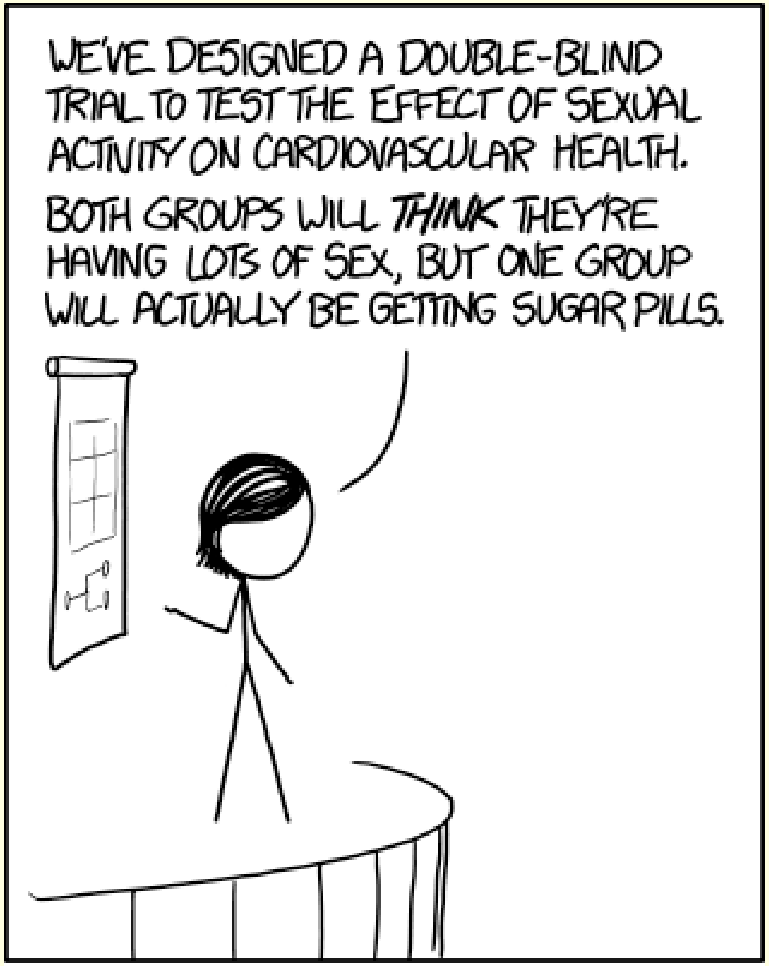
Experiments: a Primer on Design and Practical Tips



THE LIMITATIONS OF BLIND TRIALS

An abbreviated list of examples

- Program evaluation: SAT prep classes, weight loss programs, fundraising, diversity training, deliberative polls, virginity pledging, advertising campaigns, wilderness programs, mental exercise for elderly
- Public policy evaluation: speed traps, vouchers, alternative sentencing, job training, health insurance subsidies, tax compliance, public housing
- Behavioral Research: persuasion, mobilization, education, income, interpersonal influence, conscientious health behaviors
- Research on Institutions: rules for authorizing decisions, rules of succession, manner in which an organization is founded, monitoring performance

Consider the alternatives to experiments

- Survey analysis: sampling advantages, but problem of measurement and unobserved heterogeneity
- Process tracing: useful for developing hypotheses but often subject to uncertain inference due to limited sample size and unobserved heterogeneity
- Econometric analysis: cross-sections, time-series, and panels are all subject to uncertainty about measurement and unobserved heterogeneity

Experiments as a Solution to the Problem of Unobserved Confounders

- strategy that does not require to identify, let alone measure, all potential confounders

- Key insight in experimental design

Experiments as Fair Tests

- random assignment implies that the observed *and unobserved* factors that affect outcomes are equally likely to be present in the treatment and control groups

- Any given experiment may overestimate or underestimate the effect of the treatment, but if the experiment were conducted repeatedly under similar conditions, the average experimental result would accurately reflect the true treatment effect.

Replicability

- Experiments are fair in another sense: they involve transparent, reproducible procedures. The steps used to conduct a randomized experiment may be carried out by any research group.

- Because treatments assignment precedes the measurement of outcomes, it is also possible to spell out beforehand the way in which the data will be analyzed.

Creativity vs excessive discretion

- Advantage of experimentation: Making one think clearly about hypotheses *before* the empirics are launched
- What is the independent variable? What is the dependent variable?
- What are the practical or theoretical implications of estimating this causal relationship?

Other Advantages

- Experiments allow us to study the effects of phenomena that seldom occur naturally
  - (e.g., out-of-equilibrium phenomena)
- Systematic experimental inquiry can lead to the discovery and development of new interventions
- Momentum of experimental programs: each finding leads to new questions

What is an experiment?

- “Experiment” in common parlance vs. experiment as a term of art: *random assignment of observations to treatment and control conditions such that every unit has the same ex ante probability of receiving the treatment*

- Perfectly controlled experiments in the physical sciences versus randomized experiments in the behavioral sciences

- Confusion between random *sampling* and random *assignment*

Different types of experiments

- Randomized experiment vs Natural experiment
- Laboratory experiments vs field experiments
- Artefactual vs Framed vs Natural field experiment

Different degrees of Fieldness

- Authenticity of treatments
- Participants
- Contexts
- Outcome measures

Lab vs field

- “program evaluations” (e.g. effect of ads)
- Test of theoretical propositions
- Difficulty of implementation

Are experiments feasible?

- Lost history of field-based interventions in social science
- Collaborating with those who allocate resources (sometimes at cross-purposes with other academics who regard randomization as optional or impractical)
- Seizing opportunities presented by naturally occurring randomized assignment

Are experiments necessary?

- Not when the counterfactual outcome is obvious (e.g., the effects of parachutes on the well-being of skydivers)
  - By the way, how do we know it's obvious?
- Not when there is little or no heterogeneity among the units of observation (e.g., consumer product testing)
- Not when the effect is so large that there is no reason to suppose that it is due to unobserved heterogeneity
- …for most behavioral science applications, experiments are indispensable

Natural experiments and quasi-experiments

- Randomization conducted by subjects other than the researcher
- Downstream experiment
- Quasi experiments not involving random assignment

Potential Outcomes

- suppose we wish to study the budgetary consequences of having women, rather than men, head Indian village councils, which govern rural areas in West Bengal and Rajasthan

- assume that each village either receives the treatment (a woman as village council head) or remains untreated (its village council is headed by a man).

- Assume that we observe the share of the local council budget that is allocated to providing clean drinking water

7 villages

- Each village is identified by a subscript , which ranges from 1 to 7

- Let y_i (1) be the outcome if village is exposed to the treatment (a woman as village head), and let y_i (0) be the outcome if this village is not exposed to the treatment.

- Causal effect of the treatment $\tau_i \equiv Y_i(1) - Y_i(0)$.

Illustration of potential outcomes

Village i	y_i (0) Budget Share if Village Head is Male	y_i (1) Budget Share if Village Head is Female	τ_i Treatment Effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	**15**	**20**	**5**

The empirical challenge

- At any given time one can observe y_i (1) or y_i (0) but not both.

- Define y_i as the observed outcome in each village and D_i as the treatment that is delivered in each village. D_i =1 when a woman is village head and 0 otherwise.

- Observed budget: $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$.

Average Treatment Effect

- From the rightmost column of Table, we can calculate the average treatment effect for the seven villages.
- Sum of treatments divided by number of villages, 5 percentage points
- The ATE is a parameter, a constant that determines the causal relationship between treatment and outcome.

Random sampling and expectations

- Suppose that instead of calculating the average potential outcome for all villages, we drew a simple random sample of villages and calculated the average among the villages we sampled.

- By *simple random sample*, we mean a selection procedure in which ***v*** villages are selected from the complete list of ***N*** villages, and every set of villages is equally likely to be selected.

Number of possible random samples

- For example, if we select one village at random from a list of seven villages, seven possible samples are equally likely.

- If we select three villages at random from a list of seven villages the number of possible samples is:

$$\frac{N!}{v!(N-v)!} = \frac{7!}{3!4!} = 35$$

Sample average as random variable

- If potential outcomes vary from one village to the next, the average potential outcome in the villages we sample will vary, depending on which of the possible samples we happen to select.

- The average in our sample may be characterized as a *random variable*, a quantity that varies from sample to sample.

Expected value of a random sample average

- Under simple random sampling, the expected value of a sample average is equal to the average of the population from which the sample is drawn.

Example

- Suppose we sample two villages at random from the list of seven villages and calculate the average value of y_i (0) for the two selected villages.

- Total average value of y_i (0) is 15

- given sample of 2 villages might contain average values that are higher or lower than 15

- but the expectation refers to what we would obtain on average if we were to examine all possible random samples

Example

- Total number of possible way to sample 2 villages:

$$\frac{N!}{v!\,(N-v)!} = \frac{7!}{2!\,5!} = 21$$

- Each sample is equally likely

- The 21 possible samples, on average, produce an estimate of 15.

- In other words, the expected value of the average $y_i(0)$ obtained from a random sample of 2 villages is 15.

Expectations and averages

- the expected value of y_i (0):

$$E[Y_i(0)]$$

- The average value of a population of y_i (0):

$$\bar{Y}(0)$$

- Due to the property of expected values of a random sample average:

$$E[Y_i(0)]=\bar{Y}(0)$$

Conditional expectations

- Expectation in a subgroup
- the expectation of y_i (1) for those villages that receive the treatment:

$$E[Y_i(1)|D_i = 1]$$

Average treatment effect (ATE)

$$ATE \equiv E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)].$$

- In our village example, we see that the column of numbers representing the treatment effect is, on average, 5.
- If we were to select villages at random from this list, we would expect their average treatment effect to be 5.

A missing data problem

- The challenge of estimating the average treatment effect is that at a given point in time each village is either treated or not
- Example: village 1 and 7 assigned to treatment

Illustration of potential outcomes

Village i	y_i (0) Budget Share if Village Head is Male	y_i (1) Budget Share if Village Head is Female	τ_i Treatment Effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	**15**	**20**	**5**

Example: 1 and 7 treated

Village i	$y_i(0)$ Budget Share if Village Head is Male	$y_i(1)$ Budget Share if Village Head is Female	τ_i Treatment Effect
Village 1	?	15	?
Village 2	15	?	?
Village 3	20	?	?
Village 4	20	?	?
Village 5	10	?	?
Village 6	15	?	?
Village 7	?	30	?
Estimated Average Based on Observed Data	**16**	**22.5**	**6.5**

A missing data problem

- Random assignment addresses the “missing data” problem by creating two groups of observations that are, in expectation, identical prior to application of the treatment.

- In expectation, the treatment group’s potential outcomes are the same as the control group’s

More formally

- In the treatment group:

$$E[Y_i(1)|D_i = 1] = E[Y_i(1)].$$

$$E[Y_i(1)|D_i = 0] = E[Y_i(1)].$$

$$E[Y_i(1)|D_i = 1] = E[Y_i(1)|D_i = 0]. \qquad (2)$$

- Same logic in the control groups

$$E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1]. \qquad (3)$$

ATE

- Using:

$$ATE \equiv E[\tau_i] = E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)].$$

- Together with (2) and (3)

$$ATE \equiv E[\tau_i] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0].$$

An empirical strategy!

- If we want to estimate the average treatment effect, the previous equation suggests that we should take the difference between two sample means: the average outcome in the treatment group minus the average outcome in the control group.

Random Assignment and Unbiased Inference

- An estimator is unbiased if it generates the right answer, on average.

- In other words, if the experiment were replicated an infinite number of times under identical conditions, the average estimate would equal the true parameter.

Practice vs theory

- In practice, we will not be able to perform an infinite number of experiments. In fact, we might just perform one experiment and leave it at that.

- in theory we can analyze the properties of our estimation procedure to see whether, on average, it recovers the right answer.

Unbiasdness

- Because the units assigned to the control group are a random sample of all units, the average of the control group outcomes is an unbiased estimator of the average value of y_i (0)

- The same goes for the treatment group

Formally

- If we order the first m observation in the treatment group, and the remaining N-m are in the control

$$E\left[\frac{\sum_{1}^{m} Y_i}{m} - \frac{\sum_{m+1}^{N} Y_i}{N-m}\right] = E\left[\frac{\sum_{1}^{m} Y_i}{m}\right] - E\left[\frac{\sum_{m+1}^{N} Y_i}{N-m}\right] =$$

$$= E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] =$$

$$= E[Y_i(1)] - E[Y_i(0)] = E[\tau_i] = ATE.$$

Mechanisms of randmization

- How to randomize
- Threat to randomization: the human factor

Selection bias when randomization is not used

$$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] =$$

$$= E[Y_i(1) - Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0].$$

- Under random assignment, the selection bias term is zero, and the ATE among the (randomly) treated villages is the same as the ATE among all villages.

Example

- Suppose that, if left to their own devices, Village 5 and Village 7 each elect a woman due to villagers' pent-up demand for water sanitation. Self-selection in this case leads to an exaggerated estimate of the ATE because receiving the treatment is associated with lower-than-average values of y_i (0) and and higher-than-average values of y_i (1)

Illustration of potential outcomes

Village i	y_i (0) Budget Share if Village Head is Male	y_i (1) Budget Share if Village Head is Female	τ_i Treatment Effect
Village 1	10	15	5
Village 2	15	15	0
Village 3	20	30	10
Village 4	20	15	-5
Village 5	10	20	10
Village 6	15	15	0
Village 7	15	30	15
Average	**15**	**20**	**5**

Self-selection

- The average outcome in the treatment group is 25, and the average outcome in the control group is 16. The estimated ATE is therefore 9, whereas the actual ATE is 5.

- In this example, self-selection is related to potential outcomes;

- the comparison of treated and untreated villages recovers neither the ATE for the sample as a whole nor the ATE among those villages that receive treatment.

Two key assumptions

- each potential outcome depends *solely* on whether the subject *itself* receives the treatment.
- Excludability
- Non interference (SUTVA)

Excludability

- Define Z_i a variable that indicates which observations have been allocated to treatment or control. {allocated treatment}

- Recall the meaning of D_i {actual treatment}

Exclusion restriction

- Let $Y_i\ (z,d)$ be the potential outcome when $Z_i=z$ and $D_i=d$
- For example if $Z_i=1$ and $D_i=0$
- The exclusion restriction assumption is that $Y_i\ (1,d) = Y_i\ (0,d)$

Threat to excludability

- Example of Ngos and villages
- Asymmetry in measurement

Defending excludability

- Uniform handling of treatment and control
- Double blindness
- Parallelism
- Placebo design

Non interference (SUTVA)

- Outcome in a unit analysis in only affected by the treatment administered in that unit
- In our example it implies that the sanitation budget in one village is unaffected by the gender of the council heads in other villages

Other example

- Contagion
- Displacement
- Communication
- Social Comparison
- Signaling
- Persistence and Memory