Comparative Effectiveness Research Methods Training

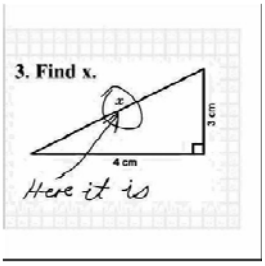
Module 1: Background

J. Michael Oakes, PhD
Associate Professor
Division of Epidemiology
University of Minnesota
oakes007@umn.edu

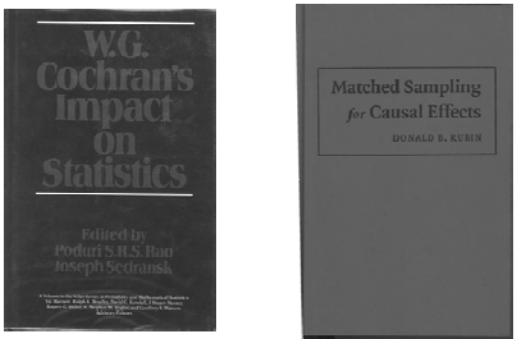
CCTS | Center for Clinical & Translational Science

A little about me.

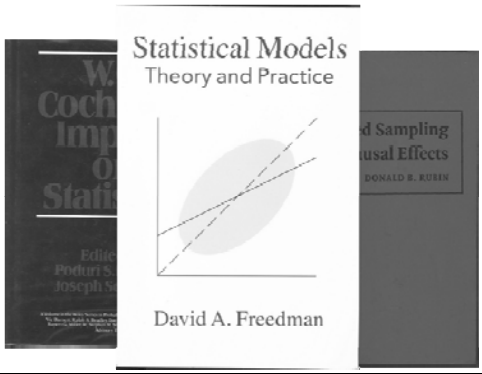
CCTS | Center for Clinical & Translational Science



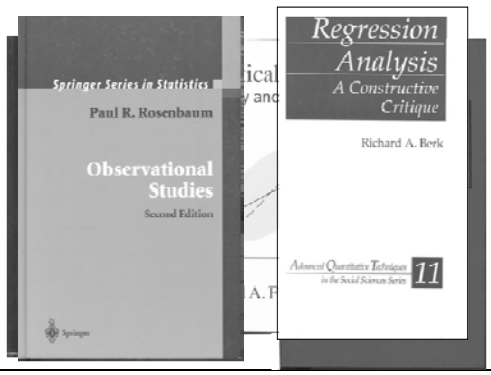
CCTS | Center for Clinical & Translational Science



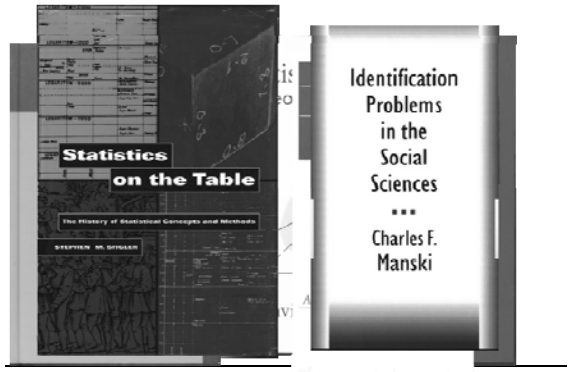
CCTS | Center for Clinical & Translational Science



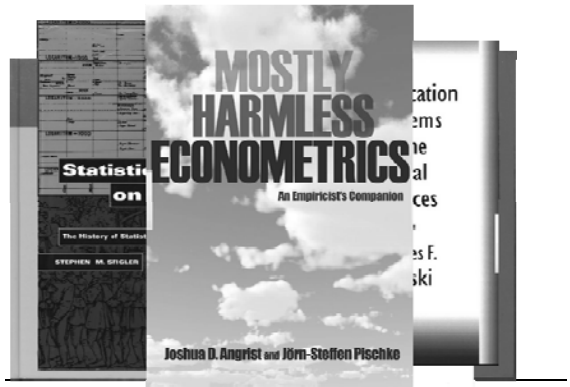
CCTS | Center for Clinical & Translational Science



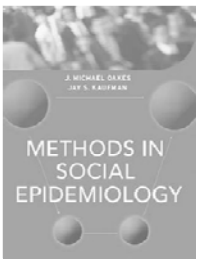
CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science

Module #1 Outline

1. Background Concepts
2. Causation and Counterfactuals
3. Effect Identification
4. Statistical Inference
5. Review
6. Questions

CCTS | Center for Clinical & Translational Science

On being asked to talk on the principles of research, my first thought was to arise… and say, “Be careful”, and to sit down.

Cornfield J. “Principles of research” Am J Ment Defic. 1959;64:240–52

CCTS | Center for Clinical & Translational Science

1. Background Concepts

Image from cartoonbank.com removed.

Image description: Picture of a distinguished speaker at a university behind podium in front of an audience with caption “I know so much that I don’t know where to begin.”

CCTS | Center for Clinical & Translational Science

Methodology

Methodological research is concerned with the logic of scientific inference.

The objective is to learn what conclusions can and cannot be drawn given specified combinations of assumptions and data.

(Manski)

CCTS | Center for Clinical & Translational Science

Science

Science (from the Latin *scientia*, meaning "knowledge") is an enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the world.

CCTS | Center for Clinical & Translational Science

Scientific Method

- Procedures for searching for truth.
- Transparency and replication are essential.
- Childlike discovery process. But scientific training aims to enhance the rate of success. Training is what makes science different from normal developmental processes.

CCTS | Center for Clinical & Translational Science

Science vs. Research

- Science is not about marshalling evidence to "prove" what you already believe or want others to believe.
- Science entails accepting unexpected if not undesirable answers.
- A scientist cannot be outcome oriented; she must be process and principle oriented.

CCTS | Center for Clinical & Translational Science

Induction & Deduction

- The idea of deduction is relatively simple.
- Deductive arguments attempt to show that a conclusion necessarily follows from a set of premises or hypotheses. A deductive argument is valid if the conclusion follows necessarily from the premises.

CCTS | Center for Clinical & Translational Science

A classic deductive argument is:

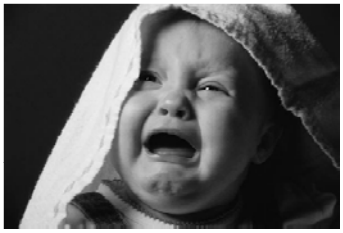
1. All men are mortal
2. Socrates is a man
3. Therefore, Socrates is mortal

CCTS | Center for Clinical & Translational Science

Induction

- Inductive inference is reasoning from the observed behavior of objects to their behavior when unobserved. It's a bottom-up empirical process.
- According to philosopher David Hume (1711-1776), we see a conjugation and name one thing a cause and another an effect. For Hume, cause is not "real"; induction can only yield correlations.
- Hume's contention calls into question all empirical claims made in everyday life or through the scientific method.

CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science

Falsification

Karl Popper (1902-1994) sought to solve the problem of induction. He argued that science does not rely on induction to generate knowledge. Instead of induction he argued that knowledge is created by conjecture and criticism. The main role of observations and experiments is marshal evidence to criticize and refute existing theories.

The process of refutation was called falsification.

You can think of this approach as a sort of Darwinian approach to scientific theories.

CCTS | Center for Clinical & Translational Science

2. Causation and Counterfactuals

CCTS | Center for Clinical & Translational Science

Statistically Speaking…

- For 100 years statistically minded researchers avoided the term "cause."
- Statisticians would reluctantly use the term in well-done randomized trials, but never in other situations.
- For those keeping up, this is no longer true. There's been a quiet revolution over the last 20+ years.

CCTS | Center for Clinical & Translational Science

Association, Correlation, and Causation

CCTS | Center for Clinical & Translational Science

Karl Pearson (1857-1936)

- Socialist Darwinist
- Student of Galton
- Correlation coefficient (derived in 1896)
- Analysis of contingency tables
- Law of large numbers ('in expectation')
- Not interested in causality: Correlation is greater than causation!

CCTS | Center for Clinical & Translational Science

Ronald A. Fisher (1890-1962)

- 1918 paper gave us ANOVA
- In 1919 rejected Pearson's offer for Galton's lab
- Joined Rothamsted (agri) experimental station
- Eugenics – as important as Darwin
- Built on work of Gosset (a beer brewer!)
- Developed randomized experiments
- Aimed to quantify uncertainty

CCTS | Center for Clinical & Translational Science

Austin B. Hill (1897-1991)

- Student of Greenwood and thus Pearson
- Pioneered randomized clinical Trial
- Randomization to mitigate bias, unlike Fisher
- With Richard Doll (1912-2005) showed smoking causes lung cancer in 1950 (developed case-control design)

CCTS | Center for Clinical & Translational Science

Pearson vs. Fisher/Hill

- Pearson: Estimate correlations and maximize R^2
- Fisher & Hill: Estimate the effects of causes.

CCTS | Center for Clinical & Translational Science

Counterfactuals

- **Idea comes from philosopher David Hume (1711-1776)**

  "…but for…"

- **Advanced by philosopher David Lewis in 1973**

  "If kangaroos had no tails, they would topple over…"

- **Statistician Don Rubin advanced ideas in statistics and epidemiology**

  Potential outcomes model.

- **Recent work takes "closet possible world" assumption seriously**

  Just what are limits to comparative inference?

CCTS | Center for Clinical & Translational Science

Ideal is to compare <u>same</u> unit under two scenarios; difference in outcomes is causal effect attributable to the changed scenario.

Let's compare Ivan (and only Ivan) in environment <u>with and without</u> McDonalds, *with all else exactly the same*. The difference in Ivan's BMI is effect of McDonald's on Ivan.

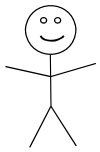
Trouble is, we cannot observe Ivan under <u>both</u> scenarios. Ivan either lives in an environment with McDonalds or he doesn't.

The scenario which Ivan does not actually live in is counter to fact, or the <u>counterfactual</u>. Rubin calls same the "potential outcome for Ivan."

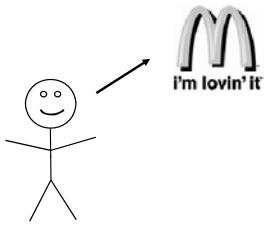
In order to have actual data, we must <u>find a substitute</u> for Ivan's unobservable counterfactual state. Finding a credible counterfactual substitute is the crux of all sound causal inference.

In fact, <u>randomization</u> to condition in this context is nothing more than technique to generate credible counterfactual substitutes.

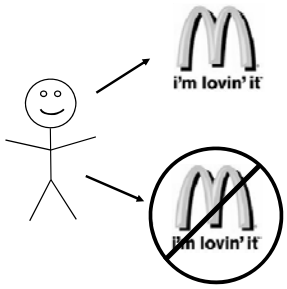
CCTS | Center for Clinical & Translational Science



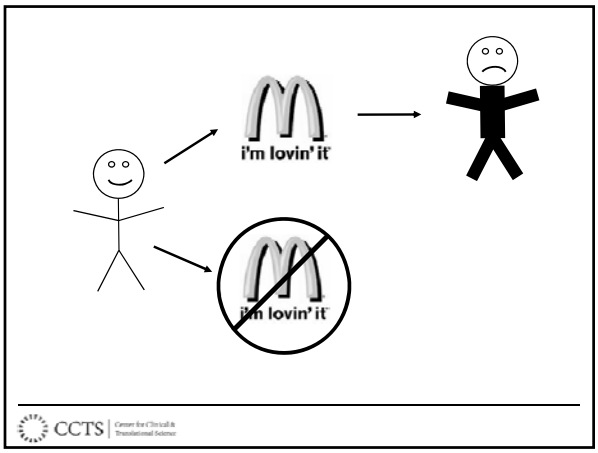
CCTS | Center for Clinical & Translational Science

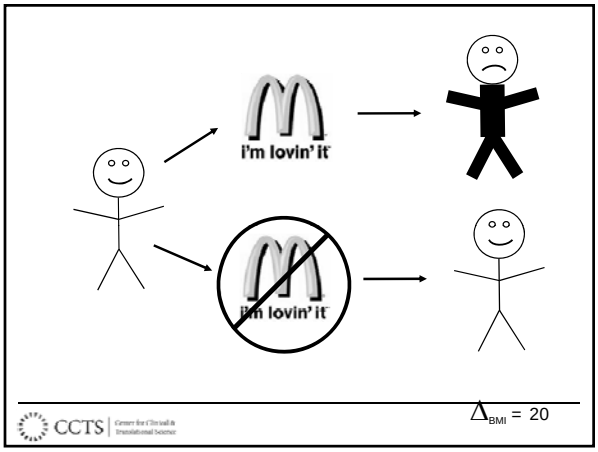


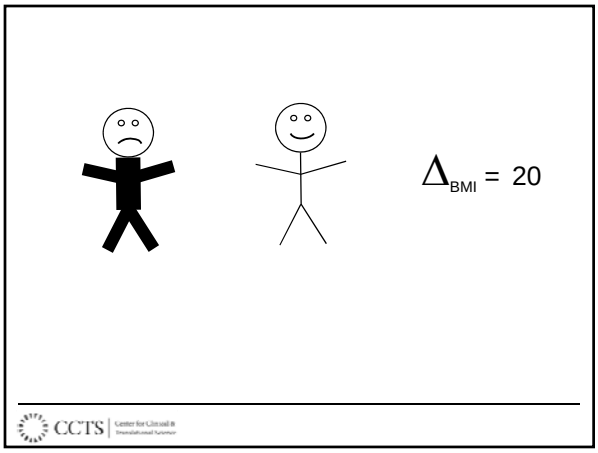
CCTS | Center for Clinical & Translational Science

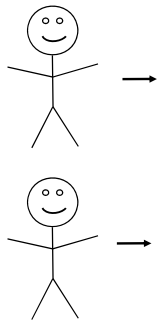


CCTS | Center for Clinical & Translational Science


i'm lovin' it
i'm lovin' it
CCTS

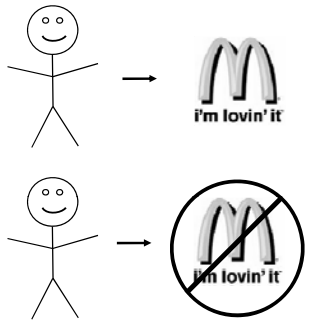

i'm lovin' it
i'm lovin' it
CCTS
Δ_{BMI} = 20


Δ_{BMI} = 20
CCTS

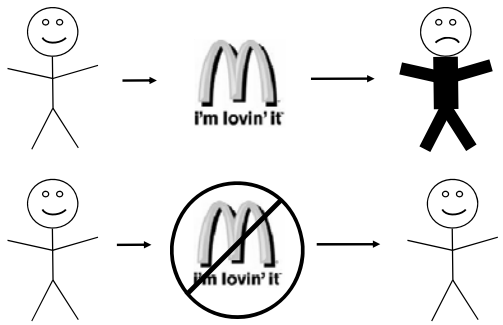


Since we cannot observe **BLACK** under both scenarios, we substitute **BLUE** for the unexposed scenario.

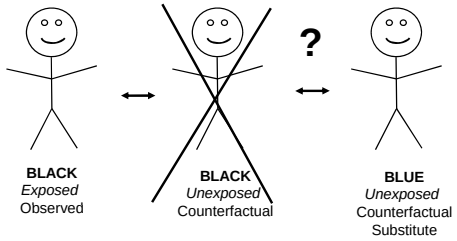
CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science



If **BLUE** is a perfect substitute for **BLACK** *but for the exposure (ie, unexposed BLACK)*, then causal inference is credible.

To extent **BLUE** is NOT a perfect substitute for **BLACK** *but for the exposure*, we have bias or confounding. Background differences complicate inference.

CCTS | Center for Clinical & Translational Science

Simply put, causal inference is all about finding the best counterfactual substitute for the unobservable counterfactual scenario.

The best ones are said to be *exchangeable*.

Randomization is a good mechanism to produce a group of subjects that are exchangeable.

Simply, it's all about the comparison group!

CCTS | Center for Clinical & Translational Science

- Confounding is lack of exchangeability or an imbalance of background characteristics between groups that effects outcomes
- Conditional exchangeability (eg, within strata)
- Presumes groups/subject *could be* exposed
  - Positivity assumption
  - Structural confounding

CCTS | Center for Clinical & Translational Science

Potential Outcomes

Condition Assigned	Outcome if Treated, Y^1	Outcome if not Treated, Y^0
Treatment	Observed	Unobservable Counterfactual
Control	Unobservable Counterfactual	Observed

CCTS | Center for Clinical & Translational Science

More formally, T has a causal effect on Y for person i if

$$Y_{i\ T=0} \neq Y_{i\ T=1}$$

But we can only observe one of these outcome for any i

At the population level, we use probabilities and assuming exchangeability,

$$\text{Prob}[Y_{T=0} = 1] \neq \text{Prob}[Y_{T=1} = 1]$$

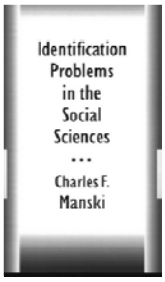
CCTS | Center for Clinical & Translational Science

Limitations of Counterfactuals

- Tends to focus our attention on single/sole causes
- Manipulation only? (is being tall or female causal?)
- Level of dose?
- Mechanism/process of cause?
- Closest possible world assumption?
- Necessary and sufficient conditions/causes?
- Interactions and effect heterogeneity (some are doomed/immune)?
- Switch from individual to population-level?
- Dynamics & feedback loops
- Necessary for "explanation"?

CCTS | Center for Clinical & Translational Science

3. Effect Identification



CCTS | Center for Clinical & Translational Science

An effect is identifiable if it is theoretically possible to learn the true value of the parameter when the sample size approaches infinity.

Imagining that your sample size is infinitely large eliminates all problems of statistical uncertainty, confidence intervals and p-values. The procedure lets you concentrate on competing explanations in the parameter space.

CCTS | Center for Clinical & Translational Science

Identification is about ruling out competing hypotheses or explanations.

Apart from statistical imprecision, if more than one explanation for your effect (eg, difference b/w treatment and control) exists, then you have an identification problem. You cannot say X_1 caused Y because the cause might have been X_2, or X_3, or X_n .

CCTS | Center for Clinical & Translational Science

Akin to a Physician's Differential Diagnosis

Differential diagnosis involves first making a list of possible diagnoses, then attempting to eliminate each possible diagnosis from the list by making observations and using tests that should yield different results, depending on which diagnosis is correct.

CCTS | Center for Clinical & Translational Science

Confirmation Bias

- Conducting research, perhaps subconsciously, in order to support your current beliefs.
- Bias may arise in the way the particular question is framed, in data collection, in analysis, or in interpretation.

CCTS | Center for Clinical & Translational Science

English philosopher and scientist Francis Bacon (1561–1626) wrote in *Novum Organum*

> The human understanding when it has once adopted an opinion... draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects or despises, or else by some distinction sets aside or rejects[.]

CCTS | Center for Clinical & Translational Science

"The latter half of this [20th] century has seen an erosion in the perceived legitimacy of science as an impartial means of finding truth. Many research topics are the subject of highly politicized dispute; indeed, the objectivity of [so much of science] has been called into question."

MacCoun, R.J., 1998. Biases in the interpretation and use of research results. *Annu Rev Psychol* 49, 259-287.

CCTS | Center for Clinical & Translational Science

"Good science is more than the mechanics of research and experimentation. Good science requires that scientists look inward – to contemplate the origin of their thoughts. The failures of science do not begin with flawed evidence or fumbled statistics; they begin with personal self-deception and an unjustified sense of knowing."

Burton, R.A., 2008. *On Being Certain: Believing you are Right even when you're Not*. St. Martin's Press, New York, p 167.

CCTS | Center for Clinical & Translational Science

Trofim D. Lysenko (1898-1976)

http://en.wikipedia.org/wiki/File:Lysenko_with_Stalin.gif

CCTS | Center for Clinical & Translational Science

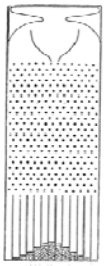
" … how can you possibly award prizes when everyone missed the target?" said Alice.

"Well" said the Queen, "Some missed by more than others, and we have a fine normal distribution of misses, so we can forget about the target."

Kennedy, P. 1988. *A Guide to Econometrics.* Cambridge, MA: MIT Press; p.292

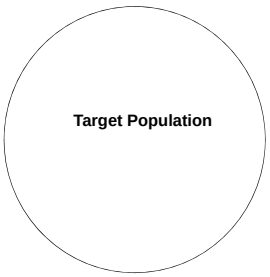
CCTS | Center for Clinical & Translational Science

4. Statistical Inference



CCTS | Center for Clinical & Translational Science

Target Population & Inference about Parameters



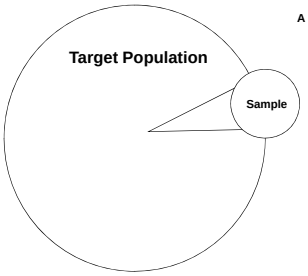
We want to know value of some <u>parameter</u>, $\overline{Y}$.

CCTS | Center for Clinical & Translational Science

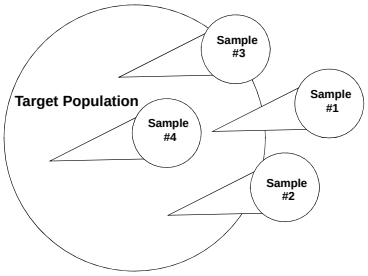
Sample == Statistic
Population == Parameter

A statistic is an estimate of a parameter

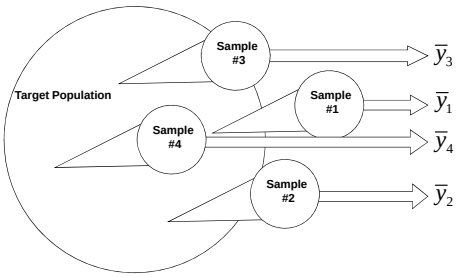
Bias and Precision are the key issues



CCTS | Center for Clinical & Translational Science

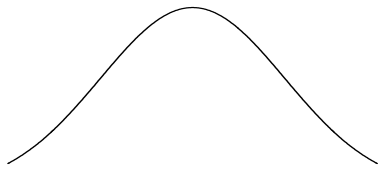


CCTS | Center for Clinical & Translational Science



CCTS | Center for Clinical & Translational Science

Sampling Distribution of Means



mean of $\bar{y}_{i=1\text{-}S} \Rightarrow \bar{Y}$

CCTS | Center for Clinical & Translational Science

P-Values

- Imagine you collected some data and estimate an effect, perhaps the difference between treatment and control group subjects.
- Given sampling variability, how do you know the effect is meaningful? That is, how do you know observed effect is not due to chance alone?

CCTS | Center for Clinical & Translational Science

Test Statistic

$$t = \frac{\Delta}{SE_{\Delta}}$$

⇦ Signal

⇦ Noise

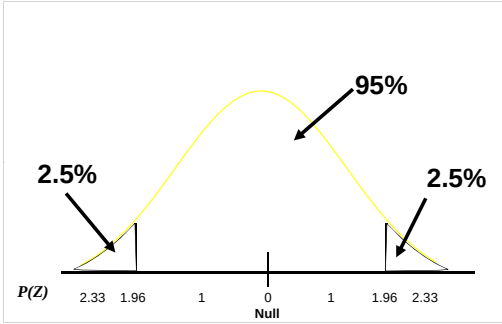
Compare this t to some known distribution and a pre-defined cut-point to determine "statistical significance".

CCTS | Center for Clinical & Translational Science

- Typically, we calculate a p-value associated with a statistical test. We then see if the calculated p-value is less than some threshold, often 0.05.
- If it is, we reject the null hypothesis of "no effect" and claim that the effect is statistically significant, which seems to mean that the estimated effect is (probably) not due to chance alone.
- OK, but what does this *really* mean?

CCTS | Center for Clinical & Translational Science

Statistical Significance



CCTS | Center for Clinical & Translational Science

P-values Today

- A mixed blend of theory from Fisher and Neyman-Pearson
- Lots of confusion and ignorance
- Many interpretive errors
- Increasing calls for improvement (eg, Bayesian approaches)

CCTS | Center for Clinical & Translational Science

- A p-value of 0.04 tells us that, if the null were true, an effect/association at least a large as the observed would occur 4 out of 100 times (4%).
- p-value represents the strength of evidence against the null hypothesis.
- p-value is *not the probability* that a given (alternative) hypothesis is true.

CCTS | Center for Clinical & Translational Science

- We very much want to know the probability that our hypothesis is true. But p-values cannot tell us this. We must use Bayesian methods to answer such questions.

Image from unknown author removed.

Image description: Picture with title "Yet Another History of Life as We Know It" showing evolution of man from apes.

Image shows 5 stages:
1. small ape with Homo Apriorius label and caption representing *ideas*
2. larger ape with Homo Pragmaticus label and caption representing *data*
3. cave man holding blade with label Homo Frequentistus and caption representing *have some data given an idea*
4. caveman fully standing with label Homo Sapiens and caption representing *have data and ideas*
5. crouching man behind computer with Homo Bayesianis label and caption representin *the idea or effect given the data*

Picture originally taken from Prof Mike West's website, Duke U.

CCTS | Center for Clinical & Translational Science

- For over 50 years every major methodologist, it seems, has rejected the use of p-values and significance testing.
- Statistical significance says nothing about importance, utility or meaningfulness; effects with small p-values may be meaningless and imply precise but negligible effects.
- You cannot compare p-values across studies or even across variables in a given study and claim one "effect" is more statistically significant than the other. In other words, p-values do not reflect the strength of a relationship.

CCTS | Center for Clinical & Translational Science

"The question is not whether the *P*-value is intrinsically bad, but whether it too easily substitutes for the thoughtful integration of evidence and reasoning. Given the *P*-value's blighted history, researchers who would employ the *P*-value take on a particularly heavy burden to do so wisely."

Editors. 2001. *Epidemiology*: Volume 12(3) p 286

CCTS | Center for Clinical & Translational Science

Confidence Intervals

- They reflect the precision or reliability of our estimates.
- To interpret 95% CIs as having probability of 0.95 of including the parameter of interest is a mistake; technically, this mixes Bayesian interpretations with Frequentist interpretations.
- We can say that if we conduct a large number of studies and calculate 95% confidence intervals for some parameter estimate in each, the true parameter of interest will be in 95% of the confidence intervals calculated.

CCTS | Center for Clinical & Translational Science

- Estimates that are the least influenced by chance are not necessarily those with small p-values but rather those with narrow confidence intervals.
  - An estimate with a wide confidence interval is imprecise and unstable no matter how low its p-value is.
- Confidence intervals give us everything a p-value significance test does, and more! Confidence intervals not only tell us whether or not we should to reject the null, but gives a range of expected values for our effect in a meaningful and understandable metric.

CCTS | Center for Clinical & Translational Science

Decisions about hypotheses

Researcher's Inference	Mother Nature or True State of Null Hypothesis: H_0 is True	H_0 is False
Reject H_0	***Type I error*** *probability* = α	Correct Inference *probability* = $1-\beta$ *Power* (H_A)
Accept H_0	Correct Inference *probability* = $1-\alpha$	***Type II error*** *probability* = β

CCTS | Center for Clinical & Translational Science

Statistical Power

$$\text{Power} = 1-\beta$$

$$Z_{1-\alpha/2} + Z_{Power} = \frac{\Delta}{SE(\Delta)}$$

CCTS | Center for Clinical & Translational Science

How to increase statistical power?

Improve power by increasing signal and/or decreasing noise

$$t = \frac{\Delta}{SE_{\Delta}}$$

> Half the "game" is to decrease SE
- **Better measurement**
- **More subjects**

> Other half of the game is to increase delta…
- **Have stronger effect**

CCTS | Center for Clinical & Translational Science

Ultimately…

CER is not difficult because of "statistics"

CCTS | Center for Clinical & Translational Science

5. Review

I. Background Concepts
   a. Scientific Method
   b. Methodology
   c. Falsification

II. Causation and Counterfactuals
   a. Pearson vs. Fisher/Hill
   b. Causal effects are the goal
   c. Seek counterfactual substitutes

III. Effect Identification
   a. Not about sample size
   b. Rule out competing explanations
   c. Confirmation bias

IV. Statistical Inference
   a. Statistics estimate parameters
   b. P-values and Statistical power
   c. Not the key problem for CER

CCTS | Center for Clinical & Translational Science

6. Questions

1. What is the difference between correlation and causation?
2. What does effect identification have to do with competing explanations for an observed phenomena?
3. How does confirmation bias among researchers undermine the validity of research?
4. Names two problems associated with drawing scientific inferences based on p-values.
5. What is the difference between science and research?

CCTS | Center for Clinical & Translational Science

References & Resources

1. Burton RA. On Being Certain: Believing you are right even when you're Not. New York, NY: St. Martin's Press; 2008:167.
2. Cornfield J. Principles of research. American Journal of Mental Deficiency. 1959; 64: (24):0–52.
3. Manski CF. Identification Problems in the Social Sciences. Cambridge, MA: Harvard University Press; 1995.
4. Editors. Epidemiology. 2001; 12: (3):286.
5. Kennedy P. A Guide to Econometrics. Cambridge, MA: MIT Press; 1988:292.
6. MacCoun RJ. Biases in the interpretation and use of research results. Annual Review of Psychology. 1998; 49:259-287.
7. Slide 54 uses an image from Soyfer V. The Consequences of political dictatorship for Russian Science. *Nature Reviews Genetics*. 2001;2:(9):723-9 (Wikimedia Commons)

CCTS | Center for Clinical & Translational Science