Causal mediation analysis

at Social and Behavioral Sciences Branch, NICHD
Methods Seminar Series

by Trang Quynh Nguyen (trang.nguyen@jhu.edu)
Department of Mental Health, JHSPH

2022-06-08

HUGE and complicated topic!

- Lots of methods development
- Applications lag behind

At Stuart lab, mediation project (R01 MH115487) focuses on translating, adapting and disseminating causal mediation methods for applied researchers

OVERVIEW

1. A snapshot of current practice
   - review paper (Epidemiologic Reviews, 2021, doi:10.1093/epirev/mxab007)
2. A touch on estimands
   - estimands paper (Psych Methods, 2021, doi:10.1037/met0000299)
3. A tiny glimpse of identification
   - identification paper (under review, arXiv:2011.09537)
4. A glimpse of estimation methods
   - estimation paper (under review, arXiv:2102.06048)

Simple setting:

$$A \quad \ldots\ldots \quad M \quad \ldots\ldots \quad Y$$

1. A snapshot

A review of about 200 papers published in psychiatry and psychology journals in 2013-2018

(results similar to another review of analyses using randomized trials)

Temporal ordering

Mediation analysis is about causal effects
but only 1/4 of the papers had appropriate A-M-Y temporal ordering

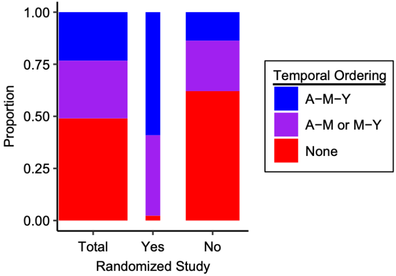


Figure 2. Proportion of studies with temporal ordering, stratified by whether exposure (A) was randomized. Horizontal bar size reflects the relative number of randomized and nonrandomized studies. M, mediator; Y, outcome.

Control for confounding

Control for confounding in mediation analysis is complicated
and this is generally not done well

adjusted for any covariates	60.7%
adjusted for covariates including baseline measure of M	13.6%
adjusted for covariates including baseline measure of Y	16.0%
adjusted for covariates including baseline measures of both M and Y	11.7%

Assumptions implicit in the analysis are often not discussed

Mediation analysis is hard

but this is under-appreciated

it's perhaps too easy to do mediation analysis
with software that "does it for you"

OVERVIEW

1. A snapshot of current practice
   - review paper (Epidemiologic Reviews, 2021, doi:10.1093/epirev/mxab007)
2. A touch on estimands
   - estimands paper (Psych Methods, 2021, doi:10.1037/met0000299)
3. A tiny glimpse of identification
   - identification paper (under review, arXiv:2011.09537)
4. A glimpse of estimation methods
   - estimation paper (under review, arXiv:2102.06048)

Simple setting:

$$A \quad \ldots\ldots \quad M \quad \ldots\ldots \quad Y$$

2. Estimands

A bit of historical view

- Original desire: understand mechanisms of effect of A on Y
  - effect through a causal pathway via an intermediate variable M
  - total effect = direct + indirect components
- With this desire
  - Effect were traditionally model-centric, eg indirect effect $= ab$, where a, b are two regression coefs
  - Causal inference revised these effects using potential outcomes, freeing them from the models – *natural (in)direct effects*
- Causal inference brings in the idea of sequential intervention
  - Another genre of effects – *interventional effects*
  - Fit a different desire: effects of hypothetical conditions – in intervention research, disparity research

Our proposal: carefully choose the target effect (*estimand*) based on what we want to learn

the estimand should drive the analysis

clarity on the estimand leads to clarity in interpreting analysis results

3 steps of analysis:

- *define*: define the target estimand – what we want to learn
- *identify*: assess its identifiability – given study design, assumptions
- *estimate*: estimate or test it – using statistical methods

Research questions $\rightarrow$ effect definitions

Many effects and effect types

Which one best matches my research question?

May require clarifying vague research questions

Quick connection

If the research question is about explaining the causal effect of exposure on outcome, the relevant estimands are natural (in)direct effects

If the research question is about *what-if* conditions (eg modifying the intervention, manipulating the mediator distribution, etc.), want to consider the class of interventional effects

Notation and consistency

$$A \quad \dots\dots \quad M \quad \dots\dots \quad Y$$

Observed variables:	A	binary exposure (0/1)
	M	mediator
	Y	outcome
Potential variables:	M_a	$a = 0, 1$
	Y_a	
	Y_{am}	m is a mediator value
	$Y_{aM_{a'}}$	

Consistency assumptions:	if $A = a$	$M = M_a$
(connecting potential and		$Y = Y_a = Y_{aM} = Y_{aM_a}$
observed variables)	if $A = a, M = m$	$Y = Y_a = Y_{aM} = Y_{am}$
	if $M_{a'} = m$	$Y_{aM_{a'}} = Y_{am}$

Natural (in)direct effects

Research question

If the research question is about explaining the causal effect of exposure on outcome, eg

- what are the mechanisms of this effect?
- what part of this effect is due to the exposure's influence on this intermediate variable and what part is not?
- is the effect partly due to the exposure's influence on this intermediate variable?

then want natural (in)direct effects

Defined at the individual level

Natural (in)direct effects decompose individual total effect

$$\begin{aligned} TE &= Y_1 - Y_0 \\ &= Y_{1M_1} - Y_{0M_0} \end{aligned}$$

Defined at the individual level

Natural (in)direct effects decompose individual total effect

$$\begin{aligned} TE &= Y_1 - Y_0 \\ &= Y_{1M_1} - Y_{0M_0} \end{aligned}$$

2 decompositions

- direct-indirect: $TE = \underbrace{Y_{1M_1} - Y_{1M_0}}_{NIE_1} + \underbrace{Y_{1M_0} - Y_{0M_0}}_{NDE_0}$
- indirect-direct: $TE = \underbrace{Y_{1M_1} - Y_{0M_1}}_{NDE_1} + \underbrace{Y_{0M_1} - Y_{0M_0}}_{NIE_0}$

NIE = an effect on the outcome *of the exposure's effect on the mediator*

NDE = an effect of the exposure when holding the mediator at a natural value

Target: average natural (in)direct effects

- direct-indirect: $\text{TE} = \underbrace{\text{E}[Y_1] - \text{E}[Y_{1M_0}]}_{\text{NIE}_1} + \underbrace{\text{E}[Y_{1M_0}] - \text{E}[Y_0]}_{\text{NDE}_0}$
- indirect-direct: $\text{TE} = \underbrace{\text{E}[Y_1] - \text{E}[Y_{0M_1}]}_{\text{NDE}_1} + \underbrace{\text{E}[Y_{0M_1}] - \text{E}[Y_0]}_{\text{NIE}_0}$

These definitions are model free

The class of interventional effects

Research questions

If the research question is a *what-if* question, eg

- in intervention development research: what if the program is modified
    - removing elements that affect the mediator
    - retaining only elements that affect the mediator
    - some other way
- in disparities research: what if could shift the distribution of a factor that contributes to disparity

then want to consider the class of interventional effects

Interventional effects

Large class, incl. total effect, controlled and generalized direct effects, interventional (in)direct effects, and many other effects, but NOT natural (in)direct effects

An effect in this class contrasts

- a (hypothetical) active intervention condition
- a comparison (intervention or no intervention) condition

An (hypothetical) intervention condition

- sets exposure and/or mediator
  to a specific value or a distribution
  that is known or is identified (based on data in current study)
- does not change anything else

Selecting an interventional effect

2 key questions:

- Which condition best matches the *what-if* condition of scientific interest?
- What is the most appropriate comparison condition?

Note that an interventional effect

- generally does not tell us exactly about a *realistic* intervention

  BUT
- does tell us about an *ideal* intervention

- our job to judge how rough or fine the approximation is

Some examples

Controlled and generalized direct effects

$$A \quad \cdots\cdots\cdots\cdots \quad M \quad \cdots\cdots\cdots\cdots \quad Y$$

traffic safety intervention	bike helmet use	injury

In the context of new law requiring helment use

assuming 100% compliance, the effect of the intervention in the new context is a controlled direct effect:

$$\text{CDE}(100) = \text{E}[Y_{1,100}] - \text{E}[Y_{0,100}]$$

assuming compliance about 75% $\pm$ 15%, and representing this distribution by $\mathcal{M}$, the intervention's effect in the new context is a generalized direct effect:

$$\text{GDE}(\mathcal{M}) = \text{E}[Y_{1,\mathcal{M}}] - \text{E}[Y_{0,\mathcal{M}}]$$

Effect of intervention if modified to remove indirect effect elements

$C \cdots\cdots\cdots\cdots A \cdots\cdots\cdots\cdots M \cdots\cdots\cdots\cdots Y$

	safer sex intervention	access to protection	protection behavior

$$E[Y_{1,\mathcal{M}(0|C)}] - E[Y_0]$$

The active intervention condition here sets the exposure to 1, but sets the mediator to the distribution of M_0 (conditional on pre-exposure covariates)

Note this is different from setting the mediator to M_0

Effect of alternative intervention that affects treatment but not screening for depression

$C \cdots\cdots\cdots\cdots A \cdots\cdots\cdots\cdots L \cdots\cdots\cdots\cdots M \cdots\cdots\cdots\cdots Y$

C	A	L	M	Y
	intervention with providers	depression screening	depression therapy	patient outcome

$$\mathrm{E}[Y_{0,L_0,\mathcal{M}(1|C,L=L_0)}] - \mathrm{E}[Y_0]$$

Interventional *(in)direct* effects

Well-known cousins of natural effects. Also called stochastic (in)direct effects

IMHO, not as relevant as some of the effects mentioned earlier

$$\mathrm{IDE}(\cdot 0) = \mathrm{E}[Y(1, \mathcal{M}(0|C))] - \mathrm{E}[Y(0, \mathcal{M}(0|C))]$$
$$\mathrm{IDE}(\cdot 1) = \mathrm{E}[Y(1, \mathcal{M}(1|C))] - \mathrm{E}[Y(0, \mathcal{M}(1|C))]$$

$$\mathrm{IIE}(0\cdot) = \mathrm{E}[Y(0, \mathcal{M}(1|C))] - \mathrm{E}[Y(0, \mathcal{M}(0|C))]$$
$$\mathrm{IIE}(1\cdot) = \mathrm{E}[Y(1, \mathcal{M}(1|C))] - \mathrm{E}[Y(1, \mathcal{M}(0|C))]$$

In special case with no intermediate confounders, equal to natural (in)direct effects

What if could reduce the frequency of traffic stops of Black folks down to half-way between their actual experience and that of non-Black folks

Another example to show the flexibility of defining effects based on research question

A	$\cdots\ \cdots\ \cdots\ \cdots$	M	$\cdots\ \cdots\ \cdots\ \cdots$	Y
being seen as Black		traffic stops		survival

$$\text{E}[Y_{1,\mathcal{M}(0.5|C)} \mid A=1] - \text{E}[Y_1 \mid A=1]$$

$\mathcal{M}(0.5|C)$ is a half-half mixture of two distributions

To sum up

Wide range of effect definitions

- natural (in)direct effects
- very broad class of interventional effects

Flexibility in selecting/defining effects to match research questions

(For references, see the estimands paper)

OVERVIEW

1. A snapshot of current practice
   - review paper (Epidemiologic Reviews, 2021, doi:10.1093/epirev/mxab007)
2. A touch on estimands
   - estimands paper (Psych Methods, 2021, doi:10.1037/met0000299)
3. A tiny glimpse of identification
   - identification paper (under review, arXiv:2011.09537)
4. A glimpse of estimation methods
   - estimation paper (under review, arXiv:2102.06048)

3. Identification

Can we learn the effect from data?

Fundamental problem of causal inference: do not observe two potential outcomes on same individual

Identify = connect the causal contrast to the observed data distribution using assumptions

Identification gives us the license to estimate the effect

Key questions:

- what assumptions are required?
- are they plausible in this study?

Three types of assumptions

- Consistency/SUTVA
- (Conditional) Exchangeability/ignorability/unconfoundedness
- Positivity/overlap

Varies depending on the estimand

Let's just consider the second assumption

Exchangeability/ignorability/unconfoundedness

- rough quick answers for several estimands
- more precise, and assumptions for flexibly defined estimands (see the identification paper)

Rough answers

1. no unmeasured A-Y confounders
2. no unmeasured A-M confounders
3. no unmeasured M-Y confounders
4. no (observed or unobserved) variables influenced by A that confound M-Y, aka no post-treatment confounders

Assumptions required:

- TE: 1
- CDE, GDE and other interventional effects where interventional mediator distributionn $\mathcal{M}$ is known: 1, 3
- interventional effects where $\mathcal{M}$ is defined based on a potential mediator distribution: 1, 2, 3
- natural (in)direct effects: 1, 2, 3, 4

e.g., natural (in)direct effects

no L case: okay

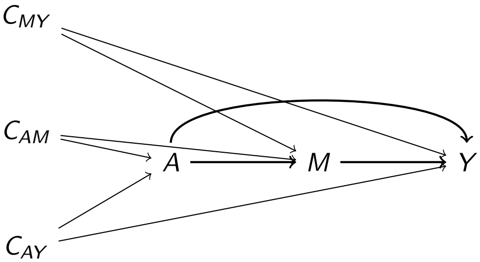
$$\begin{tikzcd}
C \arrow[r] \arrow[rr, bend left] \arrow[rrr, bend left] & A \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y
\end{tikzcd}$$

with L case: natural (in)direct effects are NOT identified

$$\begin{tikzcd}
C \arrow[r] \arrow[rr, bend left] \arrow[rrr, bend left] \arrow[rrrr, bend left] & A \arrow[r] \arrow[rr, bend left] \arrow[rrr, bend left] & L \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y
\end{tikzcd}$$

e.g., natural (in)direct effects

C is the collection of (overlapping) sets of variables



Precise system (see paper)

Relies on five types of potential outcomes

- Y_a
- Y_{am}
- $Y_{a\mathcal{M}}$ where $\mathcal{M}$ is known
- $Y_{a\mathcal{M}}$ where $\mathcal{M}$ is defined based on a potential mediator distribution
- $Y_{aM_{a'}}$

Assemble required assumptions for any flexibly defined estimand (use Table 1)

Application to different effects in an example

OVERVIEW

1. A snapshot of current practice
   - review paper (Epidemiologic Reviews, 2021, doi:10.1093/epirev/mxab007)
2. A touch on estimands
   - estimands paper (Psych Methods, 2021, doi:10.1037/met0000299)
3. A tiny glimpse of identification
   - identification paper (under review, arXiv:2011.09537)
4. A glimpse of estimation methods
   - estimation paper (under review, arXiv:2102.06048)

4. Estimation

(If effect is identified) How do we learn it from data?

Many methods, and many method papers quite technical

We use two ingredients that are familiar

- weighting
- regression

treat the identification result/estimation task as a puzzle

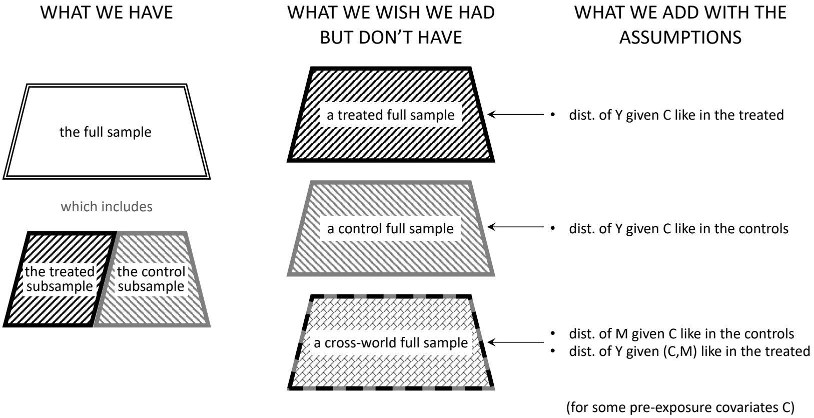
- seek solutions using the two tools
- use visualization to build intuition

consider simple and more complex solutions

- simple – nonrobust
- more complex (combining tools) – more robust to model misspecification

For now, consider natural (in)direct effects only

ID result/estimation task as a puzzle



Tool 1: weighting

Form relevant pseudo samples!

then average outcome on pseudo samples (or combine with tool 2)

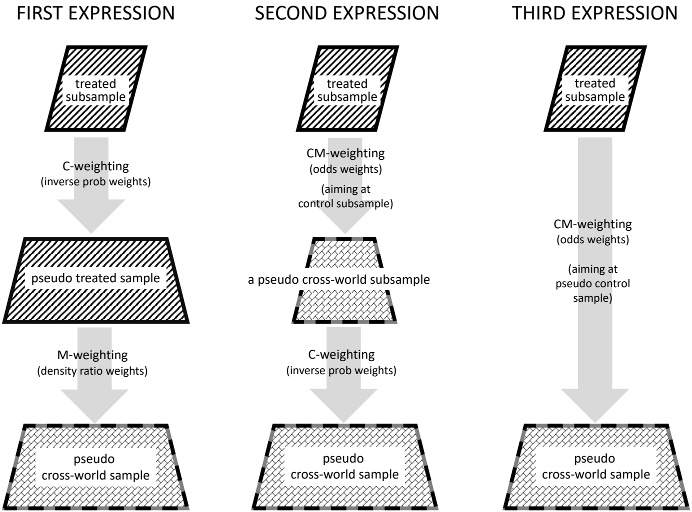
- pseudo treated sample and pseudo control sample

  inverse probability weighting: $\frac{1}{\mathrm{P}(A=1|C)}$ for treated units, $\frac{1}{\mathrm{P}(A=0|C)}$ for control units

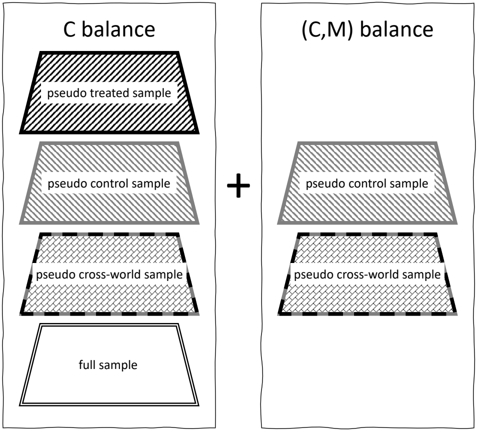
- pseudo cross-world sample
  - use treated units so the $Y \mid C, M$ dist. is like that of the treated
  - weight to mimic C dist of full sample and $M \mid C$ dist of control units

  there are different ways to estimate these weights based on different expressions

view of 3 expressions of the cross-world weight



desired balance



Tool 2: regression

specifically, regression-based prediction (or simulation)

can be used alone or combined with weighting

some combinations induce robustness

With these tools, can build pairs of estimators

A simpler estimator

- solves the puzzle
- requires all modeling components to be consistent

A more complex estimator

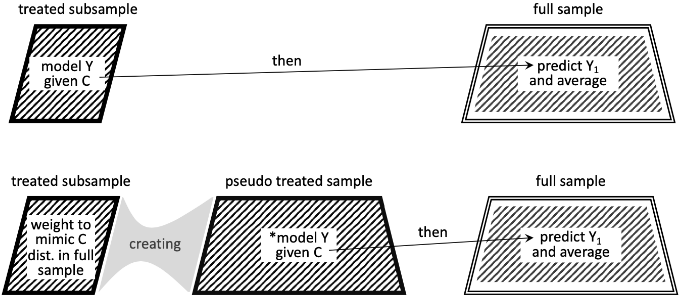
- key: replace all subsamples used to fit models with relevant pseudo samples

  (ie fit model to predictors spece where model is used for prediction)
- also: require regression model used for prediction to satisfy mean recovery

  (even if predictions are wrong, they will be right on average, if weights correct)
- more robust, ie ok if one of two components (weights or regression) correct

A few of these estimators have appeared in the literature or are related to existing estimators (see references in paper)

Simple case: estimating $E[Y_1]$



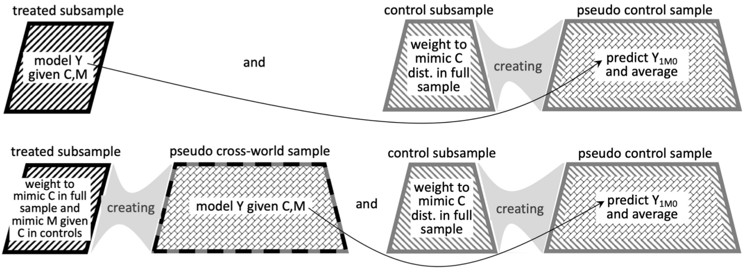
Estimating $\mathrm{E}[Y_{1M_0}]$

multiple solutions, with different properties

next slides visualize several pairs

and note which modeling components are required to be correct for consistency

Estimating E$[Y_{1M_0}]$: "outcome imputation" method



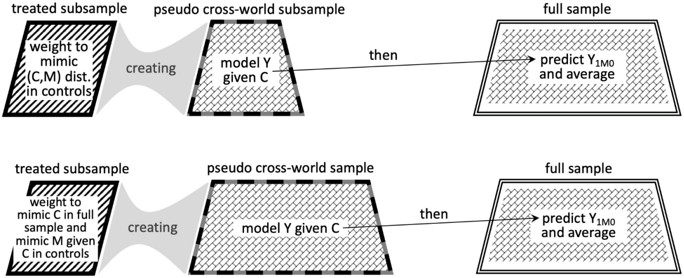
simpler version

- control weights
- outcome regression

more robust version

- control weights
- either outcome regression or cross-world weights

Estimating $E[Y_{1M_0}]$



simpler version

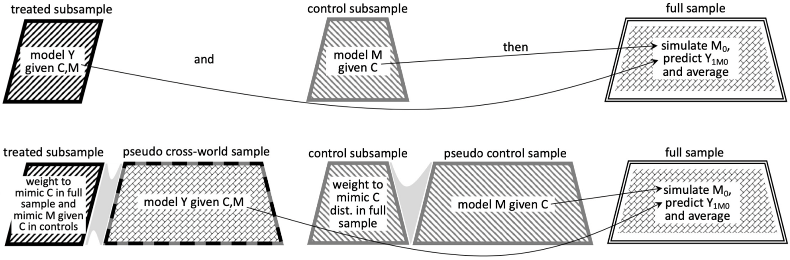
- the weights
- outcome model

more robust version

either the weights only

or the M-part of the weights +
outcome model

Estimating $E[Y_{1M_0}]$: mediator simulation method



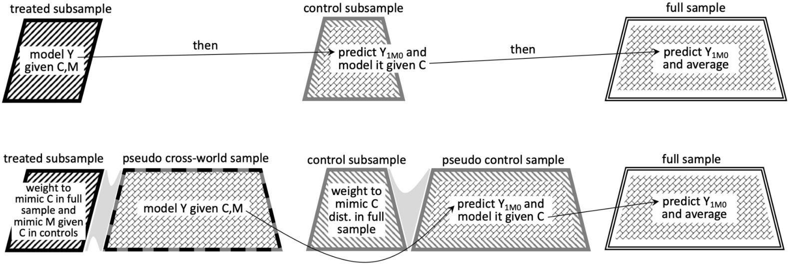
simpler version

- mediator density
- outcome regression

more robust version

- mediator density
- either outcome model or cross-world weights

Estimating $E[Y_{1M_0}]$: iterated regression method



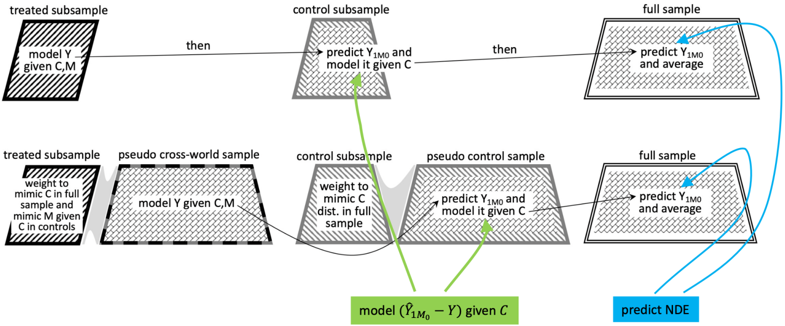
simpler version

- two regression models

robust version

- either first model or cross-world weights
- either second model or control weights

If target marginal additive effects, can modify last pair to estimate NDE_0



R-package mediationClarity

https://github.com/trangnguyen74/mediationClarity

OVERVIEW

1. A snapshot of current practice
   - review paper (Epidemiologic Reviews, 2021, doi:10.1093/epirev/mxab007)
2. A touch on estimands
   - estimands paper (Psych Methods, 2021, doi:10.1037/met0000299)
3. A tiny glimpse of identification
   - identification paper (under review, arXiv:2011.09537)
4. A glimpse of estimation methods
   - estimation paper (under review, arXiv:2102.06048)

Lots of great references (see citations in our papers)