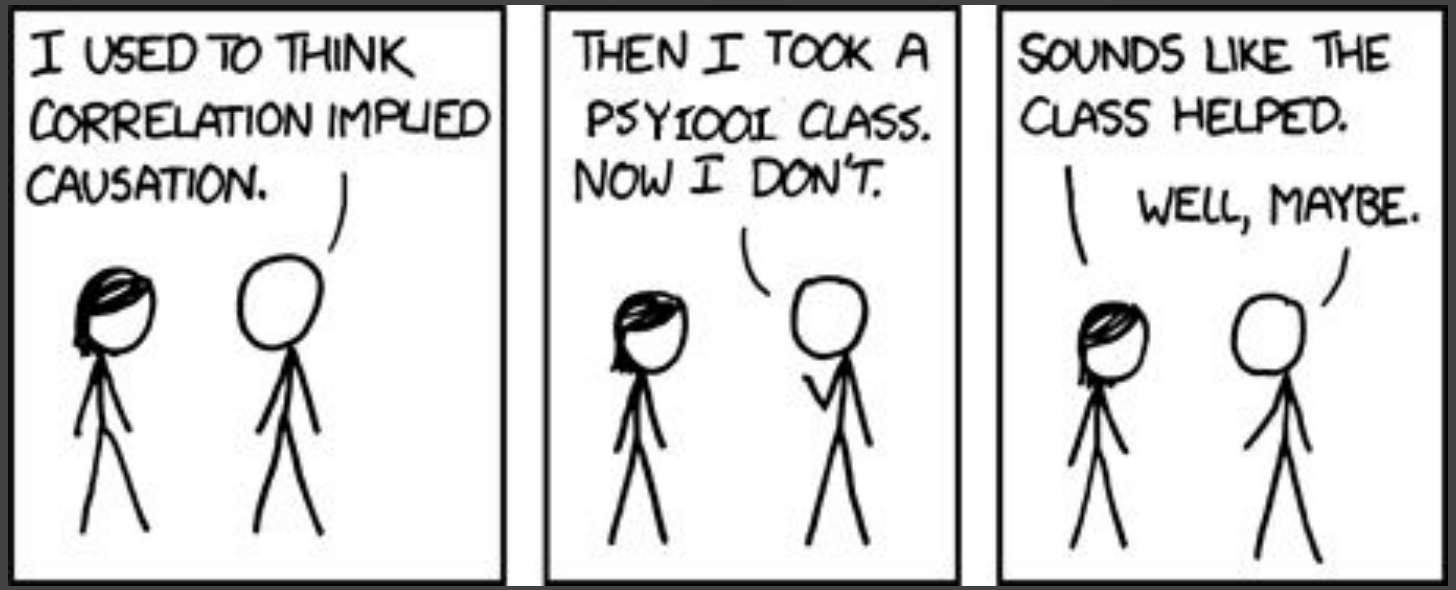
Causality

Brian Caffo, Jeff Leek, Roger Peng

@bcaffo

www.bcaffo.com

then what is causation? (XKCD/552)



How do we define causality?

“We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.”

http://en.wikipedia.org/wiki/David_Hume

For further discussion and references on the history of causality
http://plato.stanford.edu/entries/causation-counterfactual/

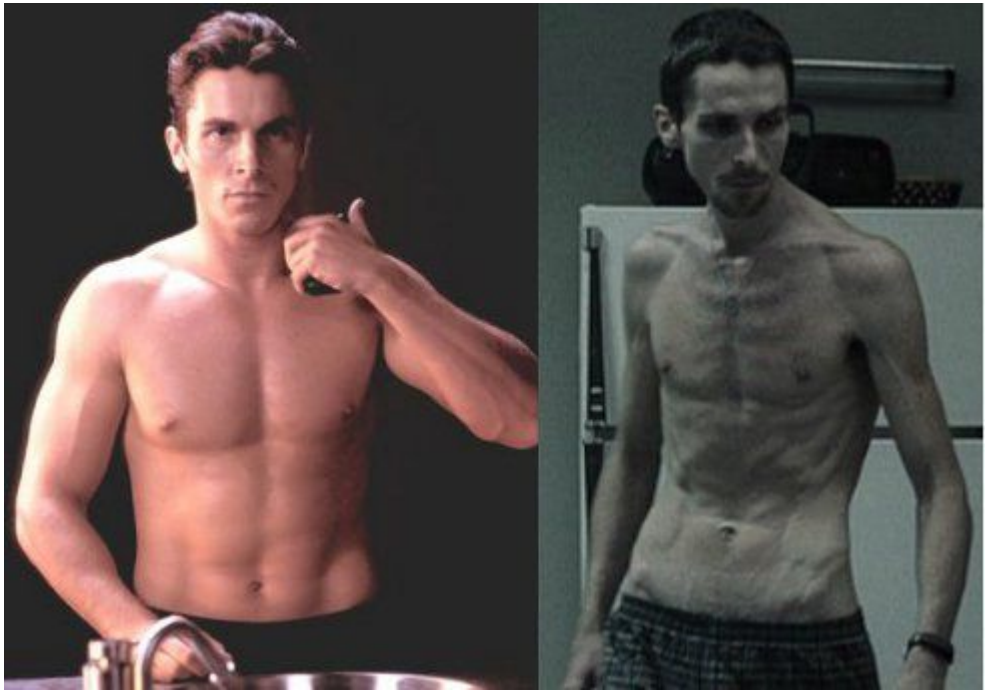
The causal effect of a treatment on a subject is the change in the outcome:

from the treatment the subject actually received and

what would of occurred had she or he received the other treatment

Twist 'N Tone
The exerciser that's fun, is easy and works. Helps you keep fit and helps take off inches. Burn up calories while you shape up waist, hips and legs. Provides same stimulating exercise as expensive equipment used in health clubs. Can be used by men and women.
1164-006-7A396—Wt. 3 lbs. $5.95
392
© London Media

Counterfactual



The average causal effect is the average of the subject-specific causal effects

This way of thinking about causal inference requires an assignable treatment or intervention

We can’t observe counterfactuals; we only get to observe one state of nature

We can, with assumptions and careful study design, make inferences about average causal effects

Causal thinking is essential for understanding the causal implications associated assumptions in our data analysis

Think about these study designs in the light of counterfactuals

Crossover trials: Give a subject a treatment, then after a suitable washout period, give the other

Consider a study of chronic migraines; give subjects one relief medication, washout period, then another

Why would this not work for an ad campaign, or a weight loss study?

How does it actualize counterfactual thinking?

http://bit.ly/1T39kyx

http://bit.ly/1cDE08b

http://bit.ly/1AUj1t9

Matching (aka finding doppelgangers)

http://bit.ly/1ARH7Fh

Randomization is our most effective tool for estimating average causal effects

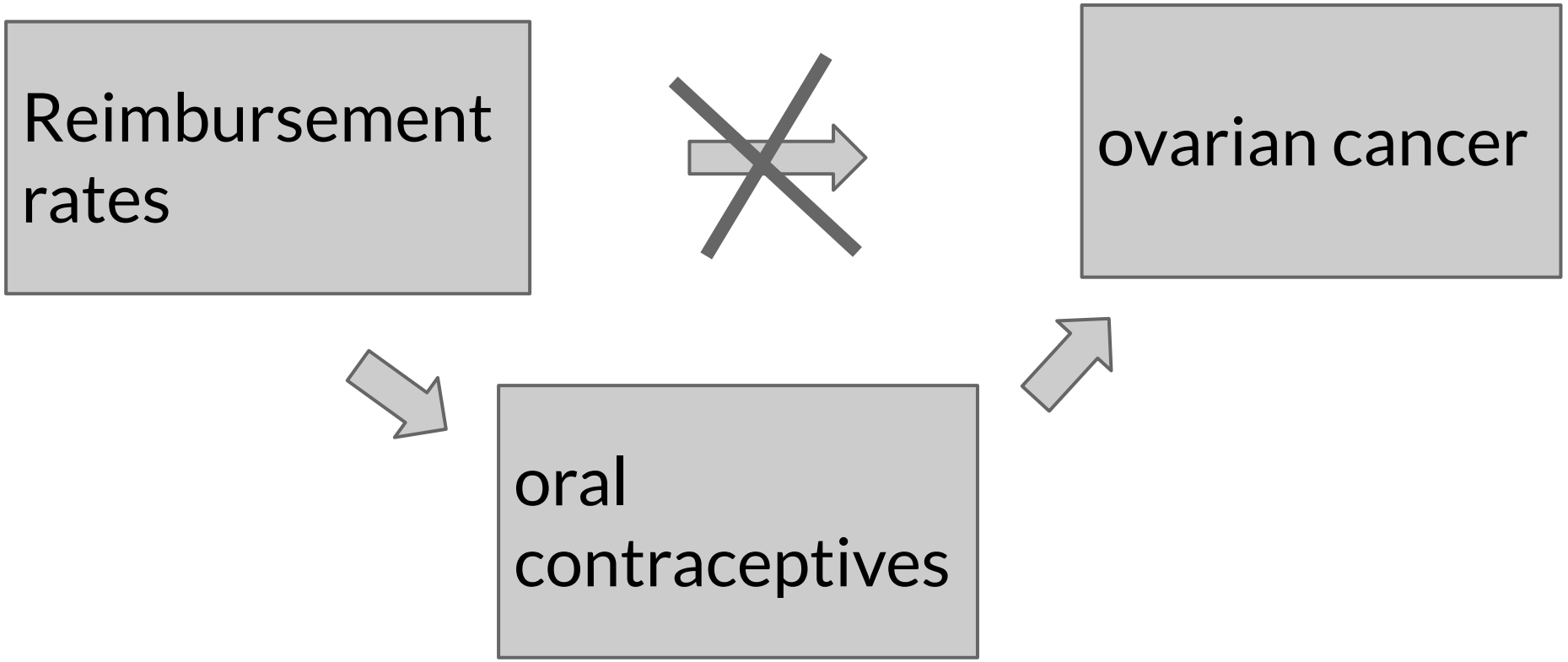
(we have an entire lecture on randomization)

Randomization, with high probability, makes the treated and untreated groups directly comparable

This can be shown to lead to causal inferences

http://skillsprojects.files.wordpress.com/2009/10/like_for_like.jpg

What else can we do?
Some examples


Reimbursement rates
ovarian cancer
oral contraceptives

Modeling comes with a lot of assumptions

Most well known modeling technique in causal inferences use propensity scores