Basic Concepts of Causal Mediation Analysis and Some Extensions

Vanessa Didelez
School of Mathematics
University of Bristol

Joint work with: Philip Dawid, Sara Geneletti, Svend Kreiner

Symposium: Causal Mediation Analysis
Ghent, January 2013
Overview

• Basic concepts of causal inference

• Basic concepts of causal mediation analysis

• Manipulable parameters and augmented systems

• Post-treatment confounding

• Estimation using augmentation

• A typical sociological study

• Conclusions
Basic Concepts of Causal Inference
Some Notation

Potential Outcomes (Counterfactuals): Rubin (1970s)

\[ Y(x) = \text{outcome if } X \text{ were set to } x. \]

\textbf{do(·)-Calculus:} Spirtes / Pearl (1990s)

\[ p(y|\text{do}(X = x)) \text{ intervention distribution.} \]

\textbf{Often:} \[ p(Y(x)) = p(y|\text{do}(X = x)), \]

but can express different assumptions/targets with different notation.

\[ \rightarrow \text{do(·)-models “⊂” potential outcomes models.} \]

\textbf{Confounding:} is present if \[ p(y|\text{do}(X = x)) \neq p(y|X = x). \]
Directed Acyclic Graphs (DAGs)

Nodes / vertices = variables $X_1, \ldots, X_K$
no edge $\Rightarrow$ some conditional independence such that

$$X_i \perp \perp X_{\text{nd}(i) \setminus \text{pa}(i)} | X_{\text{pa}(i)}$$

$\text{nd}(i) =$ ‘non-descendants of $i$', $\text{pa}(i) =$ ‘parents of $i$’.

**Example:** $X \perp \perp (Y, W)$ or $W \perp \perp (X, Z) | Y$ etc.

Equivalent: factorisation

$$p(x) = \prod_{i=1}^{K} p(x_i | x_{\text{pa}(i)})$$

**Example:**

$$p(x, y, z, w, u) = p(x)p(y)p(z | x, y)p(w | y)p(u | z, w)$$
(Locally) Causal DAGs

Example: DAG is causal wrt. $Z$ if

$$p(x, y, w, u|\text{do}(Z = \tilde{z})) = p(x)p(y)I(z = \tilde{z})p(w|y)p(u|z, w)$$

Can then show that e.g.

$$p(u|\text{do}(Z = \tilde{z})) = \sum_w p(u|\tilde{z}, w)p(w)$$

$\Rightarrow$ intervention distribution is identified.

Here, $W$ is sufficient to adjust for confounding.

Identification: can express (aspects of) the intervention distribution in terms of observable quantities.

Nonparametric Structural Equation Models (NPSEMs): (Pearl, 2000) quasi-deterministic causal DAGs "⇔" counterfactuals
Basic Concepts of Causal Mediation Analysis
Some Examples

• Socioeconomic status $\rightarrow$ health behaviour $\rightarrow$ health.

• Alcoholism $\rightarrow$ loss of social network $\rightarrow$ homelessness.

• Ethnicity/gender $\rightarrow$ qualification $\rightarrow$ job offer.

• Age at conception $\rightarrow$ gestation period $\rightarrow$ perinatal death.

• Placebo: treatment $\rightarrow$ expectation $\rightarrow$ recovery.
What is the Target of Inference?

Research questions in context of mediation analysis often vague — something to do with “causal mechanisms”.

**Ideally:** target of inference is clear if we can
— describe experiment to measure the desired quantity explicitly
— formulate decision problem that will be informed
⇒ should guide the design, collection of data, assumptions, and analysis.

← Range from less to more hypothetical / feasible →
Total Causal Effects

Set $X$ to different values $\rightarrow$ effect on distribution of $Y$.

$E(Y(x^*))$ vs. $E(Y(x))$  
$p(y|\text{do}(X = x^*))$ vs. $p(y|\text{do}(X = x))$

In (locally causal) DAG:

Observationally $p(\text{all}) = p(y|w, m, x, c)p(m|w, x)p(x|c)p(c)p(w)$

... intervention $p(\text{all}|\text{do}(X = x^*)) = p(y|w, m, x, c)p(m|w, x)I(X = x^*)p(c)p(w)$
Total Causal Effects

Identification — Assumption of “no unobserved confounding”: let $C$ be observable (pre-treatment) covariates with potential outcomes: $Y(x) \perp\!\perp X \mid C$ (for all $x$)

graphically: all ‘back–door’ paths from $X$ to $Y$ are blocked by $C$.

Then: (standardisation)

$$p(y|\text{do}(X = x)) = \sum_c p(y|C = c, X = x)p(C = c).$$
Controlled (Direct) Effects

Set $X$ to different values while holding $M$ fixed $\rightarrow$ effect on $Y$.

$$E(Y(x^*, m^*)) \text{ vs. } E(Y(x, m^*))$$

$$p(y|\text{do}(X = x^*, M = m^*)) \text{ vs. } p(y|\text{do}(X = x, M = m^*))$$

In (locally causal) DAG:

Observationally $p(\text{all}) = p(y|w, m, x, c)p(m|w, x)p(x|c)p(c)p(w)$

... intervention $p(\text{all}|\text{do}(X = x^*, M = m^*)) =

$$p(y|w, m, x, c)I(M = m^*)I(X = x^*)p(c)p(w)$$
Controlled (Direct) Effects

Identification — Assumption

**Sequential** version of “no unobserved confounding”:

let $C$ be pre-$X$ covariates and $W$ pre-$M$ covariates,

$Y(x, m) \perp \perp X | C$ and $Y(x, m) \perp \perp M | (X = x, C, W)$

graphically: sequential version of back–door criterion  (Dawid & Didelez, 2010)

**Then:** (G–Formula)

$$p(y|do(X = x^*, M = m^*)) = \sum_{c, w} p(y|c, w, x^*, m^*)p(w|x^*, m^*)p(c)$$

**Note 1:** here, $W$ allowed to depend on $X$.

**Note 2:** no model for $M$ given $X$. 
Controlled (Direct) Effects

Pro’s:
– clear practical interpretation,
– “understandable” conditions for identifiability.

Con’s
– may depend on choice of \( m^* \),
– nothing really ‘direct’ about it, as effect is the same if \( M \) precedes \( X \),
– no corresponding concept of ‘controlled indirect’ effect,
– often “impractical” to fix \( M \) at \( m^* \).
Standardised (Direct) Effects

(Geneletti, 2007; Didelez et al., 2006)

Set $X$ to different values while $M$ is made to arise from distribution $\mathcal{D}$ ($\mathcal{D}$ may depend on pre-$(X, M)$ variables) → effect on $Y$.

$$p(y|\text{do}(X = x^*), \text{draw}_\mathcal{D}(M))$$

vs. $$p(y|\text{do}(X = x), \text{draw}_\mathcal{D}(M))$$

In (locally causal) DAG:

Observationally $p(\text{all}) = p(y|w, m, x, c)p(m|w, x)p(x|c)p(c)p(w)$

... intervention $p(\text{all}|\text{do}(X = x^*), \text{draw}_\mathcal{D}(M)) =$

$$p(y|w, m, x, c)p_\mathcal{D}(M = m)I(X = x^*)p(c)p(w)$$
Standardised (Direct) Effects

More specifically: could augment the ‘system’ (DAG, model) with the random mechanism that generates $M \rightarrow$ within this system can again condition on $M$ or integrate it out etc.

Then: $p(y|\text{do}(X = x^*), \text{draw}_D(M))$

$$= \sum_{c,m,w} p(y|w, m, x^*, c)p_D(m)p(c)p(w)$$

Identification: similar to CDE, except if $D$ needs to be estimated.
Natural (In)Direct Effects

(Robins & Greenland, 1992; Pearl, 2001)

Set $M$ to $M(x^*)$ while setting $X$ to $x$, vary $x$ or $x^*$ → effect on $Y$.

Key quantity: nested counterfactual $Y(x, M(x^*))$.

Natural Direct Effect: $p(Y(x, M(x^*)))$ vs. $p(Y(x^*, M(x^*)))$

Natural Indirect Effect: $p(Y(x, M(x)))$ vs. $p(Y(x, M(x^*)))$

⇒ Total effect = NDE “+” NIE

Note 1: “additivity” not valid for other definitions of (in)direct effects.

Note 2: swap $x, x^*$ ⇒ NDE, NIE different when interaction present.
Identification via Mediation Formula

Let’s ignore pre–$X$ variables, e.g. assume $X$ was randomised.

Natural effects are identified if $W$ exists such that

$Y(x, m) \perp \perp M(x^*) \mid W$ (for all $m$).

Implied by NPSEM with DAG as shown. Not expressible in other frameworks.

Then:

$$p(Y(x, M(x^*))) = \sum_{m, w} p(y \mid w, m, x)p(m \mid w, x^*)p(w)$$

Crucial: $W$ not affected by interventions in $X$, i.e. no “post-treatment confounding” of $M$ and $Y$. 
$M - Y$ “Confounding”

Intervention in $M$ interrupts its dependence on other preceding variables.

**Pure/natural effects:**
when “setting” $M$ at $M(x^*)$ we do not interrupt its dependence on preceding variables, especially not on $W$!

$\Rightarrow M(x^*)$ & $W$ dependent — natural effects average over their joint distribution; information lost by do($M = m$).

$\Rightarrow$ stratify by the same $W$ when assessing $X \rightarrow M$ and $M \rightarrow Y$ effect.
Natural (In)Direct vs. Standardised Effects

**Standardised effect:** not the same but comes quite close:
choose $D$ to be $p(m|W, \text{do}(X = x^*))$ ($= p(m|W, X = x^*)$) when $X$ randomised).

$$p(y|\text{do}(X = x), \text{draw}_D(M)) = \sum_{m,w} p(y|w, m, x)p(m|w, X = x^*)p(w)$$

**Interestingly:** same mediation formula for natural effects earlier.

**Hence:** under certain structures and data situations, cannot empirically distinguish between natural effects and specific standardised effects.
Natural (In)Direct Effects

Pro’s:
– offers a indirect effect notion,
– “additivity” of direct and indirect effect.

Con’s:
– not guaranteed identified by a single randomised experiment,
– assumption $Y(x, m) \perp M(x^*)|W$ (for all $m$) is ‘cross–world’,
– ...hence difficult to understand or justify,
– concepts (and assumption) are thoroughly counterfactual.
Manipulable Parameters

and augmented systems
Manipulable Parameters

(Robins, 2003; Robins and Richardson, 2011)

“Any contrast between treatment regimes which could be implemented in an experiment with sequential treatment assignments, wherein the treatment given at any stage can be a function of past covariates.”

⇒ represented by (functions of) G–formula wrt. a DAG.

⇒ Natural effects are not ‘manipulable’ without extending the story.
Assume we can separate different aspects of $X$ that can be set to different values for separate pathways; other conditional distributions remain the same.

Observable system:

\[
p(y, m|x) = p(y|m, x)p(m|x)
\]

Hypothetical (augmented) system:

\[
p^{\text{aug}}(y, m|x, x^*) = p(y|m, x)p(m|x^*)
\]

Direct: $Y \rightarrow X$–association

Indirect: $Y \rightarrow X^*$–association

$\rightarrow$ manipulable wrt augm. system.
Placebo–type design

It may sometimes be actually possible to separate different aspects of treatment $X$ by design so that each pathway (direct / indirect) is affected by only one aspect. (Didelez, 2012)

In fact, this is what a double-blind placebo controlled study does.
Double–Blind Placebo Controlled Studies

\(X\) = treatment
\(M\) = patient’s / doctor’s expectation
\(W\) = disease history
\(Y\) = health outcome

Separate treatment into:
\(A\) = amount of active ingredient,
\(B\) = form of treatment (size/shape/colour/number of pills).

⇒ essentially the augmentation but as actual experiment.
Interpretation

In placebo controlled trial: no need to worry about identifiability, as we can observe the augmented system itself. (Also, no need to collect data on $W$.)

But: may want to think whether desired interpretation is achieved.

E.g.: do placebo patients truly believe they are being treated? (For ethical reasons need to tell people that they may be getting placebo.)
Mediation Formula — Again!

In augmented system

\[ p_{\text{aug}}(y|x, x^*) = \]

\[ = \sum_{m,w} p(y|w, m, x)p(m|w, x^*)p(w). \]

⇒ same formula as before!

⇒ New motivation for mediation formula.
Post Treatment $M - Y$ Confounding
Post–treatment $M \rightarrow Y$ Confounding

Mediation formula does not identify the natural effects.

$W$ has “conflict of interest”:

Nested counterfactual: $Y(x, M(x^*)) = Y(x, M(x^*, W(x^*)), W(x))$.

Difficult to get data that informs us jointly about $W(x^*), W(x)$.

(see Avin et al. (2005), “Recanting Witness” criterion.)

Usually, $W$ is assumed away... but often realistic, especially when we admit that things happen continually in continuous time.

Problem should be explored by clarifying what kind of experiment/decision problem we want to address.
Post–treatment $M \rightarrow Y$ Confounding

Placebo Study:

$W = \text{side effect}$

Plausible augmented DAG
⇒ illustrates why this is considered as “unblinding”
Corresponds to
$Y(x, M(x^*, W(x)), W(x))$
Post–treatment $M - Y$ Confounding

Placebo Study:

Could modify placebo to cause side effect?

$\Rightarrow$ yields natural direct effect of active ingredient not mediated through either expectation or side effect.

Corresponds to $Y(x, M(x^*, W(x^*)), W(x^*))$.

$\Rightarrow$ not the same as $Y(x, M(x^*))$ but sensible quantity.
Estimation Using Augmentation
Estimation Methods

Observational data, assume no post-treatment confounding of $M-Y$.

**In principle**, (baseline covariates omitted):

— estimate model for $p(y|x, m, w)$
— estimate model for $p(m|x, w)$

$\rightarrow$ plug into mediation formula

$\Rightarrow$ potential for misspecification unless saturated/nonparametric models can be fitted, may need MC integration etc.

$\Rightarrow$ various double/triple robust suggestions.

**But**: saturated models can sometimes be used!

And, (if not) can subject the above to model checking etc.

(Note: Robins & Richardson (2011) derive bounds under weaker assumptions.)
Fitting Augmented DAGs with Auxiliary Variables

Two methods:

1) Kreiner (2002, unpibl.) fits a DAG, where node $X$ (and corresponding data) is duplicated to obtain direct/indirect effects.

2) Lange et al. (2012) fit marginal natural effect models using clever weights, also based on duplicating $X$-data and individuals — can also be viewed as imputation.

Note: both methods equivalent for fully saturated models.
Fitting Augmented DAGs with Auxiliary Variables

Kreiner (2002) Method:

• sequence of loglinear models to fit conditional distributions;

• duplicate $X$ by $X^*$ (same data);

• graphical modelling software to obtain desired (possibly standardised) marginals;

• can equivalently be carried out with probability propagation software for DAG expert systems (e.g. gRain).

Note: under identifying assumptions $X$ and $X^*$ never occur together in conditioning set, so no problem with ‘duplicate’ data.
Fitting Augmented DAGs with Auxiliary Variables

Lange et al. (2012) Method

- A marginal natural effect model parameterises
  \[ E(Y(x, M(x^*))) = g(x, x^*; \beta) \]

- **augment** data for \( X \) so that \( X^* = 1 - X \) (binary case)

- fit model to the new data set, with weights for individual \( i \)
  \[ \frac{p(M = m_i | X = x^*_i, w_i)}{p(M = m_i | X = x_i, w_i)} \]

→ can be done with standard software if weights can be specified.

**Note:** models \( g(x, x^*; \beta) \) and \( p(m | x, w) \) may not be compatible.
Fitting Augmented DAGs with Auxiliary Variables

Observational system
\[ p(y, m, w | X = x) = p(y|m, X = x, w)p(m|X = x, w)p(w) \]

Hypothetical system
\[ p^{\text{aug}}(y, m, w | x^*, x) = p(y|m, X = x, w)p(m|X = x^*, w)p(w) \]

Where \( p^{\text{aug}}(y|x, x^*) = \sum_{m, w} p^{\text{aug}}(y, m, w | x, x^*) \)
\[
= \sum_{m, w} p(y, m, w | X = x) \frac{p(m|X = x^*, w)}{p(m|X = x, w)}
\]

⇒ motivate the weighting approach of Lange et al. (2012)
A Typical Sociological Study
Example: Childhood Environment and Adult Anxiety

Representative Survey of Living Conditions in Denmark

Subset of variables, \( N = 4561 \):

**Fear** of violence (yes/no); overall 18.7%

**Exposed** to violence or threats (yes/no); overall 3.6%

**Adult** environment (3 levels of urbanisation)

Socioeconomic status, **SES**, (5 levels)

**Childhood** environment (3 levels of urbanisation)

Baseline variables: **Age** and **Sex**.

**Primary analysis** (logistic regression): main predictors of fear are exposure to violence, sex, and *childhood environment*
Example: Childhood Environment and Adult Anxiety

More Detailed Analysis based on Graphical Modelling

Combination of subject matter background knowledge and statistical model selection yields this directed acyclic graph (DAG):

(Kreiner, 2002)

For now, will regard above graph as reasonable starting point.

Various questions relating to Mediation could be of interest here.
Example — Assumptions Plausible?

Survey of Living Conditions in Denmark

Potential problems: unobserved confounding, e.g. parents’ SES; also post-treatment confounding likely (childhood exposure to violence?).

⇒ take following analyses with a pinch of salt.
Motivating Example — Target of Inference

Assume we can separate, say, emotional from factual consequences of childhood environment (very hypothetical).

Note: for identification observing either “Exposed to violence” or “Adult environment” is sufficient w.r.t. above DAG.
Results: Direct Effect

Preliminary and incomplete analysis

**Total effect** (adjusting for age & sex):
\[
\hat{p}(F = 1 | do(X = \text{urban})) = 0.293 \\
\hat{p}(F = 1 | do(X = \text{suburb})) = 0.151 \\
\hat{p}(F = 1 | do(X = \text{rural})) = 0.083
\]

\(\gamma\)-coefficient: 0.414

**Standardised direct effect**: average \(X^*\) over marginal
\[
\hat{p}^{\text{aug}}(F = 1 | X = \text{urban}) = 0.280 \\
\hat{p}^{\text{aug}}(F = 1 | X = \text{suburb}) = 0.153 \\
\hat{p}^{\text{aug}}(F = 1 | X = \text{rural}) = 0.083
\]

\(\gamma\)-coefficient: 0.39
Results: Indirect Effect

Preliminary and incomplete analysis

**Total effect** (adjusting for age & sex):
\[
\hat{p}(F = 1|do(X = \text{urban})) = 0.293 \\
\hat{p}(F = 1|do(X = \text{suburb})) = 0.151 \\
\hat{p}(F = 1|do(X = \text{rural})) = 0.083 \\
\gamma-\text{coefficient: 0.414}
\]

**Standardised indirect effect:** average \(X\) over marginal
\[
\hat{p}^{\text{aug}}(F = 1|X^* = \text{urban}) = 0.18 \\
\hat{p}^{\text{aug}}(F = 1|X^* = \text{suburb}) = 0.17 \\
\hat{p}^{\text{aug}}(F = 1|X^* = \text{rural}) = 0.168 \\
\gamma-\text{coefficient: 0.027}
\]
Results: Indirect Effect of Adult Environment

Standardised indirect effect of adult environment:

\[ \hat{p}^{\text{aug}}(F = 1|X_{\text{adult}} = \text{urban}) = 0.183 \]
\[ \hat{p}^{\text{aug}}(F = 1|X_{\text{adult}} = \text{suburb}) = 0.173 \]
\[ \hat{p}^{\text{aug}}(F = 1|X_{\text{adult}} = \text{rural}) = 0.17 \]

\( \gamma \)-coefficient: 0.031
Conclusions

• Focus on manipulable parameters makes you think harder about the meaning of target of inference.

• Augmented DAGs can help to bring conceptual clarity e.g. to mediation analyses;

• ... should also be helpful when dealing with multiple mediators or for more general hypothetical scenarios.

• ... leads to straightforward methods of estimating (in)direct effects.

• More efficient and robust methods for mediation analysis are available, but incredibly more complicated and not easy to implement.

• Omitted: principal stratum direct effects — not manipulable; see discussion in IJB 2011/12. (e.g. Joffe, 2011).
References


