Natural and Observational Studies (NOS), Attrition (A), and Treatment Effects (TEs)

Ethan Moody

December 2023

Vietnam Draft Lottery

Observational esteimate

Suppose that you had not run an experiment. Estimate the "effect" of each year of education on income as an observational researcher might, by just running a regression of years of education on income (in R-ish, `income ~ years_education`). What does this naive regression suggest?

```r
# Load additional necessary packages
library(stargazer)

# Create a regression model to estimate the effect of education on income
model_observational      <- d[, lm(income ~ years_education)]

# Calculate robust standard errors (RSEs)
model_observational_rses <- sqrt(diag(vcovHC(model_observational)))

# Display stargazer table with model results and RSEs
stargazer(
  model_observational,
  se = list(model_observational_rses),
  add.lines = list(
    c("Using Robust Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Answer: This naive regression suggests that `years_education` has a statistically significant *positive* "effect" on `income` (p $\approx$ 0). More specifically, the model suggests that for each year of education attained, income increases by approximately \$5750.48. So, the more educated one is, the more income they tend to earn. It's important to note that, because we haven't run an experiment, this relationship isn't necessarily anything more than *associative* in nature (i.e., it's not necessarily causal).

Evaluating observational estimate

Continue to suppose that we did not run the experiment, but that we saw the result that you noted in part 1. Tell a concrete story about why you don't believe that observational result tells you anything causal.

Answer: Assuming that we hadn't run an experiment but observed the result above, we could only conclude that there's a *positive association* between `years_education` and `income` (i.e., as education increases, so does income — generally speaking). There are several reasons why this finding doesn't have to tell us anything

Table 1:

	Dependent variable:
	income
years_education	5,750.480***
	(84.411)
Constant	−23,354.640***
	(1,197.226)
Using Robust Standard Errors	Yes
Observations	19,567
R^2	0.196
Adjusted R^2	0.196
Residual Std. Error	26,592.180 (df = 19565)
F Statistic	4,761.015*** (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

causal about the relationship between education and income. One possible explanation for this finding is that those who make a choice to pursue higher education might also have skills, personality characteristics, and/or family or community environments (e.g., higher average intelligence, more ambition, a stronger work ethic, a history of family members pursuing advanced degrees, etc.) that are associated with higher income potential. Another possible explanation is related to geographic differences, where individuals who live in particularly populated areas of the country might have more opportunity to pursue both higher levels of education (because of the proximity of schools/universities) and jobs with higher average pay than those who live in less-populated areas. In both cases, years of education does not necessarily *cause* higher income, though a positive relationship between the two variables still exists.

Natural experiment effect on education

Now, let's get to using the natural experiment. Define "having a high-ranked draft number" as having a draft number between 1-80. For the remaining 285 days of the year, consider them having a "low-ranked" draft number). Create a variable in your dataset called `high_draft` that indicates whether each person has a high-ranked draft number or not. Using a regression, estimate the effect of having a high-ranked draft number on years of education obtained. Report the estimate and a correctly computed standard error. (*Hint: How is the assignment to having a draft number conducted? Does random assignment happen at the individual level? Or, at some higher level?)

```r
# Add `high_draft` variable to original data.table
d                         <- d[, high_draft := ifelse(draft_number <= 80, 1, 0)]

# Create two overall regression models to estimate the effect of draft number on education
# Note: first model uses `high_draft` variable, second model uses `draft_number` variable
model_education_hd        <- d[, lm(years_education ~ high_draft)]
model_education_dn        <- d[, lm(years_education ~ draft_number)]

# Create two more models to estimate the effect of draft number on education for high/low draft groups
model_education_hdn       <- d[high_draft == 1, lm(years_education ~ draft_number)]
model_education_ldn       <- d[high_draft == 0, lm(years_education ~ draft_number)]

# Calculate standard errors (clustered = CSEs, robust = RSEs) for all three models
model_education_hd_rses   <- sqrt(diag(vcovHC(model_education_hd)))
```

```r
model_education_dn_cses  <- sqrt(diag(vcovCL(model_education_dn, cluster = ~draft_number)))
model_education_hdn_rses <- sqrt(diag(vcovHC(model_education_hdn)))
model_education_ldn_rses <- sqrt(diag(vcovHC(model_education_ldn)))

# Display stargazer table with first model results
stargazer(
  model_education_hd,
  se = list(model_education_hd_rses),
  add.lines = list(
    c("Using Robust Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Table 2:

	Dependent variable:
	years_education
high_draft	2.126^{***}
	(0.038)
Constant	14.434^{***}
	(0.017)
Using Robust Standard Errors	Yes
Observations	19,567
R^2	0.138
Adjusted R^2	0.138
Residual Std. Error	2.117 (df = 19565)
F Statistic	$3{,}145.132^{***}$ (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
# Display stargazer table with second model results
stargazer(
  model_education_dn,
  se = list(model_education_dn_cses),
  add.lines = list(
    c("Using Clustered Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

```r
# Display stargazer table with cluster-specific third and fourth model results
# Note: tables show results for `high_draft` == 1 and `high_draft` == 0
stargazer(
  model_education_hdn, model_education_ldn,
  se = list(model_education_hdn_rses, model_education_ldn_rses),
  add.lines = list(
    c("High-Ranked Draft Group", "Yes", "No"),
    c("Using Robust Standard Errors", "Yes", "Yes")),
  type = "latex",
  #type = "text")
```

Table 3:

	Dependent variable:
	years_education
draft_number	-0.006^{***}
	(0.0004)
Constant	15.925^{***}
	(0.089)
Using Clustered Standard Errors	Yes
Observations	19,567
R^2	0.069
Adjusted R^2	0.068
Residual Std. Error	2.202 (df = 19565)
F Statistic	$1{,}439.698^{***}$ (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
  header = FALSE)
```

Table 4:

	Dependent variable:	
	years_education	
	(1)	(2)
draft_number	0.00001	-0.0001
	(0.001)	(0.0002)
Constant	16.559^{***}	14.448^{***}
	(0.067)	(0.049)
High-Ranked Draft Group	Yes	No
Using Robust Standard Errors	Yes	Yes
Observations	3,896	15,671
R^2	0.00000	0.00001
Adjusted R^2	-0.0003	-0.0001
Residual Std. Error	2.092 (df = 3894)	2.124 (df = 15669)
F Statistic	0.0001 (df = 1; 3894)	0.087 (df = 1; 15669)
Note:		*p<0.1; **p<0.05; ***p<0.01

Answer: A few regression models are used here to explore the results. Our first regression model suggests (via our new `high_draft` variable) that having a high-ranked draft number has a statistically significant *positive* "effect" on `years_education` ($p \approx 0$). More specifically, it shows that those with a high-ranked draft number have approximately 2.126 more years of education than those with a low-ranked draft number. Our second, third, and fourth regression models look at this "effect" from a slightly different angle (via the `draft_number` variable). Our second model suggests that `draft_number` *overall* has a statistically significant *negative* "effect" on `years_education` ($p \approx 4.775451 \times 10^{-304}$). In other words, this model indicates that as draft number increases, years of education decrease — a finding that's consistent with our first model, too. When we further investigate these results by splitting the data into two subsets — the `high_draft` group/cluster and the non-`high_draft` group/cluster — and then run a regression for each of these subsets (represented by our third and fourth models), we no longer see a statistically significant "effect" of `draft_number` on `years_education` for both groups (since they're now cluster-specific and the coefficient for `draft_number` is effectively equal to 0); however, we *do* see roughly the same magnitude of difference between the "constant" terms as we observed between the draft groups in our first model. The `high_draft` group has a "constant" value of 16.559 while the non-`high_draft` group has a "constant" value of 14.448 — approximately 2.112 years lower. These results suggest that having a high-ranked draft number has some "effect" on education: those with the higher-ranked draft numbers tend to have more years of education than those with the lower-ranked draft numbers. Although the draft numbers are supposedly assigned randomly, I actually have a hard time feeling *fully convinced* that a truly causal relationship underlies the association we see here. The draft numbers are determined by a process in which birthdays are drawn out of a hat (which has a seemingly random element to it) — but who's to say that all birthdays have an equal chance of being selected for any given draw? It may be plausible to claim that random assignment occurs at the birthday-/birth-date-to-draft-number level — i.e., hypothetically, any birthday/birth-date *could* be assigned to any draft number, pre-draw; however, once the birthdays are placed in a hat and the drawing process commences, claiming *true randomness* may no longer be fully justified. Perhaps some birthdays/birth-dates rise to the top of the stack in the hat, making them easier to draw sooner than others. Perhaps not every piece of paper on which a birthday/birth-date is written is the same size or shape, making some faster to grasp and select than others. In fact, there could be a myriad of reasons why some dates are more prone to be selected before others, or why each date might have a different probability of selection at any point in the process. Based on the limited information provided about the draft assignment process, we just don't know. And if there's anything other than *true random assignment* happening within this procedure, a critical assumption for drawing causal conclusions from this study breaks down, leaving us with (at best) strong indicators of interesting correlations/associations between variables. That's what I think is happening here with `draft_number`/`high_draft` and `years_education`, even if one can furnish a compelling explanation for the relationship we see between these variables that sounds causal (e.g., the high-ranked draft numbers actually *induce/cause* people to go pursue higher levels of education). While that explanation may be true in part, I have some reservations about drawing purely causal connections from this study based on what *could be* an assignment procedure that's not truly random.

Natural experiment effect on income

Using linear regression, estimate the effect of having a high-ranked draft number on income. Report the estimate and the correct standard error.

```r
# Create two overall regression models to estimate the effect of draft number on income
# Note: first model uses `high_draft` variable, second model uses `draft_number` variable
model_income_hd       <- d[, lm(income ~ high_draft)]
model_income_dn       <- d[, lm(income ~ draft_number)]

# Create two more models to estimate the effect of draft number on income for high/low draft groups
model_income_hdn      <- d[high_draft == 1, lm(income ~ draft_number)]
model_income_ldn      <- d[high_draft == 0, lm(income ~ draft_number)]

# Calculate standard errors (clustered = CSEs, robust = RSEs) for all three models
model_income_hd_rses  <- sqrt(diag(vcovHC(model_income_hd)))
model_income_dn_cses  <- sqrt(diag(vcovCL(model_income_dn, cluster = ~draft_number)))
model_income_hdn_rses <- sqrt(diag(vcovHC(model_income_hdn)))
model_income_ldn_rses <- sqrt(diag(vcovHC(model_income_ldn)))

# Display stargazer table with first model results
stargazer(
  model_income_hd,
  se = list(model_income_hd_rses),
  add.lines = list(
```

```r
    c("Using Robust Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Table 5:

	Dependent variable:
	income
high_draft	6,637.554***
	(545.590)
Constant	60,761.890***
	(233.387)
Using Robust Standard Errors	Yes
Observations	19,567
R^2	0.008
Adjusted R^2	0.008
Residual Std. Error	29,532.970 (df = 19565)
F Statistic	157.613*** (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
# Display stargazer table with second model results
stargazer(
  model_income_dn,
  se = list(model_income_dn_cses),
  add.lines = list(
    c("Using Clustered Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

```r
# Display stargazer table with cluster-specific third and fourth model results
# Note: tables show results for `high_draft` == 1 and `high_draft` == 0
stargazer(
  model_income_hdn, model_income_ldn,
  se = list(model_income_hdn_rses, model_income_ldn_rses),
  add.lines = list(
    c("High-Ranked Draft Group", "Yes", "No"),
    c("Using Robust Standard Errors", "Yes", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Answer: Similarly, a few regression models are used here to explore the results. Our first regression model suggests (via our new `high_draft` variable) that having a high-ranked draft number has a statistically significant *positive* “effect” on `income` ($p \approx 5.1755739 \times 10^{-36}$). More specifically, it shows that those with a high-ranked draft number have approximately $6637.554 more income than those with a low-ranked draft number. Once again, our second, third, and fourth regression models can help us look at this “effect” from a slightly different angle (via the `draft_number` variable). Our second model suggests that `draft_number` *overall* has a statistically significant *negative* “effect” on `income` ($p \approx 3.4824674 \times 10^{-16}$). In other words,

Table 6:

	Dependent variable:
	income
draft_number	-16.500^{***}
	(2.299)
Constant	$65{,}171.380^{***}$
	(527.314)
Using Clustered Standard Errors	Yes
Observations	19,567
R^2	0.003
Adjusted R^2	0.003
Residual Std. Error	29,601.330 (df = 19565)
F Statistic	66.624^{***} (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 7:

	Dependent variable:	
	income	
	(1)	(2)
draft_number	7.665	2.286
	(21.238)	(2.847)
Constant	$67{,}091.170^{***}$	$60{,}250.670^{***}$
	(983.664)	(680.705)
High-Ranked Draft Group	Yes	No
Using Robust Standard Errors	Yes	Yes
Observations	3,896	15,671
R^2	0.00003	0.00004
Adjusted R^2	-0.0002	-0.00002
Residual Std. Error	30,781.040 (df = 3894)	29,215.690 (df = 15669)
F Statistic	0.129 (df = 1; 3894)	0.657 (df = 1; 15669)
Note:		*p<0.1; **p<0.05; ***p<0.01

this model indicates that as draft number increases, income decreases — a finding that's consistent with our first model, too. Like in the previous problem, when we further investigate these results by splitting the data into two subsets — the `high_draft` group/cluster and the non-`high_draft` group/cluster — and then run a regression for each of these subsets (represented by our third and fourth models), we no longer see a statistically significant "effect" of `draft_number` on `income` for both groups (since they're now cluster-specific and the coefficient for `draft_number` is virtually 0 when accounting for the size of the SEs); however, we *do* see roughly the same magnitude of difference between the "constant" terms as we observed between the draft groups in our first model. The `high_draft` group has a "constant" value of $\$6.7091171 \times 10^4$ while the non-`high_draft` group has a "constant" value of $\$6.0250665 \times 10^4$ — approximately \$6840.506 lower. These results suggest that, like with education, having a high-ranked draft number has some "effect" on income: those with the higher-ranked draft numbers tend to have higher incomes than those with the lower-ranked draft numbers.

Instrumental variables estimate of education on income

Now, estimate the Instrumental Variables regression to estimate the effect of education on income. To do so, use `AER::ivreg`. After you evaluate your code, write a narrative description about what you learn.

```r
# Create a 2SLS regression model to estimate the effect of education and draft number on income
model_iv      <- d[, ivreg(income ~ years_education | draft_number)]

# Calculate clustered standard errors (CSEs)
model_iv_cses <- sqrt(diag(vcovCL(model_iv, cluster = ~draft_number)))

# Display stargazer table with model results and RSEs
stargazer(
  model_iv,
  se = list(model_iv_cses),
  add.lines = list(
    c("Using Clustered Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Table 8:

	Dependent variable:
	income
years_education	2,892.345***
	(336.517)
Constant	19,110.280***
	(4,993.838)
Using Clustered Standard Errors	Yes
Observations	19,567
R^2	0.147
Adjusted R^2	0.147
Residual Std. Error	27,379.790 (df = 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: Our IV model reinforces the statistically significant *positive* relationship between `years_education` and `income` with `draft_number` included as an instrumental variable ($p \approx 1.1887073 \times 10^{-18}$). More specifically, this outcome indicates that for each additional year of education attained, income increases by $2892.345. By performing IV/2SLS regression, we've addressed endogeneity within our previous models and can now more credibly make a causal claim about this relationship — that is, we can say that income differences are *attributable to* (produced by) differences in years of education. In other words, what this model allows us to do is to siphon off any effect on `income` that `draft_number` might have (working through a variable like `years_education`) and instead focus on just the effect on `income` from `years_education`. It appears — as one might suspect — that there really is a basis for claiming some sort of causal relationship between one's income and the level of education they pursue: as they pursue more education, their income also generally increases *as a result.*

Evaluating the exclusion restriction

Give one reason this requirement might not be satisfied in this context. In what ways might having a high draft rank affect individuals' income **other** than nudging them to attend more school?

Answer: It's possible that some individuals with a high draft rank were, in fact, drafted into the military and then either (A) received income *for the first time* through their service in a military role, (B) died as a result of military combat, or (C) moved out of country post-military service. Since serving in the military is a form of employment, all who serve in this capacity receive compensation from the government. For some drafted individuals (especially younger individuals), this compensation may have been their first ever income. This would have obviously affected their income in the data without any relation to education/schooling. Alternatively, some individuals who served in the military may have been killed in combat or may have moved out of country, which would have effectively reduced their IRS-reported incomes to zero. This, too, would be an income-affecting outcome influenced by a high draft rank, but not related to education/schooling.

Differential attrition

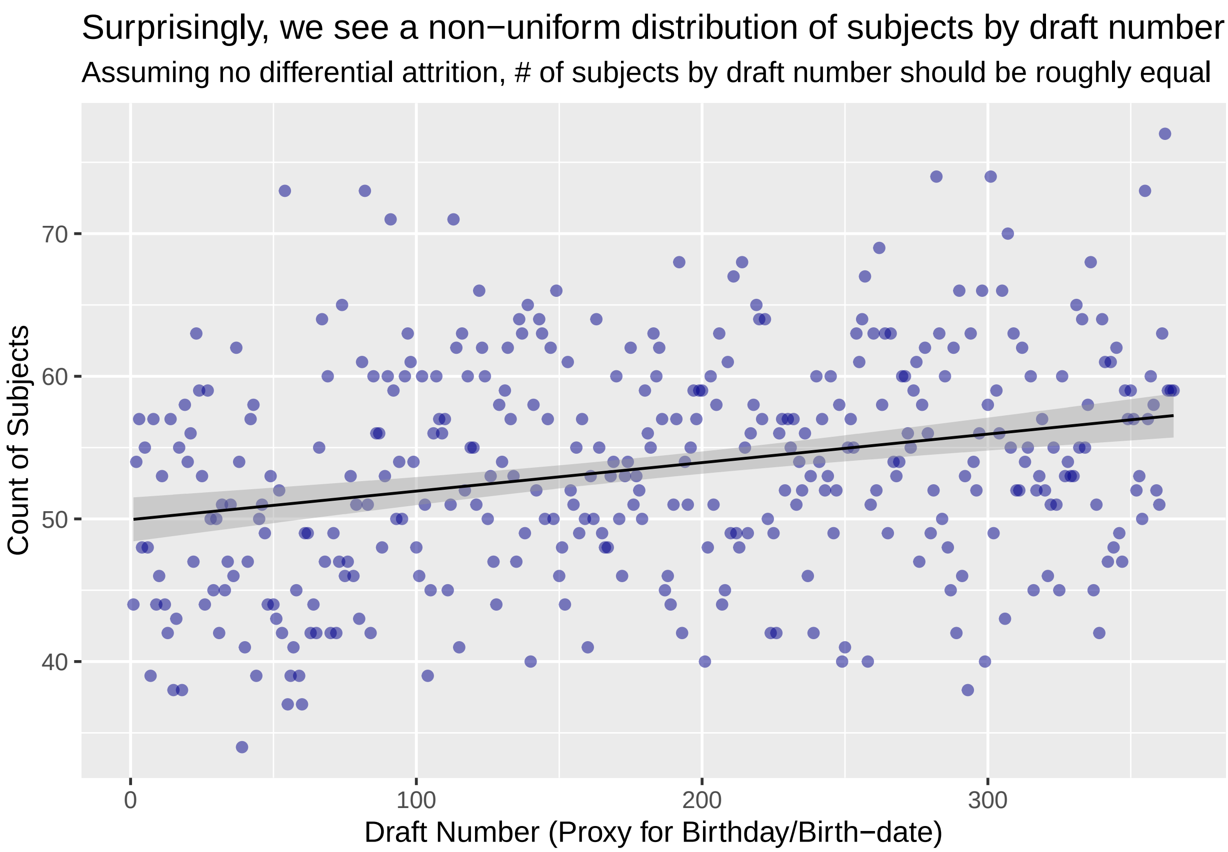
Conduct a test for the presence of differential attrition by treatment condition. That is, conduct a formal test of the hypothesis that the "high-ranked draft number" treatment has no effect on whether we observe a person's income. **(Note, that an earning of $0 *actually* means they didn't earn any money – i.e. earning $0 does not mean that their data wasn't measured. Let's be really, really specific: If you write a model that looks anything like, `lm(income == 0 ~ .)` you've gone the wrong direction.)**

```
# Add `draft_number_count` variable to original data.table
d                                 <- d[, draft_number_count := .N, by = draft_number]

# Create modified version of data.table, subsetting to the columns necessary for testing
selected_columns                  <- c("draft_number", "high_draft", "draft_number_count")
d_sub                             <- unique(d[, ..selected_columns])
d_sub_ordered                     <- d_sub[order(draft_number)]

# Create a scatterplot and first-pass trendline showing subject counts by `draft_number`
# Note 1: expecting a uniform distribution with flat trendline, unless there *is* differential attritio
# Note 2: trendline should "preview" formal regression test results -- this is just a helpful visual
ggplot(
  data = d_sub_ordered,
  aes(
    x = draft_number,
    y = draft_number_count)) +
  geom_point(
    color = "darkblue",
    alpha = 0.5) +
  geom_smooth(
    method = "lm",
```

```
    se = TRUE,
    color = "black",
    size = 0.5) +
  labs(
    title = "Surprisingly, we see a non-uniform distribution of subjects by draft number",
    subtitle = "Assuming no differential attrition, # of subjects by draft number should be roughly equa
    x = "Draft Number (Proxy for Birthday/Birth-date)",
    y = "Count of Subjects")
```




```
# Create a regression model to test whether differential attrition is at-play
model_differential_attrition      <- d_sub_ordered[, lm(draft_number_count ~ high_draft)]

# Calculate robust standard errors (RSEs)
model_differential_attrition_rses <- sqrt(diag(vcovHC(model_differential_attrition)))

# Display stargazer table with model results
stargazer(
  model_differential_attrition,
  se = list(model_differential_attrition_rses),
  add.lines = list(
    c("Using Robust Standard Errors", "Yes")),
  type = "latex",
  #type = "text")
  header = FALSE)
```

Answer: The results of our regression test indicate that having a high-ranked draft number *does* (in fact) have an effect on whether we observe a person's income. This finding suggests the presence of differential

Table 9:

	Dependent variable:
	draft_number_count
high_draft	-6.286^{***}
	(0.952)
Constant	54.986^{***}
	(0.432)
Using Robust Standard Errors	Yes
Observations	365
R^2	0.112
Adjusted R^2	0.110
Residual Std. Error	7.336 (df = 363)
F Statistic	45.861^{***} (df = 1; 363)
Note:	*p<0.1; **p<0.05; ***p<0.01

attrition by treatment condition (high-/low-ranked draft group). From our regression model, we see a statistically significant negative coefficient for those in the `high_draft` group ($p \approx 5.1277633 \times 10^{-11}$), which tells us that *a smaller proportion of subjects with high-ranked draft numbers appear in our data than would be expected if no differential attrition occurred.* In other words, in the absence of differential attrition — and under the assumption that a count of individuals by all possible birthdays/birth-dates within a year should *generally* follow a uniform distribution — we'd expect to see roughly the same count of subjects across birthdays/birth-dates in the `high_draft` group as in the non-`high_draft` group. Given that we don't, we can conclude that something is impacting the completeness of data for the `high_draft` group.

Evaluate differential attrition

Tell a concrete story about what could be leading to the result in part 7. How might this differential attrition create bias in the estimates of a causal effect?

Answer: The result we observe in part 7 could be related to the fact that *some of those with high-ranked draft numbers were actually drafted into the military, and then died during their military service.* (This is one possible explanation.) From the information given in the prompts/setup for this assignment, we know that all incomes in our dataset are reported/measured without error. We also know from the provided SQL query logic that the income data is being taken from a 1980 IRS dataset. This would logically exclude anyone who was deceased within — or prior to — the 1980 tax year, and a portion of that group could be comprised of *drafted fallen military service members.* As a result, we could expect to see a larger proportion of high-ranked draft individuals "dropped" from the data generated by the query than low-ranked draft individuals, yielding a situation where differential attrition is at-play. In this scenario, differential attrition can create bias in the estimates of a causal effect by reducing the similarity and comparability of the high-ranked and low-ranked draft groups. It's possible that the attrition across the high-ranked draft group wasn't *random*, and that those who ultimately attrited (died) may have had specific characteristics associated with higher/lower incomes, making this group *not* a truly "representative subset" of the broader study sample. To the extent this attrited subgroup didn't actually resemble all other records in the data, any observed causal effect drawn from the study might be under- or overestimated (biased) relative to what is *actually* the true effect.

Think about Treatment Effects

Throughout this course we have focused on the average treatment effect. *Why* we are concerned about the average treatment effect. What is the relationship between an ATE, and some individuals' potential outcomes? Make the strongest case you can for why this is a *good* measure.

Answer: The average treatment effect (ATE) is a fundamental concept in field experimentation. In "potential outcomes" notation, the ATE measures the arithmetic difference between (1) the expected value (average) of experimental subjects' potential outcomes to treatment and (2) the expected value (average) of experimental subjects' potential outcomes to control (i.e., $ATE = E[Y_i(1)] - E[Y_i(0)]$, or $ATE = \frac{1}{N}\sum_{i=1}^{N} \tau_i$, where τ_i represents the treatment effect $Y_i(1) - Y_i(0)$ of the ith subject). Ideally, to measure treatment effects, we'd have access to both (1) and (2) for every experimental subject. However, in the real world, this isn't possible; we can only observe *either* (1) *or* (2) for each subject based on whether they're exposed to treatment or to control. Fortunately, to overcome this obstacle, we can leverage a procedure called random assignment to randomly assign all experimental subjects to either the treatment or control condition. If this assignment procedure is truly random, it ensures that the make-up of both groups of subjects is identical across all their underlying characteristics (eliminating all unobserved heterogeneity), and it allows us to attribute any change in outcomes for the treatment group relative to the control group to the treatment itself. This provides a justification for deeming these groups to be equivalent prior to the administration of treatment, and for comparing *averages* across the groups to determine the effect of a treatment. Because of their pre-treatment equivalence, we can say that if treatment were *not* administered to those in the treatment group, the average value for the outcome variable across the treatment group should be statistically identical to the average value for the outcome variable across the control group. Thus, we can also say that any difference observed between these group-average outcome values *post-treatment* represents the ATE.

It's important to note that any individual subject's potential outcome to treatment or to control might differ from the average outcome across their membership group (treatment/control). In fact, the "average" prefix in "average treatment effect" (ATE) recognizes the presence of variability in the underlying individual subject-level outcomes. However, using the ATE as a summary measure of an experiment's treatment effect has several key advantages over merely invoking the distribution of individual subject-level treatment effects. First, the ATE represents a single number that is simple to communicate and easily understandable. It's far less cumbersome to cite a single number when reporting the outcome of an experiment than a whole range (distribution) of numbers, which improves the communicability of experimental results. The ATE helps anchor both experimenters and their target audience on a particular value that describes the overall effect of a treatment, rather than focus their attention on a wide range of possible effects across all unique experimental subjects (e.g., the minimum effect, the maximum effect, the effect at the 25th percentile, the effect at the 75th percentile, etc.). Second, the ATE is something that can be easily tested for reproducibility. One of the hallmarks of any good experiment — in terms of its design and measurement — is its ability to be reproduced or corroborated by a follow-up experiment. If the ATE were tossed aside in favor of the distribution of individual subject-level outcomes, would any experiment be considered truly reproducible unless the follow-up study had the exact same number of participants or very closely matched the distribution of effects from the original study? Perhaps that question verges on hyperbole, but more fairly, comparing ATEs across experiments allows for a more rapid/efficient and reliable test of reproducibility than comparing individualized outcomes, and reproducible experiments — as well as reproducibility tests — advance the cause of scientific progress. Third, the ATE is a generalized measure of a given treatment's effect, which helps set expectations for researchers, subjects, decision-makers, and the general population around the treatment's impact. While the ATE may not describe the unique effect that each individual subject experiences, it seeks to offer a fair estimate for the effect of the treatment across this diverse group of subjects so that the treatment's impact is understood *in general*. Business professionals might use this info to design a new product or enhancement; social scientists or research practitioners might use this info to build on prior research; politicians or economists might use this info to implement new policy; and members of the general public might use this info to make a decision about whether to take or not take an action, a medical supplement, etc. Fourth, the ATE is helpful for prioritizing or stack-ranking different treatments and making decisions around which to pursue. A researcher might test many similar treatments (e.g., a drug containing 1000mg of a substance vs. 100mg, or a message with many typos vs. none, or a visual stimulus that lasts 20 seconds vs. 5 seconds, etc.) for a particular use-case and need to rank them on the basis of affordability, availability, feasibility, public safety, etc. Having an ATE for each treatment dosage-level or scenario would allow them to quickly analyze these considerations against their associated experimental outcomes/impacts, draw comparisons, and make informed decisions. For these reasons, although the ATE isn't a perfect representation of every individualized outcome within an experiment, it provides a *good* measure of the effect of any given treatment — one that's simple to communicate and understand, reproducible, generalizable, comparable, and practical.