Statistical Learning in Mediation Analysis

**Chapter 3:
Natural direct and indirect effects**

David Benkeser
Emory University

Iván Díaz
Weill Cornell Medicine

Marco Carone
University of Washington

MODULE 13

**Summer Institute in Statistics for
Clinical and Epidemiological Research**

July 2021

Contents of this chapter

Mediation can be quantified parsimoniously via **natural effects**, in which the intervened mediator value is stochastic (defined relative to the same individual) rather than set deterministically.

We first note that the average treatment effect of A on Y in a given population can be expressed as

$$ATE \;=\; E[Y(1)] - E[Y(0)] \;=\; E[Y(1, M(1))] - E[Y(0, M(0))] \;,$$

where $Y(a)$ is the counterfactual outcome under an intervention that sets $A = a$ (but does not directly set M).

Here, we are relying on a consistency condition that ensures that $Y(a) = Y(a, M(a))$:

- if we set $A = a$ and $M = M(a)$, we see $Y(a, M(a))$;
- if we set $A = a$ and do not intervene on M, we instead see $Y(a)$.

We can decompose the total effect as

$$ATE \;=\; \underbrace{E[Y(1, M(1))] - E[Y(1, M(0))]}_{\text{natural indirect effect}} + \underbrace{E[Y(1, M(0))] - E[Y(0, M(0))]}_{\text{natural direct effect}} \;.$$

The **natural indirect effect (NIE)** of A on Y through M compares interventions:

1. set $A = 1$ and set M to its natural value under $A = 1$;
2. set $A = 1$ and set M to its natural value under $A = 0$.

Example: risk of influenza infection under different scenarios

1. administer the flu vaccine;
2. administer the flu vaccine but set neutralizing antibody (NAb) levels to the natural level the patient would have had without the vaccine.

This quantifies the effect of the vaccine mediated through NAb titers.

We can decompose the total effect as

$$ATE = \underbrace{E[Y(1, M(1))] - E[Y(1, M(0))]}_{\text{natural indirect effect}} + \underbrace{E[Y(1, M(0))] - E[Y(0, M(0))]}_{\text{natural direct effect}} .$$

The **natural direct effect (NDE)** of A on Y relative to M compares interventions:

1. set $A = 1$ and set M to its natural value under $A = 0$;
2. set $A = 0$ and set M to its natural value under $A = 0$.

Example: risk of influenza infection under different scenarios

1. administer the flu vaccine but set neutralizing antibody (NAb) levels to the natural level the patient would have had without the vaccine;
2. administer the placebo vaccine.

This quantifies the effect of the vaccine not mediated through NAb titers.

We can alternatively decompose the total effect as

$$ATE = \underbrace{E[Y(1, M(1))] - E[Y(0, M(1))]}_{\text{natural direct effect}} + \underbrace{E[Y(0, M(1))] - E[Y(0, M(0))]}_{\text{natural indirect effect}}.$$

Alternative definitions of NIE

contrast mean outcome under differing mediator values while administering treatment (total) or control (pure)?

Alternative definitions of NDE

contrast mean outcome under different exposure levels while enforcing natural mediator values arising under treatment (total) or control (pure)?

Sometimes, only one version of the NDE or NIE can plausibly be identified (if any).

Report (transparently and explicitly) on the version(s) of scientific interest.

Identification of the NDE and NIE typically requires stronger causal conditions.

Sufficient rich covariate information must be collected to allow deconfounding of key relationships — this leads to a collection of **randomization conditions**.

To identify $\psi(a, a^*) := E[Y(a, M(a^*))]$, we require:

1. $M(a^*) \perp A \mid W$;

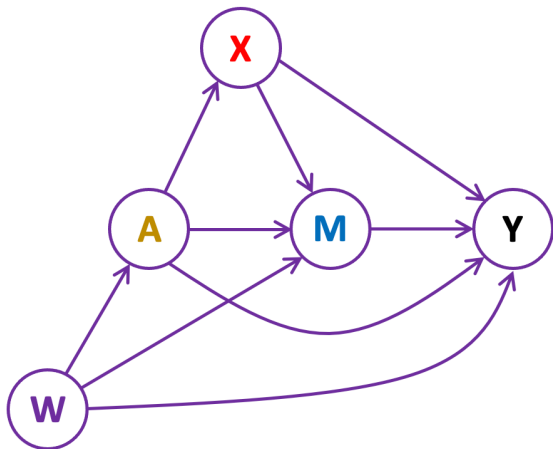for each value m that $M(a^*)$ can possibly take:

2. $Y(a, m) \perp A \mid W$;
3. $Y(a, m) \perp M \mid A = a, W$;
4. $Y(a, m) \perp M(a^*) \mid W$.

(As a consequence, there cannot exist a confounder of the $M - Y$ relationship in the causal pathway from A to Y. However, this can be relaxed — more on this later.)

Randomization conditions 1–3 should be familiar.

What is the interpretation of randomization condition 4?

Within subpopulations of patients with common W value, a patient's mediator value under $A = a^*$ gives no info about their outcome under $A = a$ and $M = m$.

This is an example of a **cross-world condition** since it involves counterfactual variables tied to incompatible interventions (i.e., defined in different hypothetical worlds).

Critical point: There is no plausible experimental design that can circumvent it.

Some question how usefulness natural mediation effects are in view of this condition.
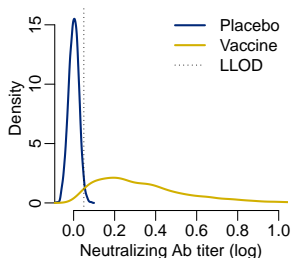(Naimi et al., 2014; Andrews & Didelez, 2020)

Certain **positivity conditions** are also required.

1. (exposure positivity) $P(A = a \mid W = w) > 0$ for every possible w;
2. (mediator overlap) for every possible w and each value m,

$$P(M = m \mid A = a^*, W = w) > 0 \ \Rightarrow \ P(M = m \mid A = a, W = w) > 0$$

Example: neutralizing antibody response to a vaccine

Under the randomization and positivity conditions stated so far, the following **G-computation** formula holds:

$$E\left[Y(a, M(a^*)) \mid W = w\right] = E\left[E\left(Y \mid M, A = a, W = w\right) \mid A = a^*, W = w\right]$$
$$= \sum_m E\left(Y \mid M = m, A = a, W = w\right) P\left(M = m \mid A = a^*, W = w\right) .$$

Within the subpopulation of patients with $W = w$, we must average over the distribution of M among patients following $A = a^*$ the mean outcome among patients following $A = a$ and with mediator value M.

This readily implies an identification for $\psi(a, a^*)$ as

$$\psi(a, a^*) = E\left[Y(a, M(a^*))\right] = E\left[E\left[Y(a, M(a^*)) \mid W\right]\right]$$
$$= E\left[E\left[E\left(Y \mid M, A = a, W\right) \mid A = a^*, W\right]\right] .$$

The G-computation identification formula

How is this result formally established?

$E[Y(a, M(a^*)) \mid W = w]$

$$= \sum_m E[Y(a, m) \mid M(a^*) = m, W = w] P(M(a^*) = m \mid W = w) \tag{1}$$

$$= \sum_m E[Y(a, m) \mid W = w] P(M(a^*) = m \mid W = w) \tag{2}$$

$$= \sum_m E[Y(a, m) \mid A = a, W = w] P(M(a^*) = m \mid W = w) \tag{3}$$

$$= \sum_m E[Y(a, m) \mid A = a, M = m, W = w] P(M(a^*) = m \mid A = a^*, W = w) \tag{4}$$

$$= \sum_m E(Y \mid A = a, M = m, W = w) P(M = m \mid A = a^*, W = w) \tag{5}$$

(1): by law of total expectation; (3): by randomization cond. 2; (5): by consistency cond.
(2): by randomization cond. 4; (4): by randomization cond. $1 + 3$;

Special case: linear outcome and mediator regression models

If the partially linear models

$$E(Y \mid M = m, A = a, W = w) = \beta_M m + \beta_A a + f_1(w)$$
$$E(M \mid A = a, W = w) = \alpha_A a + f_2(w)$$

hold for unspecified functions f_1 and f_2, then it follows that

$$NDE = \beta_A \quad \text{and} \quad NIE = \alpha_A \beta_M \ .$$

This is true for both the pure and total versions of these estimands. This is referred to as the classical **product of coefficients** method of Baron & Kenney (1986).

What if the regression models above include interaction terms?

We now consider estimation strategies based on the G-computation formula

$$E\left[E\left[E\left(Y \mid M, A = a, W\right) \mid A = a^*, W\right]\right] = E\left[\sum_m \bar{Q}(m, a, W) g_M(m \mid a^*, W)\right].$$

Approach #1: via estimation of the mediator distribution

1. obtain estimates $\bar{Q}_n$ of $\bar{Q}$ and $g_{M,n}$ of g_M;
2. define final estimate $\frac{1}{n} \sum_{i=1}^n \sum_m \bar{Q}_n(m, a, W_i) p_n(m \mid a^*, W_i)$.

If M is binary, this traditional approach is easy to use.

Otherwise, restrictive (i.e., parametric) models for g_M are typically used, and even then, the problem can involve a difficult (numerical) summation/integration step.

We now consider estimation strategies based on the G-computation formula

$$E\left[E\left[E\left(Y\mid M,A=a,W\right)\mid A=a^*,W\right]\right]=E\left[\sum_m \bar{Q}(m,a,W)g_M(m\mid a^*,W)\right].$$

Approach #2: via sequential regression

1. obtain estimate $\bar{Q}_n$ of $\bar{Q}$;
2. regress $\bar{Q}_n(M,a,W)$ on W using only observations with $A=a^*$, and compute the regression prediction $\bar{Q}_{M,n,i}$ for patient $i=1,2\dots,n$;
3. define final estimate as the average of all predictions.

This modern approach works well irrespective of the type of variable M is, and avoids having to estimate more than is needed.

Using the data simulated previously, we demonstrate how to estimate natural direct and indirect effects when estimating the (binary) mediator distribution.

```r
# fit outcome regression (include interaction because we can)
or_fit <- glm(Y ~ A + M + W1 + W2 + A*M + M*W1,
              family = binomial(), data = full_data)

# need E(Y | A = 0/1, M = 0/1, W1 = W1i, W2 = W2i)
get_EY_a_m_Wi <- function(full_data, or_fit, a, m){
    data_Aa_Mm_Wi <- full_data
    data_Aa_Mm_Wi$A <- a; data_Aa_Mm_Wi$M <- m
    predict(or_fit, newdata = data_Aa_Mm_Wi, type = "response")
}

EY_A0_M0_Wi <- get_EY_a_m_Wi(full_data, or_fit, a = 0, m = 0)
EY_A0_M1_Wi <- get_EY_a_m_Wi(full_data, or_fit, a = 0, m = 1)
EY_A1_M0_Wi <- get_EY_a_m_Wi(full_data, or_fit, a = 1, m = 0)
EY_A1_M1_Wi <- get_EY_a_m_Wi(full_data, or_fit, a = 1, m = 1)
```

We have computed $\bar{Q}_n(m, a, W_i)$ for $a = 0, 1$, $m = 0, 1$ and $i = 1, 2, \ldots, n$, and must now compute

$$\sum_{m=0}^{1} \bar{Q}_n(m, a, W_i) g_{M,n}(m \mid a^*, W_i) \ .$$

We first estimate the mediator distribution $g_M(m \mid a^*, w) = P(M = m \mid A = a^*, W)$.

```
# include interactions -- why not?
med_fit <- glm(M ~ A*W1 + W1*W2, family = binomial(), data = full_data)
```

Next, for $m = 0, 1$, we evaluate $g_{M,n}(m \mid a^*, W_i)$.

```r
# estimates of P(M = m | A = a, W = W_i)
get_Pm_a_Wi <- function(full_data, med_fit, a, m){
    data_Aa_Wi <- full_data; data_Aa_Wi$A <- a
    p <- predict(med_fit, newdata = data_Aa_Wi, type = "response")
    if(m == 1){
        p
    }else{
        1 - p
    }
}
PM0_A0_Wi <- get_Pm_a_Wi(full_data, med_fit, a = 0, m = 0)
PM1_A0_Wi <- get_Pm_a_Wi(full_data, med_fit, a = 0, m = 1)
PM0_A1_Wi <- get_Pm_a_Wi(full_data, med_fit, a = 1, m = 0)
PM1_A1_Wi <- get_Pm_a_Wi(full_data, med_fit, a = 1, m = 1)
```

Estimation based on the G-computation formula

We can then compute $\sum_{m=0}^{1} \bar{Q}_n(m, a, W_i) g_{M,n}(m \mid a^*, W_i)$ for $i = 1, 2, \ldots, n$.

```
# E(E(Y | A = 1, M, W) | A = 1, W)
EY1M1_Wi <- EY_A1_M1_Wi * PM1_A1_Wi + EY_A1_M0_Wi * PM0_A1_Wi
# E(E(Y | A = 0, M, W) | A = 1, W)
EY0M1_Wi <- EY_A0_M1_Wi * PM1_A1_Wi + EY_A0_M0_Wi * PM0_A1_Wi
# E(E(Y | A = 1, M, W) | A = 0, W)
EY1M0_Wi <- EY_A1_M1_Wi * PM1_A0_Wi + EY_A1_M0_Wi * PM0_A0_Wi
# E(E(Y | A = 0, M, W) | A = 0, W)
EY0M0_Wi <- EY_A0_M1_Wi * PM1_A0_Wi + EY_A0_M0_Wi * PM0_A0_Wi
```

Estimation based on the G-computation formula

Finally, we average over distribution of W to get effect estimates.

```
# estimate of E[Y(1, M(1))]
E_Y1M1 <- mean(EY1M1_Wi)
# estimate of E[Y(0, M(1))]
E_Y0M1 <- mean(EY0M1_Wi)
# estimate of E[Y(1, M(0))]
E_Y1M0 <- mean(EY1M0_Wi)
# estimate of E[Y(0, M(0))]
E_Y0M0 <- mean(EY0M0_Wi)
```

These values can be combined to estimate the desired natural direct or indirect effects.

Using the data simulated previously, we demonstrate how to estimate natural direct and indirect effects using sequential regression.

We will illustrate estimation of $E[Y(1, M(0))]$.

Note that we have already estimated $\bar{Q}(m, a, w)$, and we now use $\bar{Q}_n(M_i, a, W_i)$ as outcome in the sequential regression.

```
# estimate of E(Y | A = 1, W = W_i, M = M_i)
data_Aa_Mi_Wi <- full_data
data_Aa_Mi_Wi$A <- 1
full_data$Qbar2 <- predict(or_fit, newdata = data_Aa_Mi_Wi,
                           type = "response")
# estimate E(E(Y | A = 1, W, M) | A, W)
seq_or_fit <- glm(Qbar2 ~ A*W1 + A*W2, family = binomial(),
                  data = full_data)
```

We then obtain the prediction from the fitted sequential regression by plugging in $(A, W) = (0, W_i)$, $i = 1, 2, \ldots, n$, and finally average all obtained predictions.

```
data_A0 <- full_data; data_A0$A <- 0
Qbar1 <- predict(seq_or_fit, newdata = data_A0, type = "response")
mean(Qbar1)

## [1] 0.509928
```

Multiply-robust estimation

The G-computation plug-in estimation procedures typically only provide valid inference when simple (i.e., parametric or empirical) regression techniques are used. When flexible learning is used instead, more sophisticated approaches are needed.

Let $\bar{Q}_M(w) := \sum_m \bar{Q}(m, a, w) g_M(m \mid a^*, w)$ and denote by $\bar{Q}_{M,n}$ the estimator of $\bar{Q}_M$ based on $\bar{Q}_n$ and $g_{M,n}$ so that
$$\psi_{n,G}(a, a^*) := \frac{1}{n} \sum_{i=1}^{n} \bar{Q}_{M,n}(W_i)$$
is the G-computation plug-in estimator of $\psi(a, a^*)$ based on approach #1.

An **AIPW estimator** of $\psi(a, a^*)$ is then given by
$$\psi_{n,AIPW} := \psi_{n,G}(a, a^*) + B_n(\bar{Q}_n, p_n, g_{A,n}) \ ,$$
where the augmentation term is defined as

$$
\begin{aligned}
B_n(\bar{Q}_n, g_{M,n}, g_{A,n}) \ := \ & \frac{1}{n} \sum_{i=1}^{n} \frac{I(A_i = a)}{g_{A,n}(a \mid W_i)} \frac{g_{M,n}(M_i \mid a^*, W_i)}{g_{M,n}(M_i \mid a, W_i)} [Y_i - \bar{Q}_n(M_i, a, W_i)] \\
& + \frac{1}{n} \sum_{i=1}^{n} \frac{I(A_i = a^*)}{g_{A,n}(a^* \mid W_i)} [\bar{Q}_n(M_i, a, W_i) - \bar{Q}_{M,n}(W_i)] \ .
\end{aligned}
$$

Properties of the AIPW estimator: (Tchetgen Tchetgen & Shpitser, 2011)

- Doubly-robust consistency

$\psi_{n,AIPW}(a, a^*) \xrightarrow{P} \psi(a, a^*)$ provided at least 2 of these 3 statements hold:

$$\text{(i) } \bar{Q}_n \xrightarrow{P} \bar{Q}, \quad \text{(ii) } g_{A,n} \xrightarrow{P} g_A \text{ and } \text{(iii) } g_{M,n} \xrightarrow{P} g_M .$$

- Asymptotic normality

Under certain regularity conditions (allowing some flexible learning), we have that

$$\psi_{n,AIPW}(a, a^*) - \psi(a, a^*) \approx \frac{1}{n} \sum_{i=1}^{n} D_i ,$$

where $\quad D_i := \dfrac{I(A_i = a)}{g_A(a \mid W_i)} \dfrac{g_M(M_i \mid a^*, W_i)}{g_M(M_i \mid a, W_i)} [Y_i - \bar{Q}(M_i, a, W_i)]$

$$+ \frac{I(A_i = a^*)}{g_A(a^* \mid W_i)} [\bar{Q}(M_i, a, W_i) - \bar{Q}_M(W_i)] + \bar{Q}_M(W_i) - \psi(a, a^*) .$$

This implies asymptotic normality, but also much more. . .

Can the AIPW estimator be re-expressed in terms of the sequential regression approach that circumvents explicit estimation of the mediator distribution?

Yes! To do so, first define the expanded treatment propensity

$$\tilde{g}_A(m, w) := P(A = a \mid M = m, W = w) \, .$$

Using Bayes' Theorem, it can be shown that

$$\frac{g_M(m \mid a^*, w)}{g_M(m \mid a, w)} = \frac{g_A(a \mid w)}{g_A(a^* \mid w)} \cdot \frac{\tilde{g}_A(a^* \mid m, w)}{\tilde{g}_A(a \mid m, w)} \, .$$

Multiply-robust estimation

This allows us to define a revised **AIPW estimator** with augmentation

$$\tilde{B}_n(\bar{Q}_n, g_{A,n}, \tilde{g}_{A,n}) := \frac{1}{n} \sum_{i=1}^n \frac{I(A_i = a)}{g_{A,n}(a^* \mid W_i)} \frac{\tilde{g}_{A,n}(a^* \mid M_i, W_i)}{\tilde{g}_{A,n}(a \mid M_i, W_i)} [Y_i - \bar{Q}_n(M_i, a, W_i)]$$

$$+ \frac{1}{n} \sum_{i=1}^n \frac{I(A_i = a^*)}{g_{A,n}(a^* \mid W_i)} [\bar{Q}_n(M_i, a, W_i) - \bar{Q}_{M,n}(W_i)] \ ,$$

where $\tilde{g}_{A,n}$ is an estimator of $\tilde{g}_A$.

Properties of this revised AIPW estimator and the corresponding TMLE estimator are described in Zheng & van der Laan (2012).

In this procedure, instead of estimating g_M, we estimate an additional binary regression $\tilde{g}_A$. This is most useful when M is multivariate and/or continuous-valued.

The `medoutcon package in R` has an implementation of these methods.

References:

Naimi, AI, Kaufman, JS, MacLehose, RF (2014). Mediation misgivings: ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology*, 43(5):1656-1661. doi: 10.1093/ije/dyu107

Andrews, RM, Didelez, V (2020). Insights into the cross-world independence assumption of causal mediation analysis. *Epidemiology*, 32(2):209-219. 10.1097/EDE.0000000000001313

Baron, RM, Kenny, DA (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182. 10.1037/0022-3514.51.6.1173

Tchetgen Tchetgen, E, Shpitser, I (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of Statistics*, 43(3):1816-1845. doi: 10.1214/12-AOS990

Zheng, W, van der Laan, MJ (2012). Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, 8(1). doi: 10.2202/1557-4679.1361

Additional reading:

Rudolph, KE, Goin, DE, Paksarian, D, Crowder, R, Merikangas, KR, Stuart, EA (2019). Causal mediation analysis with observational data: considerations and illustration examining mechanisms linking neighborhood poverty to adolescent substance use. *American Journal of Epidemiology*, 188(3):598-608. doi: 10.1093/aje/kwy248