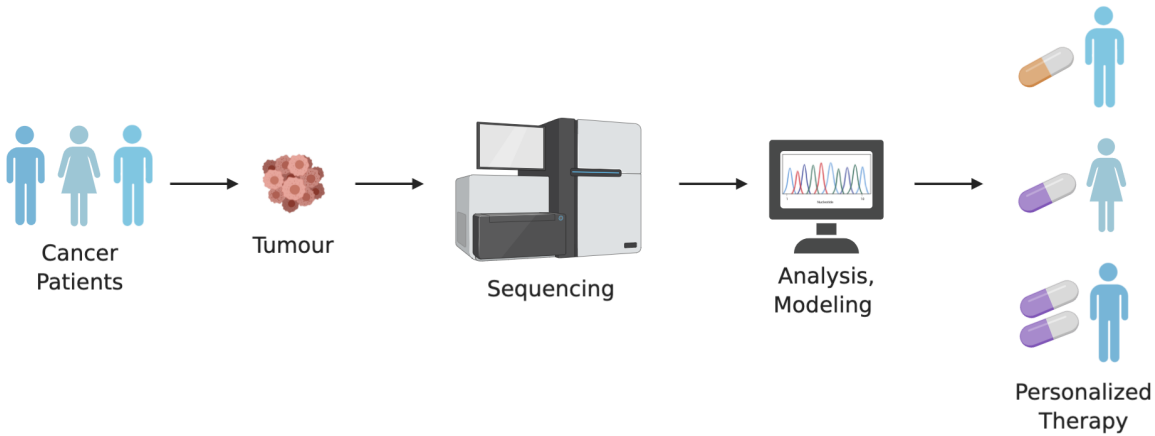
Section 33

Lecture 12

Precision medicine is a buzz word



My claim:

Modelling the disease process is of secondary importance in precision medicine, except when it helps support the identification (and estimation) of optimal regimes.

Precision medicine is a buzz word, and the idea is simple

- The idea is to tailor treatment decisions to patient characteristics.
- The premise: *individual heterogeneity* can be leveraged to *individualize therapy*.
- Work on causal inference gives us theory for *optimizing* individual decisions.
  - What if patient i receives treatment A vs. treatment B?
    That is, what is the causal effect of taking A vs. B...

Algorithmic vs. human decisions

- Decision rules might be algorithmically individualized.
- Yet these rules will be implemented under supervision of humans (e.g., doctors).
- Are optimal algorithmic regimes better than human-decision rules?
  - Care providers may have information that is not recorded in the observed data.
    $\Longrightarrow$ unmeasured confounding in the data.
  - So, when should we let humans override algorithmic treatment recommendations?

...but causal inference requires strong assumptions, no?

- We need to take the causal question seriously.
  Scientists who choose not to give up causal inference must understand that, without selecting a definition of a causal effect, it is impossible to evaluate whether we have reasonably estimated one.
- Can we deal with unmeasured confounding?
  - Sometimes we can point identify effects in the presence of unmeasured confounding.
    Instrumental variables, front-door variables, negative controls (proximal inference) ...
  - Other times we can bound the causal effects.

  Transparency about study goals and the assumptions we make to justify an analysis are required to discuss bias, refine our questions and improve our answers.

Section 34

Unmeasured confounding and instrumental variables

We have derived results under identification assumptions, but what do we do when these assumptions are violated?

- We have relied on the key assumptions that we have:
  - measured a sufficient set of variables to adjust for confounding, and
  - we have avoided selection bias.
- If these assumptions are incorrect, our estimation strategies will yield bias.
- Now we will discuss *alternative* strategies that can validly estimate causal effects under an alternative set of assumptions that do not require that our conventional exchangeability conditions hold.
- Our first example is instrumental variable (IV) methods.
  - Instrumental variables are *very* popular in economics and the social sciences. Angrist, Imbens and Card were awarded the 2021 Nobel medal in Economics for their work on instrumental variables.

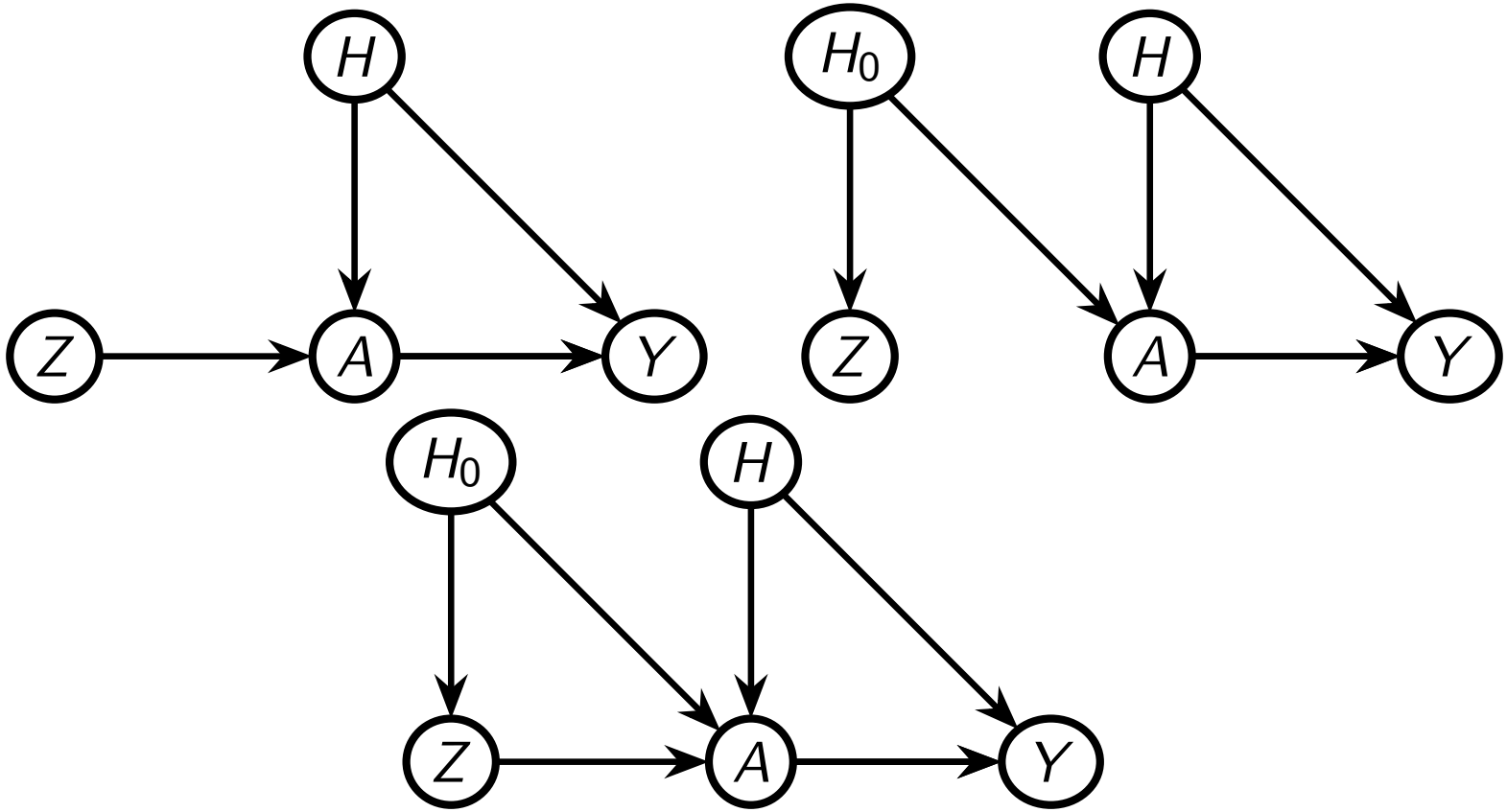
Definition of IVs

Definition (Main IV assumptions)

1. $\text{cor}(Z, A) \neq 0$ (instrument strength)
2. $Y^{z,a} = Y^a$ for all a, z (exclusion restriction)
3. $Z \perp\!\!\!\perp Y^a$ for all a (unconfoundedness of Z)

- The main IV assumptions are not themselves sufficient to identify the average effect of A on Y; thus, we need additional assumptions, and people have suggested several different ones; these different conditions are often called *homogeneity conditions*.
- If the unconfoundedness assumption holds, then there are no common causes of Z and Y in the DAG.

3 graphs that satisfy the main IV assumptions



In the first graph, we have a *causal instrument*, in the second graph we have a *proxy instrument* and the third graph is a combination.

Example of IV studies

- Several studies in economics.
  - A seminal example is on the effect of education on future earnings by Joshua D Angrist and Alan B Krueger. "Estimating the payoff to schooling using the Vietnam-era draft lottery". In: *NBER working paper* w4067 (1992);
    It seems very difficult to adjust for common causes of education and future earnings, but the authors used the result of a lottery that determined priority for the US military during the Vietnam war.
- Randomised controlled trials when treatment is blinded.
- Some non-blinded studies. For example, American economist gave families vouchers Z to reduce the costs from moving from a neighbourhood with high poverty to a neighbourhood with low poverty. A denotes moving. Y is psychological stress.
- Mendelian Randomization
  - A genetic variant Z that is associated with treatment A and is not associated with the outcome Y. outside of A.
- Applied researchers use them in a range of other settings too.

More on examples

- In our smoking cessation study: The price of cigarettes in the population could be an instrument if:
  - Cigarette price affects the decision to quit smoking,
  - cigarette price affects weight change only through its effect on smoking cessation, and
  - no common causes of cigarette price and weight change exist.

Hernan and Robins, *Causal inference: What if?*

Additional IV assumptions: linear structural equation model

Suppose that the structural equation for Y is linear,

Definition (Linear SEM)

$$Y = f_y(A, H, \epsilon_Y) = \beta A + g(H, \epsilon_Y).$$

Clearly, this linear structure is stronger than what we have previously imposed when we have done identification (think about all the things we did on non-parametric structural equations, DAGs and SWIGs).
However, some disciplines have almost only considered linear models. And there is a lot of disagreement about whether these assumptions are justified.
We will leave these issues aside for a moment (but we will get back to them), and notice that if follows from the linear SEM that

$$Y^{a'} - Y^{a} = \beta(a' - a).$$

IV theorem

Theorem (IV theorem 1)

If $cor(Z, A) \neq 0$, $\mathbb{E}(Y^{a=0} \mid Z) = \mathbb{E}(Y^{a=0})$, *and the linear SEM hold, then*

$$\psi \equiv \frac{cov(Y, Z)}{cov(A, Z)} = \beta$$

The first two assumptions are implied by the main IV assumptions
We will call ψ the IV functional.

Proof of IV theorem 1

Proof.

Under the linear SEM,

$$Y - Y^{a=0} = \beta A + g(H, \epsilon_Y) - g(H, \epsilon_Y) = \beta A, \ \textit{i.e.} \ Y - \beta A = Y^{a=0}$$

Thus, $\mathbb{E}(Y - \beta A \mid Z) = \mathbb{E}(Y^{a=0} \mid Z) = \mathbb{E}(Y^{a=0})$, that is, $\mathbb{E}(Y - \beta A \mid Z)$ is independent of Z. Thus,

$$\begin{aligned}
&\mathbb{E}[\{Z - \mathbb{E}(Z)\}(Y - \beta A)] \\
=&\mathbb{E}[\{Z - \mathbb{E}(Z)\}\mathbb{E}\{(Y - \beta A) \mid Z\}] \text{ iterative expectation} \\
=&\mathbb{E}[\{Z - \mathbb{E}(Z)\}\mathbb{E}\{Y - \beta A\}] \text{ the independence above} \\
=&\mathbb{E}\{Z - \mathbb{E}(Z)\}\mathbb{E}\{Y - \beta A\}, \text{ the independence above} \\
=&0,
\end{aligned}$$

and therefore $\text{cov}(Y - \beta A, Z) = 0$ (you'll see this by using the definition of covariance), and $0 = \text{cov}(Y - \beta A, Z) = \text{cov}(Y, Z) - \beta \text{cov}(A, Z)$. □

An intuitive interpretation

In the "causal instrument" graph, the IV functional has an intuitive interpretation:

- Consider the coefficient of the population least squares of a dependent variable W on $(1, Z)$. This coefficient say β_{WZ}, is indeed $\beta_{WZ} = \frac{\text{cov}(W,Z)}{\text{var}(Z)}$.
  That is, we can think about β_{WZ} as the limit in probability of the least squares coefficient in the regression model $\mathbb{E}(W \mid Z; \alpha, \beta) = \alpha + \beta Z$.
- By dividing the definition of ψ by $\text{var}(Z)$ in the numerator and denominator it follows that

$$\psi = \frac{\beta_{YZ}}{\beta_{AZ}}.$$

- If the instrument $Z \in \{0, 1\}$ then

$$\psi = \frac{\mathbb{E}(Y \mid Z = 1) - \mathbb{E}(Y \mid Z = 0)}{\mathbb{E}(A \mid Z = 1) - \mathbb{E}(A \mid Z = 0)},$$

  in words, the average additive effect of Z on Y divided by average effect of Z on A in our IV graph.

The problem with the linear SEM

- The linear SEM is a very strong restriction. Essentially, we are saying that all individuals have the same effect of the treatment, which is very unlikely.
  - In fact, the key idea of "personalised medicine" is that different people respond different to treatments.
  - In the smoking cessation example, this assumption would only hold if smoking cessation made every individual in the population gain (or lose) the same amount of body weight!
- In the homework you will show that ψ has a causal interpretation even under a relaxation of the linear SEM assumption, where

$$Y = f_y(A, H, \epsilon_Y) = h(\epsilon_Y)A + g(H, \epsilon_Y),$$

where h and g are unspecified functions. However, now we've made the strong assumption that H does not modify the causal effect of A (on the additive scale) because h does not have H as argument.

A further relaxation (we will not study this one in detail)

Definition (Robins's IV assumptions)

1. $\text{cor}(Z, A) \neq 0$ (instrument strength) (as before)
2. $Y^{z,a} = Y^a$ for all a, z (exclusion restriction) (as before)
3. $Z \perp\!\!\!\perp Y^{a=0}$ (unconfoundedness of Z) (Slightly weaker)
4. There exists a β such that

$$\mathbb{E}(Y \mid Z, A) - \mathbb{E}(Y^{a=0} \mid Z, A) = \beta A$$

On Robins's IV assumptions

Theorem

Under Robins's IV assumptions, $\beta = \psi$

Proof.

Using assumption 4,

$$\mathbb{E}(Y - \beta A \mid Z, A) = \mathbb{E}(Y^{a=0} \mid Z, A),$$

and thus we integrate out A,

$$\mathbb{E}(Y - \beta A \mid Z) = \mathbb{E}(Y^{a=0} \mid Z),$$

and follow the steps in the proof of the first theorem on IVs. □

*Robins's IV assumptions (continuation)

- The last assumption is an example of a so-called structural nested model,

$$\mathbb{E}(Y \mid Z, A) - \mathbb{E}(Y^{a=0} \mid Z, A) = h(A, Z; \beta)$$

  satisfying $h(0, Z; \beta) = 0$ for all β.

- The model from the previous slide does not assume effect homogeneity, but it does (only) assume no effect modification by Z on the additive scale.

- However, β does not (without extra assumptions) have the interpretation as the average (additive) treatment effect, but when A is binary it quantifies an average treatment effect of the treated (use consistency to prove this),

$$\mathbb{E}(Y^{a=1} \mid Z, A = 1) - \mathbb{E}(Y^{a=0} \mid Z, A = 1) = \mathbb{E}(Y^{a=1} - Y^{a=0} \mid A = 1) = \beta.$$

- Does this model generalize the linear structural equation model? Yes.

But is it a plausible assumption that we can reason about?

- How can a scientist (or other expert) argue in support of a constant average causal effect within levels of the proposed instrument Z and the treatment A in any particular study? Hernan and Robins, *Causal inference: What if?*
- Yet another possibility is to assume that, for any level of the unmeasured variable H, the effect of A on Y is the same, i.e.

$$\mathbb{E}(Y^{a=1} \mid H) - \mathbb{E}(Y^{a=0} \mid H) = \mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}),$$

  but this assumption is not plausible either because the unmeasured variables can often be effect modifiers.

  - For example, weight gain after smoking cessation can vary with prior intensity of smoking, which may itself be an unmeasured confounder for the effect of smoking cessation on weight gain.[42]

[42] Hernan and Robins, *Causal inference: What if?*

Another common alternative

The final criterion we will study is not a criterion about homogeneity.

Definition (Imbens and Angrist's IV assumptions)

1. $\text{cor}(Z, A) \neq 0$ (valid instrument)
2. $Y^{z,a} = Y^a$ for all a, z (exclusion restriction)
3. $Z \perp\!\!\!\perp Y^a$ and $Z \perp\!\!\!\perp A^z$ (strong unconfoundedness of Z)
4. $A^{z=1} \geq A^{z=0}$ (Monotonicity)

- These assumptions are often used in practice
- Note that the 3rd assumption is violated in two of our example graphs
- The 4th assumption is strong but sometimes plausible. I will give some intuition why.

Intuition (Robins)

We can only can estimate the effect of treatment on those whose behavior was actually affected by the instrument, so Compliers and Defiers are the only relevant sets. If we have both, then we get mixed up. If there are only Compliers, things work out OK

More on monotonicity

Suppose $Z, A \in \{0, 1\}$. Then we can divide the population into 4 mutually exclusive groups

- $(A^{z=1} = 0, A^{z=0} = 0)$, the never-takers
- $(A^{z=1} = 1, A^{z=0} = 1)$, the always-takers
- $(A^{z=1} = 0, A^{z=0} = 1)$, the defiers
- $(A^{z=1} = 1, A^{z=0} = 0)$, the compliers

Monotonicity assumes no defiers in the entire population. That is, nobody does exactly the opposite of what they are told to do.

Definition (Local Average Treatment Effect)

The local average treatment effect in stratum $A^{z=1} = a, A^{z=0} = a'$

$$\text{LATE} = \mathbb{E}(Y^{a=1} - Y^{a=0} \mid A^{z=1} = a, A^{z=0} = a')$$

In particular, the complier average treatment effect is

$$\text{CACE} = \mathbb{E}(Y^{a=1} - Y^{a=0} \mid A^{z=1} = 1, A^{z=0} = 0)$$

Monotonicity example

As a simplified example, consider a physician who generally prefers Treatment A, but prescribes Treatment B for more physically active patients (e.g., because Treatment A is associated with risk of motor-skill impairment), and another physician who generally prefers Treatment B, but makes exceptions for patients with a family history of diabetes (e.g., because a new study suggests such patients might respond better to Treatment A). Any physically active patient with a family history of diabetes who could potentially have seen either of these providers would "defy" both preferences and thus violate the monotonicity assumption Sonja A Swanson et al. "Definition and evaluation of the monotonicity condition for preference-based instruments". In: *Epidemiology (Cambridge, Mass.)* 26.3 (2015), p. 414

Returning to our smoking example

- the compliers are those who would quit smoking when the high cigarette price is high and who would not quit smoking when the cigarette price is low. Conversely, the defiers are those who would *not* quit smoking when the high cigarette price is high and who would quit smoking when the cigarette price is low.[43]

[43] Hernan and Robins, *Causal inference: What if?*

Result on Imbens and Angrist's IV

Theorem

Under conditions 1-4 of Angrist and Imbens,

$$CACE = \psi$$

*Proof of Angrist and Imbens

Proof.

$$
\begin{aligned}
& Y^{z=1} - Y^{z=0} \\
&= Y^{z=1, A^{z=1}} - Y^{z=0, A^{z=0}} \text{ consistency} \\
&= Y^{A^{z=1}} - Y^{A^{z=0}} \text{ exclusion restriction} \\
&= (Y^{a=1} - Y^{a=0}) A^{z=1} + Y^{a=0} - \{(Y^{a=1} - Y^{a=0}) A^{z=0} + Y^{a=0}\} \text{ consistency} \\
&= (Y^{a=1} - Y^{a=0})(A^{z=1} - A^{z=0})
\end{aligned}
$$

Note that because $A^{z=1} \geq A^{z=0}$, $(A^{z=1} - A^{z=0}) \in \{0, 1\}$. Thus,

$$
\begin{aligned}
\mathbb{E}(Y^{z=1} - Y^{z=0}) &= \mathbb{E}[(Y^{a=1} - Y^{a=0})(A^{z=1} - A^{z=0})] \\
&= \mathbb{E}[(Y^{a=1} - Y^{a=0}) \mid A^{z=1} - A^{z=0} = 1] P(A^{z=1} - A^{z=0} = 1) \\
&= \mathbb{E}[(Y^{a=1} - Y^{a=0}) \mid A^{z=1} > A^{z=0}] P(A^{z=1} > A^{z=0}).
\end{aligned}
$$

Thus, $\mathbb{E}[(Y^{a=1} - Y^{a=0}) \mid A^{z=1} > A^{z=0}] = \frac{\mathbb{E}(Y^{z=1} - Y^{z=0})}{P(A^{z=1} > A^{z=0})}$. □

*Proof of Angrist and Imbens

Proof.

Furthermore, by assumption 3,

$$\mathbb{E}(Y^{z=1} - Y^{z=0}) = \mathbb{E}(Y \mid Z = 1) - \mathbb{E}(Y \mid Z = 0)$$

and

$$\begin{aligned}
P(A^{z=1} > A^{z=0}) &= P(A^{z=1} = 1, A^{z=0} = 0) \\
&= P(A^{z=1} = 1) - P(A^{z=1} = 1, A^{z=0} = 1) \text{ law of total probability} \\
&= P(A^{z=1} = 1) - P(A^{z=0} = 1) \text{ monotonicity} \\
&= P(A = 1 \mid Z = 1) - P(A = 1 \mid Z = 0) \text{ assumption 3 and consist.}
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbb{E}[(Y^{a=1} - Y^{a=0}) \mid A^{z=1} > A^{z=0}] &= \frac{\mathbb{E}(Y^{z=1} = Y^{z=0})}{P(A^{z=1} > A^{z=0})} \\
&= \frac{\mathbb{E}(Y \mid Z = 1) - \mathbb{E}(Y \mid Z = 0)}{P(A = 1 \mid Z = 1) - P(A = 1 \mid Z = 0)}
\end{aligned}$$

IV estimation

"IV estimation requires modeling assumptions (such as monotonicity) even if infinite data were available. This is not the case for previous methods like IP weighting or standardization: If we had treatment, outcome, and confounder data from all individuals in the superpopulation"[44].

[44] Hernan and Robins, *Causal inference: What if?*

The monotonicity assumption was considered to be a salvation, but...

- Hard to use the complier average treatment effect by decision makers, because it only tells us something about a subset of the population. Suppose, for example, 10% of the population are compliers. Then,
  - can we justify to make recommendations based on the CATE to everyone in the population?
  - Unfortunately, we cannot observe the compliers, so we cannot target the intervention to the compliers.
  - What is the right thing to do if the treatment is not beneficial in always-takers and never-takers?
  - I agree with Hernan & Robins that it is often better to be more honest and accept that "interest in this estimand is not the result of its practical relevance, but rather of the (often erroneous) perception that it is easy to identify..."

Angus Deaton

"This goes beyond the old story of looking for an object where the light is strong enough to see; rather, we have control over the light, but choose to let it fall where it may and then proclaim that whatever it illuminates is what we were looking for all along."
Angus Deaton

In practice

- "Second, relatively minor violations of conditions (i)-(iv) for IV estimation may result in large biases of unpredictable or counterintuitive direction. The foundation of IV estimation is that the denominator blows up the numerator. Therefore, when the conditions do not hold perfectly or the instrument is weak, there is potential for explosive bias in either direction."
- "As a result, an IV estimate may often be more biased than an unadjusted estimate. In contrast, previous methods tend to result in slightly biased estimates when their identifiability conditions are only slightly violated, and adjustment is less likely to introduce a large bias. The exquisite sensitivity of IV estimates to departures from its identifiability conditions makes the method especially dangerous"

Hernan and Robins, *Causal inference: What if?*

More positively

- "IV estimation is better reserved for settings with lots of unmeasured confounding, a truly dichotomous and time-fixed treatment A, a strong and causal proposed instrument Z, and in which either effect homogeneity is expected to hold, or one is genuinely interested in the effect in the compliers and monotonicity is expected to hold."[45]
- Causal inference relies on transparency of assumptions and results from analyses that rely on different assumptions. In that sense, IV is an attractive approach because it depends on a different set of assumptions than other methods.

[45] Hernan and Robins, *Causal inference: What if?*