"Challenges to Randomization: Noncompliance and Missing Data"

ICPSR 2023 Session 1

Jake Bowers, Ben Hansen & Tom Leavitt

June 24, 2024

Today

1. Agenda: One step away from easy to interpret experiments: non-random doses/compliance (Gerber and Green, 2012) Chapter 5, non-random missing data (Gerber and Green, 2012) Chapter 7 and the Threats module of The Theory and Practice of Field Experiments.
2. Recap: We use statistics to **infer** about unobservable counterfactual quantities (functions of potential outcomes); we can estimate unobservable averages; we can test unobservable hypotheses; we can test unobservable hypotheses about averages.
3. Questions arising from the reading or assignments or life?

Today

1. Agenda: One step away from easy to interpret experiments: non-random doses/compliance (Gerber and Green, 2012) Chapter 5, non-random missing data (Gerber and Green, 2012) Chapter 7 and the Threats module of The Theory and Practice of Field Experiments.
2. Recap: We use statistics to **infer** about unobservable counterfactual quantities (functions of potential outcomes); we can estimate unobservable averages; we can test unobservable hypotheses; we can test unobservable hypotheses about averages.
3. Questions arising from the reading or assignments or life?

Today

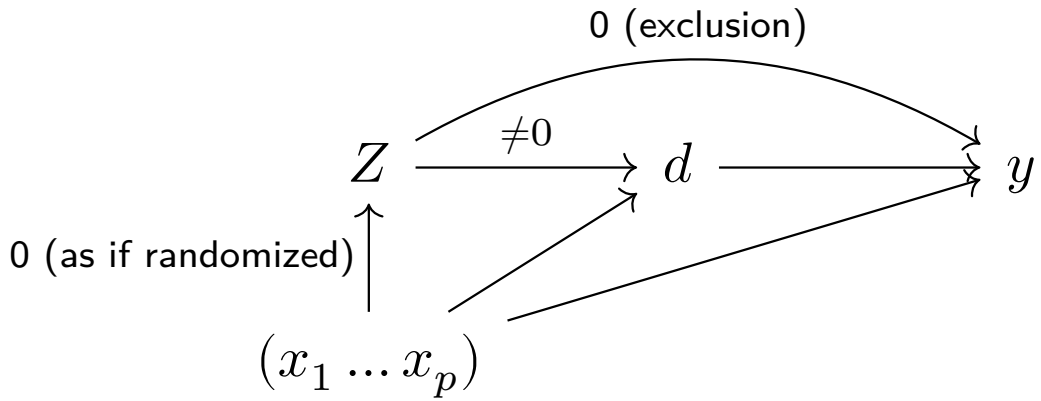
1. Agenda: One step away from easy to interpret experiments: non-random doses/compliance (Gerber and Green, 2012) Chapter 5, non-random missing data (Gerber and Green, 2012) Chapter 7 and the Threats module of The Theory and Practice of Field Experiments.
2. Recap: We use statistics to **infer** about unobservable counterfactual quantities (functions of potential outcomes); we can estimate unobservable averages; we can test unobservable hypotheses; we can test unobservable hypotheses about averages.
3. Questions arising from the reading or assignments or life?

1. Causal effects when we do not control the dose
2. Hypothesis Tests about Complier causal effects
3. Learning about causal effects when data are missing

Defining causal effects I

Imagine a door-to-door communication experiment where some houses are randomly assigned to receive a visit. Note that we now use Z and d instead of T.

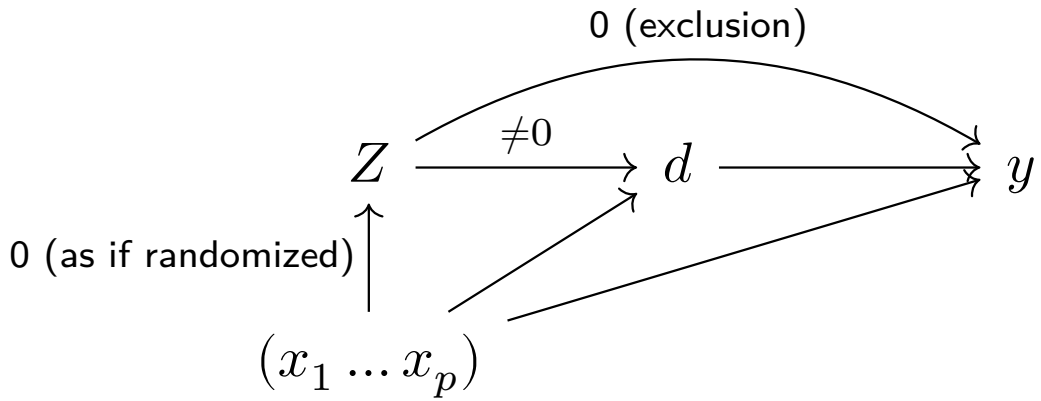
- Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- $d_{i,Z_i=1} = 1$ means that person i would open the door to have a conversation when assigned a visit.
- $d_{i,Z_i=1} = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- Opening the door is an outcome of the treatment.



Defining causal effects I

Imagine a door-to-door communication experiment where some houses are randomly assigned to receive a visit. Note that we now use Z and d instead of T.

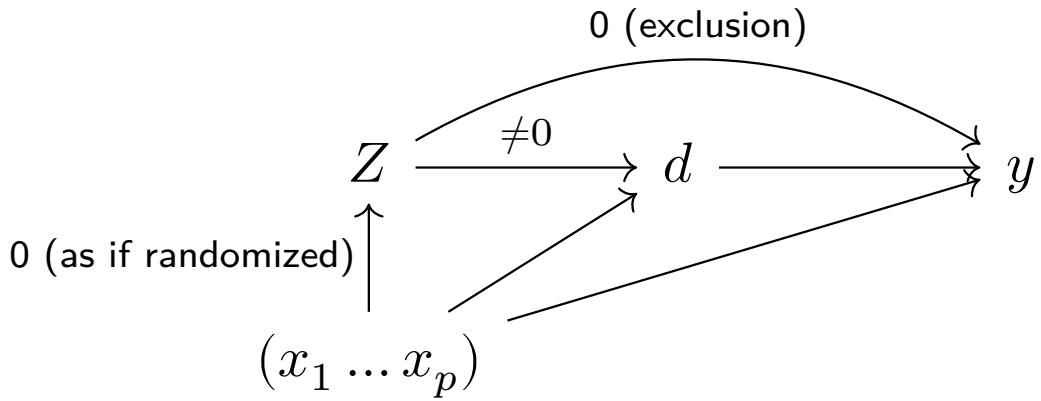
- Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- $d_{i,Z_i=1} = 1$ means that person i would open the door to have a conversation when assigned a visit.
- $d_{i,Z_i=1} = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- Opening the door is an outcome of the treatment.



Defining causal effects I

Imagine a door-to-door communication experiment where some houses are randomly assigned to receive a visit. Note that we now use Z and d instead of T.

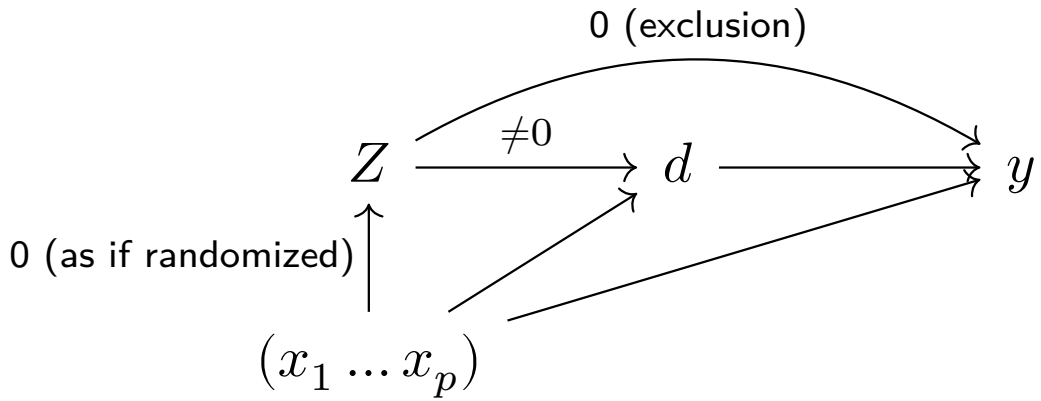
- Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- $d_{i,Z_i=1} = 1$ means that person i would open the door to have a conversation when assigned a visit.
- $d_{i,Z_i=1} = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- Opening the door is an outcome of the treatment.



Defining causal effects I

Imagine a door-to-door communication experiment where some houses are randomly assigned to receive a visit. Note that we now use Z and d instead of T.

- Z_i is random assignment to a visit ($Z_i = 1$) or not ($Z_i = 0$).
- $d_{i,Z_i=1} = 1$ means that person i would open the door to have a conversation when assigned a visit.
- $d_{i,Z_i=1} = 0$ means that person i would not open the door to have a conversation when assigned a visit.
- Opening the door is an outcome of the treatment.



Defining causal effects II

We can simplify the ways in which people get a dose of the treatment like so (where d is lower case reflecting the idea that whether you open the door when visited or not is a fixed attribute like a potential outcome).

- Y : outcome ($y_{i,Z}$ or $y_{i,Z_i=1}$ for potential outcome to treatment for person i, fixed)
- X : covariate/baseline variable
- Z : treatment assignment ($Z_i = 1$ if assigned to a visit, $Z_i = 0$ if not assigned to a visit)
- D : treatment received ($D_i = 1$ if answered door, $D_i = 0$ if person i did not answer the door) (using D here because $D_i = d_{i,1}Z_i + d_{i,0}(1 - Z_i)$)

Defining causal effects II

We can simplify the ways in which people get a dose of the treatment like so (where d is lower case reflecting the idea that whether you open the door when visited or not is a fixed attribute like a potential outcome).

- Y : outcome ($y_{i,Z}$ or $y_{i,Z_i=1}$ for potential outcome to treatment for person i, fixed)
- X : covariate/baseline variable
- Z : treatment assignment ($Z_i = 1$ if assigned to a visit, $Z_i = 0$ if not assigned to a visit)
- D : treatment received ($D_i = 1$ if answered door, $D_i = 0$ if person i did not answer the door) (using D here because $D_i = d_{i,1}Z_i + d_{i,0}(1 - Z_i)$)

Defining causal effects II

We can simplify the ways in which people get a dose of the treatment like so (where d is lower case reflecting the idea that whether you open the door when visited or not is a fixed attribute like a potential outcome).

- Y : outcome ($y_{i,Z}$ or $y_{i,Z_i=1}$ for potential outcome to treatment for person i, fixed)
- X : covariate/baseline variable
- Z : treatment assignment ($Z_i = 1$ if assigned to a visit, $Z_i = 0$ if not assigned to a visit)
- D : treatment received ($D_i = 1$ if answered door, $D_i = 0$ if person i did not answer the door) (using D here because $D_i = d_{i,1}Z_i + d_{i,0}(1 - Z_i)$)

Defining causal effects II

We can simplify the ways in which people get a dose of the treatment like so (where d is lower case reflecting the idea that whether you open the door when visited or not is a fixed attribute like a potential outcome).

- Y : outcome ($y_{i,Z}$ or $y_{i,Z_i=1}$ for potential outcome to treatment for person i, fixed)
- X : covariate/baseline variable
- Z : treatment assignment ($Z_i = 1$ if assigned to a visit, $Z_i = 0$ if not assigned to a visit)
- D : treatment received ($D_i = 1$ if answered door, $D_i = 0$ if person i did not answer the door) (using D here because $D_i = d_{i,1}Z_i + d_{i,0}(1 - Z_i)$)

Defining causal effects III

We have two causal effects of Z: $Z \to Y$ (known as δ, ITT, ITT_Y), and $Z \to D$ (known as ITT_D, p_c).
And different types of people can react differently to the attempt to move the dose with the instrument.

		$Z = 1$	
		$D = 0$	$D = 1$
$Z = 0$	$D = 0$	Never taker	Complier
	$D = 1$	Defier	Always taker

Defining causal effects IV

The $ITT = ITT_Y = \delta = \bar{y}_{Z=1} - \bar{y}_{Z=0}$.

But, in this design, $\bar{y}_{Z=1} = \bar{y}_1$ is split into pieces: the outcome of those who answered the door (Compliers and Always-takers and Defiers). Write p_C for the proportion of compliers in the study, p_A for proportion always-takers, etc... The proportions have to sum to 1. So, we have weighted averages:

$$\bar{y}_1 = (\bar{y}_1|C)p_C + (\bar{y}_1|A)p_A + (\bar{y}_1|N)p_N + (\bar{y}_1|D)p_D.$$

And $\bar{y}_0$ is also split into pieces:

$$\bar{y}_0 = (\bar{y}_0|C)p_C + (\bar{y}_1|A)p_A + (\bar{y}_0|N)p_N + (\bar{y}_0|D)p_D.$$

Defining causal effects V

So, the ITT itself is a combination of the effects of Z on Y within these different groups.

People who are compliers tend to be different types of people than people who are always takers: comparisons across types would raise questions about how to interpret the results — interpretations that would focus more on differences in types than in differences caused by Z.

But, we can still estimate it because we have unbiased estimators of $\bar{y}_1$ and $\bar{y}_0$ within each type.

Learning about the ITT I

First, let's learn about the effect of the policy itself.

Let's assume we have no defiers ($p_D = 0$). The we can write the ITT more simply.

$$\bar{y}_1 = (\bar{y}_1|C)p_C + (\bar{y}_1|A)p_A + (\bar{y}_1|N)p_N$$

$$\bar{y}_0 = (\bar{y}_0|C)p_C + (\bar{y}_0|A)p_A + (\bar{y}_0|N)p_N$$

Learning about the ITT II

First, let's learn about the effect of the policy itself. We assume no defiers ($p_D = 0$), which allows us to write the ITT more simply.

$$\begin{aligned} ITT =& \bar{y}_1 - \bar{y}_0 \\ =& ((\bar{y}_1|C)p_C + (\bar{y}_1|A)p_A + (\bar{y}_1|N)p_N) - \\ & ((\bar{y}_0|C)p_C + (\bar{y}_0|A)p_A + (\bar{y}_0|N)p_N) \end{aligned}$$

collecting each type together — to have an ITT for each type

$$\begin{aligned} =& ((\bar{y}_1|C)p_C - (\bar{y}_0|C)p_C) + ((\bar{y}_1|A)p_A - (\bar{y}_0|A)p_A) + \\ & ((\bar{y}_1|N)p_N - (\bar{y}_0|N)p_N) \\ =& \underbrace{((\bar{y}_1|C) - (\bar{y}_0|C))}_{\text{ITT among Compliers}} p_C + \\ & \underbrace{((\bar{y}_1|A) - (\bar{y}_0|A))}_{\text{ITT among Always-Takers}} p_A + \underbrace{((\bar{y}_1|N) - (\bar{y}_0|N))}_{\text{ITT among Never-Takers}} p_N \end{aligned}$$

Learning about the ITT III

$$
\begin{aligned}
ITT =& \bar{y}_1 - \bar{y}_0 \\
=& ((\bar{y}_1|C)p_C + (\bar{y}_1|A)p_A + (\bar{y}_1|N)p_N) - \\
& ((\bar{y}_0|C)p_C + (\bar{y}_0|A)p_A + (\bar{y}_0|N)p_N) \\
=& ((\bar{y}_1|C)p_C - (\bar{y}_0|C)p_C) + ((\bar{y}_1|A)p_A - (\bar{y}_0|A)p_A) + \\
& ((\bar{y}_1|N)p_N - (\bar{y}_0|N)p_N) \\
=& ((\bar{y}_1|C) - (\bar{y}_0|C))p_C + ((\bar{y}_1|A) - (\bar{y}_0|A))p_A + \\
& ((\bar{y}_1|N) - (\bar{y}_0|N))p_N \\
=& (\text{ITT among compliers})(\text{proportion of compliers}) + \\
& (\text{ITT among always takers})(\text{proportion of always takers}) + \ldots
\end{aligned}
$$

Learning about the ITT IV

And, if the effect of the dose can only occur for those who open the door, and you can only open the door when assigned to do so then:

$$((\bar{y}_1|A) - (\bar{y}_0|A))p_A = 0 \text{ and } ((\bar{y}_1|N) - (\bar{y}_0|N))p_N = 0$$

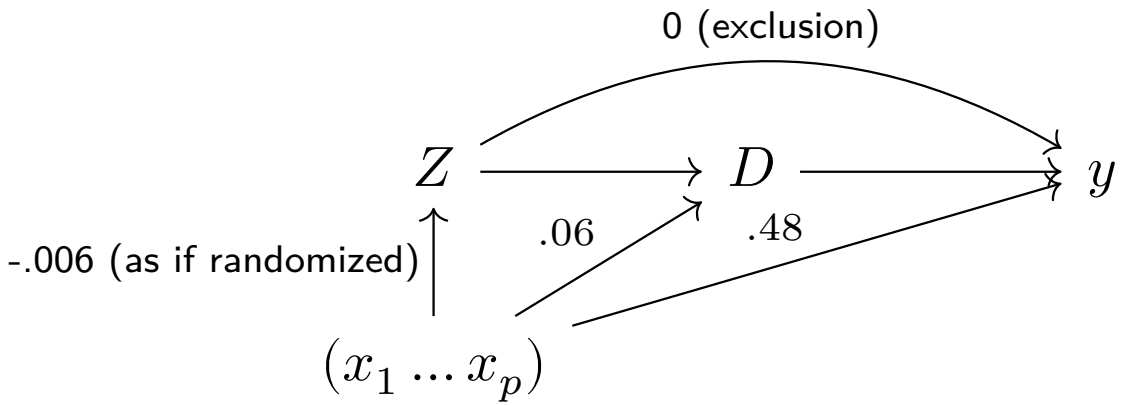
And so, under these assumptions, the ITT is a simple function of the ITT among compliers and the proportion of compliers.

$$ITT = ((\bar{y}_1|C) - (\bar{y}_0|C))p_C = (CACE)p_C.$$

The complier average causal effect I

If we want to can learn about the the causal effect of answering the door and having the conversation why not just compare people who answer the door to people who do not?

The problem with this "as-treated" or "per-protocol" comparison is that this comparison is confounded by x: a simple $\bar{Y}|D=1 - \bar{Y}|D=0$ comparison tells us about differences in the outcome due to x in addition to the difference caused by D. (Numbers below from some simulated data)



The complier average causal effect II

In actual data:

```
with(dat, cor(Y, x)) ## can be any number
with(dat, cor(d, x)) ## can be any number
with(dat, cor(Z, x)) ## should be near 0
```

And we just saw that, in this design, and with these assumptions (including a SUTVA assumption) that $ITT = ((\bar{y}_1|C) - (\bar{y}_0|C))p_C = (CACE)p_C$, so we can define $CACE = ITT/p_C$. That is, we can learn about the effect of answering the door without worrying about the bias from x (or any set of x's).

VERY COOL You can learn about the causal effect of a non-random intervention (deciding to open the door) without "controlling for" $x_1, x_2, ...$ in this case.

How to calculate the ITT and CACE/LATE I

Some example data (where we know all potential outcomes):

	ID	u0	u	type	D_Z_0	D_Z_1	Y_D_0	Y_D_1	Y_D_0_Z_0	Y_D_1_Z_0	Y_D_0_Z_1	Y_D
1	084	-0.35785	0.00000	Always-Taker	1	1	0.00000	0.2972	0.00000	0.00000	0.1486	
2	088	-0.64348	0.00000	Never-Taker	0	0	0.00000	0.2972	0.00000	0.00000	0.1486	
3	058	0.55112	0.55112	Never-Taker	0	0	0.55112	0.8483	0.55112	0.55112	0.6997	
4	056	0.43293	0.43293	Never-Taker	0	0	0.43293	0.7301	0.43293	0.43293	0.5815	
5	079	-0.64878	0.00000	Never-Taker	0	0	0.00000	0.2972	0.00000	0.00000	0.1486	
6	037	0.26708	0.26708	Never-Taker	0	0	0.26708	0.5643	0.26708	0.26708	0.4157	
7	005	0.93277	0.93277	Complier	0	1	0.93277	1.2300	0.93277	0.93277	1.0814	
8	069	1.21382	1.21382	Always-Taker	1	1	1.21382	1.5110	1.21382	1.21382	1.3624	
9	015	0.08881	0.08881	Always-Taker	1	1	0.08881	0.3860	0.08881	0.08881	0.2374	
10	073	1.82298	1.82298	Complier	0	1	1.82298	2.1202	1.82298	1.82298	1.9716	
11	040	-0.79857	0.00000	Complier	0	1	0.00000	0.2972	0.00000	0.00000	0.1486	
12	081	0.97619	0.97619	Complier	0	1	0.97619	1.2734	0.97619	0.97619	1.1248	
13	042	-1.08643	0.00000	Never-Taker	0	0	0.00000	0.2972	0.00000	0.00000	0.1486	
14	098	0.53097	0.53097	Complier	0	1	0.53097	0.8282	0.53097	0.53097	0.6796	
15	052	0.62987	0.62987	Complier	0	1	0.62987	0.9271	0.62987	0.62987	0.7785	

How to calculate the ITT and CACE/LATE II

The ITT and CACE (the parts)

```
itt_y <- difference_in_means(Y ~ Z, data = dat0)
itt_y
```

```
Design:  Standard
  Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
Z   0.2024     0.1194   1.695  0.09319 -0.03452   0.4392 97.51
```

```
itt_d <- difference_in_means(D ~ Z, data = dat0)
itt_d
```

```
Design:  Standard
  Estimate Std. Error t value      Pr(>|t|) CI Lower CI Upper    DF
Z      0.5    0.07861    6.36 0.00000001764   0.3432   0.6568 70.32
```

How to calculate the ITT and CACE/LATE III

All together (the version dividing an unbiased estimator of ITT by an unbiased estimator of Proportion Compliers is often called Bloom's method from Bloom (1984)):[1]

```
cace_est <- iv_robust(Y ~ D | Z, data = dat0)
cace_est
```

```
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper DF
(Intercept)   0.3672     0.1019   3.604 0.0004944  0.16504   0.5695 98
D             0.4047     0.2423   1.670 0.0980802 -0.07616   0.8856 98
```

```
## Notice same as below:
coef(itt_y)[["Z"]] / coef(itt_d)[["Z"]]
```

```
[1] 0.4047
```

[1] works when $Z \to D$ is not weak see Imbens and Rosenbaum (2005) for a cautionary tale

Variance of IV estimator

- Recall that there exist analytic expressions for $\mathrm{Var}\left[\widehat{\mathrm{ITT}}_Y\right]$ and $\mathrm{Var}\left[\widehat{\mathrm{ITT}}_D\right]$
- We can conservatively estimate $\mathrm{Var}\left[\widehat{\mathrm{ITT}}_Y\right]$ and $\mathrm{Var}\left[\widehat{\mathrm{ITT}}_D\right]$ via $\widehat{\mathrm{Var}}\left[\widehat{\mathrm{ITT}}_Y\right]$ and $\widehat{\mathrm{Var}}\left[\widehat{\mathrm{ITT}}_D\right]$
- However, in general, there is no closed-form analytic expression for the variance of a random ratio
- We do not have an estimator for $\mathrm{Var}\left[\dfrac{\widehat{\mathrm{ITT}}_Y}{\widehat{\mathrm{ITT}}_D}\right]$ that is known to be unbiased, consistent or conservative
- Bloom (1984) proposed treating $\widehat{\mathrm{ITT}}_D$ as fixed
- Others use Delta method (Taylor series approximation), e.g., in `AER` or `estimatr` package in `R`

How do our estimators perform?

First, setup estimands and estimators:

```
estimands <- declare_inquiry(
  CACE = mean(Y_D_1[type == "Complier"] - Y_D_0[type == "Complier"]),
  ITT_y = mean(((Y_D_1_Z_1 + Y_D_0_Z_1) / 2) - ((Y_D_1_Z_0 + Y_D_0_Z_0) / 2)),
  ITT_d = mean(D_Z_1) - mean(D_Z_0)
)

estimator_cace <- declare_estimator(Y ~ D | Z, .method = iv_robust, inquiry = c("CACE"), label =
estimator_itt_y <- declare_estimator(Y ~ Z, inquiry = "ITT_y", .method = lm_robust, label = "diff
estimator_pp <- declare_estimator(Y ~ D, inquiry = "CACE", .method = lm_robust, label = "per-prot
estimator_itt_d <- declare_estimator(D ~ Z, inquiry = "ITT_d", .method = lm_robust, label = "diff



full_design <- base_design + estimands +
  estimator_cace + estimator_itt_y + estimator_itt_d + estimator_pp

draw_estimands(full_design)
```

```
  inquiry estimand
1    CACE   0.3462
2   ITT_y   0.1731
3   ITT_d   0.4300
```

```
draw_estimates(full_design)[, c("estimator", "term", "estimate", "std.error", "outcome", "inquiry
```

```
         estimator term estimate std.error outcome inquiry
1        iv_robust    D  0.41648   0.24444       Y    CACE
2   diff means ITT    Z  0.16659   0.09036       Y   ITT_y
3 diff means ITT_D    Z  0.40000   0.08122       D   ITT_d
4     per-protocol    D  0.04589   0.09725       Y    CACE
```

How do our estimators perform?

Then repeat the design many times:

```
full_designs_by_size <-
  redesign(full_design, N = c(50, 100, 200, 1000), prop_comply = c(.2, .5, .8))

dat_n20 <- draw_data(full_designs_by_size[["design_1"]])

my_diagnosands <-
  declare_diagnosands(
    mean_estimand = mean(estimand),
    mean_estimate = mean(estimate),
    bias = mean(estimate - estimand),
    rmse = sqrt(mean((estimate - estimand)^2)),
    ## power = mean(p.value <= alpha),
    coverage = mean(estimand <= conf.high & estimand >= conf.low),
    sd_estimate = sqrt(pop.var(estimate)),
    mean_se = mean(std.error)
  )

library(future)
library(future.apply)
plan(strategy = "multicore") ## won't work on Windows
which_to_sim <- rep(1, length = length(full_design))
names(which_to_sim) <- names(full_design)
which_to_sim["the_assign"] <- 1000
set.seed(12345)
results <- diagnose_design(full_designs_by_size,
  bootstrap_sims = 0,
  sims = 1000, # which_to_sim,
  diagnosands = my_diagnosands
```

Summary of Encouragement/Complier/Dose oriented designs:

- Analyze as you randomized: even when you don't control the dose you can learn something.
- The danger of per-protocol analysis: you give up the benefits of the research design (i.e. randomization)
- Variance calculations approximate (and can be untrustworthy in small samples, with weak instruments, and in other cases where we would worry about consistency (rare binary outcomes, very skewed outcomes, interdependence, ...)).

Summary of Encouragement/Complier/Dose oriented designs:

- Analyze as you randomized: even when you don't control the dose you can learn something.
- The danger of per-protocol analysis: you give up the benefits of the research design (i.e. randomization)
- Variance calculations approximate (and can be untrustworthy in small samples, with weak instruments, and in other cases where we would worry about consistency (rare binary outcomes, very skewed outcomes, interdependence, ...)).

Summary of Encouragement/Complier/Dose oriented designs:

- Analyze as you randomized: even when you don't control the dose you can learn something.
- The danger of per-protocol analysis: you give up the benefits of the research design (i.e. randomization)
- Variance calculations approximate (and can be untrustworthy in small samples, with weak instruments, and in other cases where we would worry about consistency (rare binary outcomes, very skewed outcomes, interdependence, ...)).

1. Causal effects when we do not control the dose
2. Hypothesis Tests about Complier causal effects
3. Learning about causal effects when data are missing

Hypothesis Tests about Complier causal effects

- We can test the sharp null hypothesis no effect among all units
- We know by random assignment that this test
  1. will have a type I error probability at least as small as α
  2. will have power greater than α for a class of alternative hypotheses
- Under what conditions /assumptions is a test of the sharp null of no effect among all units equivalent to a test of the sharp null of no effect among Compliers?
  1. Exclusion restriction
  2. No Defiers
  3. Non-zero proportion of Compliers
  4. Non-interference

Sharp null hypothesis testing example

The null hypothesis of no complier causal effect states that the individual causal effect of $\mathbf{Z}$ on $\mathbf{Y}$ is 0 among units who are Compliers.

Along with the exclusion restriction (i.e., that the individual causal effect is 0 for Always Takers and Never Takers) and the assumption of no Defiers, we can "fill in" missing potential outcomes according to the null hypothesis of no complier causal effect as follows:

$$Y_{c,0,i} = \begin{cases} Y_i - D_i\tau_i & \text{if } D_i = 1 \\ Y_i + (1 - D_i)\,\tau_i & \text{if } D_i = 0 \end{cases}$$

$$Y_{t,0,i} = \begin{cases} Y_i - D_i\tau_i & \text{if } D_i = 1 \\ Y_i + (1 - D_i)\,\tau_i & \text{if } D_i = 0, \end{cases}$$

where $\tau_i = 0$ for all i.

Sharp null hypothesis testing example

Imagine that our observed data is as follows:

$\mathbf{z}$	$\mathbf{y}$	$\mathbf{y_c}$	$\mathbf{y_t}$	$\mathbf{d}$	$\mathbf{d_c}$	$\mathbf{d_t}$
1	14	?	14	0	?	0
0	22	22	?	0	0	?
1	21	?	21	1	?	1
1	36	?	36	1	?	1
0	23	23	?	0	0	?
0	12	12	?	1	1	?
0	25	25	?	1	1	?
1	27	?	27	0	?	0

Observed experimental data

The observed Difference-in-Means test statistic, $\widehat{\bar{\tau}}\left(\mathbf{Z}, \mathbf{Y}\right)$, is 16.75. What is the distribution of that test statistic under the null hypothesis of no effects for any complier?

Sharp null hypothesis testing example

We can represent the sharp null hypothesis of no effect for all units without hypothesizing about non-random compliance (this is like the ITT_Y in that both can be assessed safely in a randomized experiment).

$\mathbf{z}$	$\mathbf{y}$	$\mathbf{y_c}$	$\mathbf{y_t}$	$\mathbf{d}$	$\mathbf{d_c}$	$\mathbf{d_t}$	Principal stratum
1	14	?	14	0	?	0	Never Taker or Defier
0	22	22	?	0	0	?	Complier or Never Taker
1	21	?	21	1	?	1	Complier or Always Taker
1	36	?	36	1	?	1	Complier or Always Taker
0	23	23	?	0	0	?	Complier or Never Taker
0	12	12	?	1	1	?	Always Taker or Defier
0	25	25	?	1	1	?	Always Taker or Defier
1	27	?	27	0	?	0	Never Taker or Defier

Sharp null of no effect for all units

Sharp null hypothesis testing example

We can represent the sharp null hypothesis of no effect for all units without hypothesizing about non-random compliance (this is like the ITT_Y in that both can be assessed safely in a randomized experiment).

$\mathbf{z}$	$\mathbf{y}$	$\mathbf{y_c}$	$\mathbf{y_t}$	$\mathbf{d}$	$\mathbf{d_c}$	$\mathbf{d_t}$	Principal stratum
1	14	14	14	0	?	0	Never Taker or Defier
0	22	22	22	0	0	?	Complier or Never Taker
1	21	21	21	1	?	1	Complier or Always Taker
1	36	36	36	1	?	1	Complier or Always Taker
0	23	23	23	0	0	?	Complier or Never Taker
0	12	12	12	1	1	?	Always Taker or Defier
0	25	25	25	1	1	?	Always Taker or Defier
1	27	27	27	0	?	0	Never Taker or Defier

Sharp null of no effect for all units

Sharp null hypothesis testing example

The null hypothesis of no effect among compliers under excludability (only a complier in the treatment group can have a causal effect), no defiers and nonzero proportion of compliers assumptions:

$\mathbf{z}$	$\mathbf{y}$	$\mathbf{y_c}$	$\mathbf{y_t}$	$\mathbf{d}$	$\mathbf{d_c}$	$\mathbf{d_t}$	Principal stratum
1	14	14	14	0	0	0	Never Taker ~~or Defier~~
0	22	22	22	0	0	?	Complier or Never Taker
1	21	21	21	1	?	1	Complier or Always Taker
1	36	36	36	1	?	1	Complier or Always Taker
0	23	23	23	0	0	?	Complier or Never Taker
0	12	12	12	1	1	1	Always Taker ~~or Defier~~
0	25	25	25	1	1	1	Always Taker ~~or Defier~~
1	27	27	27	0	0	0	Never Taker ~~or Defier~~

Sharp null of no effect among Compliers

We don't need to know which of units 2 – 5 are Compliers, only that at least one of these 4 units is a Complier.
Excludability means that the effect must be 0 for all units who are not compliers (i.e. implying the sharp null).

Sharp null hypothesis testing example

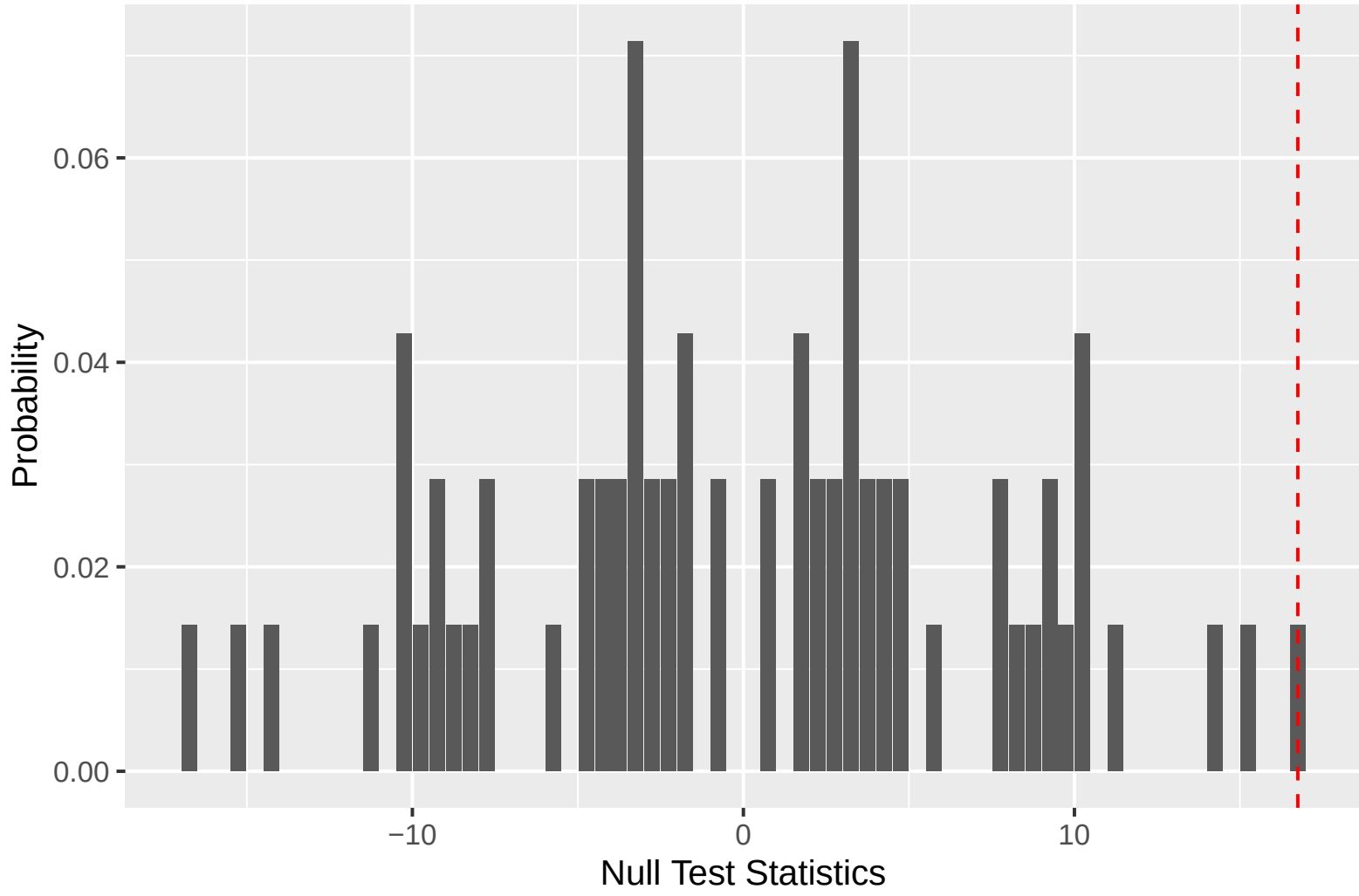
The null hypothesis of no effect among compliers under excludability (meaning that only a complier in the treatment group can have a causal effect), no defiers and nonzero proportion of Compliers assumptions:

$\mathbf{z}$	$\mathbf{y}$	$\mathbf{y_c}$	$\mathbf{y_t}$	$\mathbf{d}$	$\mathbf{d_c}$	$\mathbf{d_t}$	Principal stratum
1	14	14	14	0	0	0	Never Taker ~~or Defier~~
0	22	22	22	0	0	?	Complier or Never Taker
1	21	21	21	1	?	1	Complier or Always Taker
1	36	36	36	1	?	1	Complier or Always Taker
0	23	23	23	0	0	?	Complier or Never Taker
0	12	12	12	1	1	1	Always Taker ~~or Defier~~
0	25	25	25	1	1	1	Always Taker ~~or Defier~~
1	27	27	27	0	0	0	Never Taker ~~or Defier~~

Sharp null of no effect among Compliers

So: a regular test of the sharp null of no effects **is also a test of the sharp null of no effects among compliers** (under the assumptions of no defiers, non-zero compliers, exclusion, and no interference). The fact that $\tau_i = 0$ for Never Takers and Always Takers is by assumption, not a hypothesis.

Sharp null hypothesis testing example



Distribution of the Difference-in-Means test statistic under the sharp null of no effect: under the assumptions of excludability (no effects on Always Takers and Never Takers), no defiers, at least one complier, and SUTVA, this is a test of the hypothesis of no effects on compliers.

Summary

- The sharp null of no effects is meaningful and can be tested in a randomized experiment using assignment to treatment and ignoring compliance.
- The assumptions of excludability, no defiers, and at least one complier mean that we can interpret the test of the sharp null of no effects as a test of the sharp null of no effects on compliers: those assumptions require no effects among always-takers and never-takers, and there are no defiers in the data (again, all of this by assumption).
- Notice: no need for approximations; weak instruments do not threaten the validity of the statistical inferences.

Summary

- The sharp null of no effects is meaningful and can be tested in a randomized experiment using assignment to treatment and ignoring compliance.
- The assumptions of excludability, no defiers, and at least one complier mean that we can interpret the test of the sharp null of no effects as a test of the sharp null of no effects on compliers: those assumptions require no effects among always-takers and never-takers, and there are no defiers in the data (again, all of this by assumption).
- Notice: no need for approximations; weak instruments do not threaten the validity of the statistical inferences.

Summary

- The sharp null of no effects is meaningful and can be tested in a randomized experiment using assignment to treatment and ignoring compliance.
- The assumptions of excludability, no defiers, and at least one complier mean that we can interpret the test of the sharp null of no effects as a test of the sharp null of no effects on compliers: those assumptions require no effects among always-takers and never-takers, and there are no defiers in the data (again, all of this by assumption).
- Notice: no need for approximations; weak instruments do not threaten the validity of the statistical inferences.

1. Causal effects when we do not control the dose
2. Hypothesis Tests about Complier causal effects
3. Learning about causal effects when data are missing

Review of core assumptions from randomized experiments

1. Excludability: Potential outcomes depend only on assigned treatment (and not other factors)
2. Non-interference
3. Random assignment of treatment

Review of core assumptions from randomized experiments

1. Excludability: Potential outcomes depend only on assigned treatment (and not other factors)
2. Non-interference
3. Random assignment of treatment

Review of core assumptions from randomized experiments

1. Excludability: Potential outcomes depend only on assigned treatment (and not other factors)
2. Non-interference
3. Random assignment of treatment

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

Attrition (missing data on outcomes)

- Some units may have missing data on outcomes (= units attrit) when:
  - some respondents can't be found or refuse to participate in endline data collection.
  - some records are lost.
- This is a problem when treatment affects missingness.
  - For example, units in control may be less willing to answer survey questions.
  - For example, treatment may have caused units to migrate and cannot be reached
- If we analyze the data by dropping units with missing outcomes, then we are no longer comparing similar treatment and control groups. (We have trouble analyzing as we randomized!)
- Dropping the missing observations brings us closer to per-protocol analysis and confounding.

What can we do?

- Check whether attrition rates are similar in treatment and control groups.
- Check whether treatment and control groups have similar covariate profiles.
- Do not drop observations that are missing outcome data from your analysis.
- Analyze missingness on outcome as another outcome: could treatment have caused missing outcomes?
- When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

- Check whether attrition rates are similar in treatment and control groups.
- Check whether treatment and control groups have similar covariate profiles.
- Do not drop observations that are missing outcome data from your analysis.
- Analyze missingness on outcome as another outcome: could treatment have caused missing outcomes?
- When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

- Check whether attrition rates are similar in treatment and control groups.
- Check whether treatment and control groups have similar covariate profiles.
- Do not drop observations that are missing outcome data from your analysis.
- Analyze missingness on outcome as another outcome: could treatment have caused missing outcomes?
- When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

- Check whether attrition rates are similar in treatment and control groups.
- Check whether treatment and control groups have similar covariate profiles.
- Do not drop observations that are missing outcome data from your analysis.
- Analyze missingness on outcome as another outcome: could treatment have caused missing outcomes?
- When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

- Check whether attrition rates are similar in treatment and control groups.
- Check whether treatment and control groups have similar covariate profiles.
- Do not drop observations that are missing outcome data from your analysis.
- Analyze missingness on outcome as another outcome: could treatment have caused missing outcomes?
- When outcome data are missing we can sometimes **bound** our estimates of treatment effects.

What can we do?

- But the best approach is to try to anticipate and prevent attrition.
  - Blind people to their treatment status.
  - Promise to deliver the treatment to the control group after the research is completed.
  - Plan ex ante to reach all subjects at endline.
  - Budget for intensive follow-up with a random sample of attriters (Gerber and Green, 2012) Chapter 7.

See also Chapter on Threats to Internal Validity in Experiments

What can we do?

- But the best approach is to try to anticipate and prevent attrition.
  - Blind people to their treatment status.
  - Promise to deliver the treatment to the control group after the research is completed.
  - Plan ex ante to reach all subjects at endline.
  - Budget for intensive follow-up with a random sample of attriters (Gerber and Green, 2012) Chapter 7.

See also Chapter on Threats to Internal Validity in Experiments

What can we do?

- But the best approach is to try to anticipate and prevent attrition.
  - Blind people to their treatment status.
  - Promise to deliver the treatment to the control group after the research is completed.
  - Plan ex ante to reach all subjects at endline.
  - Budget for intensive follow-up with a random sample of attriters (Gerber and Green, 2012) Chapter 7.

See also Chapter on Threats to Internal Validity in Experiments

What can we do?

- But the best approach is to try to anticipate and prevent attrition.
  - Blind people to their treatment status.
  - Promise to deliver the treatment to the control group after the research is completed.
  - Plan ex ante to reach all subjects at endline.
  - Budget for intensive follow-up with a random sample of attriters (Gerber and Green, 2012) Chapter 7.

See also Chapter on Threats to Internal Validity in Experiments

What can we do?

- But the best approach is to try to anticipate and prevent attrition.
  - Blind people to their treatment status.
  - Promise to deliver the treatment to the control group after the research is completed.
  - Plan ex ante to reach all subjects at endline.
  - Budget for intensive follow-up with a random sample of attriters (Gerber and Green, 2012) Chapter 7.

See also Chapter on Threats to Internal Validity in Experiments

Missing data on covariates is not as problematic

- Missing **background covariates** (i.e.,variables for which values do not change as a result of treatment) for some observations is less problematic.
  - We can still learn about the causal effect of an experiment without those covariates.
  - We can also use the background covariate as planned by imputing for the missing values.
  - We can also condition on that missingness directly: we could assess causal effects for the subgroup of those missing on income and compare it to the subgroup of those not missing on income.

Missing data on covariates is not as problematic

- Missing **background covariates** (i.e.,variables for which values do not change as a result of treatment) for some observations is less problematic.
  - We can still learn about the causal effect of an experiment without those covariates.
  - We can also use the background covariate as planned by imputing for the missing values.
  - We can also condition on that missingness directly: we could assess causal effects for the subgroup of those missing on income and compare it to the subgroup of those not missing on income.

Missing data on covariates is not as problematic

- Missing **background covariates** (i.e.,variables for which values do not change as a result of treatment) for some observations is less problematic.
  - We can still learn about the causal effect of an experiment without those covariates.
  - We can also use the background covariate as planned by imputing for the missing values.
  - We can also condition on that missingness directly: we could assess causal effects for the subgroup of those missing on income and compare it to the subgroup of those not missing on income.

Missing data on covariates is not as problematic

- Missing **background covariates** (i.e.,variables for which values do not change as a result of treatment) for some observations is less problematic.
  - We can still learn about the causal effect of an experiment without those covariates.
  - We can also use the background covariate as planned by imputing for the missing values.
  - We can also condition on that missingness directly: we could assess causal effects for the subgroup of those missing on income and compare it to the subgroup of those not missing on income.

Summary about Missing data and Experiments.

- Missing outcomes or missing treatment assignment (or missing blocking information) are all big problems. How might those with missing outcomes have behaved in treatment versus control? We don't know.
- Missing covariate information is not a problem: it is fixed, same proportion should be missing covariate information in both treatment and control conditions

Summary about Missing data and Experiments.

- Missing outcomes or missing treatment assignment (or missing blocking information) are all big problems. How might those with missing outcomes have behaved in treatment versus control? We don't know.
- Missing covariate information is not a problem: it is fixed, same proportion should be missing covariate information in both treatment and control conditions

References


Gerber, Alan S and Donald P Green (2012). Field Experiments: Design, Analysis, and Interpretation. New York, NY: W.W. Norton.


Imbens, Guido W and Paul R Rosenbaum (2005). "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education". In: Journal of the Royal Statistical Society: Series A (Statistics in Society) 168.1, pp. 109–126.