

Best Refuted Causal Claims from Observational Studies

<http://bit.ly/refutedCausalClaims>

Ronny Kohavi, Distinguished Engineer, General Manager,
Analysis and Experimentation, Microsoft



Motivation

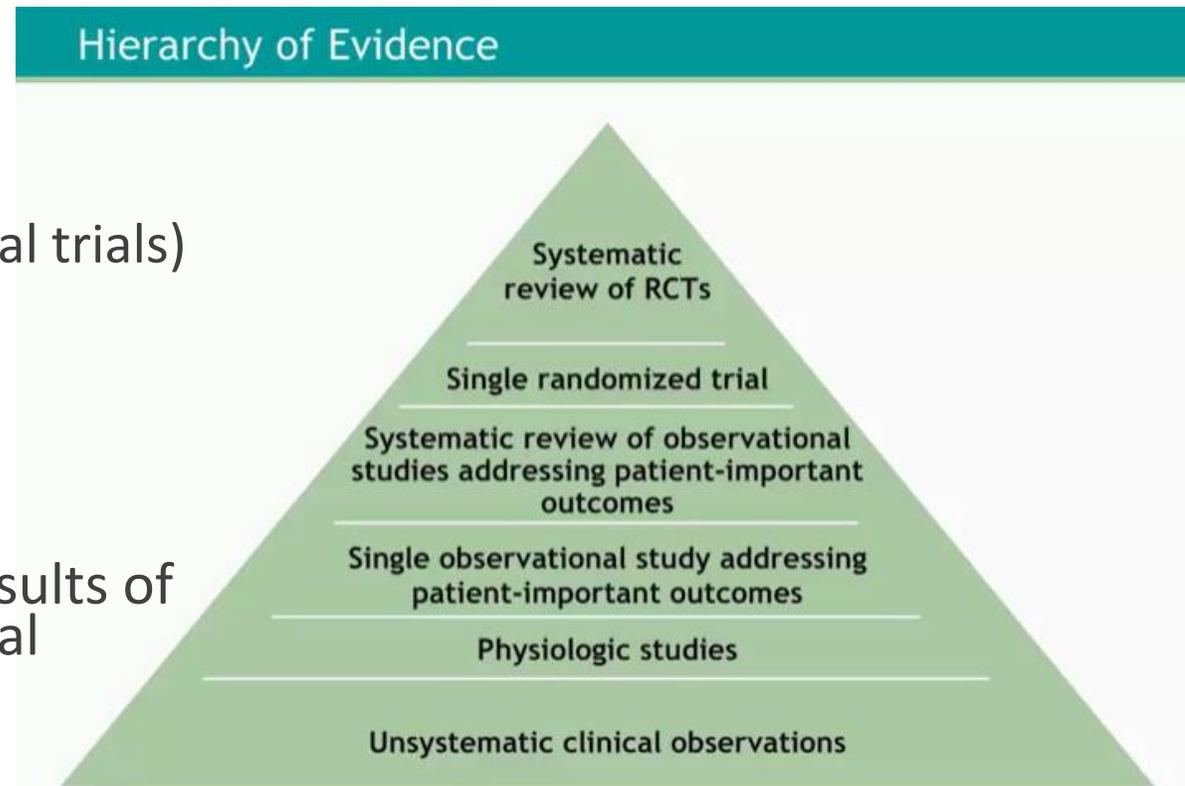
- I have seen too many reports where someone shows a correlation and claims causality
- Most recently, two observational studies in Microsoft Office, which were distributed widely, claimed that a new Office feature reduced attrition
- Almost any advanced feature will show such correlation because heavy users
 - Use advanced features
 - Have lower attritionSelecting users who use an advanced feature is similar to selecting heavy users

Claims of a causal relationship from such studies are simply bad science

- Users getting Office errors also have lower attrition, for the same reason (heavy users). Obviously we should not increase errors to reduce attrition
- Talking about correlations, causal sufficiency, selection-bias is fine for a statistic-savvy audience. Here I share some good examples of real mistakes in peer-reviewed journals.

Hierarchy of Evidence and Observational Studies

- All claims are not created equally
- Observational studies are UNcontrolled studies
- Be very skeptical about unsystematic studies or single observational studies
- The table on the right is from a Coursera lecture: [Are Randomized Clinical Trials Still the Gold Standard?](#)
- At the top are the most trustworthy
 - Controlled Experiments (e.g., RCT randomized clinical trials)
 - Even higher: multiple RCTs—replicated results
- The next few slides highlight common problems with observational studies
- Why show this?
Our users the experimentation platform trusted results of controlled experiments. When seeing observational studies, they did not realize the trustworthiness is MUCH lower



Warmup: Common Cause

- Example Observation (highly stat-sig)

Palm size correlates with your life expectancy

The larger your palm, the less you will live, on average

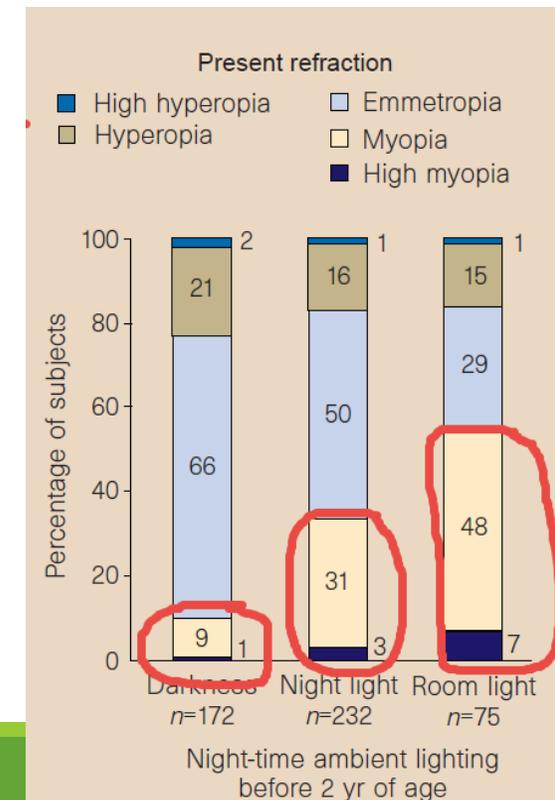
- Try it out - look at your neighbors and you'll see who is expected to live longer
- But...don't try to bandage your hands, as there is a **common cause**
- Women have smaller palms and live 6 years longer on average
- Obviously you wouldn't have believed that palm size is causal, but how about observational studies about features in products reducing churn?

Night Light Causes Myopia?

- May 1999: **CBS News Health Consultant Dr. Bernadine Healy** reports based on new study in the journal Nature that children who sleep with a night light ... until the age of two have a higher incident of nearsightedness - also known as myopia

Sleeping condition	% of children developing myopia
Darkness	10%
Night Light	34%
Lamp on	55%

- Dr. Graham Quinn, the study's lead author... urged parents to provide sleeping infants and toddlers with a dark bedroom -- within reason
- That last statement implies causality



Night Light Causes Myopia? Probably Not

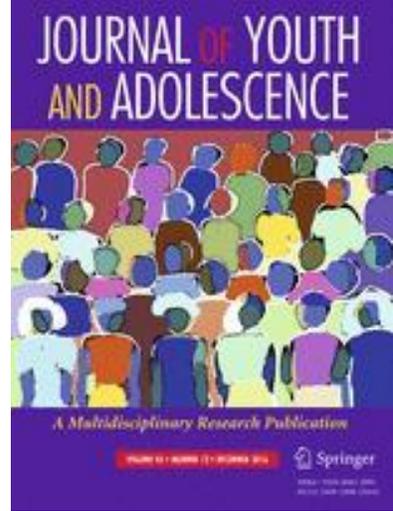
- Two studies published in Nature a year later failed to replicate the result and saw no such correlation
- Both made a crucial observation about a **common cause** they found:
 - Myopic parents are more likely to employ night-time lighting aids for their children
 - There is an association between myopia in parents and their children

Confounders

- Observational study in *The Lancet* showed that Vitamin C reduces Coronary Heart Disease (CHD)
- An RCT study appeared later in the same journal showing Vitamin C increases Coronary Heart Disease (CHD)
- Which one do we trust?
The controlled experiment
- Nice paper analyzed the reasons:
[Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?](#)
- The people who took Vitamin C are different on many attributes. The following were stat-sig differences at the $p < 0.0001$ level
 - Socioeconomic indicators: Social class, number bathrooms in house, shared bedroom, car access
 - Behavioral factors: current smoker, exercise, low fat diet, BMI >30 (obesity), alcohol consumption
 - Biomarkers: adult height
- If an observational study does not control for confounders, it is not trustworthy.
Problem: we may not know that we controlled for enough confounders

Twin Studies

- Observational study claimed
 - Youngsters who lose their virginity earlier than their peers are more likely to become juvenile delinquents
- Considered a well run study, which controlled for
 - Gender, Race, Receipt of public assistance, Parental education, Family structure, Previous substance use and depression, Importance of religion, School GPA, Relative pubertal status,. Virginity pledge status
- Paige Harden, a PhD student, used *the same database* and found 534 same-sex twins
 - Twins studies effectively control for many unknown factors
 - Her publication showing the OPPOSITE result, was considered superior and accepted to the same journal
- **Causal Sufficiency** is impossible to prove



[Summary from Washington Post 2007](#)

Hormone Replacement Therapy

- Large observational studies suggested a reduced risk of Coronary Heart Disease (CHD) among postmenopausal women (e.g., [pubmed 1996](#), [pubmed 2000](#))
- Randomized Control Trial showed the opposite
- Great [Coursera lecture](#) summarizes this fairly complex confounder
 - Time of usage of Hormone Replacement Therapy (HRT)
 - The risk of CHD is highest when you start HRT
- The problem with the observational study?
 - The women who died early on were less likely to get into the study

Systematic Studies of Observational Studies

- Jim Manzi in the book [Uncontrolled](#) summarized papers by Ioannidis showing that 90 percent of large randomized experiments produced results that stood up to replication, as compared to only 20 percent of nonrandomized studies
- [Young and Carr](#) looked at 52 claims made in medical observational studies, which were grouped into 12 claims of beneficial treatments (Vitamin E, beta-carotene, Low Fat, Vitamin D, Calcium, etc.)
- These were not random observational studies, but ones that had follow-on controlled experiments (RCTs)
- NONE (zero) of the claims replicated in RCTs, 5 claims were stat-sig in the opposite direction in the RCT
- Their summary
Any claim coming from an observational study is most likely to be wrong