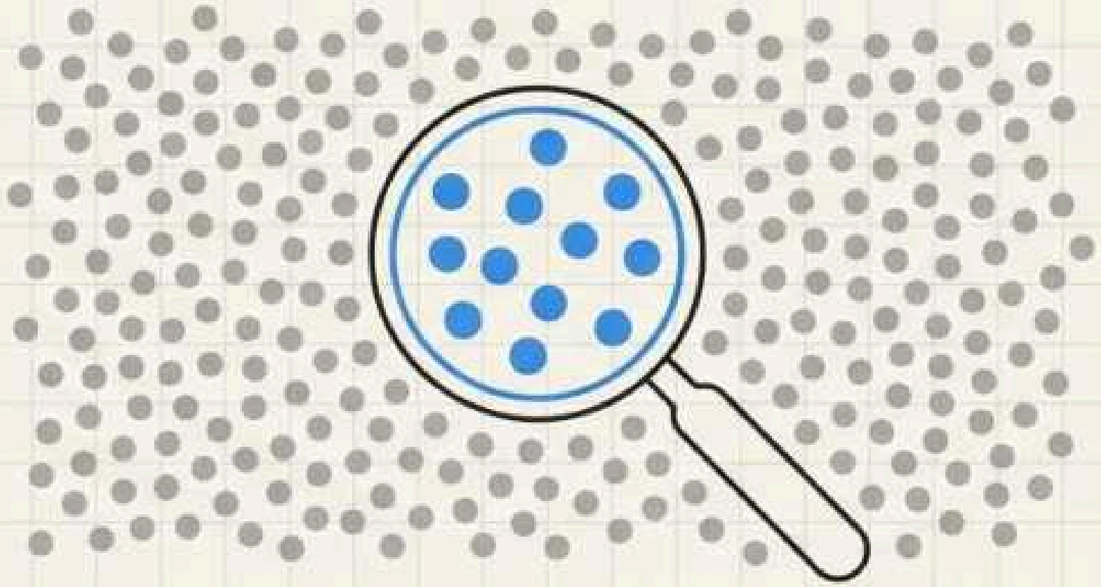
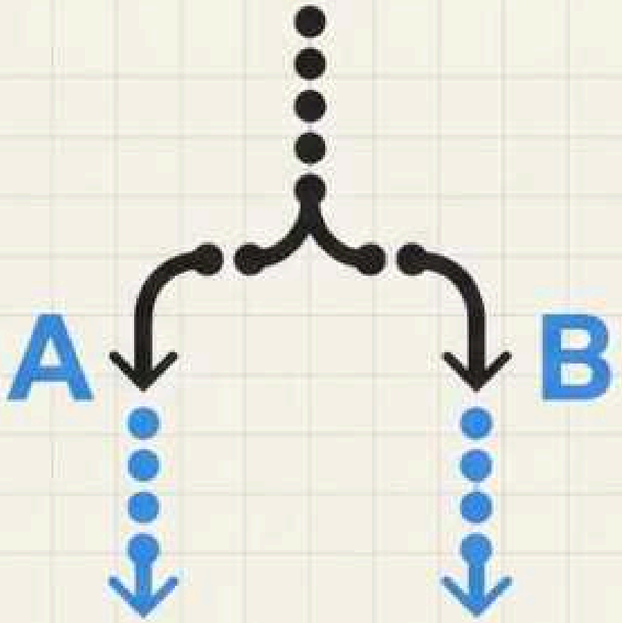
5.1 How should data be collected so conclusions are trustworthy?



SAMPLING: Randomness protects us from bias.



EXPERIMENTS: Randomness lets us find causes.

Together, these ideas are the most practical in all of statistics.

5.2

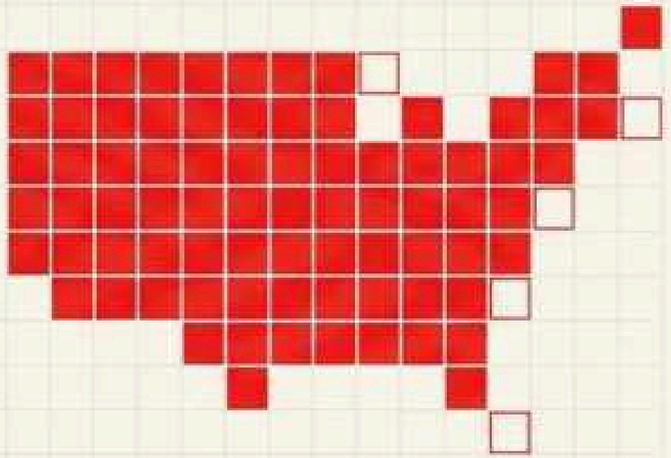
The 1936 Literary Digest Failure

2,400,000

The magazine predicts Alf Landon will win the presidency.

Out of 10 million mailed ballots.

61%



Franklin Roosevelt wins in a landslide.

How could an enormous sample of 2.4 million people be so wrong?

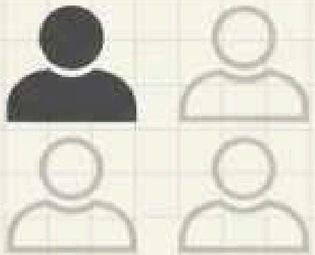
5.3

The Anatomy of a Flawed Sample



Source Bias

Names drawn from phone directories and car registrations. During the Great Depression, this skewed heavily toward wealthier demographics who favored Landon.



Response Bias

Only a ~25% response rate. Those who took the time to mail it back were not typical of the average voter.

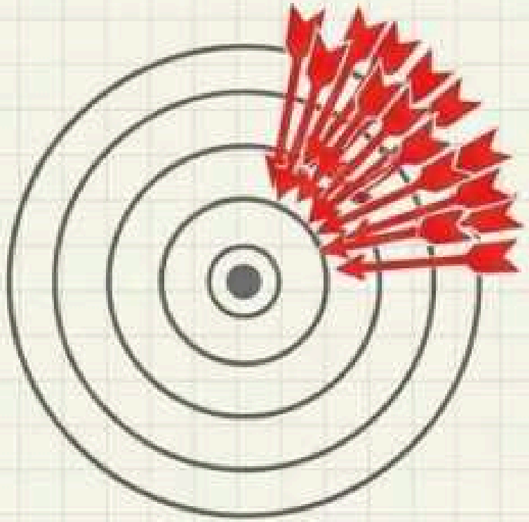
A huge sample cannot repair a biased way of choosing it.

5.4

Bias vs. Random Error

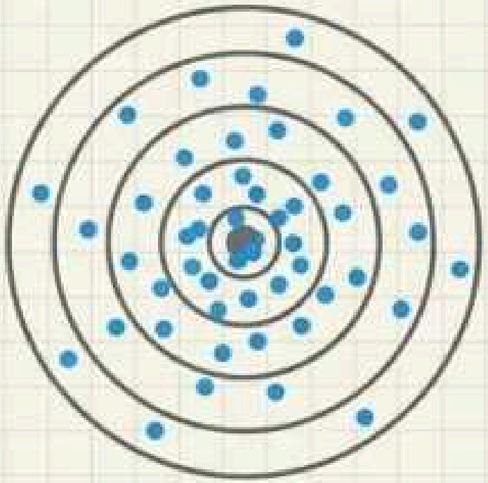
Bias

A systematic precision bullseye in acquorcinatal target.



A systematic tendency to miss the truth in one direction.

Requires a better methodology, not a larger sample.



Unpredictable variation that goes up or down but averages out.

Random Error

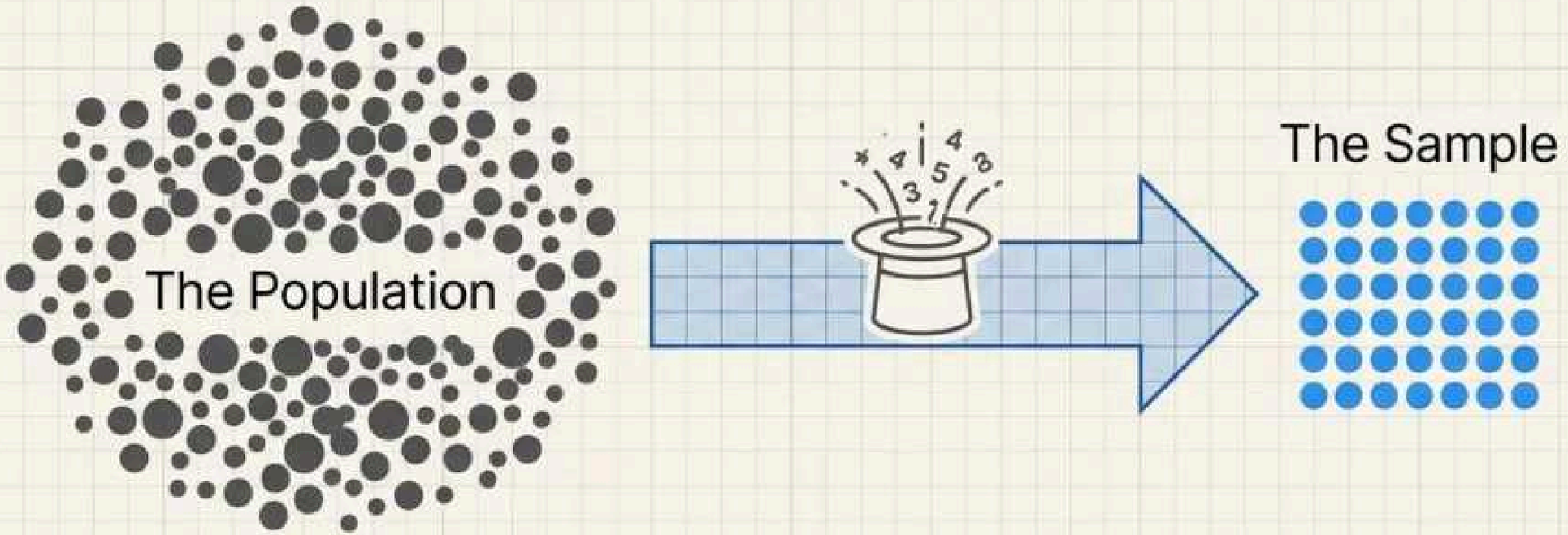
Unpredictable doets in averaging up dout out to bullseye.

Shrinks with a bigger sample size.

George Gallup correctly predicted FDR's win with a carefully chosen sample sample of just 50,000 people—proving methodology beats sheer volume.

5.5

The Remedy for Bias is Chance



	1. Every individual has the same chance of being chosen. 2. Every group of the given size is equally likely to be the sample.	

Random choice has no favorites,
so it cannot lean in one direction.

5.6

Non-Random Methods (What to Avoid)



Convenience Sample

Taking whoever is easiest to reach.



Voluntary Response Sample

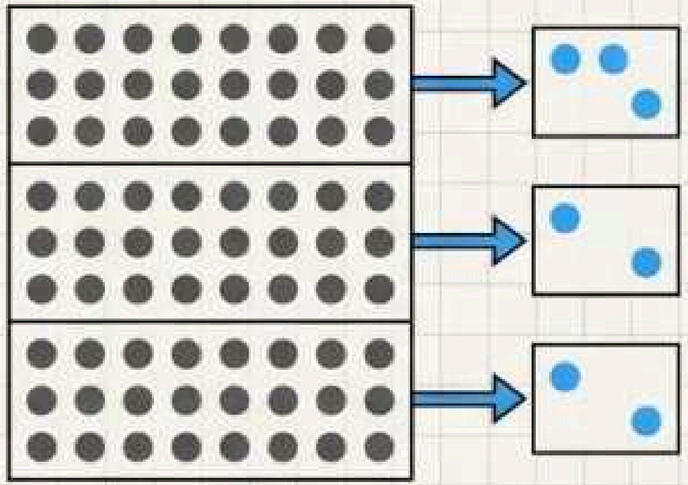
Letting people choose to participate.
Attracts extreme/strong opinions.

These methods tell us absolutely nothing reliable about the whole population.

5.7

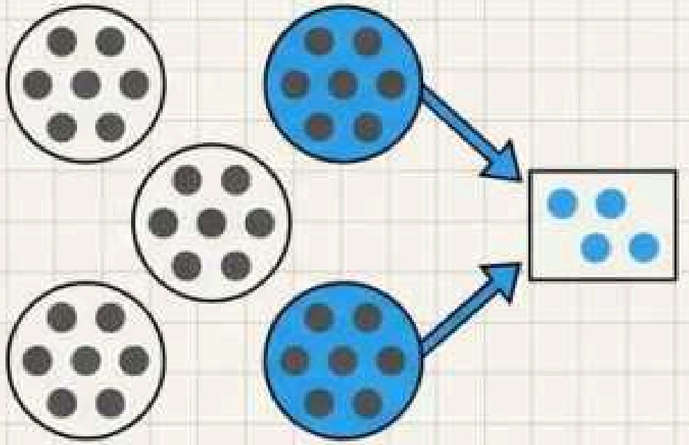
Elaborate Random Designs

Stratified Sample



Split the population into strata (groups), then take a random sample from each.

Cluster Sample



Randomly choose whole groups (clusters), and survey everyone inside them.

Systematic Sample

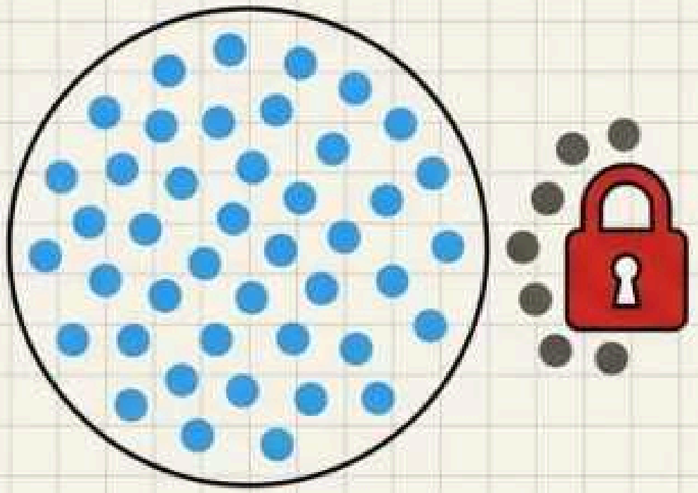


Pick every nth (e.g., 10th) name from a randomly started list.

5.8

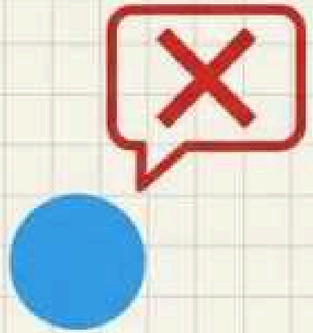
When Good Designs Go Wrong

Undercoverage



Segments of the population have zero chance of being chosen (e.g., people without phones).

Nonresponse



Chosen people simply do not answer the survey.

Response Bias



Untrue answers, or steering via question wording.

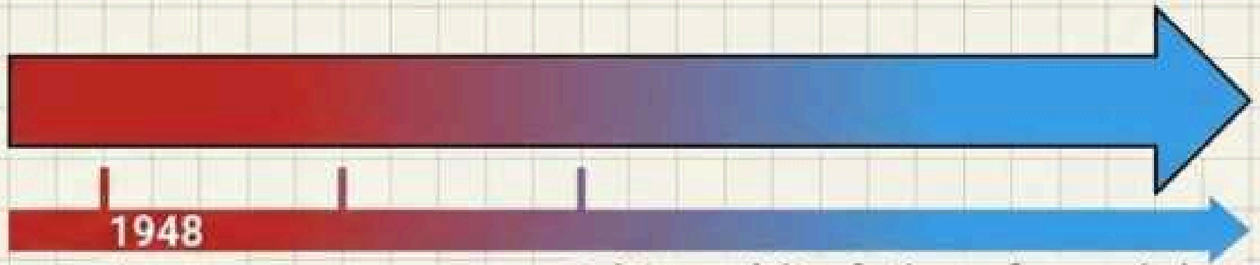
Asking about a Death Tax yields entirely different data than asking about an Estate Tax.

5.9

The 1948 Polling Failure



Pollsters used **quota sampling** (interviewers choosing people to fill demographic buckets). The result was a catastrophic false prediction.

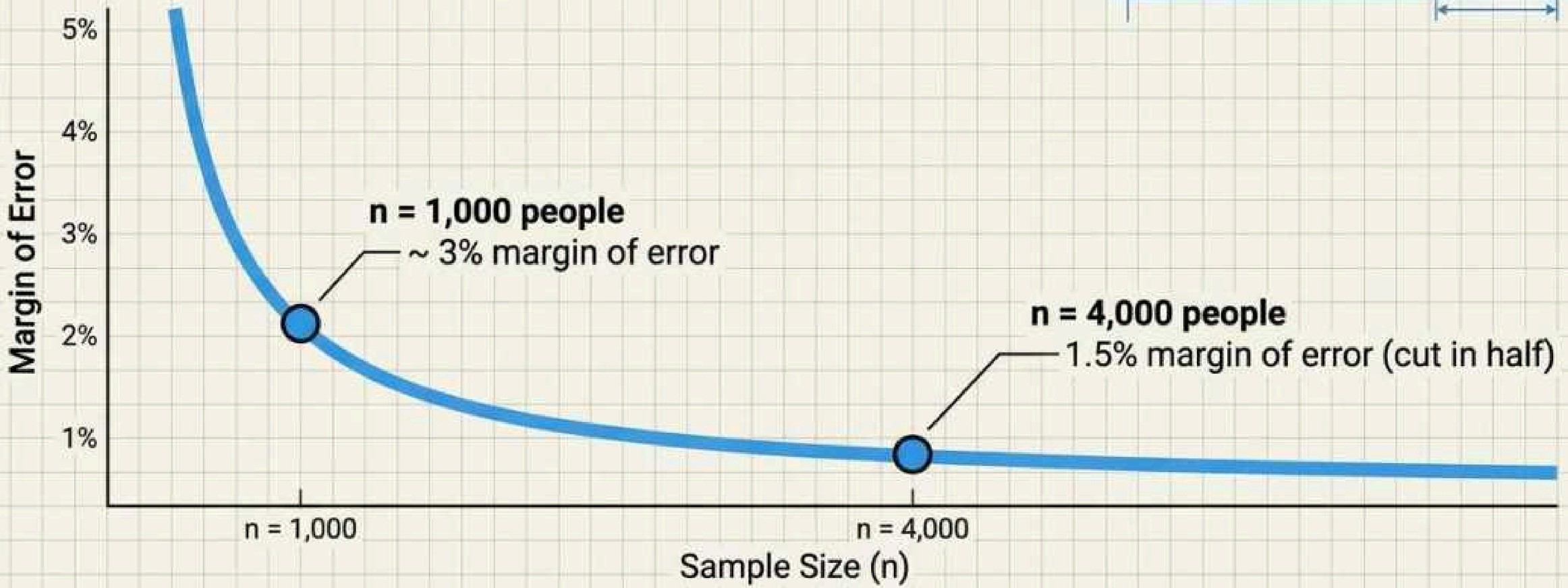


This public failure forced the industry to abandon quota sampling. Today, good surveys rely strictly on random selection and aggressive follow-up with nonresponders.

5.10

Sample Size & Margin of Error

$$\text{Margin of Error} \approx \frac{1}{\sqrt{n}}$$



Error shrinks as the sample grows, but slowly.

5.11

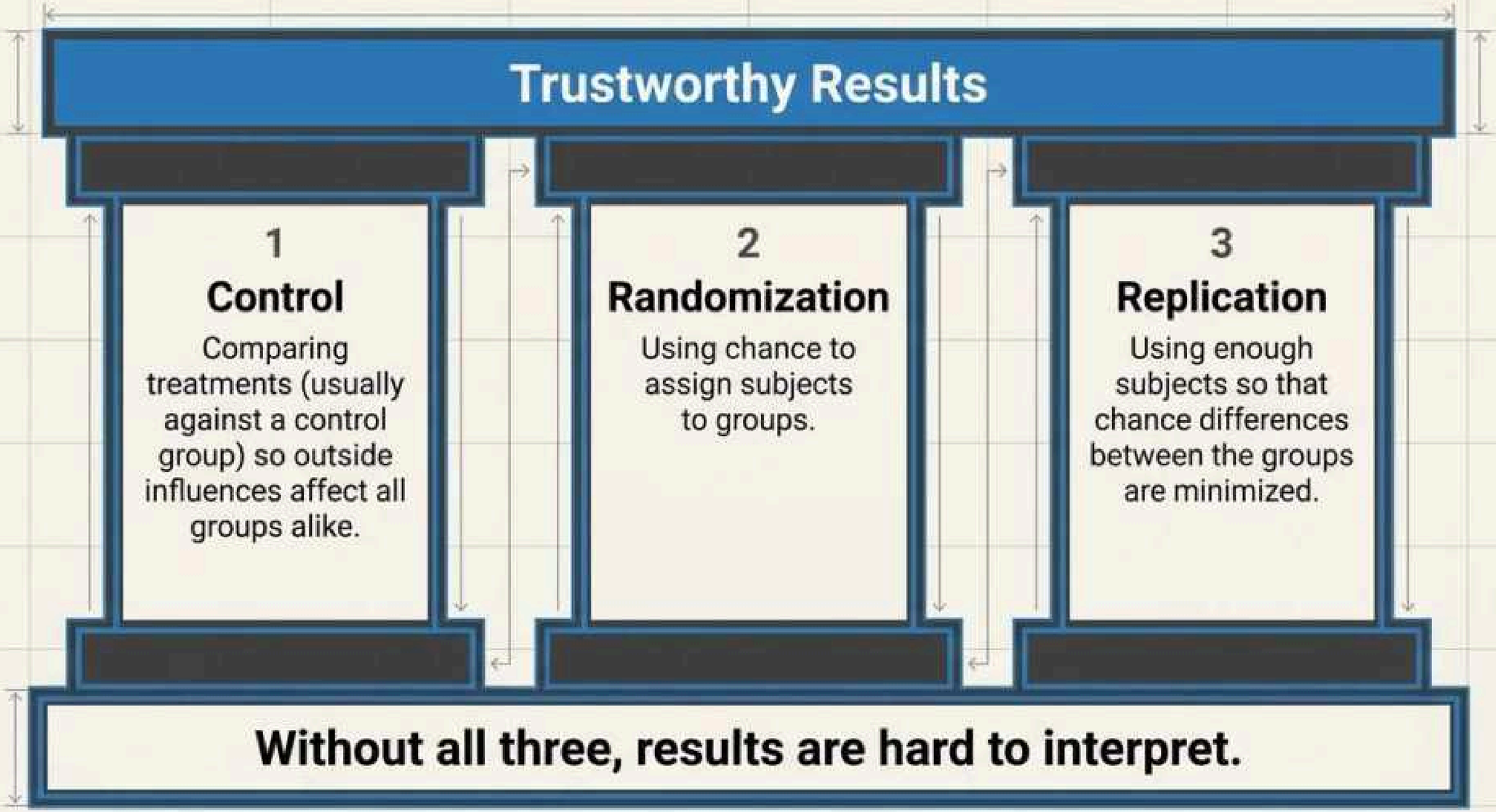
Moving to Experiments (Studying Cause)



Specific values applied are called Treatments.

Because we impose the treatment, we can isolate its effect and uncover true causation.

5.12 The Three Principles of Experiments

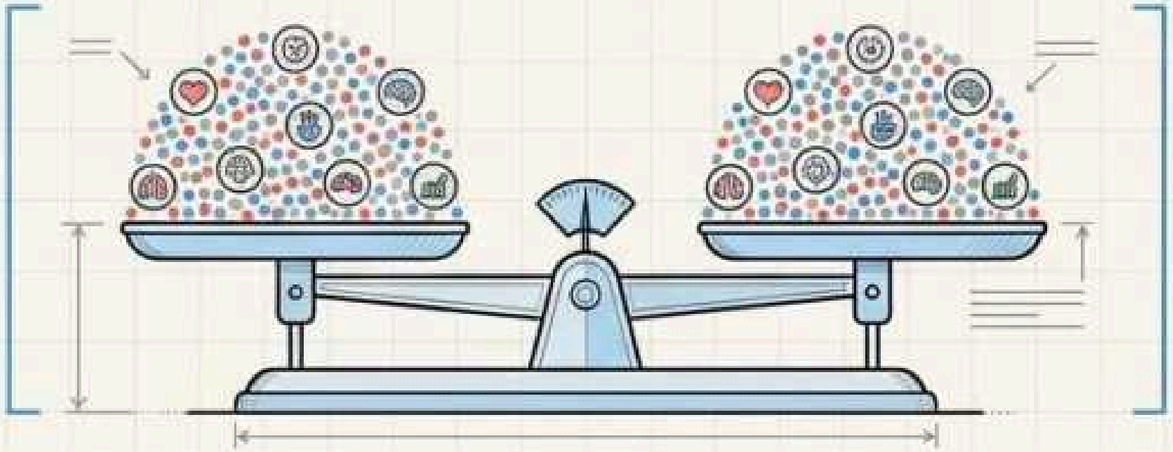


5.13

Why Randomize? (Balancing Lurking Variables)



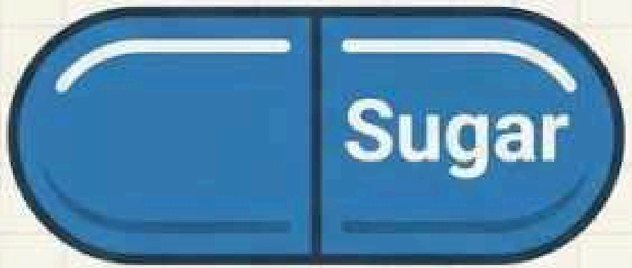
If doctors choose, they may pick healthier patients for the drug. The drug appears to work even if it does nothing (false positive).



Chance assigns healthy/sick, young/old, and motivated/unmotivated equally across groups. It balances factors **we haven't even thought of.**

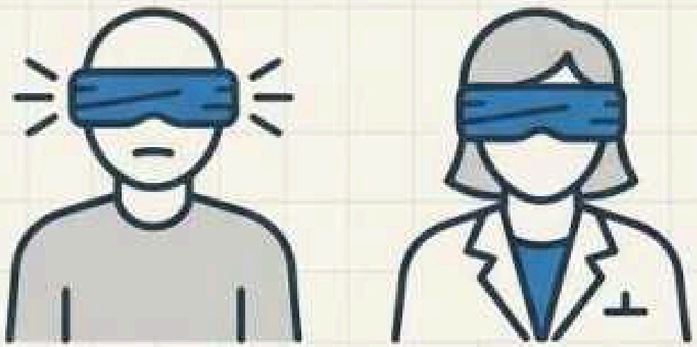
5.14

The Placebo Effect & Blinding



The Placebo Effect

People improve simply from the idea of being treated, even with no active ingredient. Addressed by giving the control group a placebo.



Double-Blind Design

Neither the subjects nor the staff measuring outcomes know who got which treatment.

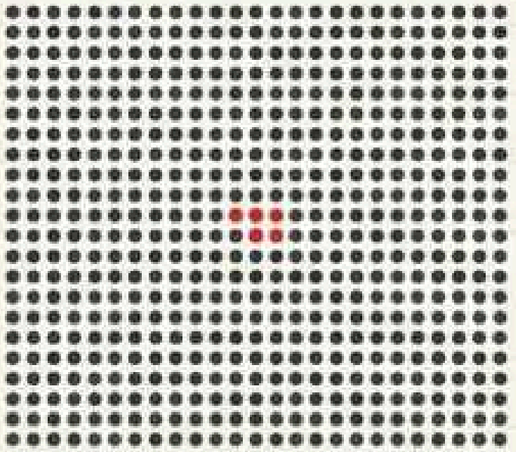
Expectations cannot color the results.

5.15

The 1954 Salk Polio Vaccine Trial

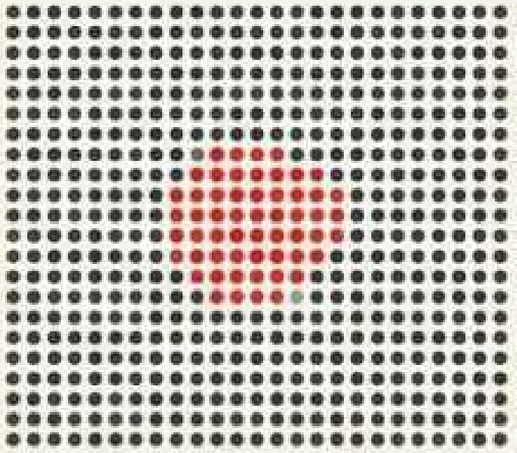
400,000 children in a randomized, double-blind experiment.

Vaccine Group (200,000 kids)



Polio struck **28 per 100,000**.

Placebo Group (200,000 kids)

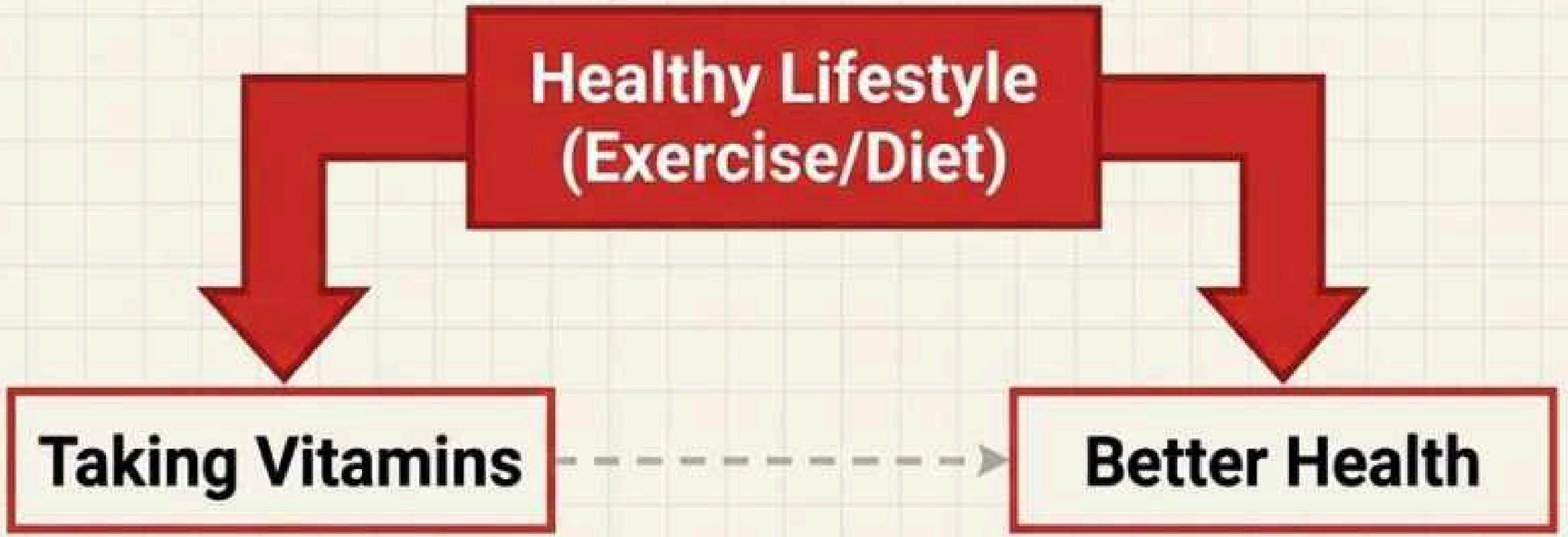


Polio struck **71 per 100,000** (salt water injection).

The resulting gap was far too massive to be attributed to chance. The vaccine worked.

5.16

The Danger of Observational Studies



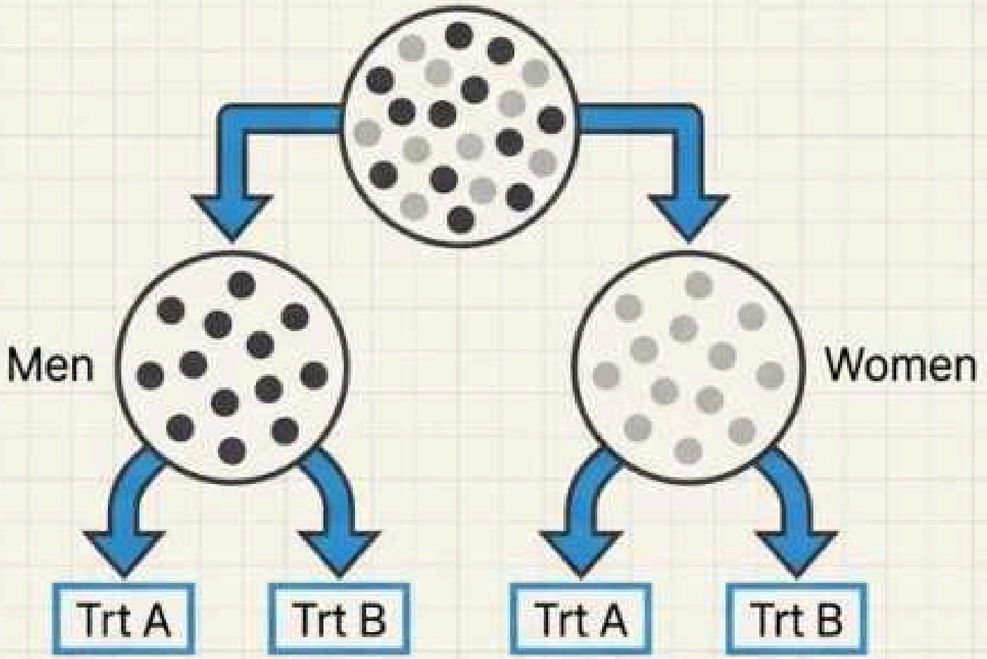
Several famous observational studies suggested hormone therapy protected women's hearts. When randomized trials were finally done, the benefit vanished and harm appeared.

Choice is a lurking variable.

5.17

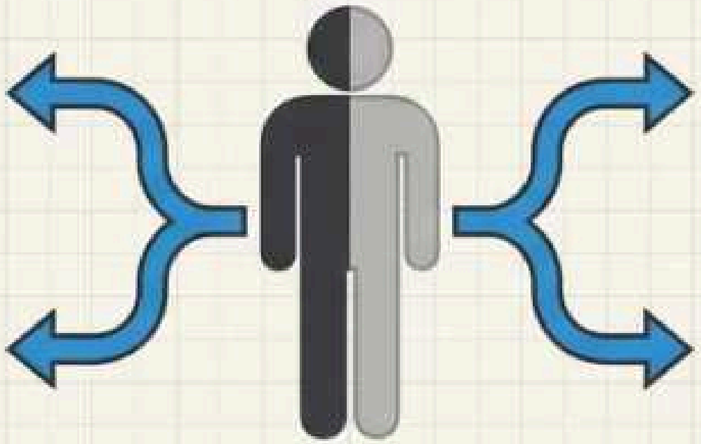
Reducing Variation: Blocks & Pairs

Blocking



Group similar subjects into blocks first, then randomize treatments within those blocks.

Matched Pairs



Each subject receives both treatments in random order, or twin pairs are split.

Both designs compare like with like to detect smaller differences.

5.18

The Two Uses of Randomness

Random Sampling (WHO)

Choosing the subjects from a larger pool.

Lets us GENERALIZE to the population.

Random Assignment (WHAT)

Deciding which group the chosen subjects go into.

Lets us CONCLUDE CAUSE.

A study can have one, both, or neither—and its conclusions strictly depend on which. Keep them apart!

5.19 Ethical Duties in Human Experiments

1. Informed Consent

2. Weigh Risks vs. Benefits

3. Data Privacy

The US Public Health Service Syphilis Study (1932–1972). Black men were left untreated and deceived about their condition, even after penicillin was available.

↓

Public outrage led to strict rules requiring Institutional Review Boards (IRBs) for all research on people.

5.20

Producing Good Data: The Summary

The Enemy & The Remedy

Bias, not size, is the great enemy of a sample.

Defeated by: Random selection.

The Mechanism

Experiments require Control, Randomization, and Replication (plus placebos/blinding).

Purpose: Defeats lurking variables to find causes.

The Ultimate Output

- Random Sampling → **Generalization**
- Random Assignment → **Causation**

Next Up: Lecture 6 introduces the mathematics of chance that underlies all of this.