Identification Without Strong Ignorability and Unmeasured Confounding

Hyunseung Kang

Stat 992: Topics in Causal Inference
Apr. 10, 2024

Review: Strong Ignorability and Observational Studies

In the previous lecture, we identified various causal estimands under the following set of assumptions:

- (A1, SUTVA): $Y = AY(1) + (1 - A)Y(0)$
- (A2, Conditional randomization of A): $A \perp Y(1), Y(0) | X$
- (A3, Positivity/Overlap): $0 < \mathbb{P}(A = 1 | X = x) < 1$ for all x

Let $\mu_a(X) = \mathbb{E}[Y \mid A = a, X]$. Under conditions (A1)-(A3), we showed that the ATE can be identified as

$$\begin{aligned}\mathrm{ATE} &= \mathbb{E}[Y(1) - Y(0)] \\ &= \mathbb{E}[\mathbb{E}[Y \mid A = 1, X]] - \mathbb{E}[\mathbb{E}[Y \mid A = 0, X]] \\ &= \mathbb{E}[\mu_1(X)] - \mathbb{E}[\mu_0(X)].\end{aligned}$$

In observational studies, these assumptions state that the investigator measured all variables that make assumptions (A2) and (A3), also known as **strong ignorability**, plausible.

Clearly, this is a simplification of actual practice and strong ignorability is unlikely to hold in an observational study.

Some Examples of When Strong Ignorability Fails

Unmeasured confounding in an observational study: People select themselves into treatment (or control) based on what we can measure X and what we cannot measure U. Had we measured both X and U, strong ignorability would hold, i.e.

$$(\mathrm{B2}) : A \perp Y(1), Y(0) | X, U \quad \text{and} \quad (\mathrm{B3}) : 0 < \mathbb{P}(A = 1 | X = x, U = u)$$

Imperfect randomized experiment: Suppose we run a randomized experiment to study the causal effect of a new drug.

- Individuals are randomized to get the new drug or the placebo.
- After randomization, individuals may decide to not take the new drug (or the placebo)
- If we want to identify the causal effect of actually taking the new drug, **the treatment receipt** is no longer random.
- If we want to identify the causal effect of being assigned to to the new drug, **the treatment assignment** is random and (A1), (A2c), (A3c) holds if A is defined as the initial trt. assignment.

Consequences of Violating Strong Ignorability

In both examples, we no longer have
$\mathrm{ATE} = \mathbb{E}[Y(1) - Y(0)] \neq \mathbb{E}[\mu_1(X)] - \mathbb{E}[\mu_0(X)]$.

To understand the consequences of violating strong ignorability, suppose conditions (B2) and (B3) hold and let
$\mu_a(X, U) = \mathbb{E}[Y \mid A = a, X, U]$.

Then, from lecture notes, we get that the *causal bias* of identifying the ATE based on assumptions (A1), (A2), and (A3) (i.e. via $\mathbb{E}[\mu_1(X) - \mu_0(X)]$) even though in reality, assumptions (B2) and (B3) hold is

$$\underbrace{\mathrm{ATE} - \mathbb{E}[\mu_1(X) - \mu_0(X)]}_{\text{"Causal bias"}}$$
$$= \mathbb{E}[\underbrace{\{\mu_1(X, U) - \mu_0(X, U)\}}_{\text{CATE of } X \text{ and } U} - \underbrace{\{\mu_1(X) - \mu_0(X)\}}_{\text{CATE of } X}]$$

CATE stands for conditional average treatment effect that we discussed from the previous lecture.

Consequences of Violating Strong Ignorability

There are some interesting implications of this "causal bias" formula.

- If the CATE of X and U does not vary too much as a function of the unmeasured variable U, $\mu_1(X, U) - \mu_0(X, U)$ will be close to $\mu_1(X) - \mu_0(X)$ and we would have a small causal bias.
- Suppose the difference between $\mu_a(X, U)$ and $\mu_a(X)$ is at most $\Gamma \geq 0$, i.e. $|\mu_a(X, U) - \mu_a(X)| \leq \Gamma$ for all X, U, a. Then, we can get a lower and upper bound of the ATE based on only observed data:

$$\mathbb{E}[\mu_1(X) - \mu_1(X)] - 2\Gamma \leq \mathrm{ATE} \leq \mathbb{E}[\mu_1(X) - \mu_1(X)] + 2\Gamma$$

Identification Without Strong Ignorability: Instrumental Variables (IVs)

Identification Without Strong Ignorability: Instrumental Variables (IVs)

Instrumental variables (IVs) are popular approaches to identify a causal estimand when (A2) and (A3) does not hold; see Hernán and Robins (2006) and Baiocchi, Cheng, and Small (2014) for a more completely review.

Roughly speaking, an instrument relies on finding a variable Z, called an **instrument** where

- Z is related to the treatment,
- Z is independent from all unmeasured confounders that affect the outcome and the treatment, and
- Z is related to the outcome via the treatment.

Here, we discuss two approaches to making the above statements precise: (1) randomized encouragement designs (i.e. monotonicity) and (2) no additive interactions approach

Motivation: Causal Effect of Smoking During Pregnancy (Sexton and Hebel (1984), Permutt and Hebel (1989))

Sexton and Hebel (1984) wanted to study the causal effect of maternal smoking on birth weight.

Because randomizing pregnant mothers to smoking (or non-smoking) is unethical, they considered an experimental design that randomized the encouragement to pregnant mothers to quit smoking.

1. Randomly assign some mothers to an encouragement intervention (i.e. $Z = 1$) or the usual care (i.e. $Z = 0$). The encouragement intervention involved encouraging mothers to not smoke through information, support, etc. We refer to Z as the treatment assignment variable.
2. Observe their smoking status, denoted as $A \in \{0, 1\}$ where $A = 1$ denotes not smoking and $A = 0$ denotes smoking. We refer to A as treatment receipt.
3. Observe the weight of the newborn, denoted as Y.

Data

	Encouragement Group	Usual Care Group
% Smoking	15.5%	44.5%
Birth Weight	123.65	120.12
Number of subjects	1000	1000

Defining Counterfactuals

Let $A(z)$ denote the counterfactual treatment receipt under treatment variable z.

Let $Y(a, z)$ denote the counterfactual outcome under treatment variable z and treatment receipt a. In the maternal smoking example:

- $A(1)$: Counterfactual smoking status if the mother was encouraged to stop smoking
- $A(0)$: Counterfactual smoking status if the mother was not encouraged to stop smoking (i.e. usual care)
- $Y(1, 1)$: Counterfactual birthweight if the mother was encouraged to stop smoking and the mother stopped smoking
- $Y(1, 0)$: Counterfactual birthweight if the mother was under the usual care and the mother stopped smoking
- $Y(0, 1)$: Counterfactual birthweight if the mother was encouraged to stop smoking and the mother kept smoking
- $Y(0, 0)$: Counterfactual birthweight if the mother was under the usual care and the mother kept smoking

Assumptions Behind Randomized Encouragement Designs

Randomized encouragement designs satisfy the following:

- (IV1, SUTVA): $A = ZA(1) + (1 - Z)A(0)$ and $Y = ZY(A(1), 1) + (1 - Z)Y(A(0), z)$
- (IV2, Ignorable instrument): $Z \perp Y(1,1), Y(1,0), Y(0,1), Y(0,0), A(1), A(0)$
- (IV3, Overlap, positivity on instrument): $0 < P(Z = 1) < 1$

Assumption (IV1) says we observe counterfactuals that correspond to instrument/encouragement Z. Notably, we only get to observe $Y(a, z)$ that corresponds to $Z = z$ and $a = A(Z)$.

Assumption (IV2) says that the instrument/encouragement was completely randomized.

Assumption (IV3) says that all values of the instrument has a non-zero probability of being realized.

Assumptions (IV1)-(IV3) are similar to (A1)-(A3) where Z is replaced by A and we have more counterfactual outcomes.

Missing Data Perspective

We can also interpret assumptions (IV1), (IV2), and (IV3) using the data table that includes both counterfactuals and observed variables:

	$Y(1,1)$	$Y(1,0)$	$Y(0,1)$	$Y(0,0)$	$A(1)$	$A(0)$	Z	Y	A
Chloe	15	NA	NA	NA	1	NA	1	30	1
Sally	NA	NA	20	NA	0	NA	1	20	0
Kate	NA	NA	NA	18	NA	0	0	18	0
Julie	NA	25	NA	NA	NA	1	0	25	1

The variables Z and A both serve as missing indicators.

But, we only assume something about the missingness on Z via (IV2) and (IV3); we don't make any assumptions about the missingness on A.

Missing Data Perspective

	$Y(1,1)$	$Y(1,0)$	$Y(0,1)$	$Y(0,0)$	$A(1)$	$A(0)$	A	Z	Y
Chloe	15	NA	NA	NA	1	NA	1	1	30
Sally	NA	NA	20	NA	0	NA	0	1	20
Kate	NA	NA	NA	18	NA	0	0	0	18
Julie	NA	25	NA	NA	NA	1	1	0	25

- (IV2) and (IV3) say that the missingness in the columns $A(1)$ and $A(0)$ are completely at random (MCAR). This implies that identifying the causal effect of Z on A can be identified.
- But, the missingness in the columns of $Y(\cdot)$ may not be entirely random. For the missingness in all $Y(\cdot)$ columns to be random, we have to assume that $A, Z \perp Y(1,1), Y(1,0), Y(0,1), Y(0,0)$.

Conditional Versions of (IV2) and (IV3)

Assumptions (IV2) and (IV3) can have conditional counterparts that condition on X, i.e.

- (IV2c): $Z \perp Y(1,0), Y(0,1), Y(0,0), A(1), A(0) \mid X$
- (IV3c): $0 < \mathbb{P}(Z = 1 \mid X = x) < 1$ for all x

These conditional versions of (IV2) and (IV3) would be plausible if the investigator conducted a stratified randomized encouragement design where randomization of Z was done within pre-defined blocks of individuals defined by X.

This is akin to the stratified randomized experiment from the previous lecture.

IV Under Randomized Encouragement Designs

One way to formalize the definition of an instrument under a randomized encouragement design is as follows:

- (IV4, Non-zero causal effect): $\mathbb{E}[A(1) - A(0)] \neq 0$
- (IV5, Exclusion restriction): $Y(a, 1) = Y(a, 0) = Y(a)$ for all a
- (IV6, Monotonicity/No Defiers): $\mathbb{P}(A(1) \geq A(0)) = 1$

Assumption (IV4) states that the instrument has a non-zero, average effect on the treatment receipt.

In the maternal smoking example, (IV4) states that the encouragement intervention caused more mothers to quit smoking during pregnancy.

Under (IV1)-(IV3), this assumption can be re-written based on the observed data,
i.e. $\mathbb{E}[A(1) - A(0)] = \mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0] \neq 0$,

(IV5, Exclusion Restriction): $Y(a,1) = Y(a,0) = Y(a)$

(IV5) that after fixing a, the counterfactual outcomes are identical between $z = 1$ and $z' = 0$.

In the maternal smoking example, (IV5) states that after fixing the mother's smoking status, whether the mother was encouraged or not does not affect the birthweight of the newborn.

- Unlike (IV4), (IV5) cannot be written as a function of the observed data because we cannot obseve both $Y(a,1)$ and $Y(a,0)$.
- In other words, (IV5) cannot be directly assessed with the observed data, but testable implications exist
- (IV5) is the most controversial assumption in IV as the other assumptions (IV1)-(IV4) and (IV6) can be satisfied by a randomized encouragement design or be directly tested (e.g. (IV4)).

(IV6, Monotonicity) $\mathbb{P}(A(1) \geq A(0)) = 1$ and Compliance

It's useful to interpret (IV6) by partitioning individuals based on their counterfactual $A(1), A(0)$

$A(0)$	$A(1)$	Type
1	1	Always-taker
0	1	Complier
1	0	Defier
0	0	Never-taker

- Always-takers are mothers who smoke irrespective of whether they were under the encouragement intervention or not.
- Compliers are mothers who do not smoke when they were under the encouragement intervention, but smoke if they were under the usual care.
- Never-takers are mothers who never smoke irrespective of encouragement status.
- Defiers are mothers who do not smoke when they are under the usual care, but smoke when they are encouraged to not smoke.

(IV6, Monotonicity)

$A(0)$	$A(1)$	Type
1	1	Always-taker
0	1	Complier
1	0	Defier
0	0	Never-taker

Assumption (IV6) rules out the existence of defiers in the study population.

An important point from the table is that **we cannot classify everyone in the population as always-takers, compliers, and never-takers** as this requires observing both $A(1)$ and $A(0)$.

However, we can identify the column means of $A(1)$ and $A(0)$ from (IV1)-(IV3), which allows us to identify the proportion of compliance types under (IV6).

Formal Proof of Identifying Proportion of Compliance Types

Under (IV1)-(IV3), we have $\mathbb{E}[A(1)] = \mathbb{E}[A \mid Z = 1]$ and $\mathbb{E}[A(0)] = \mathbb{E}[A \mid Z = 0]$

Under (IV6) where defiers do not exist, we can identify the proportion of always-takers and never-takers as

$$\begin{aligned}
\mathbb{P}(\mathrm{Always-takers}) &= \mathbb{P}(A(0) = 1) && \text{(IV6)} \\
&= \mathbb{E}[A \mid Z = 0] && \text{(IV1)-(IV3)} \\
\mathbb{P}(\mathrm{Never-takers}) &= \mathbb{P}(A(1) = 0) && \text{(IV6)} \\
&= 1 - \mathbb{P}(A(1) = 1) && \\
&= 1 - \mathbb{E}[A \mid Z = 1] && \text{(IV1)-(IV3)}
\end{aligned}$$

We can identify the proportion of compliers as one minus the proportion of always-takers and never-takers:

$$\begin{aligned}
&\mathbb{P}(\mathrm{Compliers}) \\
=&1 - (\mathbb{P}(\mathrm{Always-takers}) + \mathbb{P}(\mathrm{Never-takers})) && \text{(IV6)} \\
=&\mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0] && \text{See above}
\end{aligned}$$

One-Sided Noncompliance

In some experimental designs, we can enforce (IV6) by blocking access to treatment for all individuals who are randomized to the control $Z = 0$, i.e.,

- (IV6.0 One-Sided Noncompliance): $A(0) = 0$

One-sided non-compliance is plausible in settings where Z represents a new program under evaluation and A represents the actual enrollment into the new program.

- In these settings, those who are not randomized into the new program (i.e $Z = 0$) usually cannot enroll into the new program.
- In contrast, those who are randomized into the new program (i.e $Z = 1$) can choose to enroll (i.e $A = 1$) or not enroll (i.e. $A = 0$) into the program.

Note that (IV6.0) implies (IV6).

Causal Estimand: The Local Average Treatment Effect (LATE)

With assumptions (IV1)-(IV6), we can identify the average treatment effect among compliers. This quantity is sometimes referred to the **local average treatment effect (LATE):**

$$\mathrm{LATE} = \mathbb{E}[Y(1) - Y(0) \mid A(1) = 1, A(0) = 0]$$

In the maternal smoking example, LATE is the average effect of smoking during pregnancy among complier mothers

- We cannot identify the different types of individuals (i.e. compliers, always-takers, and never-takers).
- In other words, LATE identifies the average treatment effect among a subgroup defined by latent classes.
- In contrast, the CATE identifies the average treatment effect among a subgroup of individuals defined by observed X.
- There is a healthy debate about whether LATE is a useful estimand or not; see lecture notes.

Proof of Identifying the LATE

We will show that

$$\mathrm{LATE} = \mathbb{E}[Y(1) - Y(0)|A(1) - A(0) = 1] = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0]}$$

First, we have

$$\begin{aligned}
\mathbb{E}[Y \mid Z = 1] &= \mathbb{E}[ZY(A(1), 1) + (1 - Z)Y(A(0), 0) \mid Z = 1] && \text{(IV1)} \\
&= \mathbb{E}[Y(A(1), 1) \mid Z = 1] \\
&= \mathbb{E}[Y(1, 1)A(1) + Y(0, 1)(1 - A(1)) \mid Z = 1] \\
&= \mathbb{E}[Y(1, 1)A(1) + Y(0, 1)(1 - A(1))] && \text{(IV2)} \\
&= \mathbb{E}[Y(1)A(1) + Y(0)(1 - A(1))] && \text{(IV5)}
\end{aligned}$$

Note that (IV3) is needed to ensure that the conditional expectation that conditions on $\{Z = 1\}$ is well-defined.

By a similar argument, we have
$\mathbb{E}[Y \mid Z = 0] = \mathbb{E}[Y(1)A(0) + Y(0)(1 - A(0))]$.

Proof of Identifying the LATE

Second, we take the difference between the two expectations

$$
\begin{aligned}
&\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0] \\
=&\mathbb{E}[\{Y(1)A(1) + Y(0)(1 - A(1))\} - \{Y(1)A(0) + Y(0)(1 - A(0))\}] \\
=&\mathbb{E}[Y(1)\{A(1) - A(0)\} - Y(0)\{A(1) - A(0)\}] \\
=&\mathbb{E}[\{Y(1) - Y(0)\}\{A(1) - A(0)\}] \\
=&\mathbb{E}[\{Y(1) - Y(0)\}I(A(1) - A(0) = 1) \\
&\quad + \{Y(1) - Y(0)\}I(A(1) - A(0) = -1)] \\
=&\mathbb{E}[Y(1) - Y(0)|A(1) - A(0) = 1]\mathbb{P}(A(1) - A(0) = 1) \quad \text{(IV6)}
\end{aligned}
$$

The last equality also uses the definition of conditional expectation.

Proof of Identifying the LATE

We can also take the difference between $\mathbb{E}[A \mid Z = 1]$ and $\mathbb{E}[A \mid Z = 0]$:

$$\begin{aligned} & \mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0] \\ =& \mathbb{E}[A(1) - A(0)] && \text{(IV1)-(IV3)} \\ =& \mathbb{P}(A(1) - A(0) = 1) && \text{(IV6)} \end{aligned}$$

Finally, under (IV4), we can take the ratio of the two differences and the denominator of this ratio is non-zero and arrive at

$$\begin{aligned} & \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0]} \\ =& \frac{\mathbb{E}[Y(1) - Y(0) | A(1) - A(0) = 1] \mathbb{P}(A(1) - A(0) = 1)}{\mathbb{P}(A(1) - A(0) = 1)} \\ =& \mathbb{E}[Y(1) - Y(0) | A(1) - A(0) = 1] \\ =& \mathrm{LATE} \end{aligned}$$

IV Defined Under No Additive Interaction Assumption

Roughly speaking, the no additive interaction framework does not necessarily assume the existence of the counterfactual $A(z)$.

Instead, I like to think of this framework as treating the instrument as a special, pre-treatment covariate Z that is endowed with the following properties.

- (JV1, Causal consistency): $Y = Y(A, Z)$
- (JV2, Exhcangeable instrument): $Z \perp Y(1,1), Y(1,0), Y(0,1), Y(0,0)$
- (JV3, Positivity): $0 < P(Z = 1) < 1$
- (JV4, Instrument relevance) $Z \not\perp A$
- (JV5, Exclusion restriction) $Y(a, 1) = Y(a, 0) = Y(a)$ for all a
- (JV6, No additive interaction) Suppose (JV5) holds. We have $\mathbb{E}[Y(1) - Y(0)|Z = 1, A = 1] = \mathbb{E}[Y(1) - Y(0)|Z = 0, A = 1]$

Assumption (JV1) and (JV2) are similar to assumptions (IV1) and (IV2), except that assumptions about the counterfactual $A(z)$ are no longer present.

Assumptions (JV1)-(JV3)

- (JV1, Causal consistency): $Y = Y(A, Z)$
- (JV2, Exhcangeable instrument):
  $Z \perp Y(1,1), Y(1,0), Y(0,1), Y(0,0)$
- (JV3, Positivity): $0 < P(Z = 1) < 1$

Assumption (JV1) and (JV2) are similar to assumptions (IV1) and (IV2), except that assumptions about the counterfactual $A(z)$ are no longer present.

Assumption (JV3) and (IV3) are identical.

Also, similar to assumptions (IV2.c) and (IV3.c), we can create conditional versions of (JV2) and (JV3), i.e.:

- (JV2.c) $Z \perp Y(1,1), Y(1,0), Y(0,1), Y(0,0) \mid X$
- (JV3.c) $0 < P(Z = 1 \mid X = x) < 1$ for all x

Assumptions (JV4) and (JV5)

- (JV4, Instrument relevance) $Z \not\perp A$
- (JV5, Exclusion restriction) $Y(a, 1) = Y(a, 0) = Y(a)$ for all a

Assumption (JV4) states that the instrument is associated with A.

In contrast to assumption (IV4), we do not necessarily need to have a causal effect of Z on A.

Assumption (JV5) is identical to (IV5).

Assumption (JV6)

- (JV6, No additive interaction) Suppose (JV5) holds. We have $\mathbb{E}[Y(1) - Y(0)|Z = 1, A = 1] = \mathbb{E}[Y(1) - Y(0)|Z = 0, A = 1]$

Assumption (JV6) can be interpreted by writing out a *saturated* model of the conditional expectation in (JV6).

$$\mathbb{E}[Y(1) - Y(0) \mid Z = z, A = 1] = \beta_0 + \beta_1 z$$

The term β_0 represents the ATT among individuals with $Z = 0$ and the term $\beta_0 + \beta_1$ represents the ATT among individuals with $Z = 1$.

Assumption (JV6) can be rewritten as

$$\mathbb{E}[Y(1)-Y(0) \mid Z = 1, A = 1]-\mathbb{E}[Y(1)-Y(0) \mid Z = 0, A = 1] = \beta_1 = 0$$

In other words, the no additive interaction effect says that the ATT effect is the same among individuals with $Z = 0$ and $Z = 1$.

Formal Proof that No Additive Interaction IV Identifies the ATT

Under (JV1-JV6), we can identify the ATT using the same ratio:

$$\text{ATT} = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[A \mid Z = 1] - \mathbb{E}[A \mid Z = 0]}$$

We begin with the numerator of this ratio.

$$
\begin{aligned}
&\mathbb{E}[Y \mid Z = z] \\
=&\mathbb{E}[Y(A, Z) \mid Z = z] \quad \text{(JV1)} \\
=&\mathbb{E}[Y(A) \mid Z = z] \quad \text{(JV5)} \\
=&\mathbb{E}[Y(1) \mid Z = z, A = 1]\mathbb{P}(A = 1 \mid Z = z) \\
&\quad + \mathbb{E}[Y(0) \mid Z = z, A = 0]\mathbb{P}(A = 0 \mid Z = z) \\
=&\mathbb{E}[Y(1) - Y(0) \mid Z = z, A = 1]\mathbb{P}(A = 1 \mid Z = z) \\
&\quad + \mathbb{E}[Y(0) \mid Z = z, A = 1]\mathbb{P}(A = 1 \mid Z = z) \\
&\quad + \mathbb{E}[Y(0) \mid Z = z, A = 0]\mathbb{P}(A = 0 \mid Z = z) \\
=&\mathbb{E}[Y(1) - Y(0) \mid Z = z, A = 1]\mathbb{P}(A = 1 \mid Z = z) + \mathbb{E}[Y(0) \mid Z = z] \\
=&\mathbb{E}[Y(1) - Y(0) \mid Z = z, A = 1]\mathbb{P}(A = 1 \mid Z = z) + \mathbb{E}[Y(0)] \quad \text{(JV2)}
\end{aligned}
$$

(JV3) ensures well-defined conditional expectation.

Proof Continued

Taking the difference $\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]$ yields

$$\begin{aligned}
&\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0] \\
=&\mathbb{E}[Y(1) - Y(0) \mid Z = 1, A = 1]\mathbb{P}(A = 1 \mid Z = 1) \\
&\quad - \mathbb{E}[Y(1) - Y(0) \mid Z = 0, A = 1]\mathbb{P}(A = 1 \mid Z = 0) \\
=&\mathbb{E}[Y(1) - Y(0) \mid Z = 0, A = 1]\left(\mathbb{P}(A = 1 \mid Z = 1) - \mathbb{P}(A = 1 \mid Z = 0)\right) \quad \text{(JV6)}
\end{aligned}$$

Dividing this by the $\mathbb{P}(A = 1 \mid Z = 1) - \mathbb{P}(A = 1 \mid Z = 0)$, which must be non-zero by assumption (JV4).

Finally, under (JV6), we have $\mathbb{E}[Y(1) - Y(0) \mid Z = 0, A = 1] = \mathbb{E}[Y(1) - Y(0) \mid A = 1]$ and we arrive at the desired result