Causal Inference: Identification of Causal Effects (Missing Data & Randomized Experiment)

Hyunseung Kang

2025-01-28

Concepts Covered Today

- Three causal identifying assumptions: (a) SUTVA/consistency, (b) strong ignorability, and (c) positivity/overlap.
- Motivation of assumptions based on missing data (MCAR, MAR)
- Motivation of assumptions based ona randomized experiment
- Covariance balance
- References
  - Chapter 2 of Hernán and Robins (2020)

Review from Last Week I

We are interested in the causal effect of a treatment (versus no treatment/control) on an outcome.

We used the counterfactual/potential outcomes to define causal effects.

- $Y(1)$: the **counterfactual** outcome if, contrary to fact, the study unit was treated.
- $Y(0)$: the **counterfactual** outcome if, contrary to fact, the study unit was not treated (i.e., control)

	$Y(1)$	$Y(0)$
John	0.54	0.94
Sally	0.91	0.91
Kate	0.81	0.60
Jason	0.60	0.84

Review from Last Week II

From the fundamental problem of causal inference, we generally cannot observe both $Y(1), Y(0)$.

Also, the counterfactual outcomes differ from the **observed** outcomes Y in that the observed outcome is a realization for a particular value of the treatment assignment A.

	Y	A
John	0.94	0
Sally	0.91	0
Kate	0.81	1
Jason	0.60	1

Review from Last Week III

How do we learn about $Y(1), Y(0)$ from the observable data Y, A?

Our First Assumption for Causal Identification: SUTVA/Causal Consistency I

First, let's make the following assumption known as **stable unit treatment value assumption (SUTVA)** (Rubin (1980)) or **causal consistency** (page 4 of Hernán and Robins (2020)):

$$Y = AY(1) + (1 - A)Y(0).$$

It's also common to rewrite the above assumption as

$$Y = Y(A), \quad \text{or if } A = a, \text{ then } Y = Y(a).$$

The latter version covers the case when the treatment A is not binary (e.g., discrete, continuous)

Our First Assumption for Causal Identification: SUTVA/Causal Consistency II

In words, SUTVA states the observed outcome Y is one realization of the counterfactual outcomes $Y(a)$ based on the observed value of the treatment A.

More subtly, SUTVA implies two "mini" assumptions.

- There are **no multiple versions of treatment.**
- There is **no interference**, a term coined by Cox (1958).

No Multiple Versions of Treatment I

It's useful to understand this assumption by studying the case when the assumption is violated.

Let's go back to the smoking example from the first lecture where we defined the causal effect of daily smoking (i.e., treatment) versus never smoking (i.e., control) on lung function.

Daily smoking may include different type of daily smokers.

a. Is a daily smoker a person who smokes at least one cigarette per day?
b. Is a daily smoker a person who smokes at least one pack of cigarettes per day?
c. Is a daily smoker a person who smokes during all time in their lives, including during pregnancy?
d. Is a daily smoker a person who vapes every day?
e. ...

No Multiple Versions of Treatment II

We can define counterfactual outcomes for all types of daily smokers:

- $Y(1)$: counterfactual outcome under definition (a) of a daily smoker
- $Y(2)$: counterfactual outcome under definition (b) of a daily smoker
- ...
- $Y(k)$: counterfactual outcome under definition (k) of a daily smoker

By assuming SUTVA, we eliminate these variations in the counterfactuals. Formally,

$$Y(1) = Y(2) = ... = Y(k).$$

No Multiple Versions of Treatment III

In words, SUTVA implies that the lung function of a daily smoker who smokes at least one cigarette per day (i.e., $Y(1)$) is equal to the lung function of the same daily smoker living in the same environment except that he/she smokes at least one pack of cigarettes per day (i.e., $Y(2)$).

- No multiple versions of treatment assumption *does not* imply that the counterfactual outcome of the control, $Y(0)$, is equal to the counterfactual outcome for the treatment , i.e., $Y(1) = Y(0)$.
- But, the assumption implies that there are *no multiple versions of control*.
  - In the data example, the counterfactual outcomes of different types of "never-smokers" are identical.
  - For example, a never-smoke could be someone who never smoked since birth or someone who only smokes "rarely."
  - Formally, $Y(0) = Y(0') = Y(0'')$ where $0, 0', 0''$ represent different types of never-smokers.

No Multiple Versions of Treatment IV

There are settings where no multiple versions of treatment is plausible. For example,

- The casual effect of taking Wegovy/semaglutides weight loss drug (i.e., treatment). Taking this drug is a well-define intervention. Similarly, not taking the drug (i.e., control) is also a well-defined intervention.
- In general, **randomized controlled trials (RCTs)** usually have unambiguous definitions of treatment and control.
- The causal effect of increasing graduate student's stipends in Fall 2024 (i.e., the treatment). The increase is a well-defined policy (e.g., 5% increase).

Broadly speaking, SUTVA forces you to define meaningful $Y(a)$; see the first lecture.

- Some authors restrict counterfactual outcomes to be based on well-defined interventions or "no causation without manipulation"

No Multiple Versions of Treatment V

- See Holland (1986),Hernán and Taubman (2008),Cole and Frangakis (2009), VanderWeele (2009), and first lecture notes on the "causal effect" of race.

No Interference I

It is useful to understand this assumption with a counterexample.

Suppose we want to study the causal effect of getting the varicella vaccine (i.e., chickenpox vaccine) on getting the chickenpox. Let's define the following counterfactual outcomes:

- $Y(1)$: John's counterfactual chickenpox status if John gets vaccinated.
- $Y(0)$: John's counterfactual chickenpox status if John doesn't get vaccinated.
  - If $Y(0) = 0$, John did not get the chickenpox in the universe where he's not vaccinated against it.
  - If $Y(0) = 1$, John got the chickenpox in the universe where he's not vaccinated against it.

No Interference II

If the chickenpox vaccine is 100% effective for everyone, it's likely that $Y(1) = 0$.

Now, suppose John has a sibling Sally and let's consider John's counterfactual universe where he is not vaccinated and Sally's vaccination status varies.

a. John's counterfactual chickenpox status when John is not vaccinated, but Sally is vaccinated.
b. John's counterfactual chickenpox status when John and Sally are both unvaccinated.

We can redefine the counterfactual outcomes to incorporate Sally's vaccination status.

a. $Y(0, 1)$: John's counterfactual chickenpox status where the first 0 refers to John's vaccination status (i.e., not vaccinated) and the second 1 refers to Sally's vaccination status (i.e., vaccinated)

No Interference III

b. $Y(0,0)$: John's counterfactual chickenpox status where the first 0 refers to John's vaccination status (i.e., not vaccinated) and the second 0 refers to Sally's vaccination status (i.e., not vaccinated)

SUTVA implies that John's counterfactual outcome only depends on John's vaccination status, not Sally's vaccination status. Formally,

$$Y(0,1) = Y(0,0) = Y(0)$$

- From our understanding of chickenpox and how contagious it is, no interference is an implausible assumption.
- For example, if Sally is vaccinated, John will be less likely to get the chickenpox compared to when Sally is not vaccinated.
  - We can express this as $Y(0,1) \leq Y(0,0)$
  - Remember, the study unit's outcome is 1 if the unit gets the chickenpox and 0 otherwise.

No Interference IV

- In general, no interference is unlikely to hold in vaccine studies and studies of peer/neighborhood/carryover effects.
  - Rosenbaum (2007) has a nice set of examples of when the no interference assumption is implausible.
  - There is *a lot* of ongoing work on this topic (e.g., Li and Wager (2022), Sävje, Aronow, and Hudgens (2021)).

There are settings where the no interference assumption is plausible.

- The causal effect of taking Lipitor/atorvastatin cholesterol drug (i.e., treatment) on total cholesterol levels (i.e., outcome).
  - John's cholesterol level will unlikely be affected by whether Sally takes the drug or not.
- The causal effect of enrolling in a job training program (i.e., treatment) on employment.
  - John's employment status will unlikely be affected by whether Sally enrolls in the training program or not.

Motivating the Other Assumptions for Causal Identification: A Missing Data Perspective

Once we assume SUTVA (i.e. $Y = AY(1) + (1 - A)Y(0)$), the other assumptions for causal identification can be motivated by a connection to a missing data problem.

	$Y(1)$	$Y(0)$	Y	A
John	NA	0.94	0.94	0
Sally	NA	0.91	0.91	0
Kate	0.81	NA	0.81	1
Jason	0.60	NA	0.60	1

Under SUTVA, we only see one of the two counterfactual outcomes based on A.

- A serves as the "missingness" indicator where $A = 1$ implies $Y(1)$ is observed and $A = 0$ implies $Y(0)$ is observed.
- Y is the "observed" value.

Assumption on Missingness Pattern

Suppose we are interested in the causal estimand $\mathbb{E}[Y(1)]$ (i.e. the mean of the first column).

One approach to study it is to take the average of the "complete cases" (i.e., Kate and Jason's $Y(1)$s).

- Formally, we would *identify* $\mathbb{E}[Y(1)]$ with $\mathbb{E}[Y|A=1]$, the population mean of the observed outcome Y among $A=1$.
- This approach is valid if the entries of the first column are **missing completely at random** (MCAR).
  - For each row, the missingness of $Y(1)$ depends on a missingness indicator A where the value of this indicator is based on the result of a random, independent, and identical coin flip.
  - Someone essentially had a blindfold on and randomly erased some values of $Y(1)$ values; the entries of $Y(1)$ are missing completely by chance.

See here for an introduction to missing data.

Formal Statement of MCAR

Formally, MCAR can be stated as

$$A \perp Y(1) \text{ and } 0 < \mathbb{P}(A = 1)$$

- $A \perp Y(1)$ states that missingness is independent of $Y(1)$
  i. Missingness occurs completely at random in the rows of the first column, say by a flip of a random coin.
  ii. Missingness doesn't occur more frequently for lower values of $Y(1)$; this would violate $A \perp Y(1)$.
  iii. Used in the context of causal inference, this assumption is sometimes referred to as **(complete) exchangeability** or **ignorabiltiy** or **complete randomization**
- $0 < \mathbb{P}(A = 1)$ states that you have a non-zero probability of observing some entries of the column $Y(1)$
  i. If $\mathbb{P}(A = 1) = 0$, then all entries of the column $Y(1)$ are missing and we can't learn anything about its column mean.
  ii. Used in the context of causal inference, this assumption is sometimes referred to as **positivity** or **overlap**.

Formal Proof of Causal Identification of $\mathbb{E}[Y(1)]$

Suppose SUTVA and MCAR hold:

- (A1, SUTVA): $Y = AY(1) + (1-A)Y(0)$
- (A2, Complete randomization): $A \perp Y(1)$
- (A3, Positivity): $0 < \mathbb{P}(A = 1)$

Then, we can identify the causal estimand $\mathbb{E}[Y(1)]$ by writing it as the following function of the observed data $\mathbb{E}[Y|A=1]$:

$$\begin{aligned}
\mathbb{E}[Y \mid A=1] &= \mathbb{E}[AY(1) + (1-A)Y(0) \mid A=1] && \text{(A1)} \\
&= \mathbb{E}[Y(1) \mid A=1] && \text{Algebra} \\
&= \mathbb{E}[Y(1)] && \text{(A2)}
\end{aligned}$$

(A3) is used to ensure that $\mathbb{E}[Y|A=1]$ is a well-defined quantity.

Causal Identification of the ATE

In a similar vein, to *identify* the ATE $\mathbb{E}[Y(1) - Y(0)]$, a natural approach would be to use $\mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]$.

This approach would be valid under the following variation of the MCAR assumption:

$$A \perp Y(0), Y(1), \quad 0 < \mathbb{P}(A = 1) < 1$$

- The first part states that the treatment A is independent of $Y(1), Y(0)$. This is also referred to **(complete) exchangeability**, **ignorability**, or **complete randomization** in causal inference.
- $0 < \mathbb{P}(A = 1) < 1$ states that there is a non-zero probability of observing some entries from the columns of $Y(1)$ and $Y(0)$. This is (again) referred to **positivity** or **overlap** in causal inference.

Formal Proof of Causal Identification of the ATE

Suppose SUTVA and MCAR hold:

- (A1, SUTVA): $Y = AY(1) + (1-A)Y(0)$
- (A2, Igornability): $A \perp Y(1), Y(0)$
- (A3, Positivity): $0 < \mathbb{P}(A=1) < 1$

Then, we can identify the ATE from the observed data via:

$$
\begin{aligned}
&\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0] \\
=&\mathbb{E}[AY(1) + (1-A)Y(0)|A=1] \\
&\quad - \mathbb{E}[AY(1) + (1-A)Y(0)|A=0] && \text{(A1)} \\
=&\mathbb{E}[Y(1)|A=1] - \mathbb{E}[Y(0)|A=0] && \text{Algebra} \\
=&\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] && \text{(A2)}
\end{aligned}
$$

(A3) ensures that the conditioning events in $\mathbb{E}[\cdot|A=0]$ and $\mathbb{E}[\cdot|A=1]$ are well-defined.

Interpreting the Causal Identification of the ATE I

$$\underbrace{\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0]}_{\text{Measure of Association}} = \underbrace{\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]}_{\text{Measure of Causation}}$$

This equality implies that under (A1,SUTVA), (A2, Ignorability), and (A3, Positivity), a measure of association between A and Y based on difference in means (i.e., the left-hand-side) is equal to a measure of causation based on difference in counterfactual means (i.e., the right-hand side).

- More concretely, suppose the difference in the population means of Y is 0.5 (i.e., $\mathbb{E}[Y|A=1] - \mathbb{E}[Y|A=0] = 0.5$)
- Then the difference in the means of the counterfactual outcomes is also 0.5 (i.e., $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = 0.5$)
- In other words, under (A1,SUTVA), (A2, Ignorability), and (A3, Positivity), **association can imply causation**.

Interpreting the Causal Identification of the ATE II

We can take this result a bit further by considering the setting where there is no association between A and Y.

- No association is equivalent to $A \perp Y$.
  - This implies that there is no (linear) correlation between A and Y (i.e., the population $\text{corr}(A, Y) = 0$)
  - In general, there is no dependence of any kind (linear or non-linear) between A and Y.
- An implication of $A \perp Y$ is that $\mathbb{E}[Y|A = 1] = \mathbb{E}[Y|A = 0] = \mathbb{E}[Y]$
- By the result above, if (A1,SUTVA), (A2, Ignorability), and (A3, Positivity) hold, $\mathbb{E}[Y(1)] = \mathbb{E}[Y(0)]$.
- In short, under (A1,SUTVA), (A2, Ignorability), and (A3, Positivity), **no association implies no causation.**

More broadly, as illustrated by both examples, association can imply certain causal claims **if** additional assumptions hold (e.g., (A1,SUTVA), (A2,Ignorability), and (A3, Positivity)).

Interpreting the Causal Identification of the ATE III

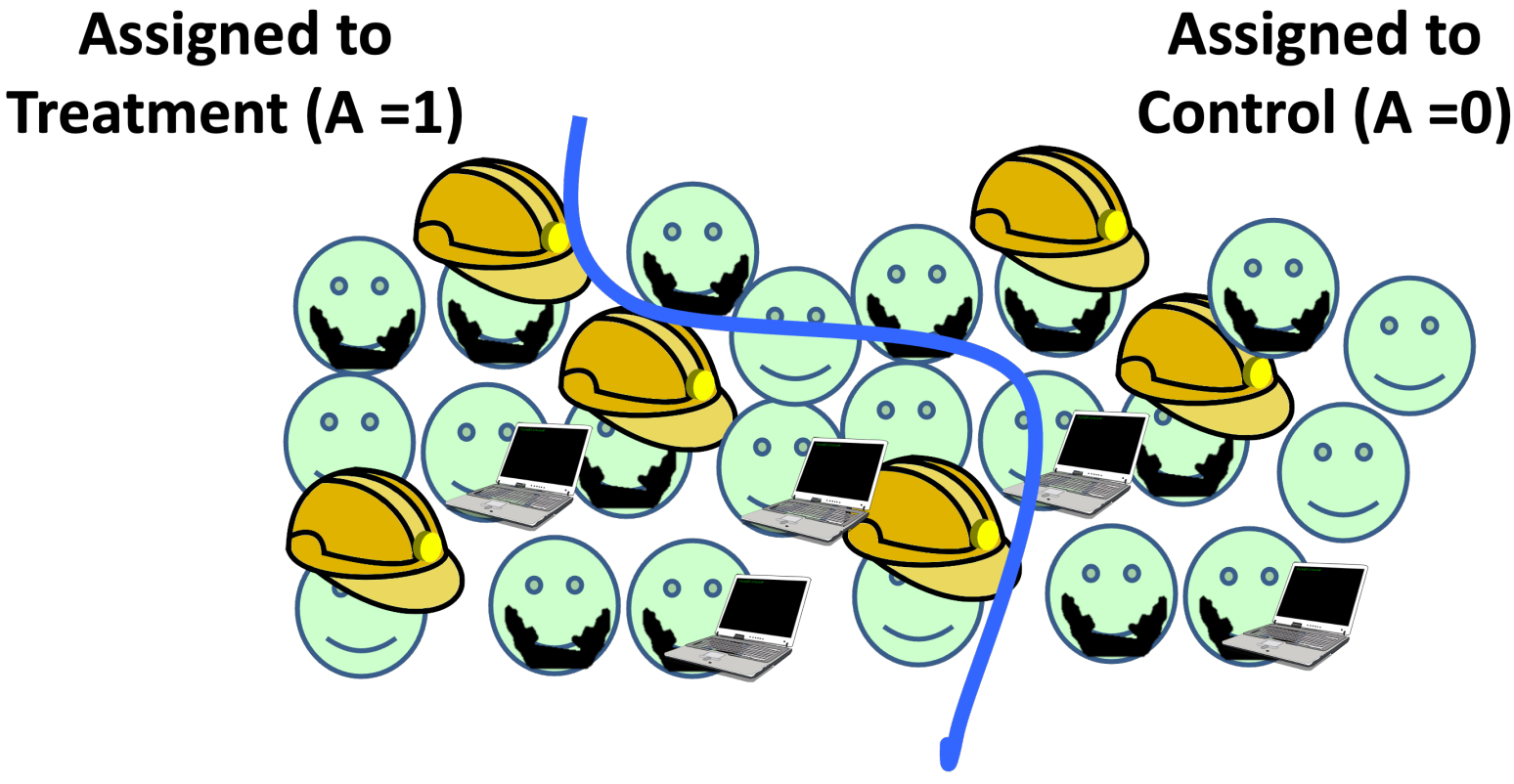
- Importantly, if SUTVA does not hold, there is no way to link the observed values A, Y to the counterfactual outcomes $Y(1), Y(0)$
- Thus, as the old saying goes, association may not imply causation unless additional assumptions hold.

Motivating the Other Assumptions for Causal Identification: A Randomized Experiment I

Consider an ideal, completely randomized trial/experiment (RCT) to assess the causal effect of a new drug (versus a control/placebo) on an outcome of interest.

1. Enroll individuals to the experiment based on some enrollment criterion.
2. Randomly assign some individuals to treatment (i.e., $A = 1$) and others to control (i.e., $A = 0$)
3. Observe outcomes Y from both groups.

Motivating the Other Assumptions for Causal Identification: A Randomized Experiment II



RCTs have been referred to as the **gold standard** to study causal effects of a treatment on an outcome of interest. But why?

- At a high level, RCTs recreates the parallel universe analogy.

Motivating the Other Assumptions for Causal Identification: A Randomized Experiment III

- Specifically, by randomization, all features about the study units are similar between the treated and the control groups.
  - The two groups are similar with respect to their measurable traits (X)
  - The two groups are also similar with respect to their unmeasurable traits (U)
- Then, any difference in the outcome between the two groups can only be attributed to a difference in the treatment status, thus recreating the parallel universe analogy from our first lecture.

This was the "big" idea from Fisher in 1935 where he used randomization as the "reasoned basis'' for causal inference. Paul Rosenbaum explains this more beautifully than I do in Chapter 2.3 of Rosenbaum (2020).

Formalizing RCTs with Counterfactual Outcomes I

Consider the following data table.

	$Y(1)$	$Y(0)$	Y	A	X (Measured; age)	U (Unmeasured; environment)
John	NA	0.94	0.94	0	23	U_{John}
Sally	NA	0.91	0.91	0	27	U_{Sally}
Kate	0.81	NA	0.81	1	32	U_{Kate}
Jason	0.60	NA	0.60	1	30	U_{Jason}

Formalizing RCTs with Counterfactual Outcomes II

If the treatment A is completely randomized (as in an RCT), we would also have $A \perp X, U$. More generally, we have

$$\text{(A2, Ignorability) } A \perp Y(1), Y(0), X, U$$

.

Also, because there is at least one control unit and treated unit in an RCT, we have

$$\text{(A3, Positivity) } 0 < \mathbb{P}(A = 1) < 1$$

.

Even with the change in (A2, Ignorability) from before, the proof to identify the ATE in an RCT remains the same as before, i.e., $\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]$. This is because only need $A \perp Y(1), Y(0)$ for identification.

Covariate Balance I

An important implication from randomization of treatment assignment is **covariate balance**.

Roughly speaking, we say that covariate X (measured or unmeasured) is "balanced'' between treated and control groups if

$$\mathbb{P}(X|A=1) = \mathbb{P}(X|A=0)$$

- In words, covariate balance states that the treated group and the control group have similar distribution of covariates.
- Suppose measured and unmeasured covariates are balanced between the treated group and the control group.
- Then on average, any difference in the outcome between the two groups can be attributed to the difference in their treatment status.

Covariate Balance II

From the RCT motivation, it's very obvious that covariate balance holds for both X and U, i.e.

$$\mathbb{P}(X, U|A = 1) = \mathbb{P}(X, U|A = 0)$$

.

In general, covariates should be balanced between treated and control groups to make causal claims about the relationship between the outcome and the treatment.

- As a result, it's common to check for covariate balance in causal inference by comparing the means of Xs among treated and control units (e.g. two-sample t-test of the mean of X).
- It's also common to do this for RCTs to verify that the randomization was done successfully.

Note About Pre-treatment Covariates

We briefly mentioned that covariates X must precede treatment assignment, i.e.

1. We collect X (i.e. baseline covariates)
2. We assign treatment/control A
3. We observe outcome Y

If they are post-treatment covariates, then the treatment can have a causal effect on both the outcome Y and the covariates X.

In this case, it's not unclear whether Y has a causal effect because of a causal effect in X. Studying this type of question is called **causal mediation analysis.**

In general, we don't want to condition on post-treatment covariates X when the goal is to estimate the average treatment effect of A on Y.

References I

Cole, Stephen R, and Constantine E Frangakis. 2009. "The Consistency Statement in Causal Inference: A Definition or an Assumption?" *Epidemiology* 20 (1): 3–5.

Cox, David. 1958. *Planning of Experiments*. Wiley.

Hernán, Miguel, and James Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.

Hernán, Miguel, and Sarah Taubman. 2008. "Does Obesity Shorten Life? The Importance of Well-Defined Interventions to Answer Causal Questions." *International Journal of Obesity* 32 (3): S8–14.

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60.

References II

Li, Shuangning, and Stefan Wager. 2022. "Random Graph Asymptotics for Treatment Effect Estimation Under Network Interference." *The Annals of Statistics* 50 (4): 2334–58.

Rosenbaum, Paul. 2007. "Interference Between Units in Randomized Experiments." *Journal of the American Statistical Association* 102 (477): 191–200.

———. 2020. *Design of Observational Studies*. Springer.

Rubin, Donald B. 1980. "Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment." *Journal of the American Statistical Association* 75 (371): 591–93.

Sävje, Fredrik, Peter Aronow, and Michael Hudgens. 2021. "Average Treatment Effects in the Presence of Unknown Interference." *Annals of Statistics* 49 (2): 673.

VanderWeele, Tyler J. 2009. "Concerning the Consistency Assumption in Causal Inference." *Epidemiology* 20 (6): 880–83.

References III