Identifying Direct Causal Effects Under Unmeasured Confounding

Philippe Boileau*[1], Nima S. Hejazi*[2], Ivana Malenica*[1], Sandrine Dudoit[1], Mark J. van der Laan[1]

[1]University of California, Berkeley; [2]Weill Cornell Medicine

Introduction & Motivations

- Developing *mechanistic* understandings of causal effects is a ubiquitous goal across scientific disciplines.
- The natural direct and indirect effects are common target causal parameters — both identifiable without structural assumptions.
- Identification assumes absence of unmeasured confounders of exposure–mediator, mediator–outcome, and exposure–outcome pathways — but this is *not* completely necessary.
- The natural direct and indirect effects arise from a decomposition of the average treatment effect (ATE), defined by a
  - *joint* intervention on exposure (A) and mediators (Z), where
  - natural indirect effect (NIE) captures the portion of the ATE that passes only through the mediators, while the
  - natural direct effect (NDE) captures the remainder of the ATE, that is, through all paths excluding Z.

The Statistical Problem

Full (unobserved) data is $X = (W, V, A, Z, Y)$, and we observe $O = (W, A, Z, Y)$. Exposure–mediator confounder V **not** part of O.

Let $O \sim P_0 \in \mathcal{M}$, with $\mathcal{M}$ being the nonparametric *statistical model*. The likelihood of the observed data under $\mathcal{M}$ is:

$$p(o) = \prod_{i=1}^{n} q_Y(y_i \mid w_i, a_i, z_i) q_Z(z_i \mid w_i, a_i) g_A(a_i \mid w_i) q_W(w_i) \ .$$

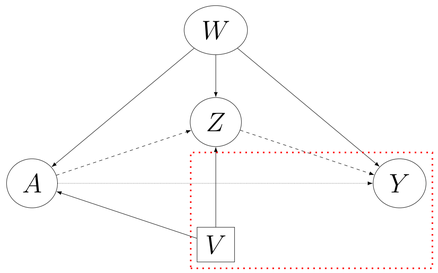
The average treatment effect may be decomposed as

$$\begin{aligned}\Psi_{\text{ATE}}(P_{X,0}) =& \mathbb{E}_{X,0}[Y(1) - Y(0)] \\ =& \underbrace{\mathbb{E}_{X,0}[Y(1, Z(1)) - Y(1, Z(0))]}_{\text{NIE}} \\ &+ \underbrace{\mathbb{E}_{X,0}[Y(1, Z(0)) - Y(0, Z(0))]}_{\text{NDE}} \ ,\end{aligned}$$

where $Y(1, Z(0))$ arises from a joint intervention on the exposure and mediators, setting them to *incompatible* values. The NDE is

$$\Psi_{\text{NDE}}(P_{X,0}) = \int_{\mathcal{W},\mathcal{Z}} \mathbb{E}[Y(1, z) - Y(0, z) \mid W = w] \\ p_Z(z \mid A = 0, w) p_W(w) \ d\mu(z) \ d\nu(w) \ .$$

Causal Identification



(A1) No unmeasured endogenous pathways:
$f_Y(Z, A, W, V, U_Y) \equiv f_Y(Z, A, W, U_Y)$

(A2) Conditional expectation equivalence:
$\mathbb{E}(Y \mid Z, A = 1, W, V) \equiv \mathbb{E}(Y \mid Z, A = 1, W)$

Theorem

Under assumptions A1 and A2, $\Psi_{\text{NDE}}(P_{X,0})$ is identified by

$$\Psi(P_0) = \mathbb{E}_{P_0}\mathbb{E}_{P_0}\{\mathbb{E}_{P_0}(Y \mid W, A = 1, Z) \\ - \mathbb{E}_{P_0}(Y \mid W, A = 0, Z) \mid A = 0, W\} \ .$$

Estimation & Inference

Existing estimation and testing approaches are compatible with this relaxed identification strategy. For example, one could use the

- targeted maximum likelihood estimator of Zheng and van der Laan [2012], or the
- one-step estimator of Tchetgen Tchetgen and Shpitser [2011].

Both are *multiple-robust* (being based on the efficient influence function), *asymptotically linear* under fairly non-restrictive assumptions, and compatible with *cross-fitting* of nuisance estimators.

Both estimators are implemented in the `medoutcon` `R` package (check it out at `https://github.com/nhejazi/medoutcon`).

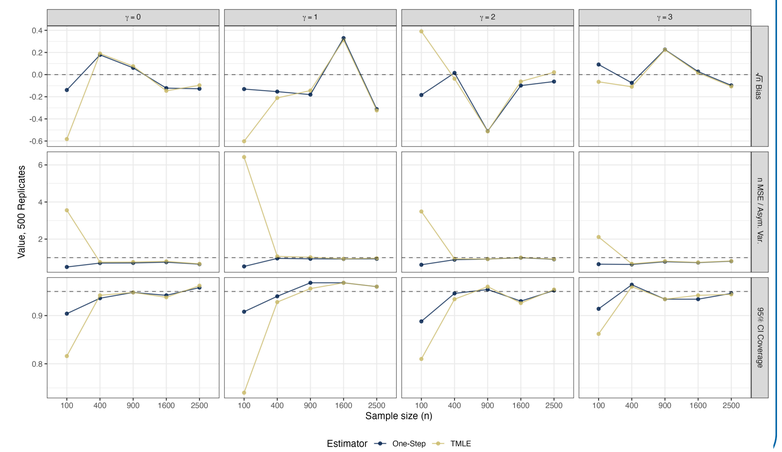
Funding

PB's gratefully acknowledges the support of the FRQNT and NSERC. NSH's work was supported by NSF DMS 2102840.

Numerical Experiments

We consider the following data-generating process:

$$\begin{aligned} W_1 &\sim \text{Unif}(-1, 1); \ W_2, V \sim \text{Norm}(0, 1) \\ A|W, V &\sim \text{Bern}\left((1 + \exp\{-W_1 - W_2 - V\})^{-1}\right) \\ Z|A, W, V &\sim \text{Bern}\left((1 + \exp\{-W_1 - W_2 - \gamma V - 3A\})^{-1}\right) \\ Y|Z, A, W, V &\sim \text{Norm}(3A + W_1 + W_2 + Z, 1) \ . \end{aligned}$$



Conclusions

- The NDE is identifiable under unmeasured exposure–mediator confounding by a *well-studied* statistical functional!
- The NDE has recently been used to study the *proportion of vaccine effect mediated* through candidate immune correlates.
  - Our results strengthen "out-of-the-box" uses of the NDE.
  - Exposure–mediator confounders in vaccine studies: prior infection history, immunocompromised status, genetics.

References

Wenjing Zheng and Mark J. van der Laan. Targeted maximum likelihood estimation of natural direct effects. *The International Journal of Biostatistics*, 8(1):1–40, 2012. doi: doi:10.2202/1557-4679.1361. URL `https://doi.org/10.2202/1557-4679.1361`.

Eric J Tchetgen Tchetgen and Ilya Shpitser. Semiparametric estimation of models for natural direct and indirect effects. Working Paper 129, Harvard University, 2011. URL `https://biostats.bepress.com/harvardbiostat/paper129/`.

* *indicates shared first-authorship, ordered alphabetically.*