Review Session 6

RCTs and Conditional Independence

Ben Berger

March 3, 2023

Announcement

Next week's review session will be held on Wednesday 3/8 3:00-4:15.

Randomized Control Trials

- Randomly assign treatment to participants.

$$\text{Independence: } \{Y_{0i}, Y_{1i}\} \perp D_i$$

- Pros:
  - If properly administered, RCTs capture the average causal effect of a treatment.
- Cons:
  - Expen$ive.
  - Ethical dilemmas.
- How does random assignment guarantee causality?
  - It makes treatment assignment independent of potential outcomes. Therefore, selection bias $= 0$.
  - For example, by randomly assigning vaccinations in a clinical trial, we make it so that vaccination is uncorrelated with unobserved factors like willingness to social distance.
  - This helps achieve ***internal validity*** so that estimates reflect a true causal relationship.

Non-Compliance

- Even in RCTs there are some issues...
- **Non-compliance** is when participants do not actually take the treatment they are assigned.
- Directly comparing people who actually received treatment to people who actually did not can threaten internal validity.
  - Reasons for non-compliance are correlated with treatment and may be correlated with the outcome, leading to OVB.
  - For example, I might opt out of a randomized worker training program if I expect to gain little from it, and conversely the people who remain to receive treatment may expect it to have positive impacts on them.
  - Thus, treatment is correlated with potential outcomes and independence no longer holds.
- Solutions?
  - Estimate intent to treat: the effect of being *offered* treatment.
    - ITT effect is likely attenuated towards zero.
  - Use instrumental variables. Coming soon to an API-203Z near you.

Other issues

Attrition

- Different from non-compliance: participants receive treatment but are lost to follow-up, so can't observe outcome at some point in time.
- Some attrition is expected, but is it different across treatment and control arms?
  - If so, remaining sample could differ in baseline characteristics across arms.
  - Test for difference in attrition rates across treatment/control.
  - Even if attrition rate is the same, causes of attrition could vary across arms.

Spillovers

- Violation of stable-unit treatment value assumption (SUTVA).
- Control units receive some amount of treatment.
- Minimize spillovers by randomizing treatment by clusters (e.g. schools) rather than individuals.

External Validity

External Validity: the extent to which the results of a study can be generalized to other contexts.

What do you think about the external validity of the following experiment?

> *Suppose that we are interested in the impact of job training on wages among the American labor force. An internally valid RCT finds that prisoners who were assigned to a worker training program had higher wages after release. (Made-Up Example, 2021)*

While retraining prisoners may be important, they make up only a small fraction of participants in worker training programs and are in very different circumstances from most of the labor force. Thus the external validity of the study with respect to the general efficacy of job programs may be questionable.

External Validity

What do you think about the external validity of the following experiment?

> *Intestinal helminths... infect more than one-quarter of the world's population... We evaluate a Kenyan project in which school-based mass treatment with deworming drugs was randomly phased into schools... The program reduced school absenteeism in treatment schools by one-quarter, and was far cheaper than alternative ways of boosting school participation. Deworming substantially improved health and school participation among untreated children in both treatment schools and neighboring schools... (Miguel and Kremer, 2004)*

Time and place matter a lot. Extreme flooding in 1998 due to El Niño and proximity to Lake Victoria are believed by the authors to have contributed to a high rate of infections. Thus deworming may have been unordinarily effective in this context.

External validity is hard to prove, but internally valid studies contribute to the literature, helping to build a consensus among researchers.

Pfizer-BioNTech Vaccine Study

New England Journal of Medicine Study:

> *In an ongoing multinational, placebo-controlled, observer-blinded, pivotal efficacy trial, we randomly assigned persons 16 years of age or older in a 1:1 ratio to receive two doses, 21 days apart, of either placebo or the BNT162b2 vaccine candidate ... The primary end points were efficacy of the vaccine against laboratory-confirmed Covid-19 and safety.*

What do all of these terms mean, and how to they relate to the RCTs discussed in class?

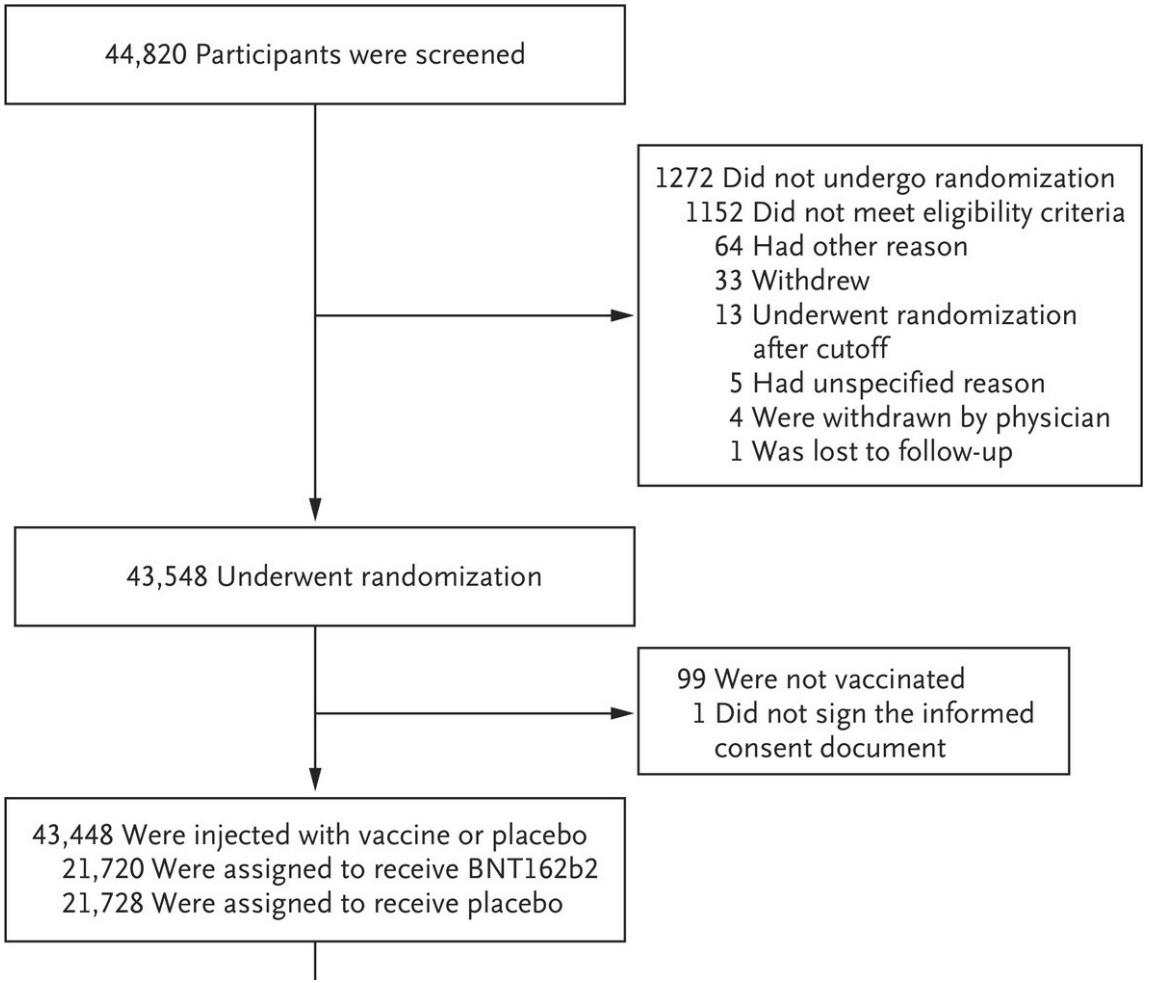
- *Placebo-controlled*: the control group is given an inactive substance rather than no treatment.
- *Blinded/masked*: treatment assignment is unknown to some parties (e.g. participants/care providers/investigators.)
- *End points*: dependent variables, usually specified in advance.

Methodology

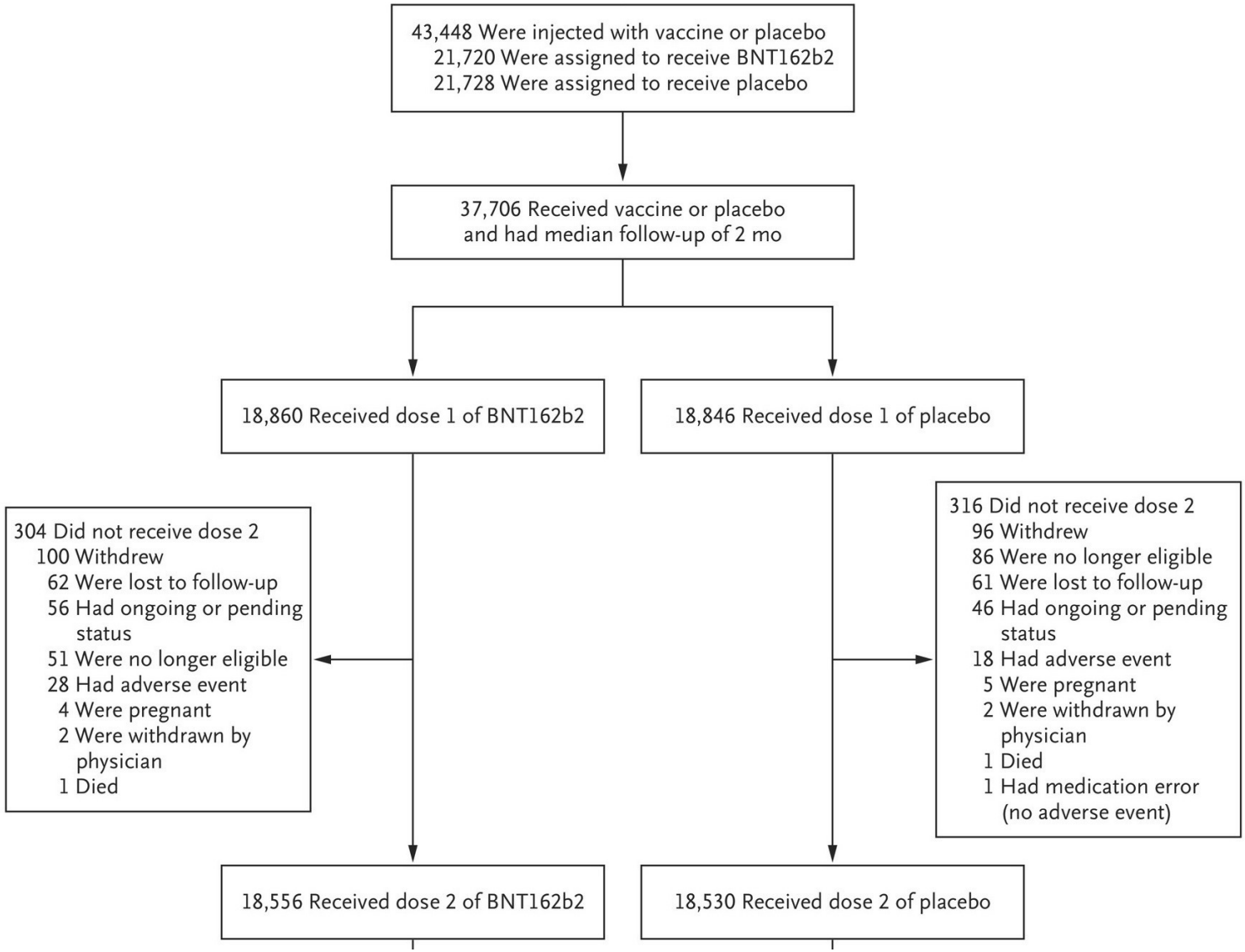
Questions:

- How does use of a placebo control relate to the treatment effect of the RCT?
  - It estimates a treatment effect relative to a different counterfactual (placebo counterfactual vs. no injection counterfactual).
- Are there any alternatives to placebo? For example, can we do better than giving a placebo to cancer patients?
  - Yes, we can give patients the "standard of care" which may be an active treatment. Frequently employed in cancer trials.
- How does trial masking mitigate unwanted sources of bias?
  - Placebo-control likely wouldn't be effective if individuals know they have the placebo.
  - Placebo-assignees may leave study or try to get switched to treatment.
  - Vaccinated patients might change their behavior (decreased social distancing) if they knew for sure they received the vaccine.

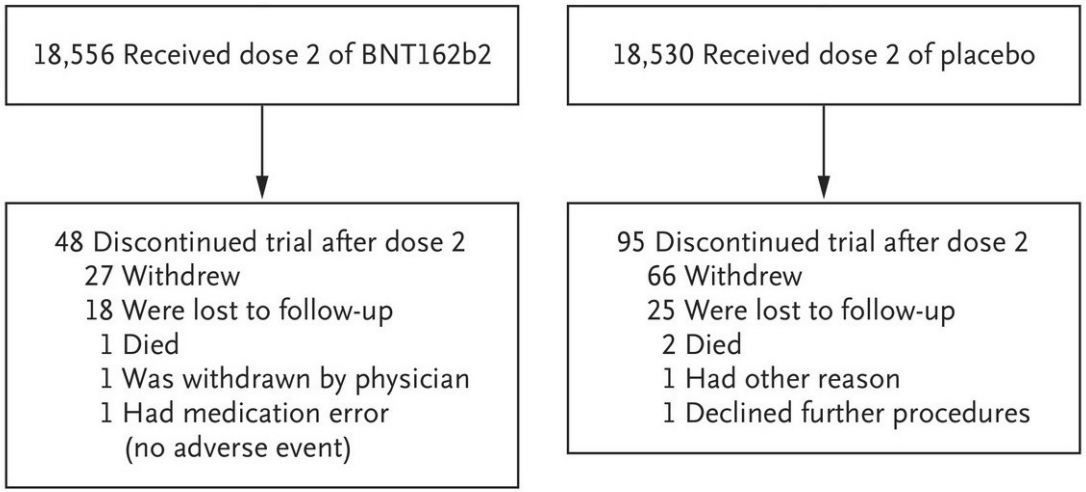
Enrollment and Randomization



Enrollment and Randomization



Enrollment and Randomization



Enrollment and Randomization

What does this figure tell us?

- Little attrition after randomization and before first dose (100 people out of >40,000).
- More attrition between doses, but similar in treatment and control groups.

Which steps of the protocol threaten the study's internal validity?

- Attrition?
  - Attrition may be concerning because it could be correlated with treatment assignment. However, attrition appears to be similar in placebo and treatment groups.
- FDA-mandated sample (participants enrolled by October 9, 2020)?
  - As long as enrollment date is not correlated with treatment assignment (it shouldn't be), this won't threaten internal validity.

Baseline Characteristics/Balance Table

Question: Did randomization work?

Table 1. Demographic Characteristics of the Participants in the Main Safety Population.*

Characteristic	BNT162b2 (N=18,860)	Placebo (N=18,846)	Total (N=37,706)
Sex — no. (%)			
Male	9,639 (51.1)	9,436 (50.1)	19,075 (50.6)
Female	9,221 (48.9)	9,410 (49.9)	18,631 (49.4)
Race or ethnic group — no. (%)†			
White	15,636 (82.9)	15,630 (82.9)	31,266 (82.9)
Black or African American	1,729 (9.2)	1,763 (9.4)	3,492 (9.3)
Asian	801 (4.2)	807 (4.3)	1,608 (4.3)
Native American or Alaska Native	102 (0.5)	99 (0.5)	201 (0.5)
Native Hawaiian or other Pacific Islander	50 (0.3)	26 (0.1)	76 (0.2)
Multiracial	449 (2.4)	406 (2.2)	855 (2.3)
Not reported	93 (0.5)	115 (0.6)	208 (0.6)
Hispanic or Latinx	5,266 (27.9)	5,277 (28.0)	10,543 (28.0)
Country — no. (%)			
Argentina	2,883 (15.3)	2,881 (15.3)	5,764 (15.3)
Brazil	1,145 (6.1)	1,139 (6.0)	2,284 (6.1)
South Africa	372 (2.0)	372 (2.0)	744 (2.0)
United States	14,460 (76.7)	14,454 (76.7)	28,914 (76.7)
Age group — no. (%)			
16–55 yr	10,889 (57.7)	10,896 (57.8)	21,785 (57.8)
>55 yr	7,971 (42.3)	7,950 (42.2)	15,921 (42.2)
Age at vaccination — yr			
Median	52.0	52.0	52.0
Range	16–89	16–91	16–91
Body-mass index‡			
≥30.0: obese	6,556 (34.8)	6,662 (35.3)	13,218 (35.1)

* Percentages may not total 100 because of rounding.
† Race or ethnic group was reported by the participants.
‡ The body-mass index is the weight in kilograms divided by the square of the height in meters.

Efficacy

Calculate efficacy using the ratio of the probability of contracting Covid-19 if vaccinated to the probability if unvaccinated.

$$
\begin{aligned}
Efficacy &= \left(1 - \frac{CasesVax/TotalVax}{CasesUnvax/TotalUnvax}\right) \times 100\% \\
&= \left(1 - \frac{8/17411}{162/17511}\right) \times 100\% \\
&= (1 - 0.05) \times 100\% \\
&= 95\%
\end{aligned}
$$

Efficacy

Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
		(N=18,198)		**(N=18,325)**		
Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.3–97.6)	>0.9999
		(N=19,965)		**(N=20,172)**		
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	9	2.332 (18,559)	169	2.345 (18,708)	94.6 (89.9–97.3)	>0.9999

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.
† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.
‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.
§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

Efficacy by Subgroup

Table 3. Vaccine Efficacy Overall and by Subgroup in Participants without Evidence of Infection before 7 Days after Dose 2.

Efficacy End-Point Subgroup	BNT162b2 (N=18,198)		Placebo (N=18,325)		Vaccine Efficacy, % (95% CI)†
	No. of Cases	Surveillance Time (No. at Risk)*	No. of Cases	Surveillance Time (No. at Risk)*	
Overall	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.0–97.9)
Age group					
16 to 55 yr	5	1.234 (9,897)	114	1.239 (9,955)	95.6 (89.4–98.6)
>55 yr	3	0.980 (7,500)	48	0.983 (7,543)	93.7 (80.6–98.8)
≥65 yr	1	0.508 (3,848)	19	0.511 (3,880)	94.7 (66.7–99.9)
≥75 yr	0	0.102 (774)	5	0.106 (785)	100.0 (–13.1–100.0)
Sex					
Male	3	1.124 (8,875)	81	1.108 (8,762)	96.4 (88.9–99.3)
Female	5	1.090 (8,536)	81	1.114 (8,749)	93.7 (84.7–98.0)
Race or ethnic group‡					
White	7	1.889 (14,504)	146	1.903 (14,670)	95.2 (89.8–98.1)
Black or African American	0	0.165 (1,502)	7	0.164 (1,486)	100.0 (31.2–100.0)
All others	1	0.160 (1,405)	9	0.155 (1,355)	89.3 (22.6–99.8)
Hispanic or Latinx	3	0.605 (4,764)	53	0.600 (4,746)	94.4 (82.7–98.9)
Non-Hispanic, non-Latinx	5	1.596 (12,548)	109	1.608 (12,661)	95.4 (88.9–98.5)
Country					
Argentina	1	0.351 (2,545)	35	0.346 (2,521)	97.2 (83.3–99.9)
Brazil	1	0.119 (1,129)	8	0.117 (1,121)	87.7 (8.1–99.7)
United States	6	1.732 (13,359)	119	1.747 (13,506)	94.9 (88.6–98.2)

* Surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.
† The confidence interval (CI) for vaccine efficacy is derived according to the Clopper–Pearson method, adjusted for surveillance time.
‡ Race or ethnic group was reported by the participants. "All others" included the following categories: American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, multiracial, and not reported.

Trial Results

Is the vaccine effective?

- Yes.

What are some shortcomings regarding internal validity?

- The study is not adequately powered to detect differences in efficacy between subgroups.
  - While no black participants who received the vaccine had evidence of infection, the lower bound on the 95% CI is 31% effectiveness.
  - Even less precise for individuals ≥ 75 yrs: cannot reject 0% efficacy.
- Attrition complicates analysis, but likely has minimal effect on results.

External validity?

- Vaccine waned in efficacy over time, necessitating booster injections.
- Vaccine was less effective against new variants.

Selection on Observables

- Outside of randomized control trials, it is often hard to believe that potential outcomes are independent of treatment status.

$$\text{CIA} : \{Y_{0i}, Y_{1i}\} \perp D_i | W_i$$

- A weaker (more plausible) assumption is that potential outcomes are independent of treatment status D_i conditional on some variables W_i. This is called the **Conditional Independence Assumption**.
  - Given some known set of characteristics, treatment is as good as randomly allocated.
  - Example: an individual's Powerball winnings are as good as random conditional on the number of tickets they buy.
- In this case we can calculate average treatment effects conditional on W_i and then aggregate to a single average treatment effect.
- Implement with regression by controlling for W_i as covariates or using matching.

Omitted Variable Bias

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \text{ (Long Regression)}$$
$$Y = \alpha_0 + \alpha_1 X_1 + v \text{ (Short Regression)}$$
$$X_2 = \gamma_0 + \gamma_1 X_1 + w \text{ (Auxillary Regression)}$$

Omitted Variable Bias $= \alpha_1 - \beta_1 = \beta_2 \times \gamma_1$.

Both the magnitude and sign of bias depend on (1) the relationship between the omitted variable and the included variable and (2) the relationship between the omitted variable on the outcome.

- Bias and the true relationship have same sign $\rightarrow$*overstate* the effect.
- Opposite signs
  - Bias is smaller than the true effect $\rightarrow$*understate* the effect.
  - Bias is larger than the true effect $\rightarrow$*flip the sign*.

Omitted Variable Bias: Example

The built-in R dataset `airquality` contains the following variables:

- `Temp` – Temperature in degrees Fahrenheit
- `Ozone` – Ozone concentration in parts per billion
- `Wind` – Wind speed in miles per hour

Using R, estimate the following regressions:

$$Temperature_i = \alpha_0 + \alpha_1 Ozone_i + v_i$$

$$Temperature_i = \beta_0 + \beta_1 Ozone_i + \beta_2 Wind_i + u_i$$

Omitted Variable Bias: Example

```
library(tidyverse)
library(fixest)
# Estimate Regressions
bivariate <- feols(Temp ~ Ozone, airquality, vcov = "hetero")
multivariate <- feols(Temp ~ Ozone + Wind, airquality, vcov = "hetero")

summary(bivariate)
```

```
OLS estimation, Dep. Var.: Temp
Observations: 116
Standard-errors: Heteroskedasticity-robust
             Estimate Std. Error  t value   Pr(>|t|)
(Intercept) 69.41072   1.300595 53.36842  < 2.2e-16 ***
Ozone        0.20081   0.029406  6.82894 4.3971e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 6.75987   Adj. R2: 0.483213
```

Omitted Variable Bias: Example

```
summary(multivariate)
```

```
OLS estimation, Dep. Var.: Temp
Observations: 116
Standard-errors: Heteroskedasticity-robust
             Estimate Std. Error  t value   Pr(>|t|)
(Intercept) 74.180348   2.935869 25.26692  < 2.2e-16 ***
Ozone        0.176149   0.033401  5.27381 6.5018e-07 ***
Wind        -0.378289   0.208574 -1.81369 7.2379e-02 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 6.67375   Adj. R2: 0.49184
```

Omitted Variable Bias: Example

Bivariate	Multivariate
$\hat{\alpha}_0 = 69.4$	$\hat{\beta}_0 = 74.2$
$\hat{\alpha}_1 = 0.201$	$\hat{\beta}_1 = 0.176$
	$\hat{\beta}_2 = -0.378$

What is the sign of omitted variable bias in short regression?

Positive. $\hat{\alpha}_1 > \hat{\beta}_1$*, so the short regression estimate is positively biased.*

If wind speed and temperature are negatively correlated ($\hat{\beta}_2 < 0$), how is the sign of OVB positive?

Wind speed and ozone concentration are also negatively correlated. Negative times negative equals positive, so the OVB is positive.

Propensity Score Matching

- Matching is another method used to condition on variables.
  - Simple in concept: pick controls that are similar to treated observations on observables.
  - Complex in practice: many decisions to make.
- Exact matching is ideal in theory because it exactly balances characteristics across treatment and control groups. However, it is plagued by the curse of dimensionality: too few individuals have exact matches on all dimensions.
- Propensity score matching is a popular alternative because it boils the matching problem down to one dimension and makes it easier to match control observations to treated observations.

Propensity Score Matching

1. Estimate a predictive model of treatment. Calculate fitted values of treatment (predicted propensity scores) using the predictive model.
   - You can try different models for this step. You can use a linear probability model, but binary outcome models like logit and probit are typically more popular.
2. For each individual in the treatment group, match them (potentially with replacement) to the individual in the control group with the closest propensity score.
   - You may want to match only within a certain caliper, say within 0.05 more or less.
3. Using the matched sample, test the difference in means of the outcome variable.

Propensity Score Matching

Matching Identification Assumption 1: Selection on Propensity Score

Conditional on a propensity score, treatment assignment is unrelated to an observation's potential outcomes.

Example:
Suppose we were interested in the impact of a scholarship on college graduation rates. The factors that influenced whether a candidate received the scholarship were high school GPA, extracurricular activities, and a written assessment of the candidate by an alumnus who interviewed the candidate.

If we calculate a propensity score using GPA and extracurriculars, what source of bias remains?

Positive alumni assessments may identify qualities of the candidates that make them more likely to succeed in college. Because this isn't accounted for in the propensity score, the estimate will be biased.

Propensity Score Matching

Matching Identification Assumption 2: Common Support

For any set of covariates X (and correspondingly for any propensity score $p(X)$) in the treatment group, there must be at least one unit in the donor pool.

There must be some overlap in the treated and untreated groups in our sample. We can verify this in the data.

Matching Exercise

Let's use the dataset `smoking.dta` and propensity score matching to evaluate the impact of smoking (`smoker`) on psychological distress (`psyc_distress`). Load the data:

```
library(tidyverse)
library(haven)
library(MatchIt)

smoking <- read_dta("smoking.dta")
```

Matching Exercise

- Our goal is to create a control group of non-smokers who are similar to smokers on observable dimensions.
- Package: `MatchIt`; Function: `matchit()`
  - Matching metric?
  - Match with replacement?
  - Exact matches?
  - Caliper?
- Steps:
  1. Run the matching algorithm.
  2. Verify covariate balance.
  3. Create the matched dataset.
  4. Run a regression to estimate the treatment effect.

Step 1: Matching

Use `matchit` for nearest neighbor matching on a propensity score estimated using all control variables. Match each smoker to 2 non-smokers with replacement (i.e. non-smokers can be matched to multiple smokers). Also match exactly on sex and high school completion.

```
match <-
  matchit(smoker ~ sex + indigeneity + high_school + partnered +
            remoteness + language + risky_alcohol + age,
          data = smoking,
          exact = ~ sex + high_school, # exact match variables
          replace = TRUE, # whether to use controls multiple times
          ratio = 2 # controls to match to each treated
          )
```

Step 1: Matching

```
match
```

```
A matchit object
 - method: 2:1 nearest neighbor matching with replacement
 - distance: Propensity score
             - estimated with logistic regression
 - number of obs.: 8000 (original), 2279 (matched)
 - target estimand: ATT
 - covariates: sex, indigeneity, high_school, partnered, remoteness
```

Step 2: Verify Covariate Balance

```
summary(match, un = F, standardize = F)$sum.matched[, 1:3]
```

```
              Means Treated Means Control    Mean Diff.
distance          0.1850228     0.1847535  2.692519e-04
sex               0.4938398     0.4938398  3.830269e-15
indigeneity       0.0523614     0.0426078  9.753593e-03
high_school       0.4219713     0.4219713  3.275158e-15
partnered         0.4630390     0.4517454  1.129363e-02
remoteness        0.6119097     0.5651951  4.671458e-02
language          0.9579055     0.9712526 -1.334702e-02
risky_alcohol     0.6427105     0.6524641 -9.753593e-03
age              51.6057495    51.8886037 -2.828542e-01
```

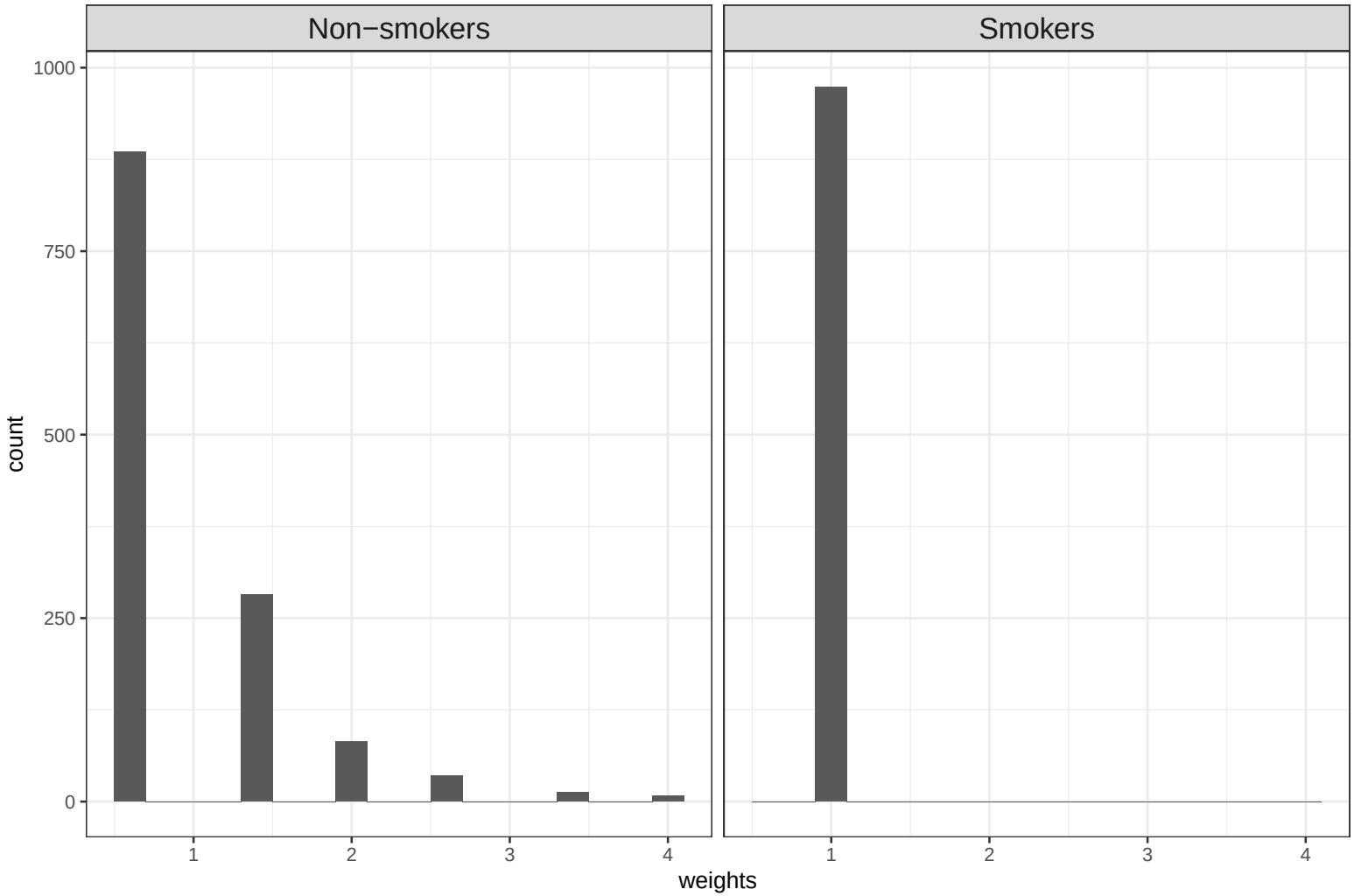
Step 3: Create Matched Data

Now use the `match.data` function to create a matched dataset. Note that this will create a new variable called `weights` that accounts for the number of times non-smokers are used as controls.

```
matched_smoking <- match.data(match)
matched_smoking %>% select(smoker, weights) %>% head()
```

```
# A tibble: 6 x 2
  smoker weights
   <dbl>   <dbl>
1      0   0.670
2      0   0.670
3      0   0.670
4      0   1.34
5      0   0.670
6      1   1
```

Step 3: Create Matched Data



Step 4: Estimate Effect with (Weighted) Regression

```
feols(psyc_distress ~ smoker,
      data = matched_smoking,
      vcov = "hetero",
      weights = ~weights)
```

```
OLS estimation, Dep. Var.: psyc_distress
Observations: 2,279
Weights: weights
Standard-errors: Heteroskedasticity-robust
             Estimate Std. Error  t value   Pr(>|t|)
(Intercept) 15.59908   0.199671 78.12374  < 2.2e-16 ***
smoker       1.71407   0.314697  5.44672 5.6826e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 6.77021   Adj. R2: 0.015012
```