HESP 180
REVIEW

Block 1 *Defining Science, Research, and Epidemiology*

https://epistatistics.com

https://epistatistics.com

CLINICAL EPIDEMIOLOGY LECTURES

Block One: Defining Science, Research, and Epidemiology

Lecture 1: Syllabus and the Etymology of "Science"

Lecture 2: Fundamentals of Science

Office Hours (Week 1)

Lecture 3: Fundamentals of Study Design

Lecture 4: Developing a Research Question

Lecture 5: From Bench to Bedside

Lecture 6: Introduction to Epidemiology

Office Hours (Week 2)

Lecture 7: Epidemiological Tools and Questions

Lecture 8: Drawing Inappropriate Conclusions

Office Hours (Week 3)

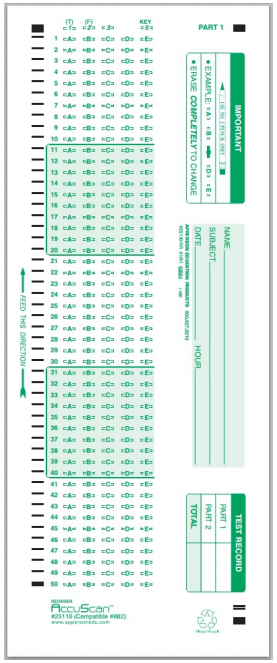
Lecture 9: Current Trends in Epidemiology

Lecture 10: Future of Epidemiology

Lecture 11: Review of Block One

Review Slides (PDF)

Block 1 *Defining Science, Research, and Epidemiology*



EXAM ONE

Scientific method, history of epidemiology, and fundamentals of evidence-based practice. The material you could be tested on includes everything presented during Lectures 1 – 10, especially (but not exclusively) what was focused on in Lecture 11… and potentially some material from Chapter 6 of the ACSM book.

Not on an actual Scantron, but all multiple choice.

Block 1 *Defining Science, Research, and Epidemiology*

Everything that you could potentially be accountable for

Lecture 2:

ACSM Research Methods: Chapter 1 (1-8)

Lecture 3:

ACSM Research Methods: Chapter 5 (73-90)

ACSM Research Methods: Chapter 7 (121-141)

Epidemiology 101: Chapter 6 (105-119) (2nd Edition: Chapter 7)

Lecture 4:

Epidemiology 101: Chapter 4 (65-87) (2nd Edition: Chapter 5)

ACSM Research Methods: Chapter 3 (29-49)

Lecture 5:

ACSM Research Methods: Chapter 6 (93-118)

Lecture 6:

Epidemiology 101: Chapter 1 (1-23)

RECAP SEQUENCE

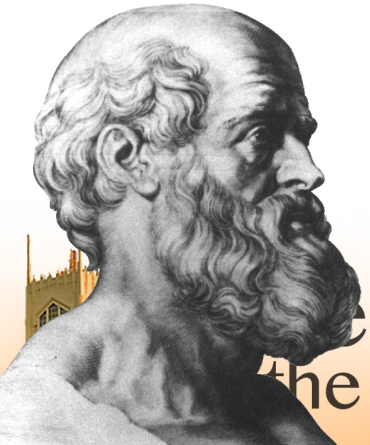
For the exam, I'll be drawing from everything we've talked about so far (Lectures 2–10). I've compiled the most important slides here; if you know all of this stuff, you're off to a really good start.

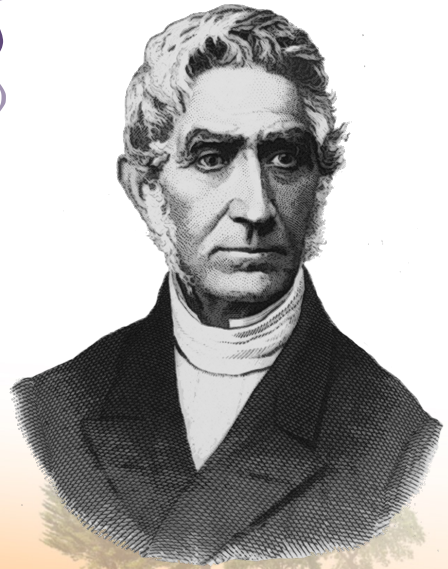
RECAP SEQUENCE

RECAP SEQUENCE

Epidemiology: A long time ago, Hippocrates discussed exposures and outcomes, and offered us some epidemiological nomenclature: *epidemic* and *endemic*. In the 17th century, "political arithmetic" entered public health, with the work of demographers such as John Graunt (1620–1674). In the middle of the 19th century, John Snow establishes epidemiology as a powerful tool for evaluating relationships between exposures and outcomes.

HIPPOCRATES
(460BC-ish TO 370 BC-ish)





JOHN SNOW
(AD 1813 TO AD 1858)

RECAP SEQUENCE

Epidemiology: In the years after John Snow, epidemiology developed and deformed itself. It became a mighty weapon against ignorance, but was often discharged with terrible aim.

Table 1. We have found 12 papers in which claims coming from observational studies were tested in randomised clinical trials. Many of the trials are quite large. In most of the observational studies multiple claims were tested, often in factorial designs, e.g. vitamin D and calcium individually and together along with a placebo group. Note that none of the claims replicated in the direction claimed in the observational studies and that there was statistical significance in the opposite direction five times

ID no.	Pos.	Neg.	No. of claims	Treatment(s)	Reference
1	0	1	3	Vit E, beta-carotene	*NEJM* 1994; **330**: 1029–1035
2	0	3	4	Hormone Replacement Ther.	*JAMA* 2003; **289**: 2651–2662, 2663–2672, 2673–2684
3	0	1	2	Vit E, beta-carotene	*JNCI* 2005; **97**: 481–488
4	0	0	3	Vit E	*JAMA* 2005; **293**: 1338–1347
5	0	0	3	Low Fat	*JAMA*. 2006; **295**: 655–666
6	0	0	3	Vit D, Calcium	*NEJM* 2006; **354**: 669–683
7	0	0	2	Folic acid, Vit B6, B12	*NEJM* 2006; **354**: 2764–2772
8	0	0	2	Low Fat	*JAMA* 2007; **298**: 289–298
9	0	0	12	Vit C, Vit E, beta-carotene	*Arch Intern Med* 2007; **167**: 1610–1618
10	0	0	12	Vit C, Vit E	*JAMA* 2008; **300**: 2123–2133
11	0	0	3	Vit E, Selenium	*JAMA* 2009; **301**: 39–51
12	0	0	3	HRT + Vitamins	*JAMA* 2002; **288**: 2431–2440
Totals	0	5	52		

significance september2011

Deming, data and observational studies

A process out of control and needing fixing

RECAP SEQUENCE

Good

The New York Times

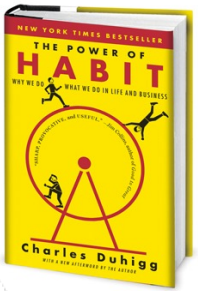
BUSINESS DAY

What Wal-Mart Knows About Customers' Habits

By CONSTANCE L. HAYS NOV. 14, 2004

RECAP SEQUENCE

Good & Bad



TARGET®

TARGET COUPON EXPIRES 2/15/16

$1 off 20-ct. or larger Pampers diapers

Item Coupon. Valid in store only. Limit one manufacturer and one Target coupon per item. Limit of 4 identical coupons per household, per day. Void if copied, scanned, altered, transferred, purchased, sold or prohibited by law. Item(s) may not be available at all stores. Quantities limited; no rain checks. Coupon value may not exceed value of item purchased. No cash value.

9856-0112-1554-4755-0205-2604-69

RECAP SEQUENCE

Positive Side of Neutral

Table 1. We have found 12 papers in which claims coming from observational studies were tested in randomised clinical trials. Many of the trials are quite large. In most of the observational studies multiple claims were tested, often in factorial designs, e.g. vitamin D and calcium individually and together along with a placebo group. Note that none of the claims replicated in the direction claimed in the observational studies and that there was statistical significance in the opposite direction five times

ID no.	*Pos.*	*Neg.*	*No. of claims*	*Treatment(s)*	*Reference*
1	0	1	3	Vit E, beta-carotene	*NEJM* 1994; **330**: 1029–1035
2	0	3	4	Hormone Replacement Ther.	*JAMA* 2003; **289**: 2651–2662, 2663–2672, 2673–2684
3	0	1	2	Vit E, beta-carotene	*JNCI* 2005; **97**: 481–488
4	0	0	3	Vit E	*JAMA* 2005; **293**: 1338–1347
5	0	0	3	Low Fat	*JAMA*. 2006; **295**: 655–666
6	0	0	3	Vit D, Calcium	*NEJM* 2006; **354**: 669–683
7	0	0	2	Folic acid, Vit B6, B12	*NEJM* 2006; **354**: 2764–2772
8	0	0	2	Low Fat	*JAMA* 2007; **298**: 289–298
9	0	0	12	Vit C, Vit E, beta-carotene	*Arch Intern Med* 2007; **167**: 1610–1618
10	0	0	12	Vit C, Vit E	*JAMA* 2008; **300**: 2123–2133
11	0	0	3	Vit E, Selenium	*JAMA* 2009; **301**: 39–51
12	0	0	3	HRT + Vitamins	*JAMA* 2002; **288**: 2431–2440
Totals	0	5	52		

significance september2011

Deming, data and observational studies

A process out of control and needing fixing

RECAP SEQUENCE

Good

RECAP SEQUENCE

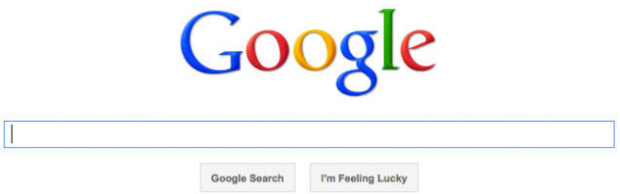
Mostly Good

nature Vol 457|19 February 2009|doi:10.1038/nature07634

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

[1]Google Inc., 1600 Amphitheatre Parkway, Mountain View, California 94043, USA. [2]Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA.



The final model was validated on 42 points per region of previously untested data from 2007 to 2008, which were excluded from all previous steps. Estimates generated for these 42 points obtained a mean correlation of 0.97 (min = 0.92, max = 0.99, $n = 9$ regions) with the CDC-observed ILI percentages.

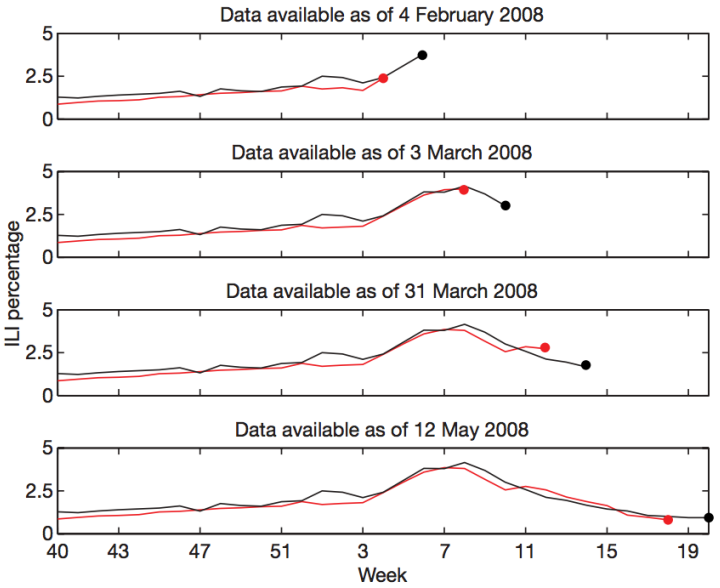
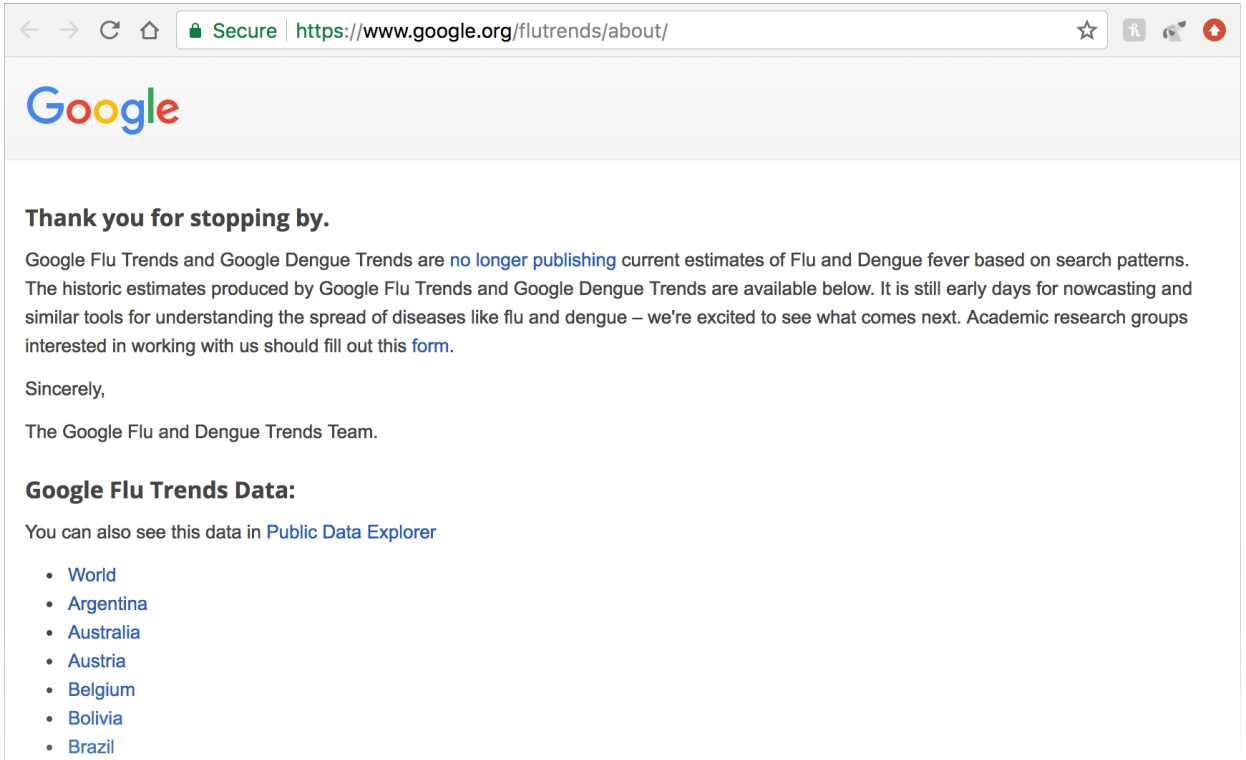


Figure 3 | ILI percentages estimated by our model (black) and provided by the CDC (red) in the mid-Atlantic region, showing data available at four points in the 2007-2008 influenza season.

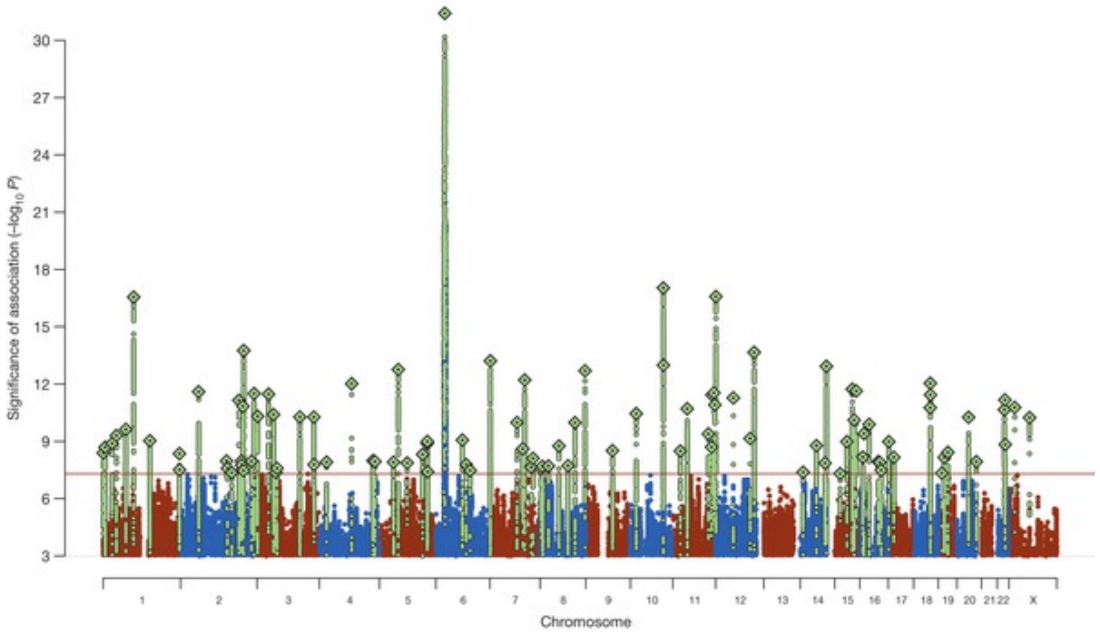
RECAP SEQUENCE

Mostly Good



RECAP SEQUENCE

Mostly Good, but…

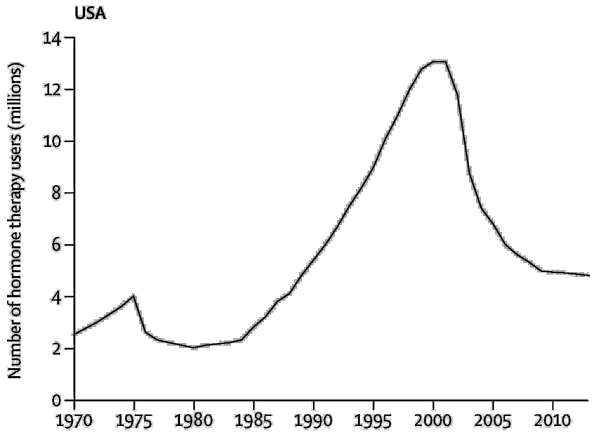


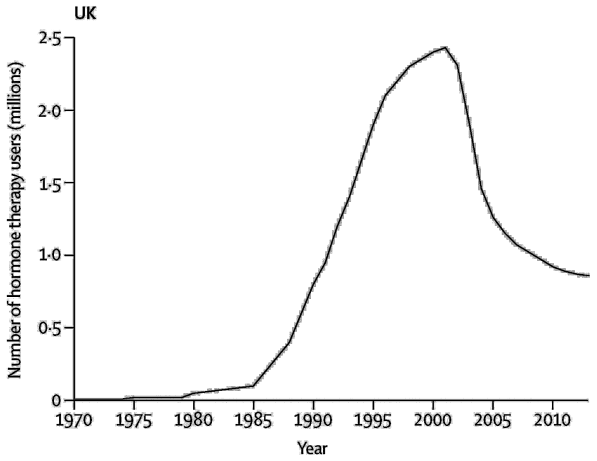
JULY 2014 | VOL 511 | NATURE

Biological insights from 108 schizophrenia-associated genetic loci

RECAP SEQUENCE

A Bit More Good Than Bad





RECAP SEQUENCE

Barely Leaning Bad

RECAP SEQUENCE

Positive Side of Neutral

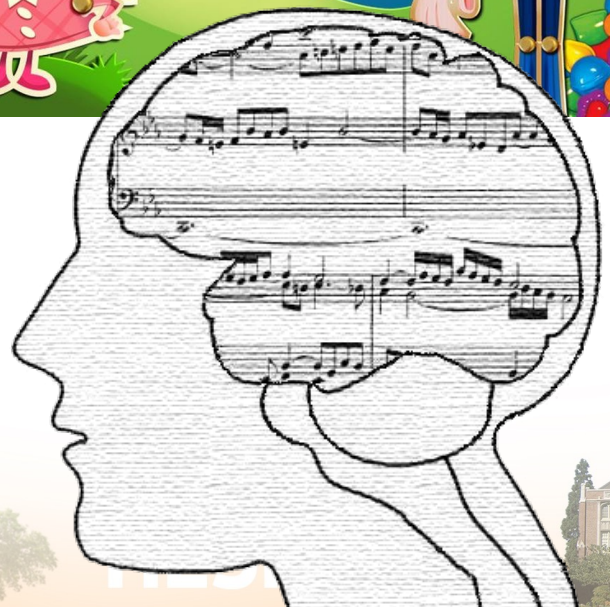
Maternal Grandmother's Smoking Correlated With Autism Diagnosis

40.5K
SHARES

Share on Facebook | Share on Twitter | +

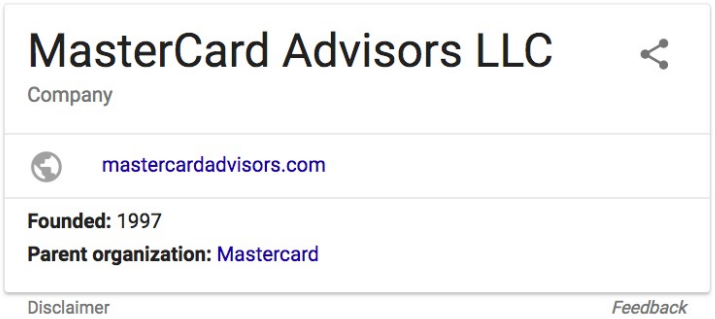
RECAP SEQUENCE

Good (Excepting Artistic Value)



RECAP SEQUENCE

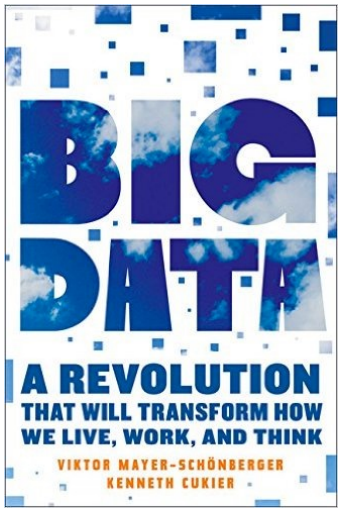
Mostly Good



MasterCard Advisors is a division of the company that aggregates and analyzes tons of transactions (>65 billion) from tons of card holders (>1 billion) in lots of countries (>250) to understand and predict consumer trends.

RECAP SEQUENCE

Mostly Good



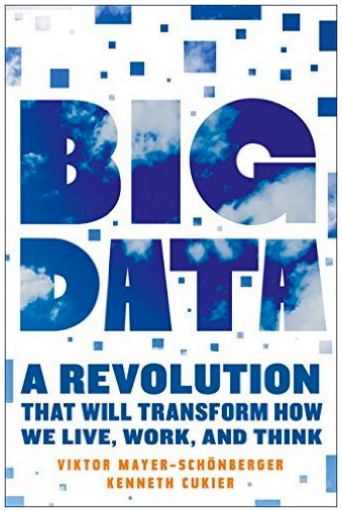
7 IMPLICATIONS 127

Analysts identified a trend: People filled up their gas tanks around 4 in the afternoon had a high likelihood of spending $35 to $50 at a restaurant or grocery store within an hour.

How can that finding be used?

RECAP SEQUENCE

Mostly Good



7 IMPLICATIONS 127

Sell the information to gas stations so they can sell advertising to restaurants and stores, printing coupons on the back of gas receipts.

Epidemiology is true here, but all it's doing is characterizing patterns, not outlining cause and effect.

RECAP SEQUENCE

Mostly Good, But…

Fitbit is tracking your sleep. Your steps. Your heart rate. Not just for you.

The Health Insurance Portability and Accountability Act (HIPAA) was written in 1996. Its writers didn't anticipate the iPhone or the FitBit. The collecting and sharing of data doesn't fall under HIPAA restrictions. The Federal Food, Drug, and Cosmetic Act discusses devices, but not of this variety. Your information is being analyzed, just like your Amazon purchases.

RECAP SEQUENCE

Mostly Good

RECAP SEQUENCE

Mostly Good

Once upon a time, there were in-house reviewers. Then, there wasn't anymore. Algorithmic recommendations recommend better. And that dries way (way) more sales. What are those algorithms doing?

(Hint: epidemiolog……)

Today, those recommendations and other algorithm-driven personalization features create about a third of Amazon's sales.

R SEQUENCE

To succeed in practically anything in the 21st century, understanding of statistics is critical. It's every bit as important as literacy was just a few centuries ago. Technically these aren't my words. I never said anything like this. Courtney is just pretending I did.

ood

RECAP SEQUENCE

Mostly Bad

Probability and Preemptive Punishment

In the year 2054, there's a specialized police department called "PreCrime". They imprisoned people for crimes that they were predicted to commit pretty soon.

(It was some weird, hairless clairvoyant women-ish things, and not data, but data can do it too… sort of.)

RECAP SEQUENCE

Mostly Bad

Probability and Preemptive Punishment

Non-cinematic world:

Parole boards in more than half of the states use data analyses to predict whether to release someone from prison or keep them locked up.

Certain precincts (e.g., LA) use predictive policing. Algorithms predict people and places that are more likely to need policing. So the police are a bit more vigilant with those regions and people.

RECAP SEQUENCE

Depends How You Use It

Probability and Preemptive Punishment

Richard Berk (a full professor of stats and criminology at Upenn) makes predictions on whether a person who gets released on parole will then kill someone or be killed. Using his predictors, he does so with ~75% accuracy.

RECAP SEQUENCE

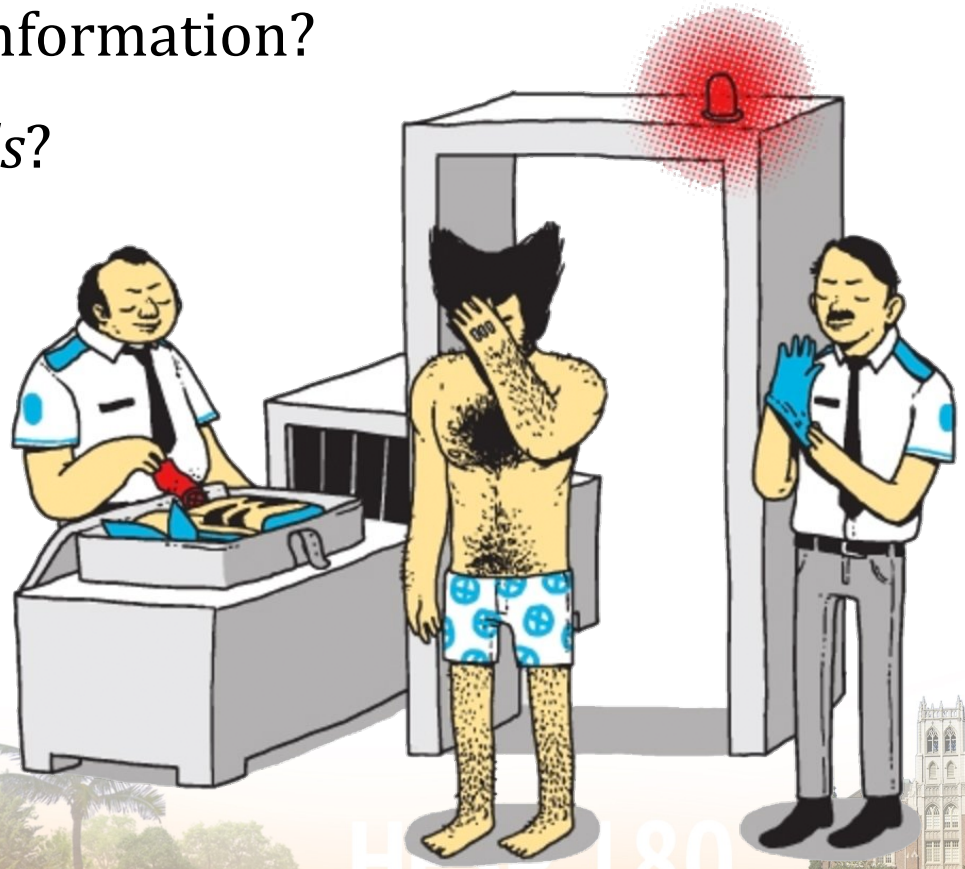
Depends How You Use It

NSA intercepts 1.7 billion communications every day.

(according to the Washington Post in 2010)

What do you do with that information?

Should we target *individuals*?



RECAP SEQUENCE

Use It

Epidemiology doesn't discriminate.
The population is the population.
There's no such thing as an individual.

very day.

n Post in 2010)

When we prescribe, it gets risky…

R E C A P S E Q U E N C E

What's science?

Science is the systematic study of the structure and behavior of the physical and natural world through observation and experimentation.

RECAP SEQUENCE

What's science?

Why is science important?

How is science typically communicated?

1988. Not famous. Nobody liked this.

1993. Ridiculously simple. Sold millions.

RECAP SEQUENCE

What's science?

Why is science important?

How is science typically communicated?

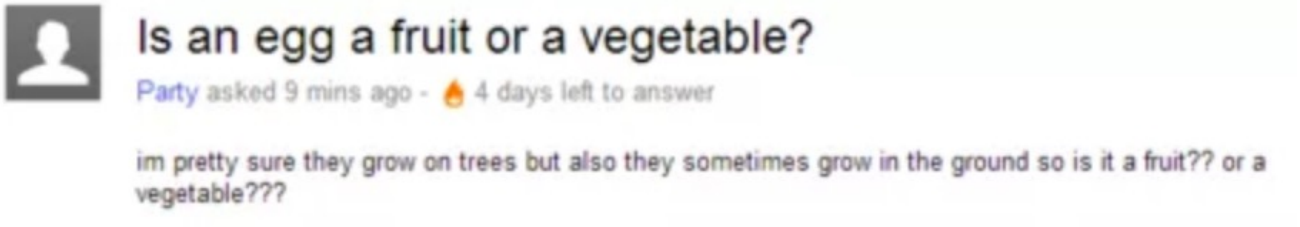
Why is this (or why is it not) a problem?

How do you fix it?

What's the purpose of research?

What's a good place to start with research?

RECAP SEQUENCE



What's a good place to start with research?

RECAP SEQUENCE

Worth asking: *What am I good at?*
What's a useful thing to know?

I have decided to research what others have already researched. And I've decided to do it in exactly the same way so that my work adds nothing. I want to be boring and pointless. ☺

I'm horrible at working with people. Everyone deeply hates me. I should do research that involves lots and lots of human interaction!

RECAP SEQUENCE

FROM BENCH TO BEDSIDE

95 Although each of these examples occupies a different stage in the bench-to-bedside continuum, the respective researchers have a similar set of challenges to overcome. First, they must identify the purpose of their research (as well as the end user it proposes to affect). Only then can one properly decide how the data are to be collected. Once the data are collected, the nature of those data must guide the choice of [illegible]nalysis. Statistics is really a process of coping with uncertainty.

RECAP SEQUENCE

First, they must identify the purpose of their research (as well as the end user it proposes to affect).

BENCH TO BEDSIDE CONTINUUM

RECAP SEQUENCE

How does exercise affect an individual's health?

Is this a valuable, thought out, researchable question to ask?

RECAP SEQUENCE

How does eight weeks of thrice weekly aerobic exercise (40 min sessions on a cycle ergometer at 60% VO_2 max) affect serum lipoprotein levels (HDLs, LDLs) of post-menopausal women?

RECAP

All right, maybe. Maybe you do.
The most basic, fundamental thing to know is what variables are. Stuff like diet and health. You can't claim to know anything about science without a solid understanding of this foundation.
So what do you know about variables?

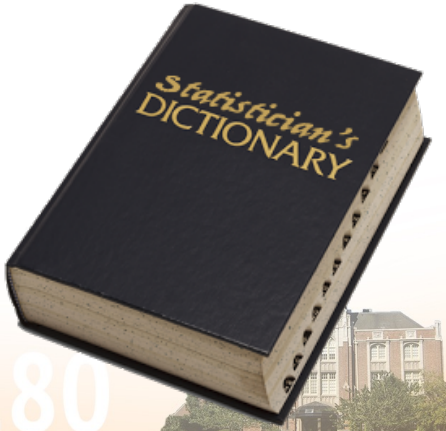
I know all about cheese cancer.

RECAP SEQUENCE

Variable: A *thing* in research that has different values. This includes just about anything that can be measured.

Independent Variable: This is the measurable thing you're controlling or manipulating to produce an effect.

Dependent Variable: This is the measurable thing that *gets* affected *by* the independent variable.

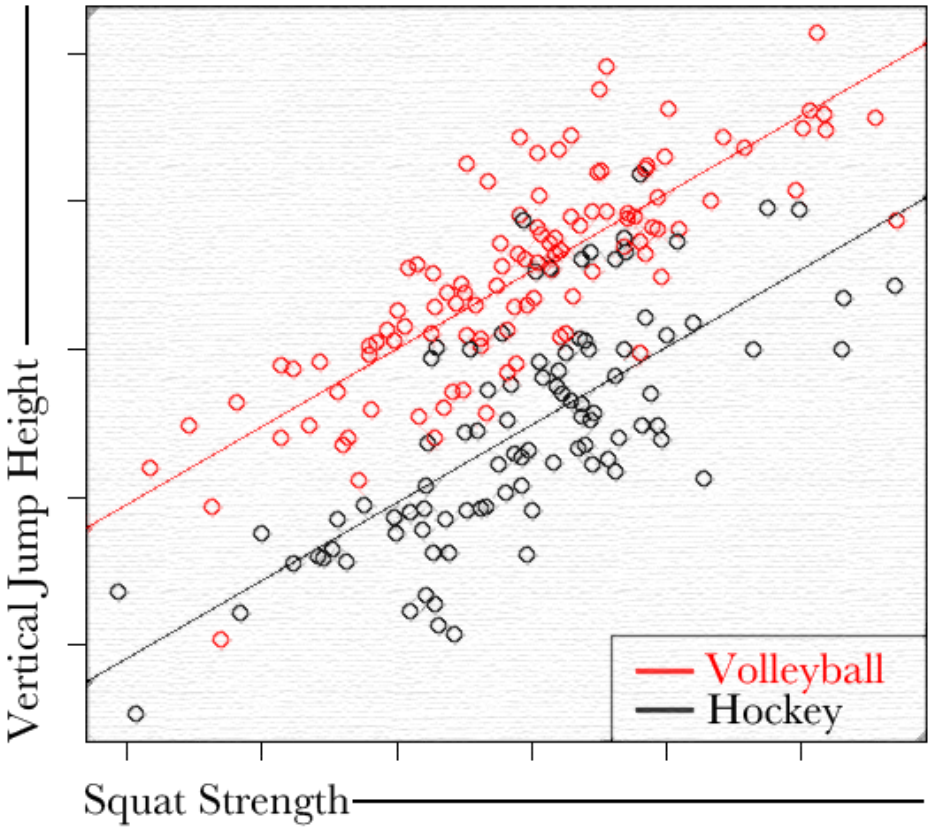


RECAP SEQUENCE

Scattered results

Variables vary, so they have "**variance.**"

In other words, the results will be very scattered.



RECAP SEQUENCE

Perfect results

High school algebra:

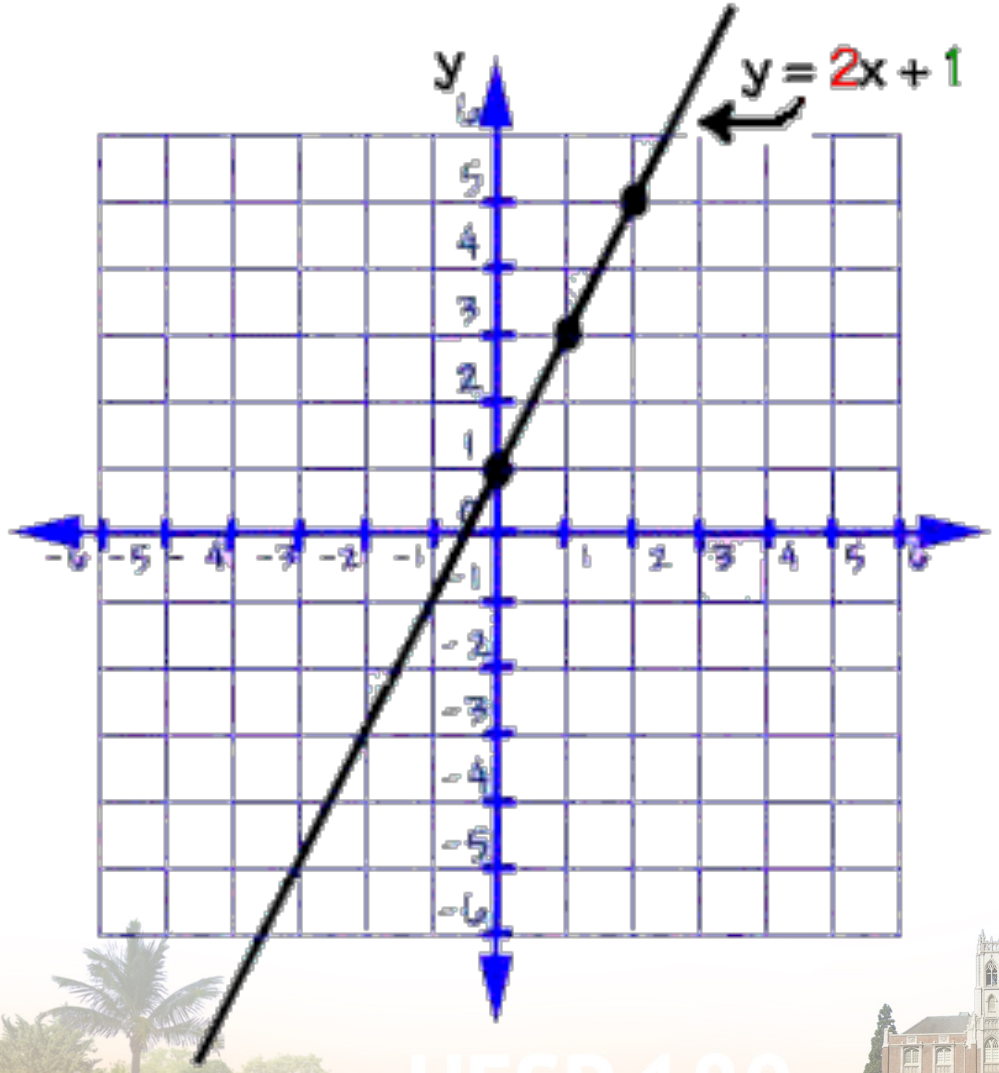
The equation of a straight line (i.e., a "linear equation") is written as:

$$y = mx + b$$

m = slope of the line

b = y-intercept.

X and Y (the two variables) would have a perfect correlation there.

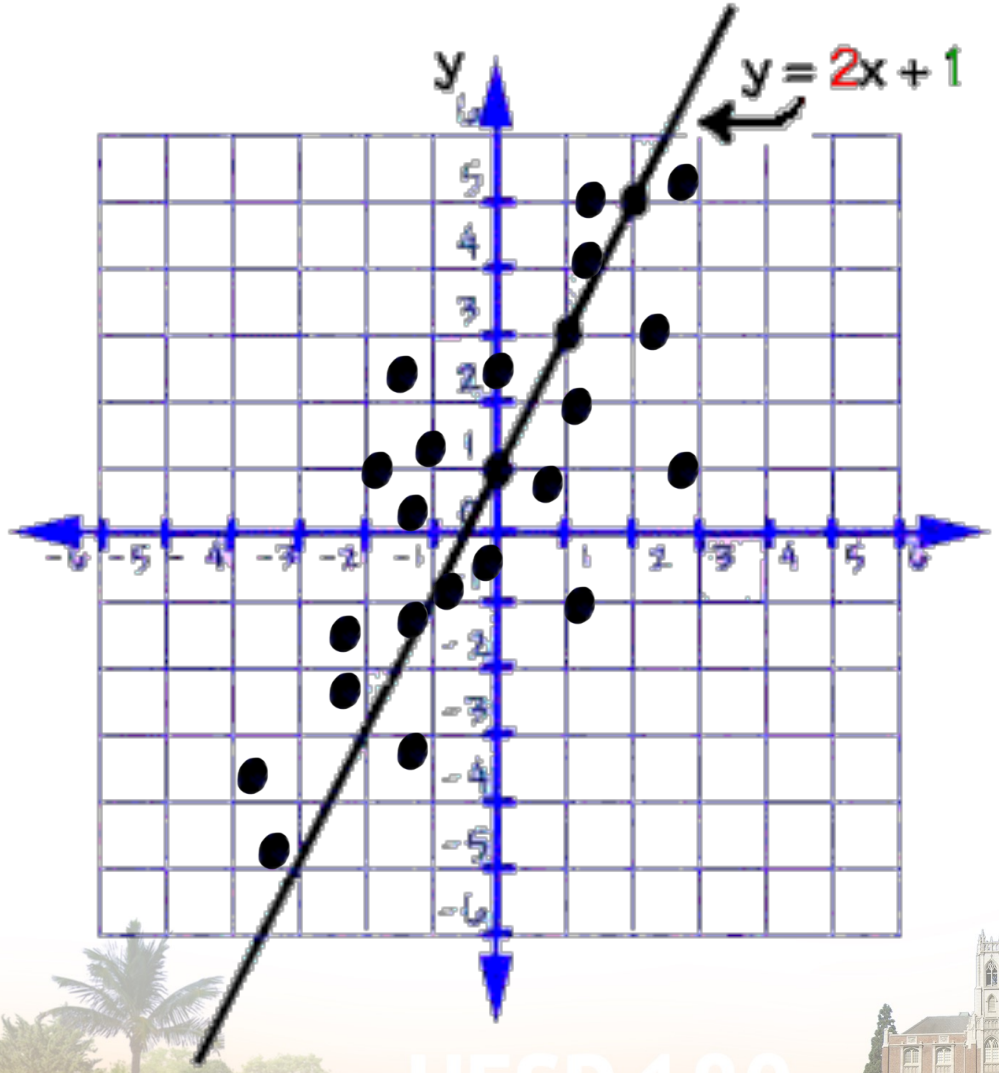


RECAP SEQUENCE

Imperfect results

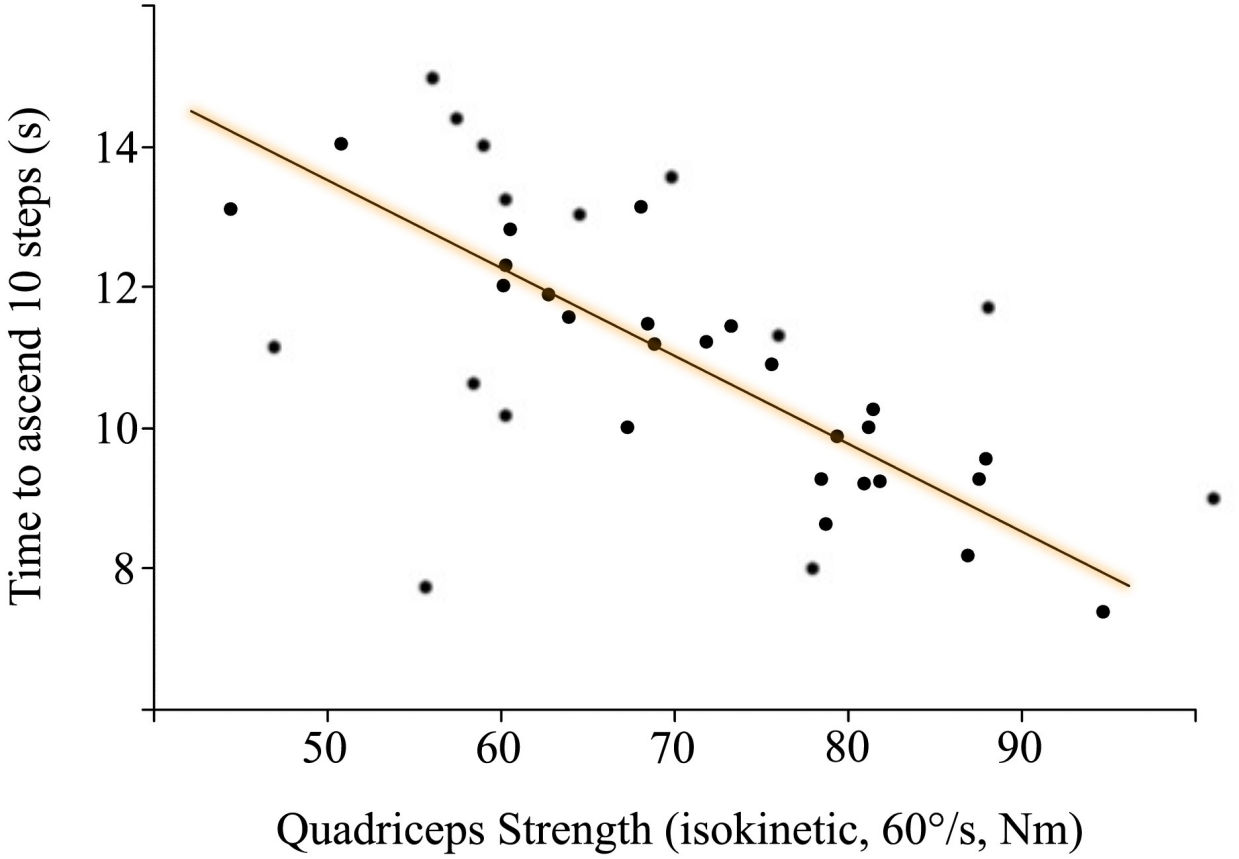
High school algebra:

In nature (in reality) there are very few perfect correlations. Almost everything is scattered.



RECAP SEQUENCE

You have to (statistically) make sense of all of that variance

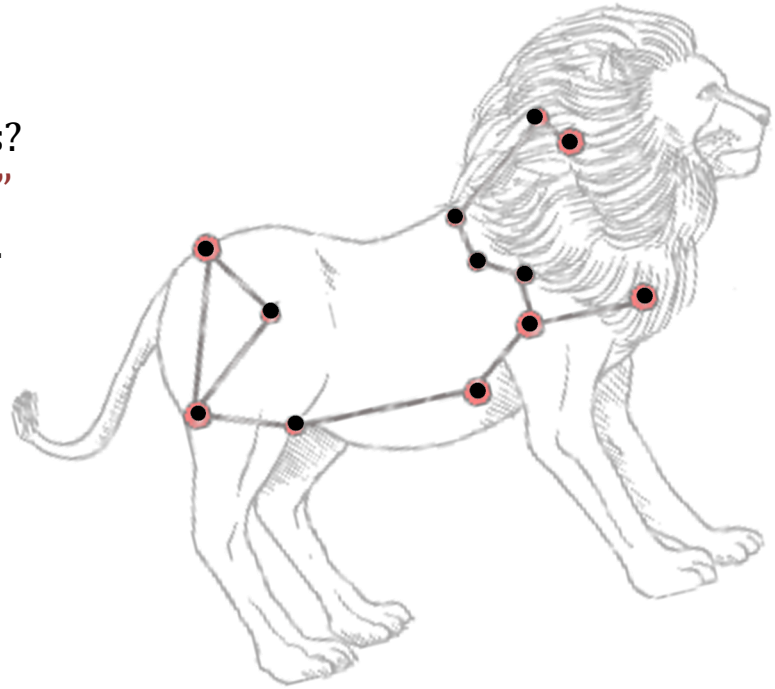


RECAP SEQUENCE

Variance: how spread out are the values?

"How scattered are the data points?"

This question requires a reference point.



RECAP SEQUENCE

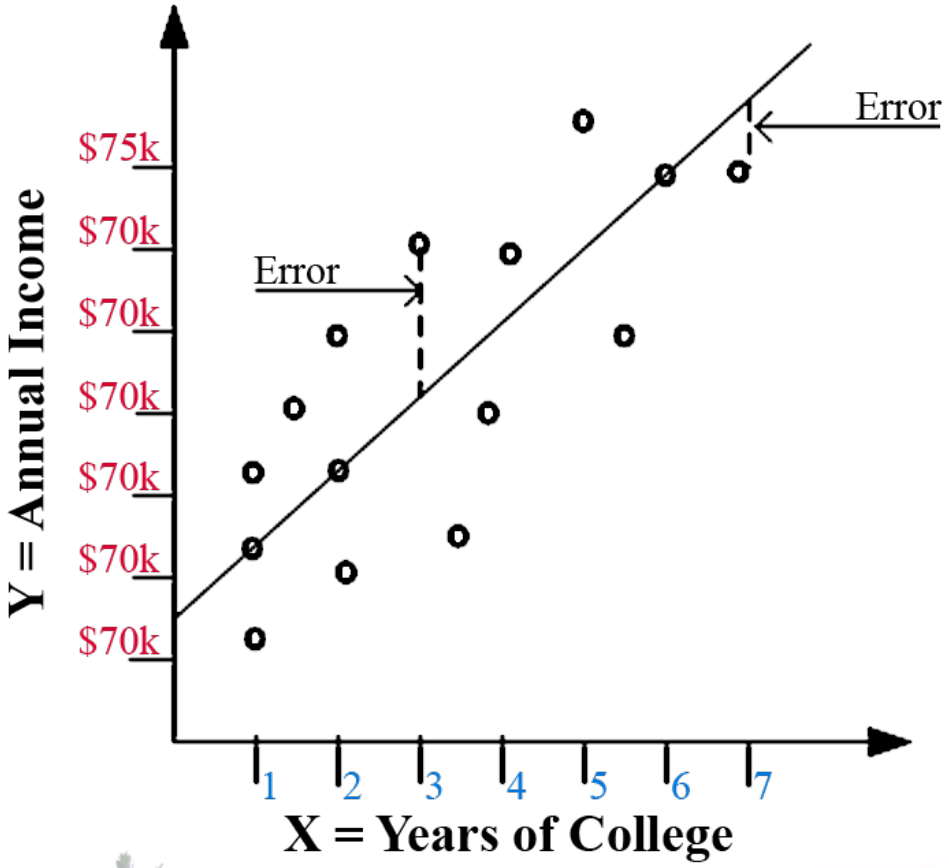
Variance: how sprea
"How scattered are
This question require

Bill Herrin
Director of SIS

RECAP SEQUENCE

variance:

How far are **the data** scattered from the **mean**?



RECAP SEQUENCE

Very, very occasionally, your results won't be all that scattered:

Group A

Group B

RECAP SEQUENCE

Binary outcome: You live or you die. Two **categorical** options.

Group A

Group B

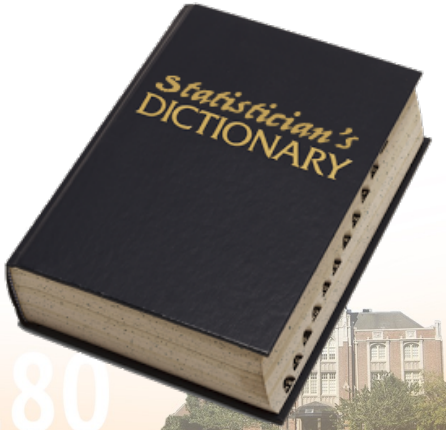
RECAP SEQUENCE

Types of variables:

Categorical variables. Dichotomous, nominal, ordinal.

Discrete random variables. These are countable. 1, 2, 3, 4, 5, etc. Goals in soccer, runs in baseball, times you've eaten donkey meat, etc. Infinite possible numbers, but they're all integers.

Continuous random variables. These values can't be counted. 1.2492, 3¾, 32.12589747. Weight, volume, duration, distance, etc. *Seriously* infinite number of possibilities.



RECAP SEQUENCE

Scatter in a binary outcome:

1972: W L L W L L W W W W L W W W W W L L L W L L L L W W W W L L L W W W L

1973: W W W L L W L L W W W L L L W L W L W L L W L W L L L W W L L W L L W W W W W W W

1974: W L W L L L W W W L W W L W W L W W W W L L L W W L W L W L W W L L W L W W W L W W

RECAP SEQUENCE

Scatter in a binary outcome:

1972: W L L W L L W W W W L W W W W W L L L W L L L L W W W W L L L W W W L

1973: W W W L L W L L W W W L L L W L W L W L L W L W L L L W W L L W L L W W W W W W W

1974: W L W L L L W W W L W W L W W L W W W W L L L W W L W L W L W W L L W L W W W L W W

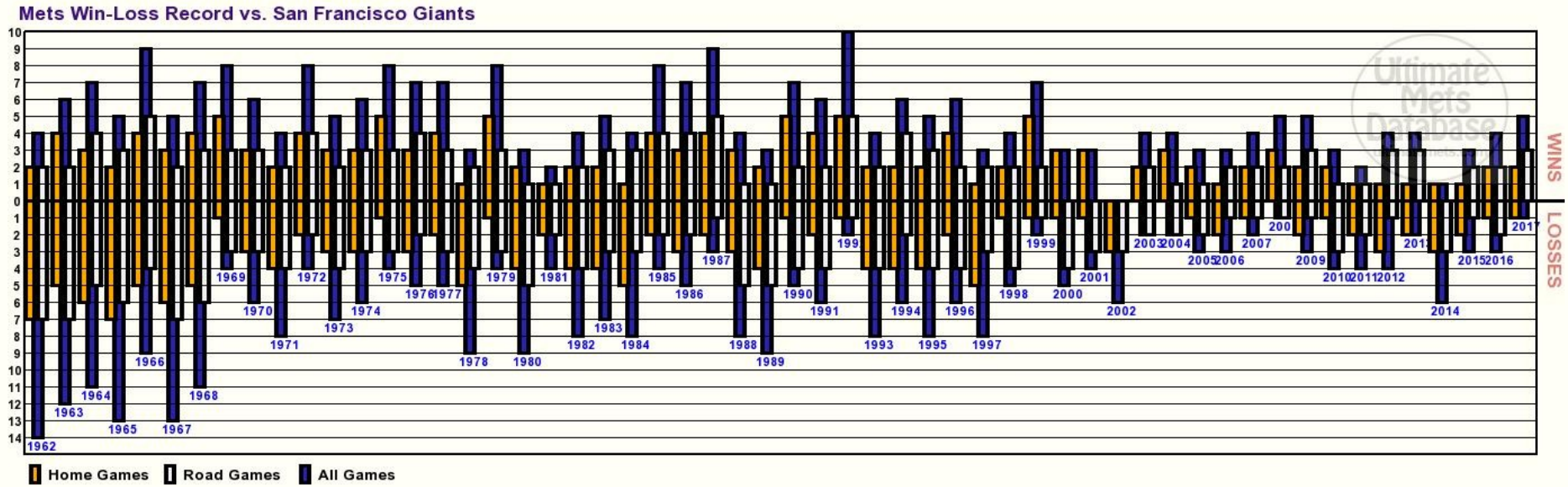
Types of questions you could ask:

What variables predict wins and losses? Who you're playing against? Time of game? Temperature? Miles traveled? How many runs your offense scored? How many errors your defense mad How do wins and losses affect salary?

1972: $27,000 salary
1973: $47,000 salary
1974: $100,000 salary

RECAP SEQUENCE

Binary nominal data (win/loss) converted into ratio data:



RECAP SEQUENCE

The goal in statistics is to account for and explain that variance.
Explain how an independent variable makes a dependent vary.

Statistical **infer**ence.

Your goal is to infer things.
Take a *sample* (e.g., 100 people)
from the population (7.5 billion).
Analyze that sample of 100 people.
And then make some inferences
from your analyses about the
larger population.

RECAP SEQUENCE

CHAPTER 6 *Understanding Research*

FROM BENCH TO BEDSIDE

95 Although each of these examples occupies a different stage in the bench-to-bedside continuum, the respective researchers have a similar set of challenges to overcome. First, they must identify the purpose of their research (as well as the end user it proposes to affect). Only then can one properly decide how the data are to be collected. Once the data are collected, the nature of those data must guide the choice of statistical analysis. **Statistics is really a process of coping with uncertainty.**

RECAP SEQUENCE

CHAPTER 6 *Understanding Research*

UNCERTAINTY AND CONSEQUENCES 106–107

"In this world nothing can be said to be certain, except death and taxes" (a letter from Benjamin Franklin to Jean-Baptiste Leroy, 1789) (15). The passage of more than two centuries has not altered the truth in Franklin's words. Uncertainty is all around us, and it is the pursuit of "truth" or "proof" that drives research. Research yields data, but we are never sure that the data collected from a sample reflect the whole of the population from which the sample was drawn. This uncertainty led to the development of **inferential statistics** or analyses that permit researchers to estimate the probability that the data gathered are reflective of the population. To some extent, these processes manage the uncertainty inherent in research.

RECAP SEQUENCE

CHAPTER 6 *Understanding Research*

UNCERTAINTY AND CONSEQUENCES 107

Uncertainty, however, is a messy subject. Consider cigarette smoking. The link between smoking and lung cancer is well known and widely accepted. More than 85% of lung cancer cases are attributable to smoking (19). However, fewer than 20% of people who smoke cigarettes develop lung cancer. In southwest England, the risk of dying from lung cancer among all lifetime smokers has been estimated to be 15.9% for men and 9% for women. If those lifetime smokers have spent their lives smoking more than 25 cigarettes a day, the risks rise to 24.4% for men and 18.5% for women (14). In Canada, lifetime risks for developing lung cancer among current smokers were 17.2% for men and 11.6% for women (20). At the population level, the link between smoking and an increased risk of developing lung cancer is approaching the certainty of death and taxes, but at the individual level, the fact that someone smokes does not assure the development of the disease.

RECAP SEQUENCE

CHAPTER 6 *Understanding Research*

UNCERTAINTY AND CONSEQUENCES

Some of this uncertainty can be managed by minimizing error in the study design, both systematic and random. What's the difference between the two types?

RECAP SEQUENCE

In conducting research, there are:
Systematic Errors and Random Errors

Measurement error:

- Instruments aren't working right (e.g., improperly calibrated)
- Human error (e.g., people collecting data are bad at data collection)

Generally manageable by better preparation, maintenance of tools, training of staff, supervision to ensure technician consistency (e.g., interrater reliability), appropriate instruments and software, etc.

RECAP SEQUENCE

In conducting research, there are: Systematic Errors and Random Errors

Unknown deviations that disrupt conclusions:

- Undefined, unpredictable, or otherwise unknown effects
- Setting/environmental factors, lack of enforced controls

Generally manageable if the researcher has enough foresight to identify possible confounders and create proper inclusion/exclusion criteria.

RECAP SEQUENCE

After controlling for as much error as possible, we remain a bit uncertain about how variables interact and affect each another. What effect will a collection of independents have on a dependent variable? The world is packed with variables. How do we use them to test predictions?

RECAP SEQUENCE

- Years of participation in an activity
- Presence of a specific allele
- Free throw percentage
- Race/ethnicity
- Hours of sleep per night
- Incidence of the flu
- Number of bowls of crispy hexagons
- Change in body fat percentage
- Which doctor performed the procedure
- Marital status
- Use of anti-inflammatory drugs
- Age of onset of Alzheimer's
- Severity of stress
- Sex/gender
- Mile time

*Which are most likely to be **independent variables** and which ones are **dependents**?*

RECAP SEQUENCE

- Years of participation in an activity
- Presence of a specific allele
- Free throw percentage
- Race/ethnicity
- Hours of sleep per night
- Incidence of the flu
- Number of bowls of crispy hexagons
- Change in body fat percentage
- Which doctor performed the procedure
- Marital status
- Use of anti-inflammatory drugs
- Age of onset of Alzheimer's
- Severity of stress
- Sex/gender
- Mile time

*How might you control for the influence of **these** on your outcome variable?*

RECAP SEQUENCE

- Presence of a specific allele
- Race/ethnicity
- Which doctor performed the procedure
- Marital status
- Sex/gender

*How might you control for the influence of **these** on your outcome variable?*

Make them "**constants**" (only look at one gender, one race, etc.).

RECAP SEQUENCE

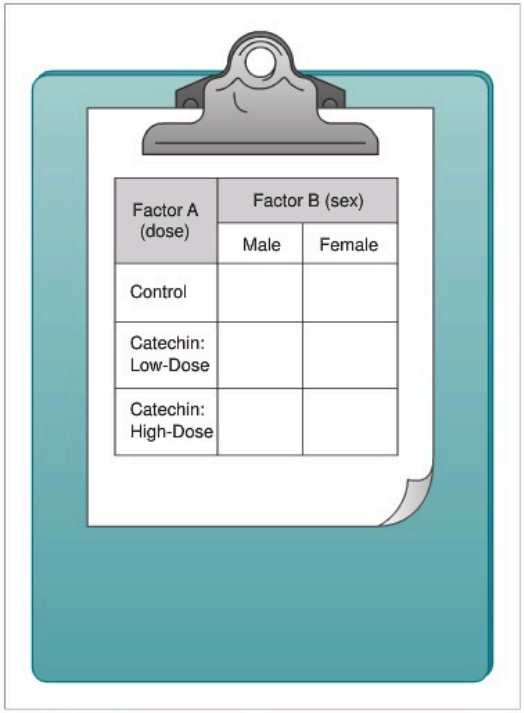


FIGURE 7-1: Schematic of a two-factor design with three levels of factor A (dose of catechin) and two levels for the between factor B of sex (male or female).

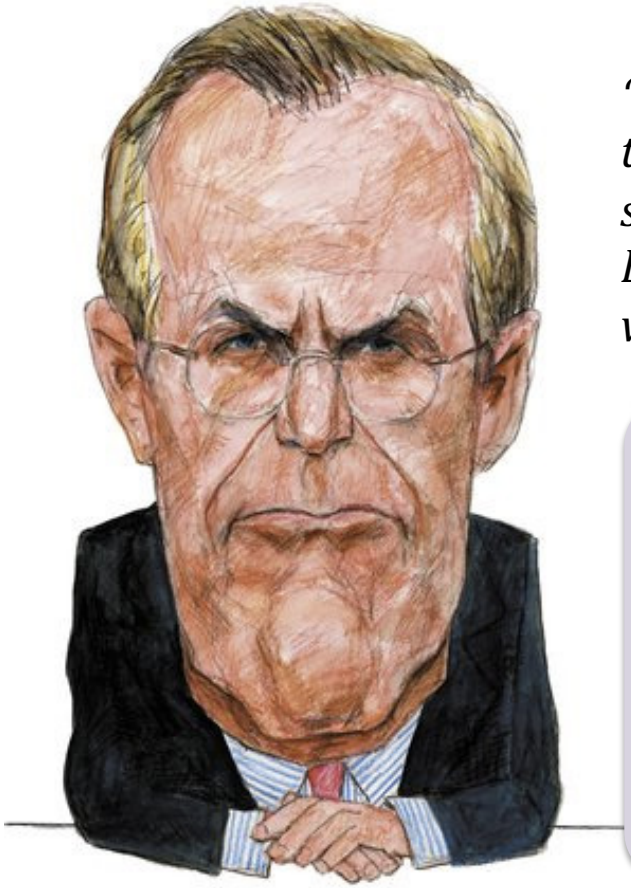
*How might you control for the influence of **these** on your outcome variable?*

Or you can just evaluate them as independents.

RECAP SEQUENCE

Nominal	Ordinal	Interval	Ratio
Categorical, no intrinsic order	Categorical, level or rank	Categorical, level or rank, uniform spacing, no zero point	Categorical, level or rank, uniform spacing, zero as reference point
E.g.: gender, race, religion, political party	E.g.: white belt, brown belt, black belt	E.g.: Borg RPE scale (6–20), degrees Celsius	E.g.: height, speed, weight, area (acreage), distance

RECAP SEQUENCE

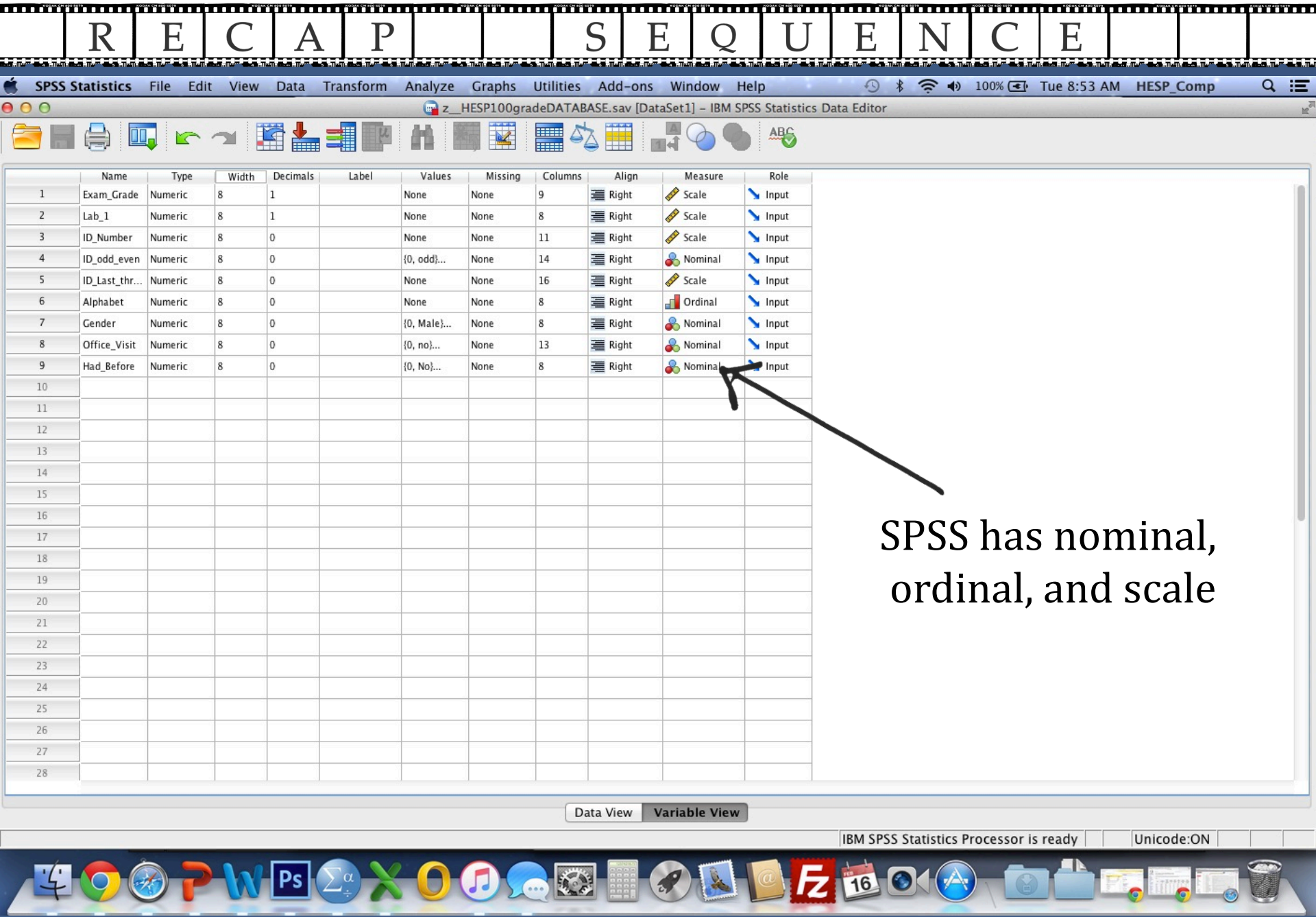


"There are known knowns; there are things we know that we know. There are known unknowns; that is to say, there are things that we now know we don't know. But there are also unknown unknowns – there are things we do not know we don't know."

Extraneous

- Also called "confounding" or "intervening"
- Not controlled or accounted for in study design
- E.g.: nutrition status, level of fatigue

RECAP SEQUENCE



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Exam_Grade	Numeric	8	1		None	None	9	Right	Scale	Input
2	Lab_1	Numeric	8	1		None	None	8	Right	Scale	Input
3	ID_Number	Numeric	8	0		None	None	11	Right	Scale	Input
4	ID_odd_even	Numeric	8	0		{0, odd}...	None	14	Right	Nominal	Input
5	ID_Last_thr...	Numeric	8	0		None	None	16	Right	Scale	Input
6	Alphabet	Numeric	8	0		None	None	8	Right	Ordinal	Input
7	Gender	Numeric	8	0		{0, Male}...	None	8	Right	Nominal	Input
8	Office_Visit	Numeric	8	0		{0, no}...	None	13	Right	Nominal	Input
9	Had_Before	Numeric	8	0		{0, No}...	None	8	Right	Nominal	Input
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											
26											
27											
28											

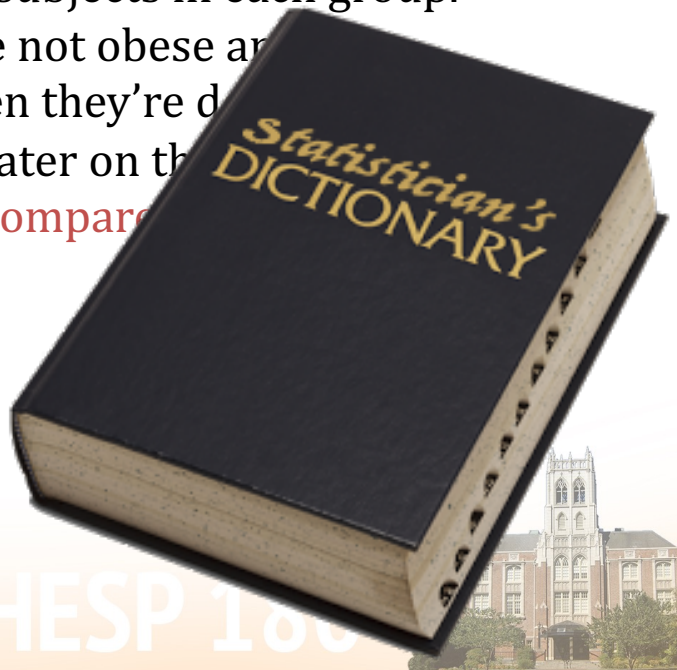
SPSS has nominal, ordinal, and scale

RECAP SEQUENCE

There are different kinds of independent variables:

Between-group variable: You're looking at different subjects in each group. Men vs. women. Obese vs. non-obese. Republicans vs. democrats. You can run these analyses with a single episode of data collection (one time point).

Within-group variable: You're looking at the same subjects in each group. People when they were obese and then when they're not obese ar
People when they're republicans, and then later when they're d
People in a drug trial while they're on the drug and later on th
You need at least two episodes of data collection to compar
subjects under different conditions.



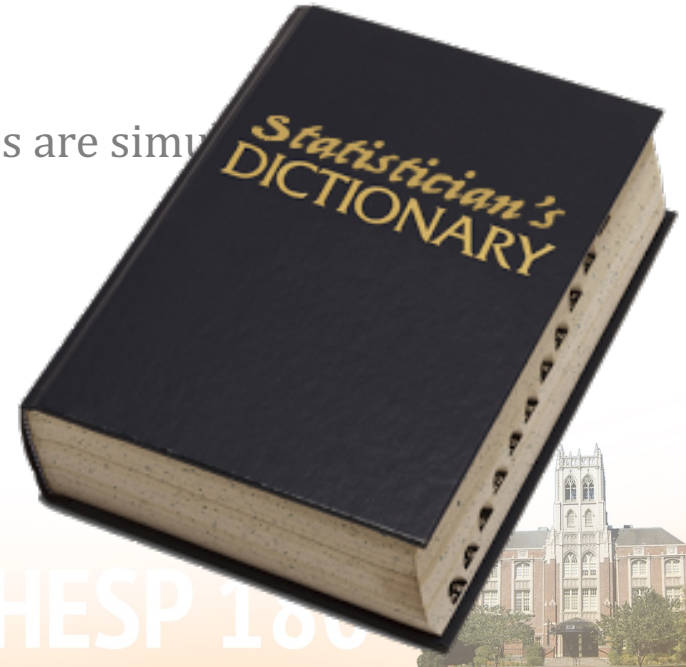
RECAP SEQUENCE

There are different kinds of research models:

In a statistical model, you can have any number of independent (explanatory) variables explaining a dependent (outcome) variable. The number of *dependent* variables affects the class of analysis:

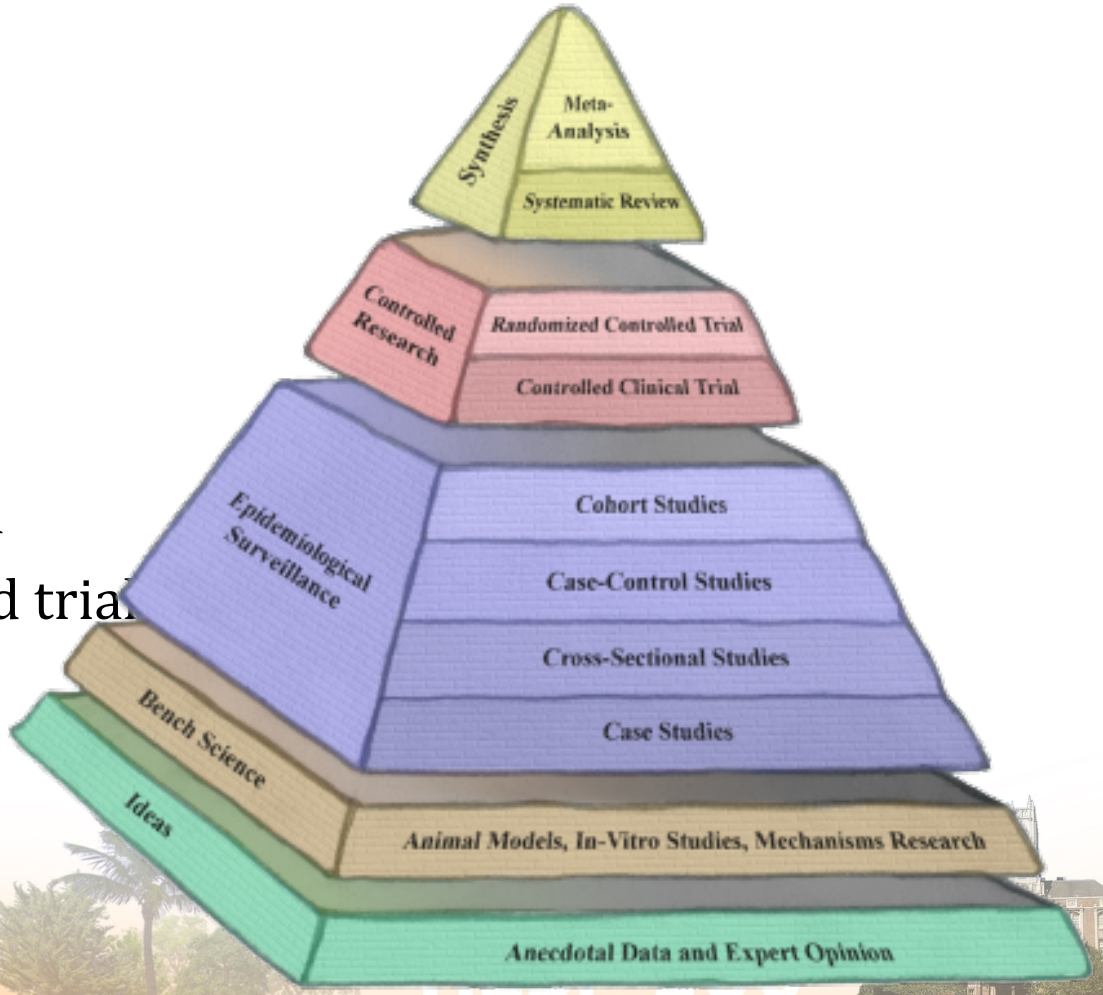
Univariate analysis: One dependent variable is examined (with as many independents as you want, sort of).

Multivariate analysis: Multiple dependent variables are simu evaluated (using a bunch of independents).



RECAP SEQUENCE

- Observational research
- Descriptive research
- Ecological study
- Correlational study
- Cross-sectional study
- Case-control study
- Cohort study
- Experimental research
- Randomized controlled trial



RECAP SEQUENCE

Types of epidemiological investigations

- [ ] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

Ecological studies evaluate relationships between exposures and outcomes. Those exposures can either be geographical or temporal (i.e., time-related).

Needs to include as many risk-modifying factors as possible to avoid the effect of unknown unknowns.

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

1980: Using the map as a predictor of cancer:

Int J Epidemiol. 1980 Sep;9(3):227-31.

Do sunlight and vitamin D reduce the likelihood of colon cancer?

Garland CF, Garland FC.

Abstract

It is proposed that vitamin D is a protective factor against colon cancer. This hypothesis arose from inspection of the geographic distribution of colon cancer deaths in the U.S., which revealed that colon cancer mortality rates were highest in places where populations were exposed to the least amounts of natural light--major cities, and rural areas in high latitudes. The hypothesis is supported by a comparison of colon cancer mortality rates in areas that vary in mean daily solar radiation penetrating the atmosphere. A mechanism involving cholecalciferol (vitamin D3) is suggested. The possibility that an ecological fallacy or other indirect association explains the findings is explored.

Sure, sunlight increases skin cancer (just ask a Floridian); but it may also reduce colon cancer. Are vitamin D pills anti-carcinogenic then?

RECAP SEQUENCE

Types of epidemiological investigations

- [ ] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

2006: An update to the map's effect:

LETTER TO THE EDITOR

Do sunlight and vitamin D reduce the likelihood of colon cancer? Time for a paradigm shift?

From PIET HEIN JONGBLOET

19 September 2006
doi:10.1093/ije/dyl206

Correlation studies between higher colon cancer rates and higher geographic latitude and, thus, less vitamin D have been extended to other cancer sites, including breast, bladder, corpus uteri, oesophageal, kidney, lung, ovary, pancreas, prostate, rectum, stomach, multiple myeloma, and non-Hodgkin lymphoma. And, if that was not enough, also to other diseases—including hypertension, type 1 diabetes mellitus, multiple sclerosis and osteoporosis—that are linked to vitamin D insufficiency. This hypothesis, therefore, has found adherence and has been tested in different ways. However, it is broadly accepted that there was no vitamin D deficiency in the US.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

2006: An update to the map's effect:

Metropolitan vs. non-metropolitan?
The sunlight exposure will be different.

Seasonal Preovulatory Overripeness Ovopathy:
During some seasons, the preovulatory phase of egg development is protracted, and that can increase the likelihood of congenital anomalies. Maybe. And maybe this increases the further away from the equator one is. And maybe the immune system anomalies are responsible for the elevated cancer rates.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

2006: An update to the map's effect:

This proposed causal relation of seasonally bound non-optimally matured oocytes to cancer, therefore, not only agrees with the geographical latitude effect under discussion but also with the disproportional month-of-birth deviations apparent in childhood leukaemia,[13–15] premenopausal breast cancer,[16–18] lung cancer,[19] and brain tumours in children[20,21] and adults.[22] These deviations in general correspond with the major total birth peak or major ovulatory season.[11,12] Further, it is in line with the enigmatic phenomenon of shorter life expectancy of people born in the first half of the year when compared with the second half, at least in Europe and upside down in Australia.[23]

RECAP SEQUENCE

Types of epidemiological investigations

- [ ] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

2006: An update to the map's effect:

> An analogous persistent paradigm regards the so-called Mediterranean diet allegedly protective for cardiovascular diseases.[24] Olive oil in the diet has been argued to be a similarly confounding or indirect factor, which is associated with geographical latitude. Seasonally bound ovopathy and its pathophysiological consequences related to geographical latitude again explains the presumed association and needs a paradigm shift. *'Not the olive is important, but where the olive grows'*. A similar paradigm shift related to vitamin D might be necessary. It would discard some inconsistencies encountered, such as fortification or addition of vitamin D in the nutrition still under debate. The epidemiological findings do not prove that higher levels of vitamin D would lower risk of colorectal cancer, while overexposure to UVA is harmful.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

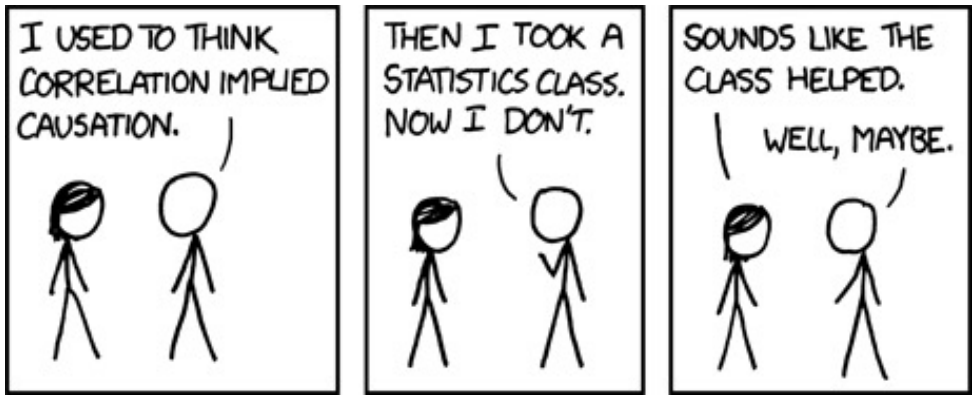
Correlational studies measure simple statistical relationships between variables. Do variables move together? If you attend class more, does that associate with a better grade? How strong is that relationship?

Strength of relationship is calculated as a correlation coefficient between -1 and +1. A correlation of zero means there's no relationship; the variables don't move together. A strong positive correlation means the variables move in the same direction (a greater squat max associates with a better vertical jump). A strong negative correlation means the variables move in opposite directions: The older you are, the lower your vertical jump. In general, anything less than 0.3 can be thought of as a bit weak.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study



* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [ ] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

PLOS ONE

Published: July 8, 2015

Mortality Attributable to Low Levels of Education in the United States

Patrick M. Krueger[1,2]*, Melanie K. Tran[1], Robert A. Hummer[3], Virginia W. Chang[4]

Variables

Our outcomes include all-cause mortality, cardiovascular disease mortality (ICD-10 codes I00 through I99) and cancer mortality (ICD-10 codes C00 through C97). We use broad cause-of-death categories to limit the impact of misclassification of underlying causes of death on death certificates. Age at baseline is measured in quarter-years (range 25.0 to 84.75). Birth cohort is also measured in quarter-years (range 1901.25 to 1980.5). Because we use 19 cross-sectional waves of the NHIS, follow vital status for up to 21 years, and allow age to increase throughout the follow-up period, the age and birth cohort variables are not perfectly collinear.

Educational attainment is coded categorically as less than a high school degree, high school degree or GED, some college but no baccalaureate degree, baccalaureate degree, and any post-baccalaureate education. Our variable captures key educational milestones. The NHIS ascertains educational attainment as years completed prior to 1997 and as highest degree completed in 1997 and later; we recode years of education completed in the 1986–1996 waves of data to be consistent with the highest degree measure used in later waves. This coding strategy has been used previously to estimate educational disparities among U.S. adults [6, 7]. Race/ethnicity is coded categorically as non-Hispanic white, non-Hispanic black, non-Hispanic Asian/Pacific Islander, non-Hispanic Native American, Mexican American, Puerto Rican, Cuban, and all others. Sex is measured dichotomously.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

These three are collectively referred to as **observational studies** and fall within the scope of **epidemiology**.

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

Cross-sectional studies evaluate prevalence, i.e., the number of cases of some condition at a given point in time. Some specific population (that whole population) at one point in time.

For example, look at people with a particular exposure (e.g., those who smoke) and calculate the proportion of that population who have a condition (e.g., COPD).

One point in time. Were you exposed? And do you have the condition?

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Cross-sectional study:

BMJ 1998;316:1643–6

Prevalence of serious eye disease and visual impairment in a north London population: population based, cross sectional study

A Reidy, D C Minassian, G Vafidis, J Joseph, S Farrow, J Wu, P Desai, A Connolly

Objective: To estimate the magnitude of serious eye disorders and of visual impairment in a defined elderly population of a typical metropolitan area in England, and to assess the frequency they were in touch with, or known to, the eye care services.
Design: Cross sectional survey using two stage cluster random sampling.
Setting: General practices in north London.
Subjects: Random sample of people aged 65 and older, drawn from a defined population of elderly people registered with 17 general practice groups.
Main outcome measures: Proportions and population prevalence estimates were determined for visual acuity, assessed with the person's own spectacles (if any), classified into four categories: prevalence of cataract, age related macular degeneration, and refractive error causing visual impairment and of definite primary open angle glaucoma; and status of contact with eye services.
Results: 1547 of 1840 (84%) eligible people were examined. The population prevalence of bilateral visual impairment (visual acuity $<6/12$) was 30%, of which 72% was potentially remediable. 92 of these 448 cases (21%) had visual acuity $<6/60$ ("blindness") in one or both eyes. Prevalence of cataract causing

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

Cross-sectional study:

BMJ 1998;316:1643–6

Prevalence of serious eye disease and visual impairment in a north London population: population based, cross sectional study

A Reidy, D C Minassian, G Vafidis, J Joseph, S Farrow, J Wu, P Desai, A Connolly

Conclusions: Untreated visual impairment and eye disorders affect a substantial proportion of people aged 65 years and older. These findings should contribute to the setting up of future strategies for preservation of sight and eye health services in general.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [ ] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Cross-sectional study you can conduct:

A cross-sectional study of your Pacific peers:

What percentage of students exercise at Baun?

What percentage of students have scholarships?

What percentage of students use tutoring?

Then go one step further: what predicts the use of tutoring? Is it GPA? Age? Gender? Language barriers? Declared major? What are the independent variables here?

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Cohort studies are generally prospective, but if you've already collected the data for one purpose, and then you come up with a new purpose (a different research question that can be answered using the same database), you can obviously analyze that same database retrospectively for your new question.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Cohort studies:

Step 1: Identify a population that does *not* have a disease (or whatever condition of interest).

Step 2: Identify two different samples within that population. Half of them are being exposed to something you think will cause the condition/disease and the other half of them do not have that exposure.

Step 3: Follow those samples and see if who gets the condition/disease. Is there a difference in prevalence between the two samples?

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

Cohort studies:

Step 4: Calculate relative risk of getting the disease/condition.

Step 5: Report your findings, but be careful in how cause-and-effecty you are when phrasing those findings.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Prospective cohort study:

BMJ 1994;309:901-11

Mortality in relation to smoking: 40 years' observations on male British doctors

Richard Doll, Richard Peto, Keith Wheatley, Richard Gray, Isabelle Sutherland

***Objective*—To assess the hazards associated with long term use of tobacco.**

***Design*—Prospective study of mortality in relation to smoking habits assessed in 1951 and again from time to time thereafter, with causes sought of deaths over 40 years (to 1991). Continuation of a study that was last reported after 20 years' follow up (1951-71).**

***Subjects*—34 439 British male doctors who replied to a postal questionnaire in 1951, of whom 10 000 had died during the first 20 years and another 10 000 have died during the second 20 years.**

***Results*—Excess mortality associated with smoking was about twice as extreme during the second half of the study as it had been during the first half. The death rate ratios during 1971-91 (comparing continuing cigarette smokers with life-long non-smokers) were approximately threefold at ages 45-64 and twofold at ages 65-84. The excess mortality was chiefly from diseases that can be**

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Prospective cohort study:

BMJ 1998;316:1631–5 **BMJ** VOLUME 316 30 MAY 1998

Adverse socioeconomic conditions in childhood and cause specific adult mortality: prospective observational study

George Davey Smith, Carole Hart, David Blane, David Hole

Objective: To investigate the association between social circumstances in childhood and mortality from various causes of death in adulthood.
Design: Prospective observational study.
Setting: 27 workplaces in the west of Scotland.
Subjects: 5645 men aged 35-64 years at the time of examination.
Main outcome measures: Death from various causes.
Results: Men whose fathers had manual occupations when they were children were more likely as adults to have manual jobs and be living in deprived areas. Gradients in mortality from coronary heart disease, stroke, lung cancer, stomach cancer, and respiratory disease were seen (all $P < 0.05$), generally increasing from men whose fathers had professional and managerial occupations (social class I and II) to those

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials
(and their less-controlled, less random companions)

Retrospective cohort study:

J Accid Emerg Med 1997;**14**:371–374

How do individuals with diabetes use the accident and emergency department?

Elizabeth C Goyder, Stephen W Goodacre, Johannes L Botha, Gautam G Bodiwala

***Objective*—To determine whether the frequency and pattern of use of the accident and emergency (A&E) department by individuals with diabetes is different from that of the general population.**
***Methods*—A historical cohort of 696 individuals with diabetes from six randomly selected general practices and a non-diabetic comparison cohort matched on age, sex, and general practice were identified. The use of an urban A&E department by the two cohorts was compared for number of visits between 1984 and 1996 for injuries, diabetes related and non-diabetes related illness, proportion referred by a general practitioner, proportion arriving by ambulance, and proportion admitted.**
***Results*—More visits were made by the diabetic cohort (1002 *v* 706, P = 0.0001); 121 visits were directly related to diabetes, including 52 for hypoglycaemia. The diabetic cohort also had more visits for medical illness unrelated to diabetes (35**

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [ ] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Cohort study you can conduct:

Dependent variable: Admission to graduate school (odds ratio; how likely is it?).

Population: Students at Pacific.

Primary independent variable: (main exposure that you think will affect the odds of admission): Whether students have athletic scholarships.

Other confounding variables to include: Age, gender, parental education, participation in extracurricular activities, declared major, etc.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [x] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Case-control studies are always retrospective. You select a sample based on the presence or absence of a disease. Then look at a specific exposure as a predictor of the current state of the samples.

You'll evaluate a minority of the population (in a cross-sectional study, it's one point in time, looking at the whole population).

Then you describe absolute and relative risk via prevalences of stuff. For example, what's the prevalence of brain damage toady based on sport participation a while ago?

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [x] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Case-control study:

Lancet 2006; 368: 647–58

Tobacco use and risk of myocardial infarction in 52 countries in the INTERHEART study: a case-control study

Koon K Teo, Stephanie Ounpuu, Steven Hawken, MR Pandey, Vicent Valentin, David Hunt, Rafael Diaz, Wafa Rashed, Rosario Freeman, Lixin Jiang, Xiaofei Zhang, Salim Yusuf, on behalf of the INTERHEART Study Investigators

Background Tobacco use is one of the major avoidable causes of cardiovascular diseases. We aimed to assess the risks associated with tobacco use (both smoking and non-smoking) and second hand tobacco smoke (SHS) worldwide.

Methods We did a standardised case-control study of acute myocardial infarction (AMI) with 27 089 participants in 52 countries (12 461 cases, 14 637 controls). We assessed relation between risk of AMI and current or former smoking, type of tobacco, amount smoked, effect of smokeless tobacco, and exposure to SHS. We controlled for confounders such as differences in lifestyles between smokers and non-smokers.

Findings Current smoking was associated with a greater risk of non-fatal AMI (odds ratio [OR] 2·95, 95% CI 2·77–3·14, p<0·0001) compared with never smoking; risk increased by 5·6% for every additional cigarette smoked. The OR associated with former smoking fell to 1·87 (95% CI 1·55–2·24) within 3 years of quitting. A residual excess risk remained 20 or more years after quitting (1·22, 1·09–1·37). Exclusion of individuals exposed to SHS in the never smoker reference group raised the risk in former smokers by about 10%. Smoking beedies alone (indigenous to South Asia) was associated with increased risk (2·89, 2·11–3·96) similar to that associated with cigarette smoking. Chewing tobacco alone was associated with OR 2·23 (1·41–3·52), and smokers who also chewed tobacco had the highest increase in risk (4·09, 2·98–5·61). SHS was associated with a graded increase in risk related to exposure; OR was 1·24 (1·17–1·32) in individuals who were least exposed (1–7 h per week) and 1·62 (1·45–1·81) in people who were most exposed (>21 h per week). Young male current smokers had the highest population attributable risk (58·3%; 95% CI 55·0–61·6) and older women the lowest (6·2%, 4·1–9·2). Population attributable risk for exposure to SHS for more than 1 h per week in never smokers was 15·4% (12·1–19·3).

Conclusion Tobacco use is one of the most important causes of AMI globally, especially in men. All forms of tobacco use, including different types of smoking and chewing tobacco and inhalation of SHS, should be discouraged to prevent cardiovascular diseases.

RECAP SEQUENCE

Types of epidemiological investigations

- [x] Ecological Study
- [x] Correlational Study
- [x] Cross-Sectional Study
- [x] Cohort Study
- [ ] Case-Control Study

* Non-epidemiological work: randomized, controlled trials (and their less-controlled, less random companions)

Case-control study you can conduct:

Dependent variable: Graduation.

Population: Pacific students after graduation (some graduated; some failed to do so).

Primary independent variable: (main exposure that you think will affect graduation rate): Whether students live on campus.

Other confounding variables to include: Age, gender, parental education, participation in athletics, declared major, etc.

RECAP SEQUENCE

Difference between a cohort study and a case-control study

Retrospective study that selects subjects based on their disease status.

	Cases	Controls
Exposed	A	B
Unexposed	C	D

Sick people ("disease positive"): case group.

Not sick people ("disease negative"): control group.

They should all (ideally) come from the same population. And the control group should have a similar distribution of exposure as the case group.

RECAP SEQUENCE

Difference between a cohort study and a case-control study

Retrospective study that selects subjects based on their disease status.

	Cases	Controls
Exposed	A	B
Unexposed	C	D

Sick people ("disease positive"): case group.

Not sick people ("disease negative"): control group.

Compare A, B, C, and D (cases vs. controls; exposed vs. not exposed).

Generate an odds ratio:

Ratio of odds of exposure in cases (A ÷ C) to odds of exposure in controls (B ÷ D), i.e., (A × D) ÷ (B × C).

RECAP SEQUENCE

Difference between a cohort study and a case-control study

If the odds ratio is > 1:
Then the infected people are more likely to have been exposed.

If the odds ratio is roughly 1ish:
Then the disease (condition/infection/whatever) isn't correlated with the exposure.

If the odds ratio is < 1:
Then the exposure seems to have a protective effect.

RECAP SEQUENCE

Difference between a cohort study and a case-control study

Odds ratio of 1.1:

.....	Cases	Controls
Exposed	1732	1652
Unexposed	1652	1732

Odds ratio of 1.5:

.....	Cases	Controls
Exposed	103	84
Unexposed	84	103

RECAP SEQUENCE

Difference between a cohort study and a case-control study

Subjects are selected based on exposure status.

At the beginning of the study, all subjects *should* be disease free and at risk of acquiring whatever disease.

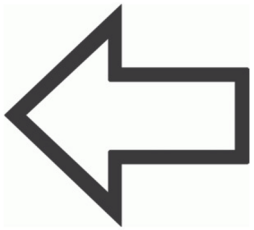
Follow those subjects over time and reassess them later.

RECAP SEQUENCE

Difference between a cohort study and a case-control study

For example: follow smokers and non-smokers over time.
Test lung cancer incidence after that period of time.

.....	Case	Non-case	Total
Exposed	*A*	*B*	(*A* + *B*)
Unexposed	*C*	*D*	(*C* + *D*)



Similar 2 × 2 table
is constructed

RECAP SEQUENCE

Difference between a cohort study and a case-control study

But you don't calculate an odds ratio; you estimate **relative risk.**

Relative risk: probability of disease in the exposed group (A ÷ (A + B)) over the disease over the probability of disease in the unexposed group (C ÷ (C + D)).

Or this:

$$\frac{\text{Probability of disease in exposed group}}{\text{Probability of disease in unexposed group}}$$

RECAP SEQUENCE

EPI (UPON)

DEMOS (PEOPLE)

LOGOS (STUDY)

EPIDEMIOLOGY: THE STUDY OF WHAT IS UPON PEOPLE.

Primary purpose of epidemiological research:

Collect information about populations that can be used to prevent, control, or treat various health problems.

R E C A P S E Q U E N C E

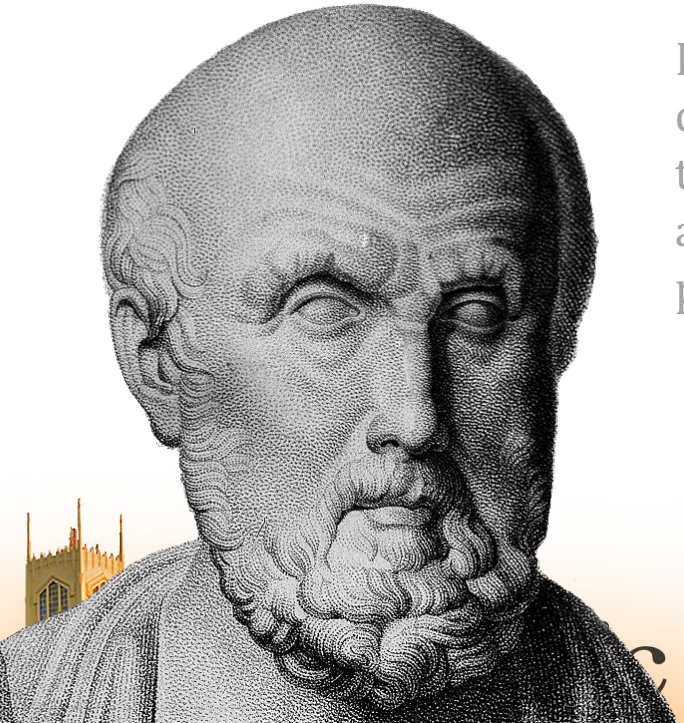
HIPPOCRATES

Hippocrates' understanding of medicine was given up, but his principles of understanding remained.

Epidemiological terms coined by Hippocrates:

Endemic: diseases generally found in certain locations, but not others.

Epidemic: Diseases generally seen at certain times, but not others.

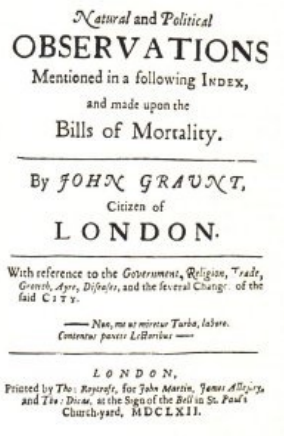


Hippocrates was the first to distinguish between the diseases that *visit* populations (epidemic) and those that reside *within* populations (endemic).

RECAP SEQUENCE

Eventually, *political arithmetic* becomes a thing.

Late-Renaissance effort by fancy wig-wearing Europeans to characterize demographic and economic data. Mostly a tool to estimate the size and growth of their populations (as well as the life expectancies of those being summed). The increasingly complex analyses of these data, aimed at the creation of a healthier state, were coined "statistics."



Natural and *Political*

OBSERVATIONS

Mentioned in a following INDEX,

and made upon the

Bills of Mortality.

By JOHN GRAUNT,

Citizen of

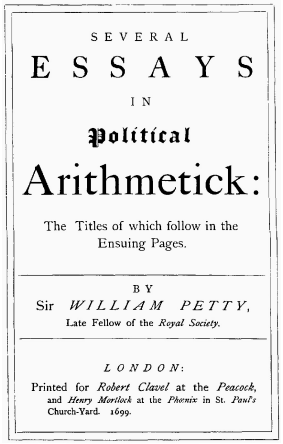
LONDON.

With reference to the *Government*, *Religion*, *Trade*, *Growth*, *Ayre*, *Diseases*, and the several Changes of the said CITY.

—— *Nec, me ut mirentur Turba, laboro,*
Contentus paucis Lectoribus ——

LONDON,

Printed by *Tho: Roycroft*, for *John Martin*, *James Allestry*, and *Tho: Dicas*, at the Sign of the *Bell* in St. *Paul's* Church-yard, MDCLXII.



SEVERAL

ESSAYS

IN

Political

Arithmetick:

The Titles of which follow in the Ensuing Pages.

BY

Sir *WILLIAM PETTY*,

Late Fellow of the *Royal Society*.

LONDON:

Printed for *Robert Clavel* at the *Peacock*, and *Henry Mortlock* at the *Phenix* in St. *Paul's* Church-Yard. 1699.

RECAP SEQUENCE

John Graunt (1620 – 1674) – British Haberdasher.

Also one of the first demographers.

Graunt made the first life table (likelihood that you'll die based on your age). Probability of survival if you're 70, probability of survival if you're 76, etc.

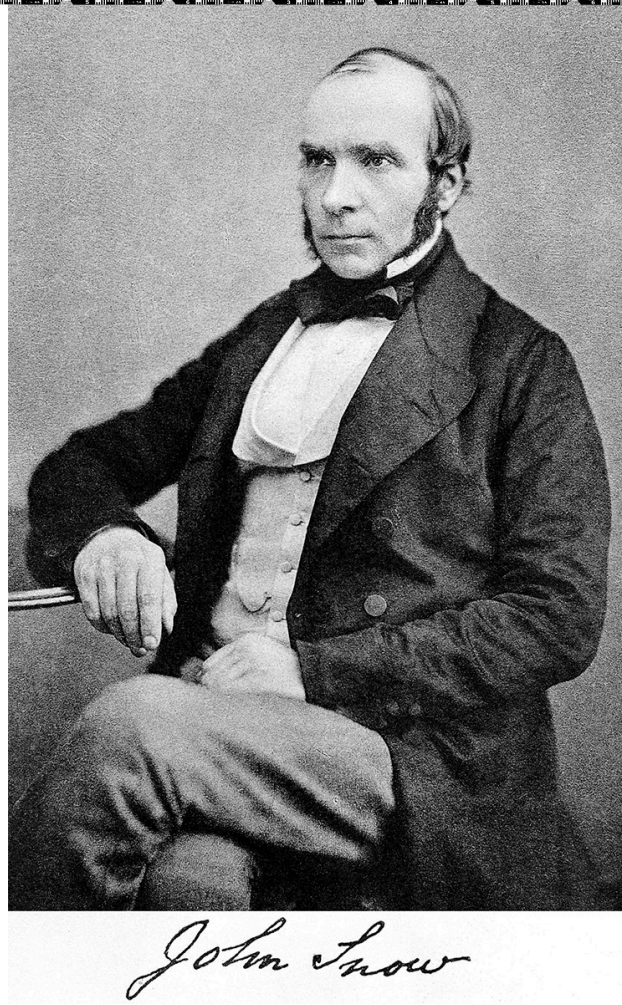
Table 1. Graunt's Life Table.

Age Interval	Prop. Deaths in Interval	Prop. Surviving til start of Interval
0-6	0.36	1.00
7-16	0.24	0.64
17-26	0.15	0.40
27-36	0.09	0.25
37-46	0.06	0.16
47-56	0.04	0.10
57-66	0.03	0.06
67-76	0.02	0.03
77-86	0.01	0.01

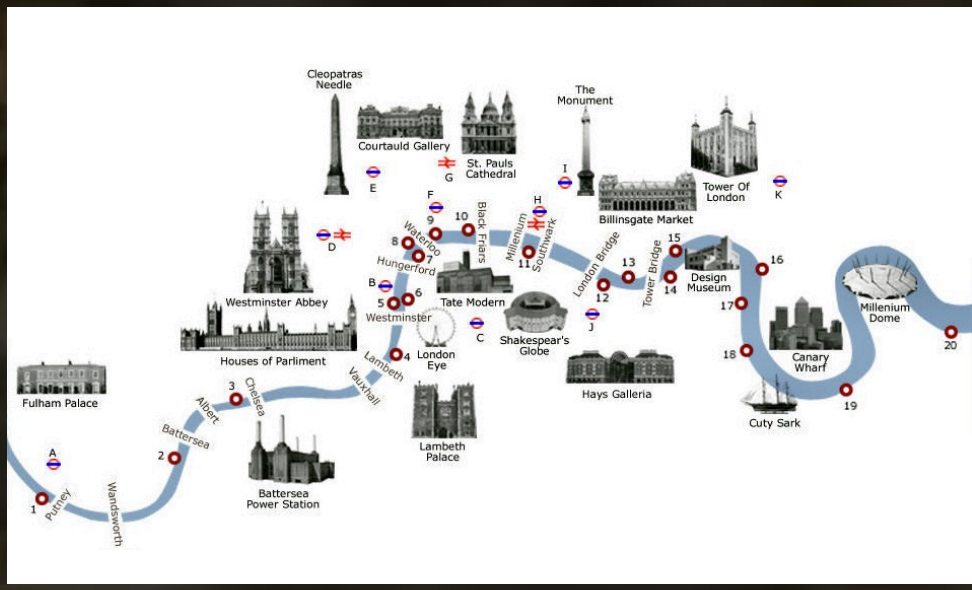
RECAP SEQUENCE

On August 31, 1854, lots of Londoners began to get cholera: the "1854 Broad Street Cholera Outbreak".

During the first three days, 127 people on or near Broad Street died. By September 10th, there were 500 deaths (after ¾ of residents fled the area). Mortality rate was >12% in parts of the city. Final death count: 616–668.

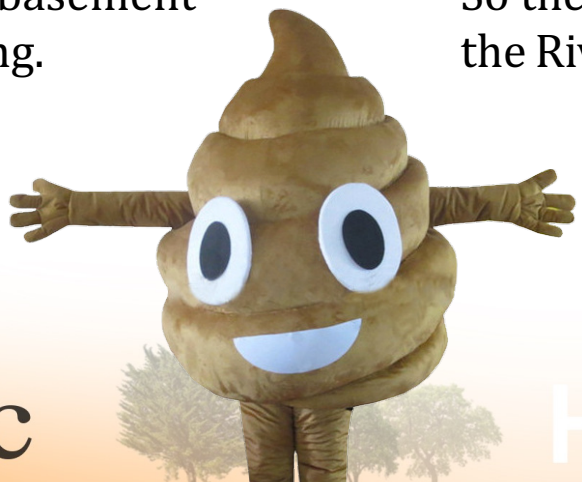


RECAP SEQUENCE



Cesspools beneath people's basement floorboards were overflowing.

So they dumped the overflow into the River Thames.



RECAP SEQUENCE

People started dying.

RECAP SEQUENCE

"From inhaling the odour of beef the butcher's wife obtains her obesity."
Professor H Booth, writing in the *Builder*, July 1844

This assertion is perhaps the most extravagant manifestation of a belief that prevailed in the medical profession for much of the 19th century and survived in some quarters into the 20th century. This belief held that most, if not all, disease was caused by inhaling air that was infected through exposure to corrupting matter. Such matter might be rotting corpses, the exhalations of other people already infected, sewage, or even rotting vegetation. The "miasmatic" explanation of the cause of disease figured prominently in the long debates among the people who were responsible for combating the cholera epidemics that afflicted Britain, and particularly London, between 1831 and 1866.

BMJ VOLUME 323 22–29 DECEMBER 2001

Death and miasma in Victorian London: an obstinate belief

Stephen Halliday

It must be the stench that's killing them…

RECAP SEQUENCE

All smell is, if it be intense, immediate acute disease; and eventually we may say that, by depressing the system and rendering it susceptible to the action of other causes, all smell is disease.

Metropolitan Sewage Committee proceedings. *Parliamentary Papers*. (1846). 10: 651.

It must be the stench that's killing them…

RECAP SEQUENCE

John Snow observed that cholera struck especially hard around a water pump on Broad Street. He persuaded the local parish to remove the pump handle. And he *suggested* that drinking water should be piped in from regions waterways that were free from pollution.

RECAP SEQUENCE

Snow J. (1857). Cholera and the water supply in the southern districts of London. *British Medical Journal* 2: 864

1857: John Snow published the first modern epidemiological paper. He compared cholera-related mortality rates between customers of two water companies:

1. Southwark Water Company (water came from the most polluted region of the River Thames).

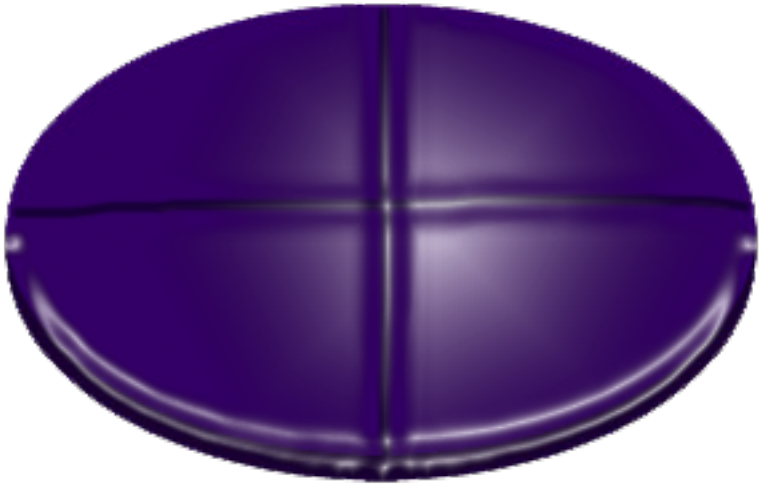
2. Lambeth Water Company (water came from Thames Ditton, where there was no sewage in the tideway).

Cholera-related mortality among Southwark customers was six times higher.

He was correct. So what obviously happened?

RECAP SEQUENCE

Things John Snow knows:



- Nothing

RECAP SEQUENCE

Epidemiology after Snow

People see
cause and effect
in everything.

RECAP SEQUENCE

Epidemiology after Snow

People see
cause and effect
in everything.

RECAP SEQUENCE

Epidemiology after Snow

People see
cause and effect
in everything.

RECAP SEQUENCE

We're great at pattern reception and we apply it to everything.

Michael Shermer:
"Humans are pattern-seeking, story-telling animals, and we are quite adept at telling stories about patterns, whether they exist or not."

RECAP SEQUENCE

CAPTCHA

Correlation just finds *that* there's a relationship, not the nature of it. Computers are bad at this. Remember the captcha.

RECAP SEQUENCE

People are bad too (in terms of accuracy), but we do it anyway.

My sister said she'd pick me up at noon. She just got here. The movie starts in about three minutes. I'm furious.

It's easy for us to figure out why she's pissed. Not easy for a computer (pissed because the movie starts in three minutes?).

RECAP SEQUENCE

I seek patterns and meanings in the patternless and meaningless!

That's called apophenia!

Another name for it is patternicity!

RECAP SEQUENCE

A human brain seeks causal relationships.

We seek (and find, however falsely) a cause where none exists.

1. You went out to eat.
2. You got sick the next day.
3. It must have been the food.

1. You went outside in the cold rain and didn't bundle up well enough.
2. You got the flu.

(That's not how the flu is transmitted, but people connect the two anyway.)

Incorrect causal intuitions. That's what you want to avoid.

RECAP SEQUENCE Epidemiology:

1. Take a population.

2. Observe them (their diets, habits, behaviors, etc.).

3. Notice patterns related to health and disease of that population.

4. Don't change anything; just document all the things you notice.

5. Make conjectures / surmise causes.

6. Publish wrong stuff.

Recap Sequence: Epidemiology:

Major epidemiological domains:

- Origin/outbreak of disease
- Disease transmission
- Screening and biomonitoring
- Comparisons of treatment effects

RECAP SEQUENCE Epidemiology:

Overarching purpose:

- Who?
- What?
- Where?
- When?

} **of health-related events.**

RECAP SEQUENCE

Traditional and Modern Epidemiology

. . . a shift in the level of analysis from the population to the individual. Most modern epidemiologists still do studies in populations, but they do so in order to study decontextualized individual risk factors, rather than to study population factors in their social and historical context.

American Journal of Public Health
May 1996, Vol. 86, No. 5

Traditional Epidemiology, Modern Epidemiology, and Public Health

Neil Pearce, PhD

R E C A P S E Q U E N C E

Post-Snow epidemiology was an attempt to apply science to the world around us:

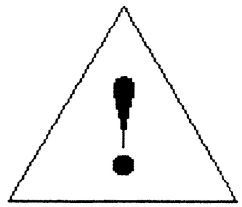
American Journal of Public Health March 1999, Vol. 89, No. 3

Cardiovascular Disease and Risk Factor Epidemiology: A Look Back at the Epidemic of the 20th Century

F. Javier Nieto
Department of Epidemiology
Johns Hopkins University
School of Public Health
Baltimore, Md

RECAP SEQUENCE

Post-Snow epidemiology was an attempt to apply science to the world around us:



GOVERNMENT WARNING: (1) According to the Surgeon General, women should not drink alcoholic beverages during pregnancy because of the risk of birth defects. (2) Consumption of alcoholic beverages impairs your ability to drive a car or operate machinery, and may cause health problems.

RECAP SEQUENCE

Epidemiology continues to have its place:
It tries to make sense of epidemics.

Proposed in California and New York:

STATE OF CALIFORNIA SAFETY WARNING: Drinking beverages with added sugar(s) contributes to obesity, diabetes, and tooth decay.

npr

Is It Time For A Warning Label On Sugar-Loaded Drinks?

PONCIE RUTSCH Published April 9, 2015 · 3:52 PM ET

RECAP SEQUENCE

Breast Cancer
Vol. 5 No. 4 October 1998

Risk Factors for Breast Cancer among Japanese Women: A Case-Control Study in Ibaraki, Japan

Masaru Ueji*[1], Ei Ueno*[2], Douglas Osei-Hyiaman*[1], Tomoko Saito*[1], Hideto Takahashi*[3], and Katsumi Kano*[3]

Results: After adjustment for potential confounders, heavy weight and higher body mass index were associated with an increased risk of breast cancer among postmenopausal women (OR=1.76, 95% CI=0.69, 4.48; OR=1.57, 95% CI=0.61, 3.99, respectively). Current or ex-smokers were found to be at an increased risk for breast cancer (OR=3.33; 95% CI=1.63, 6.80). Women who take hot baths had a decreased risk for breast cancer (OR=0.67; 95% CI=0.43, 1.06). Recreational physical activity was associated with a reduced risk of breast cancer (P_{Trend}=0.005). OR for breast cancer among physically active women was 0.36 (95% CI=0.19, 0.70), as compared with inactive women. Taller women had an increased risk of breast cancer relative to shorter women (OR=1.49; 95% CI=0.83, 2.70). No significant association between alcohol consumption and breast cancer risk was detected.

Conclusions: Our results suggest that several potentially modifiable lifestyle factors may be useful for the prevention of breast cancer.

RECAP SEQUENCE

An IRB is unlikely to approve a study in which participants are randomly assigned to one of two groups:

Group A: Smoke-free control

Group B: Two packs per day.

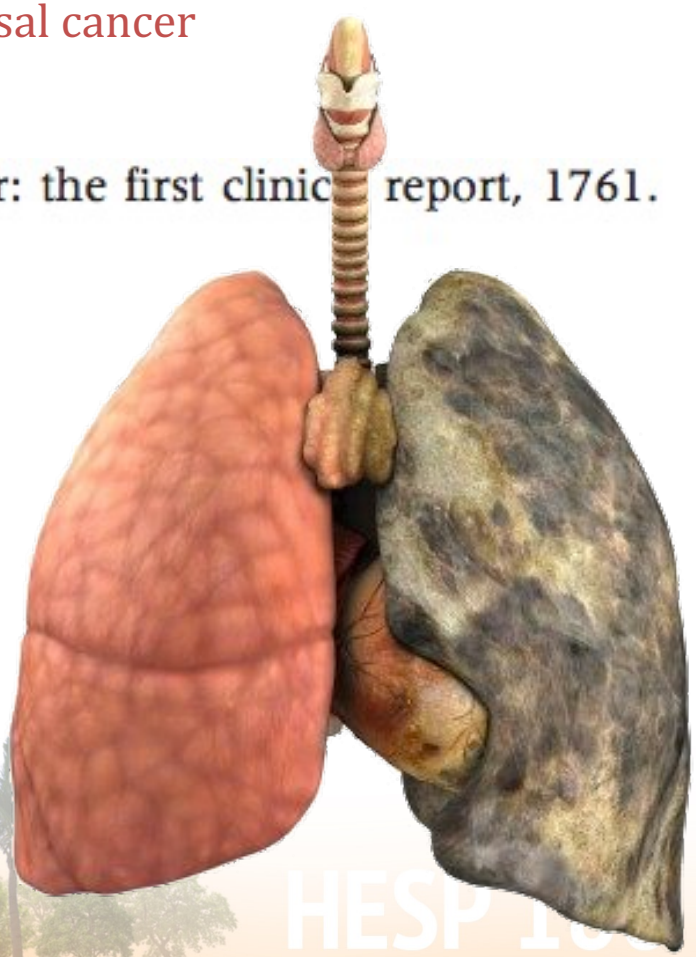
So epidemiological principles are used.

R E C A P S E Q U E N C E

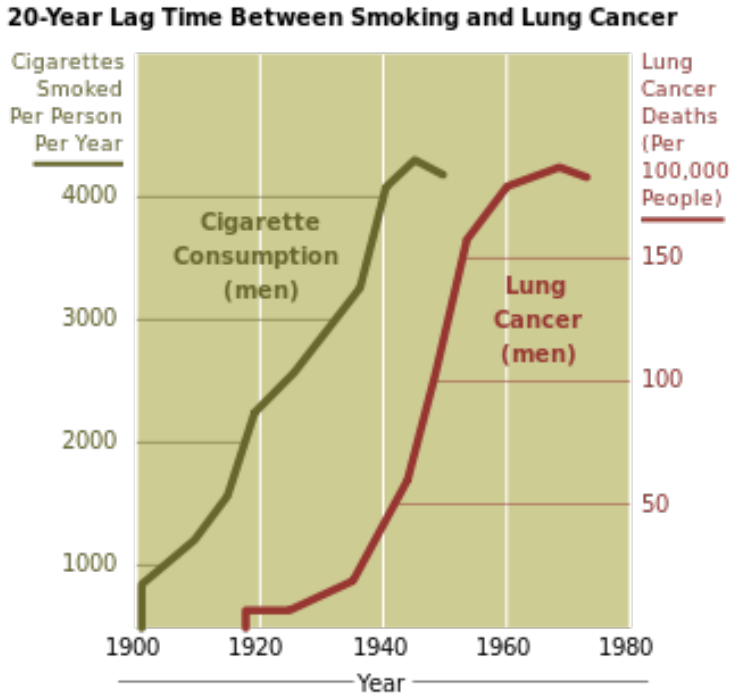
Smoking and cancer.

The relationship between snuff and nasal cancer was identified as early as 1761:

Redmond DE JR. Tobacco and cancer: the first clinic report, 1761. *N Engl J Medicine* 1970;**282**:18–23.



RECAP SEQUENCE



Fritz Lickint evaluated time-trend data, which implicated smoking as a cause of lung cancer.

Lickint F. Der Bronchialkrebs der Raucher. *Munch Med Wschr* 1935; **82**:122–24.

RECAP SEQUENCE

Findings supported by case-control studies in the US and UK:

Doll R, Hill AB. Smoking and carcinoma of the lung. Preliminary report. *Br Med J* 1950;**2:**739–48.

Levin ML, Goldstein H, Gerhardt PR. Cancer and tobacco smoking. *Journal of the American Medical Association* 1950;**143:**336–38.

Mills CA, Porter MM. Tobacco smoking habits and cancer of the mouth and respiratory system. *Cancer Research* 1950;**10:**539–42.

Schrek R, Baker LA, Ballard GP, Dolgoff S. Tobacco smoking as an etiologic factor in disease. *Cancer Research* 1950;**10:**49–58.

Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma. *Journal of the American Medical Association* 1950;**143:**329–36.

RECAP SEQUENCE

This is where we wind up:

SURGEON GENERAL'S WARNING: Smoking Causes Lung Cancer, Heart Disease, Emphysema, And May Complicate Pregnancy.

RECAP SEQUENCE

But we have to be careful:

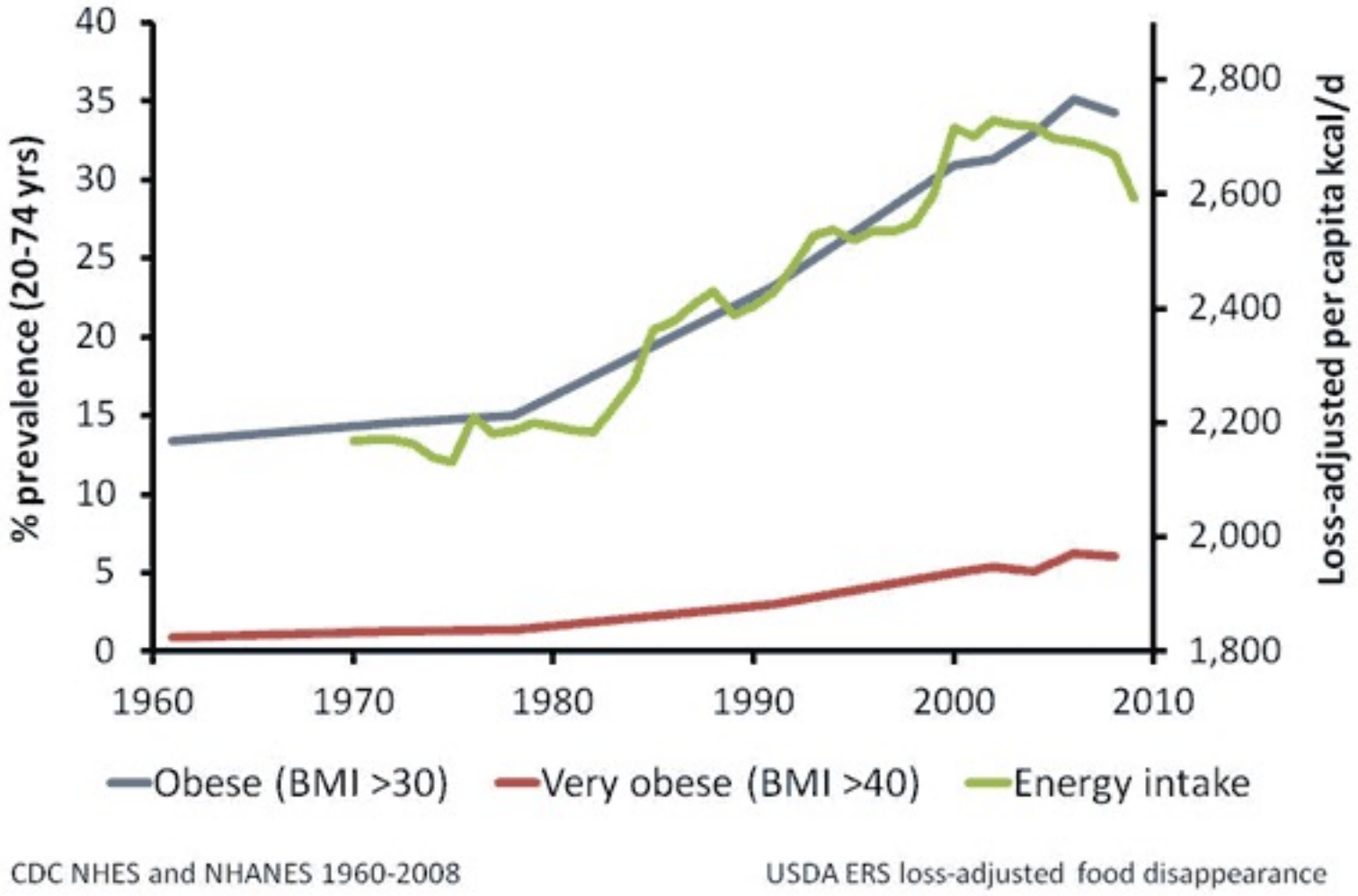
RECAP SEQUENCE

>4000 possible interactions

How many random correlations would you expect

R E C A P S E Q U E N C E

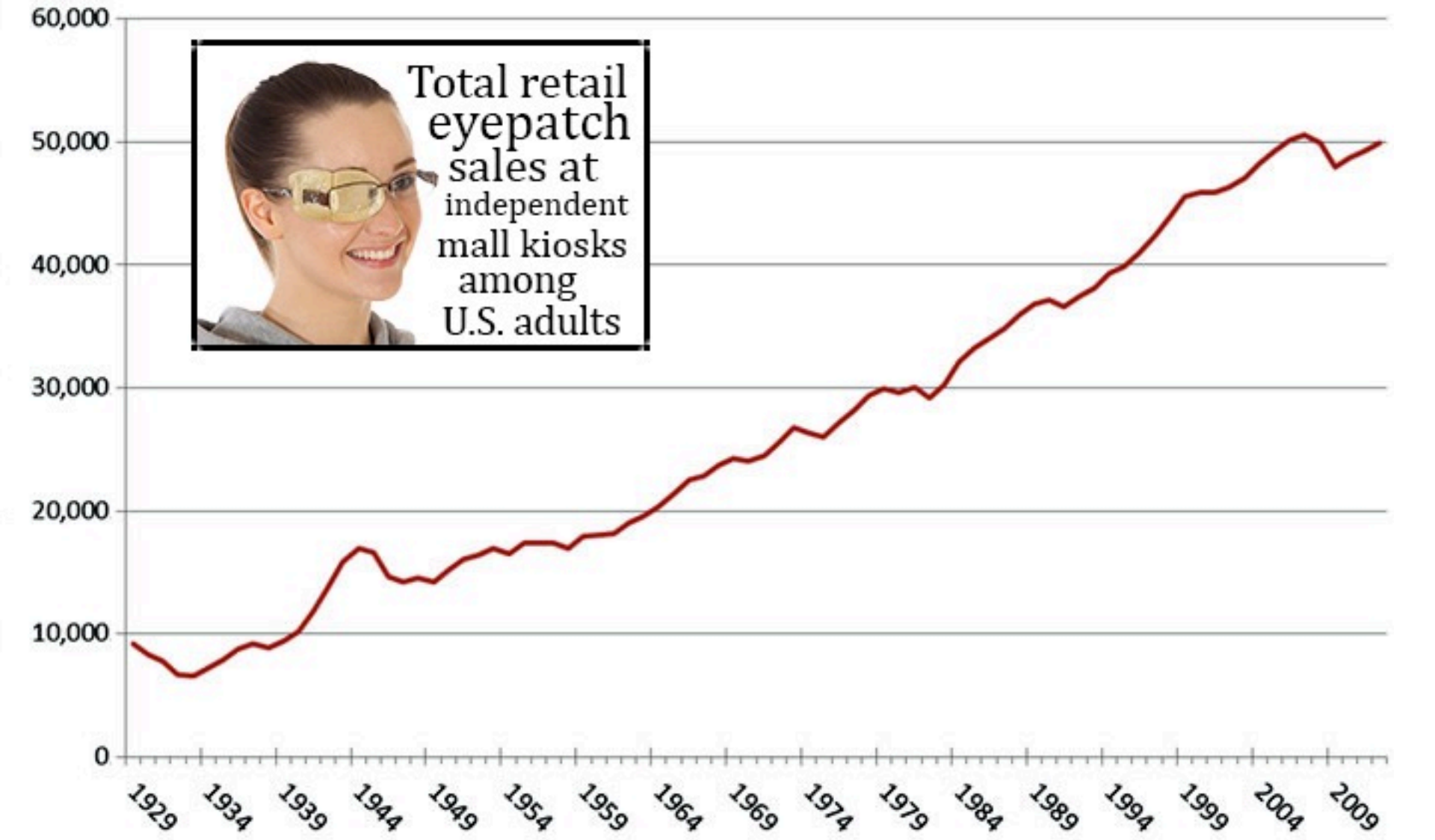
Obesity and Energy Intake in the US, 1961-2009



CDC NHES and NHANES 1960-2008

USDA ERS loss-adjusted food disappearance

RECAP SEQUENCE

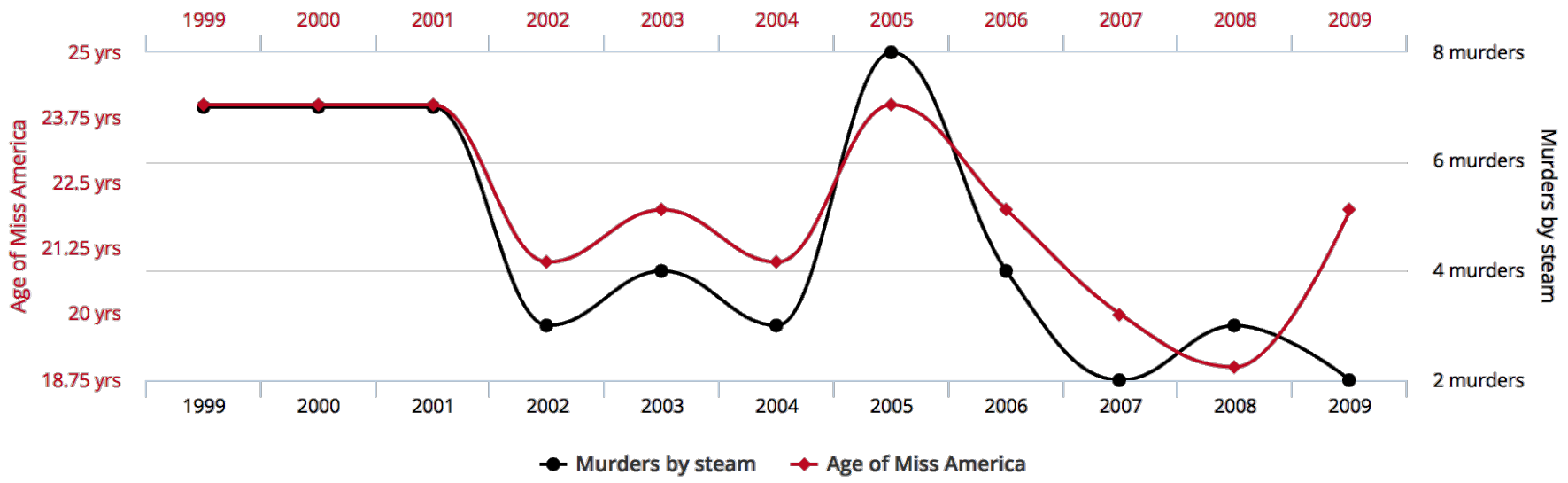


RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects

Correlation: 87.01% (r=0.870127)



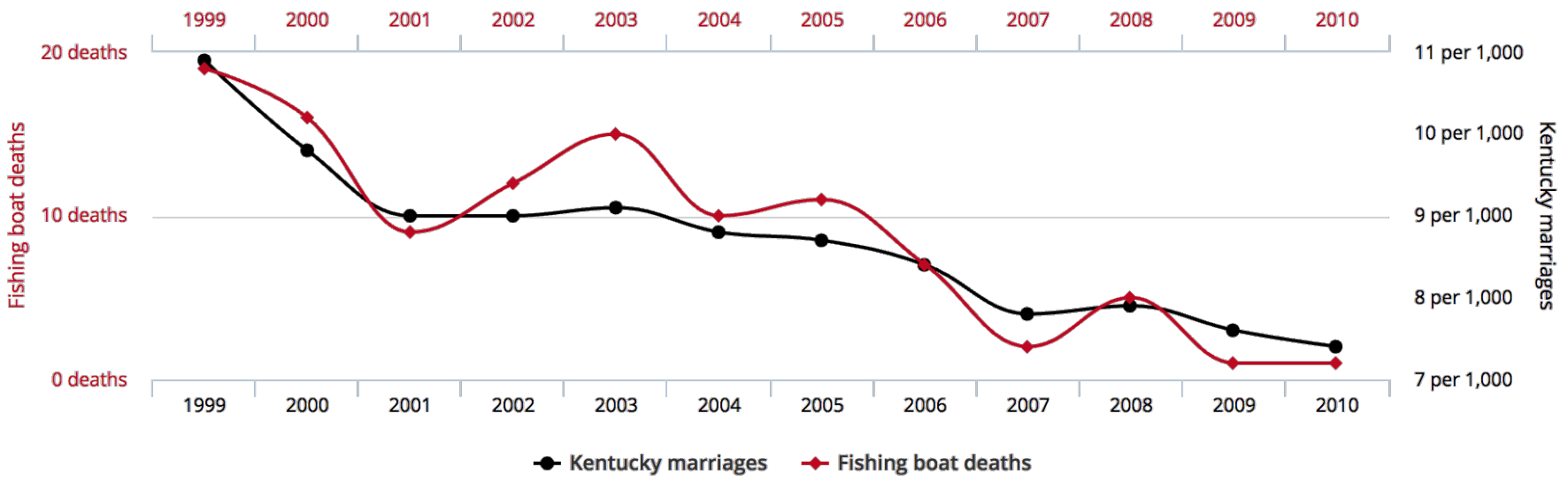
Data sources: Wikipedia and Centers for Disease Control & Prevention

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

People who drowned after falling out of a fishing boat
correlates with
Marriage rate in Kentucky

Correlation: 95.24% (r=0.952407)



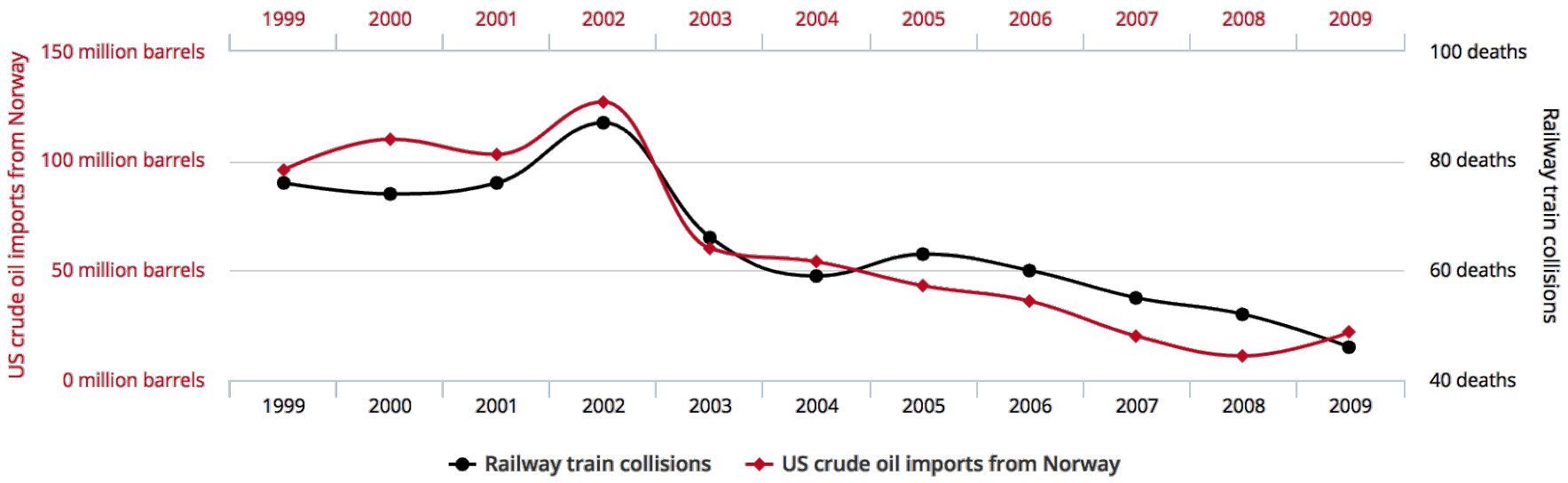
Data sources: Centers for Disease Control & Prevention and National Vital Statistics Reports

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

US crude oil imports from Norway
correlates with
Drivers killed in collision with railway train

Correlation: 95.45% (r=0.954509)



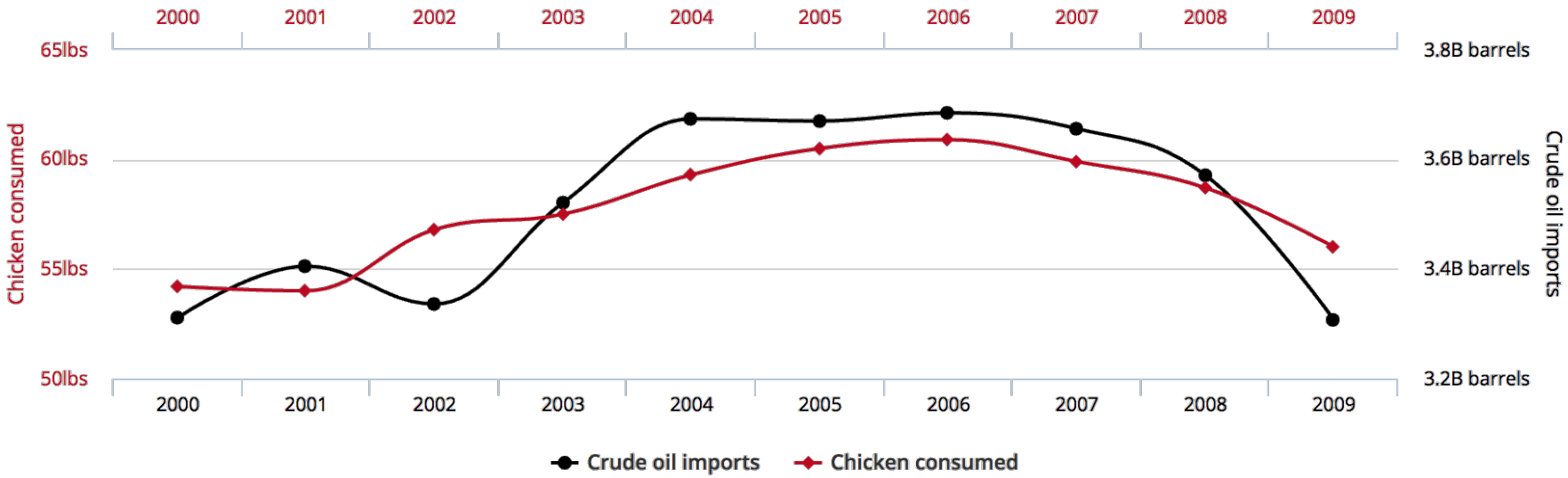
Data sources: Dept. of Energy and Centers for Disease Control & Prevention

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

Per capita consumption of chicken
correlates with
Total US crude oil imports

Correlation: 89.99% (r=0.899899)



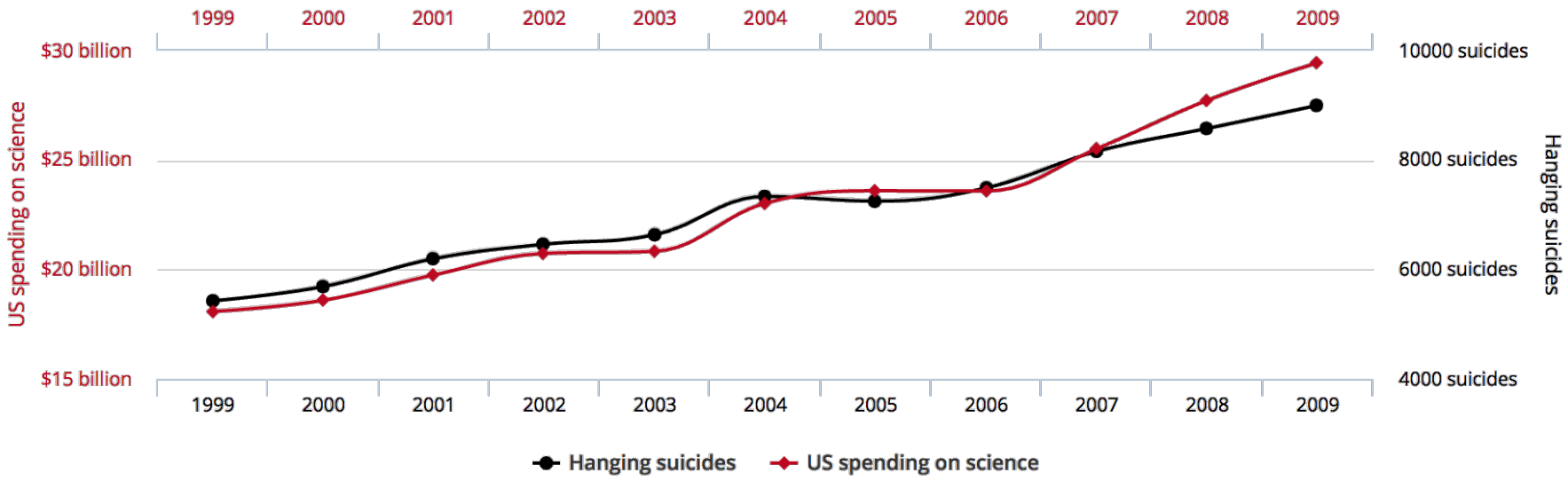
Data sources: U.S. Department of Agriculture and Dept. of Energy

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation

Correlation: 99.79% (r=0.99789126)



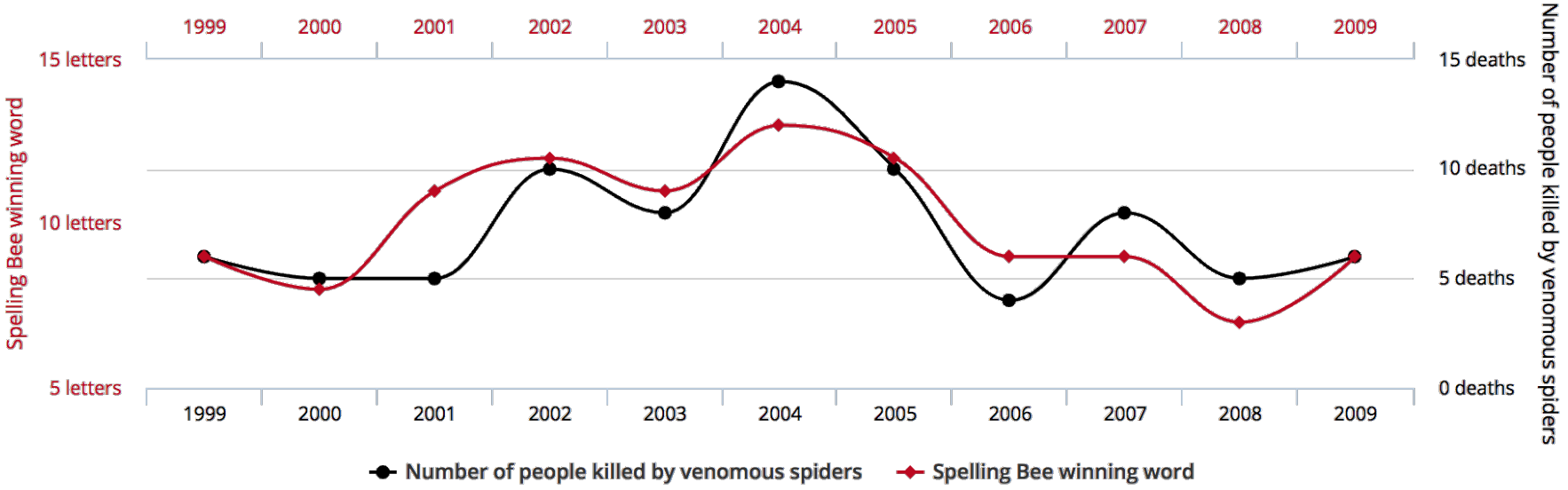
Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

Letters in Winning Word of Scripps National Spelling Bee
correlates with
Number of people killed by venomous spiders

Correlation: 80.57% (r=0.8057)



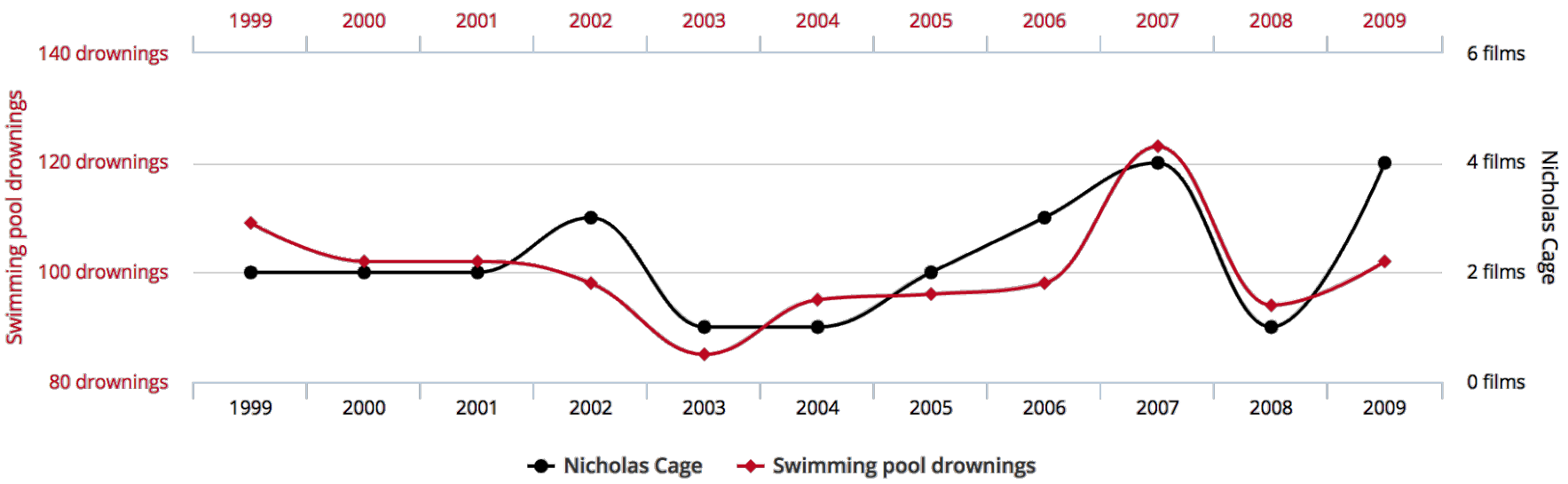
Data sources: National Spelling Bee and Centers for Disease Control & Prevention

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in

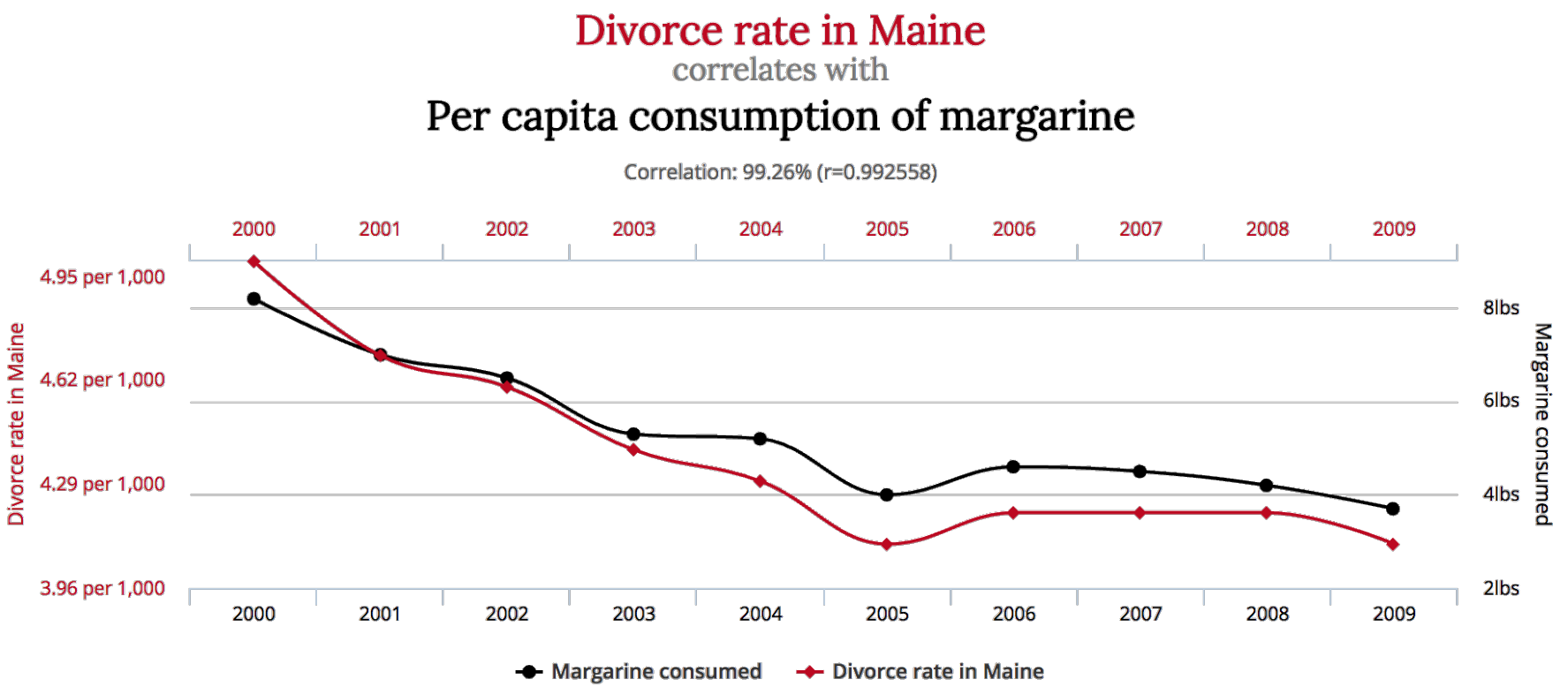
Correlation: 66.6% (r=0.666004)



Data sources: Centers for Disease Control & Prevention and Internet Movie Database

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations



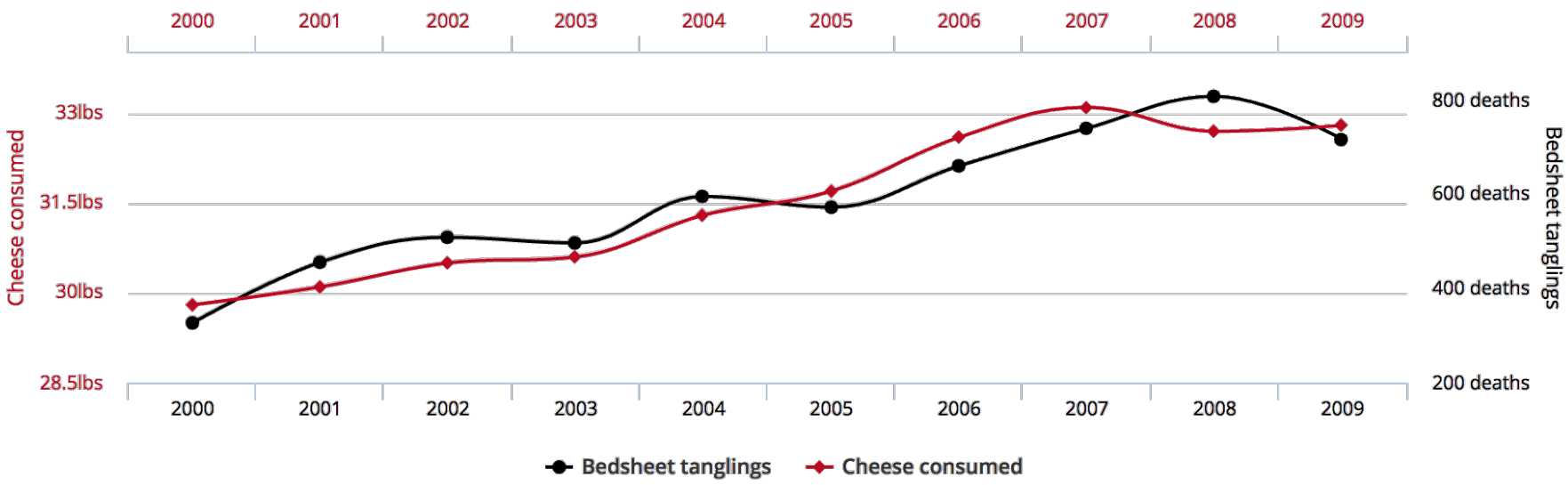
Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

RECAP SEQUENCE

www.tylervigen.com/spurious-correlations

Per capita cheese consumption
correlates with
Number of people who died by becoming tangled in their bedsheets

Correlation: 94.71% (r=0.947091)



Data sources: U.S. Department of Agriculture and Centers for Disease Control & Prevention

RECAP SEQUENCE

One view of the value of epidemiology

RECAP SEQUENCE

RECAP SEQUENCE

A big clue:

Most comparisons in "meat studies" compare the bottom quintile (vegetarians) to the top quintile (serious carnivores).

As you move from the bottom to the top, what else do you see?

Bad Box

More smoking, drinking, more sedentary lifestyles, and pretty much every other bad behavior, in addition to high BMI, high BP, etc.

Good Box

Vegetarians are more concerned with their health in a number of ways. More likely to visit the doctor, do yoga, stay hydrated, get more sleep; they probably brush their teeth more.

RECAP SEQUENCE

Table 1. We have found 12 papers in which claims coming from observational studies were tested in randomised clinical trials. Many of the trials are quite large. In most of the observational studies multiple claims were tested, often in factorial designs, e.g. vitamin D and calcium individually and together along with a placebo group. Note that none of the claims replicated in the direction claimed in the observational studies and that there was statistical significance in the opposite direction five times

ID no.	*Pos.*	*Neg.*	*No. of claims*	*Treatment(s)*	*Reference*
1	0	1	3	Vit E, beta-carotene	*NEJM* 1994; **330**: 1029–1035
2	0	3	4	Hormone Replacement Ther.	*JAMA* 2003; **289**: 2651–2662, 2663–2672, 2673–2684
3	0	1	2	Vit E, beta-carotene	*JNCI* 2005; **97**: 481–488
4	0	0	3	Vit E	*JAMA* 2005; **293**: 1338–1347
5	0	0	3	Low Fat	*JAMA*. 2006; **295**: 655–666
6	0	0	3	Vit D, Calcium	*NEJM* 2006; **354**: 669–683
7	0	0	2	Folic acid, Vit B6, B12	*NEJM* 2006; **354**: 2764–2772
8	0	0	2	Low Fat	*JAMA* 2007; **298**: 289–298
9	0	0	12	Vit C, Vit E, beta-carotene	*Arch Intern Med* 2007; **167**: 1610–1618
10	0	0	12	Vit C, Vit E	*JAMA* 2008; **300**: 2123–2133
11	0	0	3	Vit E, Selenium	*JAMA* 2009; **301**: 39–51
12	0	0	3	HRT + Vitamins	*JAMA* 2002; **288**: 2431–2440
Totals	0	5	52		

12 RCTs tested 52 observational claims: 0 findings upheld; 10% found an effect in the *opposite* direction.

RECAP SEQUENCE

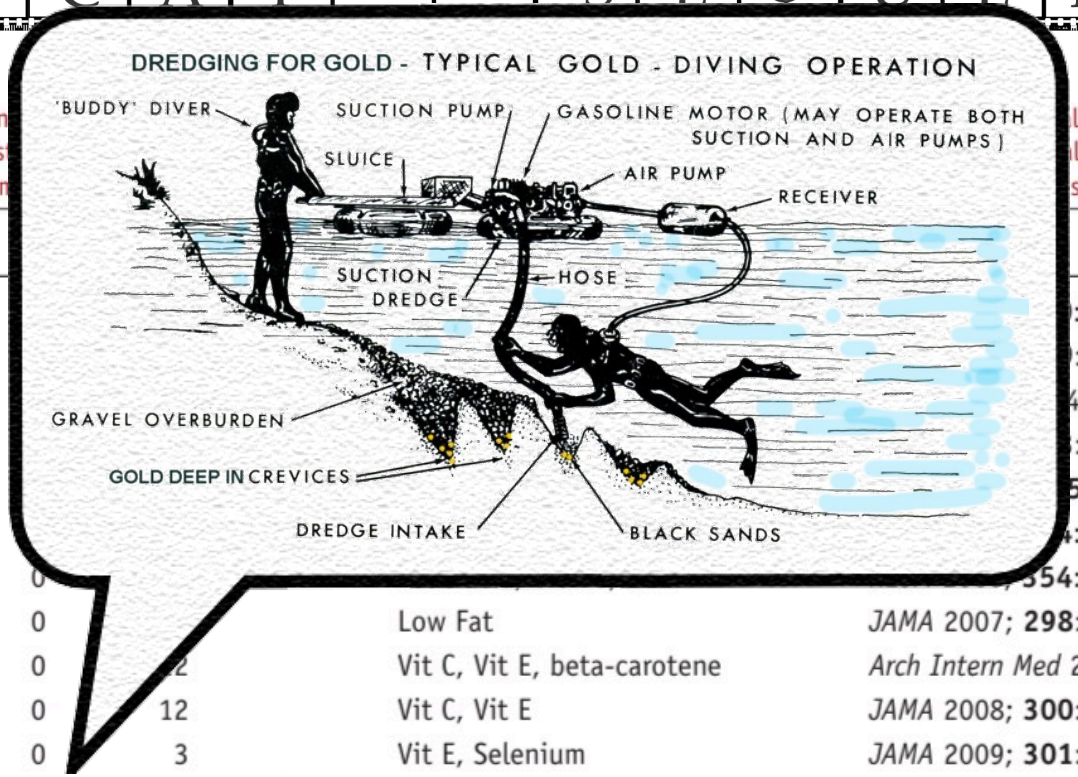


Table 1. We have foun … l trials. Many of the trials are quite large. In most of the observational s… …lly and together along with a placebo group. Note that none of the clai… …significance in the opposite direction five times

ID no.	Pos.				
1	0				…: 1029–1035
2	0				…: 2651–2662, 2663–2672, 2673–2684
3	0				…481–488
4	0				…: 1338–1347
5	0				…**5**: 655–666
6	0				…**4**: 669–683
7	0	0			…**354**: 2764–2772
8	0	0		Low Fat	*JAMA* 2007; **298**: 289–298
9	0	0	2	Vit C, Vit E, beta-carotene	*Arch Intern Med* 2007; **167**: 1610–1618
10	0	0	12	Vit C, Vit E	*JAMA* 2008; **300**: 2123–2133
11	0	0	3	Vit E, Selenium	*JAMA* 2009; **301**: 39–51
12	0	0	3	HRT + Vitamins	*JAMA* 2002; **288**: 2431–2440
Total			52		

2 observational claims: 0 findings upheld; 10% found an effect in the *opposite* direction.

RECAP SEQUENCE

Table 1. We have foun … l trials. Many of the trials are quite large. In most of the observational st… …lly and together along with a placebo group. Note that none of the clai… …significance in the opposite direction five times

ID no.	*Pos.*			
1	0			: 1029–1035
2	0			: 2651–2662, 2663–2672, 2673–2684
3	0			481–488
4	0			: 1338–1347
5	0			5: 655–666
6	0			: 669–683
7	0	0		54: 2764–2772
8	0	0		8: 289–298
9	0	0		007; **167**: 1610–1618
10	0	0	12	2123–2133
11	0	0	3	9–51
12	0	0	3	31–2
Total			52	

2 observational claims: 0 findings upheld; 10% found an effect … n.

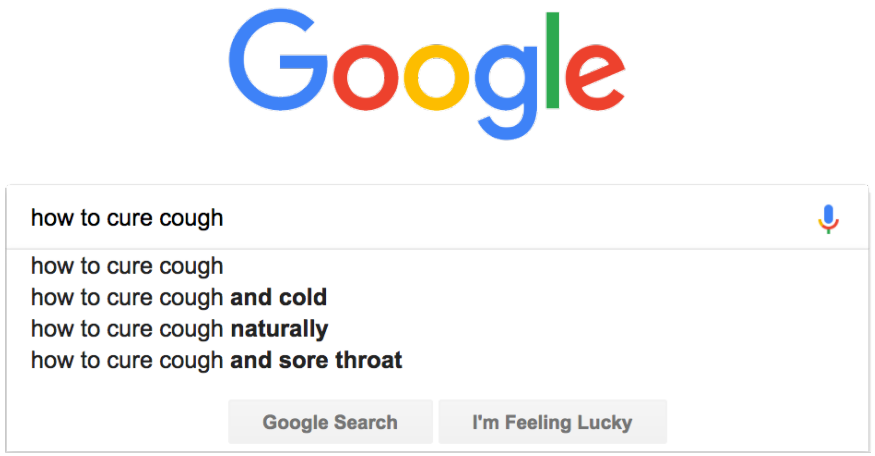
…versity
…e Pacific

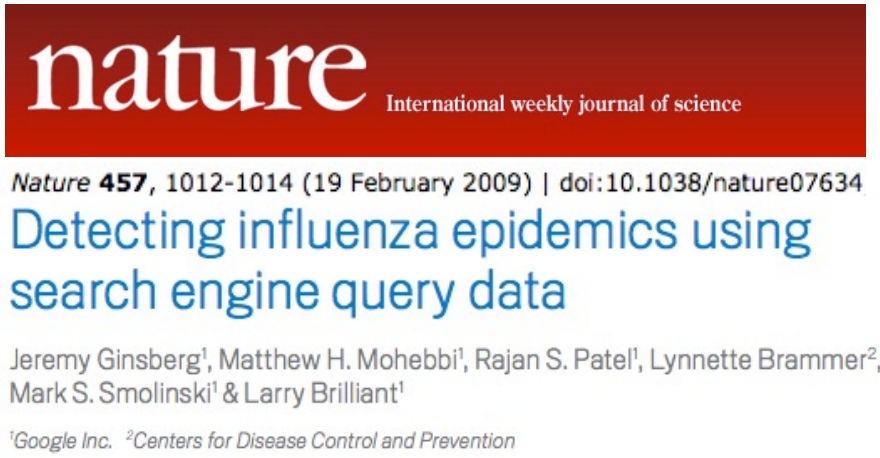
HESP 1…

RECAP SEQUENCE

21st century epidemiology:

RECAP SEQUENCE





Google receives 3 billion search queries every day and saves (and analyzes) them all. Analyses of those queries engenders a new form of public health surveillance.

RECAP SEQUENCE

Google Vs.

H1N1 outbreak. How bad is it?
Where is it spreading?

Traditional mode of reporting:

1. Patients would wait a while and then go to the doctor.
2. Doctors would wait a while and then report the cases.
3. The CDC would wait a while and then tabulate all cases.

All CDC reports were at least a week old, probably two.

.

RECAP SEQUENCE

Google *Vs.*

H1N1 outbreak. How bad is it? Where is it spreading?

”Big data” mode of reporting:

Google evaluates search terms and analyzes for terms that have a strong correlation with the spread of disease. They’re able to watch and report the spread in real time, not a week or two late.

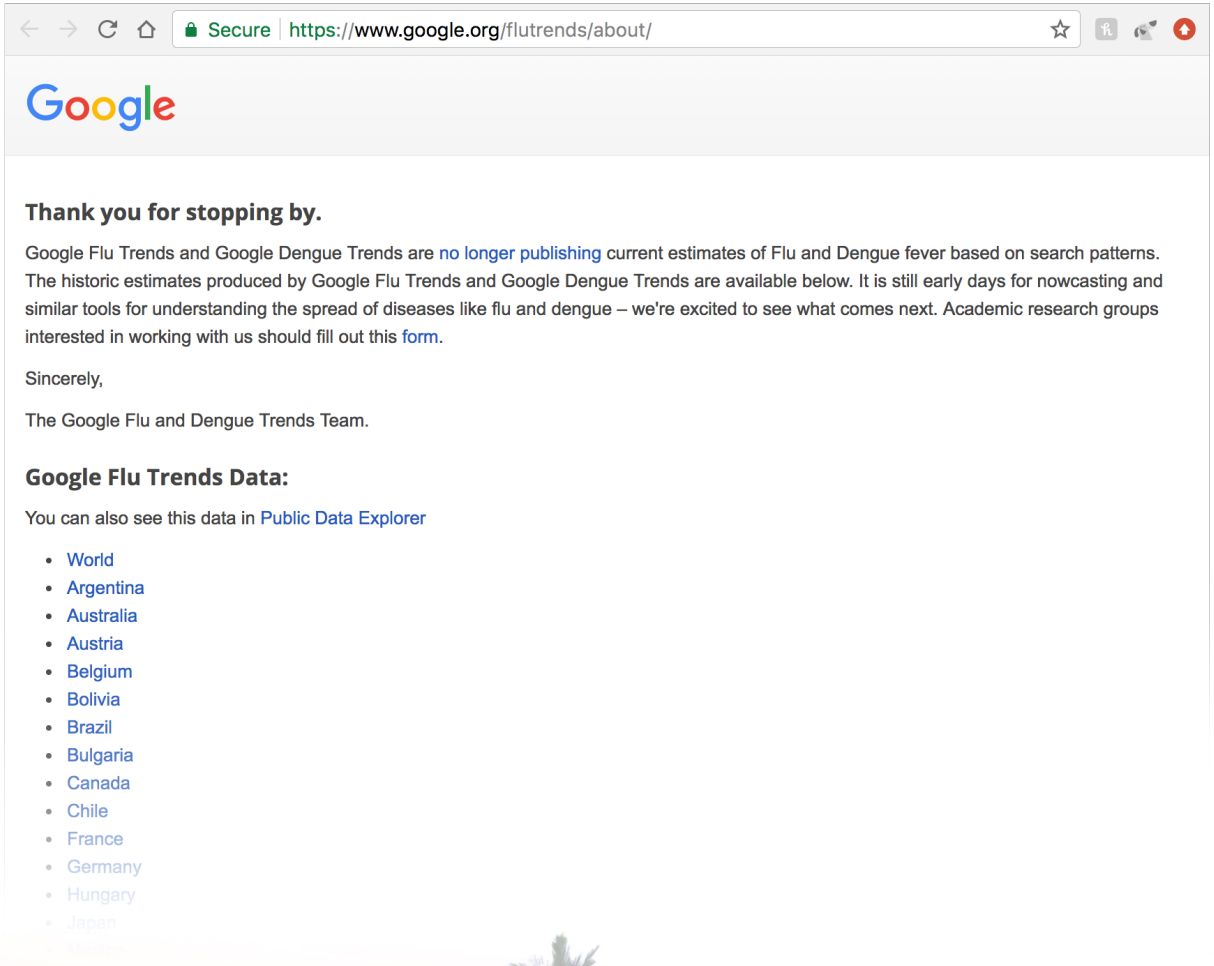
WIRED

DAVID LAZER AND RYAN KENNEDY SCIENCE 10.01.15 7:00 AM

WHAT WE CAN LEARN FROM THE EPIC FAILURE OF GOOGLE FLU TRENDS

And then, GFT failed—and failed spectacularly—missing at the peak of the 2013 flu season by 140 percent. When Google quietly euthanized the program, called Google Flu Trends (GFT), it turned the poster child of big data into the poster child of the foibles of big data. But GFT's failure doesn't erase the value of big data. What it does do is highlight a number of problematic practices in its use—what we like to call "big data hubris." The value of the data held by entities like Google is almost limitless, if used correctly. That means the corporate giants holding these data have a responsibility to use it in the public's best interest.

RECAP SEQUENCE



RECAP SEQUENCE

Digital information that can be "crunched" is relatively new.

1986: ~40% of the world's general purpose computing power was in pocket calculators. At the time, pocket calculators had more processing power than all personal computers combined.

RECAP SEQUENCE

Oren Etzioni

First undergrad to major in Computer Science at Harvard (graduated 1986).

Ph.D. at Carnegie Mellon University (1991)

Currently a professor at the University of Washington (director of the Turing Center) https://www.cs.washington.edu/people/faculty/fulltime and http://turing.cs.washington.edu/

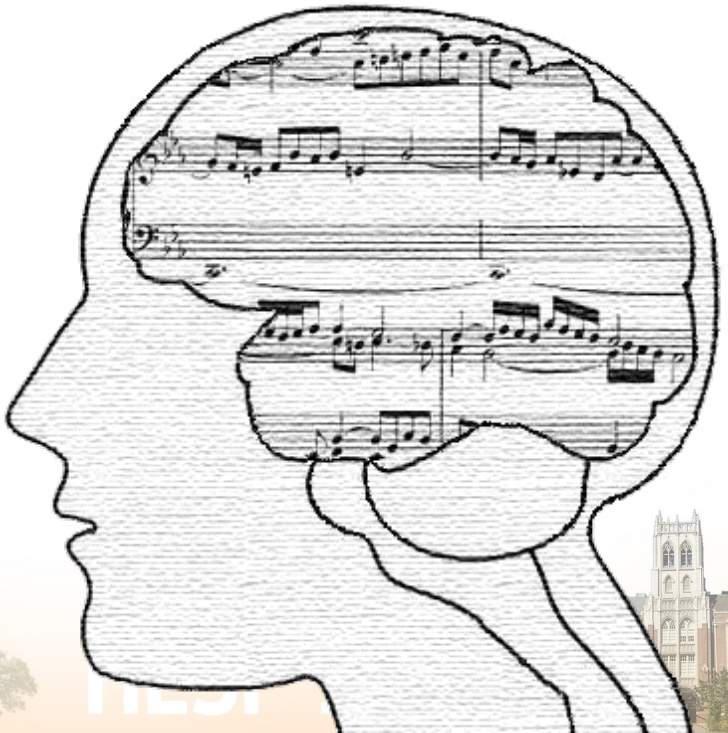
Current CEO of the Allen Institute for Artificial Intelligence (research institute funded by Microsoft Co-Founder Paul Allen). Here: http://oren-etzioni.people.allenai.org/

RECAP SEQUENCE

Music, Sports, and the Entertainments…

Music is less art than science now.

Sports are almost completely data driven.



RECAP SEQUENCE

The New York Times

BUSINESS DAY

What Wal-Mart Knows About Customers' Habits

By CONSTANCE L. HAYS NOV. 14, 2004

By its own count, Wal-Mart has 460 terabytes of data stored on Teradata mainframes, made by NCR, at its Bentonville headquarters. To put that in perspective, the Internet has less than half as much data, according to experts.

RECAP

WIRED

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

CHRIS ANDERSON 06.23.08 12:00 PM

Chris Anderson (*Wired* Editor–in-Chief) argues "big data" (epidemiology) is all you need. In a data driven world, there's no need for theories.

> The Petabyte Age is different because more is different. Kilobytes were stored on floppy disks. Megabytes were stored on hard disks. Terabytes were stored in disk arrays. Petabytes are stored in the cloud.

Scientific method replaced by statistical correlations, devoid of theory.

RECAP

WIRED

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

CHRIS ANDERSON 06.23.08 12:00 PM

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

RECAP

WIRED

THE END OF THEORY: THE DATA DELUGE MAKES THE SCIENTIFIC METHOD OBSOLETE

CHRIS ANDERSON 06.23.08 12:00 PM

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

RECAP SEQUENCE

Is this a good position?
Is "big data" all we need?

RECAP SEQUENCE

Epidemiology is very useful to answer *some* questions:

What conditions (sleeping position, etc.) maximize the odds of SIDS?

How much asbestos exposure is necessary to induce mesothelioma?

How many volts are necessary to ensure death by electrocution in well-behaved prepubescent children who love balloons?

What else? I might ask you to invent a question on a test (that we haven't previously discussed) that would be best investigated by epidemiology.

RECAP SEQUENCE

You can only study what you can measure.

You can only correct for what you can measure.

You can only make claims about what was measured and controlled for.

Epidemiology isn't doing that.

RECAP SEQUENCE

INTRODUCTION: JOIN THE SCIENTIFIC CONVERSATION

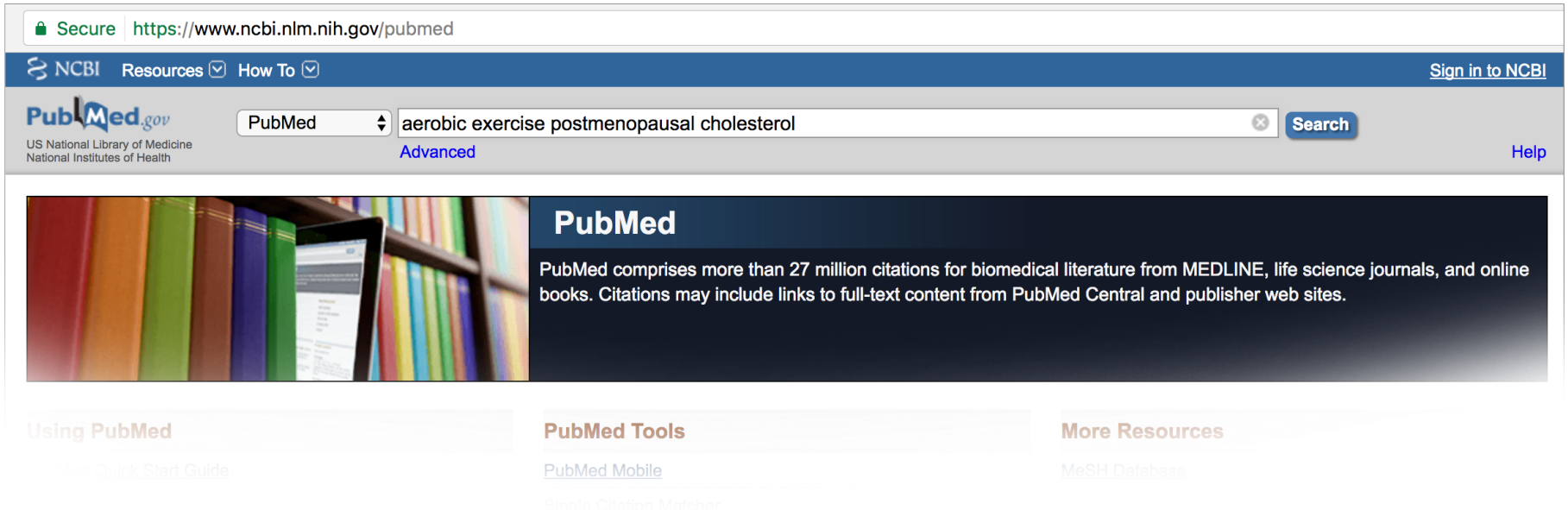
When it is time for you to design, complete, and share the findings of your own research project, whether you are a graduate or advanced undergraduate student, you are inserting yourself into the conversation. A seamless transition requires that you first learn what has been discussed in the previous conversation. Fortunately, the historical transcripts are archived in the p[illegible]. By taking the time to familiarize yourself with the existing li[illegible]nt insight into the questions that remain to be asked in your [illegible] foundation for developing an impactful research project.

Where do I fit in to all of this ?

RECAP SEQUENCE

We already talked about that! You're being a bore. Either say something new or remain in life's audience.

RECAP SEQUENCE



Once your topic is sufficiently narrowed, and all of your variables are identified, start exploring the currently-existing literature. This will require a lot of work. It won't just be a few hours. Count on it taking nearly forever.

RECAP SEQUENCE

The foundation of a good research project is a good **research question**, which originates from a strong understanding of the past and current literature in your field. When you have identified a field of study that interests you, it is then your duty as a researcher to become familiar with the findings of other researchers in your field. One of the benefits that comes with dedicating time to aggressively learning the relevant literature is that you will begin to detect **gaps in the literature**, and you will inevitably find yourself asking a question about a phenomenon in the field that has yet to be answered.

R E C A P S E Q U E N C E

Where in the continuum do your gaps reside?

CHAPTER 6 *Understanding Research*

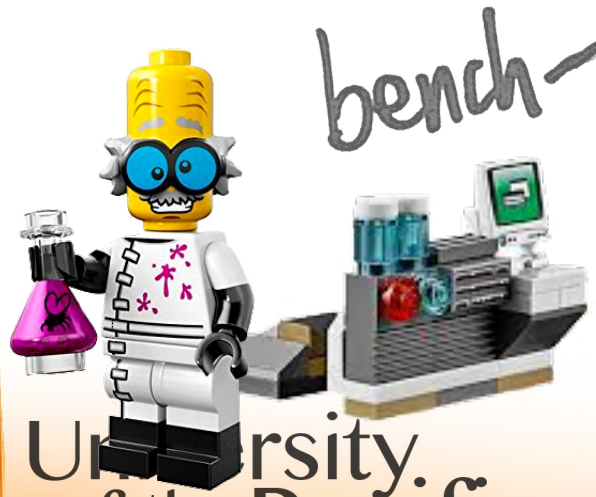
94 However, Hoover's 20th century dichotomy has expanded into a much broader gamut. It has evolved into a continuum that begins at bench science (animal models, human labs, etc.) and ends at the bedside (research aimed at informing clinical decisions).

RECAP SEQUENCE

Where in the continuum do your gaps reside?

Basic (or pure) research: Bench science. You're a microbiologist working with Petri dishes. You're a chemist who uses Bunsen burners and thinks about petroleum. You're a geneticist studying adaptations in fruit flies. Originally limited to university settings.

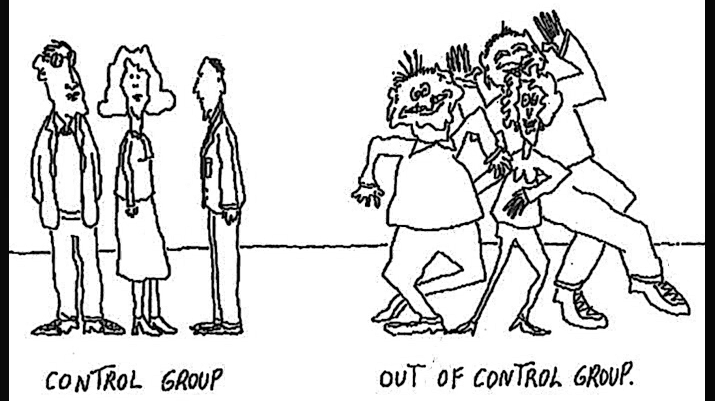
Applied research: Research that translates into "best practices". Originally an "industry" thing.



RECAP SEQUENCE

Gaps are bound to exist where data collection is difficult to accomplish.

Well-controlled studies need control groups and few pro athletes (or Olympians) are willing to be in the control.



"We're testing a conditioning program that should be the most effective thing you can do to prepare for your sport. We need 60 world-class volunteers. We'll randomize 20 to the group that gets the program we think is best; we'll randomize another 20 to a comparison program, which we believe will be a bit less effective; and the remaining 20 will be a control group; none of them will be allowed to exercise for the duration of the experiment."

Why does this not happen?

RECAP SEQUENCE

There's a shortage of evidence-based training programs for elite athletes.

"I was hired to produce winners. But I decided to run experiments on my athletes, which ensured that many of them would be losers."

This would be malpractice.

RECAP SEQUENCE

Internal vs. external validity.

Internal validity is soundness of one's methods (controlling for extraneous variables that might confound findings). With perfect internal validity, your claim is very probably true, but a narrow question with a narrow answer.

External validity is the ability to generalize. "Zumba causes a drop in body fat" (or whatever). Broad question, broad answer.

Consider blood pressure:

On average, exercise reduces blood pressure by 4 to 9 mmHg.

But if you look at the individual people who make up that average, you'll see a different story.

R E C A P S E Q U E N C E

Internal vs. external validity.

MAYO CLINIC

Diseases and Conditions

High blood pressure (hypertension)

Becoming more active can lower your systolic blood pressure — the top number in a blood pressure reading — by an average of 4 to 9 millimeters of mercury (mm Hg). That's as good as some blood pressure medications. For some people, getting some exercise is enough to reduce the need for blood pressure medication.

For example:

If you regularly perform aerobic exercise, will that lower your blood pressure?

RECAP SEQUENCE

723 subjects.

~400 responded to exercise with reductions in SBP.

Mean and median changes were reductions in SBP. Thus "exercise reduces blood pressure!"

> 300 subjects experienced an *increase* in SBP.

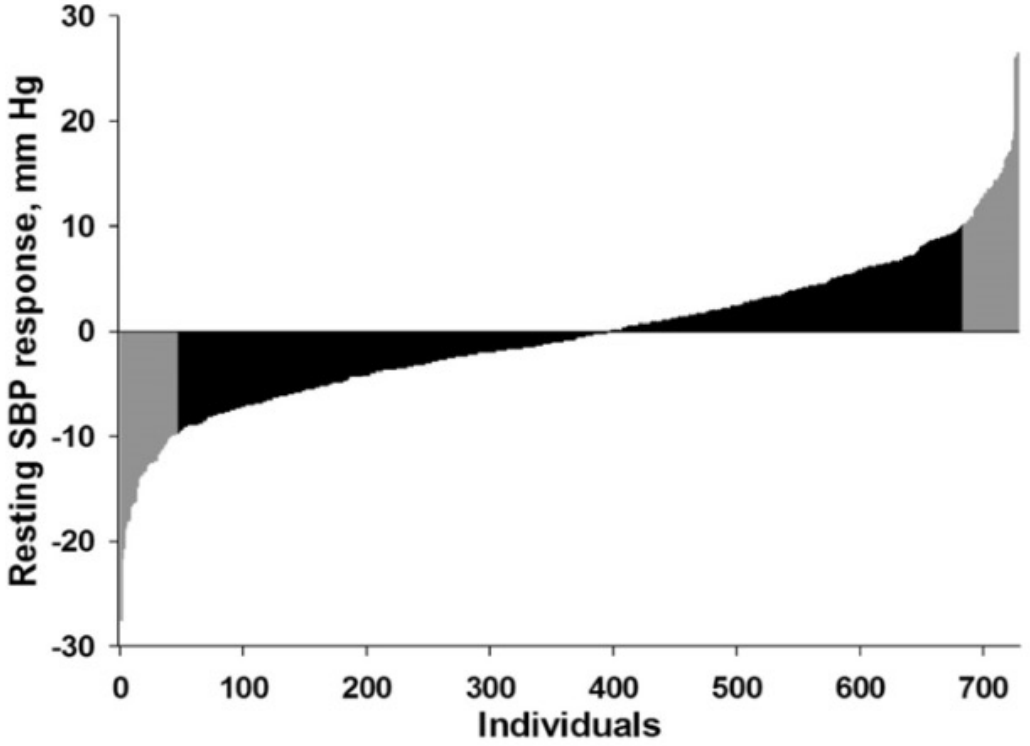


Fig. (6). Distribution of response to the HERITAGE exercise program with adverse and favorable responders highlighted in gray. SBP, systolic blood pressure. Adapted with permission from Bouchard C, Blair SN, Church TS, Earnest CP, Hagberg JM, Hakkinen K, *et al*. Adverse metabolic response to regular exercise: is it a rare or common occurrence? PLoS One 2012; 7(5): e37887.

Current Hypertension Reviews*, 2013, *Vol. 9, No. 2

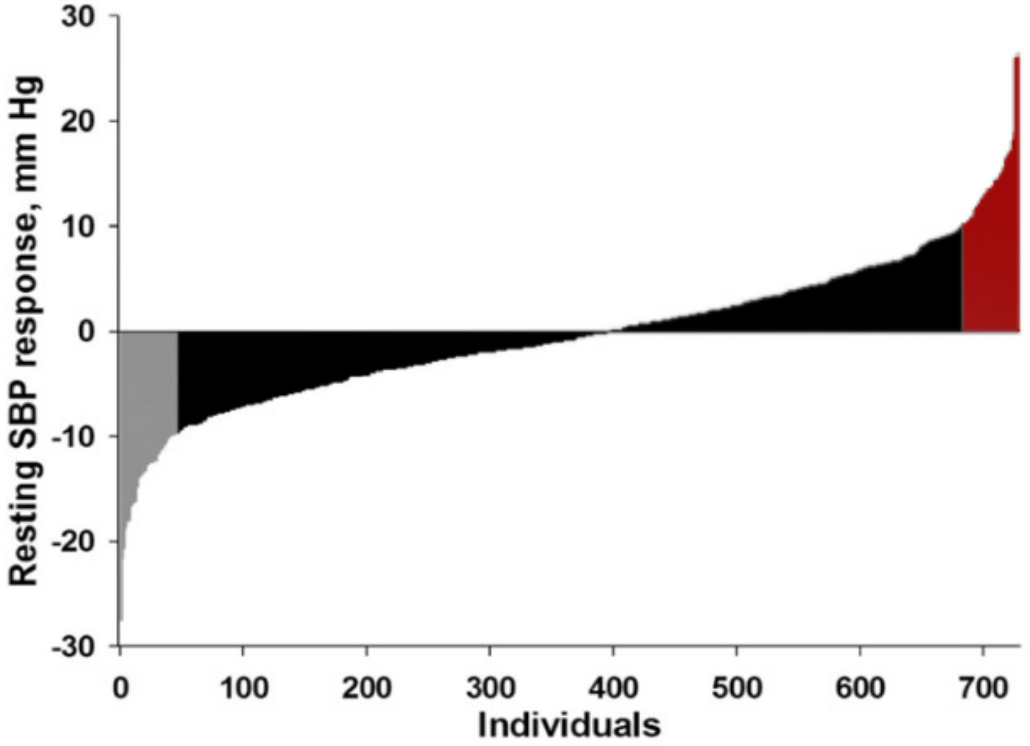
The Promises and Challenges of the Use of Genomics in the Prescription of Exercise for Hypertension: The 2013 Update

Garrett I. Ash[a,*], John D. Eicher[b] and Linda S. Pescatello[a]

RECAP SEQUENCE

12% of subjects experienced blood pressure increases of at least 10 mmHg.

Similar distributions were found to exist among these same subjects in insulin, cholesterol, and triglycerides.



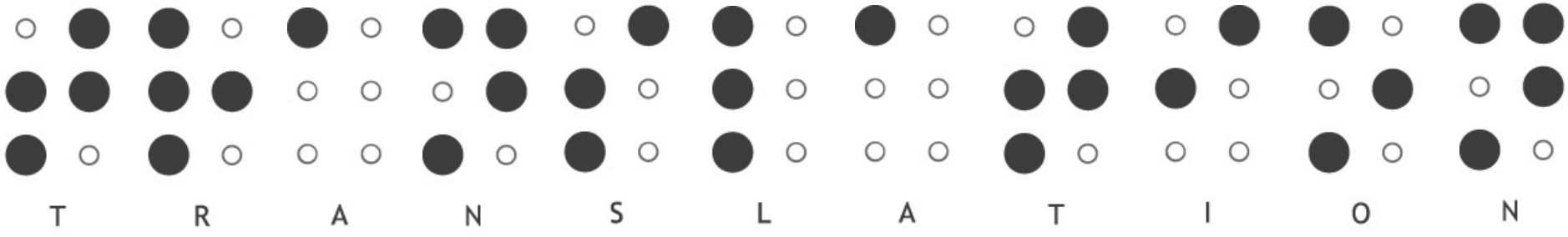
Current Hypertension Reviews, 2013, *Vol. 9, No. 2*

The Promises and Challenges of the Use of Genomics in the Prescription of Exercise for Hypertension: The 2013 Update

Garrett I. Ash[a,*], John D. Eicher[b] and Linda S. Pescatello[a]

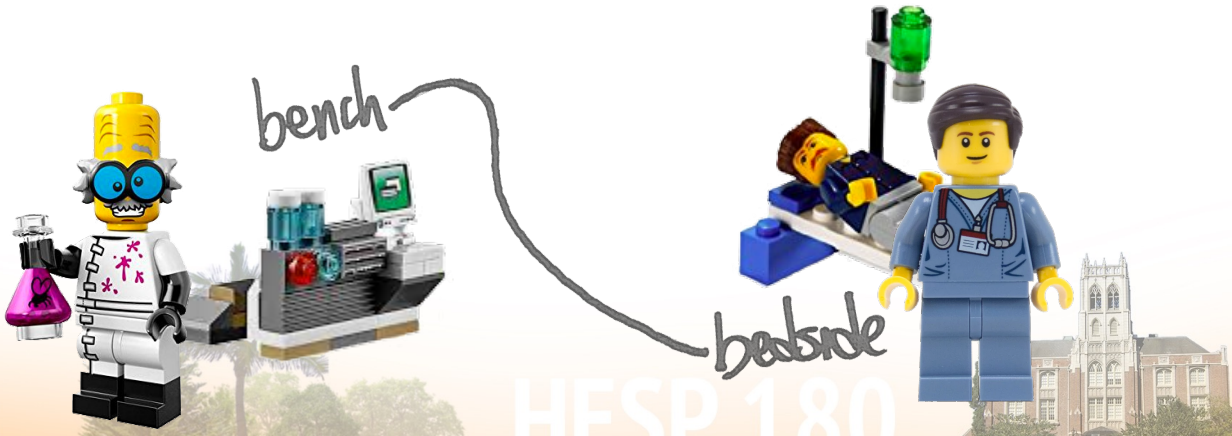
Fig. (6). Distribution of response to the HERITAGE exercise program with adverse and favorable responders highlighted in gray. SBP, systolic blood pressure. Adapted with permission from Bouchard C, Blair SN, Church TS, Earnest CP, Hagberg JM, Hakkinen K, *et al*. Adverse metabolic response to regular exercise: is it a rare or common occurrence? PLoS One 2012; 7(5): e37887.

RECAP SEQUENCE

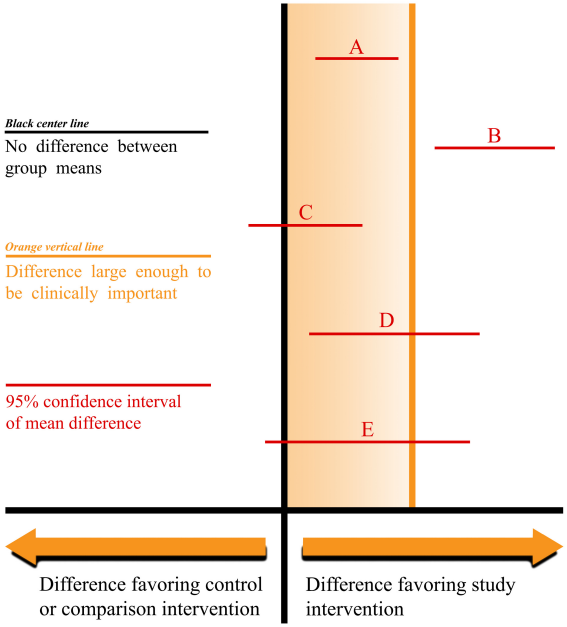


94 Translation:

Once the data have been analyzed, the results are to be interpreted and reported to the scientific, medical, and lay communities.



RECAP SEQUENCE



Magnitude of Effect

Not only is uncertainty not fully mitigated through inferential statistical analyses, it is often exacerbated. We attribute much of this problem to the use and misuse of the word "significant." It is not uncommon to read that investigators found "significant differences" between groups, suggesting a benefit of an intervention being studied or read about the "significant findings" of a study. The statistical meaning of significant is that the probability that a type I error has been committed is lower than the risk the investigators were willing to take. The word does not imply that the differences observed were of a magnitude worthy of attention.

Increasingly, the certainty or uncertainty of the magnitude of an effect is being conveyed through the reporting of **confidence intervals** (CIs). These statistics provide a range in which the population value is believed to lie at a given level of certainty. A 95% CI of mean differences provides knowledge (with 95% certainty) as to where a population mean difference may lie.

RECAP SEQUENCE

What is *variance*?

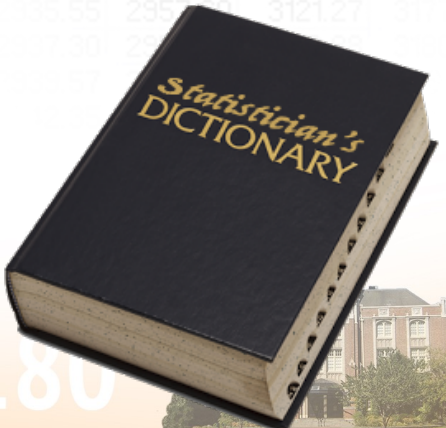
Tell me all about *variables*.

Independent and dependent variables? Extraneous variables?

What's the difference between a univariate and a multivariate analysis?

Nominal data? Binary nominal data? Ordinal data? Interval? Ratio? Scale?

Between-group? Within-group?



RECAP SEQUENCE

CHAPTER 6 *Understanding Research*

- Evidence-based practice
- What are the differences between basic (pure) and applied research?
- What does "bench to bedside" mean?

RECAP SEQUENCE

GLOSSARY 118

Biopsychosocial model: a framework by which human health and disease is evaluated in terms of biologic, psychologic, and social factors (and the interplay between each)

Confidence interval: in statistics, a CI estimates a population parameter. After measuring a variable in a sample of that population and calculating the mean, one can create a range in which the true mean of the population is likely to fall. This range is the CI. Are you 99% confident the true population mean lies within a particular range? That's a 99% CI. Are you 90% confident? That's a 90% CI (which is likely to span a much broader range)

Descriptive statistics: a quantitative characterization of a particular set of information. Unlike inferential statistics, it does not seek to make inferences based on probability but merely to describe or summarize what exists (quantitatively)

Frequentist: the inferential alternative to Bayesian statistics; the framework that allows for the testing of a null hypothesis and the calculation of CIs

Inferential statistics: unlike descriptive statistics, which summarize and characterize information, inferential statistics make inferences about collections of data, testing predictions about the larger population based on the distribution of the collected sample

Null hypothesis: a statistical starting point for most scientific investigations, stating that there is no difference between two measurable phenomena. Researchers then collect and analyze data so that they can either confirm or reject this hypothesis

RECAP SEQUENCE

What's political arithmetic?

Remember all the stuff about variables…

What's epidemiology?

Describe epidemiology in the 19th century.

Describe epidemiology in the 20th century.

Today: epidemiology in the 20th and 21st centuries.

What are the three major types of epidemiological studies?

Every student has the exact same access to the exact same readings, slideshows, and course materials. Every student has the exact same access to the exact same lectures, office hours, and GAs/tutors. And every student takes the exact same test.

If you don't have a drink in your hand, it's your own fault.

Some students get As, some get Bs, some get Cs, others get Ds, and there might be an F or two. This distribution has everything to do with the student and nothing to do with the nature of instruction or examination. You are fully in control of your grade. And life.

I'm not an "easy A". And anyone who is doesn't care about you; they only care about themselves. Professors who assign As to students who didn't earn them do so for three reasons: 1) So you'll like them, 2) So you won't complain, and 3) So you'll give them friendlier student evaluations at the end of the semester. Issuing an easy A is a selfish and terrible thing to do.

Life is hard and unforgiving, you'll find. Much more difficult than school. It grades way harder than I do. Pretty much everyone in America has (or is getting) a college degree. No matter what you want to do with your life, you're going to have to outcompete huge crowds of educated, talented, motivated people. And it's nearly impossible. So if you spend your undergrad years comfortably coasting your way to As, there's little chance you'll find success after college. Life really is that hard. So anyone who gives you easy grades is not setting you up for success. My life would be easier (nobody would complain about anything) if I gave out As for free. But the future net worth of my students would suffer summin' fierce. So I won't do that. BUT: everything you need to get your A is in the slides and lectures, and I tell you which parts of the books are helpful. That doesn't make the A "easy", but it *definitely* makes it achievable. If you want it, it's yours. And it's useful to have.