Hector Bahamonde, PhD, Docent

Experimental Methods in Social Science

INWS0059

Today's Agenda

Causal Inference in Social Sciences

- **Housekeeping:**
  - Introduce myself.
  - You will introduce yourself.
  - Go over the syllabus.
- **Today's lecture:**
  - General overview of how experimentalists work.
  - The counterfactual model of causal inference.
  - Econometrics and experiments: what's "best"?

Brief Introduction

Experimental Methods in Social Sciences—INWS0059

- Hector Bahamonde.
- PhD in **Political Science**.
- **Senior Researcher** at INVEST.
- Title of **Docent** in Political Science (UTU's Faculty of Social Sciences).
- I have taught in the United States, Chile and Finland before.
- At UTU, I **teach** **statistical** and **experimental methods**.
- I study the **political consequences of economic inequality**.
- My data are usually **experimental**. I usually use lots of **econometrics** too.
- More info: **www.HectorBahamonde.com**

Students introduce themselves

State your name, your dependent variable and your program (MA/PhD)

Go over syllabus

Causal Inference

in Social Sciences

Hector Bahamonde, PhD

Overview

Causal Inference

- Do qualitative methods (**?**) care about "**studying a population**"**?**
- Experimentalists should specify the **population**: "*yes, but*…"
  What happens if you're interested in, e.g., MP's?
- Make sure the sample **reflects the population**: **why is this important?**

Overview

Causal Inference

- **Context** in which the experiment takes place must be “*realistic*.”
  What does this mean, and how an experiment *cannot* be “realistic" (**?**)
  - Can be “*more realistic context*” be more more of a *problem* than an *advantage*?
- **Salience**: Topic must be something the studied sample cares about.
  Think about, e.g., “monetary policy.”
  Do citizens, in general, understand/have consistent preferences about, say, the interest rate?

Causal Inference and Experiments

The “Ideal Experiment”

- **Let’s device an experiment**. Topics**?**
  - What’s the **research question?**
  - What’s the “***treatment***" group**?**
  - What’s the “***control***” group**?**
    But…do we *always* need a control group**?**

How do we calculate the effects of a treatment?

Quantities of Interest

In the potential outcomes framework

The Treatment Effect τ

A "*naive*" version

- The treatment effect is essentially a **subtraction**:

$$\tau_i = y_i(1) - y_i(0)$$

- It's the **difference** between the **treatment state** $y_i(1)$ for individual i and the **control state** $y_i(0)$ for the same individual.

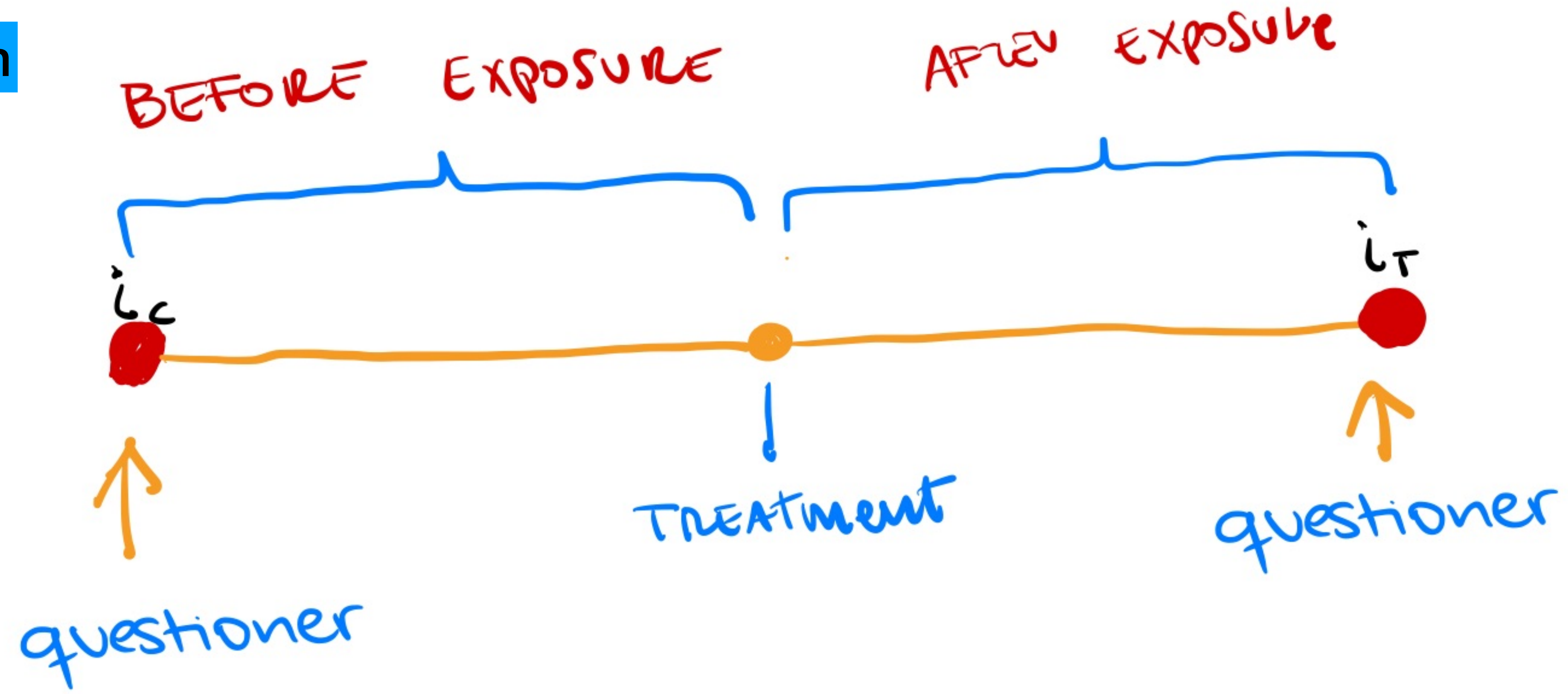
- Can you name an example**?**

The Treatment Effect τ

The Fundamental Problem of Causal Inference

- Based on this naive calculation of the treatment effect τ, there is a "*fundamental*" problem (**?**).
- What's the "*Fundamental Problem of Causal Inference*"**?**
  - It's **impossible** to observe the value of the treatment state $y_i(1)$ and the control state $y_i(0)$ at the same time. Why**?**
  - $\tau_i = y_i(1) - y_i(0)$
    Both states cannot be observed at the same time.
- How do we solve this problem**?**

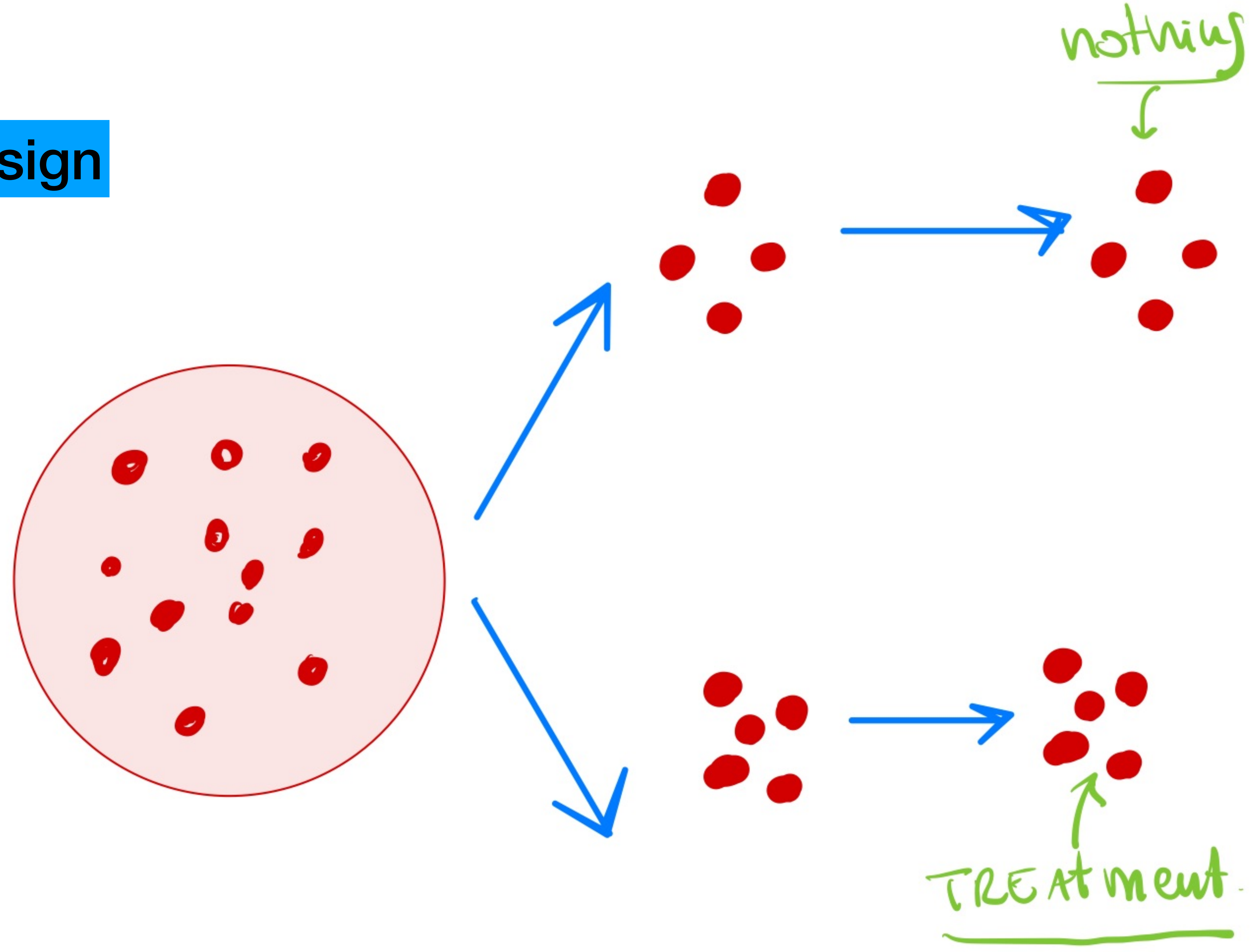
Within-subjects design



$$\boxed{\text{Treatment Effect} = i_T - i_c}$$

How can we estimate the treatment effect in this design?

Between-subjects design



How can we estimate the treatment effect in this design?

The Treatment Effect τ

Assumptions

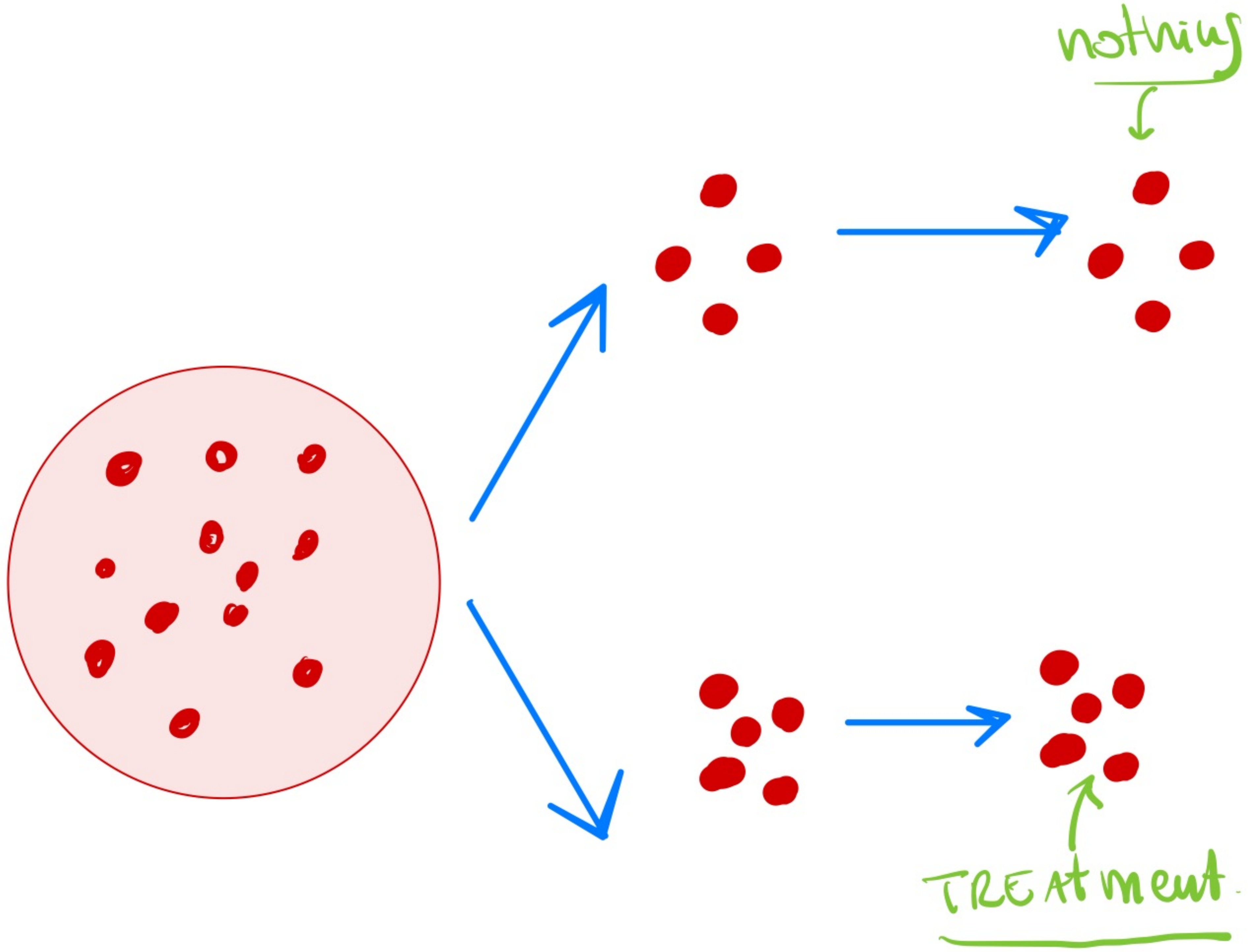
- Other important assumptions:

    1. No "**spillover**" effects (**?**) and **SUTVA** (**?!**).

    2. **Independence** (treatment status is independent of potential outcomes. "**Ignorable**."

    $$(Y^0, Y^1) \perp D$$

    Why is this important**?**

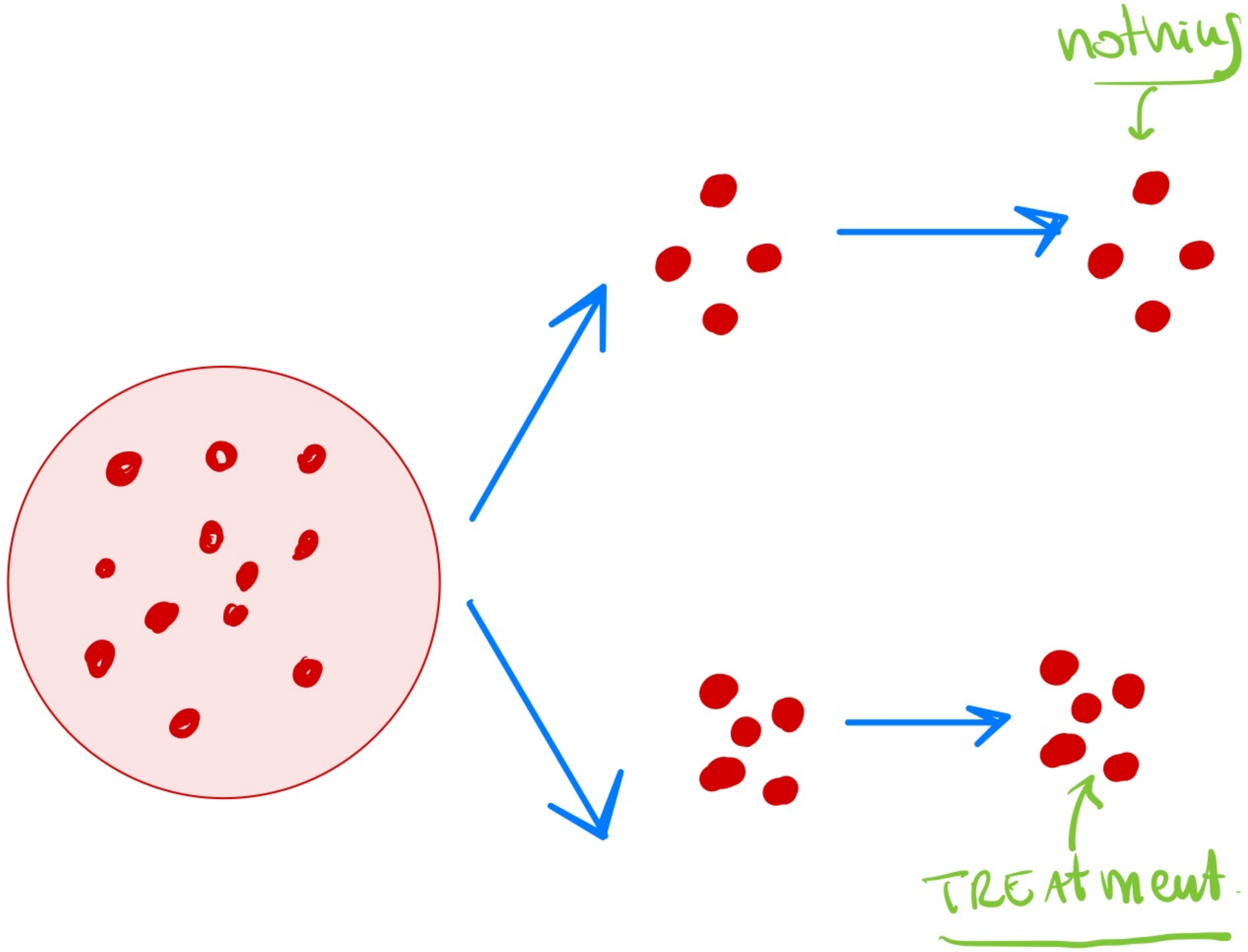


The **Average** Treatment Effect δ

ATE

- **The question still stands (fundamental problem in causal inference).**
  How do we **compute** the treatment effect using this model**?**

- The **Average Treatment Effect** (ATE):
  - $E[\delta] = E[Y^1] - E[Y^0]$
  - Notice we dropped the **individual** notation (i), and now we talk about **groups**.



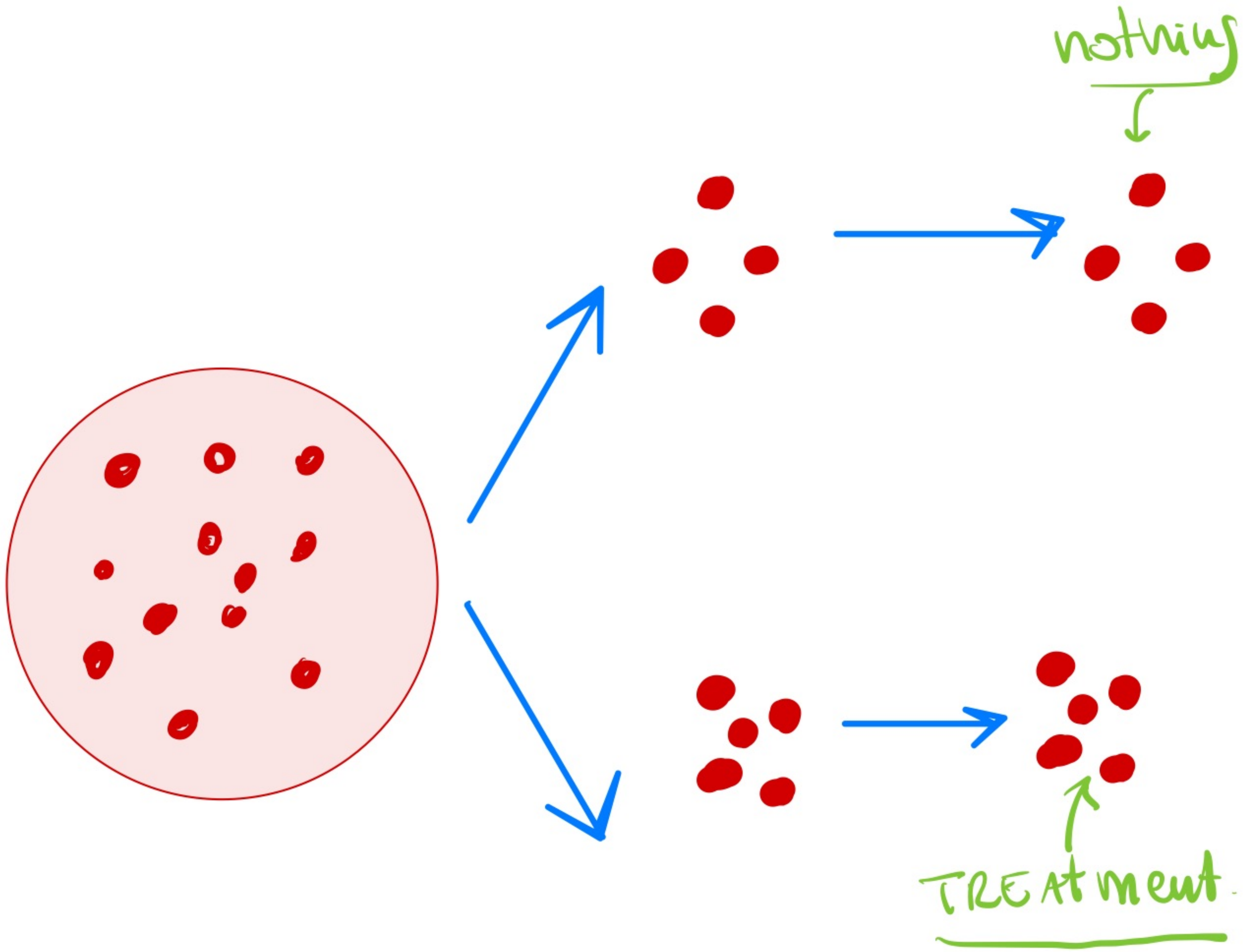
The Average Treatment Effect solves the Fundamental Problem

$$\tau_i = y_i(1) - y_i(0)$$

$$E[\delta] = E[Y^1] - E[Y^0]$$

Randomization

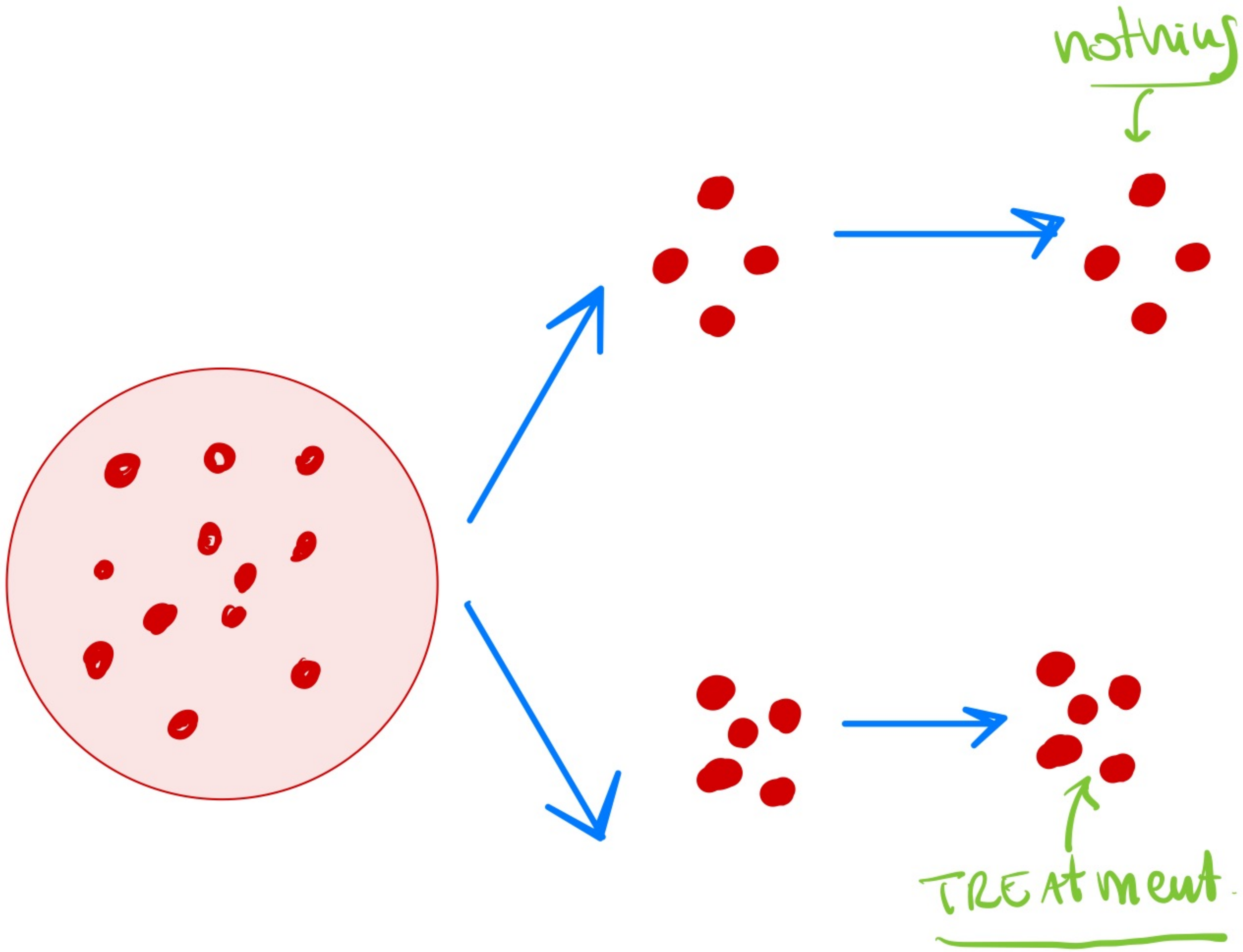
And Covariate Balance



Covariate Balance

Definition

- What's "**covariate balance**"**?**
  - **When the distribution of covariates (characteristics) is similar or balanced across different treatment groups.**
- Why do we **like** covariate balance**?**
- How do we **achieve** covariate balance**?**



Covariate Balance

Regression and Experiments

- **Regression setup**:
  - Balance is achieved by adding **control variables** (x_1):

    $$y_i = \beta_0 + \mathrm{T}\tau_1 + x_1\beta_1 + \epsilon_i$$

- What can go wrong**?**
  - **Omitted Variable Bias: Missing control variables *will* bias our results.**

☑ **Regression are based on assumptions.**

- **Experimental setup**:
  - Balance is achieved by **randomization**.
    Randomization ensures that *ALL* **observables and not observables** are balanced across treatment and control groups.
  - **Hence, balance is achieved by controlling for NOTHING**.

☑ **Experiments are based on designs.**

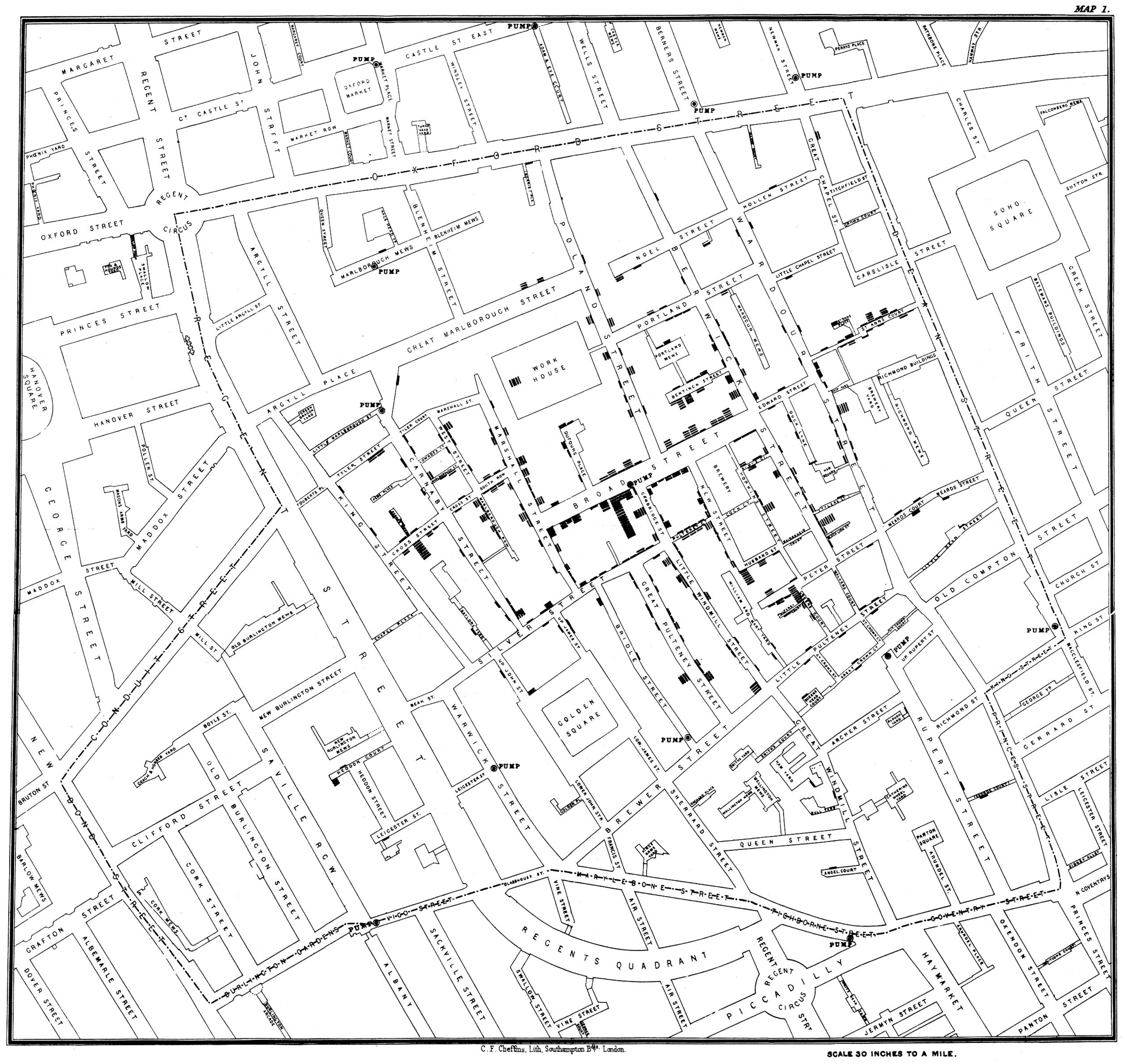
That's why experiments are the "shoe leather" of experimental methods

Ok, I get. It’s all about “*good designs*”

Let’s review an example of a good design.

What *causes* cholera?

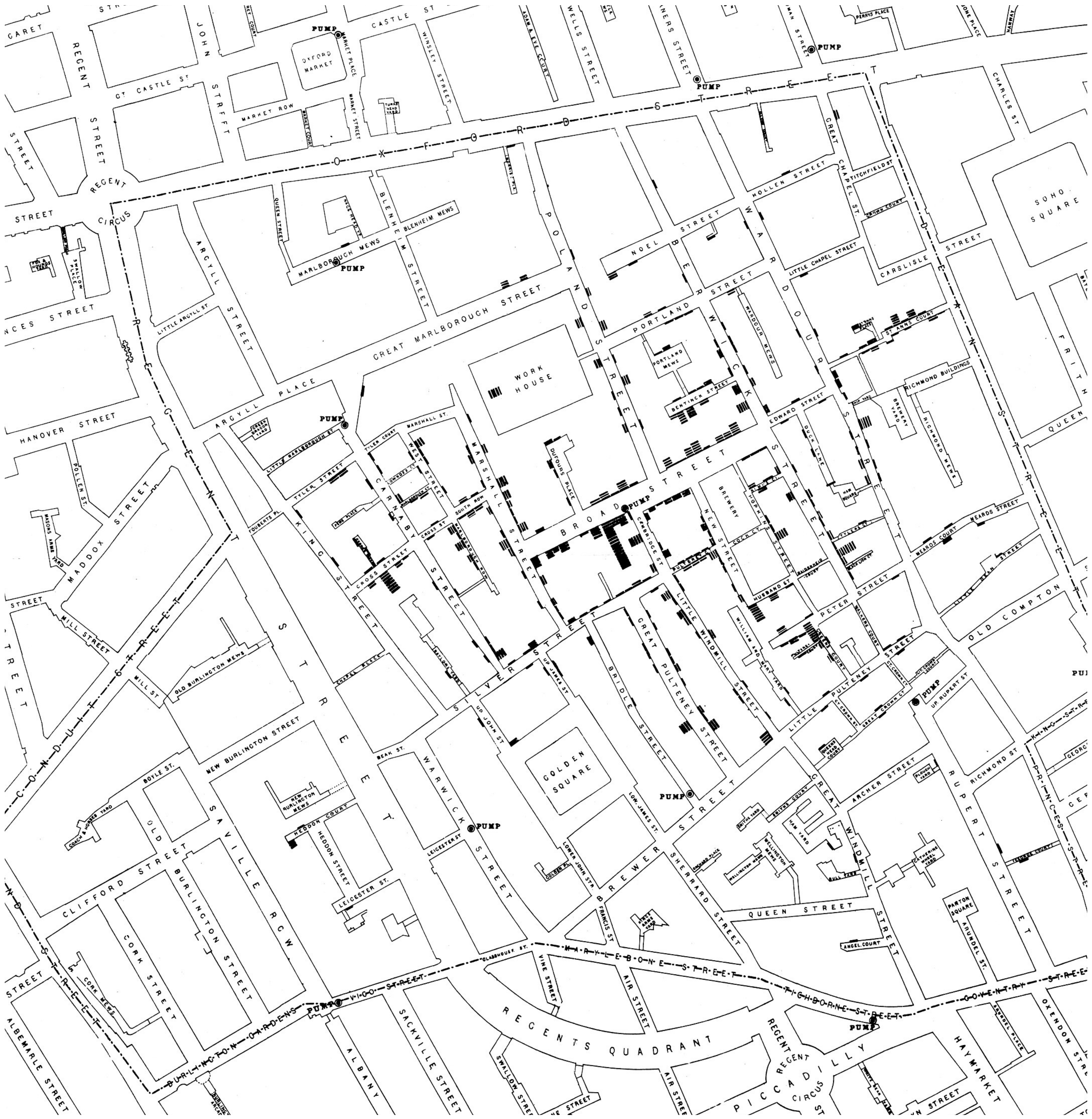
A study of water supply and miasma in London, 1854



Covariate Balance

Regression and Experiments

“In many cases a single house has a supply different from that on either side. Each company supplies **both rich and poor**, **both large houses and small**; there is no difference either in the **condition** or **occupation** of the persons receiving the water of the different Companies.”



Covariate Balance

Regression and Experiments

"In many cases a single house has a supply different from that on either side. Each company supplies **both rich and poor**, **both large houses and small**; there is no difference either in the **condition** or **occupation** of the persons receiving the water of the different Companies."

Control variables are included to obtain covariate balance

$$\text{cholera}_i = \beta_0 + \text{supply}\tau_1 + \text{income}\beta 1 + \text{size}\beta 2 + \text{condition}\beta 3 + \text{occupation}\beta 4 + \epsilon_i$$

Treatment

Covariate Balance

Regression and Experiments

- **Advantage of experimental methods over econometric technics**
  Experiments are based on designs, econometrics technics on assumptions.

- If both groups' characteristics (covariates) are balanced, then **the differences in outcomes** (cholera) **should be attributed to the treatment only** (water supply).

TABLE 1
Snow's Table IX

	Number of Houses	Deaths from Cholera	Deaths Per 10,000 Houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Covariate Balance

Regression and Experiments

- Main take aways:

    1. Regression analysis, even when "**statistical fixes**” are applied (structural equations, matching, robust estimators, GLS, etc.), they **CANNOT provide causal explanations**.

    2. While **regression** relies on statistical **assumptions** (**?**), **experiments** rely on transparent **designs** (**?**).

    3. **Random** assignment to treatment is the *only* way to get **causal** explanations:
       It achieves covariate balance of **observable *and* unobservable** characteristics—**no omitted variable bias** (**?**).