0D 2019 Univ. New Mexico 1

Teaching Confounder-Based Statistical Literacy

Milo Schield, US
Fellow: American Statistical Assoc.
US Rep: International Statistical Literacy Project

June 19, 2019
Dept. Math & Statistics
University of New Mexico
www.StatLit.org/pdf/2019-Schield-UNM-Slides.pdf

0D 2019 Univ. New Mexico 2

Confounding: Common Misuse

Confounding is used to show that "association is not causation". We then spend an entire semester on randomness (never mentioning confounding again).

This is "Bait and Switch".
"Bait and switch" is unethical!
"Bait and switch" is professional negligence!

This is arguably why most students see less value in 'statistics' after taking the intro research-methods course – than they did before taking the course.

0D 2019 Univ. New Mexico 3

My First Day #1: Coincidence (Chance)

Do some people have special powers?
Let's find out. Who gets longest run?
Q1. Could the winner have special powers?
Q2. What's another explanation?

Luck, coincidence, chance or "skill"?
Q3. How can we find out right now?

Do it again (Repeat)

0D 2019 Univ. New Mexico 4

First Day #2: Confounding

Studies show: "People that read home and fashion magazines are much more likely to get pregnant than people that read car and sport magazines."
Q1 What's an alternate explanation? Gender
Q2 How can we see this in the data? Stratify!

Suppose the best hospital had the highest death rate.
Q3. Is this strong evidence it's a bad hospital?
Q4. What's an alternate explanation? Patient health
Q5. How can we see this in the data? Stratify!

0D 2019 Univ. New Mexico 5

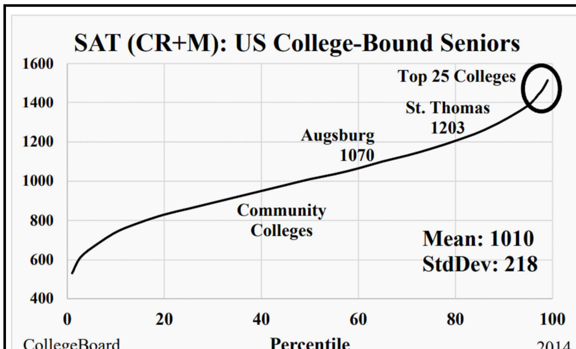
Statistical Literacy: Two Approaches

Mathematics: Rationalism
Define statistical literacy. Show what follows.

Business: Empirical*/Teleological
Who is the customer? What do they need?

Today: Empirical first; Rationalist second.

* See Schield (2008). Von Mises' Frequentist Approach to Statistics
www.statlit.org/pdf/2008SchieldBurnhamASA.pdf
3 citations, 2 recommendations, 500+ reads on ResearchGate.



0D 2019 Univ. New Mexico 7

Math-Stat Teachers vs. Non-Math Students

Math majors have highest Math SATs

		SAT MATH SCORE			
Score	Pctile	MAJOR	Score	Pctile	MAJOR
613	80	Math/Stats	550	61	Social Sc.
575	72	Physical Sc.	522	51	Bus; English
579	70	Engineering	498	43	Commun.
554	62	Computer Sc.	489	40	Psychology
551	61	Biological Sc.	482	38	Education
Business Insider (2014)					College Board (2015)

0D 2019 Univ. New Mexico 2015 8

Harvard Biz Review Cases 42K Word Prevalence: Abstract

CONTROL/CONFOUND	580
234	"control of"
137	standardize (ed)
64	"take into account"
49	"taking into account"
22	"taken into account"
14	"control for"
17	"adjust(ing, ed) for"
4	"controlling for"
1	"controlled for"
1	"took into account"
8	"common cause"
3	"sampling bias"
15	confound (ed)
9	confounding
1	confounder
1	"another explanation"
0	"alternate explanation"
0	"lurking variable"
0	"effect size:
0	z-score

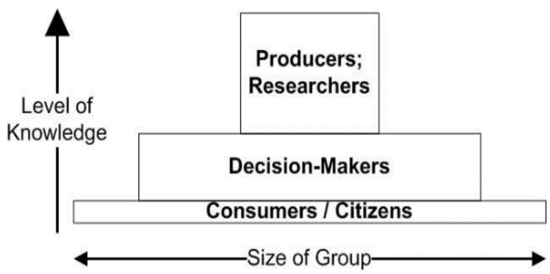
STUDY DESIGN	435
400	experiment
22	"clinical trial"
7	"longitudinal study"
3	"randomly assigned"
0	"random assignment"
0	"observational study"
1	"controlled study"
0	"clinical study"
2	"clinically proven"

#	SAMPLE 240*
145	sample
88	random (ly)
7	randomness
2	"random sample"
0	"stratified sample"
0	"cluster sample"
0	"systematic sample"
0	"convenience sample"
0	"sample statistic"

INFERENTIAL	33
7	ANOVA
7	"statistical significance"
4	"statistically significant"
4	t-test
3	"standard error"
2	chi-squared
2	"hypothesis test"
1	"sampling error"
1	"margin of error"
1	"statistical power"
1	p-value
0	"not statistically significant"
0	"statistically insignificant"
0	"sampling distribution"
0	"null hypothesis"
0	"alternate hypothesis"
0	"research hypothesis"
0	"reject the null"
0	"rejection region"
0	"confidence level (interval)"

0D 2019 Univ. New Mexico 9

Most students take statistics. Our Audience/Customers



0D 2019 Univ. New Mexico 10

Statistical Literacy: NOT

Reading numbers in the news.

Not just pedagogy in traditional stats:

- *Including major projects in statistics*
- *Using resampling to create confidence intervals*
- *Use of resampling to run hypothesis tests*
- *Analyzing results of clinical trials*
- *Analyzing results of random surveys/polls*

0D 2019 Univ. New Mexico 11

Statistical Literacy: An Overview

Statistical literacy studies statistics as evidence in arguments.

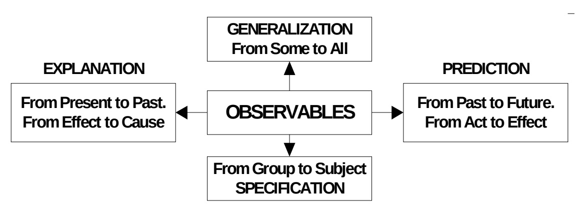
Most statistical arguments involve observational statistics. These are easily *confounded*.

Confounding is what connects statistical literacy to the humanities, the liberal arts, the social sciences, the professions and the soft physical sciences (geology, astronomy, epidemiology, etc.)

0D 2019 Univ. New Mexico 12

Statistical Literacy: Four Kinds of Arguments

Statistical literacy studies statistics as evidence in arguments.



0D 2019 Univ. New Mexico 13

Defining a "statistic" in Traditional Statistics

In order to unpack this definition, we must be clear on how we define "statistic".

In traditional inference-based courses,
a statistic is a property of a sample.

- Typically random samples.
- Typically small random samples.

0D 2019 Univ. New Mexico 14

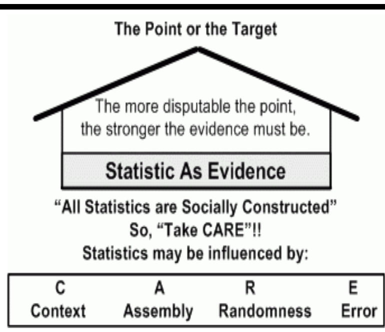
Defining a "statistic" in Statistical Literacy

1. Statistics are different from numbers.
2. Statistics are between number and words.
3. Statistics are numbers in context – where the context matters.
4. Statistics are counts and measures of real things

Consequence: Statistics can be influenced.
StatLit studies ALL the influences on a statistic

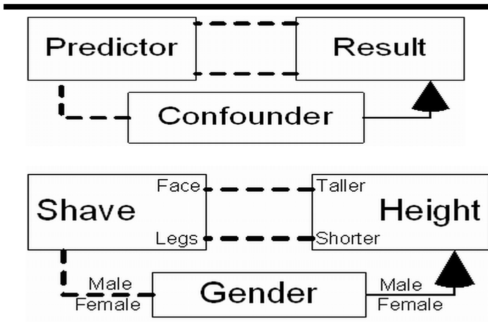
0D 2019 Univ. New Mexico 15

Statistical Literacy Studies ALL Influences



0D 2019 Univ. New Mexico 16

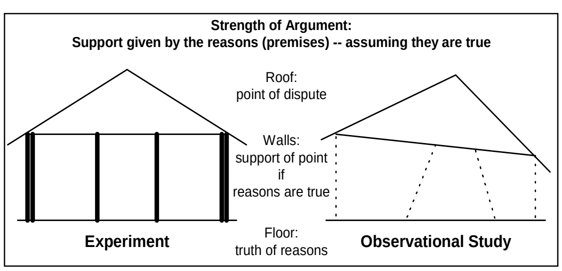
People that shave their face are taller…



0D 2019 Univ. New Mexico 17

Study Design Can Ward Off Confounders

Experiments vs. Observational Study



0D 2019 Univ. New Mexico 18

Study Design Can Ward Off Confounders

CONTROL OF CONFOUNDERS				
Physical Control (Grade = Quality)				
Experiment			Observational Study	
A+	Scientific		C	Longitudinal
A-	Random Assign		D	Cross-sectional
B	Quasi-Exper		F	Anecdotal story

0D 2019 Univ. New Mexico 19

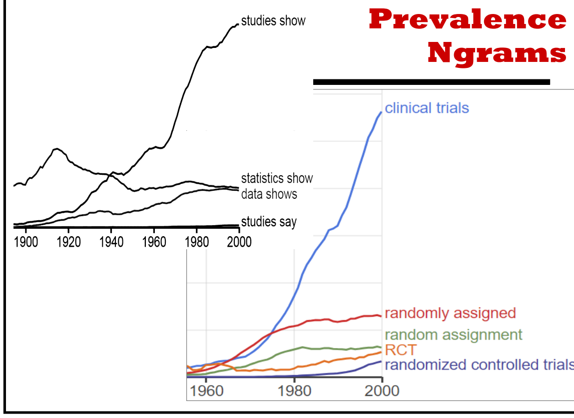
Ward off Confounders: Quasi-Experiments

Quasi (Queasy)-Experiment
Nature or humans intervene on pre-existing groups

Nature intervenes	*Humans intervene*
Epidemics Plagues, outbreaks	**Wars/Politics** Change laws & policies
Natural disasters Earthquakes, tornadoes	**Business/Education** Change pricing/teaching

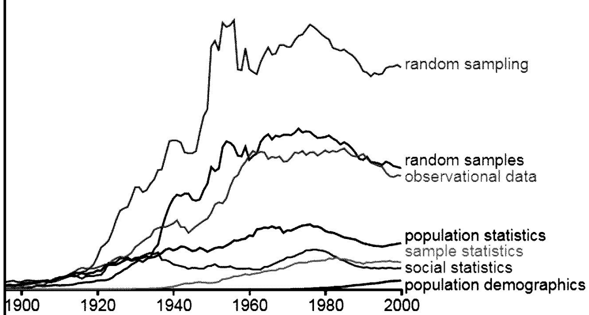
0D 2019 Univ. New Mexico 20

Prevalence Ngrams



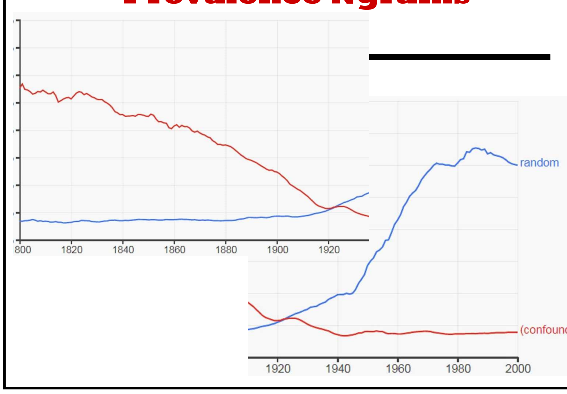
0D 2019 Univ. New Mexico 21

Prevalence Ngrams



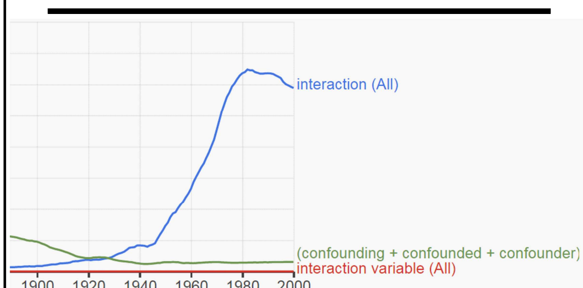
0D 2019 Univ. New Mexico 22

Prevalence Ngrams



0D 2019 Univ. New Mexico 23

Prevalence Ngrams



0D 2019 Univ. New Mexico 24

Three Big Contributions to Human Knowledge

1. Association is not causation
2. The Central Limit theorem; the formula for Standard Error: statistical average error. Howard Wainer calls this "the most dangerous equation" (next to $E=mc^2$)
3. Fisher's uses of random assignment to control for pre-existing confounders

Biggest Omissions relative to Human Knowledge

What are statistical educators biggest sins in teaching introductory statistics?

1. Ignoring multivariate data, observational studies and confounding.
2. Failing to show that controlling for a confounder can change statistical significance.
3. Ignoring the Cornfield conditions.
4. Ignoring how definitions can influence Stat. Sig.

Statistical Literacy: Four Biggest Problems

1. Lack of focus on confounding
2. Students
   - may become cynics about every statistic.
   - will have less respect for our discipline.
3. Teachers:
   Math/stat teachers: not trained to teach literacy.
   Math/stat teachers: don't want to teach literacy.
4. Textbook and teacher training materials

Statistical Literacy and Confounding

Statistical literacy: the discipline that studies:
* all the influences on a statistic.

In observational studies, confounding is arguably the most common – most important – influence.

The statistical literacy "debate" is ultimately between the 'pro' and the 'anti' confounders.

Schield is – and has always been – pro-confounder. See Schield (1998) for "confounding factors".

Confounding Almost Absent in GAISE 2005

K-12 report: The first line: "The ultimate goal: statistical literacy". Confounding is mentioned twice: once to define and once to note it may create patterns that are not a "reliable basis for statistical inference".

College report: Confounding is mentioned only once. It is not defined; it appears in a sample problem in a list of words that may apply in analyzing data from an observational study.

Confounding mentioned in GAISE 2016 Update

Plus: Confounding shown 18 times (big increase):

Twice up front:

Goal 9: Ethics: "with large data sets, … under-standing confounding … becomes even more relevant." p 11

- Recommendation: Multivariable thinking. Examples "show how confounding plays an important role…" p.15

Nine times in appendix B: 34, 38 (3), 40 (2), 41(3)

Seven times later: Footnote 105; 113, 120, 122 (4).

Minus: Not in any one-line recommendations/goals

Silence on Smoking and Lung Cancer

Extremely important observational study.
Smoking is a most-likely cause of cancer.

MINUS:

- Not in most statistics textbooks.
- Not mentioned in GAISE 2005 College.

PLUS:

- Discussed in detail in GAISE 2004 K-12.
  But *confounding* was never used in the discussion

Why are we silent on confounding?

1. Confounding is not an issue in predicting.
2. There is no test for confounding. Judea Pearl
3. Using association as evidence for causation is a matter for subject-matter experts. Statisticians have no professional opinion on the subject.
4. Discussing confounding would bring disrepute on our discipline.

How can we change the present?

We need to go back to the past.

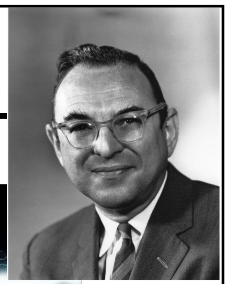
We need to revisit the Fisher-Cornfield dialogue on whether smoking caused lung cancer.

We need to revisit Cornfield's conditions for a confounder to nullify or reverse an association.

We need to see how to change statistical education to include Cornfield's criteria for confounding.

Back to the Future: Here we Go!

Back to 1958.



Back to the Future: Jerome Cornfield: 1912-1979

Jerome Cornfield got his BA and MA in history. He studied statistics at the US Dept of Agriculture.

He worked for USDA on sampling and study design

He created two common statistical measures: Relative risk (RR) and the Odds Ratio (OR).

He carfully compared prospective (cohort) and retrospective (case control) studies.

He was elected President of the ASA in 1974.

Back to the Future: Cornfield

Statisticians are subdued in talking about Cornfield; they are silent on the Cornfield conditions.

1. Not listed in RSS statistical timeline. Not listed in Wikipedia Timeline of Statistics
2. Wikipedia: Nothing on his work on confounding in the Smoking-Cancer studies.
3. Nothing in most of statisticians' commentaries about the Cornfield condition.

Why are We Silent on Confounding?

Because we don't know Cornfield's conditions!

Cornfield conditions: Minimum confounder size to nullify or reverse an observed association.

Impact: Allowed statisticians to say that "Smoking causes cancer" using data from an observational study.

0D 2019 Univ. New Mexico 37

Resist X% of Confounders

Confounders have no single analytical distribution. There is no way to say that a given effect size will resist X% of all relevant confounders.

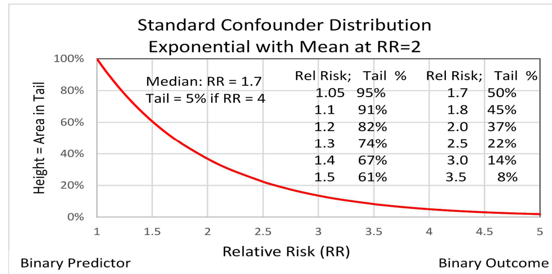
But we can postulate a standard distribution of confounders: say an exponential distribution of relative risks with a mean of 2 (median of 1.69).

An RR of 4 will resist 95% of these standard confounders. 1.5 resists 40%; 1.2 resists 20%.

0D 2019 Univ. New Mexico 38

Confounder Resistance: Propose a Standard

Arbitrary, but simple and fits existing data.



0D 2019 Univ. New Mexico 39

Summary Need Cornfield Conditions

With Cornfield conditions, we can

1. Show that the larger the effect size, the more resistance an association has to causation. (Schield and Burnham, 1998)
2. Show how to use Cornfield's conditions as necessary conditions. Schield (2012).
3. Show how to work problems controlling for a binary confounder. Schield ().

0D 2019 Univ. New Mexico 40

Conclusions

By featuring confounding in introductory statistics we can change our destiny. Statistical literacy can help untangle the confusion in many political debates.

Distinguishing between a crude association and a standardized association would be a big step forward

We are at a fork in the road. Which one will statistical educators take? Their choice will influence what most college graduates will study in decades to come.

Teaching Confounder-Based Statistical Literacy

Milo Schield, US

Fellow: American Statistical Assoc.

US Rep: International Statistical Literacy Project

June 19, 2019

Dept. Math & Statistics

University of New Mexico

www.StatLit.org/pdf/2019-Schield-UNM-Slides.pdf

Confounding: Common Misuse

Confounding is used to show that “association is not causation”. We then spend an entire semester on randomness (never mentioning confounding again).

This is “Bait and Switch”.
“Bait and switch” is unethical!
“Bait and switch” is professional negligence!

This is arguably why most students see less value in ‘statistics’ after taking the intro research-methods course – than they did before taking the course.

My First Day
#1: Coincidence (Chance)

Do some people have special powers?

Let's find out. Who gets longest run?

Q1. Could the winner have special powers?

Q2. What's another explanation?

Luck, coincidence, chance or "skill"?

Q3. How can we find out right now?

Do it again (Repeat)

First Day
#2: Confounding

Studies show: “People that read home and fashion magazines are much more likely to get pregnant than people that read car and sport magazines.”

Q1 What’s an alternate explanation? Gender

Q2 How can we see this in the data? Stratify!

Suppose the best hospital had the highest death rate.

Q3. Is this strong evidence it’s a bad hospital?

Q4. What’s an alternate explanation? Patient health

Q5. How can we see this in the data? Stratify!

Statistical Literacy: Two Approaches

Mathematics: Rationalism

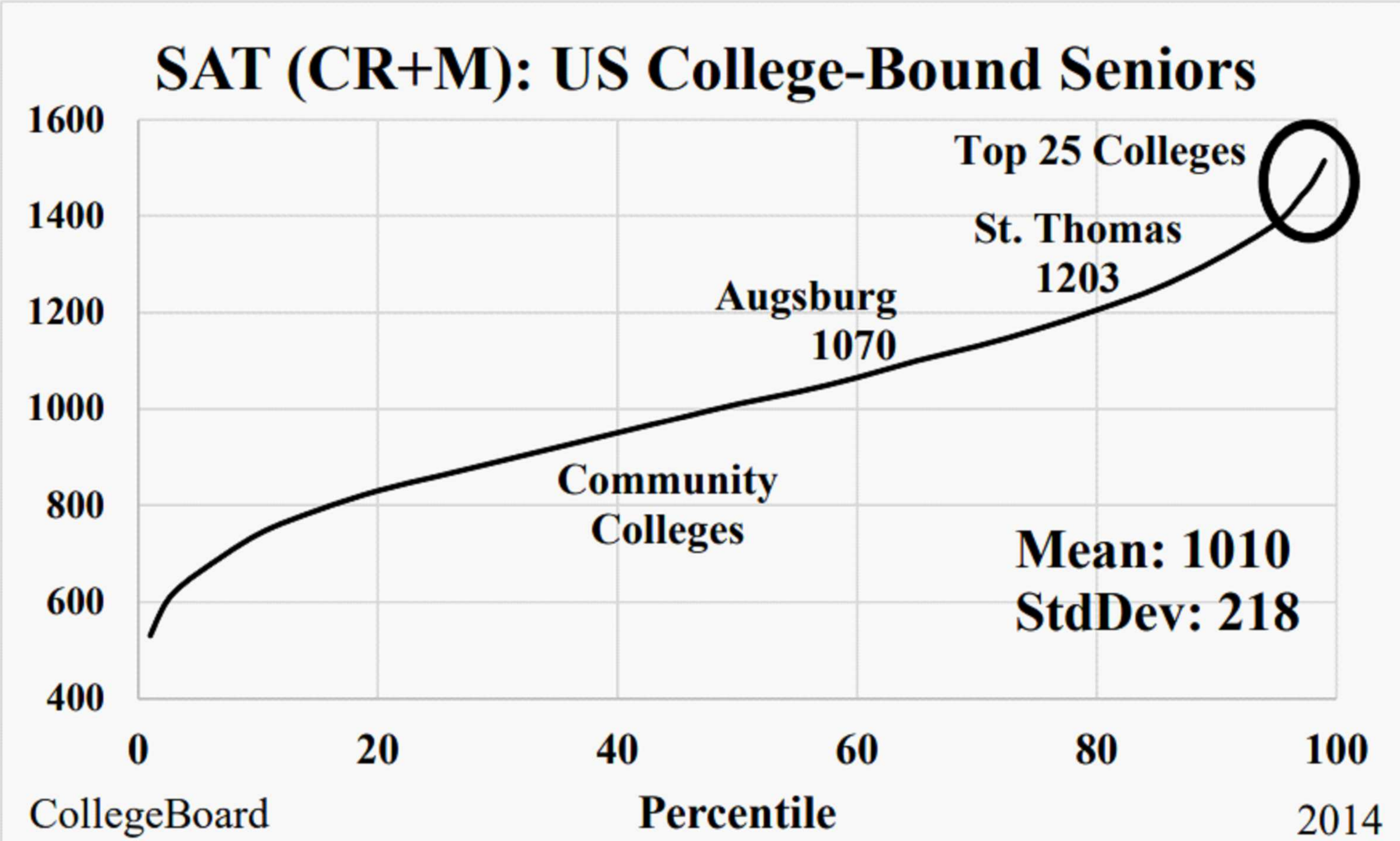
Define statistical literacy. Show what follows.

Business: Empirical*/Teleological

Who is the customer? What do they need?

Today: Empirical first; Rationalist second.

* See Schield (2008). Von Mises' Frequentist Approach to Statistics
www.statlit.org/pdf/2008SchieldBurnhamASA.pdf
3 citations, 2 recommendations, 500+ reads on ResearchGate.


SAT (CR+M): US College-Bound Seniors
1600
1400
1200
1000
800
600
400
0
20
40
60
80
100
Top 25 Colleges
St. Thomas
1203
Augsburg
1070
Community
Colleges
Mean: 1010
StdDev: 218
CollegeBoard
Percentile
2014

Math-Stat Teachers vs. Non-Math Students

Math majors have highest Math SATs

		SAT MATH SCORE			
Score	Pctile	MAJOR	Score	Pctile	MAJOR
613	80	Math/Stats	550	61	Social Sc.
575	72	Physical Sc.	522	51	Bus; English
579	70	Engineering	498	43	Commun.
554	62	Computer Sc.	489	40	Psychology
551	61	Biological Sc.	482	38	Education
Business Insider (2014)				College Board (2015)	

Harvard Biz Review Cases 42K Word Prevalence: Abstract

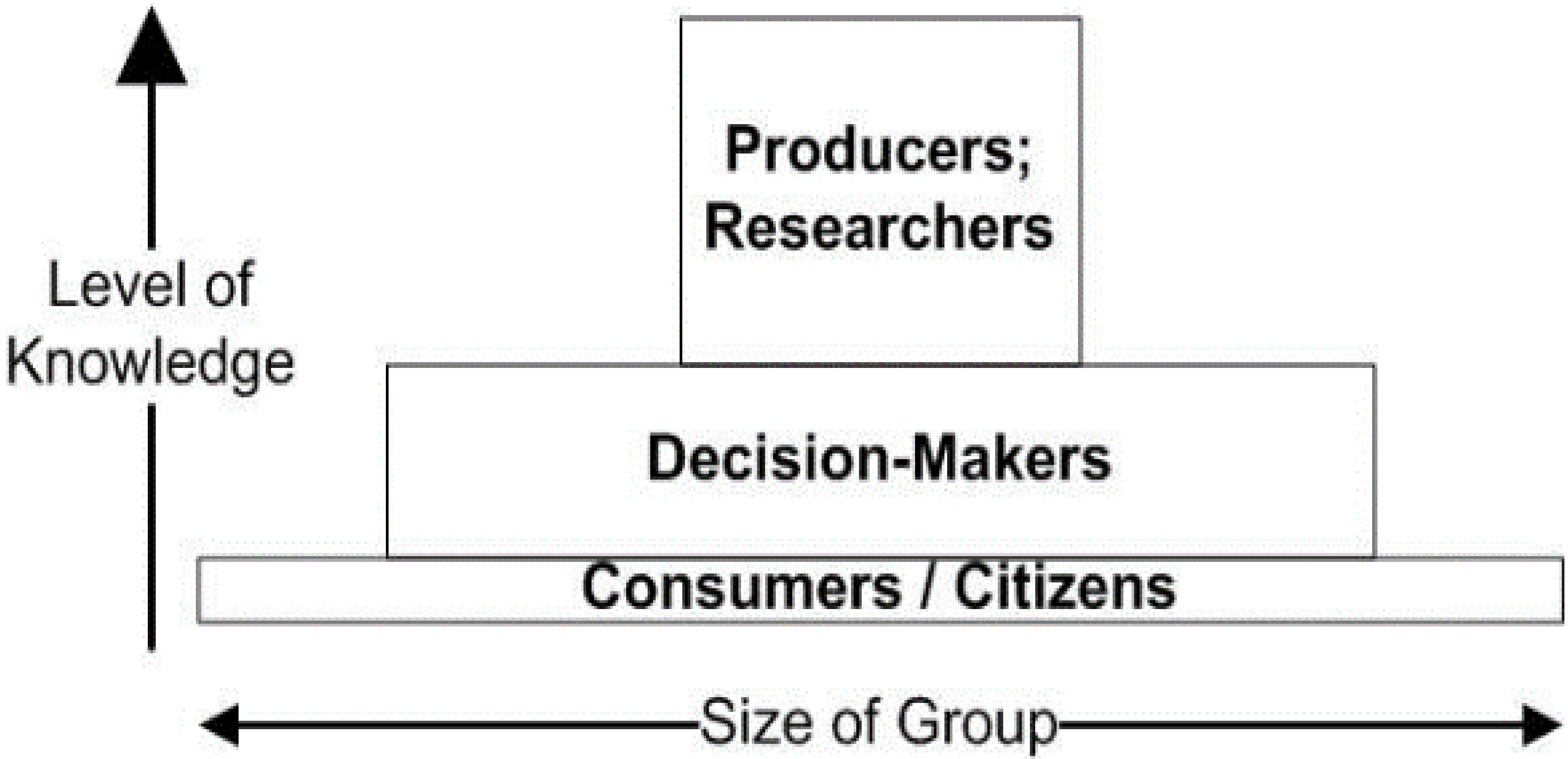
CONTROL/CONFOUND	580
234	"control of"
137	standardize (ed)
64	"take into account"
49	"taking into account"
22	"taken into account"
14	"control for"
17	"adjust(ing, ed) for"
4	"controlling for"
1	"controlled for"
1	"took into account"
8	"common cause"
3	"sampling bias"
15	confound (ed)
9	confounding
1	confounder
1	"another explanation"
0	"alternate explanation"
0	"lurking variable"
0	"effect size:
0	z-score

STUDY DESIGN	435
400	experiment
22	"clinical trial"
7	"longitudinal study"
3	"randomly assigned"
0	"random assignment"
0	"observational study"
1	"controlled study"
0	"clinical study"
2	"clinically proven"

# SAMPLE	240*
145	sample
88	random (ly)
7	randomness
2	"random sample"
0	"stratified sample"
0	"cluster sample"
0	"systematic sample"
0	"convenience sample"
0	"sample statistic"

INFERENTIAL	33
7	ANOVA
7	"statistical significance"
4	"statistically significant"
4	t-test
3	"standard error"
2	chi-squared
2	"hypothesis test"
1	"sampling error"
1	"margin of error"
1	"statistical power"
1	p-value
0	"not statistically significant"
0	"statistically insignificant"
0	"sampling distribution"
0	"null hypothesis"
0	"alternate hypothesis"
0	"research hypothesis"
0	"reject the null"
0	"rejection region"
0	"confidence level (interval)"

Most students take statistics. Our Audience/Customers



Statistical Literacy: NOT

Reading numbers in the news.

Not just pedagogy in traditional stats:

- *Including major projects in statistics*
- *Using resampling to create confidence intervals*
- *Use of resampling to run hypothesis tests*
- *Analyzing results of clinical trials*
- *Analyzing results of random surveys/polls*

Statistical Literacy: An Overview

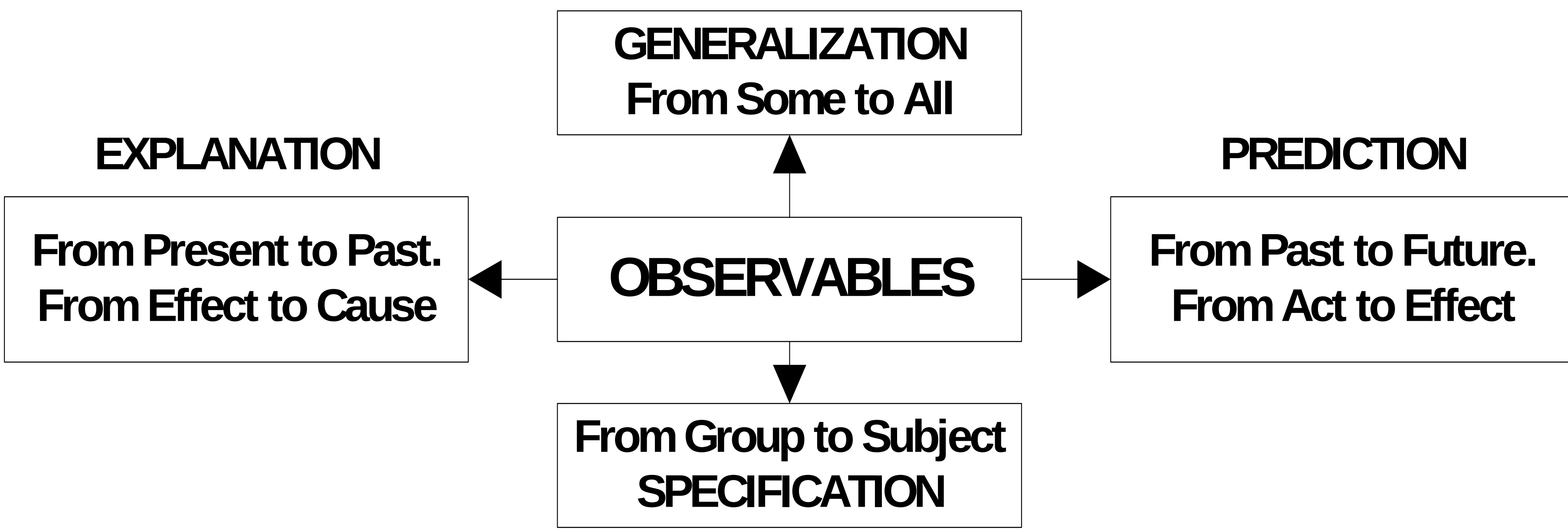
Statistical literacy studies statistics as evidence in arguments.

Most statistical arguments involve observational statistics. These are easily *confounded*.

Confounding is what connects statistical literacy to the humanities, the liberal arts, the social sciences, the professions and the soft physical sciences (geology, astronomy, epidemiology, etc.)

Statistical Literacy: Four Kinds of Arguments

Statistical literacy studies statistics as evidence in arguments.



Defining a "statistic" in Traditional Statistics

In order to unpack this definition, we must be clear on how we define "statistic".

In traditional inference-based courses,
a statistic is a property of a sample.

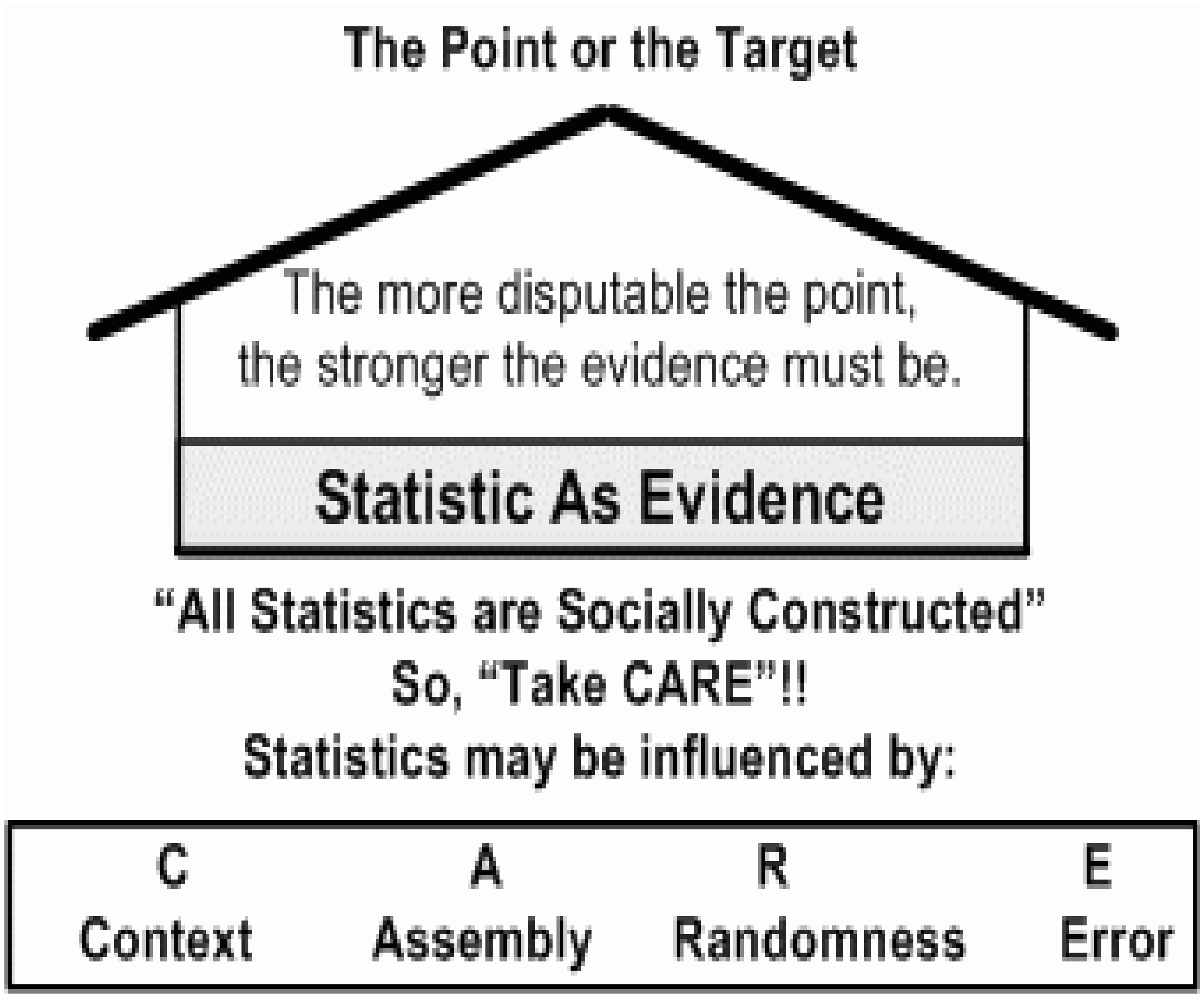
- Typically random samples.
- Typically small random samples.

Defining a "statistic" in Statistical Literacy

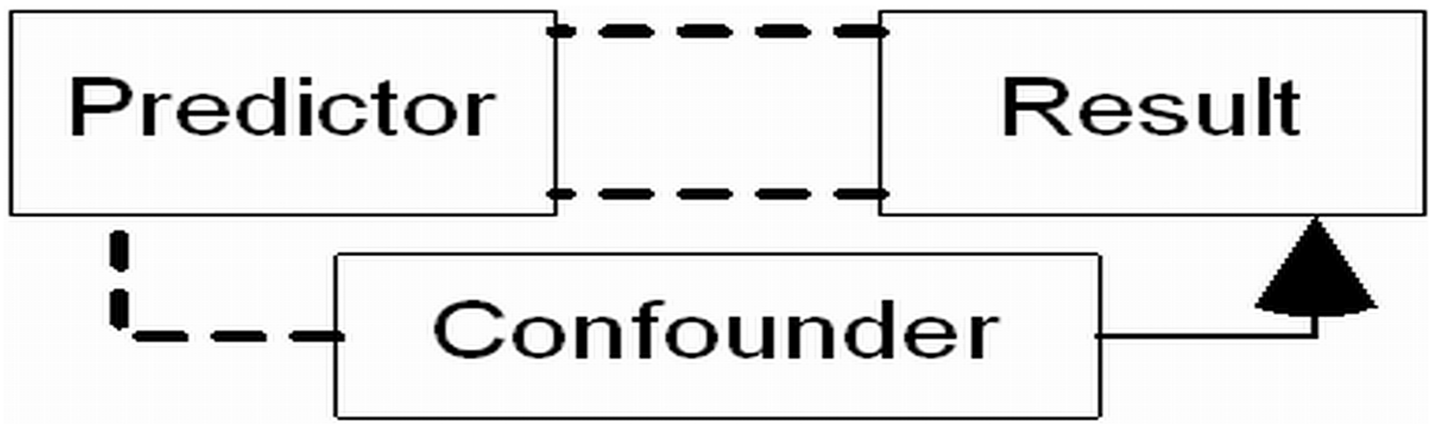
1. Statistics are different from numbers.
2. Statistics are between number and words.
3. Statistics are numbers in context –
   where the context matters.
4. Statistics are counts and measures of real things

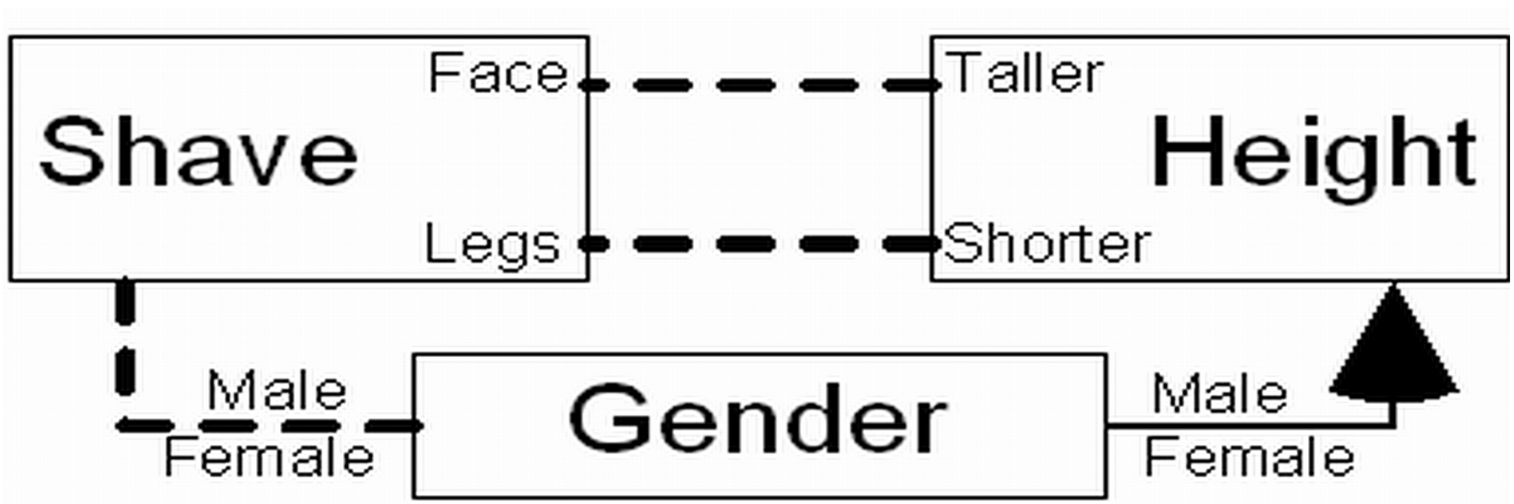
Consequence: Statistics can be influenced.
StatLit studies ALL the influences on a statistic

Statistical Literacy Studies ALL Influences



People that shave their face are taller…

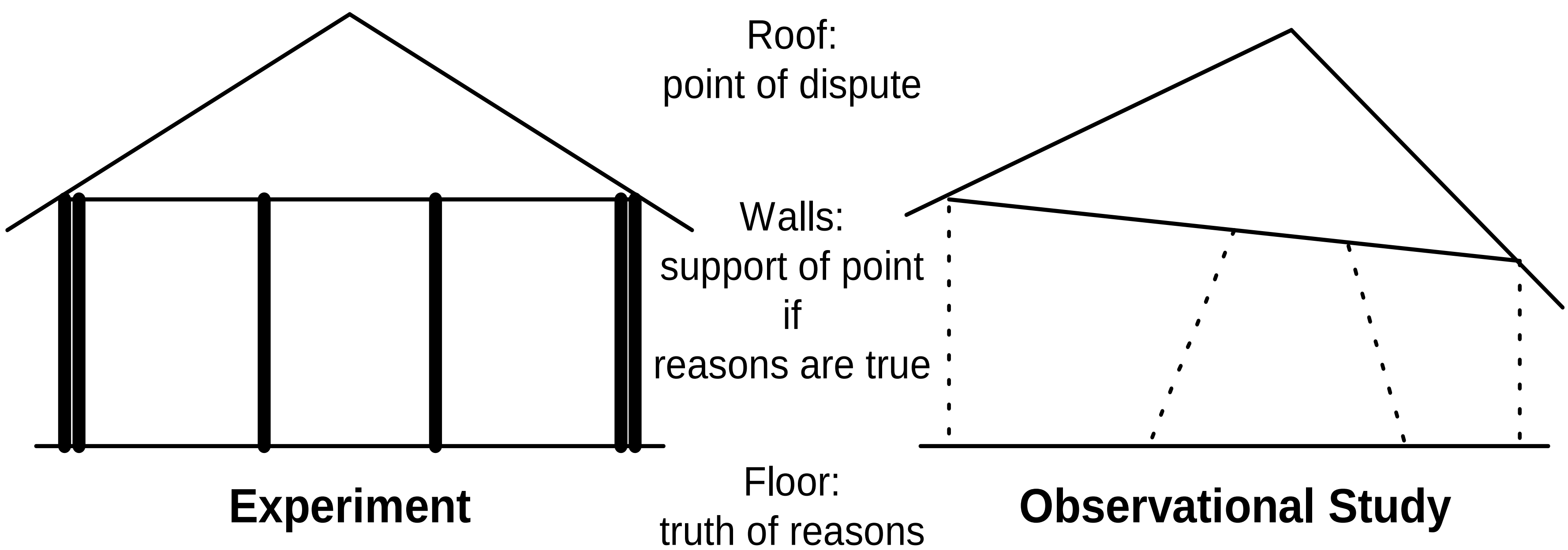




Study Design Can Ward Off Confounders

Experiments vs. Observational Study

Strength of Argument:
Support given by the reasons (premises) -- assuming they are true



Study Design Can Ward Off Confounders

	CONTROL OF CONFOUNDERS			
Physical Control (Grade = Quality)				
Experiment			Observational Study	
A+	Scientific		C	Longitudinal
A-	Random Assign		D	Cross-sectional
B	Quasi-Exper		F	Anecdotal story

Ward off Confounders: Quasi-Experiments

Quasi (Queasy)-Experiment

Nature or humans intervene on pre-existing groups

Nature intervenes

Epidemics
Plagues, outbreaks

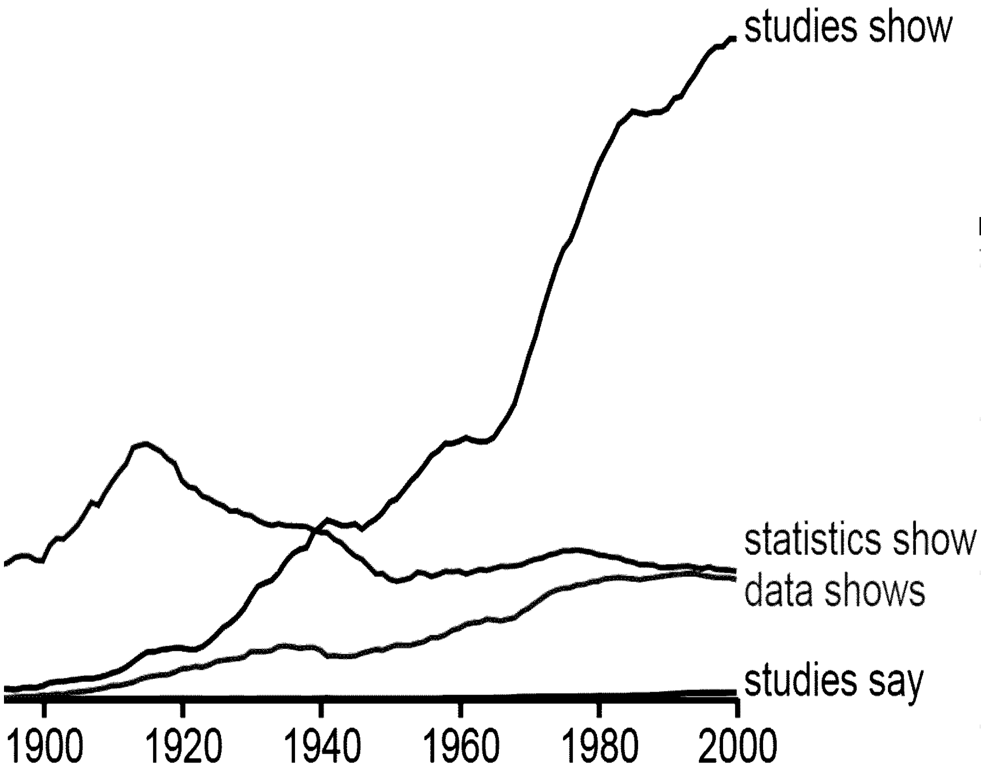
Natural disasters
Earthquakes, tornadoes

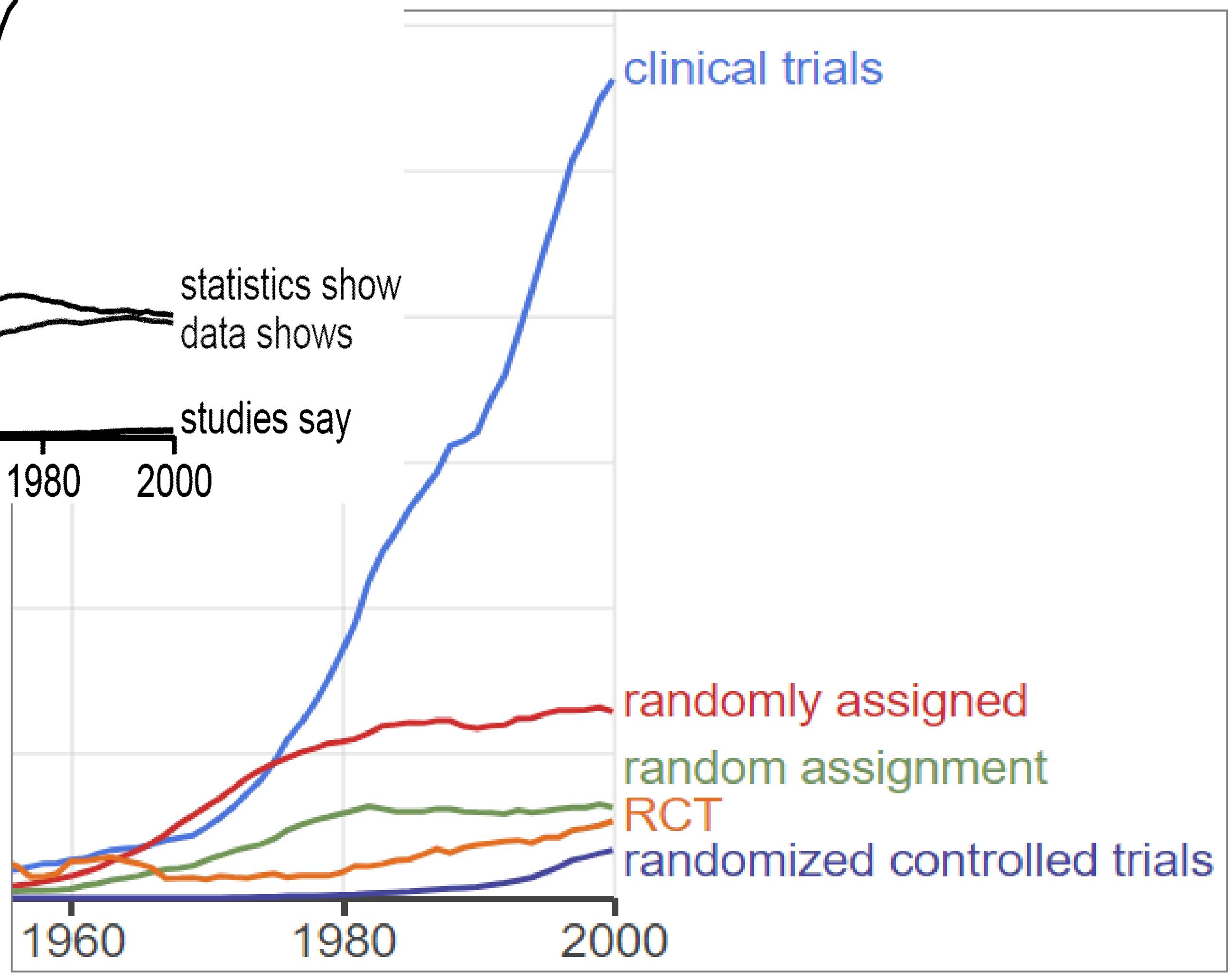
Humans intervene

Wars/Politics
Change laws & policies

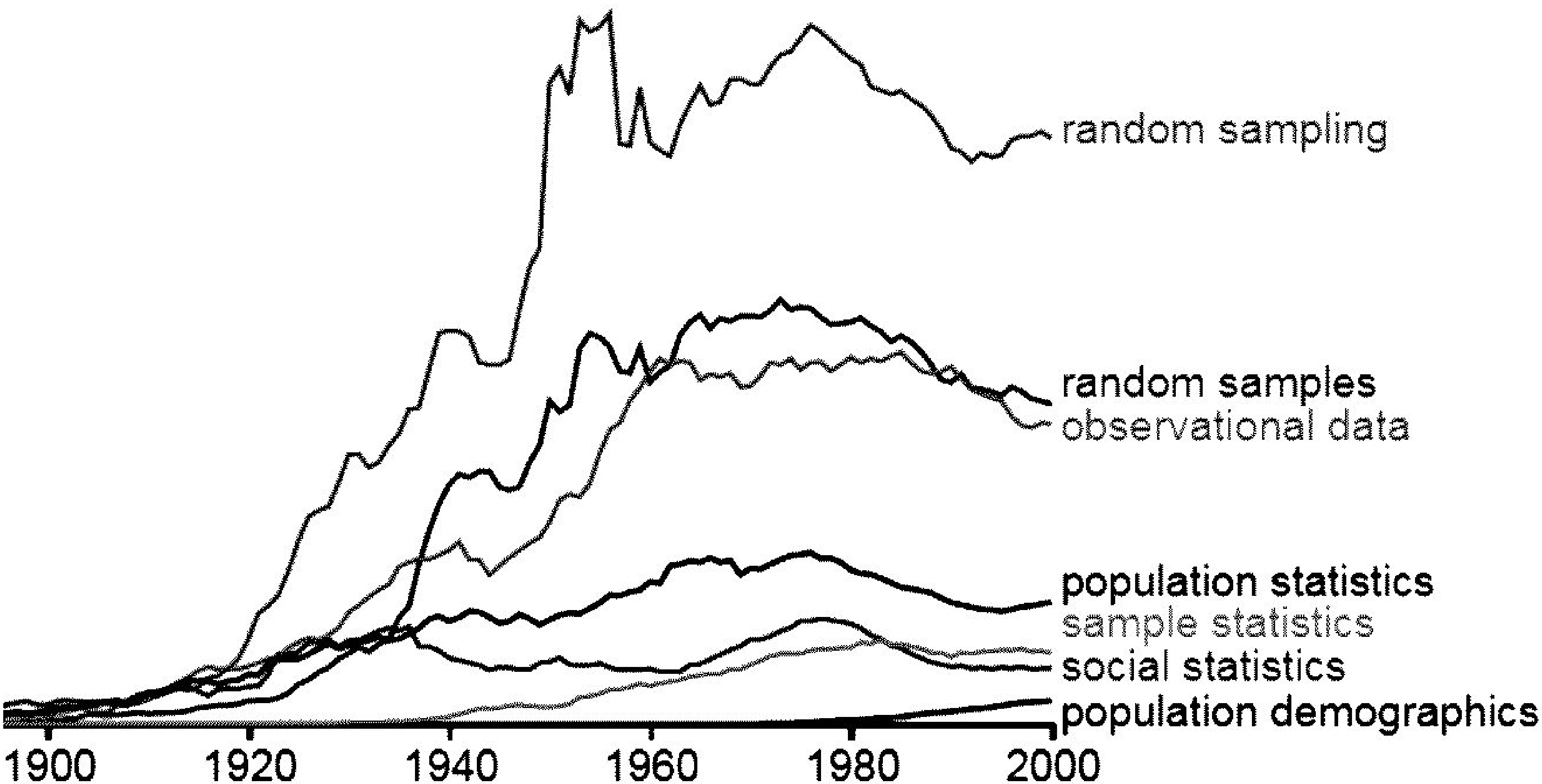
Business/Education
Change pricing/teaching

Prevalence Ngrams

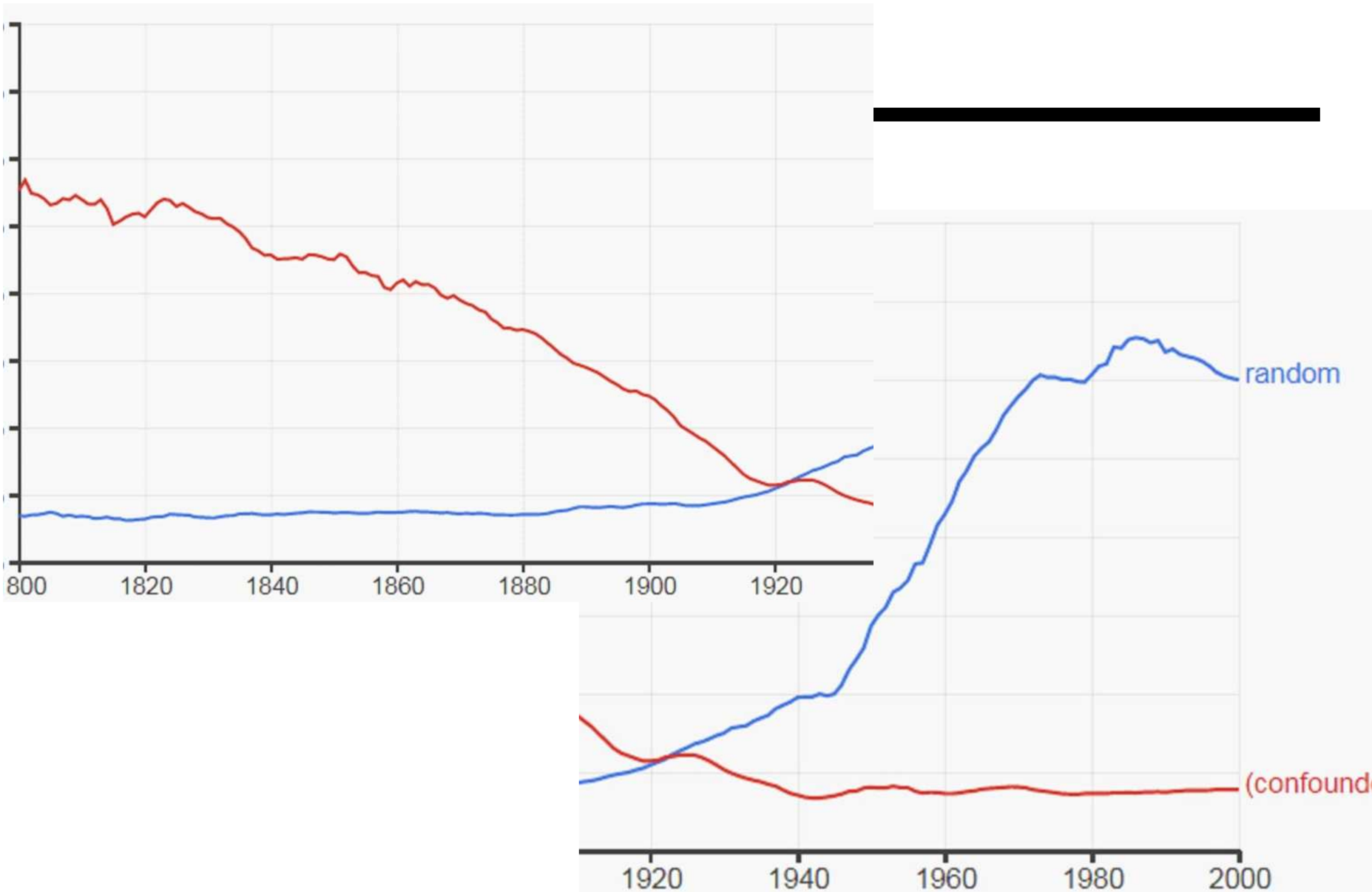




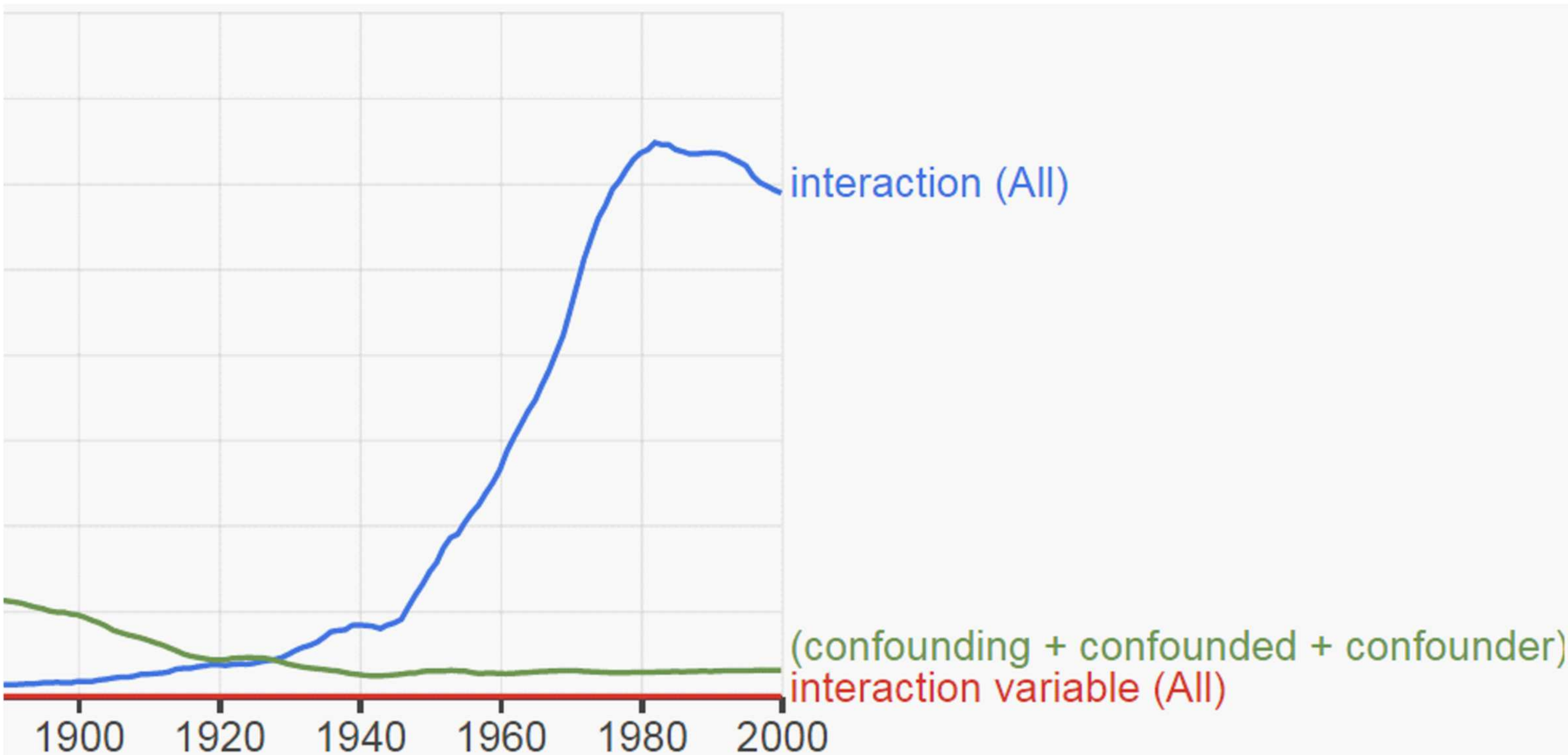
Prevalence Ngrams



Prevalence Ngrams



Prevalence Ngrams



Three Big Contributions to Human Knowledge

1. Association is not causation
2. The Central Limit theorem; the formula for Standard Error: statistical average error. Howard Wainer calls this “the most dangerous equation” (next to $E=mc^2$)
3. Fisher’s uses of random assignment to control for pre-existing confounders

Biggest Omissions relative to Human Knowledge

What are statistical educators biggest sins in teaching introductory statistics?

1. Ignoring multivariate data, observational studies and confounding.
2. Failing to show that controlling for a confounder can change statistical significance.
3. Ignoring the Cornfield conditions.
4. Ignoring how definitions can influence Stat. Sig.

Statistical Literacy: Four Biggest Problems

1. Lack of focus on confounding
2. Students
   - may become cynics about every statistic.
   - will have less respect for our discipline.
3. Teachers:
   Math/stat teachers: not trained to teach literacy.
   Math/stat teachers: don’t want to teach literacy.
4. Textbook and teacher training materials

Statistical Literacy and Confounding

Statistical literacy: the discipline that studies:
* all the influences on a statistic.

In observational studies, confounding is arguably the most common – most important – influence.

The statistical literacy "debate" is ultimately between the 'pro' and the 'anti' confounders.

Schield is – and has always been – pro-confounder. See Schield (1998) for "confounding factors".

Confounding Almost Absent in GAISE 2005

K-12 report: The first line: "The ultimate goal: statistical literacy". Confounding is mentioned twice: once to define and once to note it may create patterns that are not a "reliable basis for statistical inference".

College report: Confounding is mentioned only once. It is not defined; it appears in a sample problem in a list of words that may apply in analyzing data from an observational study.

Confounding mentioned in GAISE 2016 Update

Plus: Confounding shown 18 times (big increase):

Twice up front:

Goal 9: Ethics: "with large data sets, … under-standing confounding … becomes even more relevant." p 11

- Recommendation: Multivariable thinking. Examples "show how confounding plays an important role…" p.15

Nine times in appendix B: 34, 38 (3), 40 (2), 41(3)

Seven times later: Footnote 105; 113, 120, 122 (4).

Minus: Not in any one-line recommendations/goals

Silence on Smoking and Lung Cancer

Extremely important observational study.
Smoking is a most-likely cause of cancer.

MINUS:

- Not in most statistics textbooks.
- Not mentioned in GAISE 2005 College.

PLUS:

- Discussed in detail in GAISE 2004 K-12.
  But *confounding* was never used in the discussion

Why are we silent on confounding?

1. Confounding is not an issue in predicting.
2. There is no test for confounding. Judea Pearl
3. Using association as evidence for causation is a matter for subject-matter experts. Statisticians have no professional opinion on the subject.
4. Discussing confounding would bring disrepute on our discipline.

How can we change the present?

We need to go back to the past.

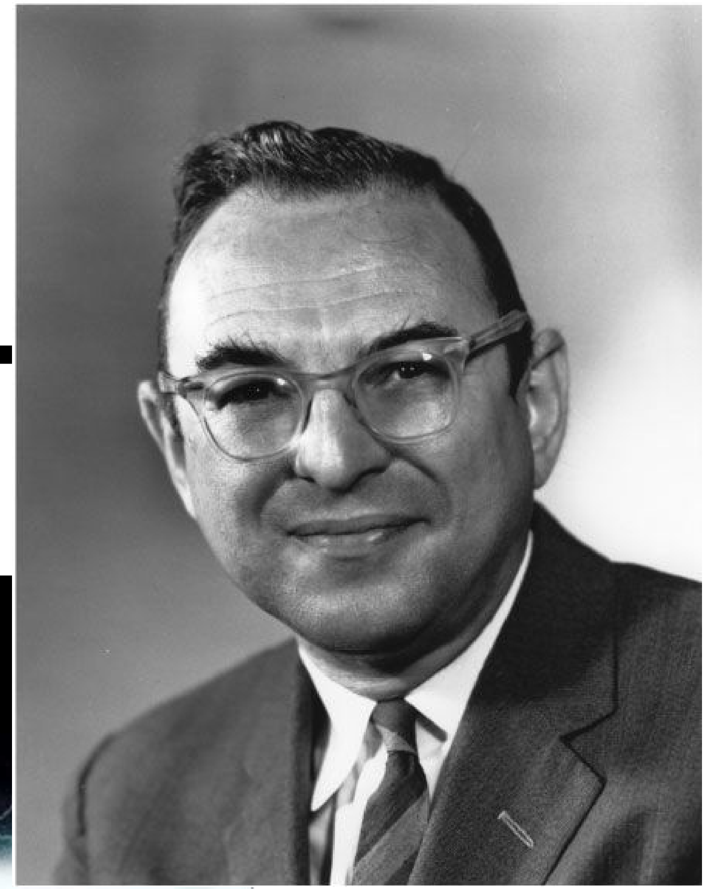
We need to revisit the Fisher-Cornfield dialogue on whether smoking caused lung cancer.

We need to revisit Cornfield's conditions for a confounder to nullify or reverse an association.

We need to see how to change statistical education to include Cornfield's criteria for confounding.

Back to the Future: Here we Go!

Back to 1958.



Back to the Future: Jerome Cornfield: 1912-1979

Jerome Cornfield got his BA and MA in history. He studied statistics at the US Dept of Agriculture.

He worked for USDA on sampling and study design

He created two common statistical measures: Relative risk (RR) and the Odds Ratio (OR).

He carfully compared prospective (cohort) and retrospective (case control) studies.

He was elected President of the ASA in 1974.

Back to the Future: Cornfield

Statisticians are subdued in talking about Cornfield; they are silent on the Cornfield conditions.

1. Not listed in RSS statistical timeline.
   Not listed in Wikipedia Timeline of Statistics
2. Wikipedia: Nothing on his work on confounding in the Smoking-Cancer studies.
3. Nothing in most of statisticians' commentaries about the Cornfield condition.

Why are We Silent on Confounding?

Because we don’t know Cornfield’s conditions!

Cornfield conditions: Minimum confounder size to nullify or reverse an observed association.

Impact: Allowed statisticians to say that “Smoking causes cancer” using data from an observational study.

Resist X% of Confounders

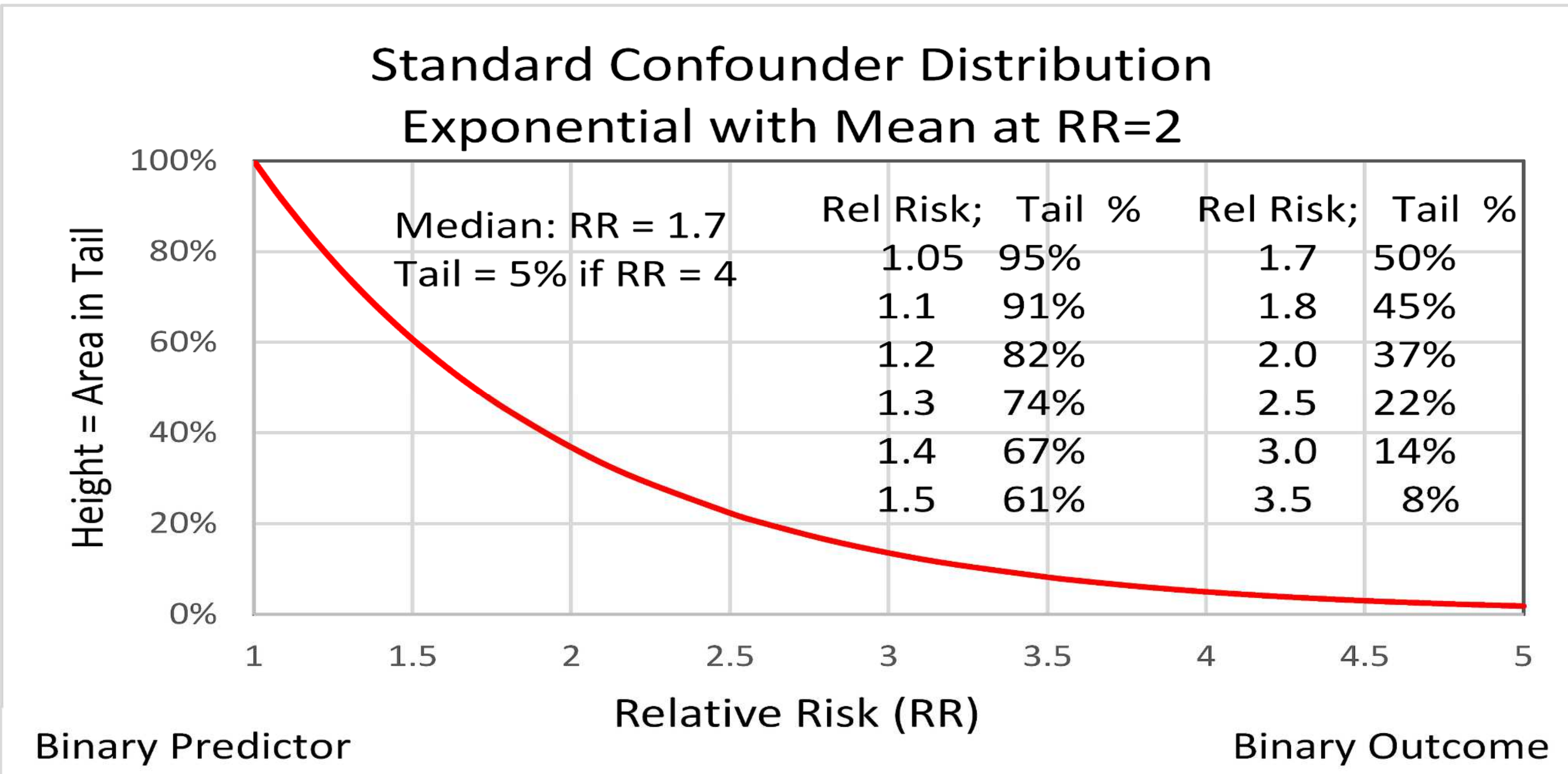
Confounders have no single analytical distribution. There is no way to say that a given effect size will resist X% of all relevant confounders.

But we can postulate a standard distribution of confounders: say an exponential distribution of relative risks with a mean of 2 (median of 1.69).

An RR of 4 will resist 95% of these standard confounders. 1.5 resists 40%; 1.2 resists 20%.

Confounder Resistance: Propose a Standard

Arbitrary, but simple and fits existing data.



Summary
Need Cornfield Conditions

With Cornfield conditions, we can

1. Show that the larger the effect size, the more resistance an association has to causation. (Schield and Burnham, 1998)
2. Show how to use Cornfield's conditions as necessary conditions. Schield (2012).
3. Show how to work problems controlling for a binary confounder. Schield ().

Conclusions

By featuring confounding in introductory statistics we can change our destiny. Statistical literacy can help untangle the confusion in many political debates.

Distinguishing between a crude association and a standardized association would be a big step forward

We are at a fork in the road. Which one will statistical educators take? Their choice will influence what most college graduates will study in decades to come.