

Week 1: Selection Bias and Experiments

From causal diagrams to a credible comparison

[Class slides](#) · [Download slides as PDF](#) · [Ungraded exercise](#) · [Download notes as PDF](#) · [Download Quarto source](#)

Discussant assignments are available on Moodle.

Learning Objectives

By the end of this topic, you should be able to:

- Use a causal diagram and potential outcomes to explain selection bias, and derive omitted-variable bias in the schooling example.
- Explain how random assignment creates a credible comparison and identify the assumptions needed to interpret an experimental effect.
- Explain how power, balance checks, and pre-registration contribute to designing and assessing an experiment.
- Interpret experimental estimates and their uncertainty, and distinguish internal from external validity in the vaccination study.

Before class

Read these notes and Campos-Mercade et al. (2021), *Monetary incentives increase COVID-19 vaccinations*. The [open-access text](#) includes links to the supplementary materials. Read Sections 2 and 3.1 of the [pre-analysis plan](#), and compare the outcome definition with supplementary Section 1.2.2. The [trial registry](#) records registration and subsequent versions.

We begin with the causal diagrams introduced in PP5000. Our question throughout is: **what makes the comparison group a credible counterfactual?**

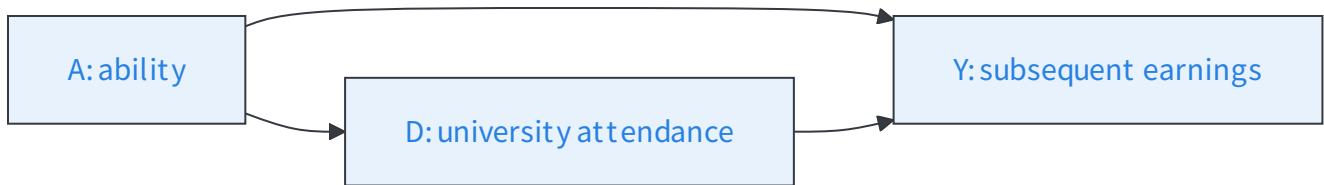
Textbook references

- **The Effect:** Chapters 6-8 (DAG review), 10 (treatment effects), and 13 (regression and omitted-variable bias). [Online book](#).
- **Causal Inference: The Remix:** Chapters 3 (Directed Acyclic Graphs) and 4 (Potential Outcomes and Randomization). [Online book](#).
- **Mastering 'Metrics:** Chapters 1 (Randomized Trials) and 2 (Regression). [Mastering Econometrics videos](#) (Marginal Revolution University).
- **Mostly Harmless Econometrics (MHE), optional advanced reading:** Chapter 2 (The Experimental Ideal) and relevant parts of Chapter 3 (Making Regression Make Sense). [Contents](#).

These are complementary references for review and clarification, rather than a requirement to read all four treatments.

1. Start with the causal diagram

Return to the example from PP5000: does university attendance raise subsequent earnings? Let D indicate university attendance, Y denote subsequent earnings, and A denote ability measured before university. Before looking at any regression, consider this account of how the variables are related:



The arrow $D \rightarrow Y$ represents the effect we want to learn about. The path $D \leftarrow A \rightarrow Y$ is a **backdoor path**: those who attend university and those who do not may differ in ability, which also predicts later earnings. Their observed earnings difference can therefore mix the effect of university attendance with differences that would have existed without attending university.

A DAG represents causal assumptions. The data do not tell us the direction of an arrow simply because two variables are correlated. Here we assume that ability affects both university attendance and earnings. Ability need not be observed in our dataset for it to affect the comparison.

Which variables should we control for?

In this diagram, conditioning on A blocks the backdoor path. That conclusion depends on the diagram being an adequate account of confounding.

Recall two other structures from PP5000:

- A **mediator**, $D \rightarrow M \rightarrow Y$, carries part of the effect of treatment. Controlling for it generally changes the question from the total effect of treatment.
- A **collider**, $D \rightarrow C \leftarrow U \rightarrow Y$, is a common effect. Conditioning on C can open a path that was previously blocked.

💡 Pause and think

For university attendance and subsequent earnings, how would controlling for ability measured **before** university differ from controlling for occupation **after** university? Explain the causal role of each, rather than just whether it predicts earnings.

2. From the DAG to omitted-variable bias

For a simple linear illustration, suppose

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 A_i + \varepsilon_i, \quad E[\varepsilon_i \mid D_i, A_i] = 0.$$

Here β_1 is a constant causal effect of university attendance. The condition on ε_i says that, once we account for A_i , the remaining determinants of earnings have conditional mean zero. This is an assumption about the model, not something that OLS guarantees.

Parameters, fitted values, and residuals

Unhatted $\beta_0, \beta_1, \beta_2$ denote parameters of the true population relationship. If ability is observed, the fitted full regression is

$$\widehat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 D_i + \hat{\beta}_2 A_i, \quad \hat{\varepsilon}_i = Y_i - \widehat{Y}_i.$$

Hats denote sample estimates; a residual is not the unobserved DGP disturbance. Omitting ability gives a different regression:

$$\widehat{Y}_i^S = \hat{\beta}_0^S + \hat{\beta}_1^S D_i.$$

The superscript S means **short regression**. Its population coefficients are β_0^S, β_1^S ; its sample estimates carry hats. In particular, $\hat{\beta}_1^S$ estimates β_1^S , which need not equal the causal parameter β_1 .

Deriving the short-regression coefficient

If we omit A_i and regress Y_i on an intercept and D_i , the sample OLS slope is

$$\hat{\beta}_1^S = \frac{\sum_i (D_i - \bar{D})(Y_i - \bar{Y})}{\sum_i (D_i - \bar{D})^2}.$$

To see what this estimates, consider its population counterpart and substitute the full model:

$$\begin{aligned} \beta_1^S &= \frac{\text{Cov}(D_i, Y_i)}{\text{Var}(D_i)} \\ &= \frac{\text{Cov}(D_i, \beta_0 + \beta_1 D_i + \beta_2 A_i + \varepsilon_i)}{\text{Var}(D_i)} \\ &= \frac{\beta_1 \text{Var}(D_i) + \beta_2 \text{Cov}(D_i, A_i) + \text{Cov}(D_i, \varepsilon_i)}{\text{Var}(D_i)} \\ &= \beta_1 + \beta_2 \frac{\text{Cov}(D_i, A_i)}{\text{Var}(D_i)}. \end{aligned}$$

Here are the steps. The covariance of D_i with the constant β_0 is zero. Covariance is linear in its second argument, so β_1 and β_2 can be taken outside. The covariance of D_i with itself is its variance. Finally, $E[\varepsilon_i | D_i, A_i] = 0$ implies $\text{Cov}(D_i, \varepsilon_i) = 0$. The remaining covariance between D_i and A_i need not be zero.

The notation distinguishes the **causal parameter** β_1 , the **population short-regression coefficient** β_1^S , and its **sample estimate** $\hat{\beta}_1^S$. Even a very large sample cannot remove the difference between β_1^S and β_1 when the omitted characteristic is correlated with treatment.

The auxiliary-regression interpretation

Write the population regression of A_i on D_i as

$$A_i = \gamma_0 + \gamma_1 D_i + v_i, \quad E[v_i] = 0, \quad \text{Cov}(D_i, v_i) = 0.$$

If ability is observed, estimating this auxiliary regression gives

$$A_i = \hat{\gamma}_0 + \hat{\gamma}_1 D_i + \hat{v}_i, \quad \widehat{A}_i = \hat{\gamma}_0 + \hat{\gamma}_1 D_i.$$

For binary D , $\gamma_0 = E[A_i | D_i = 0]$ and $\gamma_1 = E[A_i | D_i = 1] - E[A_i | D_i = 0]$. Their sample estimates are $\hat{\gamma}_0 = \bar{A}_{D=0}$ and $\hat{\gamma}_1 = \bar{A}_{D=1} - \bar{A}_{D=0}$. This is a descriptive projection of pre-existing ability on attendance, not a causal model in that direction.

Its slope is $\gamma_1 = \text{Cov}(D_i, A_i) / \text{Var}(D_i)$. Substituting into the full earnings equation gives

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 D_i + \beta_2 (\gamma_0 + \gamma_1 D_i + v_i) + \varepsilon_i \\ &= (\beta_0 + \beta_2 \gamma_0) + (\beta_1 + \beta_2 \gamma_1) D_i + (\beta_2 v_i + \varepsilon_i). \end{aligned}$$

The new disturbance is uncorrelated with D_i , so the population short-regression slope is $\beta_1^S = \beta_1 + \beta_2\gamma_1$. This auxiliary regression describes an association; it does not claim that university attendance causes pre-existing ability.

With a vector of omitted characteristics Z_i , the same reasoning gives $\beta_1^S = \beta_1 + \theta'\gamma_1$, where θ contains their earnings coefficients and each element of γ_1 is the slope from projecting one omitted characteristic on D . With several omitted characteristics, individual contributions can reinforce or offset each other.

The second term is omitted-variable bias. Two features produce it in this model: A predicts earnings ($\beta_2 \neq 0$), and university attendance is associated with A . They correspond to the two links along the confounding path.

If ability raises earnings and those who attend university have **higher** ability on average, then $\beta_2 > 0$ and $\gamma_1 > 0$: the short regression overstates the effect of university attendance. If the ability difference runs the other way, the bias is negative. The diagram alone does not establish these signs.

With a binary treatment, the formula has a simpler interpretation

For $D \in \{0, 1\}$, write $p = P(D = 1)$ with $0 < p < 1$. Since DA is zero when $D = 0$,

$$E[DA] = pE[A \mid D = 1].$$

The law of total expectation gives

$$E[A] = pE[A \mid D = 1] + (1 - p)E[A \mid D = 0].$$

Substituting into $\text{Cov}(D, A) = E[DA] - E[D]E[A]$ and collecting terms,

$$\begin{aligned}\text{Cov}(D, A) &= p(1 - p)(E[A \mid D = 1] - E[A \mid D = 0]), \\ \text{Var}(D) &= p(1 - p).\end{aligned}$$

Consequently,

$$\boxed{\beta_1^S = \beta_1 + \beta_2(E[A \mid D = 1] - E[A \mid D = 0])}.$$

The omitted-variable bias is **how much A matters for earnings multiplied by how much the groups differ in A** . A regression on an intercept and a binary treatment is just a difference in group means.

3. The same bias in potential-outcomes language

Let Y_i^0 be earnings without attending university and Y_i^1 earnings with university attendance. We observe

$$Y_i = (1 - D_i)Y_i^0 + D_iY_i^1.$$

The unobserved potential outcome is the counterfactual. Our notation follows PP5000: the superscript denotes the treatment state, not a power.

Write the same linear illustration as

$$Y_i^0 = \beta_0 + \beta_2A_i + \varepsilon_i, \quad Y_i^1 = Y_i^0 + \beta_1.$$

Substituting these potential outcomes into observed Y_i gives exactly the regression model above. Under its conditional-mean assumption,

$$\begin{aligned} E[Y_i^0 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0] \\ = \beta_2 (E[A_i \mid D_i = 1] - E[A_i \mid D_i = 0]). \end{aligned}$$

This is the OVB term we just derived. It is also **selection bias**: the difference in what the two groups would have earned without treatment.

! The connection

In this linear, constant-effect example, omitted-variable bias and selection bias are the same quantity. Those who attend university and those who do not would have earned differently even without university attendance because they differ in ability.

Regression analysis and experiments: what is the disturbance?

We can also derive the connection directly from the observed-outcome identity. Under the constant-effect assumption,

$$\begin{aligned} Y_i &= Y_i^0 + \beta_1 D_i \\ &= \underbrace{E[Y_i^0]}_{\mu_0} + \beta_1 D_i + \underbrace{(Y_i^0 - E[Y_i^0])}_{\eta_i}. \end{aligned}$$

The intercept $\mu_0 = E[Y_i^0] = \beta_0 + \beta_2 E[A_i]$ is mean untreated potential earnings, and η_i is the part of untreated potential earnings that differs from that mean. Defining a mean-zero disturbance does not make its conditional mean zero. In particular,

$$\begin{aligned} E[Y_i \mid D_i = 1] &= \mu_0 + \beta_1 + E[\eta_i \mid D_i = 1], \\ E[Y_i \mid D_i = 0] &= \mu_0 + E[\eta_i \mid D_i = 0]. \end{aligned}$$

Subtracting the two equations gives the treatment effect plus the difference in the disturbance's conditional means. Since the same constant $E[Y_i^0]$ cancels,

$$\begin{aligned} E[\eta_i \mid D_i = 1] - E[\eta_i \mid D_i = 0] \\ = E[Y_i^0 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0]. \end{aligned}$$

In the linear model $Y_i^0 = \beta_0 + \beta_2 A_i + \varepsilon_i$, the centred disturbance is $\eta_i = \beta_2 (A_i - E[A_i]) + \varepsilon_i$. Its difference in conditional means is therefore $\beta_2 \gamma_1$, the OVB term. Centring the disturbance changes the intercept, not this group difference.

Allowing treatment effects to differ

Potential outcomes do not require the constant-effect model. Adding and subtracting $E[Y_i^0 \mid D_i = 1]$ gives the general identity

$$\begin{aligned} E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] \\ = \underbrace{E[Y_i^1 - Y_i^0 \mid D_i = 1]}_{\text{ATT}} \\ + \underbrace{E[Y_i^0 \mid D_i = 1] - E[Y_i^0 \mid D_i = 0]}_{\text{selection bias}}. \end{aligned}$$

To see the add-and-subtract step explicitly,

$$\begin{aligned}
& E[Y_i | D_i = 1] - E[Y_i | D_i = 0] \\
&= E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 0] \\
&= E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1] \\
&\quad + E[Y_i^0 | D_i = 1] - E[Y_i^0 | D_i = 0].
\end{aligned}$$

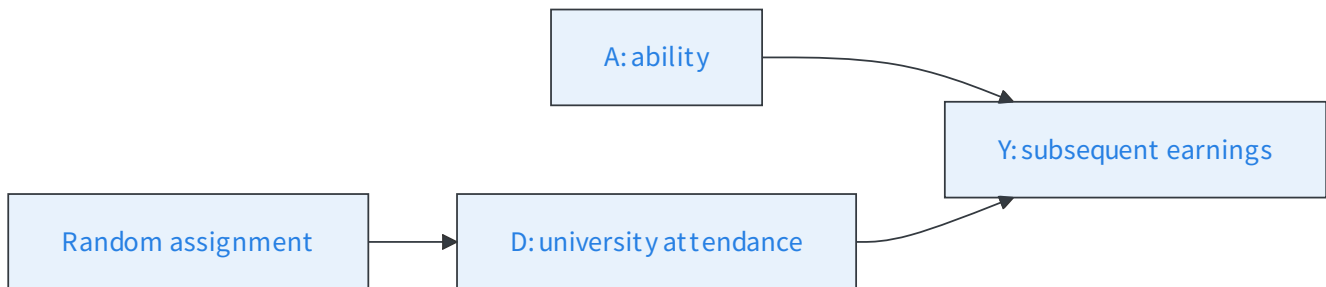
The **average treatment effect on the treated (ATT)** describes the effect for those who receive treatment. The **average treatment effect (ATE)** is $E[Y_i^1 - Y_i^0]$. They need not coincide when people select into treatment and effects vary.

4. What random assignment changes

First suppose assignment and receipt coincide. Under simple random assignment,

$$D_i \perp (Y_i^0, Y_i^1).$$

As a thought experiment, suppose university attendance itself could be randomly assigned. Assignment would replace the process through which ability influenced attendance:



Ability would still predict earnings, but it would no longer determine attendance. Randomisation also breaks selection on **unobserved** baseline determinants of potential outcomes.

Independence implies that the population selection-bias term is zero and that ATT equals ATE. For a fixed experimental sample, repeated random assignment makes the difference-in-means estimator unbiased for the sample average treatment effect.

$$\hat{\beta}_1^S = \bar{Y}_{D=1} - \bar{Y}_{D=0}.$$

Under random assignment, $\beta_1^S = \beta_1$ in the constant-effect model. The estimated short and full regressions can still differ in a finite sample.

This does **not** mean that every realised assignment produces identical groups, or that $\hat{\beta}_1^S$ equals the true effect in our one sample. Sampling and assignment uncertainty remain. Random assignment is also distinct from random sampling: an experiment need not be a representative sample of the population.

Designing the experiment

Before collecting data, specify the policy question, treatment and comparison, primary outcomes and their timing, effect sizes of policy interest, the unit of randomisation, and any planned subgroup comparisons. Randomising individuals and randomising schools or villages are different designs; the analysis must reflect the design used.

Simple randomisation leaves baseline balance to chance in a particular assignment. Stratified or blocked randomisation assigns treatment within pre-defined groups, such as regions or baseline attainment categories. Fixing treatment numbers within blocks controls their representation in the arms and can improve precision. It does not guarantee identical values of every characteristic, and differing assignment probabilities require appropriate analysis. Including region controls afterwards does not establish that assignment was blocked by region.

Pre-registration and the pre-analysis plan

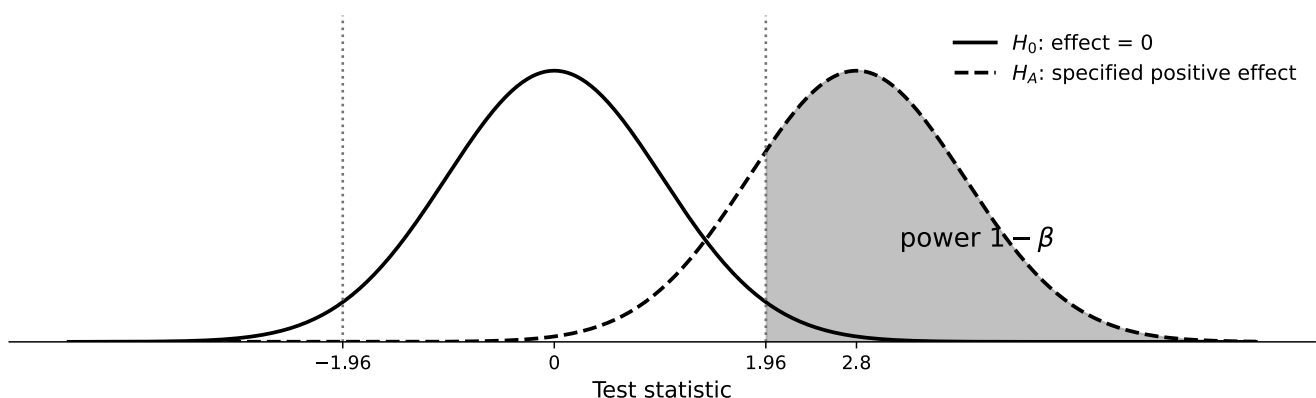
Registering an experiment records its design before the results are known. A **pre-analysis plan** specifies the primary outcomes, regressions, subgroup comparisons, sample size and its justification, and other analysis choices in advance. With many outcomes and specifications available, a plan helps distinguish anticipated tests from findings that emerged through searching.

The power calculation supplies the justification for the planned sample size: **how many observations do we need to detect an effect of policy interest?** The calculation and its assumptions belong in the plan. Work through the calculations below as part of making that design decision.

Prespecification does not prevent every implementation difficulty. Changes should be explained and dated where possible, with the original plan and relevant robustness checks available for comparison. We will examine a concrete example in the vaccination study.

Statistical power and sample-size planning

Power is the probability of rejecting the null when a specified alternative is true. It equals $1 - \beta$, where β is the probability of a Type II error at that alternative. The alternative must be specified: the same design generally has more power against a large effect than a small one.



The diagram uses a two-sided five per cent test, with normal critical values approximately -1.96 and 1.96 . Shading under the alternative in the rejection regions represents power. It does not represent the probability that the alternative hypothesis is true after seeing the data.

From power to the minimum detectable effect

For independent observations with common outcome variance σ^2 , a normal-approximation planning formula is

$$\text{MDE} = (z_{1-\alpha/2} + z_{1-\beta}) \sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}.$$

Here α is the significance level, $1 - \beta$ is the desired power, and n_T, n_C are the arm sizes. The MDE is the effect size the study is designed to detect with that power, under the planning assumptions. It is not a threshold below which significance is impossible.

Greater sample size or lower outcome variance reduces the MDE. Higher target power or a more stringent significance level raises it for a fixed sample. Covariates can help if they reduce unexplained outcome variation, but a realistic planning calculation must account for how they enter the analysis.

Deriving the required sample size

With n people in each arm,

$$\Delta = (z_{1-\alpha/2} + z_{1-\beta}) \sqrt{\frac{2\sigma^2}{n}}.$$

Squaring and rearranging gives

$$n = 2(z_{1-\alpha/2} + z_{1-\beta})^2 \frac{\sigma^2}{\Delta^2}.$$

For $\sigma = 4$, a target effect of 0.2 standard deviations means $\Delta = 0.8$. Using the rounded normal quantiles 1.96 and 0.84 for a two-sided five per cent test and 80% power gives

$$n = 2(1.96 + 0.84)^2 \frac{16}{0.64} \simeq 392 \text{ per arm.}$$

Using 90% power gives roughly 525 per arm. Holding the target effect at 0.8 and halving the standard deviation to 2 reduces the approximate requirement to 98 per arm at 80% power. Those calculations use rounded quantiles; use full-precision quantiles and round up when choosing a sample size:

```
import math
import scipy.stats as stats

sigma = 4
effect = 0.8
alpha = 0.05
power = 0.80
z_significance = stats.norm.ppf(1 - alpha / 2)
z_power = stats.norm.ppf(power)
n_per_arm = math.ceil(
    2 * (z_significance + z_power)**2 * sigma**2 / effect**2
)
print(f"Required sample using full-precision quantiles: {n_per_arm} per arm")
```

Required sample using full-precision quantiles: 393 per arm

The result is 393 per arm. The one-person difference from the slide calculation is rounding, not a different power formula. This is an illustrative independent-observation calculation, not a reconstruction of the vaccination study's original sample-size choice. With binary outcomes, the variance itself depends on the probabilities in the arms, which a tailored power calculation should incorporate.

Clustered designs

When schools or villages are randomised, outcomes within a group may be correlated. With equal group size m and intracluster correlation ρ , a simple variance inflation factor is $1 + (m - 1)\rho$. This illustrates why adding independent clusters can be more valuable than adding people within existing clusters. Cluster-level assignment also affects inference; individual-level HC1 standard errors do not account for within-cluster dependence.

5. Balance and the selection-bias connection

A balance table compares **pre-treatment** characteristics across experimental arms. Think back to

$$\beta_2 (E[A \mid D = 1] - E[A \mid D = 0]) .$$

Ability itself may be unobserved. We can nevertheless examine realised differences in observed baseline characteristics, denoted X . We should pay particular attention to characteristics likely to predict the untreated outcome. With several characteristics, both their differences and their relationships with the outcome matter.

Read a balance table by asking:

1. How large are the differences in meaningful units?
2. Which characteristics are likely to predict the outcome?
3. Are the differences plausible under the actual randomisation procedure?
4. What remains unobserved?

A standardised difference divides the difference in means by a measure of the within-group standard deviation, helping us compare variables measured in different units. A significance test asks a different question and depends on sample size. With many comparisons, some small p-values can occur by chance.

Observed balance is a diagnostic, not proof that randomisation was implemented correctly. Nor do insignificant differences prove equivalence. **The assignment mechanism supplies the justification for balance on unobservables in expectation.** Balance on observed characteristics alone would not establish this in an observational study.

6. Policy application: paying people to vaccinate

Campos-Mercade et al. study a payment of SEK 200, approximately US\$24, conditional on vaccination. The experiment took place in Sweden in 2021; its main analysis has 8,286 participants aged 18–49. Participants were assigned individually, and vaccination outcomes were linked to administrative records.

The six conditions are:

Condition	What distinguishes it?
Control	Encouragement, appointment information, and reminders
Incentives	An offer of SEK 200 conditional on timely vaccination
Social impact	Consider people who benefit from one's vaccination
Argument	Formulate arguments in favour of vaccination
Information	A quiz providing vaccine safety and effectiveness information
No reminders	Omits appointment information and reminder emails

The four active interventions build on the encouragement in the control condition. Notice that **control is not “nothing”**: the payment comparison measures the additional effect of the incentive package over that control condition. The no-reminders condition is a separate arm and must not be silently folded into control.

Two outcomes are central: stated vaccination intentions and recorded vaccination uptake. Intentions were elicited **after** assignment; they are an outcome, not a baseline balance variable or an appropriate baseline control for the total effect on uptake.

The published adjusted payment effect on uptake is about 4.2 percentage points. Interpreting it requires specifying the comparison, outcome, time window, and uncertainty. A statistically insignificant nudge estimate does not prove its effect is exactly zero. Nor does “significant here, insignificant there” itself establish that two effects differ.

Registration and the analysis plan

The plan, dated 27 May 2021, sets out the treatments, outcomes, regressions, and controls. Survey collection began on 28 May. The plan provides a record of what was intended before the results were known; it does not remove the need for judgement when implementation differs from expectations.

There is a concrete comparison to make here. The plan defined uptake within 30 days of **vaccine availability**. Supplementary Section 1.2.2 explains why the main analysis instead uses 30 days after **survey completion**, and states that this decision preceded linkage to vaccination records. Regional rollout dates were difficult to measure precisely. The supplement also examines alternative definitions.

Discuss in class

What makes this change more or less persuasive? Distinguish the reason for the change, when the decision was made, how it was disclosed, and whether alternative definitions support the conclusion.

7. The linear probability model

When the dependent variable is a 0/1 outcome, an OLS regression is called a **linear probability model (LPM)**. Since

$$E[Y_i | X_i] = P(Y_i = 1 | X_i) = p_i,$$

the model approximates the conditional probability with a linear function of its regressors. A coefficient of 0.04 on an assignment indicator represents a difference of **4 percentage points**, not a 4 per cent relative increase.

With just an intercept and one binary treatment, OLS fits the two observed group means exactly. With additional regressors, a linear model can predict outside $[0, 1]$; it is a model for interpreting average differences, not a guarantee that every individual prediction is a valid probability.

Why heteroskedasticity arises

For a Bernoulli outcome, $Y_i^2 = Y_i$. Therefore

$$\text{Var}(Y_i | X_i) = E[Y_i^2 | X_i] - (E[Y_i | X_i])^2 = p_i - p_i^2 = p_i(1 - p_i).$$

If the conditional mean is correctly specified, subtracting that mean gives

$$\text{Var}(u_i | X_i) = p_i(1 - p_i).$$

So the disturbance variance generally changes as the probability changes. At $p_i = 0.5$ it is 0.25; at $p_i = 0.9$ it is 0.09. This is **heteroskedasticity**. It affects the usual constant-variance standard-error calculation.

Use heteroskedasticity-consistent standard errors for the LPM. These change the estimated uncertainty, including confidence intervals and p-values, **not the OLS coefficients**. They do not repair confounding or an inappropriate comparison group.

```
import statsmodels.formula.api as smf

model = smf.ols("outcome ~ assigned_treatment", data=df).fit()
```

```
cov_type="HC1", use_t=True
)
```

HC1 selects a heteroskedasticity-consistent covariance estimate with a degrees-of-freedom correction. **use_t=True** uses t-based inference, as in this paper. The exercise asks you to compare this with the default constant-variance calculation.

Why include baseline controls in an experiment?

Randomisation provides identification. Baseline variables that predict the outcome can help improve precision; they can also account for chance baseline differences in the realised sample. Covariate adjustment does not guarantee greater precision in every setting, and its specification should be chosen thoughtfully. Here we reproduce the authors' prespecified controls rather than search for the combination that makes the treatment significant.

8. Internal and external validity

Internal validity: does the comparison identify the causal effect for the people studied, under the conditions of the experiment?

External validity: how far can that effect be generalised to other people, places, times, or implementations?

Table S3 compares the experimental arms with one another. Table S4 compares the experimental sample with the Swedish population aged 18–49. These answer different questions. A sample can support a credible experimental comparison while differing from the population to which a policymaker wants to apply it.

Differences in sample composition matter especially when effects vary with those characteristics. Similar demographics alone cannot establish that the same effect would hold under a different provider, payment, vaccination environment, or stage of the rollout. Read the footnotes of Table S4: some measures are not defined identically in the two sources.

9. What can complicate an experiment?

Attrition and missing outcomes

Randomisation balances groups in expectation at assignment. If observation at follow-up depends on assignment and potential outcomes, restricting analysis to observed outcomes can reintroduce selection. Similar attrition rates do not guarantee similar types of people have been lost. Administrative records can help with outcome measurement, but linkage and exclusion rules still deserve scrutiny.

Assignment versus receipt

If some people do not take up an assigned programme, comparing them by **assignment** estimates an **intention-to-treat (ITT)** effect. Comparing those who actually take it up with those who do not can reintroduce selection. We will study how assignment can serve as an instrument later in the module.

In the vaccination study, the intervention is the **offer of an incentive**; vaccination is the outcome. Comparing those who received payment with those who did not would select on the very behaviour the offer was intended to change.

SUTVA

The potential-outcomes notation we have used assumes a well-defined treatment and no interference between units: one person's outcome does not depend on someone else's assignment. These ideas are commonly grouped under the **stable unit treatment value assumption (SUTVA)**.

For this example, distinguish the effect of offering a payment on a person's vaccination from the effect of vaccination on other people's infection risk. The latter plainly raises spillovers. The former could also involve interference if participants influence each other's vaccination decisions. The relevance of interference depends on the outcome and the assignment being studied.

10. Further issues in experimental analysis

Covariate adjustment: Freedman and Lin

Conventional OLS adjustment can introduce small-sample bias and need not improve precision when effects are heterogeneous (Freedman, 2008). Lin (2013) considers an adjustment that includes interactions between treatment and demeaned baseline covariates:

$$Y_i = \beta_0 + \beta_1 D_i + (X_i - \bar{X})' \theta + D_i (X_i - \bar{X})' \lambda + u_i.$$

Under the relevant regularity conditions it does not reduce asymptotic precision relative to an unadjusted difference in means. Robust standard errors remain appropriate. This is a result about adjustment, not a reason to replace the specification we have been asked to replicate in Table S5.

Randomisation inference

Under the sharp null that treatment affects nobody, every unit's observed outcome would remain the same under every assignment. Reassign treatment according to the actual experimental procedure, recompute the statistic, and compare the observed statistic with that distribution. Enumeration provides an exact randomisation calculation; simulation provides a Monte Carlo approximation. A sharp null of no individual effects is stronger than a zero average effect when effects differ across people.

Multiple testing

If twenty true nulls are each tested at level 0.05, the expected number of false rejections is one. For independent tests, the probability of at least one false rejection is $1 - 0.95^{20} \simeq 0.64$. Dependence changes that latter calculation. Familywise-error procedures address the chance of any false rejection; false-discovery-rate procedures address the expected fraction of rejections that are false. The family of outcomes or comparisons must be defined thoughtfully. A pre-analysis plan helps establish which tests were anticipated but does not make multiplicity disappear.

Attrition bounds and spillover designs

When missing outcomes threaten identification, Lee (2009) bounds use a monotonicity assumption about observation under assignment to trim the group with higher observation rates. The resulting bounds concern the effect for people who would be observed under either assignment. They are not assumption-free estimates for everyone initially assigned.

Interference calls for a design that reflects the question. Partial-population and saturation designs can use variation in one's own assignment and others' assignment to distinguish direct effects from spillovers. At larger scales, prices or labour-market conditions may also change, making scale part of the external-validity question.

After class

Complete the [ungraded vaccination exercise](#): reconstruct the pooled nudge indicator, replicate Table S3 and Table S5, and interpret Table S4. Write explanations alongside your code. A table that matches numerically is useful only if you can explain the comparison it represents.

Reading and sources

The papers below are cited in these notes or the accompanying slides. Textbook chapters are listed in the [textbook reading guide](#).

- Allcott, H. (2015). “Site Selection Bias in Program Evaluation.” *Quarterly Journal of Economics* 130(3): 1117–1165.
- Campos-Mercade, P., A. N. Meier, F. H. Schneider, S. Meier, D. Pope, and E. Wengström (2021). “Monetary incentives increase COVID-19 vaccinations.” *Science* 374(6569): 879–882.
- Freedman, D. A. (2008). “On regression adjustments to experimental data.” *Advances in Applied Mathematics* 40(2): 180–193.
- Harrison, G. W., and J. A. List (2004). “Field Experiments.” *Journal of Economic Literature* 42(4): 1009–1055.
- Lee, D. S. (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects.” *Review of Economic Studies* 76(3): 1071–1102. (Link to author manuscript.)
- Lin, W. (2013). “Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique.” *Annals of Applied Statistics* 7(1): 295–318.
- Vivalti, E. (2020). “How Much Can We Generalize From Impact Evaluations?” *Journal of the European Economic Association* 18(6): 3045–3089.

Study materials

- [Supplementary materials](#), especially Sections 1.2.2 and 2.2–2.3.
- Pre-analysis plan and trial history; direct plan.
- Original data and replication code.