Introduction to proximal causal inference

Non-parametric and parametric methods

Keling Wang[1]

[1]Department of Epidemiology, Erasmus MC Rotterdam

02 October 2024

1. Introduction
2. From negative controls to PCI
3. Nonparametric approaches
4. Parametric approaches
5. Generalization
6. References

Known but unmeasured confounder

We never believe conditional exchangeability holds.

Known but unmeasured confounder

We never believe conditional exchangeability holds.

But things can be even worse: There is a known confounder, but unmeasured! – Then we clearly know we get biased.

Known but unmeasured confounder

We never believe conditional exchangeability holds.

But things can be even worse: There is a known confounder, but unmeasured! – Then we clearly know we get biased.

How to survive in the presence of known unmeasured confounders?

Proximal causal inference helps! ... but where does it come from?

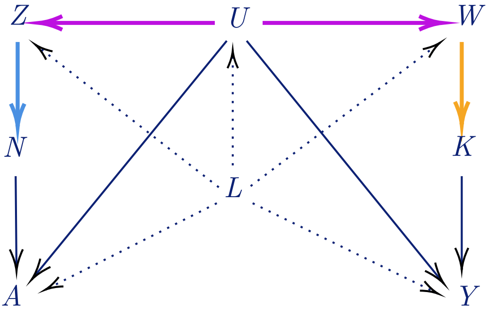
Proximal causal inference helps! ... but where does it come from?

Let's first take a glance and then see how it's formulated...

First glance at proximal causal inference

Proximal causal inference strategy is dedicated to deal with such situation: a (set of) KU- (known but unmeasured) confounder(s).

ATE can be point-identified upon considering a treatment-side proxy Z and an outcome-side proxy W, with additional assumptions.



Alternatives to "KU-confounder"

We need *more information* and alter our *assumption sets*!

Alternatives to “KU-confounder”

We need *more information* and alter our *assumption sets*!

KU-confounder is, nevertheless, easier to deal with compared to a unknown-and-unmeasured confounder. There are already alternatives for a (set of) UU-confounder(s) when conditional exchangeability doesn’t hold:

Alternatives to “KU-confounder”

We need *more information* and alter our *assumption sets*!

KU-confounder is, nevertheless, easier to deal with compared to a unknown-and-unmeasured confounder. There are already alternatives for a (set of) UU-confounder(s) when conditional exchangeability doesn’t hold:

- Instrument variable?
- Negative (population/outcome) control?
- Front door formula (causal mediation)?
- ...

Alternatives to “KU-confounder”

But all the methods listed above are for UU-confounders and are:

Alternatives to “KU-confounder”

But all the methods listed above are for UU-confounders and are:

- **NOT** considering the “known” information for a KU-confounder; and are
- adding new **strict assumptions** (*e.g.* modelling assumptions and rank preservation in NC methods) [1], or
- altering the underlying population (*e.g.* monotonicity assumptions in NC/IV methods)

Alternatives to "KU-confounder"

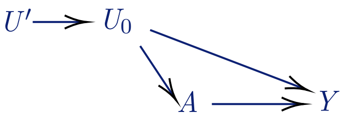
But all the methods listed above are for UU-confounders and are:

- **NOT** considering the "known" information for a KU-confounder; and are
- adding new **strict assumptions** (*e.g.* modelling assumptions and rank preservation in NC methods) [1], or
- altering the underlying population (*e.g.* monotonicity assumptions in NC/IV methods)

We therefore want another approach that integrates available information of "the known part" and does not add strong restrictions / strict assumptions that we don't believe either.

Proxy-based thinking

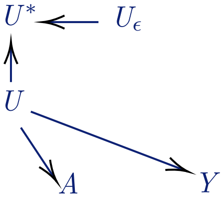
- When data for a confounder U_0 is not available, a natural alternative is to consider one measured proxy [2] of it, U';
- Intuitively, if $U' \propto U_0$ and U' is strictly only associated with U_0, the average treatment effect (ATE) can be point identified without additional assumptions.

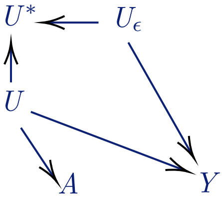


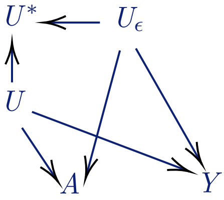
Proxy-based thinking

- However, we can never have a perfect proxy, otherwise it's equivalent to know everything about U_0.
- The actual proxies come with noises and lose information carried by the targeted confounder;
- Without additional **assumptions** a proxy may only mitigate but never eliminate bias [3];
- Error generating mechanism from targeted confounder to proxy is extremely important [2], *e.g.* coarsening, mismeasurement, etc.

Mismeasured variable as proxies

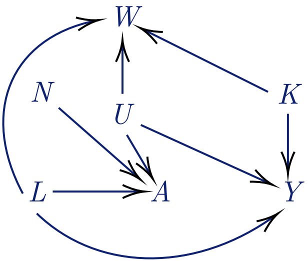


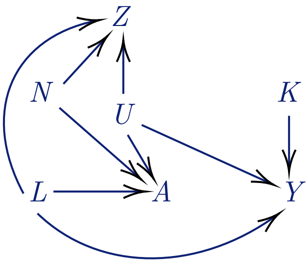




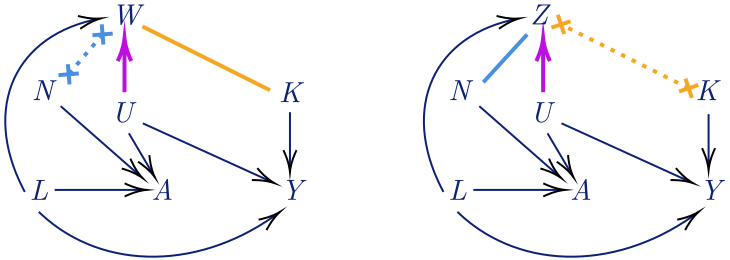
Approaching proxy-based inference (I)

- Negative control (NC) exposure (NC/e) and NC outcome (NC/o) are two most used proxy-based strategies.
- NC/o requires a proxy W that is U-comparable to Y but not caused by A or sharing a common cause [4];
- NC/e requires a proxy Z that is U-comparable to A but not causing Y or sharing a common cause [4].





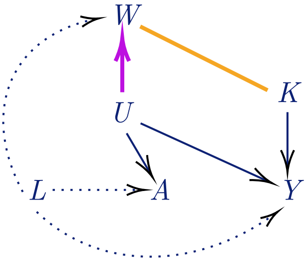
Approaching proxy-based inference (I)

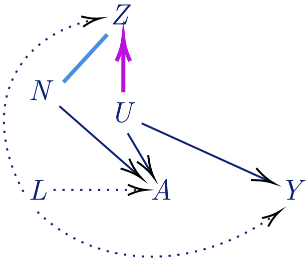


Left: NC/o proxy; proxy W is similar enough to Y, except that it is not caused by A or share common causes;

Right: NC/e proxy; proxy Z is similar enough to A, except that it cannot cause Y or share common causes.

Approaching proxy-based inference (I)



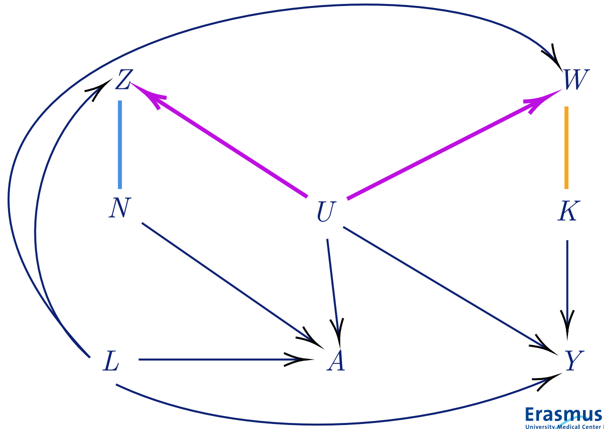


Left: NC/o proxy; proxy W is a child of U, and is (only) d-connected with Y;

Right: NC/e proxy; proxy Z is a child of U, and is (only) d-connected with A.

Approaching proxy-based inference (II)

Upon combining both the NC/o- and NC/e-proxies, we get...

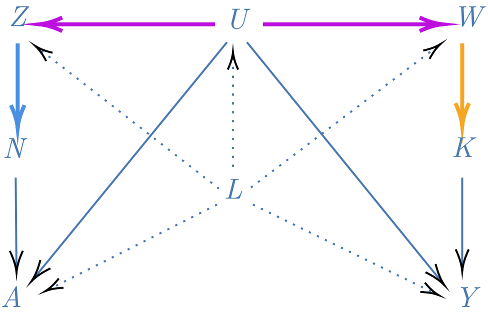


Approaching proxy-based inference (II)

And after some rearrangement, we get our final DAG...

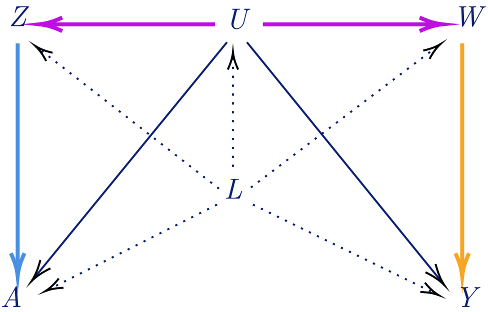
Approaching proxy-based inference (II)

And after some rearrangement, we get our final DAG...



Approaching proxy-based inference (II)

And after some rearrangement, we get our final DAG...



Nonparametric Identification (I)

Given observed individual level data with a binary treatment
$\mathcal{O}_i = \langle Y_i, A_i, \boldsymbol{L}_i, Z_i, W_i \rangle$,
an ATE $\psi = \mathbb{E}[Y^{a=1} - Y^{a=0}]$ is nonparametrically identified **without** the presence of U, if standard identifiability conditions **A1** through **A3** hold:

- **A1** Consistency: $Y_i^a = Y_i \;\; \forall a, \;\; a.s.$
- **A2** Treatment positivity over $\langle \boldsymbol{L}, Z, W \rangle$:
  $\Pr[A = a | \boldsymbol{L}, Z, W] > 0 \;\; \forall a, \;\; a.s.$
- **A3** Exchangeability over $\langle \boldsymbol{L}, Z, W \rangle$: $Y^a \perp\!\!\!\perp A | (\boldsymbol{L}, Z, W) \;\; \forall a$

Nonparametric Identification (I)

With KU-confounders $\boldsymbol{U}_i$, **A3** does not hold, and ATE cannot be identified given $\mathcal{O}_i$ and **A1**-**2**.

Proximal causal inference nonparametrically identifies ATE if additional conditions hold [1]:

- **A4** Independence of Y and Z: $Z \perp\!\!\!\perp Y|(A, \boldsymbol{U}, \boldsymbol{L})$
- **A5** Independence of W and A, W and Z: $W \perp\!\!\!\perp (Z, A)|(\boldsymbol{U}, \boldsymbol{L})$
- **A6** Exchangeability over $\langle \boldsymbol{L}, \boldsymbol{U} \rangle$: $Y^a \perp\!\!\!\perp A|(\boldsymbol{U}, \boldsymbol{L}) \ \ \forall a$
- **A7** Treatment positivity over $\langle \boldsymbol{L}, \boldsymbol{U} \rangle$: $\Pr[A = a|\boldsymbol{L}, \boldsymbol{U}] > 0 \ \forall a, \ \ a.s.$
- **A8** Statistical completeness for $\boldsymbol{U}$ given Z: $\mathbb{E}[g(\boldsymbol{U})|Z, A = a, \boldsymbol{L} = \boldsymbol{l}] = 0 \ a.s. \Leftrightarrow g(\boldsymbol{U}) = 0, \ \forall \ a, \boldsymbol{l}, $ and a square-integrable function g

G-formula (II)

Suppose there exists an function $h = h(w, a, l)$ that solves the equation *a.s.*:

$$\mathbb{E}[Y|A, Z, L] = \int h(w, A, L)\,\mathrm{d}F(w|A, Z, L)$$

then if **A1-2, A4-8** holds, ATE is nonparamentrically identified by: (outcome-side proximal g-formula)

$$\psi = \int_{\mathcal{L}} \int \Big[h(w, a = 1, l) - h(w, a = 0, l)\Big]\,\mathrm{d}F(w|l)\,\mathrm{d}F(l)$$

G-methods are all compatible and can be used for estimation.

G-formula (II)

Assumption **A9** suffices for the existence of such a "bridge" function $h = h(w, a, l)$:

A9 Statistical completeness for Z given W:
$\mathbb{E}[g(Z)|W, A = a, L = l] = 0 \ a.s. \Leftrightarrow g(Z) = 0, \ \forall \ a, l,$ and a square-integrable function g

G-formula (II)

Note: There are alternatives to completeness assumptions **A9** and/or **A8** upon considering different "flows of information" and starting points [5].

With $U|W$ completeness (alternative to **A8**) and $W|Z$ completeness (alternative to **A9**), a treatment bridge function $q = q(z, a, l)$ is valid, and ATE is identified by:
(treatment-side proximal g-formula)

$$\psi = \int_{\mathcal{L}} \int (-1)^{1-a} q(z, a, l) y \, \mathrm{d}F(y, z, a|l) \, \mathrm{d}F(l)$$

Parametric assumptions

Two-stage least squares (2SLS) are implemented the most for parametric estimation of an ATE via proximal causal inference approach.

Given that assumptions **A1, A4-6** hold, and that no structural violation of treatment positivity over $\langle \boldsymbol{L}, \boldsymbol{U} \rangle$, with additional parametric assumptions:

P1 $\mathbb{E}[Y^a | A, Z, L, U] = \beta_0 + \beta_a A + \beta_l L + \beta_u U$;

P2 $\mathbb{E}[W | A, Z, L, U] = \alpha_0 + \alpha_l L + \alpha_u U$

Parametric estimation

P1 and **P2** can be rewritten as:

$$\begin{aligned}
\mathbb{E}[Y^a|A,Z,L,U] &= \beta_0 + \beta_a A + \beta_l L + \beta_u U \Rightarrow \\
\mathbb{E}[Y^a|A,Z,L] &= \beta_0 + \beta_a A + \beta_l L + \beta_u \,\mathbb{E}[U|A,Z,L]; \\
\mathbb{E}[W|A,Z,L] &= \alpha_0 + \alpha_u \,\mathbb{E}[U|A,Z,L] + \alpha_l L \Rightarrow \\
\mathbb{E}[U|A,Z,L] &= -\alpha_u^{-1}\alpha_0 - \alpha_u^{-1}\alpha_l L + \alpha_u^{-1}\,\mathbb{E}[W|A,Z,L]; \\
&\Rightarrow \\
\mathbb{E}[Y^a|A,Z,L] &= (\beta_0 - \beta_u\alpha_u^{-1}\alpha_0) + \beta_a A + (\beta_l - \beta_u\alpha_u^{-1}\alpha_l)L \\
&\quad + \beta_u\alpha_u^{-1}\,\mathbb{E}[W|A,Z,L] \\
&= \beta_0^* + \beta_a^* A + \beta_u^*\,\mathbb{E}[\hat{W}|A,Z,L] + \beta_l^* L
\end{aligned}$$

Parametric estimation

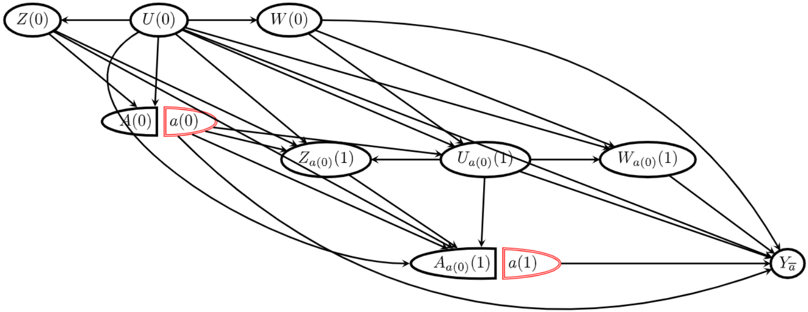
An ATE is then parametrically identifiable through a 2SLS approach [6], estimated by the coefficient β_a^* in a 2SLS regression:

$$\mathbb{E}[Y^a|A, Z, L, \hat{W}; \beta] = \beta_0^* + \beta_a^* A + \beta_u^* \mathbb{E}[\hat{W}|A, Z, L] + \beta_l^* L$$
$$\mathbb{E}[W|A, Z, L; \alpha] = \alpha_0^* + \alpha_a^* A + \alpha_z^* Z + \alpha_l^* L + \epsilon_w$$

Note: Statistical and modelling constraints/assumptions (*e.g.* distribution; the existence of MGF for error term, etc.) apply for different types of Y and W. This approach can be generalized in the presence of effect measure modification [7].

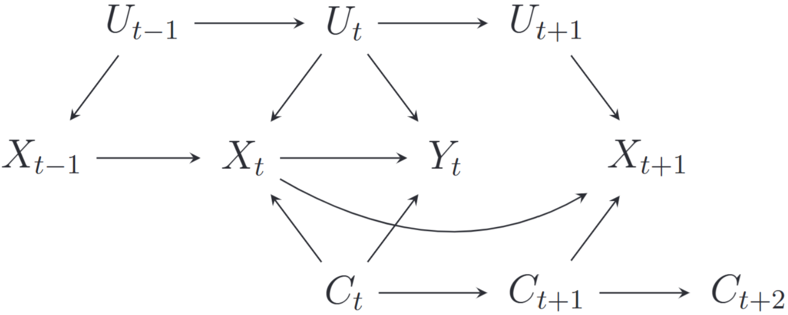
Time-varying settings

Sequential proxies can be used in a time-varying setting to allow that sequential exchangeability assumption over L not hold [7]:



Time-varying settings

Future exposures and past outcomes can serve as valid proxies [8]:



Closing and take-home message

1. PCI deals with **KU**-confounders.
2. PCI has its root in **measurement error**-based thinking and **proxy**-based approaches.
3. PCI combines **negative control** outcome and exposure proxy.
4. PCI makes the use of the residual information carried by measured proxies around the unmeasured things.
5. Nonparametric estimation of ATE via PCI is flexible and requires additional statistical assumptions (completeness); 2SLS can be used for parametric estimation.
6. Finding valid proxies is challenging. Mismeasured versions, future exposures, and past outcomes can sometimes serve as proxies.

1. Cui Y, Pu H, Shi X, Miao W, and Tchetgen Tchetgen E. Semiparametric Proximal Causal Inference. Journal of the American Statistical Association 2024;119. Publisher: ASA Website _eprint: https://doi.org/10.1080/01621459.2023.2191817:1348–59.
2. Miao W, Geng Z, and Tchetgen Tchetgen EJ. Identifying causal effects with proxy variables of an unmeasured confounder. Biometrika 2018;105:987–93.
3. Ogburn EL and Vanderweele TJ. Bias attenuation results for nondifferentially mismeasured ordinal and coarsened confounders. Biometrika 2013;100:241–8.
4. Lipsitch M, Tchetgen Tchetgen E, and Cohen T. Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. Epidemiology 2010;21:383.
5. Miao W, Shi X, Li Y, and Tchetgen Tchetgen EJ. A confounding bridge approach for double negative control inference on causal effects. Statistical Theory and Related Fields. 0. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/24754269.2024.2390748:1–12.
6. Liu J, Park C, Li K, and Tchetgen Tchetgen EJ. Regression-Based Proximal Causal Inference. American Journal of Epidemiology 2024:kwae370.
7. Tchetgen EJT, Ying A, Cui Y, Shi X, and Miao W. An Introduction to Proximal Causal Inference. Statistical Science 2024;39. Publisher: Institute of Mathematical Statistics:375–90.

8. Miao W and Tchetgen Tchetgen E. Invited Commentary: Bias Attenuation and Identification of Causal Effects With Multiple Negative Controls. American Journal of Epidemiology 2017;185:950–3.

Thanks!