Day 01

Scientific method, Critical evaluation, & Cognitive biases

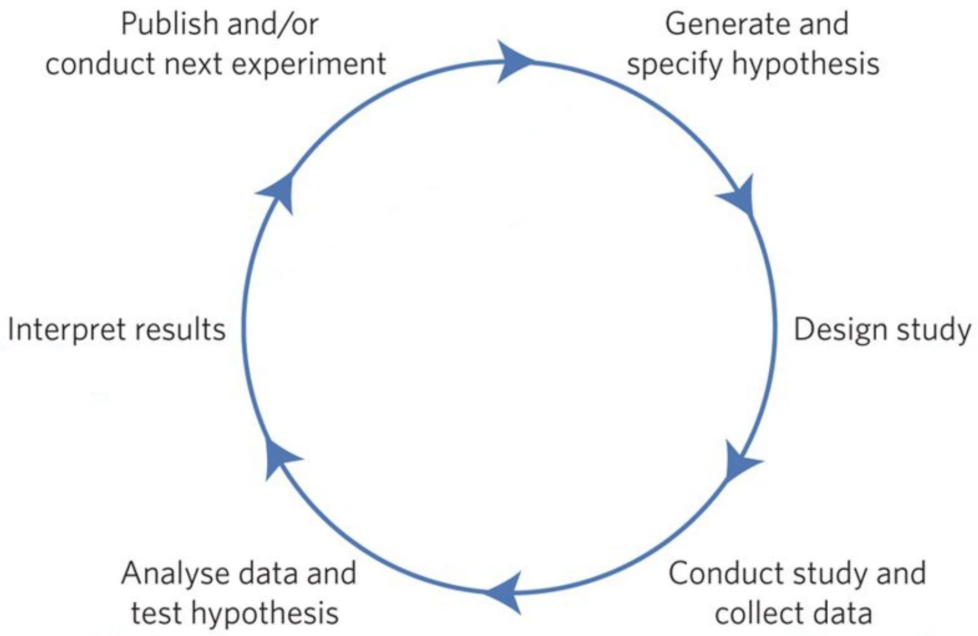
- Scientific method and hypothesis formation
- Recognizing when study designs fail
- Critical evaluation of research claims
- Cognitive biases in data interpretation
- Statistical intuition vs. blind rule-following
- What's due next week?

Scientific method & hypothesis formation

The hypothetico -deductive cycle

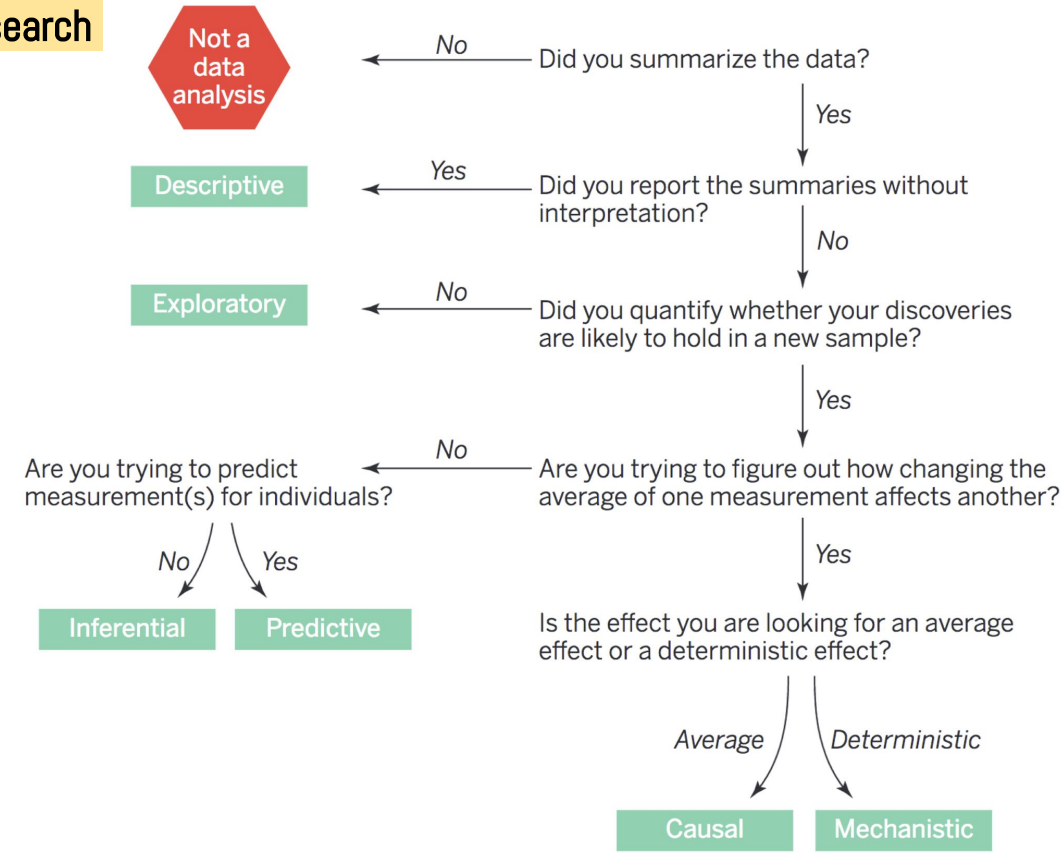
Issues:

- Relying on erroneous reports to help formulate hypotheses
- Failing to express qualitative hypotheses in quantitative form
- Using the same set of data both to formulate hypotheses & to test them



Scientific method & hypothesis formation

Confirmatory vs. e xploratory research



JT Leek, RD Peng What is the question?

Scientific method & hypothesis formation

Confirmatory research tests pre-defined hypotheses

- A **well-formulated hypothesis** is both **quantifiable** & **testable**
  - Involve measurable quantities or refer to items that may be assigned to *mutually exclusive* categories.
- It will **specify the population** to which the hypothesis will apply.

Which one's a valid hypothesis? Why?

1. For males over 40 suffering from chronic hypertension, a 100 mg daily dose of this new drug will lower diastolic blood pressure an average of 10 mm Hg.

2. All rock fans are passionate.

Exploratory research is for hypothesis generation using flexible data analysis

- Can **suggest patterns** but cannot confirm them.
- **Totally fine** ***if*** they are:
  - Reported and interpreted as such in a transparent manner
  - Especially so if they serve as the basis for a replication with pre-specified analyses.

The problems:

1. **Data dredging** : Mass-testing data to find any association → found association is stated as fact without further validation.

2. **HARKing** (Hypothesizing After Results are Known): Presenting post-hoc findings as if they were predicted from the start (creating a story around those random findings).

Scientific method & hypothesis formation

Pre-registration

State the objectives, population, and analytical plan *before* seeing the data.

- Ensures transparency
- Makes the distinction between confirmatory (hypothesis testing) and exploratory (hypothesis generating) research clearer
- Increases the credibility of findings by preventing researchers from "improving" p-values through ad-hoc data & model adjustments

Also helps in recognizing:

- That the phenomena you are investigating may have stochastic/chaotic components
- The list all possible sources of variation→Informs study design & data/metadata collection
- All alternative hypotheses→List possible experimental findings along with the conclusions you would draw & the actions you would take if this or another result should prove to be the case.

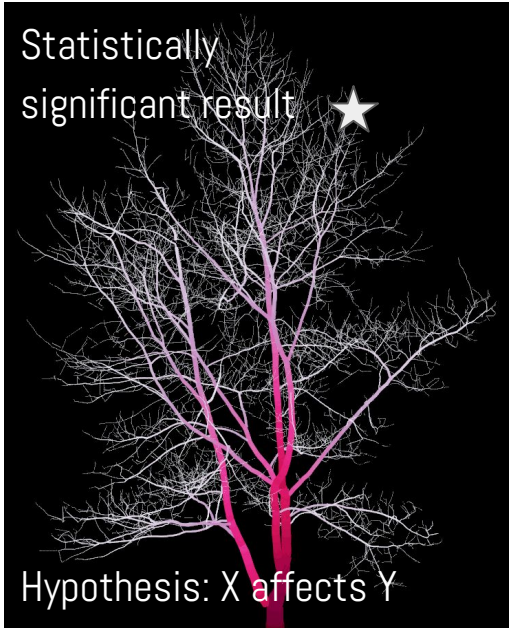
Scientific method & hypothesis formation

The garden of forking paths

A study has a **single dataset** but **many possible analytical paths** .

Each decision (which outliers to exclude, transformation to use, outcome parameters to keep, covariates to add) increases the chance of a false positive, even when researchers have good intentions.

Work in pairs: bit.ly/**phack-away**



A Gelman & E Loken 2013

Recognizing when study designs fail

When RCTs are misinterpreted

Randomized Controlled Trials (RCTs) are the "gold standard" because they ensure confounding variables are distributed evenly between groups.

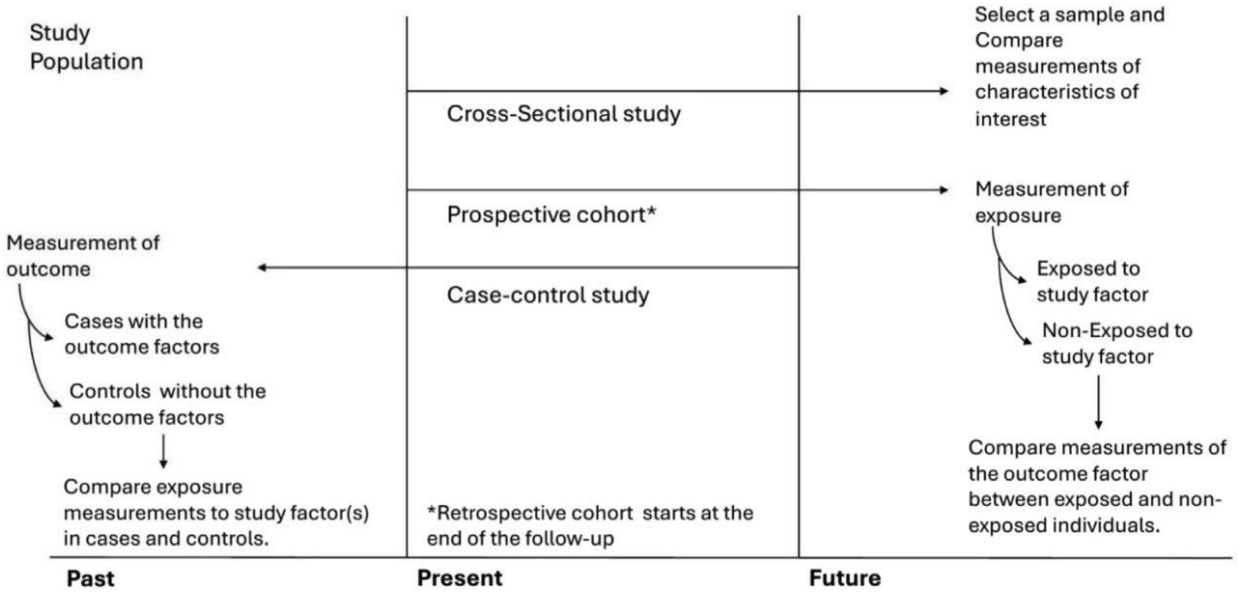
Issues:

- **Generalizability problems** : If an RCT excludes certain demographics (e.g., excluding women or only including healthy subjects), its findings cannot be extrapolated to the general population.
- **Attrition and dropout** : If participants drop out of the intervention group due to side effects, the remaining "successes" do not represent the true impact of the treatment.
- **Contamination** : If subjects in the control group inadvertently receive the treatment or change their behavior, the comparison is invalidated.

Recognizing when study designs fail

Observational studies & claims of causation

Unlike **randomized controlled trials** (RCTs), **observational studies** lack an assignment mechanism (the researcher does not control the independent variable because of ethical concerns or logistical constraints) and can usually only demonstrate associations, not causation.



Recognizing when study designs fail

Observational studies & claims of causation

Unlike **randomized controlled trials** (RCTs), **observational studies** lack an assignment mechanism (the researcher does not control the independent variable because of ethical concerns or logistical constraints) and can usually only demonstrate associations, not causation.

- **Cohort studies** : Temporal ordering (A happened before B) does not prove that A caused B.
- **Case-control studies** : Prone to recall bias (imprecise reporting of past events) and overmatching, where highly correlated predictors mask the true effect.
- **Cross-sectional studies** : They are descriptive "photographs" and cannot be used to mimic the causal inference of an RCT.

Recognizing when study designs fail

Missing, inadequate, or inappropriate controls

- **Historical controls** : These fail because the control environment (e.g., hospital standards in 2001) may differ drastically from the current intervention environment.
- **Placebo effects** : Without proper blinding, expectations of recovery can create a "mind over body" effect that mimics a treatment result.
- **No control group** : This surprisingly common failure makes it impossible to separate a treatment's effect from the passage of time or other external factors.

Recognizing when study designs fail

The fundamental question that you must constantly ask : What can this specific design actually prove?

Statistical software will not warn you if your design is inherently incapable of answering your research question.

Critical evaluation of research claims

Reading *Methods* critically (not just *Results*)

1. Use **Title & Abstract** for only selecting paper.

a. Don't be swayed by high-profile papers, media hype, or current dogma.

2. Read the **Introduction** :

a. Identify *the* question. What is the big challenge the authors are trying to solve?

b. What are the *specific* questions this paper is going to answered?

Critical evaluation of research claims

Reading *Methods* critically (not just *Results*)

3. Read **Data & Methods** : [Be critical!]

a. *Methods* should read as if the study has not been completed yet.

   i. Discussion of achieved sample sizes & p-values, missing observations, and consistency across item measures belong in *Results*.

b. For each specific Q, note data (type & source) & method (algorithms/techniques, software, & approach).

c. Are the data & methods describes sufficient to answer the Qs raised in the Intro?

d. Make detailed notes on: 1) what's unclear, 2) what you might do differently.

4. ALWAYS read the **Supplementary Materials**

Almost all the good stuff is in here!

Critical evaluation of research claims

Reading *Methods* critically (not just *Results*)

5. Read the **Results** : [Be critical!]

a. Go figure-by-figure, panel-by-panel. Based on your reading of *Data & Methods*, is there enough information to know/reproduce that analysis?

b. Try to interpret each figure/panel, then read the figure legend and the part of the results that explains it. [**Supplemental figures/tables** abound!]

   i. Do your interpretations match that of the authors'?

   ii. Are the results answering the specific Qs?

c. If "new" analytical approaches appear in *Results* without mention in *Methods*, the study is non-reproducible.

d. Make detailed notes on: 1) what's unclear, 2) what you might do differently.

Critical evaluation of research claims

6. Read the **Discussion/Conclusions** , **Title** , & **Abstract** :

a. Step back to think about contributions, limitations, open Qs, & next steps.

Red flags in scientific literature

- **Too-good-to-be-true results** : Results that are "too smooth" or perfectly replicate wiggles in noise are often signs of artifact or fraud.
- **Missing methodological details** : Be skeptical of papers that lack raw sample sizes, measures of variability, or descriptions of how treatments were allocated.
- **Suspicious citation patterns** : Over-reliance on "consensus views" to justify ignoring data inconsistencies can be a sign of motivated reasoning.

7. Read what other researchers (**papers that cite this paper**) say about this paper.

Critical evaluation of research claims

Evaluating research claims from media

Mass media often overgeneralizes findings, turning tentative associations into definitive "news"

TIME
TIME @TIME

The problem with your coffee

Hot Drinks a Probable Cancer Cause, Says WHO
time.com

4/9/17, 6:15 AM

TIME
TIME @TIME

How coffee can help you live longer

How Coffee Can Help You Live Longer
New findings add to growing evidence that co...
time.com

4/9/17, 6:45 AM

Critical evaluation of research claims

Evaluating research claims from media

Mass media often overgeneralizes findings, turning tentative associations into definitive "news"

xkcd.com/1217

https://qz.com/1595221/new-twitter-account-outs-shoddy-reporting-in-science-stories/

Let's play a game!

A Quick Puzzle to Test Your Problem Solving

By **DAVID LEONHARDT** and **YOU** JULY 2, 2015

A short game sheds light on government policy, corporate America and why no one likes to be wrong. **RELATED ARTICLE**

Here's how it works:

We've chosen a rule that some sequences of three numbers obey — and some do not. Your job is to guess what the rule is.

We'll start by telling you that the sequence 2, 4, 8 obeys the rule:

2	4	8	Obeys the rule

Now it's your turn. Enter a number sequence in the boxes below, and we'll tell you whether it satisfies the rule or not. You can test as many sequences as you want.

Enter your first sequence here:

			Check

I don't want to play; just tell me the answer.

https://www.nytimes.com/interactive/2015/07/03/upshot/a-quick-puzzle-to-test-your-problem-solving.html

Cognitive biases in data collection & interpretation

Confirmation bias

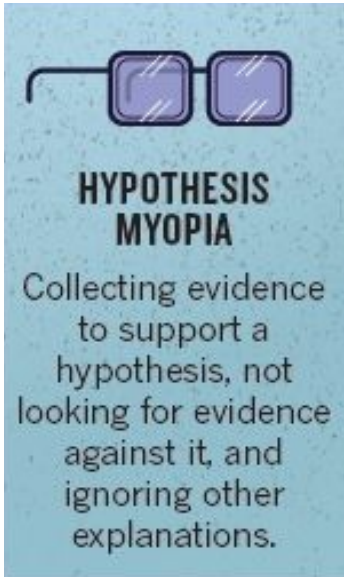
The tendency to search for, interpret, favor, and recall information in a way that confirms one's beliefs or hypotheses.

- Scientists rate studies that report findings consistent with their prior beliefs more favorably than studies reporting findings inconsistent with their previous beliefs.
- Data that conflict with the experimenter's expectations may be more readily discarded as unreliable.

"It is the peculiar and perpetual error of the human intellect to be more moved and excited by affirmatives than by negatives; whereas it ought properly to hold itself indifferently disposed towards both alike."
– Francis Bacon

https://www.nature.com/articles/nmeth.3091

Cognitive biases in data collection & interpretation



HYPOTHESIS MYOPIA

Collecting evidence to support a hypothesis, not looking for evidence against it, and ignoring other explanations.

TEXAS SHARPSHOOTER

Seizing on random patterns in the data and mistaking them for interesting findings.

ASYMMETRIC ATTENTION

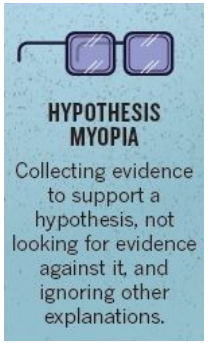
Rigorously checking unexpected results, but giving expected ones a free pass.

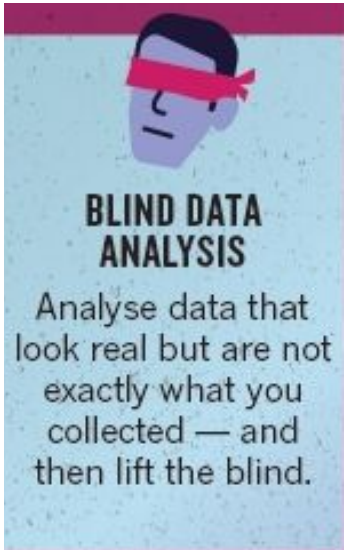
JUST-SO STORYTELLING

Finding stories after the fact to rationalize whatever the results turn out to be.

https://www.nature.com/news/how-scientists-fool-themselves-and-how-they-can-stop-1.18517

Cognitive biases in data collection & interpretation

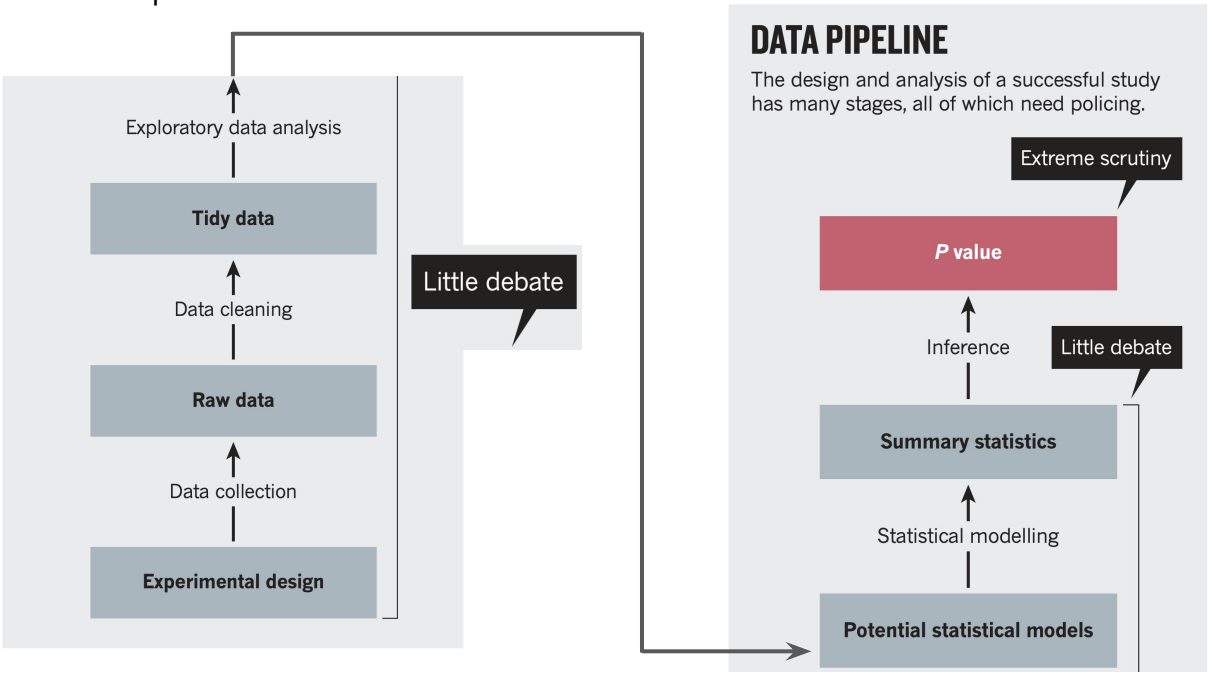




https://www.nature.com/news/how-scientists-fool-themselves-and-how-they-can-stop-1.18517

Statistical intuition vs. blind rule-following

Practical wisdom for analytical decisions : Shifting to expert thinking means understanding that all models are wrong, but some are useful. Expert intuition is "horse sense" about whether an effect size is clinically meaningful, regardless of its p-value.



JT Leek, RD Peng http://www.nature.com/news/statistics-p-values-are-just-the-tip-of-the-iceberg-1.17412

Statistical intuition vs. blind rule-following

When rules help vs. hinder : Following "Stat 101" checklists blindly is a recipe for disaster.

- Rules hinder when researchers treat the $p < 0.05$ threshold as a binary switch for truth.
- Rules help when they provide a framework for mathematical modeling under uncertainty.

The journey from novice to expert :

- From playing "statistical Mad Libs," filling in blanks in a computer-generated report, to…
- Knowing one's tools and how they break, recognizing that statistical assumptions are almost always violated in the real world.