,S}(m \mid t,z,x,1) p_{T,Z|X,S}(t,z \mid x,0) dz$$
$$= \int p_{M|T,Z,X,S}(m \mid t',z,x,1) p_{Z|X,S}(z \mid x,0) dz$$

for $t, t' \in \{0,1\}$ and x P-everywhere.

The bridge function $w^{tt'}$ reweighs the mediator strata $M, T = t \mid X, S = 0$ such that it aligns with $M(t') \mid X, S = 0$.

Remarks on bridge functions

When $t = t'$, w admits the simple solutions with $w^{tt}(z, x) = 1/P(T = t \mid z, a, x, S = 0)$. More generally, however, solutions are not unique as uniqueness holds only if Z is also complete in M.

Remarks on bridge functions

When $t = t'$, w admits the simple solutions with $w^{tt}(z, x) = 1/P(T = t \mid z, a, x, S = 0)$. More generally, however, solutions are not unique as uniqueness holds only if Z is also complete in M.

Importantly, this lack of uniqueness is harmless for identification as all valid bridge functions w will lead to the same reweighted counterfactual mediator distribution.

Identification result

Theorem 1. Suppose $E[\{w^{tt'}(Z,X)\}^2] < \infty$ for $t, t' \in \{0,1\}$. Under Assumptions 1-5 and Condition 1, NDE and NIE can be identified as $\psi^{\mathrm{NDE}}(P) = \psi^{10}(P) - \psi^{00}(P)$ and $\psi^{\mathrm{NIE}}(P) = \psi^{11}(P) - \psi^{10}(P)$, where ψ^{11}, ψ^{10}, ψ^{00}: $\mathcal{P} \to \mathbb{R}$ and

$$
\begin{aligned}
\psi^{11}(P) &:= E_P\left[E_P[Y \mid T = 1, Z, X, S = 0] \mid S = 0\right] \\
\psi^{10}(P) &:= E_P\left[E_P\left[\mathbb{1}(T = 1)w^{10}(Z,X)Y \mid X, S = 0\right] \mid S = 0\right] \\
\psi^{00}(P) &:= E_P\left[E_P[Y \mid T = 0, Z, X, S = 0] \mid S = 0\right].
\end{aligned}
$$

Identification result with unknown IV set

We have also investigated the more challenging case where the set of valid IVs must be learned.

When the index set of IVs is unknown, it is still possible to identify a subset sufficient for identification under additional assumptions. We establish the corresponding identification strategy and propose a reproducing kernel-based nonparametric IV selection algorithm.

We have identified our target estimands using two incomplete data sources.

We have identified our target estimands using two incomplete data sources.

Now let us study how to estimate them.

Constructing our estimator

General recipe for estimating $\psi(P)$:

- Suppose we have an initial estimator $\hat{P}$ of P, and a gradient $D_{\hat{P}}$ of ψ.
- We then construct a one-step estimator using observed data $O \sim P$

$$\hat{\psi} := \psi(\hat{P}) + \frac{1}{n}\sum_{i=1}^{n} D_{\hat{P}}(O_i)$$

- This estimator will be maximally efficient when $D_{\hat{P}}$ is the so-called canonical gradient of the parameter ψ.

A set of gradients

Lemma 2. Under Assumptions 1-5 and Condition 1, a set of gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}$ at P are:

$$\begin{aligned}
D_P^{11}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)}\left\{\mathbb{1}(t=1)\,\mathrm{w}^{11}(z,x)(y-\mu_1(z,x)) + \mu_1(z,x) - \psi^{11}(P)\right. \\
D_P^{10}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)}\left\{\mathbb{1}(t=1)\,\mathrm{w}^{10}(z,x)\ (y-\mu_1(z,x)) + \mu_{10}(z,x) - \psi^{10}(P)\right. \\
&\quad + \frac{\mathbb{1}(s=1)}{P(S=1)}\mathbb{1}(t=0)\mathrm{w}^{00}(z,x)\ \lambda(t,z,x)\,(\nu_1(m,x) - \mu_{10}(z,x)) \\
&\quad - \frac{\mathbb{1}(s=1)}{P(S=1)}\mathbb{1}(t=1)\lambda(t,z,x)\mathrm{w}^{10}(z,x)(\nu_1(m,x) - \mu_1(z,x)) \\
D_P^{00}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)}\left\{\mathbb{1}(t=0)\,\mathrm{w}^{00}(z,x)(y-\mu_0(z,x)) + \mu_0(z,x) - \psi^{00}(P),\right.
\end{aligned}$$

where $\mu_t(Z,X) := E_P[Y \mid T=t, Z, X, S=0]$, $\nu_t(M,X)$ be the solution to $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t(Z,X)$, where $\mathbb{M}_{\nu_t}(t',Z,X) := E_P[\nu_t(M,X) \mid T=t', Z, X, S=1]$, $\mu_{10}(Z,X) := \mathbb{M}_{\nu_1}(0,Z,X)$, and $\lambda(T,Z,X) := \frac{p(T,Z,X|S=0)}{p(T,Z,X|S=1)}$.

A set of gradients

Lemma 2. Under Assumptions 1-5 and Condition 1, a set of gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}$ at P are:

$$D_P^{11}(o) = \frac{\mathbb{1}(s=0)}{P(S=0)} \{\mathbb{1}(t=1)\,\mathrm{w}^{11}(z,x)(y-\mu_1(z,x)) + \mu_1(z,x) - \psi^{11}(P)$$

$$D_P^{00}(o) = \frac{\mathbb{1}(s=0)}{P(S=0)} \{\mathbb{1}(t=0)\,\mathrm{w}^{00}(z,x)(y-\mu_0(z,x)) + \mu_0(z,x) - \psi^{00}(P),$$

where $\mu_t(Z,X) := E_P[Y \mid T=t, Z, X, S=0]$, $\nu_t(M,X)$ be the solution to $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t(Z,X)$, where $\mathbb{M}_{\nu_t}(t',Z,X) := E_P[\nu_t(M,X) \mid T=t', Z, X, S=1]$, $\mu_{10}(Z,X) := \mathbb{M}_{\nu_1}(0,Z,X)$, and $\lambda(T,Z,X) := \frac{p(T,Z,X|S=0)}{p(T,Z,X|S=1)}$.

A set of gradients

Lemma 2. Under Assumptions 1-5 and Condition 1, a set of gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}$ at P are:

$$\begin{aligned} D_P^{10}(o) = & \frac{\mathbb{1}(s=0)}{P(S=0)} \{ \mathbb{1}(t=1)\, \mathrm{w}^{10}(z,x)\ (y - \mu_1(z,x)) + \mu_{10}(z,x) - \psi^{10}(P) \\ & + \frac{\mathbb{1}(s=1)}{P(S=1)} \mathbb{1}(t=0) \mathrm{w}^{00}(z,x)\ \lambda(t,z,x)\, (\nu_1(m,x) - \mu_{10}(z,x)) \\ & - \frac{\mathbb{1}(s=1)}{P(S=1)} \mathbb{1}(t=1) \lambda(t,z,x) \mathrm{w}^{10}(z,x) (\nu_1(m,x) - \mu_1(z,x)) \end{aligned}$$

,

where $\mu_t(Z,X) := E_P[Y \mid T=t, Z, X, S=0]$, $\nu_t(M,X)$ be the solution to $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t(Z,X)$, where $\mathbb{M}_{\nu_t}(t',Z,X) := E_P[\nu_t(M,X) \mid T=t', Z, X, S=1]$, $\mu_{10}(Z,X) := \mathbb{M}_{\nu_1}(0,Z,X)$, and $\lambda(T,Z,X) := \frac{p(T,Z,X|S=0)}{p(T,Z,X|S=1)}$.

A set of gradients

Lemma 2. Under Assumptions 1-5 and Condition 1, a set of gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}$ at P are:

$$
\begin{aligned}
D_P^{11}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)} \left\{ \mathbb{1}(t=1)\, w^{11}(z,x) \,(y-\mu_1(z,x)) + \mu_1(z,x) - \psi^{11}(P) \right\} \\
D_P^{10}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)} \left\{ \mathbb{1}(t=1)\, w^{10}(z,x) \,(y-\mu_1(z,x)) + \mu_{10}(z,x) - \psi^{10}(P) \right\} \\
&\quad + \frac{\mathbb{1}(s=1)}{P(S=1)} \mathbb{1}(t=0)\, w^{00}(z,x) \lambda(t,z,x) \,(\nu_1(m,x) - \mu_{10}(z,x)) \\
&\quad - \frac{\mathbb{1}(s=1)}{P(S=1)} \mathbb{1}(t=1) \lambda(t,z,x)\, w^{10}(z,x) \,(\nu_1(m,x) - \mu_1(z,x)) \\
D_P^{00}(o) &= \frac{\mathbb{1}(s=0)}{P(S=0)} \left\{ \mathbb{1}(t=0)\, w^{00}(z,x) \,(y-\mu_0(z,x)) + \mu_0(z,x) - \psi^{00}(P) \right\},
\end{aligned}
$$

where $\mu_t(Z,X) := E_P[Y \mid T=t, Z, X, S=0]$, $\nu_t(M,X)$ be the solution to $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t(Z,X)$, where $\mathbb{M}_{\nu_t}(t',Z,X) := E_P[\nu_t(M,X) \mid T=t', Z, X, S=1]$, $\mu_{10}(Z,X) := \mathbb{M}_{\nu_1}(0,Z,X)$, and $\lambda(T,Z,X) := \frac{p(T,Z,X \mid S=0)}{p(T,Z,X \mid S=1)}$.

Special case: mutual completeness

Condition 1 (M complete in Z). For any s, t, x, and any square-integrable function h,

$$E[h(M) \mid T = t, Z, X = x, S = s] = 0$$

almost surely implies $h = 0$ almost surely.

Condition 2 (Z complete in M). For any t, x, and any square-integrable function h,

$$E[h(Z) \mid T = t, M, X = x, S = 0] = 0$$

almost surely implies $h = 0$ almost surely.

Special case: mutual completeness

When both directions of completeness holds, bridge functions are unique.

$$w^{11}(Z, X) = \frac{1}{\pi(Z, X)}, \qquad w^{00}(Z, X) = \frac{1}{1 - \pi(Z, X)},$$

where $\pi(Z, X) = P(T = 1 \mid Z, X)$.

This clarifies when standard inverse-propensity weighting is already efficient, and when additional bridge-based information can improve efficiency.

Special case: mutual completeness

Theorem 1 (Mutual completeness implies efficiency).
Let $\mathcal{P}_{\text{complete}} \subset \mathcal{P}$ denote the set of distributions that satisfies Assumptions 1-5 and Conditions 1-2. Under a common mediator support condition, $D_P^{11}, D_P^{10}, D_P^{00}$ are the canonical gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}_{\text{complete}}$ at P.

Special case: mutual completeness

Theorem 1 (Mutual completeness implies efficiency).
Let $\mathcal{P}_{\text{complete}} \subset \mathcal{P}$ denote the set of distributions that satisfies Assumptions 1-5 and Conditions 1-2. Under a common mediator support condition, $D_P^{11}, D_P^{10}, D_P^{00}$ are the canonical gradients of ψ^{11}, ψ^{10} and ψ^{00} relative to $\mathcal{P}_{\text{complete}}$ at P.

Heuristics: model constraints are imposed on the full data model where U is observed. The tangent space $\mathcal{T}(P, \mathcal{P}_{\text{complete}}) = L_2^0(P)$.

Construction of estimators

To construct one-step estimators for $\psi^{\mathrm{NDE}}(P)$ and $\psi^{\mathrm{NIE}}(P)$, we need to estimate components of P that are needed for the evaluation of D_P^{NDE} and D_P^{NIE}.

[2] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

To construct one-step estimators for $\psi^{\mathrm{NDE}}(P)$ and $\psi^{\mathrm{NIE}}(P)$, we need to estimate components of P that are needed for the evaluation of D_P^{NDE} and D_P^{NIE}.

The list of nuisance parameters includes $(\mathbb{M}_\nu, \nu, \lambda, w)$, where we define $\mathbb{M}_\nu := (\mathbb{M}_{\nu_0}, \mathbb{M}_{\nu_1})$ and $\nu := (\nu_0, \nu_1)$. Under a generic $\widehat{P}$, we denote the corresponding estimates as $(\widehat{\mathbb{M}}_\nu, \widehat{\nu}, \widehat{\lambda}, \widehat{w})$. [2]

[2] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

$\widehat{\Psi}^{\mathrm{NDE}} = \widehat{\psi}^{10} - \widehat{\psi}^{00}$ and $\widehat{\Psi}^{\mathrm{NIE}} = \widehat{\psi}^{11} - \widehat{\psi}^{10}$ where[3]

$$\widehat{\psi}^{11} = \mathbb{P}_n\left[\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=1)\widehat{w}^{11}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right\}\right]$$

$$\begin{aligned}\widehat{\psi}^{10} = \mathbb{P}_n\Bigg[&\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=1)\widehat{w}^{10}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_1}(0,Z,G)\right\}\\ &+ \frac{\mathbb{1}(S=1)}{\widehat{P}(S=1)}\mathbb{1}(T=0)\widehat{w}^{00}(Z,G)\widehat{\lambda}(T,Z,G)\left(\widehat{\nu}_1(M,G) - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(0,Z,G)\right)\\ &- \frac{\mathbb{1}(S=1)}{\widehat{P}(S=1)}\mathbb{1}(T=1)\widehat{\lambda}(T,Z,G)\widehat{w}^{10}(Z,G)\left(\widehat{\nu}_1(M,G) - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right)\Bigg]\end{aligned}$$

$$\widehat{\psi}^{00} = \mathbb{P}_n\left[\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=0)\widehat{w}^{00}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_0}(0,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_0}(0,Z,G)\right\}\right].$$

[3] $\lambda = p(T,Z,X \mid S=0)/p(T,Z,X \mid S=1)$; $\mu_t = E[Y \mid T=t, Z, X, S=0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M,X) \mid T=t', Z, X, S=1]$; ν_t solves $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t$.

Construction of estimators

$\widehat{\Psi}^{\mathrm{NDE}} = \widehat{\psi}^{10} - \widehat{\psi}^{00}$ and $\widehat{\Psi}^{\mathrm{NIE}} = \widehat{\psi}^{11} - \widehat{\psi}^{10}$ where[3]

$$\widehat{\psi}^{11} = \mathbb{P}_n\left[\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=1)\widehat{w}^{11}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right\}\right]$$

$$\begin{aligned}\widehat{\psi}^{10} = \mathbb{P}_n\Bigg[&\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=1)\widehat{w}^{10}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_1}(0,Z,G)\right\}\\ &+ \frac{\mathbb{1}(S=1)}{\widehat{P}(S=1)}\mathbb{1}(T=0)\widehat{w}^{00}(Z,G)\widehat{\lambda}(T,Z,G)\left(\widehat{\nu}_1(M,G) - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(0,Z,G)\right)\\ &- \frac{\mathbb{1}(S=1)}{\widehat{P}(S=1)}\mathbb{1}(T=1)\widehat{\lambda}(T,Z,G)\widehat{w}^{10}(Z,G)\left(\widehat{\nu}_1(M,G) - \widehat{\mathbb{M}}_{\widehat{\nu}_1}(1,Z,G)\right)\Bigg]\end{aligned}$$

$$\widehat{\psi}^{00} = \mathbb{P}_n\left[\frac{\mathbb{1}(S=0)}{\widehat{P}(S=0)}\left\{\mathbb{1}(T=0)\widehat{w}^{00}(Z,G)\left(Y - \widehat{\mathbb{M}}_{\widehat{\nu}_0}(0,Z,G)\right) + \widehat{\mathbb{M}}_{\widehat{\nu}_0}(0,Z,G)\right\}\right].$$

How to come up with these nuisance estimates?

[3] $\lambda = p(T,Z,X \mid S=0)/p(T,Z,X \mid S=1)$; $\mu_t = E[Y \mid T=t,Z,X,S=0]$; $\mathbb{M}_{\nu_t}(t',Z,X) = E[\nu_t(M,X) \mid T=t',Z,X,S=1]$; ν_t solves $\mathbb{M}_{\nu_t}(t,Z,X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_{\nu}, \nu, \lambda, w)$$

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_\nu, \nu, \lambda, w)$$

- λ can be estimated directly using both data sources.

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_\nu, \nu, \lambda, w)$$

- λ can be estimated directly using both data sources.
- $\mathbb{M}_\nu$ can be estimated using samples from $S = 1$.

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_\nu, \nu, \lambda, w)$$

- λ can be estimated directly using both data sources.
- $\mathbb{M}_\nu$ can be estimated using samples from $S = 1$.
- ν and w solve Fredholm integral equations of the first kind.

Concretely:

- $\widehat{\mathbb{M}}_{\widehat{\nu}_1}(1, Z, G) = \widehat{\mu}_1(Z, G)$ and $\widehat{\mathbb{M}}_{\widehat{\nu}_0}(0, Z, G) = \widehat{\mu}_0(Z, G)$.
- w can be estimated from its defining equation.

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_\nu, \nu, \lambda, w)$$

- λ can be estimated directly using both data sources.
- $\mathbb{M}_\nu$ can be estimated using samples from $S = 1$.
- ν and w solve Fredholm integral equations of the first kind.

These equations also arise in nonparametric IV models (Newey and Powell, 2003; Darolles et al., 2011) and and proximal causal inference (Miao et al., 2018; Tchetgen Tchetgen et al., 2024).

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Construction of estimators

List of nuisance parameters[4]:

$$(\mathbb{M}_\nu, \nu, \lambda, w)$$

- λ can be estimated directly using both data sources.
- $\mathbb{M}_\nu$ can be estimated using samples from $S = 1$.
- ν and w solve Fredholm integral equations of the first kind.

Under our identification conditions, the solution for ν is unique, while w may admit multiple solutions. We regularize w using a variance-based penalty or a direct norm penalty.

[4] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Multiple robustness of $\widehat{\Psi}^{\mathrm{NDE}}$ and $\widehat{\Psi}^{\mathrm{NIE}}$

Theorem 2. Assume that $(\widehat{\mathbb{M}}_{\nu}, \widehat{\nu}, \widehat{\lambda}, \widehat{w}) \xrightarrow{p} (\overline{\mathbb{M}}_{\nu}, \bar{\nu}, \bar{\lambda}, \bar{w})$ with $\widehat{\mathbb{M}}_{\widehat{\nu}} \xrightarrow{p} \overline{\mathbb{M}}_{\bar{\nu}}$. Then $\widehat{\Psi}^{\mathrm{NDE}} \xrightarrow{p} \psi^{\mathrm{NDE}}(P)$ and $\widehat{\Psi}^{\mathrm{NIE}} \xrightarrow{p} \psi^{\mathrm{NIE}}(P)$ if either:[5]

1. $(\overline{\mathbb{M}}_{\nu}, \bar{\nu}) = (\mathbb{M}_{\nu}, \nu)$; or
2. $(\overline{\mathbb{M}}_{\nu}, \bar{w}) = (\mathbb{M}_{\nu}, w)$; or
3. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{w}^{10}) = (\lambda, w^{11}, w^{00}, w^{10})$; or
4. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{\nu}_1) = (\lambda, w^{11}, w^{00}, \nu_1)$.

[5] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Multiple robustness of $\widehat{\Psi}^{\mathrm{NDE}}$ and $\widehat{\Psi}^{\mathrm{NIE}}$

Theorem 2. Assume that $(\widehat{\mathbb{M}}_\nu, \widehat{\nu}, \widehat{\lambda}, \widehat{w}) \xrightarrow{p} (\overline{\mathbb{M}}_\nu, \bar{\nu}, \bar{\lambda}, \bar{w})$ with $\widehat{\mathbb{M}}_{\widehat{\nu}} \xrightarrow{p} \overline{\mathbb{M}}_{\bar{\nu}}$. Then $\widehat{\Psi}^{\mathrm{NDE}} \xrightarrow{p} \psi^{\mathrm{NDE}}(P)$ and $\widehat{\Psi}^{\mathrm{NIE}} \xrightarrow{p} \psi^{\mathrm{NIE}}(P)$ if either:[5]

1. $(\overline{\mathbb{M}}_\nu, \bar{\nu}) = (\mathbb{M}_\nu, \nu)$; or
2. $(\overline{\mathbb{M}}_\nu, \bar{w}) = (\mathbb{M}_\nu, w)$; or
3. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{w}^{10}) = (\lambda, w^{11}, w^{00}, w^{10})$; or
4. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{\nu}_1) = (\lambda, w^{11}, w^{00}, \nu_1)$.

[5] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Multiple robustness of $\widehat{\Psi}^{\mathrm{NDE}}$ and $\widehat{\Psi}^{\mathrm{NIE}}$

Theorem 2. Assume that $(\widehat{\mathbb{M}_\nu}, \widehat{\nu}, \widehat{\lambda}, \widehat{w}) \xrightarrow{p} (\overline{\mathbb{M}}_\nu, \bar{\nu}, \bar{\lambda}, \bar{w})$ with $\widehat{\mathbb{M}}_{\widehat{\nu}} \xrightarrow{p} \overline{\mathbb{M}}_{\bar{\nu}}$. Then $\widehat{\Psi}^{\mathrm{NDE}} \xrightarrow{p} \psi^{\mathrm{NDE}}(P)$ and $\widehat{\Psi}^{\mathrm{NIE}} \xrightarrow{p} \psi^{\mathrm{NIE}}(P)$ if either:[5]

1. $(\overline{\mathbb{M}}_\nu, \bar{\nu}) = (\mathbb{M}_\nu, \nu)$; or
2. $(\overline{\mathbb{M}}_\nu, \bar{w}) = (\mathbb{M}_\nu, w)$; or
3. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{w}^{10}) = (\lambda, w^{11}, w^{00}, w^{10})$; or
4. $(\bar{\lambda}, \bar{w}^{11}, \bar{w}^{00}, \bar{\nu}_1) = (\lambda, w^{11}, w^{00}, \nu_1)$.

Symmetry between ν and w: both are estimated by solving integral equations; consistency only requires one working model to be correct.

[5] $\lambda = p(T, Z, X \mid S = 0)/p(T, Z, X \mid S = 1)$; $\mu_t = E[Y \mid T = t, Z, X, S = 0]$; $\mathbb{M}_{\nu_t}(t', Z, X) = E[\nu_t(M, X) \mid T = t', Z, X, S = 1]$; ν_t solves $\mathbb{M}_{\nu_t}(t, Z, X) = \mu_t$.

Asymptotic behaviors of $\widehat{\Psi}^{\mathrm{NDE}}$ and $\widehat{\Psi}^{\mathrm{NIE}}$

Theorem 3. Under Assumptions 1-5, Condition 1, appropriate regularity conditions[6] , and

1. $\max\{\|\widehat{\nu}_0 - \nu_0\|, \|\widehat{\nu}_1 - \nu_1\|\} \cdot \max\{\|\widehat{w}^{11} - w^{11}\|, \|\widehat{w}^{10} - w^{10}\|, \|\widehat{w}^{00} - w^{00}\|\} = o_P(n^{-1/2})$; and
2. $\big\{ \max\{\|\widehat{w}^{11} - w^{11}\|, \|\widehat{w}^{10} - w^{10}\|, \|\widehat{w}^{00} - w^{00}\|\} + \|\widehat{\lambda} - \lambda\|\big\} \cdot \max\{\|\widehat{\mathbb{M}}_{\nu_0} - \mathbb{M}_{\nu_0}\|, \|\widehat{\mathbb{M}}_{\nu_1} - \mathbb{M}_{\nu_1}\|\} = o_P(n^{-1/2})$.

Then the proposed estimators are asymptotically normal with

$$\sqrt{n}(\widehat{\Psi}^{\mathrm{NDE}} - \psi^{\mathrm{NDE}}(P)) \to N(0, \mathrm{var}(\mathrm{D}_{\mathrm{P}}^{\mathrm{NDE}}))$$
$$\sqrt{n}(\widehat{\Psi}^{\mathrm{NIE}} - \psi^{\mathrm{NIE}}(P)) \to N(0, \mathrm{var}(\mathrm{D}_{\mathrm{P}}^{\mathrm{NIE}})).$$

[6] We let $\|f\| = \{\int f(o)^2 dP(o)\}^{1/2}$ denote the $L_2^0(P)$-norm. $\widehat{\mathbb{M}}_\nu, \widehat{\nu}, \widehat{\lambda}, \widehat{w}$ were estimated via cross-fitting, $\|D_{\hat{P}}^{\mathrm{NDE}} - D_P^{\mathrm{NDE}}\| = o_P(1)$ and $\|D_{\hat{P}}^{\mathrm{NIE}} - D_P^{\mathrm{NIE}}\| = o_P(1)$

Simulation setup

We investigate two instrument structures:

- a single instrument $Z \in \mathbb{R}$ where Z and M are mutually complete
- a two-instrument $Z \in \mathbb{R}^2$ where only M is complete in Z.

Common setup:

$$S \sim \mathrm{Bernoulli}(0.4), \qquad T \mid S = s \sim \mathrm{Bernoulli}\{0.4(1-s) + 0.5s\}.$$

Data-generating mechanisms

Single-IV setting:

$$Z \mid T = t, S = s \sim N\{\mu(t, s), 1\},$$

$$M = \alpha_{0t} + \alpha_{Zt} Z + U + \varepsilon_M, \qquad Y = \beta_{0t} + \beta_{Mt} M + U + \varepsilon_Y.$$

Two-IV setting:

$$Z_j \mid T = t, S = s \sim N\{\mu(t, s), 1\}, \quad j = 1, 2,$$

$$M = \alpha_{0t} + \alpha_{Zt,1} Z_1 + \alpha_{Zt,2} Z_2 + U + \varepsilon_M,$$

with the same outcome model as above.

Here $U, \varepsilon_M, \varepsilon_Y$ are standard normal, with independent errors.

Estimators compared

For each setting, we compare:

- `Correct`: all nuisance components correctly specified.
- `Mis 1`–`Mis 4`: misspecification patterns that still satisfy the multiple robustness conditions.

For the two-IV setting:

- Use an aggregated instrument

$$\widetilde{Z}_a = aZ_1 + (1-a)Z_2, \qquad a \in [0,1],$$

  with a selected by minimizing an estimated variance criterion.
- Compare against a `Single IV` estimator using only Z_2.

Monte Carlo setup: $n = 1000$ per replicate, 1000 replicates.

Simulation result

Table: Bias, SE and the coverage for different estimators.

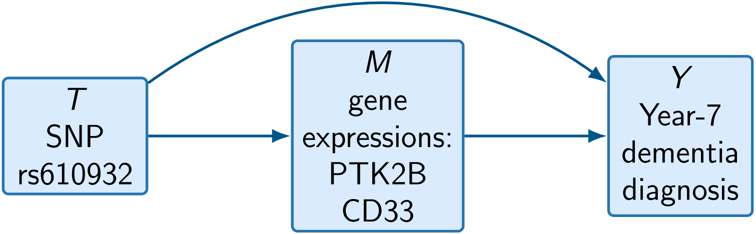
Setting	Estimator	Bias	SE	Coverage
Mutual Complete	`Correct`	0.00	0.45	93.6%
	`Mis 1`	0.0	0.60	
	`Mis 2`	-0.03	0.77	
	`Mis 3`	0.00	0.63	
	`Mis 4`	-0.03	1.03	
M complete in Z	`Correct`	0.00	0.16	94.7%
	`Mis 1`	-0.01	0.20	
	`Mis 2`	0.00	0.17	
	`Mis 3`	0.00	0.23	
	`Mis 4`	0.00	0.26	
	`Single IV`	0.00	0.59	

Application: Alzheimer's disease genetics

- We use the observational database from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI is a longitudinal, multicenter observational study launched in 2004 and continued through successive phases, including ADNI-1, ADNI-GO, ADNI-2, ADNI-3, and ADNI-4.
- Importantly, gene-expression data were not available in ADNI-1. Transcriptomic data became available only in later phases, primarily for selected ADNI-GO/2 participants and onwards.
- As a result, dementia outcomes, SNPs, and gene-expression measurements were never jointly observed for the ADNI-1 cohort (our target).

Application: Alzheimer's disease genetics

Goal: assess whether genetic effects on dementia risk operate through gene expression.



Estimand of interest

For each SNP–gene pair, estimate:

$$\mathrm{NIE} = \mathbb{E}\{Y(1, M(1)) \mid S = \text{ADNI-1}\} - \mathbb{E}\{Y(1, M(0)) \mid S = \text{ADNI-1}\}.$$

We considered:

- SNP effect mediated through CD33 expression
- SNP effect mediated through PTK2B expression

In addition,

- T = SNP rs610932, coded as indicator of ≥ 1 minor allele.
- X include age, sex, years of education, APOE genotype, diagnosis of mild cognitive impairment at enrollment, and mini-mental state examination score.

Selection of Z

- For each mediator of interest, we first selected the top 50 most correlated SNPs based on eQTL summaries (GTEx Consortium, 2020).
- We then identified SNPs overlapping with those available in the ADNI study, yielding $\{rs3735759, rs1879189, rs9644121\}$ for PTK2B and $\{rs3826656, rs273638, rs273634\}$ for CD33.
- We then applied the proposed IV selection algorithm to construct the instrument set for each mediator. The algorithm recovered the same SNP sets.

Application results

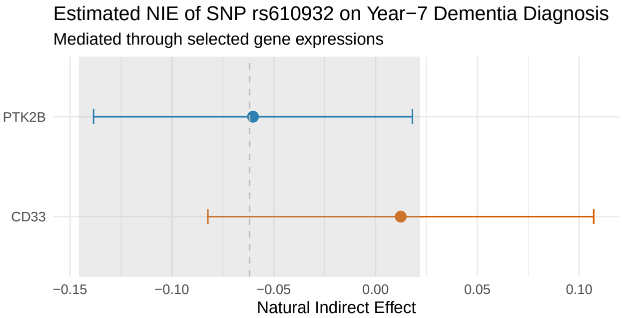


Figure: Estimated natural indirect effects mediated by selected gene expressions. The estimated average total effect is -0.06 with 95% CI $[-0.15, 0.02]$ indicated by the gray region.

I probably don't have the time for this slide

We also derived the canonical gradients of the mediation effects relative to model where

- $U = \emptyset$; and
- Only M is complete in Z

It turns out the model is semiparametric – the conditional independence involving the IVs and latent alignment in (Y, M) across data sources restrict the tangent space.

I probably don't have the time for this slide

We also derived the canonical gradients of the mediation effects relative to model where

- $U = \emptyset$; and
- Only M is complete in Z

It turns out the model is semiparametric – the conditional independence involving the IVs and latent alignment in (Y, M) across data sources restrict the tangent space.

Compared with Theorem 1 where mediator-only sample $(S = 1)$ does not improve efficiency for estimating $\psi^{11}(P)$ and $\psi^{00}(P)$, we show that such efficiency gains can arise when Z is not complete in M.

References

Li, S., and Wang, R. (2026). Semiparametric Mediation Analysis with Separately Observed Mediator and Outcome under Unmeasured Confounding. arXiv preprint arXiv:2606.17232

References I

Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87(418):328–336.

Athey, S., Chetty, R., Imbens, G. W., and Kang, H. (2025). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. *Review of Economic Studies*, page rdaf087.

Bennett, A., Kallus, N., Mao, X., Newey, W., Syrgkanis, V., and Uehara, M. (2023). Source condition double robust inference on functionals of inverse problems. *arXiv preprint arXiv:2307.13793*.

Bennett, A., Kallus, N., Mao, X., Newey, W. K., Syrgkanis, V., and Uehara, M. (2025). Inference on strongly identified functionals of weakly identified functions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf075.

References II

Chen, S., Zhang, P., and Cui, Y. (2025). Identification and debiased learning of causal effects with general instrumental variables. *arXiv preprint arXiv:2510.20404*.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565.

Derkach, A., Sampson, J. N., and Pfeiffer, R. M. (2024). Integrating incomplete data for mediation analysis. *Statistica Sinica*, 34(2).

Dong, M., Liu, L., Tang, D., Liu, G., Xu, W., and Wang, L. (2025). Marginal causal effect estimation with continuous instrumental variables. *arXiv preprint arXiv:2510.14368*.

Evans, K., Sun, B., Robins, J., and Tchetgen, E. J. T. (2021). Doubly robust regression analysis for data fusion. *Statistica Sinica*, 31(3):1285–1307.

References III

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1645–1666.

Graham, E., Carone, M., and Rotnitzky, A. (2024). Towards a unified theory for semiparametric data fusion with individual-level data. *arXiv preprint arXiv:2409.09973*.

GTEx Consortium (2020). The gtex consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330.

Guo, H., Ogburn, E. L., and Shpitser, I. (2025). Comparing two proxy methods for causal identification. *arXiv preprint arXiv:2512.00175*.

References IV

Huang, Y.-t., Yu, J.-C., and Lin, J.-H. (2023). Surrogate marker assessment using mediation and instrumental variable analyses in a case-cohort design. *The Annals of Applied Statistics*, 17(2):1239–1259.

Imai, K., Keele, L., and Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Imai, K., Keele, L., and Yamamoto, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects.

Imai, K., Tingley, D., and Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1):5–51.

Inoue, A. and Solon, G. (2010). Two-sample instrumental variables estimators. *The Review of Economics and Statistics*, 92(3):557–561.

References V

Kallus, N. and Mao, X. (2025). On the role of surrogates in the efficient estimation of treatment effects with limited outcome data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 87(2):480–509.

Li, S. and Luedtke, A. (2023). Efficient estimation under data fusion. *Biometrika*, 110(4):1041–1054.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Miao, W., Hu, W., Ogburn, E. L., and Zhou, X.-H. (2023). Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association*, 118(543):1953–1967.

References VI

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.

Rudolph, K. E., Sofrygin, O., and van der Laan, M. J. (2021). Complier stochastic direct effects: Identification and robust estimation. *Journal of the American Statistical Association*, 116(535):1254–1264.

Rudolph, K. E., Williams, N., and Díaz, I. (2024). Using instrumental variables to address unmeasured confounding in causal mediation analysis. *Biometrics*, 80(1):ujad037.

References VII

Sun, B. and Miao, W. (2022). On semiparametric instrumental variable estimation of average treatment effects through data fusion. *Statistica Sinica*, 32:569–590.

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., and Miao, W. (2024). An introduction to proximal causal inference. *Statistical Science*, 39(3):375–390.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):531–550.

References VIII

Wang, W., Scharfstein, D., Tan, Z., and MacKenzie, E. J. (2009). Causal inference in outcome-dependent two-phase sampling designs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(5):947–969.

Xu, Q., Testa, L., Lei, J., and Roeder, K. (2025a). Blockwise missingness meets ai: A tractable solution for semiparametric inference. *arXiv preprint arXiv:2509.24158*.

Xu, Z., Witten, D., and Shojaie, A. (2025b). A unified framework for semiparametrically efficient semi-supervised learning. *arXiv preprint arXiv:2502.17741*.

Zhao, Q., Wang, J., Spiller, W., Bowden, J., and Small, D. S. (2019). Two-sample instrumental variable analyses using heterogeneous samples. *Statistical Science*, 34(2):317–333.

Extension: finding canonical gradients when $U = \emptyset$

In practice, practitioners may measure a set of candidate IVs for the purpose of bridging. In that case, the dimension of Z may be larger than M, making the model $\mathcal{P}$, under which only M is complete in Z, the more realistic formulation.

While deriving the canonical gradient relative to $\mathcal{P}$ would be ideal, we have not been able to find the form of this gradient.

Extension: finding canonical gradients when $U = \emptyset$

In practice, practitioners may measure a set of candidate IVs for the purpose of bridging. In that case, the dimension of Z may be larger than M, making the model $\mathcal{P}$, under which only M is complete in Z, the more realistic formulation.

While deriving the canonical gradient relative to $\mathcal{P}$ would be ideal, we have not been able to find the form of this gradient.

In what follows, we consider the special case when $U = \emptyset$, and denote the corresponding model as $\mathcal{P}^{\emptyset}$.

Characterizing the tangent space of $\mathcal{P}^{\emptyset}$

Lemma 4. Under no unmeasured confounding, i.e. $U = \emptyset$, the tangent space of $\mathcal{P}^{\emptyset}$ at $P \in \mathcal{P}^{\emptyset}$ takes the following form:

$$\begin{aligned}\mathcal{T}(P, \mathcal{P}^{\emptyset}) = \Big\{ o \mapsto (1-s)\, &\{E_P[h_Y(Y, M, T, G) \mid y, t, z, g] + E_P[h_M(M, T, Z, G) \mid y, t, z, g]\} \\ &+ s \cdot h_M(m, t, z, g) + h_T(t, z, g, s) + h_Z(z, g, s) + h_G(g, s) + h_S(s) \\ &: h_Y \in L_2^0(P_{Y \mid M, T, G}), h_M \in L_2^0(P_{M \mid T, Z, G}), \\ &\quad h_T \in L_2^0(P_{T \mid Z, G, S}), h_Z \in L_2^0(P_{Z \mid G, S}), h_G \in L_2^0(P_{G \mid S}), h_S \in L_2^0(P_S) \Big\},\end{aligned}$$

Characterizing the tangent space of $\mathcal{P}^{\emptyset}$

Lemma 4. Under no unmeasured confounding, i.e. $U = \emptyset$, the tangent space of $\mathcal{P}^{\emptyset}$ at $P \in \mathcal{P}^{\emptyset}$ takes the following form:

$$
\begin{aligned}
\mathcal{T}(P, \mathcal{P}^{\emptyset}) = \Big\{ o \mapsto (1-s) & \{E_P[h_Y(Y, M, T, G) \mid y, t, z, g] + E_P[h_M(M, T, Z, G) \mid y, t, z, g]\} \\
& + s \cdot h_M(m, t, z, g) + h_T(t, z, g, s) + h_Z(z, g, s) + h_G(g, s) + h_S(s) \\
& : h_Y \in L_2^0(P_{Y|M,T,G}), h_M \in L_2^0(P_{M|T,Z,G}), \\
& \quad h_T \in L_2^0(P_{T|Z,G,S}), h_Z \in L_2^0(P_{Z|G,S}), h_G \in L_2^0(P_{G|S}), h_S \in L_2^0(P_S) \Big\},
\end{aligned}
$$

In the absence of unmeasured confounding and when only M is complete in Z, we now see how the conditional independence involving the IVs and latent alignment in (Y, M) across data sources restrict the tangent space.

Characterizing the tangent space of $\mathcal{P}^{\emptyset}$

Lemma 4. Under no unmeasured confounding, i.e. $U = \emptyset$, the tangent space of $\mathcal{P}^{\emptyset}$ at $P \in \mathcal{P}^{\emptyset}$ takes the following form:

$$
\begin{aligned}
\mathcal{T}(P, \mathcal{P}^{\emptyset}) = \Big\{ o \mapsto (1-s) & \left\{ E_P[h_Y(Y, M, T, G) \mid y, t, z, g] + E_P[h_M(M, T, Z, G) \mid y, t, z, g] \right\} \\
& + s \cdot h_M(m, t, z, g) + h_T(t, z, g, s) + h_Z(z, g, s) + h_G(g, s) + h_S(s) \\
& : h_Y \in L_2^0(P_{Y|M,T,G}), h_M \in L_2^0(P_{M|T,Z,G}), \\
& \quad h_T \in L_2^0(P_{T|Z,G,S}), h_Z \in L_2^0(P_{Z|G,S}), h_G \in L_2^0(P_{G|S}), h_S \in L_2^0(P_S) \Big\},
\end{aligned}
$$

We derive the canonical gradients of ψ^{NDE} and ψ^{NIE} relative to $\mathcal{P}^{\emptyset}$.