ECI: Week 7. Noncompliance and Instrumental Variables

Jaewon ("Jay-one") Yoo

National Tsing Hua University, ISS5096

Contents

---

1 Randomized Experiments with Noncompliance

1.1 Noncompliance

In the previous weeks, we covered randomized experiments (where treatment assignment is unconfounded) and observational studies under selection on observables (where conditioning on covariates removes confounding). A natural question arises: what can we do when there is unmeasured confounding, that is, when we cannot credibly argue that all confounders are observed?

The first approach we will explore is **instrumental variables** (IV). We motivate IV through a common complication in randomized experiments: **noncompliance**. The key insight is that the logic of IV generalizes far beyond experiments, and connects directly to classical econometric methods like two-stage least squares (2SLS).

Consider a get-out-the-vote (GOTV) experiment with door-to-door canvassing. Households are randomly assigned to treatment (canvassing attempted) or control (no canvassing attempted), so treatment *assignment* is unconfounded. We use the following notation:

- $Z_i = 1$ if unit i is assigned to treatment (canvassing attempted).
- $Z_i = 0$ if unit i is assigned to control (no canvassing attempted).

However, assignment does not guarantee receipt. Some households assigned to canvassing may not answer the door, and in other settings, some units assigned to control may find a way to

obtain treatment. This is **noncompliance**: units do not follow their treatment assignment. We distinguish assignment from uptake:

- $D_i = 1$ if unit i actually received treatment (heard the canvasser's message).
- $D_i = 0$ if unit i actually received control (did not hear the message).

Full compliance means $Z_i = D_i$ for all i. When this fails for some units, we have noncompliance.

1.2 How to handle noncompliance?

What do we do if we want the effect of *canvassing exposure* (D_i)? Two approaches one might consider:

Intent-to-treat (ITT) analysis. Simply use the randomized assignment Z_i as the treatment and analyze the experiment as usual. This preserves the validity of randomization, since Z_i is unconfounded by design. The downside is that the ITT captures the effect of *assignment*, not the effect of actually receiving treatment. If noncompliance is substantial, the ITT may substantially attenuate the effect of interest.

As-treated analysis. Use the actual treatment uptake D_i as if it were randomly assigned. This directly targets the effect of canvassing exposure, but it is generally invalid. Treatment uptake is a post-randomization variable that may be correlated with potential outcomes. For example, households that answer the door may differ systematically from those that do not. Using D_i as the treatment introduces unmeasured confounding between D_i and the potential outcomes.

Why the as-treated comparison fails

To see why the as-treated comparison fails, consider the GOTV example concretely. Suppose $n = 200$ households with $n_1 = n_0 = 100$ and a complier proportion of $\pi_{\text{co}} = 0.7$ under one-sided noncompliance. The as-treated estimator computes $\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$.

To visualize the problem, consider the treatment uptake status (D_i) for each group. The following table shows a schematic of the D values we would observe:

Group	D_i **values**	**Note**
$Z = 1$ group	1, 1, 1, 1, 1, 1, 1, 0, 0, 0, ...	Mix of 1s and 0s
$Z = 0$ group	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	All $D_i = 0$

In the $Z = 1$ group, D_i varies: compliers have $D_i = 1$ (they opened the door) while noncompliers have $D_i = 0$ (they did not). In the $Z = 0$ group, all units have $D_i = 0$ because under one-sided noncompliance no control units receive treatment.

Consider who ends up in each as-treated group. The "treated" group ($D_i = 1$) consists of the 70 compliers in the $Z = 1$ group, that is, the people who were assigned to canvassing *and* opened the door. This is a *selected* subgroup of the population: these are exactly the people who chose

to comply with their assignment. No one from the $Z = 0$ group can appear in the $D_i = 1$ group, because under one-sided noncompliance no control units receive treatment. So the $D_i = 1$ group is 70 units, all of whom are compliers from the treatment arm.

The "control" group ($D_i = 0$) pools together two very different populations: the 30 noncompliers from the $Z = 1$ group (people who were assigned to canvassing but did not open the door) together with all 100 units from the $Z = 0$ group (a representative mix of compliers and noncompliers who were simply never offered canvassing). This gives a control group of $30 + 100 = 130$ units.

These two groups are systematically different populations. The $D_i = 1$ group is composed entirely of people who would comply with treatment, and compliance itself may be correlated with the outcome: people who answer the door when canvassers visit are plausibly more civically engaged, more interested in politics, and more likely to vote regardless of canvassing. The $D_i = 0$ group, by contrast, is a mixture of noncompliers (who may be less engaged) and the entire control arm (which contains both would-be compliers and would-be noncompliers, but none of whom were canvassed). This is comparing apples to oranges: the groups differ not only in whether they received treatment, but in what *kind* of people they are. By conditioning on the post-treatment variable D_i, we reintroduce exactly the confounding that randomization was designed to eliminate.

Neither approach is fully satisfactory on its own. The ITT is valid but does not answer the question of primary interest (the effect of canvassing exposure). The as-treated analysis answers the right question but introduces bias. The alternative we develop in this lecture is to leverage latent strata of **compliance types** to identify the causal effect of treatment receipt under additional assumptions.

1.3 Setup and notation

Let $Z_i \in \{0, 1\}$ denote treatment assignment and $D_i \in \{0, 1\}$ denote treatment uptake. Because treatment uptake may depend on assignment, we write potential treatment uptake as $D_i(z)$:

- $D_i(1) = 1$: if assigned to canvassing, unit i *would* open the door.
- $D_i(1) = 0$: if assigned to canvassing, unit i *would not* open the door.

Noncompliance means $D_i(z) \neq z$ for some unit i.

The consistency assumption for the observed treatment is:

$$D_i = D_i(Z_i) = Z_i D_i(1) + (1 - Z_i) D_i(0)$$

The canvassing example is an instance of **one-sided noncompliance**: people might refuse treatment when offered (i.e., $D_i(1) = 0$ for some i), but no one in the control group receives treatment (i.e., $D_i(0) = 0$ for all i). **Two-sided noncompliance**, where units can refuse to comply with either treatment *or* control, is discussed in Section 4.

1.4 Potential outcomes

With noncompliance, potential outcomes depend on both assignment and uptake: $Y_i(z, d)$. For example, $Y_i(1, 1)$ asks: would unit i vote if assigned to canvassing *and* actually canvassed?

However, we can only ever observe two potential outcomes for each unit: $Y_i(1, D_i(1))$ and $Y_i(0, D_i(0))$. The quantity $Y_i(1, 1 - D_i(1))$, the outcome under assignment to treatment but the *opposite* uptake status, is a **cross-world counterfactual** that is never observable. To see why, consider a complier (a unit with $D_i(1) = 1$). The cross-world counterfactual $Y_i(1, 1-D_i(1)) = Y_i(1, 0)$ asks: what would happen if this unit were assigned to treatment but somehow did *not* receive it? This contradicts the unit's own compliance type: a complier assigned to treatment would, by definition, take treatment. The counterfactual demands that the unit simultaneously be a complier (as defined by its potential treatment uptakes) and behave as a noncomplier (not taking treatment when assigned). No real-world scenario can produce this combination.

Cross-world counterfactuals are not needed for LATE identification under the IV framework. However, they do enter the definition of causal mediation parameters such as the average causal mediation effect (ACME). Recall from Week 5 that mediation analysis decomposes the total effect into direct and indirect components, and the indirect effect is defined in terms of quantities like $Y_i(d, M_i(d'))$ where the treatment value that determines the mediator (d') differs from the treatment value that enters the outcome (d). This is structurally identical to the cross-world counterfactual here, and it is precisely what makes identification of mediation effects harder than identification of the LATE.

The consistency assumption for the outcome is:

$$Y_i = Y_i(Z_i, D_i(Z_i))$$

1.5 Notation for group averages

We use 0/1 subscripts for assignment groups and t/c subscripts for uptake groups:

$$n_1 = \sum_{i=1}^{n} Z_i, \quad n_0 = \sum_{i=1}^{n} (1 - Z_i), \quad n_t = \sum_{i=1}^{n} D_i, \quad n_c = \sum_{i=1}^{n} (1 - D_i)$$

Average outcomes and uptake rates within each assignment group are:

$$\overline{Y}_1 = \frac{1}{n_1} \sum_{i=1}^{n} Z_i Y_i, \qquad \overline{Y}_0 = \frac{1}{n_0} \sum_{i=1}^{n} (1 - Z_i) Y_i$$

$$\overline{D}_1 = \frac{1}{n_1} \sum_{i=1}^{n} Z_i D_i, \qquad \overline{D}_0 = \frac{1}{n_0} \sum_{i=1}^{n} (1 - Z_i) D_i$$

Our first assumption is **randomization**: all potential outcomes and potential treatment uptakes are jointly independent of assignment:

$$[\{Y_i(d, z), \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$$

In observational settings where the instrument is not literally randomized, this assumption may be made conditional on some covariates $\mathbf{X}_i$.

1.6 Defining ITT effects

The **intent-to-treat** (ITT) effects are simply the average treatment effects of the assignment Z_i:

$$\text{ITT}_D = \frac{1}{n}\sum_{i=1}^{n} D_i(1) - D_i(0), \qquad \text{ITT}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i(1, D_i(1)) - Y_i(0, D_i(0))$$

ITT_D is the effect of assignment on treatment uptake, and ITT_Y is the effect of assignment on the outcome. Under one-sided noncompliance, $\text{ITT}_D \geq 0$ because $D_i(0) = 0$ for all i.

The standard estimators are:

$$\widehat{\text{ITT}}_D = \overline{D}_1 - \overline{D}_0, \qquad \widehat{\text{ITT}}_Y = \overline{Y}_1 - \overline{Y}_0$$

Under randomization of Z_i, these are unbiased, and all the usual Neyman-style inference (variances, tests, confidence intervals) applies directly. This is a major advantage of the ITT: we can use the exact same machinery from Week 2.

The problem with ITT_Y is that it conflates the true causal effect of treatment D_i with the dilution caused by noncompliance. Moreover, the effect of actual treatment receipt D_i may be more *externally valid* than the effect of assignment Z_i, since the instrument is often specific to the experimental design while the treatment itself generalizes. For example, the effect of "being assigned to a canvassing attempt" is tied to the particular logistics of one GOTV campaign, but the effect of "actually hearing a canvasser's message" may generalize to other voter mobilization contexts.

2 Compliance Types

2.1 Compliance status

We can stratify units by their **compliance type**, which describes how each unit would respond to treatment assignment. Formally, it is the value of the pair $(D_i(0), D_i(1))$ for each unit.

Under one-sided noncompliance (where $D_i(0) = 0$ for all i), there are two types:

- **Compliers**: units with $D_i(1) = 1$. These are units who would take treatment when assigned to it; in our example, they answer the door when canvassers visit.
- **Noncompliers**: units with $D_i(1) = 0$. These units would not take treatment even when assigned; they do not answer the door.

A crucial observation is that compliance type is a function of the *potential* treatment uptakes $D_i(0)$ and $D_i(1)$, which are fixed characteristics of each unit. Compliance type is therefore **pretreatment**, meaning it is determined before assignment occurs. This is a subtle but essential point: whether a person would open the door if canvassers came is a fixed trait of that person, not something caused by the random assignment. Under randomization, treatment assignment is independent of compliance type: $Z_i \perp\!\!\!\perp C_i$.

2.2 ITTs among the compliance groups

Let $C_i \in \{\text{co}, \text{nc}\}$ denote the compliance type of unit i. Define $n_{\text{co}} = \sum_{i=1}^{n} \mathbf{1}(C_i = \text{co})$ as the number of compliers, with proportion $\pi_{\text{co}} = n_{\text{co}}/n$. Similarly define n_{nc} and π_{nc} for noncompliers.

The ITT on uptake within each compliance group reveals a clean structure:

$$\text{ITT}_{D,\text{nc}} = \frac{1}{n_{\text{nc}}} \sum_{i:C_i=\text{nc}} D_i(1) - D_i(0) = 0$$

$$\text{ITT}_{D,\text{co}} = \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} D_i(1) - D_i(0) = 1$$

The intuition is straightforward: assignment has no effect on uptake for noncompliers (they never take treatment regardless of what they are assigned to), and it has a perfect effect for compliers (they take treatment if and only if assigned). This immediately implies that the overall ITT on uptake equals the **proportion of compliers**:

$$\text{ITT}_D = \pi_{\text{co}} \cdot \text{ITT}_{D,\text{co}} + \pi_{\text{nc}} \cdot \text{ITT}_{D,\text{nc}} = \pi_{\text{co}} \cdot 1 + \pi_{\text{nc}} \cdot 0 = \pi_{\text{co}}$$

This is an important result: the first-stage effect of the instrument on treatment uptake directly tells us the share of compliers in the population. If we estimate $\widehat{\text{ITT}}_D = 0.7$, we learn that 70% of the population are compliers.

2.3 Numerical example: the God's-eye view

To make the compliance framework concrete, consider a small example with $n = 8$ units. The following table shows the *complete* science table, including all potential outcomes and compliance types, as if we were omniscient. This is the "God's-eye view" that we never have access to in practice.

Unit	$D_i(0)$	$D_i(1)$	Type	$Y_i(0,0)$	$Y_i(1, D_i(1))$	τ_i
1	0	1	Complier	4	7	3
2	0	1	Complier	3	7	4
3	0	0	Noncompl.	5	5	–
4	0	1	Complier	2	7	5
5	0	0	Noncompl.	6	6	–
6	0	1	Complier	1	5	4
7	0	0	Noncompl.	5	5	–
8	0	1	Complier	2	6	4

Several features are worth noting. First, this is one-sided noncompliance: $D_i(0) = 0$ for all units. There are 5 compliers and 3 noncompliers, giving $\pi_{\text{co}} = 5/8$ and $\pi_{\text{nc}} = 3/8$. Second, the individual treatment effect $\tau_i = Y_i(1,1) - Y_i(0,0)$ is only defined for compliers, because only compliers have their treatment uptake changed by assignment. For noncompliers, Z has no effect on D or Y, so there is no treatment effect to speak of (indicated by "–"). Third, the LATE is the average treatment effect among compliers:

$$\tau_{\text{LATE}} = \frac{3 + 4 + 5 + 4 + 4}{5} = 4.0$$

2.4 Numerical example: what we actually observe

Now suppose Z is randomly assigned: units 1–4 receive $Z = 1$ and units 5–8 receive $Z = 0$. The following table shows only the columns that a researcher would observe in practice. The compliance type is hidden (shown in grey for pedagogical purposes).

Unit	Z_i	D_i	Y_i	Type (Hidden)
1	1	1	7	Complier
2	1	1	7	Complier
3	1	0	5	Noncompl.
4	1	1	7	Complier
5	0	0	6	Noncompl.
6	0	0	1	Complier
7	0	0	5	Noncompl.
8	0	0	2	Complier

From the observed data, we can compute:

- $\widehat{\text{ITT}}_Y = \overline{Y}_1 - \overline{Y}_0 = \frac{7+7+5+7}{4} - \frac{6+1+5+2}{4} = 6.5 - 3.5 = 3.0$
- $\widehat{\text{ITT}}_D = \overline{D}_1 - \overline{D}_0 = 0.75 - 0 = 0.75$
- $\widehat{\tau}_{\text{iv}} = 3.0 \;/\; 0.75 = 4.0 = \tau_{\text{LATE}}$

Compliance type is *never observed* directly, yet the Wald estimator recovers the LATE. The exact match between $\widehat{\tau}_{\text{iv}}$ and τ_{LATE} is specific to this example; with small n, the Wald estimator generally does not equal the LATE exactly. It converges to the LATE asymptotically as the sample size grows.

3 Instrumental Variables

3.1 Identification assumptions

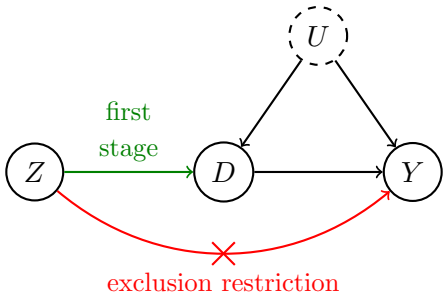
Beyond randomization (Assumption 1), two additional assumptions are needed to identify the causal effect of treatment receipt.

Assumption 2: First stage (relevance). The instrument must affect treatment uptake: $\text{ITT}_D = \pi_{\text{co}} \neq 0$. In words, at least one unit must comply with treatment assignment. Without a first stage, the instrument provides no information about the effect of treatment. This assumption is testable from the data: we can estimate $\widehat{\text{ITT}}_D$ and test whether it is significantly different from zero. A weak first stage (small π_{co}) does not technically violate this assumption, but it creates practical problems for estimation, as we discuss below.

Assumption 3: Exclusion restriction. The instrument Z_i affects the outcome Y_i *only through* treatment uptake D_i. Formally: $Y_i(z, d) = Y_i(z', d)$ for all z, z', and d. In the GOTV example, this means that being assigned to canvassing affects voter turnout only through actually hearing the canvasser's message; assignment per se has no direct effect on voting.

The exclusion restriction is **not a testable assumption** and cannot be guaranteed by experimental design. It must be justified on substantive grounds. Of the three IV assumptions, the exclusion restriction is often the most controversial, and we devote the next subsection to discussing when and how it can fail.

The following DAG summarizes the IV setup:



The instrument Z affects the treatment D (the first stage, shown in green), D affects the outcome Y, and there may be unmeasured confounders U (dashed) that affect both D and Y. The exclusion restriction rules out any direct path from Z to Y (shown as the crossed-out red arrow). Note that IV tolerates unmeasured confounding between D and Y: the dashed U node with arrows into both D and Y is perfectly compatible with valid IV estimation. This is the fundamental advantage of IV over selection-on-observables methods.

Under the exclusion restriction, potential outcomes can be written as functions of D_i alone:

$$Y_i(1) = Y_i(D_i = 1) = Y_i(Z_i = 1, D_i = 1)$$

$$Y_i(0) = Y_i(D_i = 0) = Y_i(Z_i = 1, D_i = 0)$$

The first argument (Z_i) drops out because the exclusion restriction guarantees that Y_i does not depend on Z_i once D_i is fixed.

3.2 When does the exclusion restriction fail?

The exclusion restriction states that Z_i affects Y_i only through D_i. It is not testable from the data and cannot be guaranteed by experimental design alone; it must be justified on substantive grounds. Of the three IV assumptions, the exclusion restriction is often the most controversial, and the credibility of any IV analysis depends critically on whether this assumption is plausible.

Consider the GOTV example. The exclusion restriction requires that being *assigned* to canvassing affects voter turnout only through actually *hearing* the canvasser's message. But is this plausible? There are at least two channels through which the exclusion restriction could be violated:

The flyer problem. In many GOTV campaigns, when canvassers visit a household and no one answers the door, they leave a **flyer** on the door with information about the upcoming election: polling locations, candidate summaries, or reminders of election day. This means that noncompliers (people who did not open the door) still receive election information through the flyer. If this information affects their voting behavior, then Z_i affects Y_i through a channel other than D_i

(hearing the canvasser's message), violating the exclusion restriction. In terms of the DAG, the flyer creates a direct arrow from Z to Y that bypasses D.

The knock problem. Even without a flyer, the act of knocking on the door might itself remind a household about the upcoming election. A knock signals that *someone cares enough about the election to come to your door*, which could change one's sense of civic duty or salience of the election. If the knock alone (without delivering the canvasser's message) affects voting behavior, the exclusion restriction is violated.

The Angrist (1990) draft lottery. A classic example comes from Angrist (1990), who used the Vietnam-era draft lottery as an instrument for military service to estimate its effect on lifetime earnings. The exclusion restriction requires that a low draft lottery number affects earnings only through actual military service. But receiving a low draft number could cause psychological stress, prompt changes in college enrollment plans (e.g., enrolling in graduate school to obtain a deferment), or alter career expectations, all before any military service occurs. If any of these channels affect lifetime earnings independently of serving in the military, the exclusion restriction is violated. Angrist acknowledges this concern and argues that the magnitude of such direct effects is likely small relative to the effect operating through military service.

The key practical lesson is that whenever an instrument is proposed, one should ask: is there *any* channel through which Z could affect Y other than through D? If the answer is plausibly yes, the exclusion restriction may not hold, and the IV estimate should be interpreted with caution. Unlike the first-stage assumption, which can be tested empirically, the exclusion restriction must be defended through careful reasoning about the substantive context.

3.3 Outcome ITTs and compliance types

We can decompose the ITT on the outcome by compliance type:

- $\text{ITT}_{Y,\text{co}}$: the effect of assignment on the outcome among compliers.
- $\text{ITT}_{Y,\text{nc}}$: the effect of assignment on the outcome among noncompliers.

The exclusion restriction has immediate implications: since noncompliers' treatment uptake does not change with assignment (D_i stays at 0 regardless of Z_i), the exclusion restriction implies $\text{ITT}_{Y,\text{nc}} = 0$. The logic is direct: if Z_i affects Y_i only through D_i, and D_i does not change for noncompliers, then Y_i cannot change either. For compliers, assignment *does* change uptake (from $D_i = 0$ to $D_i = 1$), so $\text{ITT}_{Y,\text{co}}$ captures the effect of treatment receipt among this group.

This allows us to connect the overall ITT_Y to the compliance groups:

$$\text{ITT}_Y = \pi_{\text{co}} \cdot \text{ITT}_{Y,\text{co}} + \pi_{\text{nc}} \cdot \underbrace{\text{ITT}_{Y,\text{nc}}}_{=0} = \text{ITT}_D \cdot \text{ITT}_{Y,\text{co}}$$

This decomposition is the key equation in IV. It tells us that the ITT on the outcome is the product of two things: the share of compliers ($\text{ITT}_D = \pi_{\text{co}}$) and the causal effect of treatment on compliers ($\text{ITT}_{Y,\text{co}}$). The noncomplier term drops out entirely because of the exclusion restriction.

3.4 The local average treatment effect (LATE)

Under the exclusion restriction, $\text{ITT}_{Y,\text{co}}$ is the causal effect of treatment receipt among compliers:

$$\text{ITT}_{Y,\text{co}} = \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} Y_i(1, D_i(1)) - Y_i(0, D_i(0)) = \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} Y_i(D_i = 1) - Y_i(D_i = 0) = \tau_{\text{LATE}}$$

The second equality uses the exclusion restriction to drop the first argument: since $Y_i(z, d) = Y_i(z', d)$, we can write $Y_i(d)$ instead of $Y_i(z, d)$. This quantity is known as the **local average treatment effect** (LATE), because it is local to the subpopulation of compliers. It is also referred to as the **complier average causal effect** (CACE). The LATE is a conditional ATE, where we condition on units being compliers.

The **LATE theorem** states that under one-sided noncompliance, with randomization, first stage, and the exclusion restriction:

$$\tau_{\text{LATE}} = \text{ITT}_{Y,\text{co}} = \frac{\text{ITT}_Y}{\text{ITT}_D}$$

This is a remarkable result: the ratio of two quantities that are identified by randomization alone (the ITT on the outcome and the ITT on uptake) identifies the causal effect of treatment among compliers. Neither quantity by itself gives us the treatment effect, but their ratio does.

3.5 Noncompliance dilutes the ITT

The relationship $\text{ITT}_Y = \pi_{\text{co}} \cdot \tau_{\text{LATE}}$ has a direct implication for the magnitude of the ITT relative to the LATE. Since $0 < \pi_{\text{co}} \leq 1$, it follows that:

$$\tau_{\text{LATE}} \geq \text{ITT}_Y \quad \text{always}$$

with equality holding only when $\pi_{\text{co}} = 1$ (full compliance). In other words, noncompliance *dilutes* the intent-to-treat effect: the ITT is attenuated relative to the true causal effect of treatment on compliers, because assignment moves only a fraction of the population into treatment.

To build intuition, consider an extreme case. Suppose the true LATE is 10 percentage points (canvassing has a large effect on compliers), but only 10% of the population are compliers ($\pi_{\text{co}} = 0.1$). The ITT would be $\text{ITT}_Y = 0.1 \times 10 = 1$ percentage point. An analyst looking only at the ITT might conclude that canvassing has a negligible effect, when in fact it has a large effect on the people it actually reaches. The dilution occurs because the ITT averages the large effect on compliers with zero effect on the 90% of noncompliers.

The Wald estimator "undoes" this dilution by dividing the ITT on the outcome by the proportion of compliers: $\tau_{\text{LATE}} = \text{ITT}_Y / \pi_{\text{co}}$. The denominator rescales the numerator to recover the effect among the subpopulation whose treatment status was actually changed by the instrument.

3.6 The Wald estimator

The **Wald estimator** (or **instrumental variable estimator**) for the LATE is:

$$\widehat{\tau}_{\text{iv}} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_D} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_0}$$

This is the ratio of two unbiased estimators. As a ratio, $\widehat{\tau}_{\text{iv}}$ is not itself unbiased, but it is **consistent** for τ_{LATE}.

Why biased?

Let $A = \widehat{\text{ITT}}_Y$ and $B = \widehat{\text{ITT}}_D$. In finite samples, both A and B are random variables, and the expectation of a ratio does not equal the ratio of expectations: $\mathbb{E}[A/B] \neq \mathbb{E}[A]/\mathbb{E}[B]$. The source of this bias is the nonlinearity of division: the denominator B is a random variable, and the function $f(B) = A/B$ is nonlinear in B. When B takes values near zero, the ratio A/B becomes very large in magnitude, creating an asymmetry that pulls $\mathbb{E}[A/B]$ away from $\mathbb{E}[A]/\mathbb{E}[B]$.

It is important to note that the numerator A being random does *not* contribute to the bias. The numerator appears in a linear position: A/B is linear in A for fixed B, so $\mathbb{E}[A/B|B] = \mathbb{E}[A|B]/B$, and the nonlinearity comes entirely from averaging over B. The bias arises specifically because B is random and sits in the denominator.

One might be tempted to invoke Jensen's inequality, which states that $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$ for a nonlinear function g. While this captures the correct intuition, the Wald estimator involves the ratio of *two correlated* random variables, not a nonlinear function of a single random variable. The precise bias formula comes from a bivariate Taylor expansion (the delta method applied to the function $g(a, b) = a/b$), which accounts for the variances and covariance of A and B.

Why consistent?

As the sample size grows, $B = \widehat{\text{ITT}}_D$ converges in probability to $\mathbb{E}[B] = \text{ITT}_D$, which is a constant (by the law of large numbers). Similarly, $A = \widehat{\text{ITT}}_Y$ converges in probability to ITT_Y. Once B is effectively constant (no longer random), the ratio A/B behaves like A divided by a constant, and the bias-creating mechanism vanishes. Formally, by the **continuous mapping theorem**, if $(A, B) \xrightarrow{p} (\text{ITT}_Y, \text{ITT}_D)$ and $\text{ITT}_D \neq 0$, then $A/B \xrightarrow{p} \text{ITT}_Y/\text{ITT}_D = \tau_{\text{LATE}}$.

The condition $\text{ITT}_D \neq 0$ is precisely the first-stage assumption. Without it, B converges to zero, and the continuous mapping theorem does not apply (division by zero). This is why a relevant first stage is essential for consistency.

Equivalence to 2SLS

The Wald estimator is algebraically equivalent to the **two-stage least squares** (2SLS) estimator:

1. **First stage**: regress D_i on Z_i to obtain fitted values $\widehat{D}_i$.
2. **Second stage**: regress Y_i on $\widehat{D}_i$.

The coefficient on $\widehat{D}_i$ in the second stage equals $\widehat{\tau}_{\text{iv}}$. This equivalence holds because with a single binary instrument and a single binary treatment, 2SLS reduces to the Wald ratio. In more general settings (multiple instruments, continuous treatments, or additional covariates), 2SLS extends the IV logic beyond the simple ratio.

Variance and weak instruments

The variance of the Wald estimator can be obtained via the delta method:

$$\mathbb{V}[\widehat{\tau}_{\text{iv}}] = \frac{1}{\text{ITT}_D^2}\,\mathbb{V}\Big[\widehat{\text{ITT}}_Y\Big] + \frac{\text{ITT}_Y^2}{\text{ITT}_D^4}\,\mathbb{V}\Big[\widehat{\text{ITT}}_D\Big] - 2\cdot\frac{\text{ITT}_Y}{\text{ITT}_D^3}\cdot\text{cov}\Big[\widehat{\text{ITT}}_Y, \widehat{\text{ITT}}_D\Big]$$

Notice that the variance is inversely proportional to ITT_D^2: a weak first stage (small π_{co}) inflates the variance of the IV estimator, sometimes dramatically. This is the well-known **weak instrument** problem. When the instrument barely moves treatment uptake, the denominator $\widehat{\text{ITT}}_D$ is small and noisy, and the ratio $\widehat{\text{ITT}}_Y/\widehat{\text{ITT}}_D$ becomes highly unstable. In practice, applied researchers use the first-stage F-statistic as a diagnostic: the common rule of thumb is $F > 10$ (Stock and Yogo, 2005), though this threshold is context-dependent.

4 Two-Sided Noncompliance

4.1 The encouragement design

In the GOTV example, noncompliance was one-sided: people could refuse canvassing, but no one in the control group could obtain it. In many settings, however, noncompliance goes both ways. Consider an **encouragement design**, where the researcher randomly assigns an encouragement to take some treatment:

- Some encouraged units may refuse the encouragement and not take treatment.
- Some non-encouraged units may take treatment on their own.

Here, Z_i is the encouragement and D_i is the actual treatment. The estimation strategy is the same as before; what changes are the identification assumptions.

A canonical example is an experiment studying the effect of a new educational program. Students are randomly encouraged to participate ($Z_i = 1$), but some encouraged students decline, while some non-encouraged students hear about the program from friends and sign up on their own. In this setting, both $D_i(1) = 0$ (encouraged but not participating) and $D_i(0) = 1$ (not encouraged but participating) are possible.

4.2 Four compliance types

With two-sided noncompliance, $D_i(0)$ is no longer fixed at 0 for all units. The pair $(D_i(0), D_i(1))$ can take four values, defining four compliance types (also called **principal strata**):

- **Complier**: $D_i(1) = 1$ and $D_i(0) = 0$. Takes treatment when encouraged, does not when not encouraged. These are the units whose behavior is moved by the instrument.
- **Always-taker**: $D_i(1) = D_i(0) = 1$. Takes treatment regardless of encouragement. The instrument has no effect on their treatment uptake.
- **Never-taker**: $D_i(1) = D_i(0) = 0$. Never takes treatment regardless of encouragement. Like always-takers, the instrument does not move them.

- **Defier**: $D_i(1) = 0$ and $D_i(0) = 1$. Does the opposite of what is encouraged. These units are contrarian: they take treatment precisely when discouraged and refuse it when encouraged.

A critical feature of compliance types is that they are **not directly observable** from the data. The connection between observed data and compliance types is summarized in the following table:

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Never-taker or Complier	Never-taker or Defier
$D_i = 1$	Always-taker or Defier	Always-taker or Complier

Each cell contains two possible compliance types, so we cannot determine any individual's type from their observed (Z_i, D_i) pair. For example, a unit with $Z_i = 1$ and $D_i = 1$ could be either a complier (who took treatment because they were encouraged) or an always-taker (who would have taken treatment regardless).

Let π_{co}, π_{at}, π_{nt}, and π_{df} denote the proportions of each type. The ITT on treatment uptake is now:

$$\text{ITT}_D = \pi_{\text{co}} - \pi_{\text{df}}$$

Defiers work in the opposite direction from compliers, making things considerably messier. If defiers exist, the first-stage effect is a net quantity: compliers push ITT_D up, defiers push it down.

4.3 IV assumptions under two-sided noncompliance

The canonical IV assumptions for Z_i to be a valid instrument under two-sided noncompliance are:

1. **Randomization** of Z_i.
2. **First stage**: $\pi_{\text{co}} \neq 0$ (presence of some compliers).
3. **Exclusion restriction**: $Y_i(z, d) = Y_i(z', d)$ for all z, z', and d.
4. **Monotonicity**: $D_i(1) \geq D_i(0)$ for all i (no defiers).

The monotonicity assumption eliminates defiers, so $\pi_{\text{df}} = 0$ and $\text{ITT}_D = \pi_{\text{co}}$. This restores the clean interpretation from the one-sided case. Note that one-sided noncompliance is a special case of monotonicity: if $D_i(0) = 0$ for all i, then $D_i(1) \geq D_i(0)$ automatically holds.

With all four assumptions, the decomposition of ITT_Y proceeds as follows:

$$\begin{aligned}\text{ITT}_Y &= \pi_{\text{co}} \cdot \text{ITT}_{Y,\text{co}} + \pi_{\text{at}} \cdot \underbrace{\text{ITT}_{Y,\text{at}}}_{=0 \text{ (ER)}} + \pi_{\text{nt}} \cdot \underbrace{\text{ITT}_{Y,\text{nt}}}_{=0 \text{ (ER)}} + \underbrace{\pi_{\text{df}}}_{=0 \text{ (mono)}} \cdot \text{ITT}_{Y,\text{df}} \\ &= \pi_{\text{co}} \cdot \text{ITT}_{Y,\text{co}}\end{aligned}$$

The exclusion restriction zeroes out the ITT on outcomes for always-takers and never-takers: their treatment status does not change with Z_i (always-takers always have $D_i = 1$, never-takers always have $D_i = 0$), so if Z_i affects Y_i only through D_i, their outcome cannot change either. Monotonicity zeroes out the defier term by assumption. The same identification result follows: $\tau_{\text{LATE}} = \text{ITT}_Y / \text{ITT}_D$.

4.4 Is the LATE useful?

The LATE is the causal effect of treatment on an *unknown* subpopulation: the compliers. In practice:

- Among the treated units ($D_i = 1$), we observe a mix of always-takers and compliers.
- Among the control units ($D_i = 0$), we observe a mix of never-takers and compliers.

We cannot identify which specific units are compliers from the observed data.

Without further assumptions, $\tau_{\text{LATE}} \neq \tau$ (the population ATE). The complier group is instrument-dependent: different instruments will generally identify different complier populations and therefore different LATEs. This has been a source of debate in applied work. For example, using the draft lottery versus proximity to a military base as instruments for military service would identify the treatment effect among different subpopulations of compliers, and there is no reason to expect these two LATEs to be the same.

However, it is important to recognize that we cannot do any better in terms of *point estimation* without imposing stronger assumptions. The LATE is what the data and the IV assumptions can deliver. Alternatives include bounding the ATE under weaker assumptions (as discussed in the partial identification lecture) or arguing that the complier population is itself policy-relevant. In many applied settings, the compliers are precisely the policy-relevant group: they are the people whose behavior the policy instrument can actually change.

Abadie (2003) shows that while we cannot identify *which* individuals are compliers, we can estimate the *characteristics* of the complier population. For any pretreatment covariate X_i, the complier mean is identified as:

$$\mathbb{E}[X_i \mid \text{complier}] = \frac{\mathbb{E}[X_i \cdot D_i \mid Z_i = 1] - \mathbb{E}[X_i \cdot D_i \mid Z_i = 0]}{\mathbb{E}[D_i \mid Z_i = 1] - \mathbb{E}[D_i \mid Z_i = 0]}$$

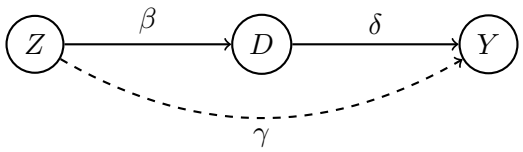
This is the same Wald ratio structure, but with $X_i \cdot D_i$ replacing Y_i. Comparing complier means to population means helps assess external validity: if compliers look similar to the population of interest, the LATE may be more broadly informative.

5 IV and Mechanisms

5.1 The structural parallel between IV and mediation

In the IV setup, D is an **intermediate outcome** that lies on the causal path between Z and Y. The DAG is $Z \to D \to Y$, which is exactly the same structure as in mediation analysis, where we decompose the total effect of a treatment into a direct effect and an indirect effect through a mediator. To make this parallel explicit, label the edges of the DAG:

- β: the effect of Z on D (the first stage, or the "a path" in mediation).
- δ: the effect of D on Y (the effect of the mediator on the outcome, or the "b path").
- γ: the direct effect of Z on Y, bypassing D (the "c' path").



In IV, D is treated as an intermediate outcome (the treatment), and we are interested in estimating δ. In mediation analysis, D is treated as a mediator, and we are interested in decomposing the total effect of Z into the indirect effect ($\beta \times \delta$) and the direct effect (γ).

5.2 Why IV cannot decompose mechanisms

This structural similarity raises a natural question: can we use the IV framework to study mechanisms, that is, to decompose the effect of Z on Y into a direct component (γ) and an indirect component ($\beta \times \delta$)? The answer is no, and the reason is instructive.

The IV framework identifies δ only when $\gamma = 0$. This is precisely the exclusion restriction: Z has no direct effect on Y. But if $\gamma = 0$, then there is no direct effect, and the decomposition becomes trivial. The entire effect of Z on Y operates through D, so the "indirect" effect *is* the total effect.

The interesting case for mediation analysis is when $\gamma \neq 0$, meaning there is a nontrivial direct effect to separate from the indirect effect. But when $\gamma \neq 0$, the exclusion restriction fails, and the IV estimator no longer identifies δ. The Wald ratio $\widehat{\text{ITT}}_Y / \widehat{\text{ITT}}_D$ would conflate δ with the direct effect γ, and the resulting estimate would not have a clean causal interpretation.

This creates a fundamental tension: IV *requires* $\gamma = 0$ to work, but mediation analysis is *most interesting* when $\gamma \neq 0$. The two frameworks are therefore answering different questions in different settings. IV tells us *whether* D affects Y (by identifying δ under the exclusion restriction), but it cannot tell us *how* Z affects Y when both direct and indirect channels are present.

5.3 What is needed instead: causal mediation analysis

Studying mechanisms requires a separate framework: **causal mediation analysis**, which relies on **sequential ignorability** (Imai, Keele, and Yamamoto, 2010; see Week 5 Notes; covered in detail later in the course). Sequential ignorability involves two conditions: (1) the treatment is unconfounded given pre-treatment covariates, and (2) the mediator is unconfounded given the treatment and pre-treatment covariates. The second condition is particularly strong because it rules out all post-treatment confounders of the mediator-outcome relationship, including unobserved ones.

Sequential ignorability is inherently untestable, just like the exclusion restriction. However, unlike the exclusion restriction, it allows for a nontrivial decomposition when $\gamma \neq 0$. The price is that the assumption is typically considered even stronger than the exclusion restriction, which is why Imai, Keele, and Tingley (2011) argue that **sensitivity analysis should be a routine part of any mediation analysis**. Their approach uses the correlation ρ between the residuals of the mediator model and the outcome model as a sensitivity parameter, assessing how robust the estimated ACME is to violations of sequential ignorability.

The takeaway is that IV and mediation analysis are complementary tools that answer different causal questions. IV identifies the effect of D on Y (among compliers) without requiring that D is unconfounded, but it cannot decompose mechanisms. Mediation analysis can decompose mechanisms, but it requires the strong assumption that the mediator is unconfounded. Neither tool subsumes the other, and the choice between them depends on the research question and the assumptions the researcher is willing to defend.

6 In R: GOTV Simulation

To tie together the concepts from this lecture, we simulate the GOTV canvassing experiment where we know the true data-generating process. This allows us to verify that the Wald estimator recovers the true LATE while the as-treated comparison does not.

6.1 Data-generating process

The simulation creates $n = 500$ households randomized to canvassing (Z_i). Sixty percent are compliers who would open the door if canvassed ($\pi_{\text{co}} = 0.6$), and forty percent are noncompliers who never open the door. Noncompliance is one-sided: $D_i(0) = 0$ for all units.

Crucially, compliers have a higher baseline turnout propensity (+3 percentage points), reflecting the fact that people who answer the door are more civically engaged and not a random sample of the population. This baseline gap is what makes the as-treated analysis biased: it confounds the true treatment effect with the pre-existing difference between compliers and noncompliers. The true LATE is 5 percentage points, and noncompliers have zero treatment effect (since they never receive treatment).

```r
set.seed(02138); n <- 500
complier <- rbinom(n, 1, 0.6)
Z <- rbinom(n, 1, 0.5)
D <- Z * complier   # one-sided noncompliance

# Compliers have higher baseline turnout (+3 pp)
Y0 <- 10 + 3 * complier + rnorm(n)
Y1 <- Y0 + 5 * complier  # true LATE = 5
Y  <- D * Y1 + (1 - D) * Y0

my_data <- data.frame(Z = Z, D = D, Y = Y)
```

Let us walk through the key features of this DGP. The baseline outcome $Y_i(0)$ is generated as $10 + 3 \times \texttt{complier}_i + \varepsilon_i$, so it averages 13 for compliers and 10 for noncompliers. The treatment effect $Y_i(1) - Y_i(0) = 5 \times \texttt{complier}_i$, so it is 5 for compliers and 0 for noncompliers. Treatment uptake $D_i = Z_i \times \texttt{complier}_i$ means that only compliers in the treatment arm actually receive treatment. After assignment, we observe only (Z_i, D_i, Y_i) and not compliance type.

Notice that the true ITT on the outcome should be approximately $\text{ITT}_Y = \pi_{\text{co}} \times \tau_{\text{LATE}} = 0.6 \times 5 = 3.0$. This is the diluted effect: even though canvassing has a 5-point effect on those it reaches,

it reaches only 60% of those assigned to treatment, so the average effect of assignment is only 3 points.

6.2 Three estimators, three answers

With the simulated data, we compute and compare the ITT, the Wald estimator, and the as-treated comparison:

```
ITT_Y <- mean(Y[Z == 1]) - mean(Y[Z == 0])   # 2.96
ITT_D <- mean(D[Z == 1]) - mean(D[Z == 0])   # 0.60
Wald  <- ITT_Y / ITT_D                        # 4.91
as_tr <- mean(Y[D == 1]) - mean(Y[D == 0])    # 7.82
```

The results are summarized in the following table:

Estimator	Estimate	Truth	Interpretation
$\widehat{\text{ITT}}_Y$	2.96	3.0	Effect of *assignment* on turnout
$\widehat{\text{ITT}}_D$ $(= \widehat{\pi}_{\text{co}})$	0.60	0.6	Proportion of compliers
$\widehat{\tau}_{\text{iv}}$ (Wald)	4.91	5.0	Effect of *canvassing* on compliers
As-treated	7.82	n/a	Confounded (+3 from selection)

The ITT estimate of 2.96 is close to its target $\pi_{\text{co}} \times \text{LATE} = 0.6 \times 5 = 3.0$. This is the diluted effect: assignment moves only 60% of units into treatment, so the average effect of assignment is attenuated. The Wald estimator of 4.91 is close to the true LATE of 5.0: it "undoes" the dilution by dividing by $\widehat{\pi}_{\text{co}} = 0.60$, recovering the effect of treatment on those whose behavior was actually changed by the instrument.

The as-treated estimator of 7.82 is far from the true treatment effect of 5.0. It picks up the +3 baseline gap between compliers and noncompliers on top of the true effect, inflating the estimate substantially. The as-treated estimator compares the $D_i = 1$ group (all compliers, with baseline turnout around 13) to the $D_i = 0$ group (a mix of noncompliers and untreated compliers, with lower baseline turnout). The resulting estimate is approximately $5 + 3 = 8$: the true LATE plus the selection bias. This is exactly the confounding that the as-treated analysis reintroduces.

6.3 Confirming with `ivreg`

The `AER` package implements two-stage least squares. With binary Z and D, the 2SLS estimator is algebraically identical to the Wald estimator:

```
library(AER)
iv_model  <- ivreg(Y ~ D | Z, data = my_data)
ols_model <- lm(Y ~ D, data = my_data)
```

The `ivreg` formula `Y ~ D | Z` specifies that D is the endogenous treatment variable and Z is the instrument. The vertical bar separates the structural equation (left: outcome regressed on treatment) from the instrument list (right: the variable used to instrument for D). The coefficient on D in the IV model equals $\widehat{\tau}_{\text{iv}} = 4.91$, which is consistent for the LATE.

The OLS coefficient from `lm(Y ~ D)` equals the as-treated estimate of 7.82, which treats D as exogenous and is biased when D is confounded. OLS recovers the causal effect only when D is unconfounded; here, D is correlated with compliance type (and hence with baseline outcomes), so OLS is inconsistent for the LATE.

Because we know the DGP, we can verify directly: IV recovers the truth, OLS does not. In practice, we never know the true DGP, which is why simulations like this are valuable for building intuition and verifying that estimators behave as theory predicts.

6.4 Reference: manual Wald estimator with delta method variance

For completeness, the following code computes the Wald estimator and its variance from scratch using the delta method, along with the 2SLS implementation. The code is structured to follow the Neyman-style variance estimation from Week 2, applied to the ratio estimator:

```r
# Proportion of compliers (= ITT_D)
pi_co <- mean(my_data$D[my_data$Z == 1]) -
         mean(my_data$D[my_data$Z == 0])

# Compute ITTs
ITT_Y <- mean(my_data$Y[my_data$Z == 1]) -
         mean(my_data$Y[my_data$Z == 0])
ITT_D <- mean(my_data$D[my_data$Z == 1]) -
         mean(my_data$D[my_data$Z == 0])
# Note: ITT_D = pi_co

# Wald estimator
Wald_est <- ITT_Y / ITT_D

# Variance of ITT_Y (Neyman estimator)
Var_ITT_Y <- var(my_data$Y[my_data$Z == 1]) / sum(my_data$Z == 1) +
             var(my_data$Y[my_data$Z == 0]) / sum(my_data$Z == 0)

# Variance of ITT_D (Neyman estimator)
Var_ITT_D <- var(my_data$D[my_data$Z == 1]) / sum(my_data$Z == 1) +
             var(my_data$D[my_data$Z == 0]) / sum(my_data$Z == 0)

# Covariance term (within the treated assignment group)
demeaned_y <- my_data$Y[my_data$Z == 1] -
              mean(my_data$Y[my_data$Z == 1])
demeaned_d <- my_data$D[my_data$Z == 1] -
              mean(my_data$D[my_data$Z == 1])
denom <- sum(my_data$Z == 1) * (sum(my_data$Z == 1) - 1)
Covar_est <- sum(demeaned_y * demeaned_d) / denom
```

```r
# Delta method variance of the Wald estimator
Var_Wald <- (1/ITT_D^2) * Var_ITT_Y +
            (ITT_Y^2 / ITT_D^4) * Var_ITT_D -
            2 * (ITT_Y / ITT_D^3) * Covar_est

# Or use TSLS with AER::ivreg
ivmodel <- AER::ivreg(Y ~ D | Z, data = my_data)
ivpack::robust.se(ivmodel)
```

The delta method variance formula has three terms. The first term, $(1/\text{ITT}_D^2) \cdot \mathbb{V}[\widehat{\text{ITT}}_Y]$, captures the contribution of sampling variability in the numerator, scaled by the first-stage strength. The second term, $(\text{ITT}_Y^2/\text{ITT}_D^4) \cdot \mathbb{V}[\widehat{\text{ITT}}_D]$, captures the contribution of sampling variability in the denominator; notice the ITT_D^4 in the denominator, which makes this term large when the first stage is weak. The third term is the covariance correction.

7 In-Class Exercise: Noncompliance in a Pharmaceutical Trial

Setting. A pharmaceutical company runs a randomized trial for a new drug. $n = 200$ patients are randomized: 100 assigned to treatment ($Z_i = 1$), 100 to placebo ($Z_i = 0$). However, some patients in the treatment group refuse to take the drug.

	n	Took drug ($\overline{D}$)	Health score ($\overline{Y}$)
Assigned treatment ($Z = 1$)	100	0.70	75
Assigned placebo ($Z = 0$)	100	0.00	70

1. Is this one-sided or two-sided noncompliance? Why?
2. Compute $\widehat{\text{ITT}}_Y$, $\widehat{\text{ITT}}_D$, and $\widehat{\tau}_{\text{iv}}$. Interpret each in words.
3. What is $\widehat{\pi}_{\text{co}}$? What proportion of the treatment group are noncompliers?
4. A colleague suggests: "Just compare patients who actually took the drug vs. those who didn't." What is wrong with this?
5. Think of a plausible violation of the exclusion restriction in this setting. Is there a channel from Z to Y that bypasses D?

8 Summary

Noncompliance in experiments creates a gap between treatment assignment (Z_i) and treatment receipt (D_i). The ITT analysis is valid but does not recover the effect of treatment itself; it captures only the diluted effect of assignment. The as-treated analysis targets the right quantity but reintroduces confounding by conditioning on the post-treatment variable D_i, creating systematically different comparison groups.

Compliance types stratify units by how they respond to assignment. Under one-sided noncompliance, there are compliers and noncompliers, and $\text{ITT}_D = \pi_{\text{co}}$. Under two-sided noncompliance,

four types emerge (compliers, always-takers, never-takers, defiers), and a fourth assumption (monotonicity, or no defiers) is needed.

Under randomization, first stage, and the exclusion restriction, the **LATE** (the ATE among compliers) is identified as $\tau_{\text{LATE}} = \text{ITT}_Y / \text{ITT}_D$. The exclusion restriction is the most controversial assumption: it is not testable and must be argued substantively by asking whether there is any channel through which the instrument affects the outcome other than through treatment uptake.

Noncompliance dilutes the ITT: $\text{ITT}_Y = \pi_{\text{co}} \cdot \tau_{\text{LATE}}$, so $\tau_{\text{LATE}} \geq \text{ITT}_Y$ always, with equality only under full compliance. The **Wald estimator** $\widehat{\tau}_{\text{iv}} = \widehat{\text{ITT}}_Y / \widehat{\text{ITT}}_D$ undoes this dilution. It is biased in finite samples (because the denominator is random and appears in a nonlinear position) but consistent for the LATE (because the denominator converges to a constant by the law of large numbers, and the continuous mapping theorem applies).

The LATE is local to compliers: different instruments identify different complier populations and potentially different treatment effects. While this limits external validity, the LATE is the best available point estimate under the IV assumptions.

IV shares the DAG structure of mediation ($Z \to D \to Y$) but cannot decompose direct and indirect effects. IV requires the exclusion restriction ($\gamma = 0$), which makes mediation decomposition trivial; when $\gamma \neq 0$ (the interesting case for mediation), IV breaks down. Mediation analysis requires a separate framework with sequential ignorability, which is a strong assumption demanding sensitivity analysis.

Up next: two-stage least squares in more general settings, with multiple instruments and covariates.