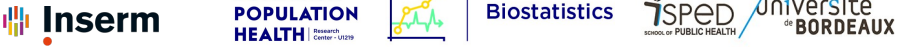
Introduction to mediation Analysis with R package CMAverse

Cécile Proust-Lima
Adapted from Linda Valeri's course notes

INSERM U1219, Bordeaux Population Health Research Center, Bordeaux, France
Univ. Bordeaux, ISPED, Bordeaux, France
cecile.proust-lima@inserm.fr

Inserm Workshop 282 - Best practices and recent advances in causal analyses
October 2025



Material available here

Session 3

Mediation analysis

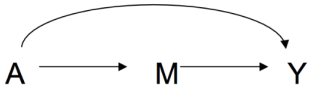
= Central causal inference approach to understand and disentangle exposure-outcome mechanisms

Notations: A exposure, M mediator, Y outcome, C / L confounders

The question: The extent to which the effect of A on Y is direct or it is mediated by an intermediate variable M

Decomposition of the total effect of exposure A on outcome Y into:

- direct effect
- indirect effect mediated through intermediate variable M



Some examples / research questions

- Scientific understanding of mechanisms
  - Do genetic variants affect lung cancer through smoking or independently?
  - Through which pathways do risk factors for dementia operate? Brain atrophy or vascular lesions?
- Assessment of a theory
  - Is the association between childhood SES and Stroke mediated by adult SES?
  - Could racial disparities in access to surgery treatment explain differences in survival of cancer patients?
- Intervening on a mediator to limit the effect of exposures
  - same examples if mediation is shown
- Refining an intervention
  - How can we explain that anti-amyloid treatments fail in dementia research? Lack of mediation through amyloid plaques?
- ...

Challenges in Mediation Analysis

- Mathematically define the causal effects
- Explore their identifiability from data
- Account for the complexity of the data in their estimation (missing data, measurement error, interactions, etc)

Traditional Approach to Mediation Analysis: the difference method

- Classically, in much epidemiological research, the mediation analysis is approached by the regression:

1. regress the outcome Y on the exposure A and confounding factors C

$$\mathbb{E}(Y|A=a, C=c) = \phi_0 + \phi_1 a + \phi_2^\top c$$

2. regress the outcome Y on the exposure A, mediator M and confounding factors C

$$\mathbb{E}(Y|A=a, C=c, M=m) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3^\top c$$

3. Compare the results!
   - ★ Total effect = ϕ_1 Direct effect = θ_1 Indirect effect = $\phi_1 - \theta_1$
   - ★ This is the difference method, used quite often in epidemiology

Traditional Approach to Mediation Analysis: the product method

- Another approach is usually preferred in social sciences: the product method (Baron and Kenny, 1986)

1. regress the mediator M on the exposure A and confounding factors C

$$\mathbb{E}(M|A=a,C=c) = \beta_0 + \beta_1 a + \beta_2^\top c$$

2. regress the outcome Y on the exposure A, mediator M and confounding factors C

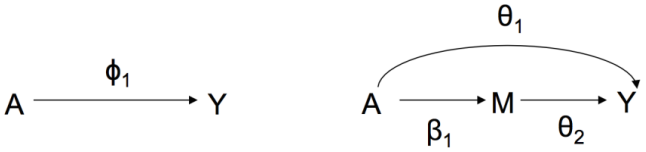
$$\mathbb{E}(Y|A=a,C=c,M=m) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3^\top c$$

3. Combine the results!
   - ★ Direct effect = θ_1 Indirect effect = $\theta_2 \times \beta_1$ Total effect = $\theta_1 + \theta_2 \times \beta_1$
   - ★ Spirit of the Structural Equation Models in social sciences

Traditional Approach to Mediation Analysis: in summary

- Three models:

$$\begin{cases} \mathbb{E}(Y \mid A=a, C=c) & = & \phi_0 + \phi_1 a + \phi_2^\top c \\ \mathbb{E}(Y \mid A=a, C=c, M=m) & = & \theta_0 + \theta_1 a + \theta_2 m + \theta_3^\top c \\ \mathbb{E}(M \mid A=a, C=c) & = & \beta_0 + \beta_1 a + \beta_2^\top c \end{cases}$$



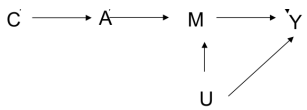
- Computation of the indirect computed effects:
  - $\theta_2 \times \beta_1$ with the product method
  - $\phi_1 - \theta_1$ with the difference method

➙ These approaches are model driven and are equivalent under joint normality (McKinnon, 1995)

Critical issues

1. Even if A is randomized or if all of the A-Y confounders are included, there may be confounders of the M-Y relationship

→ Need to control for mediator-outcome confounders, otherwise potential for high bias



2. The standard approach presupposes no interactions between the effects of the exposure and the mediator on the outcome

$$\mathbb{E}(Y|A=a,C=c,M=m) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta_4^\top c$$

→ May lead to invalid conclusions if there is an interaction

Example with binary M and true model above:

- if $\theta_1 = 0.5$ and $\theta_3 = -1$, neglecting the interaction would give a $\theta_1 \approx 0$ when the effects are +0.5 if m=0 and -0.5 if m=1. We would conclude that almost all of the effect is mediated.
- if no effet of A on M but an interaction between A and M on Y, then conclusion of a mediated effect when there is no mediation.

Alternative approach: the counterfactual framework

- define the causal effects of interest
- identify the assumptions under which they can be computed
- translate the causal effects according to the observations available
- estimate them using a working model
- eventually explore assumptions using sensitivity analyses (unmeasured confounding, measurement error)

(e.g. VanderWeele and Vansteelandt 2009, 2010; Valeri and VanderWeele 2013; Imai 2010)

Counterfactual approach: the notations

Y = outcome of interest for each individual

A = exposure or treatment of interest for each individual

M = post-treatment intermediate(s) for each individual
(potentially on the pathway between A and Y)

C = set of covariates for each individual

Y_a = counterfactual outcome (or potential outcome) Y for each individual
when intervening to set A to a

Y_{am} = counterfactual outcome Y for each individual
when intervening to set A to a and M to m

M_a = counterfactual post-treatment intermediate(s) M for each individual
when intervening to set A to a

The principle (Robins and Greenland (1992), Pearl (2001))

- Total Effect

$$TE = Y_1 - Y_0$$

- Controlled direct effect

$$CDE(m) = Y_{1m} - Y_{0m}$$

  - what would be the change in outcome with A=1 and A=0 if the mediator had been controlled to value m (so caution: it depends on the value given for m!)

- Natural direct effect

$$NDE = Y_{1M_0} - Y_{0M_0}$$

  - what would be the change in outcome with A=1 and A=0 if the mediator was set its value had no intervention been made (treatment absent)

- Natural indirect effect

$$NIE = Y_{1M_1} - Y_{1M_0}$$

  - what would be the change in outcome with A=1 (intervention been made), had the mediator been the one under treatment (A=1) or not (A=0)

Decomposition of the total effect

- The total effect decomposes into a direct and indirect effect:

$$\begin{aligned} Y_1 - Y_0 &= Y_{1M_1} - Y_{0M_0} \\ &= (Y_{1M_1} - Y_{1M_0}) + (Y_{1M_0} - Y_{0M_0}) \\ &= NIE + NDE \end{aligned}$$

- Here, there is no assumptions regarding the absence of interaction between the effects of the exposure and the mediator on the outcome
  - in the definition of the natural effects
  - in the decomposition of the total effect

⚠ Controlled direct effects are not used for effect decomposition:
  - The difference between the total effect and controlled direct effect is not a mediated effect (can be non-zero even in the absence of mediation)
  - From a policy perspective we could ask: How much of the effect would remain if we were to intervene on the mediator and fix it to some value.

Other measures

- Different measures can be derived from the direct and indirect effects:
  - Proportion mediated:

$$PM = \frac{NIE}{TE}$$

    - ★ For etiology: what proportion of the effect would have been eliminated had the mediator been fixed to its value under the reference exposure level
    - ⚠ This measure only makes sense if the NIE and NDE are in the same direction

  - Proportion eliminated:

$$PE = \frac{TE - CDE(m)}{TE}$$

    - ★ For policy: what proportion of the effect would have been eliminated had the mediator been fixed to reference value m

  - ...

Decomposition of the total effect: general notations

- Total Effect contrasting treatment levels $A = a$ to $A = a^*$:

$$TE = Y_a - Y_{a^*}$$

- Controlled direct effect

$$CDE(a, a^*, m) = Y_{am} - Y_{a^* m}$$

  - what would be the change in outcome with $A = a$ and $A = a^*$ if the mediator had been controlled to value m (so caution: it depends on the value given for m!)

- Natural direct effect

$$NDE(a, a^*; a^*) = Y_{aM_{a^*}} - Y_{a^* M_{a^*}}$$

  - what would be the change in outcome with $A = a$ and $A = a^*$ if the mediator was set its value had intervention $A = a^*$ been made

- Natural indirect effect

$$NIE(a, a^*; a) = Y_{aM_a} - Y_{aM_{a^*}}$$

  - what would be the change in outcome with $A = a$, had the mediator been the one under $A = a$ or $A = a^*$

Identification of the effects?

- Counterfactual means hypothetical:
  - Although they are all necessary for the computation, they can't be all observed: either Y_1 or Y_0
  - With natural effects, this is even worse: how can Y_{1M_0} be observed for a single individual?!
- With binary A and M:

Individual	A	M_0	M_1	Y_{00}	Y_{10}	Y_{01}	Y_{11}
1	0	0	?	0	?	?	?
2	1	?	1	?	?	?	0
3	1	?	1	?	?	?	1

→ The contrasts are impossible to compute at the individual level but we can identify them on average under certain assumptions.

Definition of average indirect and direct effects

Effects contrasting treatment levels $A = a$ to $A = a^*$

- Average total effect:

$$TE = \mathbb{E}(Y_a - Y_{a^*}|C)$$

- Average controlled direct effect:

$$CDE(a, a^*, m) = \mathbb{E}(Y_{am} - Y_{a^*m}|C)$$

- Average natural direct effect:

$$NDE(a, a^*; a^*) = \mathbb{E}(Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|C)$$

- Average natural indirect effect:

$$NIE(a, a^*; a) = \mathbb{E}(Y_{aM_a} - Y_{aM_{a^*}}|C)$$

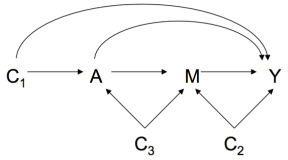
⚠ Here, effects within strata of covariates C. We could take averages over each stratum weighted by the probability $P(C = c)$ to get population averages of the effects

From estimands to estimation: identifiability assumptions

- Positivity: all the outcome values are observable under all the scenarios
- Consistency: the counterfactual equals the corresponding observation

$$Y_{am} = Y|_{A=a,M=m} \quad ; \quad M_a = M|_{A=a}$$

- Ignorability: no unmeasured confounding
  1. No unmeasured exposure-outcome confounding given C: $Y_{am} \perp A \mid C$
  2. No unmeasured mediator-outcome confounding given C: $Y_{am} \perp M \mid C, A$
  3. No unmeasured exposure-mediator confounding given C: $M_a \perp A \mid C$
  4. No effect of exposure that confounds the mediator-outcome relationship (for NDE/NIE): $Y_{am} \perp M_{a*} \mid C$

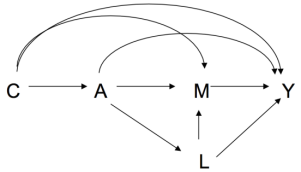


⚠These assumptions make clear and precise the no-unmeasured-confounding assumptions required for a causal interpretation

Critical assumption 4

⚠ Assumption 4 stipulates that there is no mediator-outcome confounder affected by the exposure

This is violated in the following causal diagram:



e.g. A=Childhood SES, M=Adult SES, L=Adult Risk Factors, Y=Stroke

e.g. A=prenatal care, M=pre-eclampsia, L=tobacco, Y=pre-term birth

If there are effects of A that confound the mediator-outcome relationship such as L then natural direct and indirect effects are not identified regardless of whether data is available on L

➙ See discussions of the literature on this issue (e.g., Vansteelandt 2012 (NDE/NIE among exposed) ; Avin 2005 (path-specific effects), VanderWeele 2014 (limitation to short delays, joint mediators, stochastic interventions), VanderWeele 2009 (marginal structural model for CDE))

From estimands to estimation: the G-formula

- Under the previous assumptions, we can translate the definitions based on counterfactual variables into expressions based on the actual observations:

$$
\begin{aligned}
CDE(a,a^*,m|C=c) &= \mathbb{E}(Y_{am}-Y_{a^*m}|C) \\
&= \mathbb{E}(Y|A=a,M=m,C=c)-\mathbb{E}(Y|A=a^*,M=m,C=c)
\end{aligned}
$$

$$
\begin{aligned}
NDE(a,a^*;a^*|C=c) &= \mathbb{E}(Y_{aM_{a^*}}-Y_{a^*M_{a^*}}|C) \\
&= \int_m \big(\mathbb{E}(Y|A=a,M=m,C=c)-\mathbb{E}(Y|A=a^*,M=m,C=c)\big)f(M=m|A=a^*,c)dm
\end{aligned}
$$

$$
\begin{aligned}
NIE(a,a^*;a|C=c) &= \mathbb{E}(Y_{aM_a}-Y_{aM_{a^*}}|C) \\
&= \int_m \mathbb{E}(Y|A=a,M=m,C=c)\big(f(M=m|A=a,c)-f(M=m|A=a^*,c)\big)\,dm
\end{aligned}
$$

  - to be marginalized by weighting by $P(C=c)$ and summing

→ From this, modeling techniques can be used to estimate these quantities, including classical regression techniques!

Estimating counterfactual contrasts using regression techniques

- Define direct and indirect effects under the counterfactual framework
- Choose a working model that accounts for the complex nature of the data (interactions, non-linear effects, time to event, spatial and temporal structure)
- Don't forget about the additional pitfalls of observational data (measurement error, missing data, selection bias, unmeasured confounding)
- Assess the uncertainty

(See e.g., VanderWeele and Vansteelandt (2009, 2010), Imai et al. (2010), Valeri and VanderWeele (2013))

→ Software solutions: R packages mediation (Imai, 2010), medflex (Steen, 2017), **CMAverse** (Shi, 2021) + *SAS, Stata, SPSS programs ...* (Valeri, 2013)

Example of use with continuous outcome and mediator

- Let Y denote the continuous outcome, M the continuous intermediate variables, A the exposure and C additional covariates of interest

$$\begin{cases} \mathbb{E}(Y \mid A=a, C=c, M=m) &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta_4^\top c \\ \mathbb{E}(M \mid A=a, C=c) &= \beta_0 + \beta_1 a + \beta_2^\top c \end{cases}$$

- With the previous assumptions, we have (VanderWeele and Vansteelandt, 2009)

$$CDE = (\theta_1 + \theta_3 m)(a - a^*)$$

$$NDE = (\theta_1 + \theta_3 \times \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2^\top c)(a - a^*)$$

$$NIE = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)$$

- Variance can be obtained by the Delta-Method
- marginal effects can be obtained by replacing c by E(C)

Example of use with continuous outcome and mediator: remarks

- Previous calculations reduce to product definition if no interaction (Baron and Kenny, 1986) but with the counterfactual approach, we can go further
- It follows that classical techniques (Difference and Product techniques) give valid estimates of direct and indirect effects provided:
  - (i) The model without interaction is correctly specified
  - (ii) The no-unmeasured assumptions hold

- In case of violation of these assumptions:
  - (i) The counterfactual-based regression approach can work
  - (ii) Sensitivity analyses can be carried out

Example with binary outcome: odds-ratios (VanderWeele and Vansteelandt, 2010)

- Controlled direct effect comparing treatment level A=1 to A=0 setting M=m:

$$CDE^{OR}(m|c) = \frac{P(Y_{1m} = 1|c)/P(Y_{1m} = 0|c)}{P(Y_{0m} = 1|c)/P(Y_{0m} = 0|c)}$$

- Natural Direct Effect comparing treatment level A=1 to A=0 :

$$NDE^{OR}(m|c) = \frac{P(Y_{1M_0} = 1|c)/P(Y_{1M_0} = 0|c)}{P(Y_{0M_0} = 1|c)/P(Y_{0M_0} = 0|c)}$$

- Natural Indirect Effect comparing treatment level A=1 to A=0 :

$$NIE^{OR}(m|c) = \frac{P(Y_{1M_1} = 1|c)/P(Y_{1M_1} = 0|c)}{P(Y_{1M_0} = 1|c)/P(Y_{1M_0} = 0|c)}$$

- Total Effect: $TE = NDE \times NIE$

⚠ conditional on $C = c$

Example with binary outcome: odds-ratios (VanderWeele and Vansteelandt, 2010)

- Let Y denote the binary outcome, M the continuous intermediate variables, A the exposure and C additional covariates of interest

$$\begin{cases} logit(\mathbb{P}(Y \mid A=a, C=c, M=m)) &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta_4^\top c \\ \mathbb{E}(M \mid A=a, C=c) &= \beta_0 + \beta_1 a + \beta_2^\top c \end{cases}$$

- Under the previous assumptions, we have

$$log(OR^{CDE}) = (\theta_1 + \theta_3 m)(a - a^*)$$

- And provided the outcome is rare + the error in the linear regression for M is normal and homoskedastic with variance σ^2:

$$log(OR^{NDE}) = (\theta_1 + \theta_3 \times (\beta_0 + \beta_1 a^* + \beta_2^\top c + \theta^2 \sigma^2)(a - a^*) + 0.5\theta_3^2 \sigma^2 (a^2 - a^{*2})$$

$$log(OR^{NIE}) = (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*)$$

- Variance can be obtained by the Delta-Method

Example with binary outcome: remarks

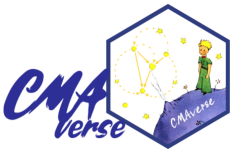
- If no interaction, the expression reduces to the one of the product method.
- It follows that classical techniques (Difference and Product techniques) give valid estimates of natural direct and indirect effects provided:
  - (i) The model without interaction is correctly specified
  - (ii) The no-unmeasured assumptions hold
  - (ii) The outcome is rare
- In case of violation of these assumptions:
  - (i) The counterfactual-based regression approach can work
  - (ii) Sensitivity analyses can be carried out
  - (iii) A log-linear model can be used instead of the logistic regression
- With case-control study, need to weight the mediator regression with weights π/p for cases and $(1-\pi)/(1-p)$ for controls where π is the prevalence and p is the proportion of cases + robust variances.

The R package CMAverse

The R package CMAverse developed by the group of Linda Valeri provides a suite of functions for reproducible causal mediation analysis:

- cmdag function for DAG visualization
- cmest for statistical modeling
- cmsens for sensitivity analysis for unmeasured confounding and measurement error

The website includes several vignettes presenting the package and describing its functionalities:
`https://bs1125.github.io/CMAverse/`



Pure *versus* Total natural effects?

- We have defined so far the natural effects as:

$$\begin{aligned} NDE &= Y_{1M_0} - Y_{0M_0} \\ NIE &= Y_{1M_1} - Y_{1M_0} \end{aligned}$$

  - These are the pure direct effect and the total indirect effect

- We could instead use:

$$\begin{aligned} NDE &= Y_{1M_1} - Y_{0M_1} \\ NIE &= Y_{0M_1} - Y_{0M_0} \end{aligned}$$

  - These are the total direct effect and the pure indirect effect

- the difference is about what we use as a reference: 1 for total (we remove the intervention), 0 for pure (we add the intervention)
- The total indirect effect is usually to be favored: it gives stronger evidence for the actual operation of mediating pathways (VanderWeele, 2011)

Computation of the contrasts in more general settings

- So far, using classical regressions, we have found explicit expressions of the natural effects.
- More generally, we have the following expressions:

$$NDE(a,a^*;a^*) = \int\int \mathbb{E}(Y|A=a,m,c) - \mathbb{E}(Y|A=a^*,m,c) f(M=m|A=a^*,c) f(c) dmdc$$

$$NIE(a,a^*;a) = \int\int \mathbb{E}(Y|A=a,m,c)(f(M=m|A=a,c) - f(M=m|A=a^*,c)) f(c) dmdc$$

- ⇝ Different models can be used, including ML approaches (Liu et al. 2024)
- These contrasts can be estimated using Bootstrap technique and Monte-Carlo technique

- ⇝ Monte-Carlo: approximate the expressions by sampling multiple M from its distribution
- ⇝ Bootstrap: Estimation on bootstrapped samples, calculation of the expressions, and average + percentiles for the 95% confidence interval

What about stochastic interventions?

- NDE / NIE identified under a very restricting assumption :
  - Assumption 4: "no confounder of mediator-outcome relation should be impacted by the exposure"
  - Required because we want to fix to what would have occured in the "other" counterfactual world with $Y_{aM_{a^*}}$: *"the natural value we would have had"*
- This assumption not required anymore if we consider another causal interpretation:
  - The stochastic intervention / randomized intervention (cf. Lin 2017, Diaz, 2020; VanderWeele et al 2014)
  - Compares the outcome by setting the mediator path to a random draw from the distribution of interest:

$$Y_{aM_{a^*}} \quad \Rightarrow \quad Y_{am} \text{ with } m \sim \mathscr{D}_{a^*}$$

  - → Particularly relevant in some settings, e.g. residual disparities (Valeri et al., 2016, Valeri 2023)

- different interpretation but exact same calculations (without the need of assumption 4)!

Data handled in CMAverse

- Outcomes: continuous, binary, nominal, ordinal, count, survival
- Mediators: continuous, binary, nominal, ordinal, count, survival
- Exposure: continuous, binary, nominal, ordinal, count

⚠ not all the methods work with all the data types

Techniques handled in CMAverse

- cmest handles much more methods than seen here:
  - regression-based approach (Valeri 2013 and VanderWeele 2014)
    with closed-form estimates, as shown previously
  - weighting-based approach (VanderWeele 2014) (if categorical only):
    inverse probability weighting
  - inverse odds-ratio weighting approach (Tchetgen Tchetgen 2013) (categorical only)
  - natural effect model (Vansteelandt, Bekaert, Lange 2012):
    expand the dataset by imputing the measure for the ("missing") counterfactual
  - marginal structural model (VanderWeele 2017) (fine when $A \rightarrow L$ + if categorical M only):
    counterfactual model for M and for Y, then combined - uses weights
  - g-formula approach (Robins 1992) (fine when $A \rightarrow L$):
    posterior calculation, compatible with stochastic effects in presence of L affected by A

- CMAverse handles other data settings:
  - survival data with multi-state setting (and G-formula) using cmest_multistate (Valeri 2023, Wang 2025)

Sensitivity Analyses using CMAverse

- Two types of sensitivity analyses can be carried out using `cmsens`

  1. Umeasured confounding: uses the E-value (Smith 2019)
     - ★ Easier with categorical outcomes. For a continuou outcome, transformation into ratio scale.
  2. Measurement error: uses regression calibration or SIMEX (Valeri et al, 2014)
     - ★ SIMEX for continuous or categorical variable
     - ★ Regression calibration with independent continuous variable

Practice with CMAverse: example

Assess whether the effect of smoking status on cardiovascular health is mediated by adiposity and inflammation

or

What are the roles of adiposity and inflammation in the effect of smoking status on cardiovascular health

- Simulated data:
  - Exposure = smoking status, binary
  - Confounders = age, sex (male, ref female)
  - First mediator/confounder = fat indicator (e.g., BMI, body fat percentage, trunk fat percentage, waist circumference), binary
  - Second mediator = Inflammation (measured by C-reactive protein (CRP)), continuous
  - Outcome = cardiovascular health, Framingham Risk Score (FRS) (could have been blood pressure too), continuous

Practice with CMAverse: two settings

- Setting 1: a single mediator (no post-exposure confounder)
  - Construct the DAG with cmdag
  - Compute "manually" the natural indirect and direct effects using classical regression
  - Compute the effects using cmest with classical regressions
  - Compute the effects using cmest with other techniques
- Setting 2: a post-exposure confounder and a mediator
  - Construct the DAG with cmdag
  - Compute the effects using cmest with g-formula
  - Investigate measurement error in the mediator with cmsens

Concluding remarks

- Traditional regression techniques do not clearly define the assumptions under which causal interpretations can be made, with potentially high bias, misleading conclusions.
- Counterfactual approach formalizes the assumptions by focusing on what we want to assess and the assumptions required to estimate it.
- Natural direct and indirect effects rely on stringent assumption of no M-O confounder affected by the exposure
- Alternatives exist: joint mediation, path-specific effects (same type of interpretation) or stochastic analogues (different interpretation)
- CMAverse offers a large set of options to be explored for "classical" mediation (with the nice feature of sensitivity checking). Other software for mediation exist!

References

Avin, Chen, Ilya Shpitser, et Judea Pearl. 2005. Identifiability of path-specific effects. Proceedings of IJCAI'05, 357-63.

Baron, Reuben M., et David A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51(6): 1173-82.

Díaz, Iván, et Nima S. Hejazi. 2020. Causal Mediation Analysis for Stochastic Interventions. JRRS B 82(3): 661-83.

Imai, K., L. Keele, D. Tingley, et T. Yamamoto. 2010. Causal Mediation Analysis Using R. In Advances in Social Science Research Using R, Springer.

Imai, Kosuke, Luke Keele, et Dustin Tingley. 2010. A general approach to causal mediation analysis. Psychological Methods 15: 309-34.

Lin, Sheng-Hsuan, Jessica Young, Roger Logan, Eric J. Tchetgen Tchetgen, et Tyler J. VanderWeele. 2017. « Parametric Mediational G-Formula Approach to Mediation Analysis with Time-Varying Exposures, Mediators, and Confounders ». Epidemiology (Cambridge, Mass.) 28 (2): 266-74. https://doi.org/10.1097/EDE.0000000000000609. Liu, Richard, Nicholas T. Williams, Kara E. Rudolph, et Iván Díaz. 2025. General targeted machine learning for modern causal mediation analysis. arXiv:2408.14620.

Mackinnon, David P., Ghulam Warsi, et James H. Dwyer. 1995. A Simulation Study of Mediated Effect Measures. Multivariate Behavioral Research 30 (1): 41.

Pearl, Judea. 2001. Direct and indirect effects. Proceedings of UAI'01, 411-20.

Robins, J. M., et S. Greenland. 1992. Identifiability and Exchangeability for Direct and Indirect Effects. Epidemiology 3(2): 143-55.

Shi, Baoyi, Christine Choirat, Brent A. Coull, Tyler J. VanderWeele, et Linda Valeri. 2021. CMAverse: A Suite of Functions for Reproducible Causal Mediation Analyses. Epidemiology 32 (5): e20-22.

Smith, Louisa H., et Tyler J. VanderWeele. 2019. Mediational E-Values: Approximate Sensitivity Analysis for Unmeasured Mediator-Outcome Confounding. Epidemiology 30(6): 835-37.

Steen, Johan, Tom Loeys, Beatrijs Moerkerke, et Stijn Vansteelandt. 2017. Medflex: An R Package for Flexible Mediation Analysis Using Natural Effect Models. JSS 76: 1-46.

Tchetgen Tchetgen, Eric J. 2014. A Note on Formulae for Causal Mediation Analysis in an Odds Ratiocontext. Epidemiologic Methods 2(1): 21-31.

Valeri, Linda, Jarvis T. Chen, Xabier Garcia-Albeniz, Nancy Krieger, Tyler J. VanderWeele, et Brent A. Coull. 2016. The Role of Stage at Diagnosis in Colorectal Cancer Black-White Survival Disparities: A Counterfactual Causal Inference Approach. Cancer Epidemiology, Biomarkers & Prevention 25(1): 83-89.

References

Valeri, Linda, Xihong Lin, et Tyler J. VanderWeele. 2014. Mediation Analysis When a Continuous Mediator Is Measured with Error and the Outcome Follows a Generalized Linear Model. Statistics in Medicine 33(28): 4875-90.
Valeri, Linda, Cecile Proust-Lima, Weijia Fan, Jarvis T. Chen, et Helene Jacqmin-Gadda. 2023. A Multistate Approach for the Study of Interventions on an Intermediate Time-to-Event in Health Disparities Research. SMMR 32(8): 1445-60.
Valeri, Linda, et Tyler J. VanderWeele. 2013. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychological methods 18 (2): 137-50.
VanderWeele, Tyler J. 2009. Marginal Structural Models for the Estimation of Direct and Indirect Effects. Epidemiology 20 (1): 18-26.
VanderWeele, Tyler J. 2011. Controlled direct and mediated effects: definition, identification and bounds. Scandinavian journal of statistics 38 (3): 551-63.
VanderWeele, Tyler J., et Eric J. Tchetgen Tchetgen. 2017. Mediation Analysis with Time Varying Exposures and Mediators. JRRS B 79 (3): 917-38.
Vanderweele, Tyler J., et Stijn Vansteelandt. 2009. Conceptual Issues Concerning Mediation, Interventions and Composition. Statistics and Its Interface 2 (4): 457-68.
Vanderweele, Tyler J., et Stijn Vansteelandt. 2010. Odds Ratios for Mediation Analysis for a Dichotomous Outcome. American Journal of Epidemiology 172 (12): 1339-48.
VanderWeele, Tyler J., Stijn Vansteelandt, et James M. Robins. 2014. Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology 25(2): 300-306.
Vansteelandt, Stijn, Maarten Bekaert, et Theis Lange. 2012. Imputation strategies for the estimation of natural direct and indirect effects. Epidemiologic Methods 1(1): 131-58. Vansteelandt, Stijn, et Tyler J. VanderWeele. 2012. Natural Direct and Indirect Effects on the Exposed: Effect Decomposition under Weaker Assumptions. Biometrics 68(4): 1019-27.
Wang, Ziqing, Baoyi Shi, Cécile Proust-Lima, Hélène Jacqmin-Gadda, et Linda Valeri. 2025. Multistate Approach for Stochastic Interventions on a Time-to-Event Mediator in the Presence of Competing Risks: A New R Command within the CMAverse R Package. Epidemiology 36(1): 139-40.