Causality & experiments

DS3

BRIAN QUISTORFF JAKE HOFMAN

MSR

Prediction

Make a forecast, leaving the world as it is

vs.

Causation

Anticipate what will happen when you make a change in the world

Prediction

Make a forecast, leaving the world as it is

(seeing my neighbor with an umbrella might predict rain)

vs.

Causation

Anticipate what will happen when you make a change in the world

(but handing my neighbor an umbrella doesn't cause rain)

Effect question

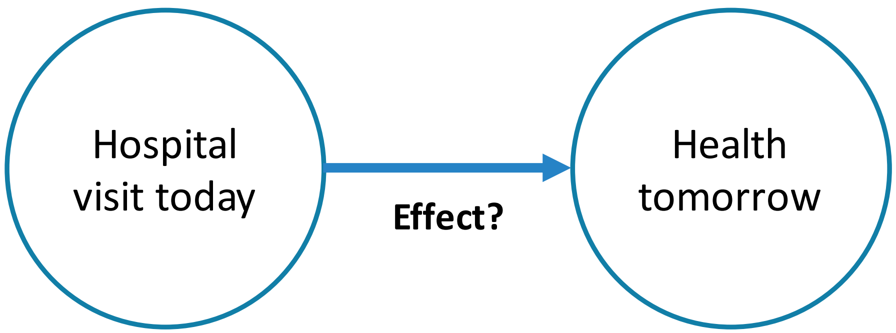
"What happens if we do X?", e.g.

- How does education impact future earnings?
- What is the effect of advertising on sales?
- How does hospitalization affect health?

John Stuart Mill (1843)

Example: Hospitalization on health

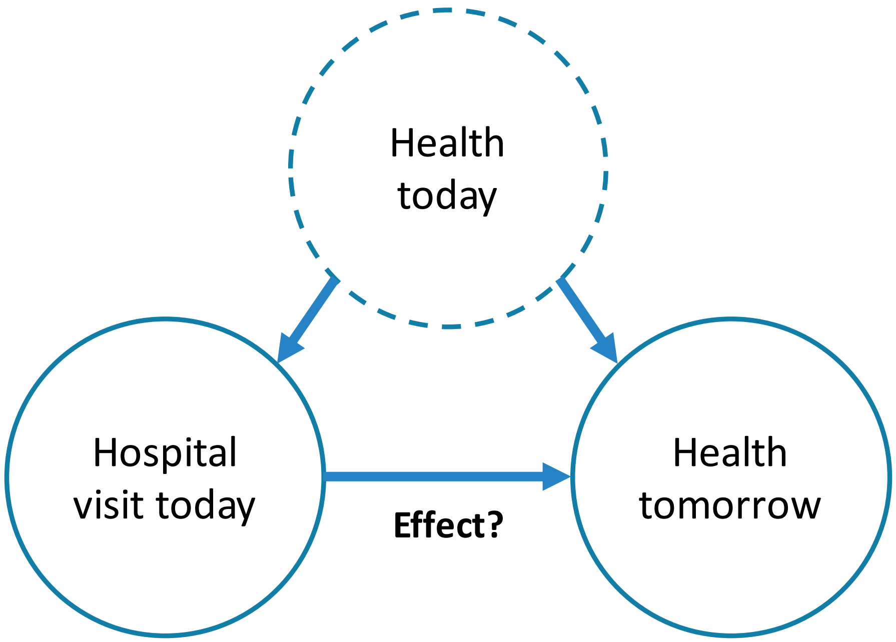
What's wrong with estimating this model from observational data?



Arrow means "X causes Y"

Confounds

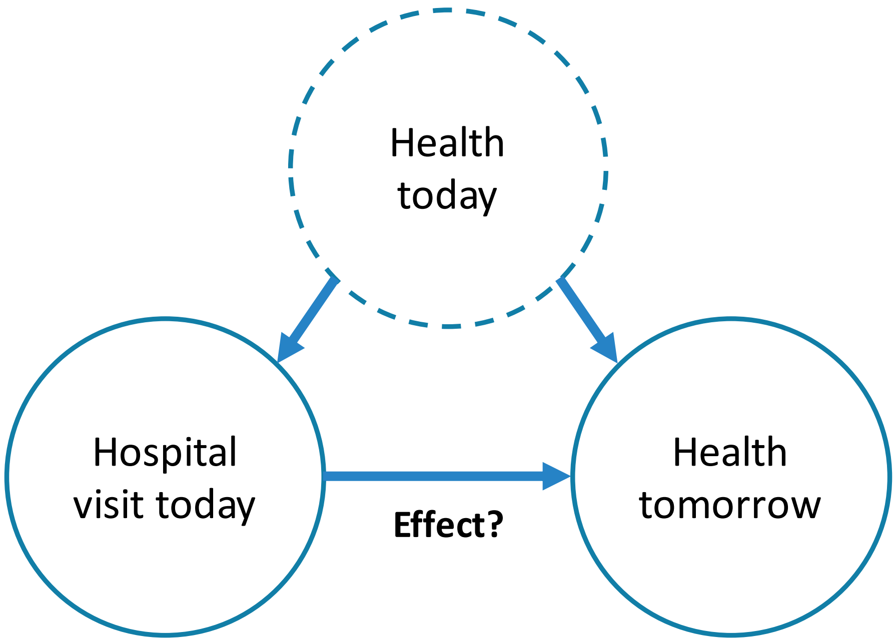
The effect and cause might be *confounded* by a common cause, and be *changing together* as a result



Dashed circle means “unobserved”

Confounds

If we *only get to observe them changing together*, we can't estimate the effect of hospitalization changing alone



Population Equation

Suppose the true model of the world is

$$y = \beta_v HospitalVisit + \beta_h Health + e$$

And that *HospitalVisit* is correlated with *Health*

- People with worse health are more likely to visit the hospital).

If we estimated a model that included *Health,* we would be fine.

- Our estimate of β_h would be close to the truth.

Omitted Variable Bias

Now suppose that we estimated a “mis-specified” model

$$y = \beta_v' HospitalVisit + u$$

The effect of health is no in the catch-all error term: $u = \beta_h Health + e$.

This means that $HospitalVisit$ is correlated with u.

Whenever a variable in a model is correlated with the error we say it is “endogenous”.

Coefficients on endogenous regressors will not be good. Called *Omitted Variable Bias.*

We should worry whenever our regressor is correlated with something in the error.

When this is caused by people’s choices we call this “Selection bias”

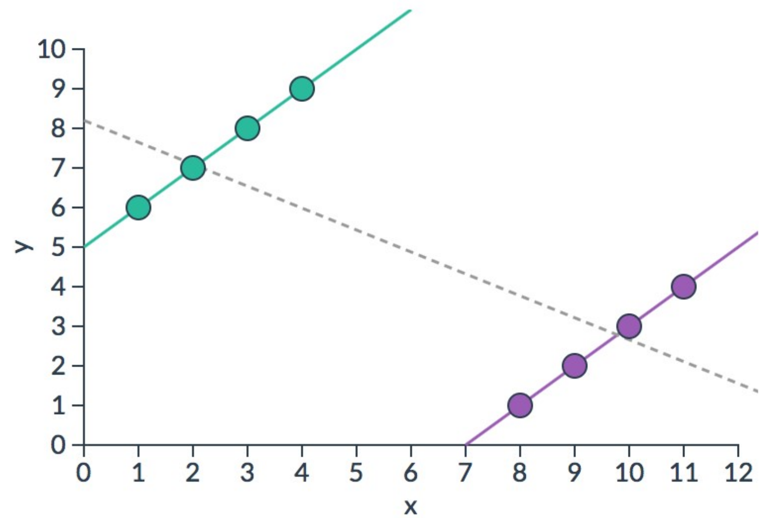
OVB

The amount of bias in our estimate of β_h is proportional to:

- The relation between the regressor and the confounder
- The relation between the confounder and the outcome.

Simpson’s paradox

Selection bias can be so large that *observational and causal estimates give opposite effects* (e.g., going to hospitals makes you less healthy)



http://vudlab.com/simpsons

Simpson's Paradox: Batting Averages

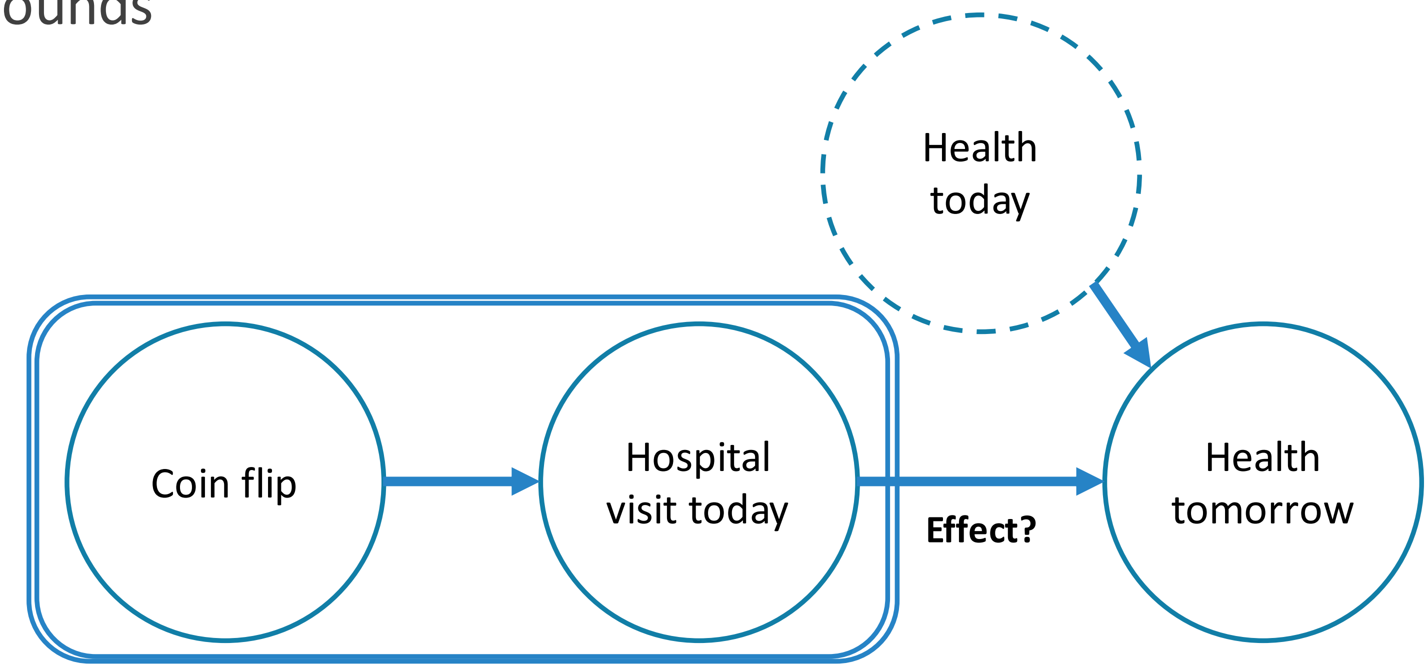
	1995		1996		Combined	
Derek Jeter	12/48	.250	183/582	.314	195/630	**.310**
David Justice	104/411	**.253**	45/140	**.321**	149/551	.270

“To find out what happens when you change something, it is necessary to change it.”

-GEORGE BOX

Random assignment

Random assignment determines the treatment independent of any confounds

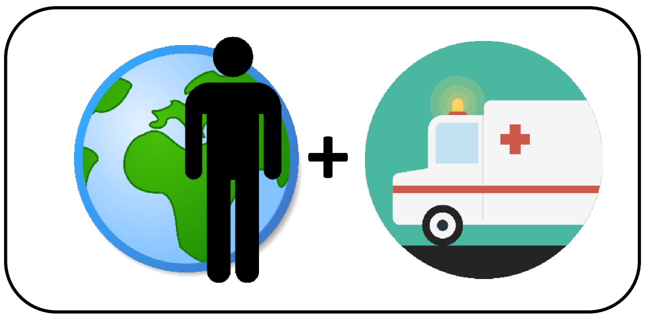


Double lines mean
“intervention”

Counterfactuals

To isolate the causal effect, we have to *change one and only one thing* (hospital visits), and compare outcomes

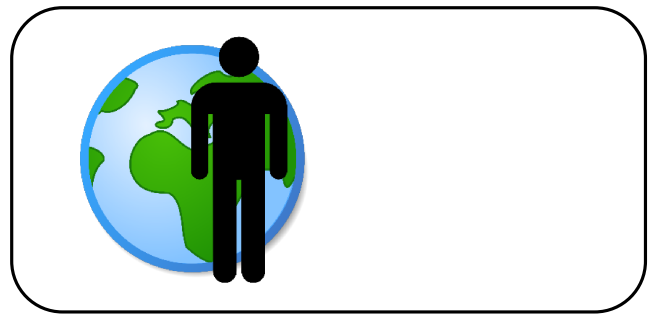
Reality



(what happened)

vs

Counterfactual

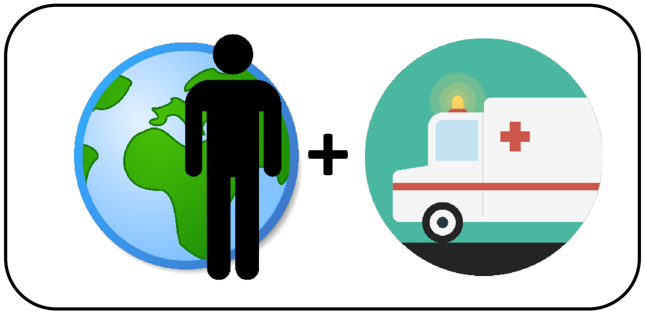


(what would have happened)

Counterfactuals

We never get to observe *what would have happened if we did something else*, so we have to estimate it

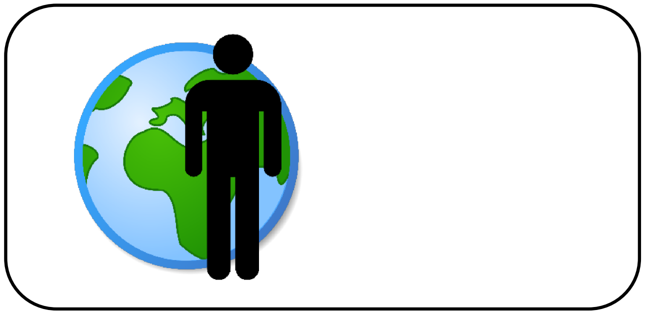
Reality



(what happened)

vs

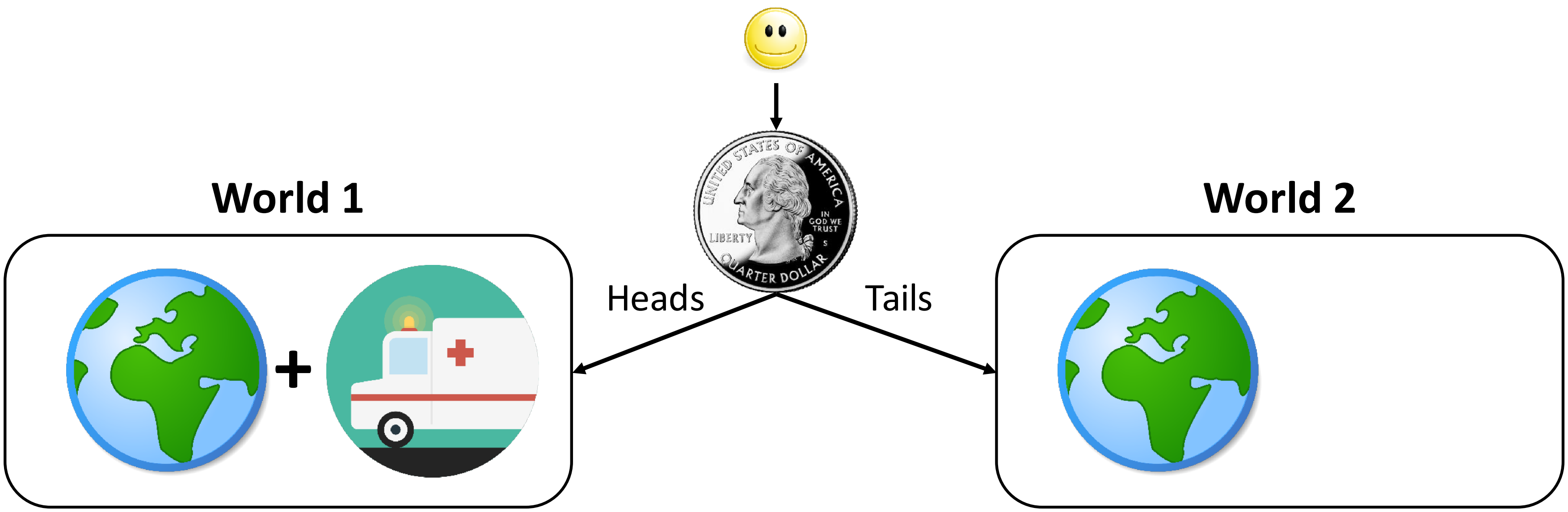
Counterfactual



(what would have happened)

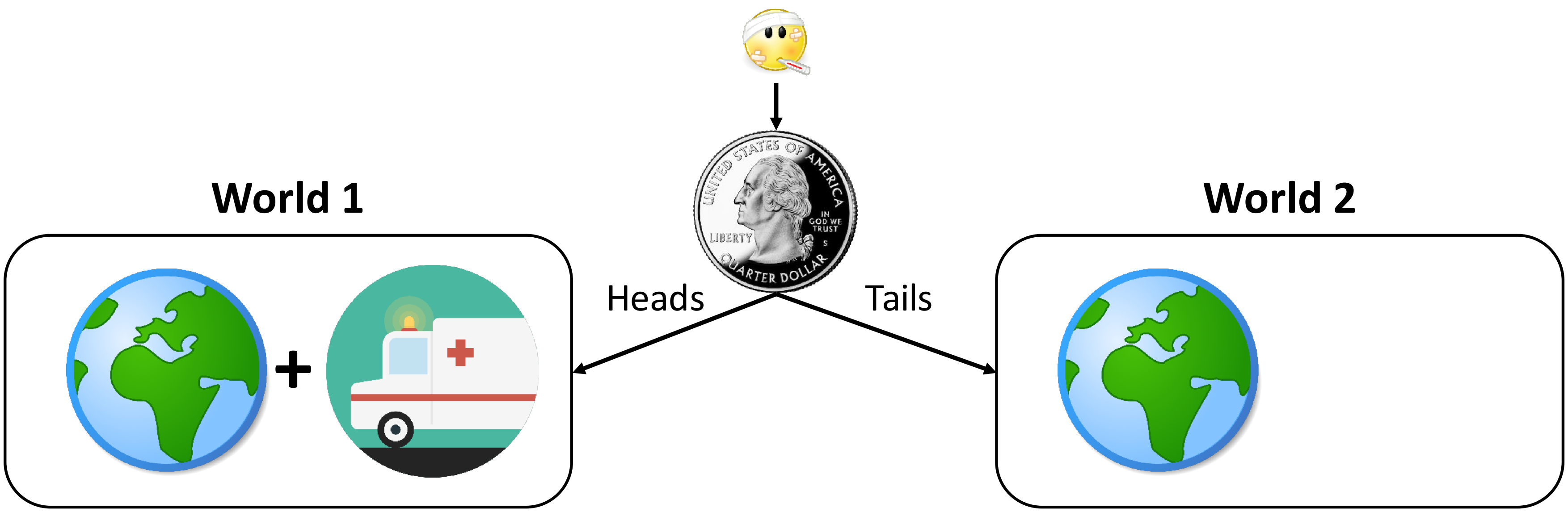
Random assignment

We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry



Random assignment

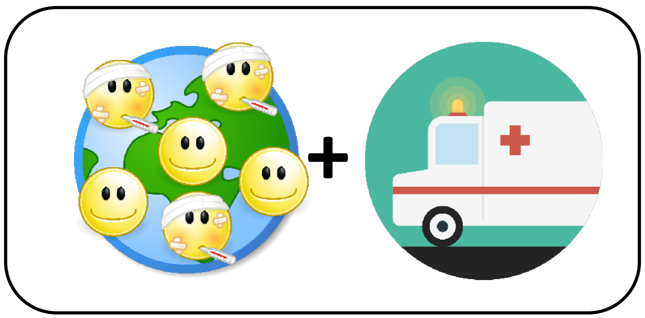
We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry



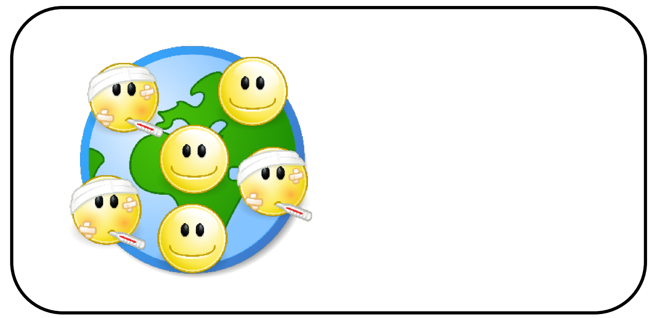
Random assignment

We can use randomization to create two groups that differ only in which treatment they receive, restoring symmetry

World 1



World 2



Basic identity of causal inference

The observed difference is now the causal effect:

Observed difference = Causal effect – ~~Selection bias~~

= Causal effect

Selection bias is zero, since there’s no difference, on average, between those who were hospitalized and those who weren’t

Caveats / limitations

Random assignment is the "gold standard" for causal inference, but it has some limitations:

- Randomization often isn't feasible and/or ethical
- Experiments are costly in terms of time and money
- Inevitably people deviate from their random assignments

Infeasibility: What researcher can’t do

Macro-political events: wars, political leaders, large policies (e.g. trade agreements, tax policies)

Natural features: earthquakes, geography

History

Ethics: Abusive studies on human subjects

Nazi human experimentation – 1940s

Tuskegee syphilis experiment – 1932-1972

Human radiation experiments – Cold War

Stanford Prison Experiment – 1971

Milgram experiment – 1963

Project MKUltra – Cold War

Ethics: Institution Review Boards (IRBs)

Most university / research organizations have an IRB

Independent of researchers

Usually requires Informed Consent (unless very low risk and notification would hamper research)

Risks to subjects are balanced by potential benefits to society

The selection of subjects presents a fair or just distribution of risks and benefits to eligible participants

Additional Protection for Vulnerable individuals: Prisoners, Children, Pregnant women (and fetuses).

Required for government funded research or for research published in most journals

Ethics: Continuing Concerns

Political Science studies effects on elections (foreign and domestic)

Facebook experimenting on the happiness of people via feed changes

Avoiding foreign IRBs due to corruption concerns

Many clinical trials are done in low-income countries due to costs.

Ethics: Is it ethical to randomize?

If we think something is better why don't we provide to all those in need?

Counter: It's good to randomize when,

- Over-subscription
- Staggered roll-out
- We often aren't sure it's better

Cost

Pharmaceutical drug clinical trials: “it’s millions, and many years, and most of them fail.”

Many of the randomized treatments are most interesting in the long-run:

- Early childhood interventions
- Chronic conditions

Contamination

Participants may drop out of an experiment. What the people doing badly preferentially leave?

Participants may switch treatments

People may behave differently if they know they are in an experiment

Experimental Design

What is the Treatment and what is the control? Is there a placebo?

At what level is this randomized (participant, cluster)?

What do I want to measure? Can I measure it effectively (e.g. privacy concerns)?

Blinding: Participants and experimental workers should not know whether in control or treatment?

Types of validity

Internal Validity

- Did I correctly estimate the effect in my sample?
- Did I not control for a confounder?

External Validity

- Will this effect hold elsewhere?
- Did I study a population that is very unique?

These answer did I learn anything potentially useful?

Closing thoughts

Experiments are called A/B testing in the business world and it is a new business process that is expanding very fast.

Closing thoughts

Large-scale *observational data* is useful for building *predictive models* of what the word has been like

But without appropriate *random variation*, it’s hard to *predict what happens when you change something* new in the world

Closing thoughts

Randomized experiments are like *custom-made datasets* to answer a specific question

Statistical Review

Significance level, α (e.g. 5%): Chance of getting Type I Error if Effect Doesn't Exist (first column)

Statistical Power: Chance of Detecting Effect when it does exist (second column)

	Effect Doesn't Exists	Effect Exist
Detect Effect	Type I Error (False Positive)	True Positive
Don't Detect Effect	True Negative	Type II Error (False Negative)

Cohen's d

http://rpsychologist.com/d3/cohend/

Appendix

Types of Estimates

Desire/Selection (D):

- *D=1: Went*=Went to Hospital
- *D=0: Stayed*=Stayed at Home

Treatment (W):

- *W=1: Treated*=Hospital Care
- *W=0: Untreated*=No Hospital Care

Health: H

Types of Estimates

Average Treatment Effect: Difference in Health between being treated and not.

$$ATE = E[H_{Treated} - H_{Untreated}]$$

Average Treatment Effect on the Treated: Difference in Health between being treated and not given they tried to get treatment.

$$ATT = E[H_{Treated} - H_{Untreated} | Went]$$

$$
\begin{gathered}
\Delta_{obs} = E[H|Went] - E[H|Stayed] \\
= E[H_{Treated}|Went] - E[H_{Untreated}|Stayed] \\
= E[H_{Treated}|Went] - E[H_{Untreated}|Went] + \\
E[H_{Untreated}|Went] - E[H_{Untreated}|Stayed] \\
= ATT + Selection
\end{gathered}
$$

Observational estimates

Δ_{obs} = (Health | Went) – (Health |Stayed)

Δ_{obs} = (Health if Treated | Went) – (Health if Untreated |Stayed)

Δ_{obs} = (Health if Treated | Went) – (Health if Untreated | Went)+

(Health if Untreated | Went) – (Health if Untreated |Stayed)

Δ_{obs} = ATET + Selection