Statistical Literacy

Mehmet Tevfik DORAK, MD PhD

Professor

School of Life Sciences, Pharmacy & Chemistry

Kingston University London

http://www.dorak.info

Kingston University London

**** Statistical Literacy ****

Harriet Hall: Science Based Medicine

JamesRandiFoundation

10 videos 94,775 views Last updated on Nov 26, 2015

Play all Shuffle

This course is presented by Harriet Hall and consists of 10 lectures:

Science-Based Medicine vs. Evidence-Based Medicine
What Is CAM?
Chiropractic
Acupuncture
Homeopathy
Naturopathy and Herbal Medicine
Energy Medicine
Miscellaneous "Alternatives"
Pitfalls in Research
Science-Based Medicine in the Media and Politics

A Course Guide is available at:
http://web.randi.org/uploads/3/7/3/7/37377621/c

1 

Lecture 1: Science Based Medicine vs.Evidence Based Medicine
JamesRandiFoundation • 97K views • 7 years ago

2 

Lecture 2: What is CAM
JamesRandiFoundation • 36K views • 7 years ago

3 

Lecture 3: Chiropractic
JamesRandiFoundation • 78K views • 7 years ago

4 

Lecture 4: Acupuncture
JamesRandiFoundation • 45K views • 7 years ago

5 

Lecture 5: Homeopathy
JamesRandiFoundation • 42K views • 7 years ago

6 

Lecture 6: Naturopathy and Herbal Medicines
JamesRandiFoundation • 37K views • 7 years ago

7 

Lecture 7: Energy Medicine
JamesRandiFoundation • 22K views • 7 years ago

8 

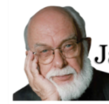
Lecture 8: Miscellaneous "Alternatives"
JamesRandiFoundation • 18K views • 7 years ago

9 

Lecture 9: Pitfalls in Research
JamesRandiFoundation • 14K views • 7 years ago

10 

Lecture 10: Science-Based Medicine in the Media and Politics
JamesRandiFoundation • 13K views • 7 years ago



James Randi Educational Foundation

HOME SWIFT ABOUT EDUCATION JREF PRESS

Search harriet hall

Filter
Everything (50)
Pages (8)
Blog posts (41)

Ask the SkepDoc: Cellphone antennas and electrodermal screening
The JREF is pleased to bring you Dr. *Harriet Hall* to answer your health-related questions about ...

Ask the Skepdoc: Pancreas miracle and Dispenza's thought healing
The JREF is pleased to bring you Dr. *Harriet Hall* to answer your health-related questions about ...

Ask the SkepDoc: Hepatitis shots, memory supplement and autism
The JREF is pleased to bring you Dr. *Harriet Hall* to answer your health-related questions about ...

Ask the Skepdoc: Qivana diet supplement claims and hyperthermi...
The JREF is pleased to bring you Dr. *Harriet Hall* to answer your health-related questions about ...

Last Week in Science-Based Medicine 22 September 2014
of The "Skepdoc" Dr. *Harriet* Hall.Here is a recap of the stories that appeared last week at Science ...

Last week in Science Based Medicine 27-October 2014
View of Medicine (*Harriet Hall*) Sandeep Jauhar wrote Doctored: The Disillusionment of an American ...

Science-based medicine content from TAM 2013
Harriet A. Hall, Skepdoc, and writer at the Science Based Medicine blog talked at TAM 2013 about ...

Last Week in Science-based Medicine 29 September 2014
By *Harriet* HallHere is a recap of the stories that appeared last week at Science-Based Medicine ...

Free Course: " Science Based Medicine" by Harriet Hall, MD - The S...
Medicine by *Harriet Hall*. James Randi said: It has never failed. Every time my path crosses that of Dr ...

Kingston University London

**** Statistical Literacy ****

More or Less: Behind the Stats

Home | More or Less on Radio 4 | More or Less on the World Service

More or Less

Tim Harford presents BBC Radio 4's surprising and refreshing guide to statistics in the news.

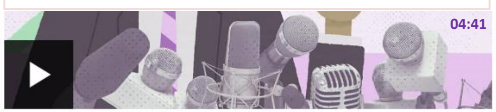
Mind Bending Statistics - Put Your Brain to the Test!

We might think that our brains are adept at dealing with numbers, possible scenarios and statistics, but how much can we really trust our intuition?

Take our statistics quiz, compiled by our OU academic experts, and test your brain against some puzzling probability problems - and learn the theory behind the (sometimes surprising!) results.

TEST YOUR BRAIN

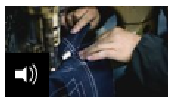
When can you trust statistics?



Tim Harford shares his tips for making sense of statistics.

Episodes to download

Are more adult nappies sold in Japan than baby ones?
SATURDAY
We examine a claim related to the perception of Japan as an ageing society.
Download

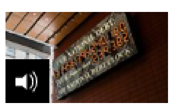


Does it take 10,000 litres of water to make a pair of jeans?
SAT 8 JUL 2023
The numbers behind water use and denim production.
Download

Immigration: A More or Less Special Programme
WED 5 JUL 2023
Who is coming to the UK - and how do we count them?
Download

Will there be just 6 grandchildren for every 100 South Koreans?
SAT 1 JUL 2023
We look at the numbers around South Korea's birth rate and the reasons why it's so low.
Download

Halving inflation, Scottish tidal power and have 1 in 3 women had an abortion?
WED 28 JUN 2023
Will Rishi Sunak's pledge to halve inflation happen?
Download



US National Debt: is $32 trillion a big number?
SAT 24 JUN 2023
How concerned should we be about how much the US government borrows?
Download

A short history of data
SAT 3 JUN 2023
How the evolution of data over centuries has shaped today's world
Download

Food prices, net migration and beef about beef
WED 31 MAY 2023
Does Britain really have the most affordable food in Europe?
Download

NHS waiting lists, Voter ID and measuring divorce
WED 24 MAY 2023
Why a fall in those waiting longest for treatment in England isn't telling the whole story
Download

Detecting Bad Science with Data
SAT 20 MAY 2023
Uncovering fraud, malpractice and incompetence in science
Download

Do 94% of marriages in Portugal really end in divorce?
SAT 13 MAY 2023
We examine a social media post that claims to outline rates of divorce across the world.
Download

Why is life expectancy falling in the USA?
SAT 6 MAY 2023
We ask why average life expectancy in America is falling so quickly.
Download

And many more episodes indefinitely available for downloading.....

Kingston University London

Statistical Literacy
Part I

Kingston
University
London

Statistical Literacy

90% of Medical Research is False !

Why Most Published Research Findings Are False

John P. A. Ioannidis

It can be proven that most claimed research findings are false.

If you want to know why and want to avoid false results or conclusions, read on!

Statistical literacy will make you a better thinker, scientist, and person who will not fall for misinformation, cheating and hoaxes.
It is not about learning statistics; it is about *statistical thinking*. Go for it!

Kingston University London

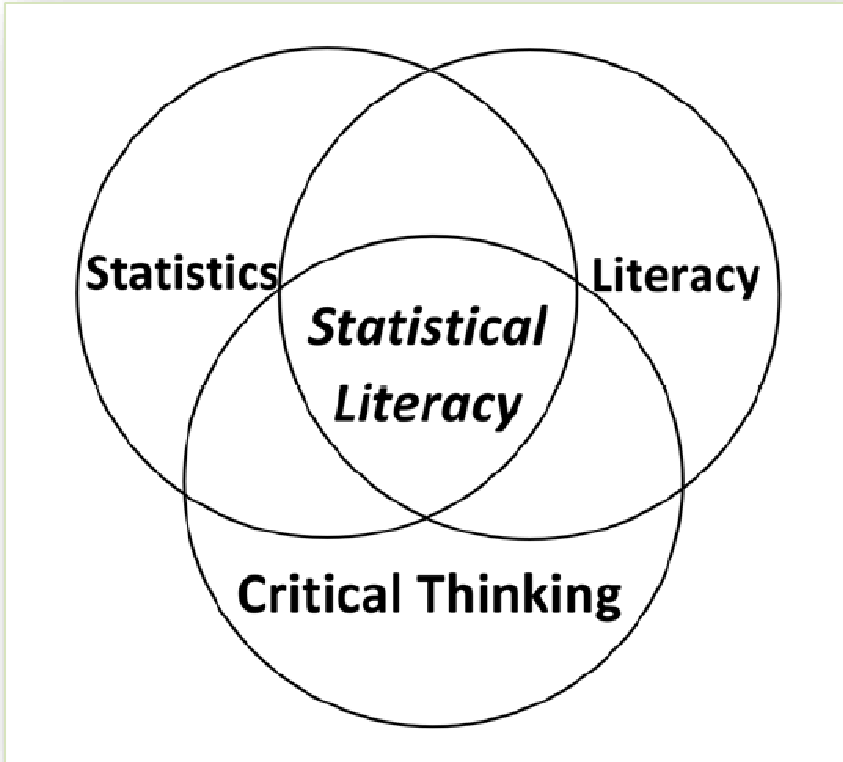
Statistical Literacy

- Statistical literacy as part of critical thinking
- Statistics and statistical thinking in our daily lives
- Statistical concepts we encounter every day
- Critical appraisal of information in the misinformation age; catching cheaters
- Unbiased thinking
- Considering confounding in causality assessment
- Thinking about effect modification and interaction
- Accuracy, precision, validity, reliability
- Sensitivity, specificity, predictive values
- Predictive value & prevalence issue
- Population level probabilities & individual
- Factfullness
- Statistical fallacies
- Statistical blunders
- Statistical Detectives: COM-PARE; StatCheck etc.

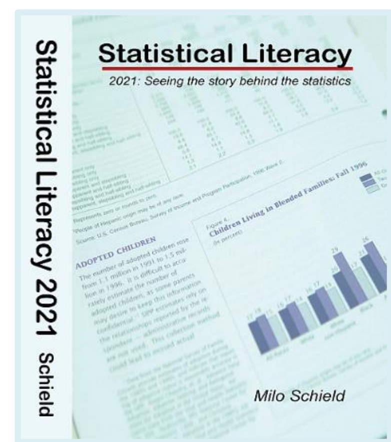
Statistical literacy will make you a better thinker, scientist, and person who will not fall for misinformation, cheating and hoaxes.
It is not about learning statistics; it is about *statistical thinking*. Go for it!

Kingston University London

Statistical Literacy



Statistical literacy will make you a better thinker, scientist, and person who will not fall for misinformation, cheating and hoaxes.
It is not about learning statistics; it is about *statistical thinking*. Go for it!



Kingston University London

Statistical Literacy

Statistical literacy skills include being able to read, understand, and communicate statistical information

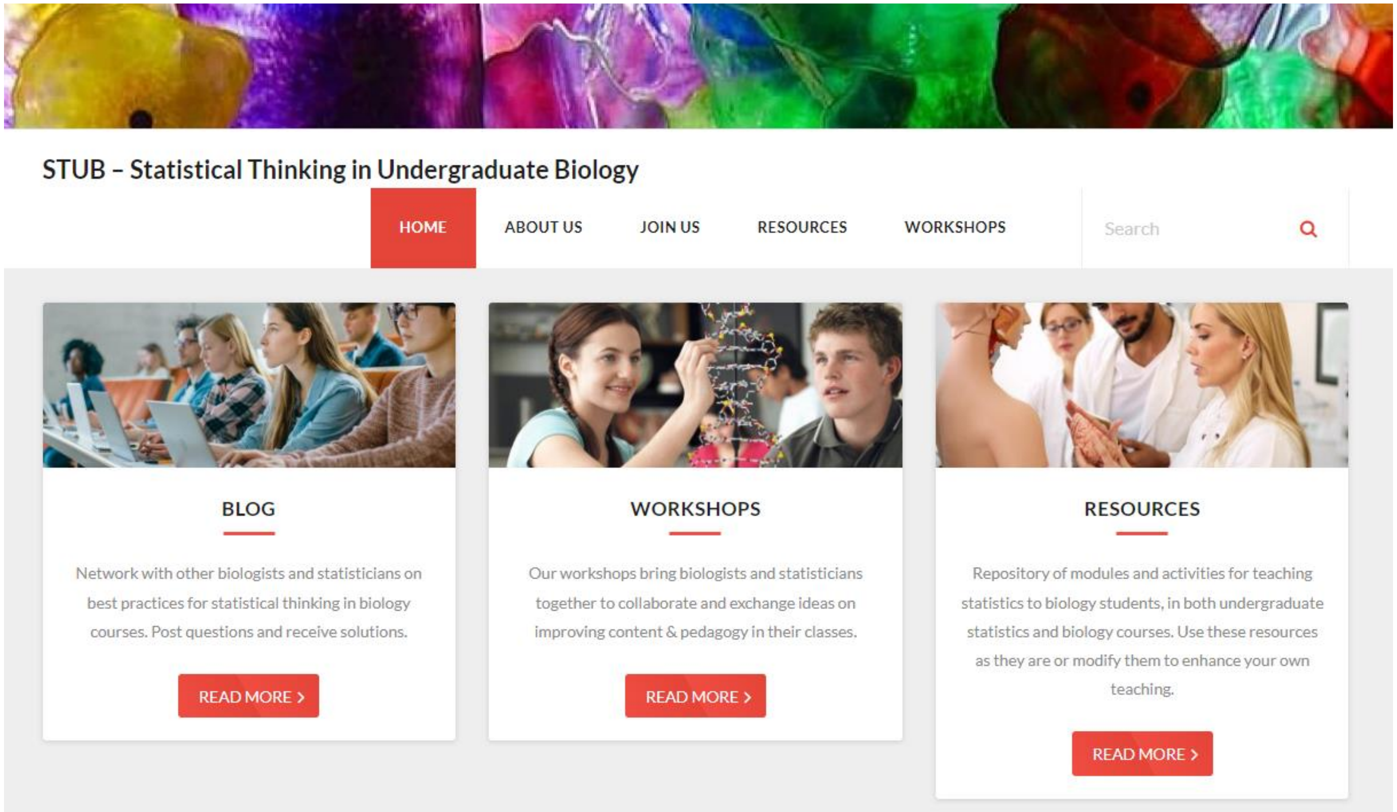
Critical Thinking at University: An Introduction

Critical thinking is a vital skill for university study whatever your discipline. Prepare for university now.

Statistical literacy will make you a better thinker, scientist, and person who will not fall for misinformation, cheating and hoaxes.
It is not about learning statistics; it is about *statistical thinking*. Go for it!

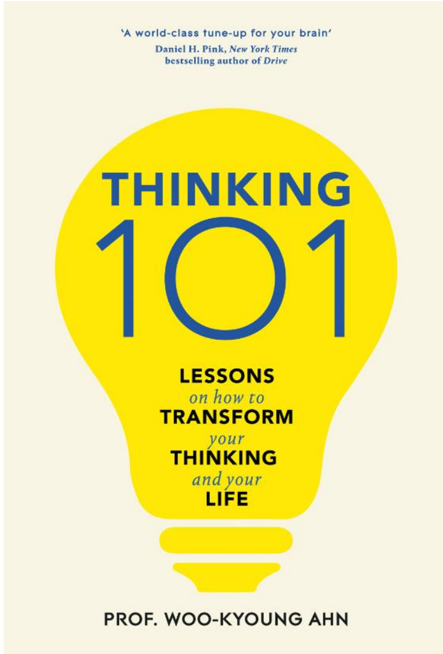
Kingston University London

Statistical Literacy



Resources: https://www.causeweb.org/stub/resources

Statistical Literacy



Statistical Literacy

Google

House of Commons Library Statistical Literacy Guide

UK Parliament

Home Research Resources Training and events About

House of Commons Library

UK Parliament > House of Commons Library > Research Briefing > Statistical Literacy Guide

Research Briefing

Statistical Literacy Guide

Published Thursday, 22 January, 2009

Research Briefing Parliament

Paul Bolton

This guide is a compilation of notes on individual subjects produced by the Social & General Statistics Section of the Library. Its aim is to help MPs and their staff better understand statistics that are used in press stories, research, debates, news releases, reports, books etc.

And …. More……….

Kingston University London

Statistical Literacy

INTERDISCIPLINARY IDEAS

Why Should We All Embrace Statistical Thinking?

[Data Literacy 101]

BY KRISTIN HUNTER-THOMSON

Way #1: We need to consider what the sample is every time we look at data

Way #2: We need to talk about uncertainty

Way #3: We support with evidence, not prove the hypothesis

Way #4: We can only make claims from the data we have, not what we want to have

Way #5: We need to think about whether a finding is truly meaningful

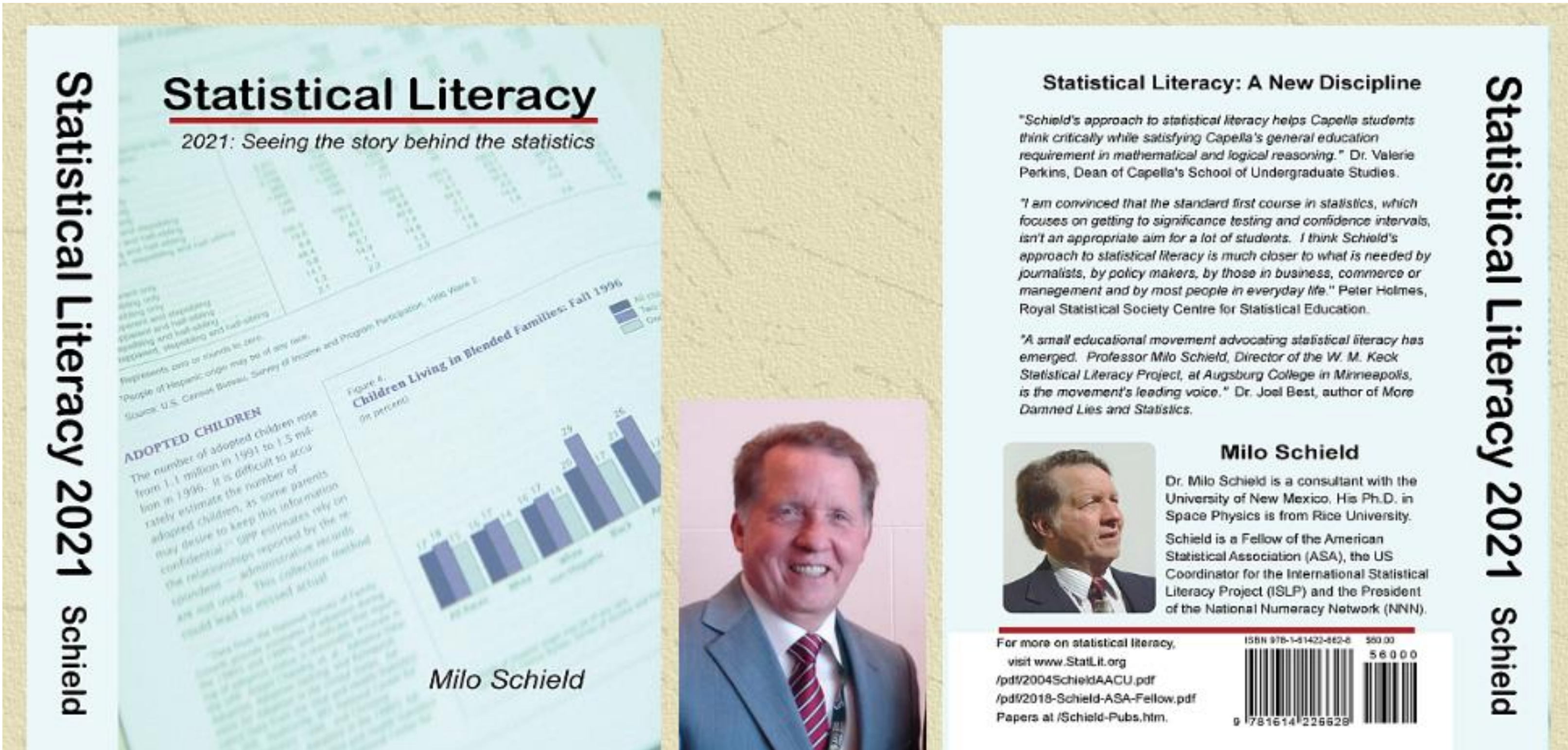
Statistical Literacy

Is That a Fact- A Field Guide for Evaluating Statistical and Scientific Information (2009)

https://www.amazon.co.uk/Evaluating-Statistical-Scientific-Information-2009-12-03/dp/B01FIY7ZJ6

Kingston University London

Statistical Literacy

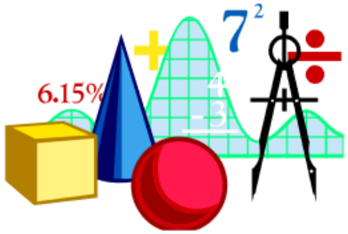


Statistical Literacy Textbook

CONTENTS **For chapter overview, download both audio and 6up overview. Play audio while watching the overview.**

Kingston University London

Statistical Literacy



STATISTICAL
LITERACY
2021B

Seeing the Story Behind the Statistics

Milo Schield

STATISTICAL LITERACY 2021B:
Seeing the Story behind the Statistics

Eighth Edition

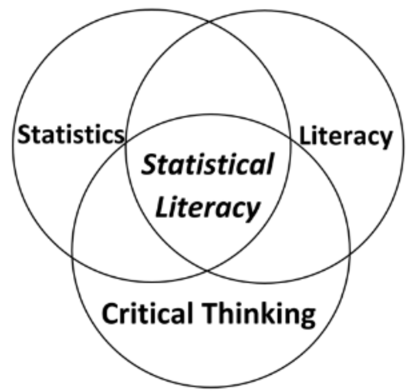
Printed in the U.S. by Instant Publisher

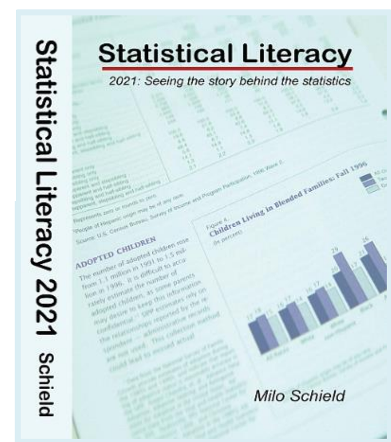
ISBN: 978-1-61422-671-0

File name: 2021B-SL4DM-V1F.docx Size: 172,280 kb Save: 12/20/2021

499 pages, 7901 paragraphs, 15,275 lines, 143,172 words, 732,405 characters.

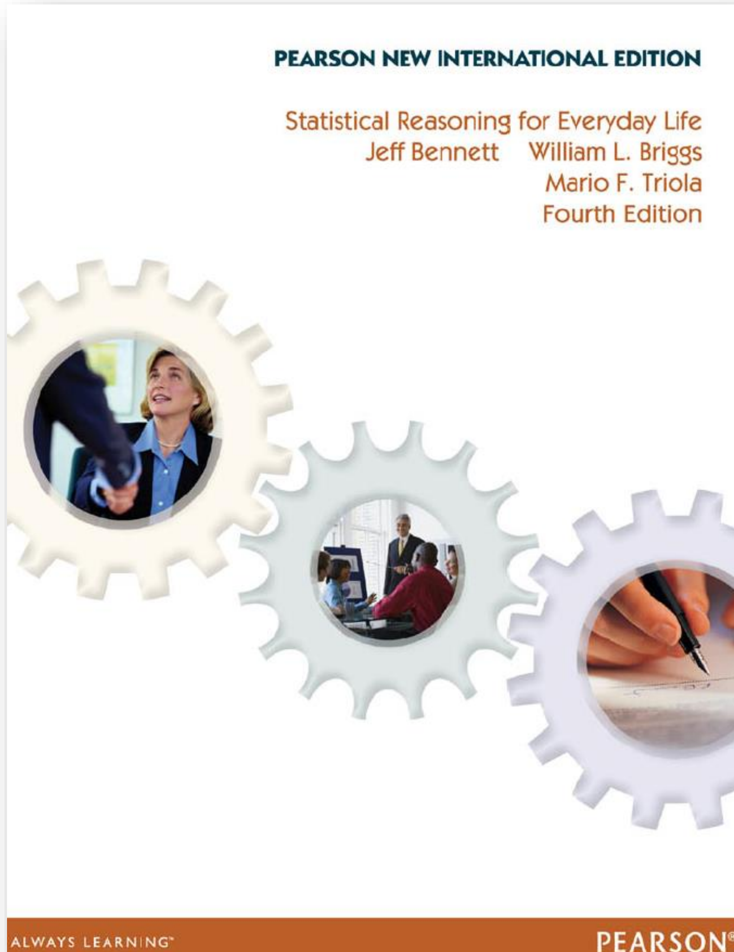
228 figures, 129 tables, 34 stories, 11 equations,
210 glossary terms, 420 index entries,





Kingston University London

Statistical Literacy



Kingston University London

Statistical Literacy

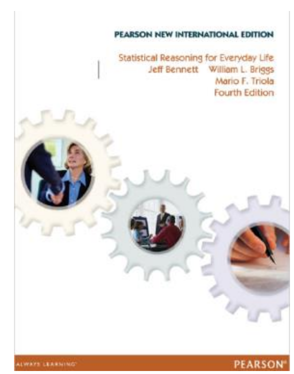
SHOULD YOU BELIEVE A STATISTICAL STUDY?

Already you know enough to achieve one of the major goals of this text: being able to answer the question "Should you believe a statistical study?"

Most researchers conduct their statistical studies with honesty and integrity, and most statistical research is carried out with diligence and care. Nevertheless, statistical research is sufficiently complex that bias can arise in many different ways, making it very important that we always examine reports of statistical research carefully. There is no definitive way to answer the question "Should I believe a statistical study?" However, in this section we'll look at eight guidelines that can be helpful. Along the way, we'll also introduce a few more definitions and concepts that will prepare you for discussions to come later.

Eight Guidelines for Critically Evaluating a Statistical Study

1. *Get a Big Picture View of the Study.* For example, you should understand the goal of the study, the population that was under study, and whether the study was observational or an experiment.
2. *Consider the Source.* In particular, look for any potential biases on the part of the researchers.
3. *Look for Bias in the Sample.* That is, decide whether the sampling method was likely to produce a representative sample.
4. *Look for Problems Defining or Measuring the Variables of Interest.* Ambiguity in the variables can make it difficult to interpret reported results.
5. *Beware of Confounding Variables.* If the study neglected potential confounding variables, its results may not be valid.
6. *Consider the Setting and Wording in Surveys.* In particular, look for anything that might tend to produce inaccurate or dishonest responses.
7. *Check That Results Are Presented Fairly.* For example, check whether the study really supports the conclusions that are presented in the media.
8. *Stand Back and Consider the Conclusions.* For example, evaluate whether study achieved its goals. If so, do the conclusions make sense and have practical significance?



Kingston University London

Statistical Literacy

SHOULD YOU BELIEVE A STATISTICAL STUDY?

CASE STUDY Cardiac Bypass Surgery

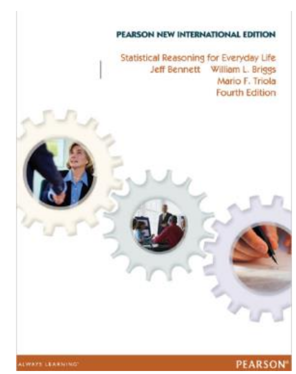
Cardiac bypass surgery is performed on people who have severe blockage of arteries that supply the heart with blood (the coronary arteries). If blood flow stops in these arteries, a patient may suffer a heart attack and die. Bypass surgery essentially involves grafting new blood vessels onto the blocked arteries so that blood can flow around the blocked areas. By the mid-1980s, many doctors were convinced that the surgery was prolonging the lives of their patients.

However, a few early retrospective studies turned up a disconcerting result: Statistically, the surgery appeared to be making little difference. In other words, patients who had the surgery seemed to be faring no better on average than similar patients who did not have it. If this were true, it meant that the surgery was not worth the pain, risk, and expense involved.

Because these results flew in the face of what many doctors thought they had observed in their own patients, researchers began to dig more deeply. Soon, they found confounding variables that had not been accounted for in the early studies. For example, they found that patients getting the surgery tended to have more severe blockage of their arteries, apparently because doctors recommended the surgery more strongly to these patients. Because these patients were in worse shape to begin with, a comparison of longevity between them and other patients was not really valid.

More important, the research soon turned up substantial differences in the results among patients who had the surgery in different hospitals. In particular, a few hospitals were achieving remarkable success with bypass surgery and their patients fared far better than patients who did not have the surgery or had it at other hospitals. Clearly, the surgical techniques used by doctors at the successful hospitals were somehow different and superior. Doctors studied the differences to ensure that all doctors could be trained in the superior techniques.

In summary, the confounding variables of *amount of blockage* and *surgical technique* had prevented the early studies from finding a real correlation between cardiac bypass surgery and prolonged life. Today, cardiac bypass surgery is accepted as a *cause* of prolonged life in patients with blocked coronary arteries. It is now among the most common types of surgery, and it typically adds *decades* to the lives of the patients who undergo it.



Statistical Literacy

SHOULD YOU BELIEVE A STATISTICAL STUDY?

Statistics need a critical eye

Rule 10

You *cannot* "prove anything with numbers"! Some people see numbers—any numbers—in an argument and conclude from that fact alone that it must be a good argument. Statistics seem to have an aura of authority and definiteness (and did you know that 88 percent of doctors agree?). In fact, though, numbers take as much critical thinking as any other kind of evidence. Don't turn off your brain!

> After an era when some athletic powerhouse universities were accused of exploiting student athletes, leaving them to flunk out once their eligibility expired, college athletes are now graduating at higher rates. Many schools are now graduating more than 50 percent of their athletes.

Fifty percent, eh? Pretty impressive! But this figure, at first so persuasive, does not really do the job it claims to do.

First, although "many" schools graduate more than 50 percent of their athletes, it appears that some do not—so this figure may well exclude the most exploitative schools that really concerned people in the first place.

The argument does offer graduation rates. But it would be useful to know how a "more than 50 percent" graduation rate compares with the graduation rate for *all* students at the same institutions. If it is significantly lower, athletes may still be getting the shaft.

Most importantly, this argument offers no reason to believe that college athletes' graduation rates are actually *improving*, because no comparison to any previous rate is offered! The conclusion claims that the graduation rate is now "higher," but without knowing the previous rates it is impossible to tell.

Numbers may offer incomplete evidence in other ways too. Rule 9, for example, tells us that knowing background rates may be crucial. Correspondingly, when an argument offers rates or percentages, the relevant background information usually must include the *number* of examples. Car thefts on campus may have doubled, but if this means that two cars were stolen rather than one, there's not much to worry about.

Another statistical pitfall is *over-precision*:

> Every year this campus wastes 412,067 paper and plastic cups. It's time to switch to reusable cups!

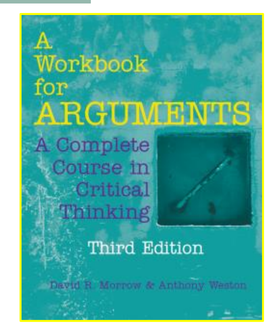
We're all for ending waste too, and we're sure the amount of campus waste is huge. But no one really knows the precise number of cups wasted—and it's extremely unlikely to be exactly the same every year. Here the appearance of exactness makes the evidence seem more authoritative than it really is.

Be wary, also, of numbers that are easily manipulated. Pollsters know very well that the way a question is asked can shape how it is answered. These days we are even seeing "polls" that try to change people's minds about, say, a political candidate, just by asking loaded questions ("If you were to discover that she is a liar and a cheat, how would that change your vote?"). Then too, many apparently "hard" statistics are actually based on guesswork or extrapolation, such as data about semi-legal or illegal activities. Since people have a major motive not to reveal or report things like drug use, under-the-counter transactions, hiring illegal aliens, and the like, beware of any confident generalizations about how widespread they are.

Yet again:

> If kids keep watching more TV at current rates, by 2025 they'll have no time left to sleep!

Right, and by 2040 they'll be watching thirty-six hours a day. Extrapolation in such cases is perfectly possible mathematically, but after a certain point it tells you nothing.



Kingston University London

Statistical Literacy

SHOULD YOU BELIEVE A STATISTICAL STUDY?

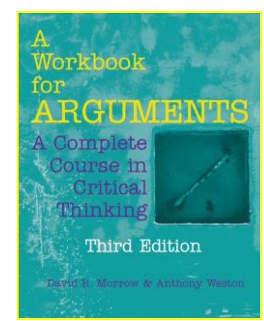
Sample

According to *U.S. News & World Report*'s compilation of statistics provided by law schools, 93 percent of law school graduates have a job nine months after finishing law school. That's up nearly ten percentage points from 1997, when law schools reported an average employment rate of 84 percent. The employment picture for law school graduates is better than ever!

Adapted from: David Segal, "Is Law School a Losing Game?" New York Times, *Jan 8, 2011, http://www.nytimes.com/2011/01/09/business/09law.html*

This argument cites two "employment rates" for recent law school graduates to show that the employment picture for law school graduates is "better than ever." There are several reasons to be skeptical of this argument. First of all, it's worth noting that these statistics come from the law schools themselves, who have an incentive to inflate employment rates. Second, the argument doesn't specify that 93 percent of graduates are employed as lawyers, which is what we really want to know about. It could be that half are employed as lawyers and 43 percent are flipping burgers and making cappuccinos. Third, the argument claims that the employment picture is "better than ever," but it offers only one point of comparison: 1997. It could be that 1997 was a particularly bad year for law school graduates. We would need more background information to evaluate the relevance of that statistic.

This response starts by explaining what the argument attempts to do with statistics. It then cites three reasons, related to those statistics, to be skeptical of the argument. Notice that the response doesn't give us a strong reason to think the conclusion itself is false. The upshot is that we just don't know. We would need to do more research to know whether law school graduates really do have good prospects. The point is that thinking critically about statistics can help prevent you from being taken in by misleading arguments.



Kingston University London

Statistical Literacy

SHOULD YOU BELIEVE A STATISTICAL STUDY?

The examples in previous slides show how you can apply statistical literacy to any information that presents numbers. Statistical literacy does not require advanced level statistics; it is about *statistical thinking*. It is not about critical appraisal of a statistics.

Statistical literacy will make you a better thinker, scientist, and person who will not fall for misinformation, cheating and hoaxes.
It is not about learning statistics; it is about *statistical thinking*. Go for it!

Kingston University London

Statistical Literacy

SHOULD YOU BELIEVE A STATISTICAL STUDY?

For more examples to begin with, visit the following links.
There will be many more during this course.

Misleading Statistics Examples – Discover The Potential For Misuse of Statistics & Data In The Digital Age

By Bernardita Calzon in Data Analysis, Jan 6th 2023

Famous Mistakes in Statistics

"A little knowledge is a dangerous thing," said Alexander Pope in 1711; he could have been speaking of the use of statistics by experts in all fields. In this article, we look at three consequential mistakes in the field of statistics. Two of them are famous, the third required a deep dive into the corporate annual reports of a U.S. software company.

Famous statistical wins and horror stories for teaching purposes

Kingston University London

Statistical Foundation

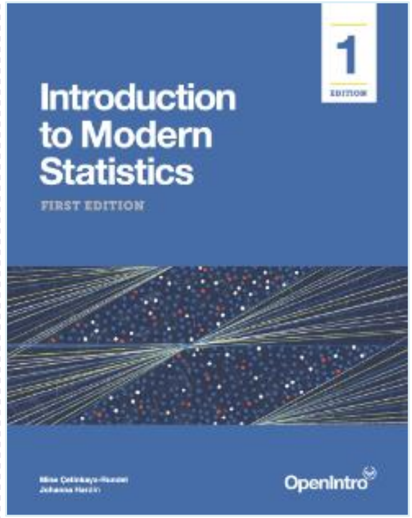
numworx

Library

- Numworx Modules Higher Education
  - Basic Math Modules
  - Math HE
  - Statistics
    - Introduction Statistics
      - 1 - Probabilities and variables
      - 2 - Descriptive statistics
      - 3 - Distributions
      - 4 - The normal distribution
      - 5 - Sampling
      - 6 - Statements about data
      - 7 - Hypothesis testing
      - 8 - Z-test and t-test
  - Old versions

Enhancing statistical literacy

Marianne van Dijke-Droogers[1], Paul Drijvers[2] and Jos Tolboom[3]



Download PDF

OpenIntro::Introduction to Modern Statistics Tutorials

Learnr tutorials supporting OpenIntro::Introduction to Modern Statistics. Each tutorial corresponds to one part of the book and is comprised of multiple lessons for a total of 35 lessons that cover the entire content of the book.

- Tutorial 1. Introduction to data
- Tutorial 2. Exploratory data analysis
- Tutorial 3. Regression modeling
- Tutorial 4. Foundations of inference
- Tutorial 5. Statistical inference
- Tutorial 6. Inferential modeling

Labs

1. Intro to R
2. Intro to data
3. Probability
4. Normal distribution
5. Foundations of inference
   a. Sampling distributions
   b. Confidence intervals
6. Inference for categorical data
7. Inference for numerical data
8. Simple linear regression
9. Multiple linear regression

OpenIntroStat / **openintro**

R package for data and supplemental functions for OpenIntro resources

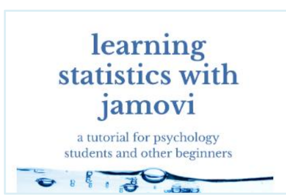
openintrostat.github.io/openintro

Online Statistics Education: An Interactive Multimedia Course of Study

Developed by Rice University , University of Houston Clear Lake, and Tufts University

Rice Virtual Lab in Statistics

Version in PDF



Statistics By Jim

Making statistics intuitive

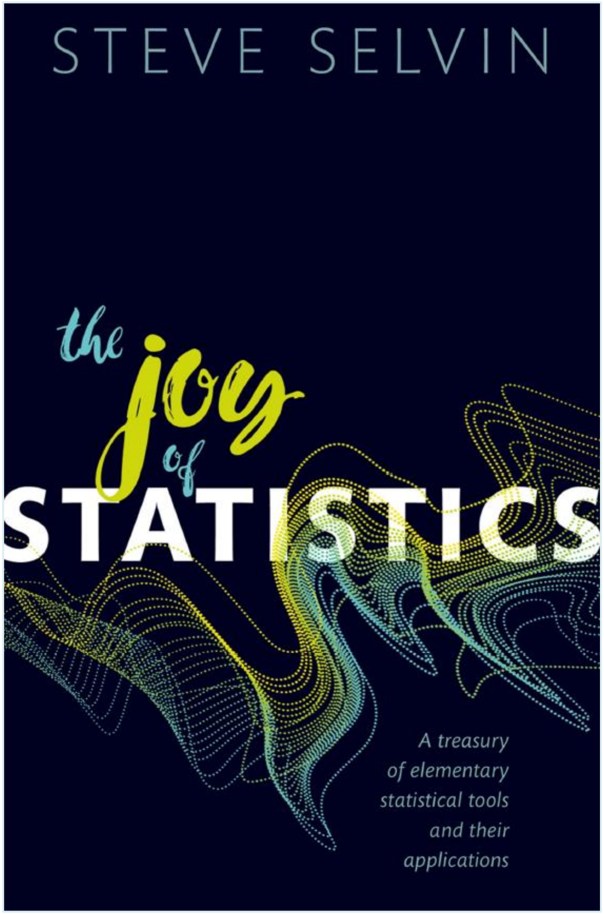
openstax Access. The future of education.

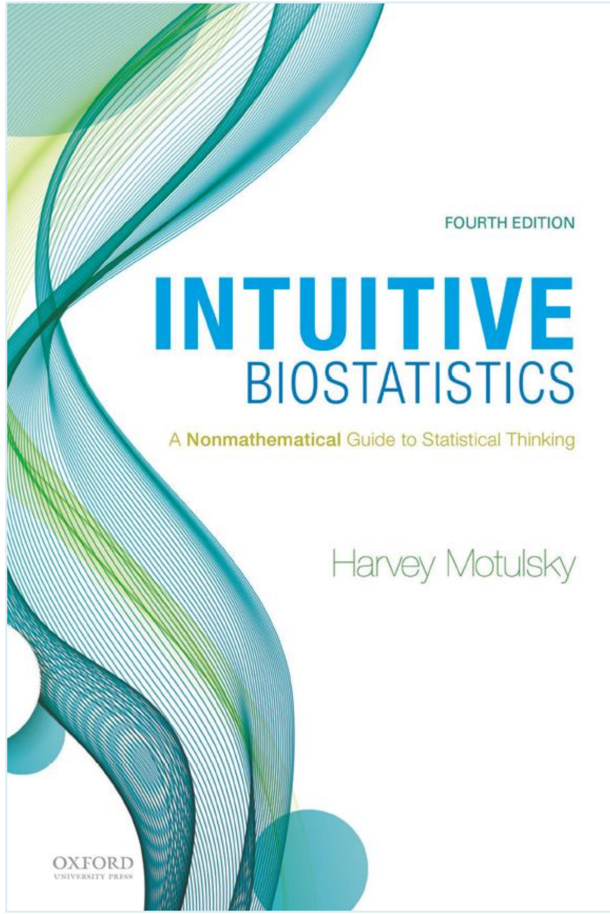
Introductory

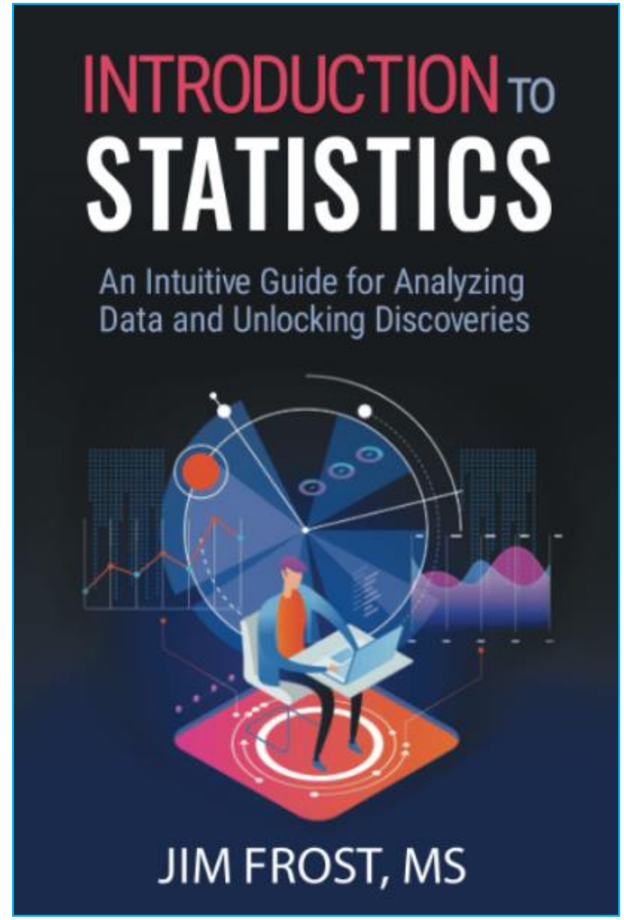
Statistics

Kingston University London

Statistical Foundation





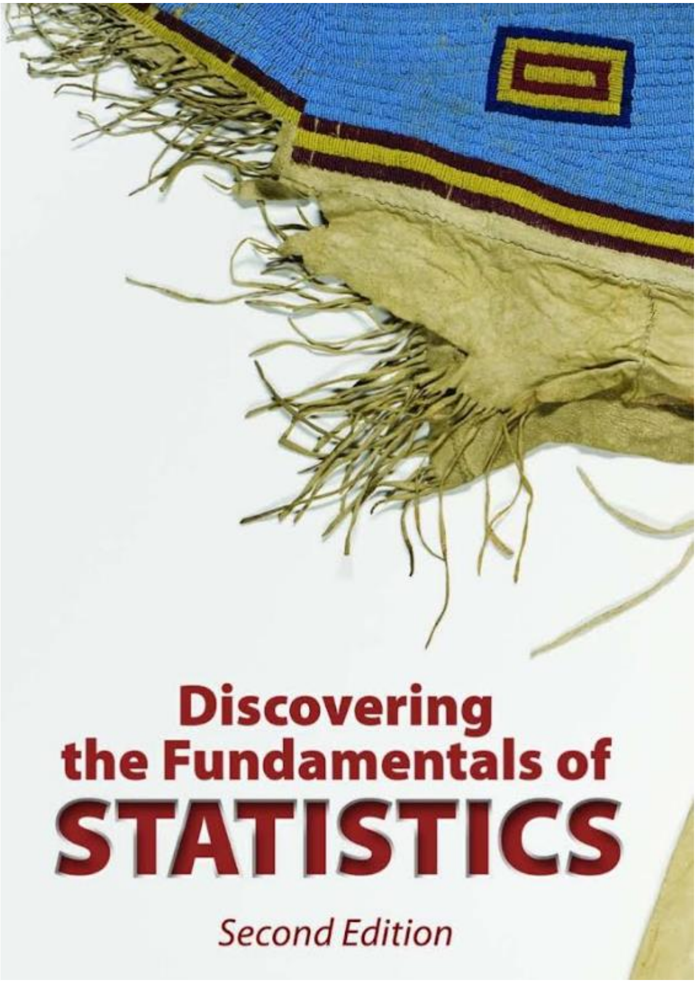


Statistics By Jim

Making statistics intuitive

Statistical Foundation

Statistical Background:



Statistical literacy: Thinking critically about statistics

As published in the Inaugural issue of the Journal "Of Significance"
Produced by the Association of Public Data Users
www.apdu.org

Milo Schield

Kingston University London

Statistical Literacy

udemy Categories Search for anything Udemy Business Teach on Udemy My learning

Teaching & Academics > Social Science > Statistics

Statistics and data literacy for non-statisticians

Learn the key terms and analysis methods in statistics

Free tutorial 4.6 ★★★★☆ (1,503 ratings) 20,594 students

1hr 36min of on-demand video

Created by Mike X Cohen

English English [Auto]

Go to course

Free online course by an expert on basic statistical background necessary for statistical literacy

Statistical Literacy

Co-learning Lounge

Sep 6, 2021 · 10 min read ★ · Listen

INTRODUCTION/STATISTICS

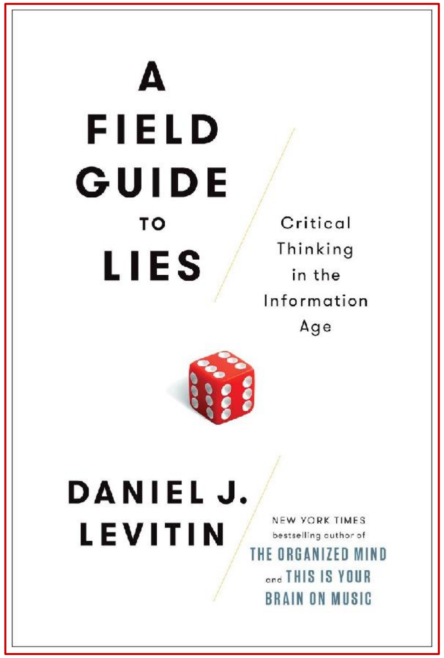
Statistics 101: A No BS Introduction for Dummies

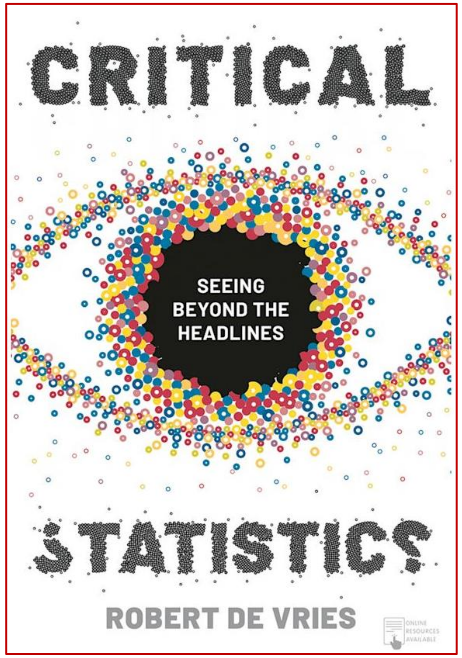
The only blog you need to understand statistics from scratch with real-life examples

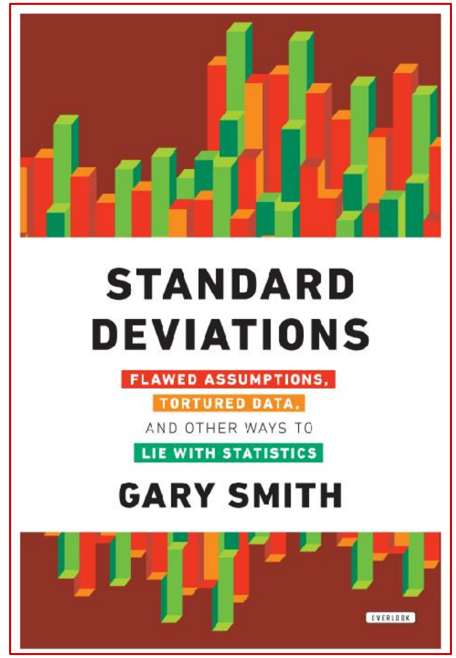
Written by Vaibhav Desai

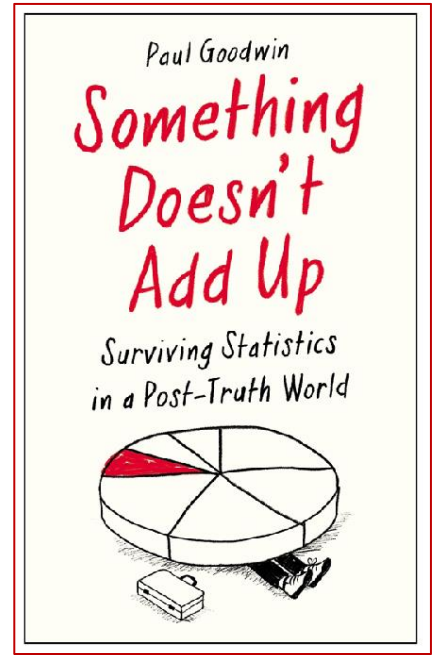
Another online course on basic statistical background necessary for statistical literacy

Statistical Literacy

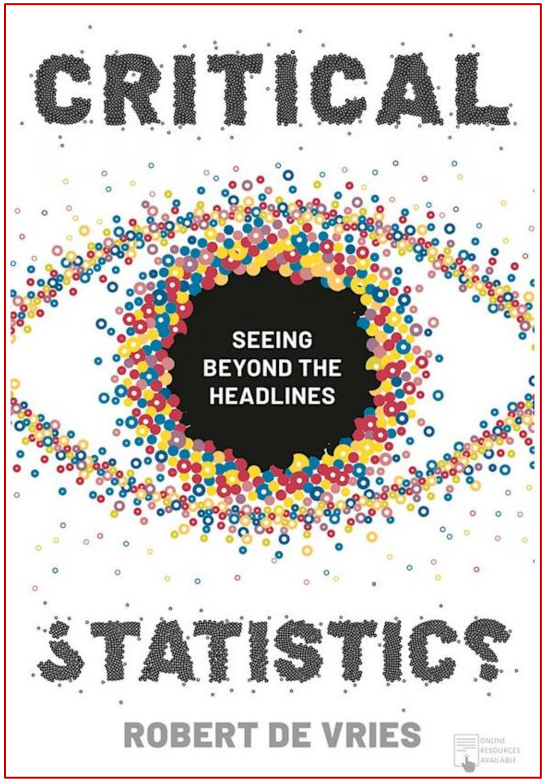








Statistical Literacy



Critical Statistics
Seeing Beyond the Headlines
by Robert de Vries

Home
Teaching Resources
Critical Statistics Course
Traditional Introductory Statistics Course
Student Resources
Exercise Material and Guidance
Extra Exercises
Other Resources
Submit your Own Examples
The Critical Statistics Hall of Shame
Updated Links

This accessible and entertaining new textbook provides students with the knowledge and skills they need to understand the barrage of numbers encountered in their everyday lives and studies. Almost all the statistics in the news, on social media or in scientific reports are based on just a few core concepts, including measurement (ensuring we count the right thing), causation (determining whether one thing causes another) and sampling (using just a few people to understand a whole population).

By explaining these concepts in plain language, without complex mathematics, this book prepares students to meet the statistical world head on and to begin their own quantitative research projects.Ideal for students facing statistical research for the first time, or for anyone interested in understanding more about the numbers in the news, this textbook helps students to see beyond the headlines and behind the numbers.

- Student Resources
- Teaching Resources

Statistical Literacy

Statistics Are Being Abused, but Mathematicians Are Fighting Back

An expert explains how numbers can mislead and what she's doing to help people understand them better

A scientist's opinion: interview with Ionica Smeets on hype in press releases

Ionica Smeets

SCIENCE COMMUNICATION & SOCIETY

Image by Sija van den Beukel

Professor of Science Communication - Leiden University

Home Publications Books Contact Dutch website

Ionica Smeets is the chair of Leiden University's research group **Science Communication and Society**. In this role she also enthusiastically teaches in the **master specialization Science Communication and Society**. Her main research interest in science communication is the gap between experts and the general public. What problems occur when those groups communicate? And what can scientists do about those problems?

The danger of mixing up causality and correlation: Ionica Smeets at TEDxDelft

TEDx Talks
37.9M subscribers
Subscribe
4K Share Download Clip Save

554K views 10 years ago
Ionica Smeets (@ionicasmeets) is joining TEDxDelft Never Grow Up: A mathematician and science journalist with plenty of media experience. Using her vast knowledge and enthusiasm, she can explain everything about her favorite topics in science and statistics. She does it well on paper and face-to-face: She writes blogs, columns and books and is also asked to appear as a speaker, live, on television and on radio shows. Show more

Ionica Smeets
Professor of Science Communication, Leiden University
Verified email at biology.leidenuniv.nl - Homepage
science communication public understanding of sci... mathematics

Google Scholar

Kingston University London

Statistical Literacy

There are even software packages that would check the statistics in published papers

Statcheck

Add languages

Article Talk | Read Edit View history

From Wikipedia, the free encyclopedia

Statcheck is an R package designed to detect statistical errors in peer-reviewed psychology articles[1] by searching papers for statistical results, redoing the calculations described in each paper, and comparing the two values to see if they match.[2] It takes advantage of the fact that psychological research papers tend to report their results in accordance with the guidelines published by the American Psychological Association (APA).[3] This leads to several disadvantages: it can only detect results reported completely and in exact accordance with the APA's guidelines,[4] and it cannot detect statistics that are only included in tables in the paper.[5] Another limitation is that Statcheck cannot deal with statistical corrections to test statistics, like Greenhouse–Geisser or Bonferroni corrections, which actually make tests more conservative.[6] Some journals have begun piloting Statcheck as part of their peer review process. Statcheck is free software published under the GNU GPL v3.[7]

Validity [edit]

In 2017, Statcheck's developers published a preprint paper concluding that the program accurately identified statistical errors over 95% of the time.[8] This validity study comprised more than 1,000 hand-checked tests among which 5.00% turned out to be inconsistent.[9] The study found that Statcheck recognized 60% of all statistical tests. A reanalysis of these data found that if the program flagged a test as inconsistent, it was correct in 60.4% of cases. Reversely, if a test was truly inconsistent, Statcheck flagged it in an estimated 51.8% of cases (this estimate included the undetected tests and assumed that they had the same rate of inconsistencies as the detected tests). Overall, Statcheck's accuracy was 95.9%, half a percentage point higher than the chance level of 95.4% expected when all tests are simply taken at face value. Statcheck was conservatively biased (by about one standard deviation) against flagging tests.[10]

More recent research has used Statcheck on papers published in Canadian psychology journals, finding similar rates of statistical reporting errors as the original authors based on a 30-year sample of such articles. The same study also found many typographical errors in online versions of relatively old papers, and that correcting for these reduced the estimated percent of tests that were erroneously reported.[11]

See also [edit]

- Abuse of statistics
- Misuse of p-values
- Metascience

Online version:

statcheck

statcheck on the web

To check a PDF, DOCX or HTML file for errors in statistical reporting, upload it below. See the FAQ page for more information about what statcheck can and cannot do.

Upload files (pdf, html, or docx):

Browse... No file selected

Kingston University London

Statistical Literacy

Stanford | ONLINE   Search Catalog   Topics   Programs   Schools & Centers   Free Content   About Us   Get More Info

Home » How to Be a Statistical Detective

How to Be a Statistical Detective

Mark Twain famously once said, “There are three kinds of lies: lies, damned lies and statistics.” People tend to view statistics with a bit of skepticism, recognizing that figures can be compiled and presented in any number of ways to support a particular argument.

According to Stanford University Associate Professor Kristin Sainani, that reputation is somewhat merited, given the abundance of misleading statistical information out in the world today. Professor Sainani recently led a webinar to discuss some of the tools of the trade that she uses to dig into statistics and find the truth, regardless of what the numbers appear to say.

statcheck

statcheck on the web

To check a PDF, DOCX or HTML file for errors in statistical reporting, upload it below. See the FAQ page for more information about what statcheck can and cannot do.

Upload files (pdf, html, or docx):

Browse... No file selected

PM&R   WILEY

Statistically Speaking | Free Access

How to Be a Statistical Detective

Kristin L. Sainani PhD

Manual statcheck 1.3.0

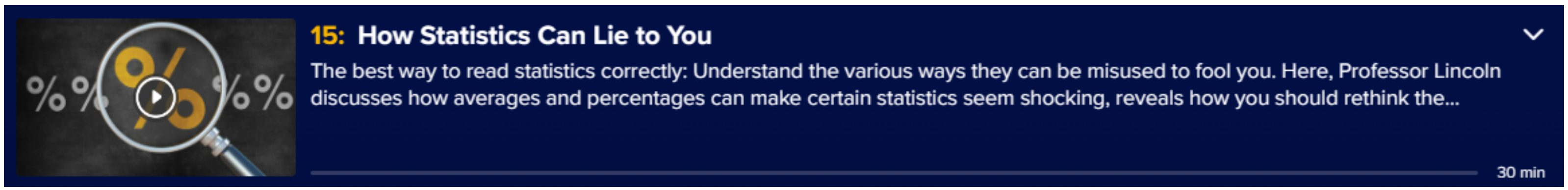
Michèle Nuijten (m.b.nuijten@uvt.nl)

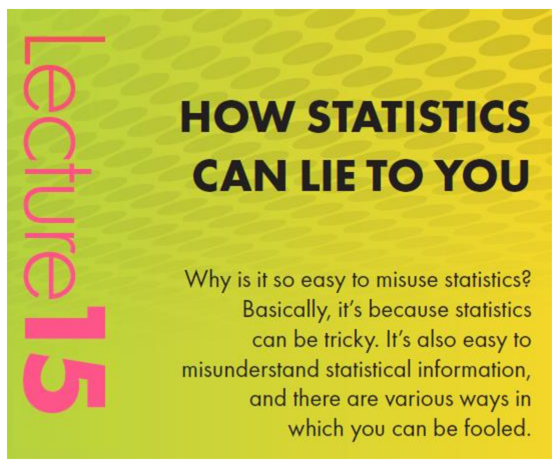
July 2018

statcheck

Kingston University London

Statistical Literacy





Kingston University London

Statistical Literacy

An Online Statistics Book

Statistical Literacy

by David M. Lane

Prerequisites

- Chapter 19:

This article describes some health effects of drinking coffee. Among the key findings were (a) women who drank four or more cups a day reduced their risk of endometrial cancer by 25% compared with those who drank less than one cup a day and (b) men who drank six or more cups had a 60% lower risk of developing the most deadly form of prostate cancer than those who drank less than one cup a day.

Each chapter ends with a section like this!

What do you think?

What is the technical term for the measure of risk reduction reported? What measures of risk reduction cannot be determined from the article? What additional information would have been helpful for assessing risk reduction?

This is called the "relative risk reduction." The article does not provide information necessary to compute the absolute risk reduction, the odds ratio, or the number needed to treat. It would have been helpful if the article had reported the proportion of women drinking less than one cup a day who developed endometrial cancer as well as the analogous statistic for men and prostate cancer.

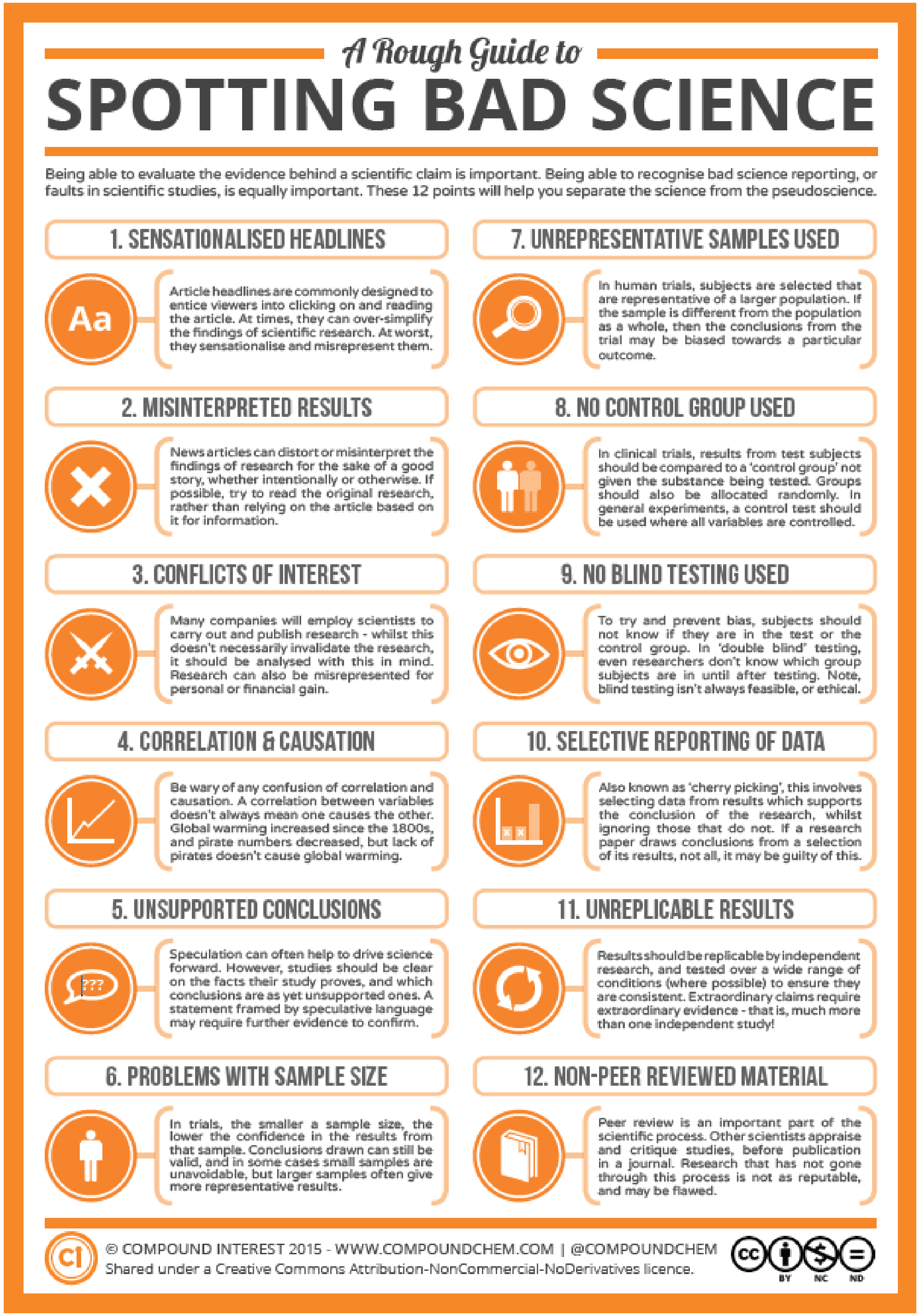
Introduction to Statistics

Online Edition

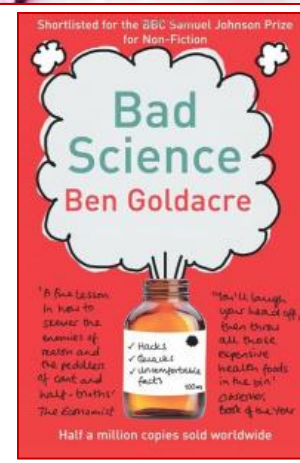
Primary author and editor:
David M. Lane[1]

Other authors:
David Scott[1], Mikki Hebl[1], Rudy Guerra[1], Dan Osherson[1], and Heidi Zimmer[2]

[1]Rice University; [2]University of Houston, Downtown Campus

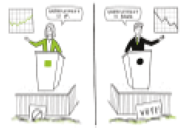


See also:



Kingston University London

DATA FALLACIES TO AVOID

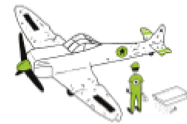


CHERRY PICKING

Selecting results that fit your claim and excluding those that don't.

DATA DREDGING

Repeatedly testing new hypotheses against the same set of data, failing to acknowledge that most correlations will be the result of chance.

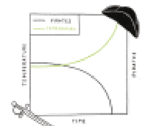


SURVIVORSHIP BIAS

Drawing conclusions from an incomplete set of data, because that data has 'survived' some selection criteria.

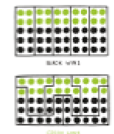
COBRA EFFECT

Setting an incentive that accidentally produces the opposite result to the one intended. Also known as a Perverse Incentive.



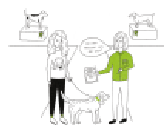
FALSE CAUSALITY

Falsely assuming when two events appear related that one must have caused the other.



GERRYMANDERING

Manipulating the geographical boundaries used to group data in order to change the result.

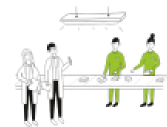


SAMPLING BIAS

Drawing conclusions from a set of data that isn't representative of the population you're trying to understand.

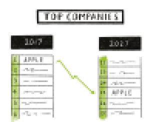
GAMBLER'S FALLACY

Mistakenly believing that because something has happened more frequently than usual, it's now less likely to happen in future (and vice versa).



HAWTHORNE EFFECT

The act of monitoring someone can affect their behaviour, leading to spurious findings. Also known as the Observer Effect.

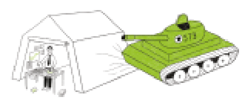


REGRESSION TOWARDS THE MEAN

When something happens that's unusually good or bad, it will revert back towards the average over time.

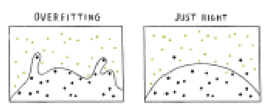
SIMPSON'S PARADOX

When a trend appears in different subsets of data but disappears or reverses when the groups are combined.



MCNAMARA FALLACY

Relying solely on metrics in complex situations and losing sight of the bigger picture.

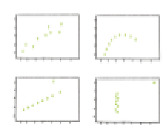


OVERFITTING

Creating a model that's overly tailored to the data you have and not representative of the general trend.

PUBLICATION BIAS

Interesting research findings are more likely to be published, distorting our impression of reality.



DANGER OF SUMMARY METRICS

Only looking at summary metrics and missing big differences in the raw data.

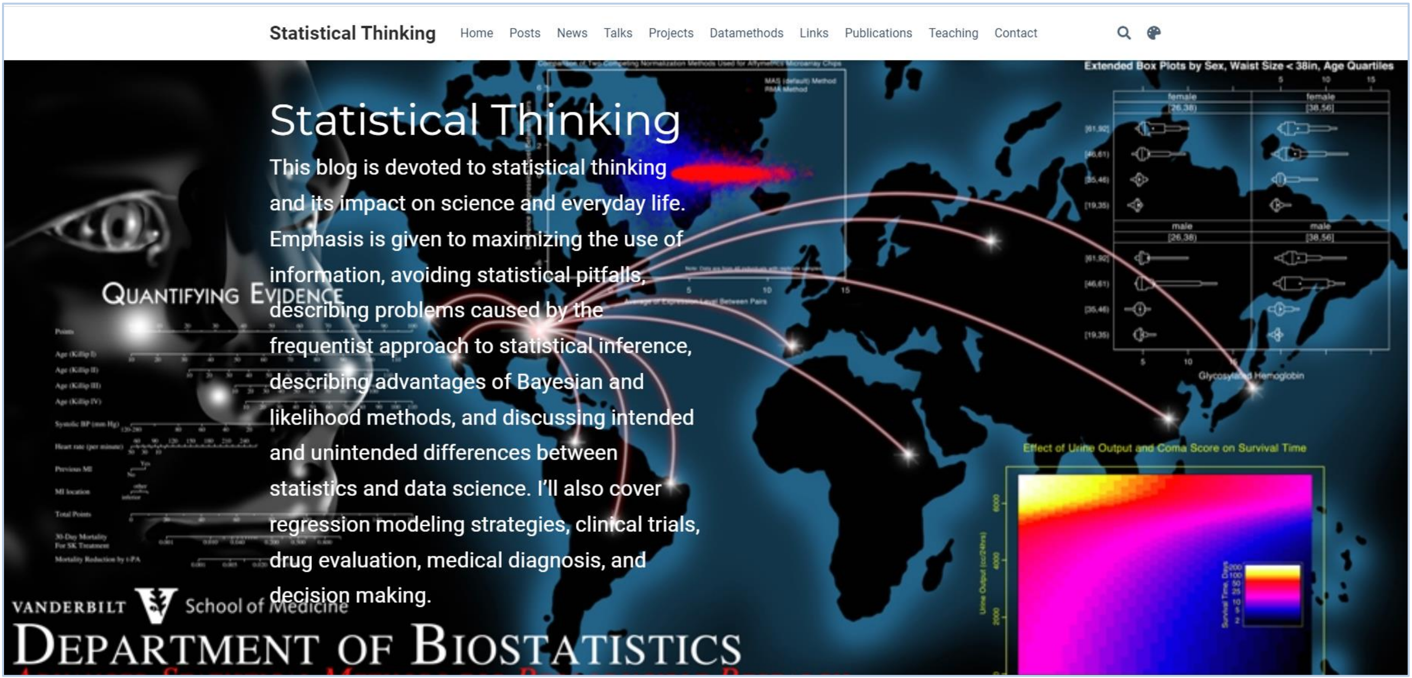
G GECKOBOARD.COM

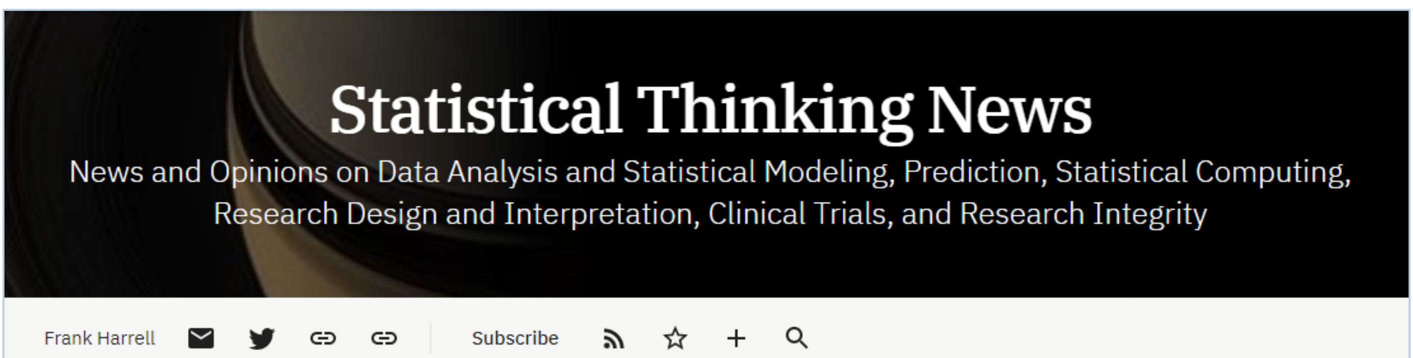
Read more at **data-literacy.geckoboard.com**

Kingston University London

Statistical Literacy

Resources to Develop Statistical Thinking





Kingston University London

Statistical Literacy

Resources to Develop Statistical Thinking

Sensible Medicine

Science-Based Medicine

Exploring issues & controversies in science & medicine

SENSE about SCIENCE

Who we are What we do What you can do Donate

Sense about Science is an independent charity that promotes the public interest in sound science and evidence

Because evidence matters

AMERICAN COUNCIL ON SCIENCE AND HEALTH

Promoting science and debunking junk since 1978.

This website is for educational purposes.

Home About Donate Publications Media/Contact

FactCheck.org®

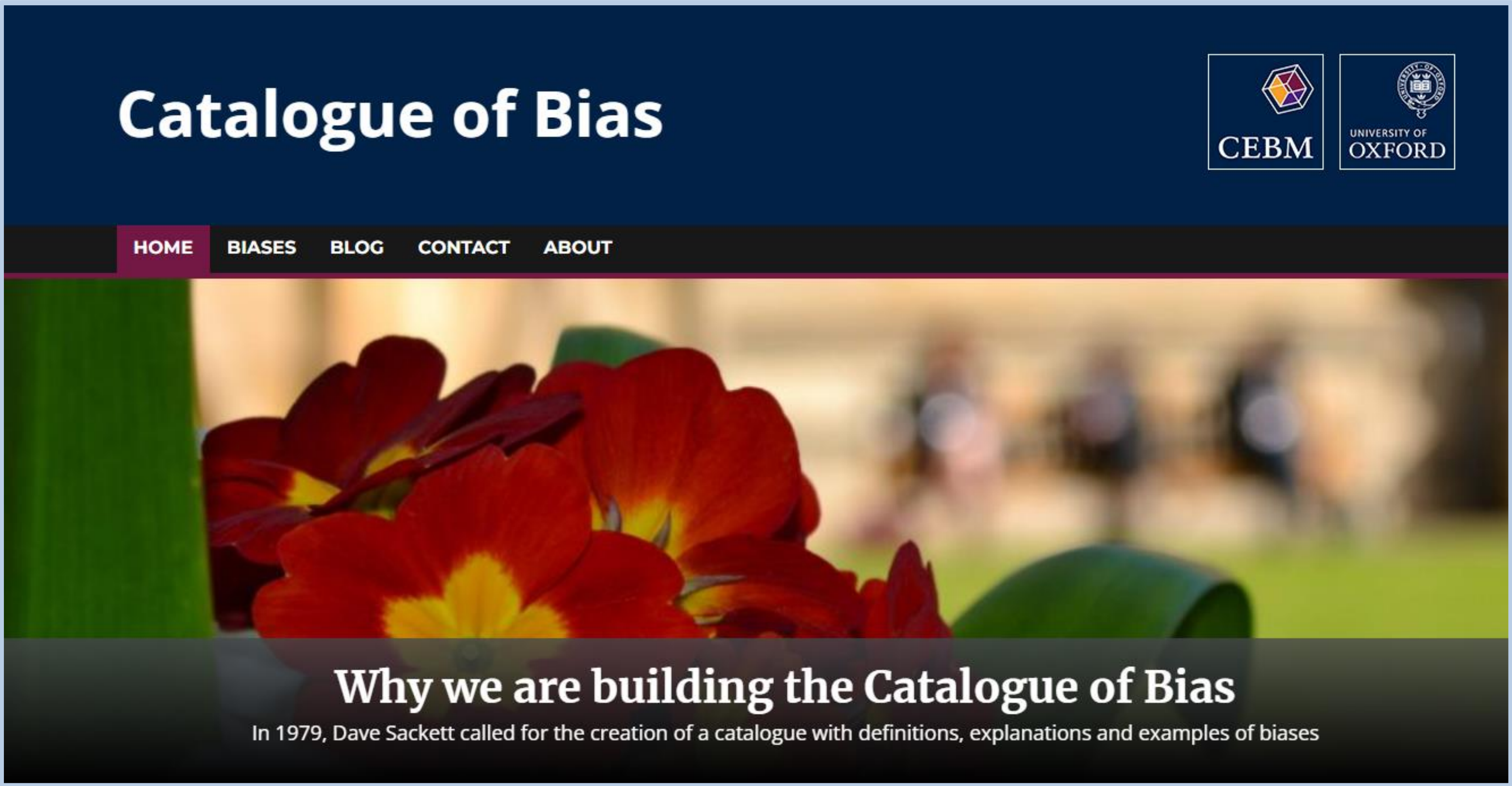
SciCheck

FactCheck.org's SciCheck feature focuses exclusively on false and misleading scientific claims that are made by partisans to influence public policy. It was launched in January 2015 with a grant from the Stanton Foundation. The foundation was founded by the late Frank Stanton, president of CBS for 25 years, from 1946 to 1971.

Kingston University London

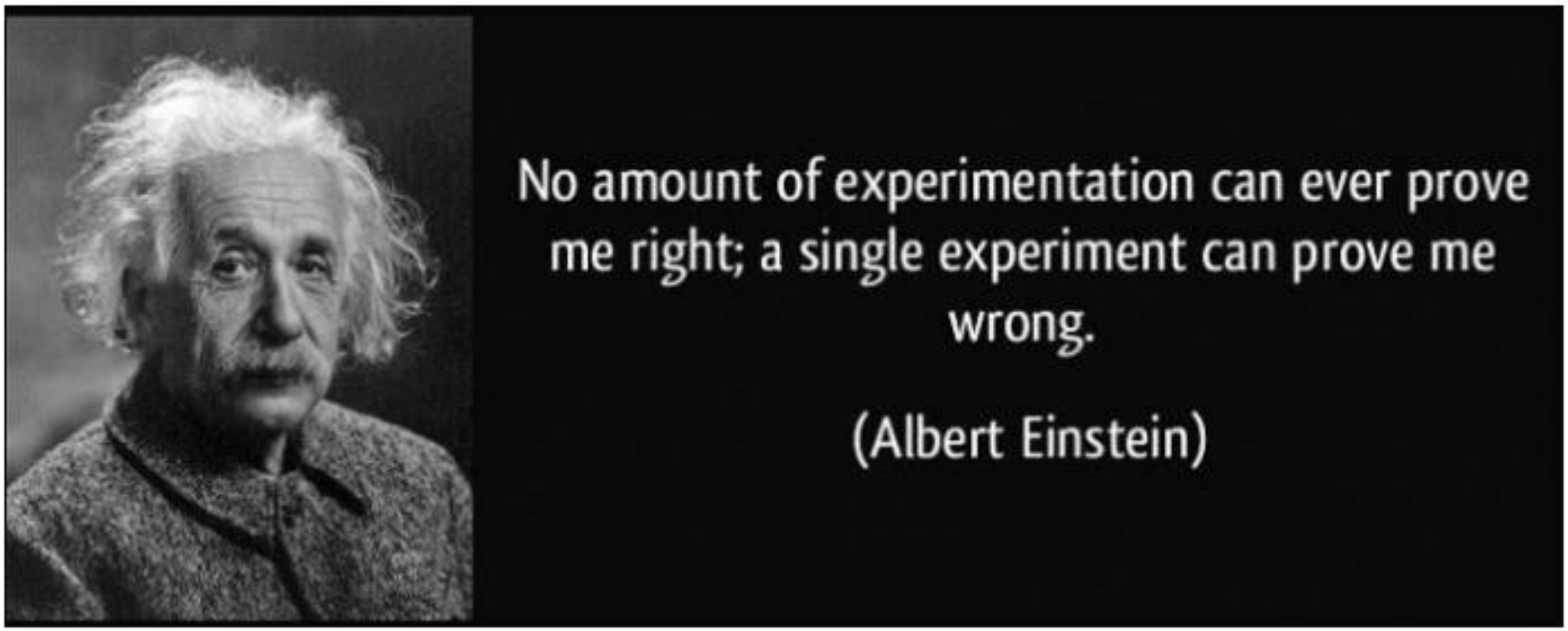
Statistical Literacy

Resources to Develop Statistical Thinking



Statistical Literacy

Statistical Thinking is More Important than Statistics Itself



Modern science training needs more philosophy of science.

Death by Statistics

"Statistics cannot substitute for clear thinking. It can't do the job of human inductive inference."

Cite as: Kamoun, S. (2022). Death by Statistics. Zenodo https://doi.org/10.5281/zenodo.7048973

Kingston University London

Statistical Literacy
Part II

Cognitive biases
Probability paradoxes
Threats to validity of research results: chance, bias, confounding
  > How to minimise them: Epidemiologic study designs
Statistical fallacies & logical fallacies
Pseudo-replication & junk science
Media hoaxes
Hyping health risks
Lying with statistics
Assessment of statistical literacy

Kingston University London

Cognitive Bias

https://www.wiseinsights.net/lp-50-hidden-influences-wrecking-decisions

WISE INSIGHTS
Navigate Toward Success

Start Here | About | Contact | Free Resources

50 Hidden Influences Impacting Your Decisions [Infographic]

The Ultimate List of Influences Distorting How You See Reality

We All Make Choices,
But In The End Our Choices Make Us.
-Ken Levine

One of the most important factors to achieving success in life is making good decisions.

Good decisions = a better life
Poor decisions = a regret filled life

Kind of hard to argue with that.

But, here's the problem: There are hidden factors distorting every decision you make—even the ones carefully made.

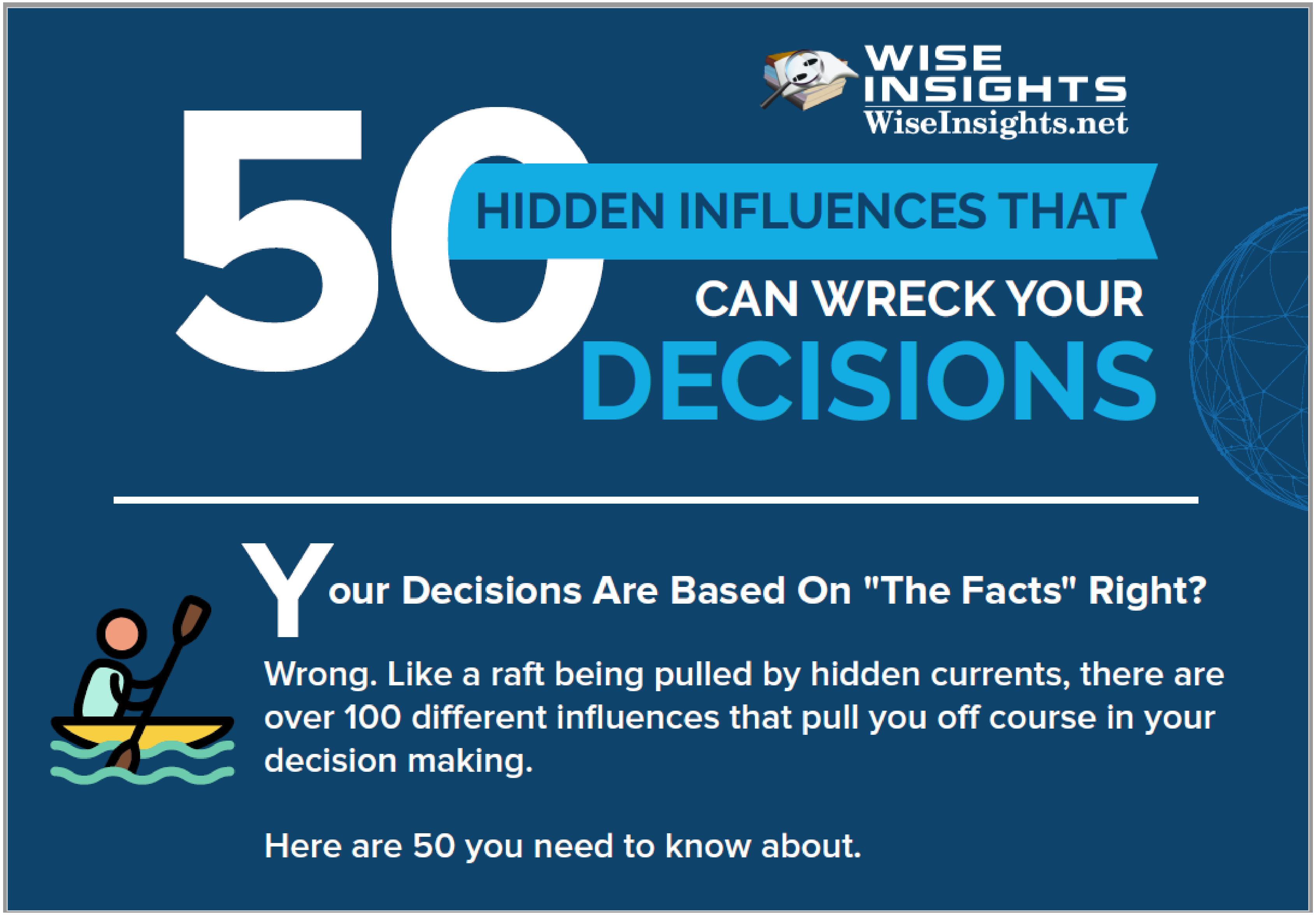
Hidden below the surface of every situation requiring a decision are **100+ different distortions, called cognitive bias, that are pushing and pulling you every which way**… *but you can't see them.*

Even worse, they distort your perception of reality.

- Your perception of yourself
- How you perceive others
- Your perception of "the facts"
- How you perceive the past
- How you look at the future
- Your perception of your current situation

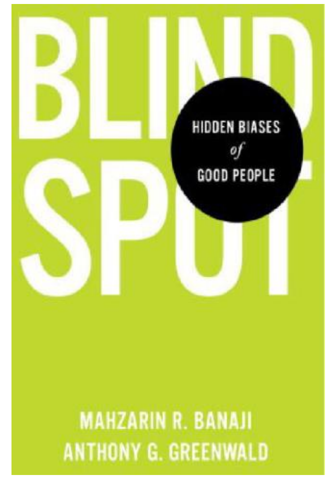
Kingston University London

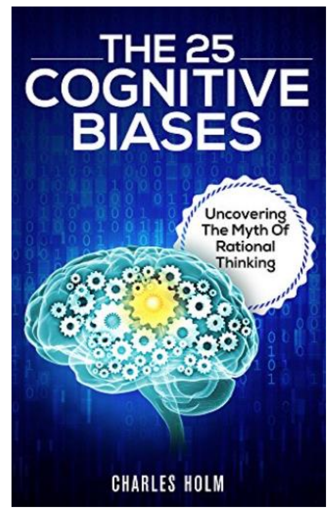
Cognitive Bias



Kingston University London

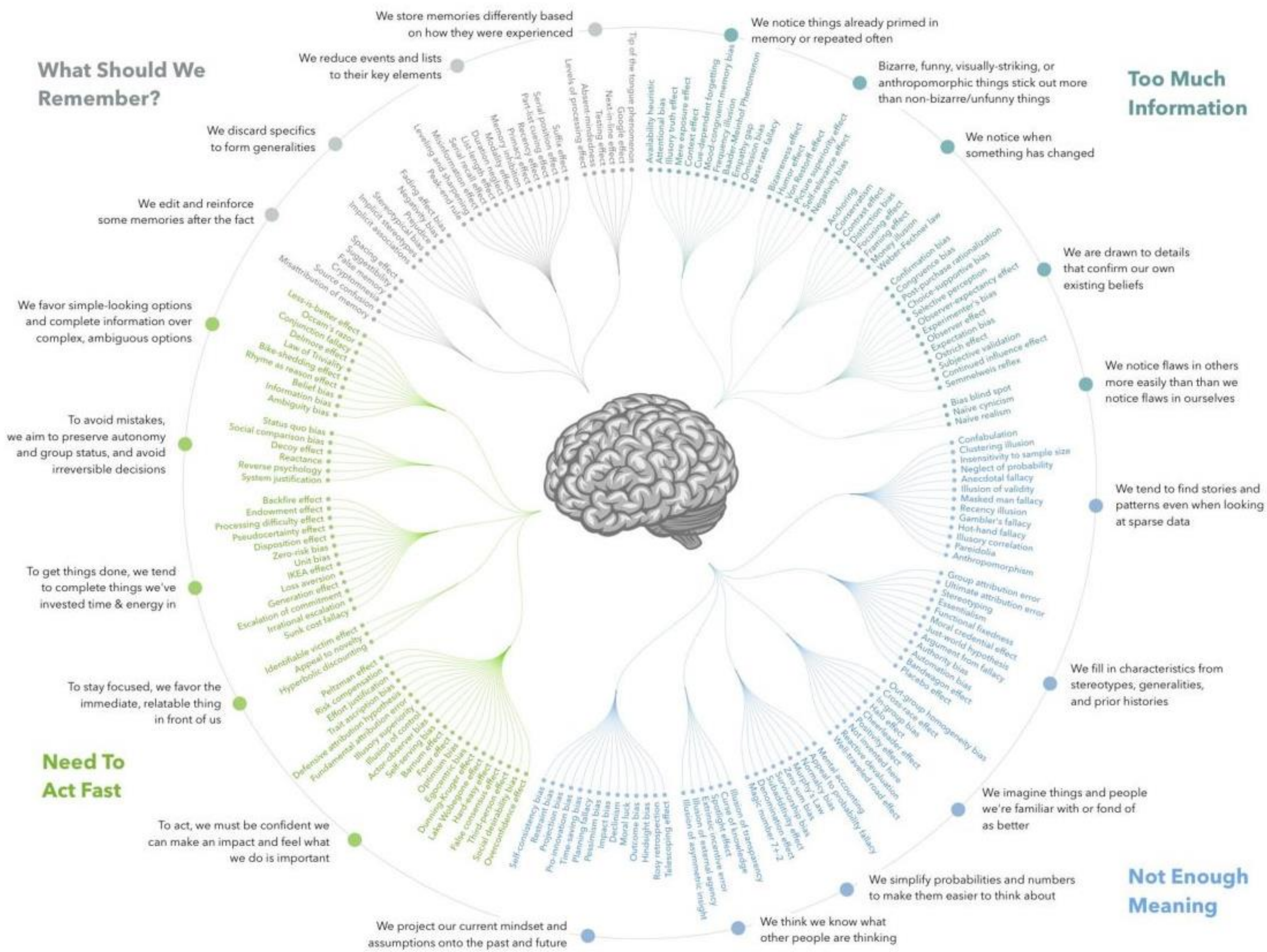
See also:





Kingston University London

COGNITIVE BIAS CODEX



DESIGNHACKS.CO · CATEGORIZATION BY BUSTER BENSON · ALGORITHMIC DESIGN BY JOHN MANOOGIAN III (JM3) · DATA BY WIKIPEDIA

Cognitive Bias

Dunning Kruger Effect

The Dunning-Kruger effect is a cognitive bias that causes people with low abilities to overestimate themselves. Conversely, people with high skills tend to underestimate themselves.

You've probably seen old high school friends on social media or had dinner guests who are "experts" on various subjects, which seem to change weekly. They'll rattle on confidently about complex topics that they know little about while being blatantly unaware of their factual shortcomings. And if you try to correct them factually, watch out! That's the Dunning-Kruger effect in action.

Why does this self-distortion occur? And how can you avoid falling victim to the Dunning-Kruger Effect?

Biases about Probability

AVAILABILITY BIAS

A tendency to assign a greater likelihood to things that are easy to remember

"It always rains on a Bank Holiday"

REPRESENTATIVENESS BIAS

A tendency to assign likelihood of an outcome based on it appearing similiar to other known outcomes

"I didn't need one yesterday so I won't need an umbrella today"

EQUIPROBABILITY BIAS

A tendency to assume different outcomes are equally likely

"It can either rain or not rain on any day, it's 50% either way"

Adapted from ideas in Pratt, D. (2000).[1]

ISSUE 37 JUNE 2021
CAMBRIDGE Mathematics

Probability: Monty Hall Problem

My Store | Glossary | Home | About Me | Contact Me

Statistics By Jim

Making statistics intuitive

Graphs | Basics | Hypothesis Testing | Regression | ANOVA | Probability | Time Series | Fun

The Monty Hall Problem: A Statistical Illusion

By Jim Frost

Who would've thought that an old TV game show could inspire a statistical problem that has tripped up mathematicians and statisticians with Ph.Ds? The Monty Hall problem has confused people for decades. In the game show, Let's Make a Deal, Monty Hall asks you to guess which closed door a prize is behind. The answer is so puzzling that people often refuse to accept it! The problem occurs because our statistical assumptions are incorrect.

The Monty Hall problem's baffling solution reminds me of optical illusions where you find it hard to disbelieve your eyes. For the Monty Hall problem, it's hard to disbelieve your common sense solution even though it is incorrect!

The comparison to optical illusions is apt. Even though I accept that square A and square B are the same color, it just doesn't seem to be true. Optical illusions remain deceiving even after you understand the truth because your brain's assessment of the visual data is operating under a false assumption about the image.

Meet Jim

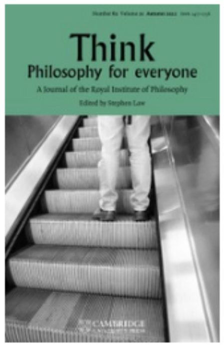
I'll help you intuitively understand statistics by focusing on concepts and using plain English so you can concentrate on understanding your results.

Read More...

Search this website

Probability: Prisoner's Dilemma

Home > Journals > Think > Volume 21 Issue 61 > THE PRISONER'S DILEMMA PARADOX: RATIONALITY, MORALITY,...



Think

THE PRISONER'S DILEMMA PARADOX: RATIONALITY, MORALITY, AND RECIPROCITY

Published online by Cambridge University Press: **17 March 2022**

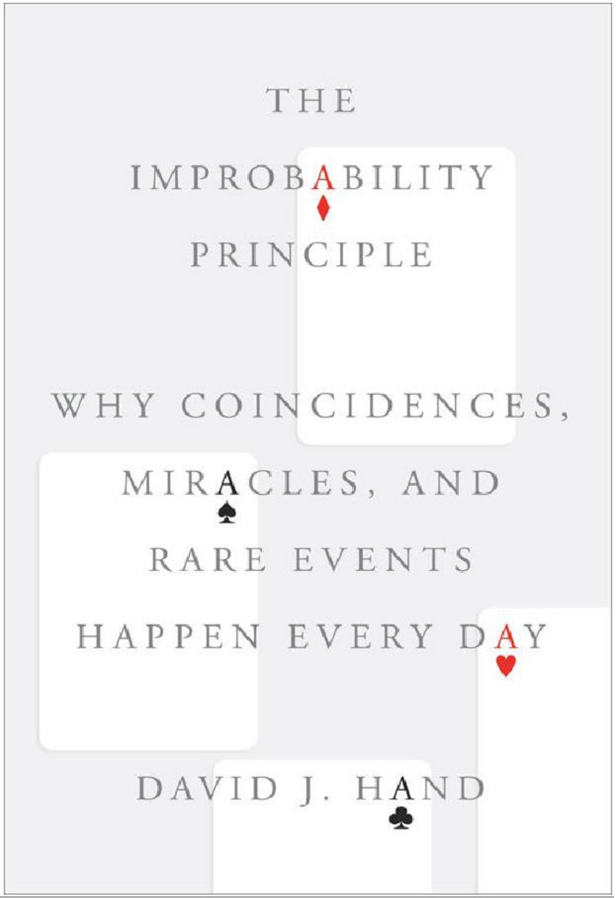
Rory W. Collins

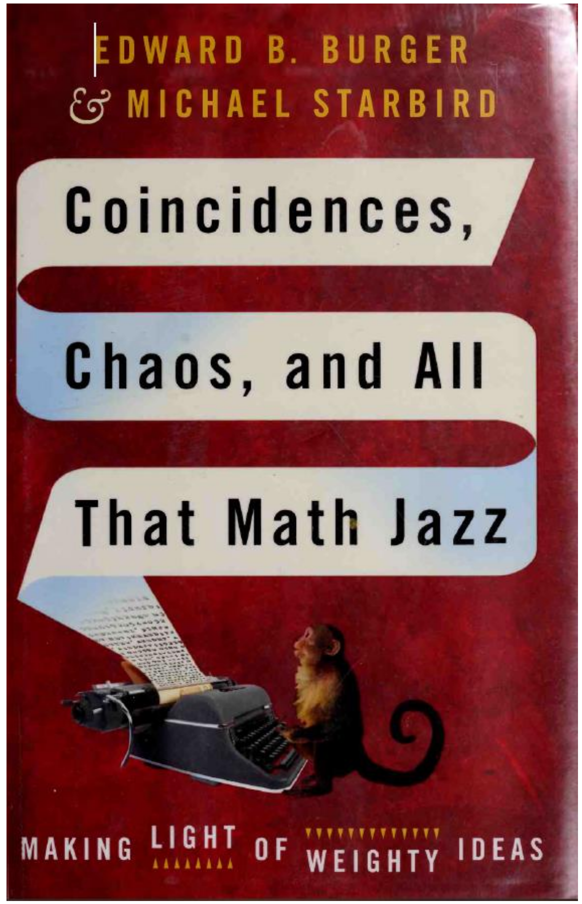
Article Figures Metrics

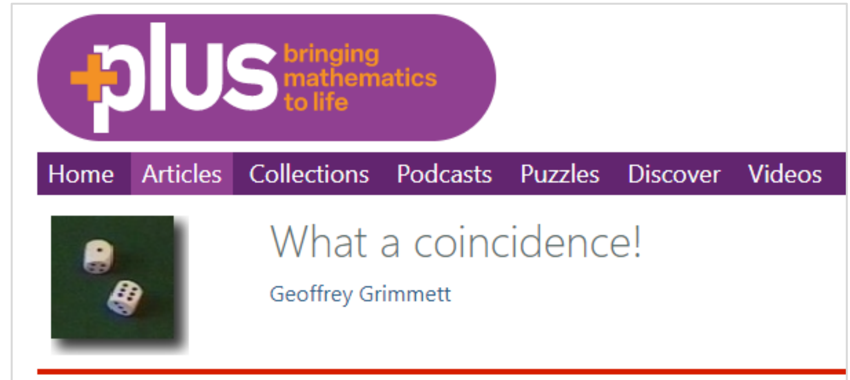
Save PDF Share Cite

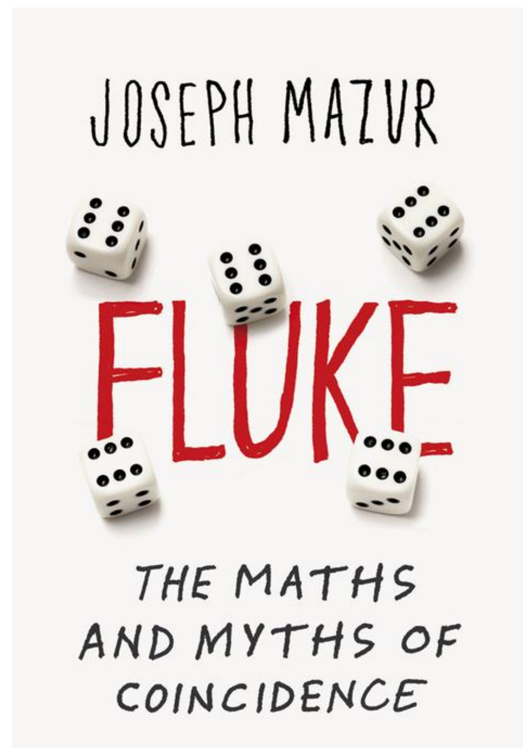
The Prisoner's Dilemma: A Mathematical Analysis

Improbability Principle









Kingston University London

Prosecutor's Fallacy

Convicted on Statistics?

by Vincent Scheurer[i]

"... we do not convict people in these courts on statistics.
It would be a terrible day if that were so."
Mr Justice Harrison, R v. Sally Clark. November 1999[ii]

Introduction

On the 9th November 1999, Sally Clark, then a solicitor and 35-year-old mother of one, was convicted of the murder of her first two children. Christopher, her first child, had died three years earlier, aged just under three months. His death had originally been treated as arising by natural causes, probably "Sudden Infant Death Syndrome" (or "SIDS"). Her second child, Harry, died just over a year later at the age of two months. His death was treated as suspicious by the Home Office pathologist who carried out the post mortem. He then revisited Christopher's death and determined that that too was suspicious. Sally Clark was arrested some weeks later and eventually tried at Chester Crown Court.

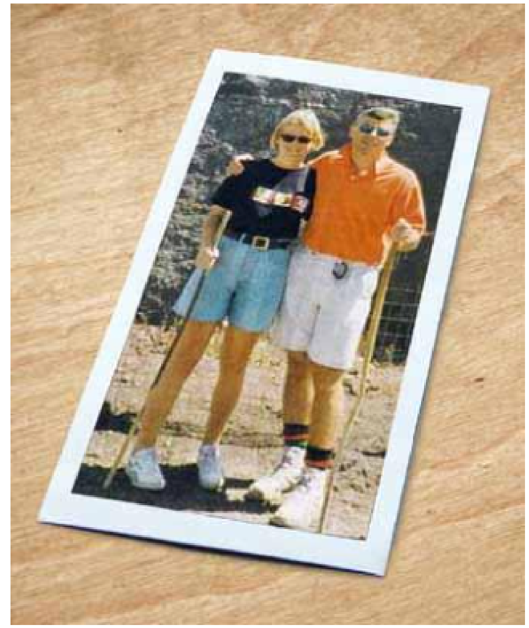
Prosecutor's Fallacy

plus bringing mathematics to life

Beyond reasonable doubt

Helen Joyce

Five years ago, a young couple from Cheshire suffered one of the most devastating losses imaginable - their baby Christopher died in his sleep, aged 11 weeks. Doctors, neighbours, all were sympathetic, and the death was certified as natural causes - there was evidence of a respiratory infection, and no sign of any failure of care.



Happier days - a family snapshot of Sally and Steve Clark when they were expecting Christopher

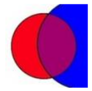
But just a year later, in what must have felt like a horribly familiar nightmare, the Clarks' second child Harry died, aged 8 weeks. This time, there was no sympathy from the professionals. Four weeks after Harry's death the couple were arrested, and eventually Sally Clark was charged with murdering both children. She was tried and convicted in 1999 and is now almost three years into a life sentence.

The forensic evidence was slim to nonexistent - certainly neither case would have stood up alone. Even the prosecution team disagreed among themselves as to how the two children had died. They claimed first that they had been shaken, then that Sally Clark had smothered them (the forensic indications of these two causes of death are normally quite distinct). There was no evidence that she had been an uncaring or violent mother - in fact, all the evidence pointed in the opposite direction. So how come Sally Clark is serving life in prison? Simply put, because the prosecution argued, and the jury accepted, that lightning does not strike twice.

This homespun piece of "wisdom" was trotted out in court in the guise of a seemingly authoritative statement by paediatrician Sir Roy Meadow, speaking as an expert witness for the prosecution. In probably the most infamous statistical statement ever made in a British courtroom, he claimed that the chance of two children in the same (affluent nonsmoking) family both dying a cot death was 1 in 73 million. This would mean that such a double death would occur less often than once a century in England. Those who attended the trial say that, in the midst of the confusion about whether any crimes had even occurred, and if so, how they had been committed, this statement provided something definite to hold on to. It was widely headlined in the national press. However, there is a problem - this statistic was wrong, irrelevant, biased and totally misleading.

Kingston University London

Prosecutor's Fallacy



Maths in a minute: Bayes' theorem

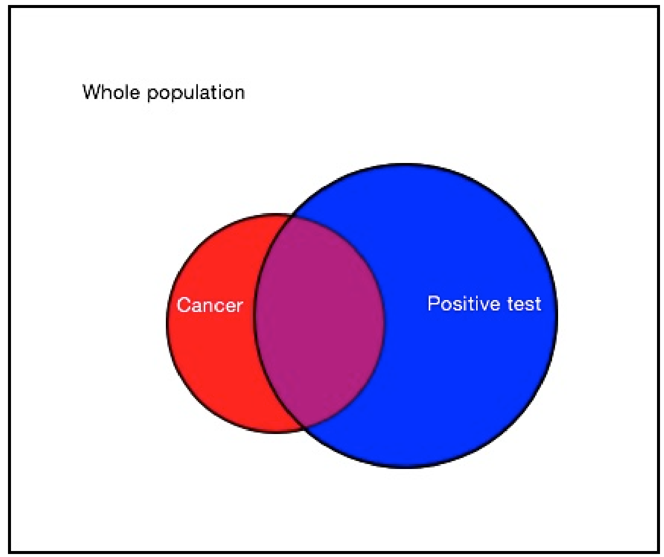
Rachel Thomas

Suppose that a particular type of cancer affects 1% of the population. There is a test for this cancer but it's not perfect: although the test gives a positive result for 90% of people who have the cancer, it also gives a positive result for 5% of the people who are cancer-free. You have just received a positive test result – what is the probability you have cancer?

Many of us would say there is now a 90% chance that we have cancer. But this isn't correct – your chances are closer to 15%. To understand why we have to call on *conditional probabilities* and a very useful result: *Bayes' theorem*.

A *conditional probability* is the probability that one thing is true (in this example, that you have this type of cancer) given another thing is true (your test result is positive). For our example we'd write the conditional probability of having this cancer given a positive test result as $P(cancer|positive)$.

Before you had the test, you believed that your probability of having this cancer was $P(cancer) = 0.01$. So, in a population of 10,000 people you'd expect 100 of them to have this cancer. This group of people is represented by the red circle in the picture.



Now you've had a positive test result. How many people out of our population of 10,000 will have had a positive test result – represented by the blue circle in the picture?

There is a 90% chance of a positive test result if you have cancer. For our example population of 10,000 people, 90 out of 100 people with this cancer will receive a positive test result – these people lie in the intersection of the blue and red circles.

And there is a 5% chance that you'll still get a positive test if you are cancer-free – these people lie in the blue circle that is outside of the red circle in the picture. So for the 9,900 cancer-free patients in our population, 495 will incorrectly test positive. This gives a total of $90 + 495 = 585$ people out of every 10,000 people expected to get a positive test.

Prosecutor's Fallacy

plus bringing mathematics to life

Home | Articles | Collections | Podcasts | Puzzles | Discover | Videos

Beyond reasonable doubt

Helen Joyce

So what is $P(cancer|positive)$, the probability of you having this cancer, given you've had a positive test result? This is the proportion of people who have cancer and have had a positive test result (the intersection of the two circles) of all the people who've had a positive test result (the blue circle): $90/585 = 0.154$. Or written in terms of probabilities

$$P(cancer|positive) = \frac{P(cancer)P(positive|cancer)}{P(positive)} = \frac{0.01 \times 0.9}{0.0585} = 0.154.$$

where $P(positive|cancer)$ is the probability of getting a positive test result given you do have cancer.

So your chance of having this cancer given you've had a positive test result is a much more encouraging 15%. This result is known as *Bayes' Theorem*, written more generally as

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Bayes' theorem allows you to update your prior belief (in this case, that your chance of having cancer was 1%) when new evidence becomes available (a positive test result).

Crucially, *P*(cancer|positive) is not equivalent to *P*(positive|cancer); that would be called "*transposed conditional*". Assuming equivalence would be the mathematical basis of "*prosecutor's fallacy*". In the above scenario, for example, *P*(cancer|positive) = 0.15 ≠ *P*(positive|cancer) = 0.90.

Prosecutor's Fallacy

Practice of Epidemiology

Epidemiology Visualized: The Prosecutor's Fallacy

Daniel Westreich* and Noah Iliinsky

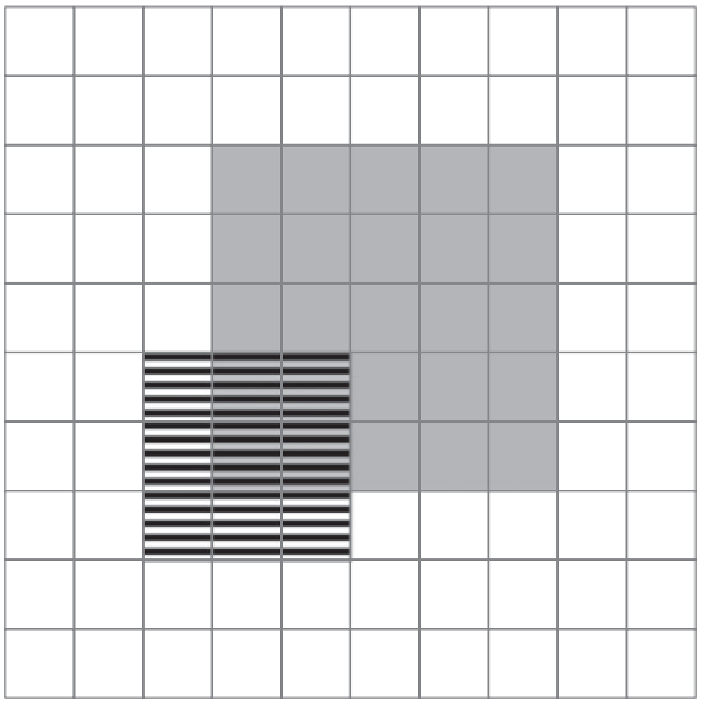


Figure 1. A population represented by 100 squares, with 9 striped squares, 25 gray squares, and 4 striped and gray squares. Pr (striped | gray) = 4/25; Pr (gray | striped) = 4/9. Assuming Pr (striped | gray) = Pr (gray | striped) is the prosecutor's fallacy.

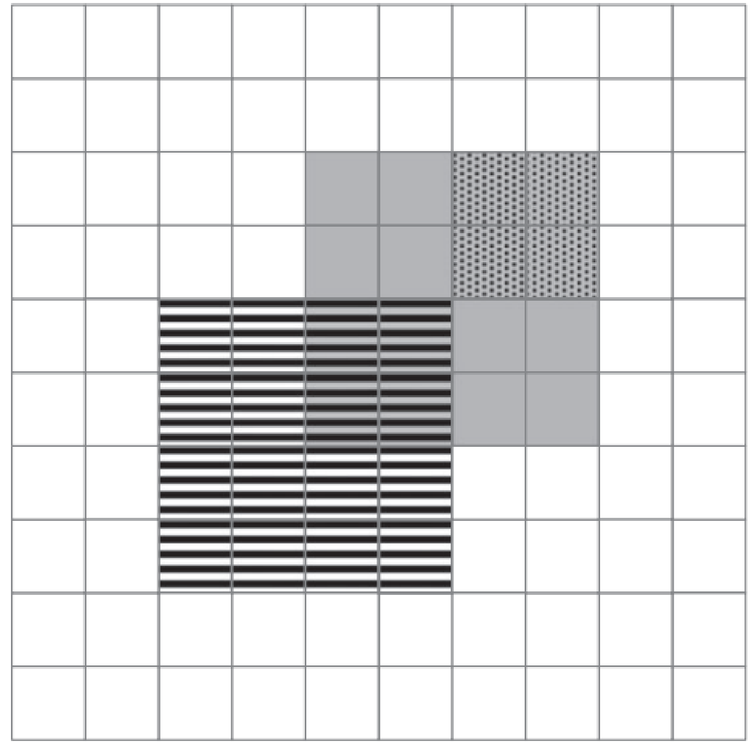


Figure 2. A population represented by 100 squares, with 16 striped squares, 16 gray squares, and 4 dotted squares. Because the overall prevalence of striped and gray squares is the same, Pr (striped | gray) = Pr (gray | striped). Also, Pr (striped | dotted) = Pr (dotted | striped), because both probabilities are 0. Both of these (similarly sized populations and populations with no overlap) are narrow exceptions to the fallacy. The dotted squares are completely contained by the gray squares; thus, Pr (gray | dotted) = 1, whereas Pr (dotted | gray) = 0.25. The prosecutor's fallacy holds when 1 group is a proper subset of the other.

Am J Epidemiol. 2014;179(9):1125–1127

Prosecutor's Fallacy

Appendix C: The Prosecutor's Fallacy

The prosecutor's fallacy is a slimy little trick born out of the difficulty we humans have in interpreting probabilities. By failing to put a probability in the proper context, a lawyer can make it appear incontrovertible that a person is guilty when the case is far from certain.

To illustrate the fallacy, let's move away from the court system for a moment. Imagine that researchers have just discovered a rare and deadly disease—Head-Exploding Syndrome (HES). It's 100 percent fatal; if you contract it, you're going to die a horrible, painful, and very messy death. Luckily, researchers have developed an extremely good blood test to tell whether you've got HES. It's exquisitely accurate—there's only a one in a million chance that the blood test gives the wrong answer. Using this test, doctors start screening the population to find HES cases.

So you go in to the doctor's office, and the doctor draws your blood and leaves the room. A few minutes later, she returns, pale as a ghost. "The test came back positive," she says. Since the test is so accurate—the chances of an error are one in a million—it's virtually certain that you've got Head-Exploding Syndrome. There's only a one in a million chance that the test is wrong. This means that there's only a one in a million chance that you don't have the disease . . . right?

Not so fast. There's a piece of information that's missing before you can conclude that you've really got HES. You need to know just how rare the disease is. As it happens, HES is extraordinarily rare—epidemiologists estimate that it afflicts one in a billion people around the world. That means that of the seven billion people on earth, we expect only seven of them to have the disease. This little bit of information is crucial, because it allows you to put your positive test in context.

The one in a million chance of an error in the test seems pretty small, but if you're screening seven billion people, the one in a million chance of error means that roughly seven thousand patients are going to get a test result that gives the wrong answer. That is, seven thousand people around the planet will test positive on the test even though they don't have the disease. Since only seven people in the entire world actually have HES, this means that the vast majority of people who get a positive test don't actually have HES. Indeed, if you test positive, the probability that you have HES is 7 divided by 7,000—or 1 in 1,000. The chances are 999 in 1,000 that the test is wrong and you don't have the disease. You can rest easy. The one in a million probability of an incorrect test was deceptive; the probability didn't mean anything on its own. Only when you put that one in a million in context—comparing it to the one in a billion incidence of the disease—can you calculate your chance of actually having HES.

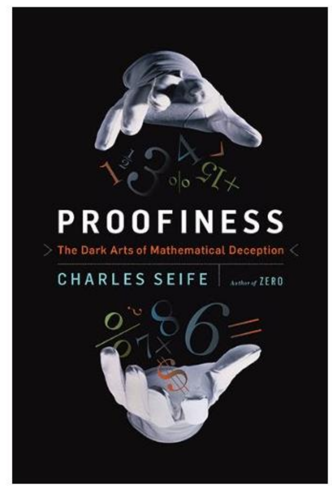
The lesson here is that you must always put probabilities in their proper context. It's a fallacy to look at the chance of the test's being wrong and equate that with the probability that you have a disease. Instead, you must compare the probability to the chances of having the disease in the first place—and when the disease is rare, it can make even a tiny probability of an error on a test loom large.

This is the prosecutor's fallacy in a different form. Instead of blood tests and disease, the fallacy deals with evidence and guilt, but the mathematics is exactly the same. A lawyer presents a very small probability without putting it in the proper context. As a result, the small probability convinces the jury that the statement must be true. Had the lawyer put the probability in the proper context, it would have been much less convincing—and perhaps even led the jury to the opposite conclusion.

For example, Alan Dershowitz argued that O. J. Simpson was innocent because there was only a one in a thousand chance that a wife-beater kills his wife. This number, he implies, means that there's only a one in a thousand chance that O. J. is guilty. This is the prosecutor's fallacy in action, because Dershowitz doesn't put the number into the proper context. The one in a thousand figure doesn't take into account that there is an extremely low probability that a thirty-five-year-old woman like Nicole Brown is murdered in a given year—about 1 in 40,000. Just as the one in a billion chance of disease made the one in a million chance of test error loom very large in comparison, the 1 in 40,000 chance of murder makes the 1 in 1,000 chance of a wife-beater turning murderer seem huge. When statisticians crunched the numbers in the proper way, their estimates of the probability of O. J.'s guilt turned out to be quite large—better than 50 percent. The fact that O. J. had previously battered his wife made it much more probable that he was the murderer, as would have been clear had Dershowitz put the number in the proper context.

Similarly, even if the probabilities Sir Roy Meadow used to convict Sally Clark of killing her children were in the ballpark (which they weren't), they would be deceptive on their own. Meadow didn't take into account that the probability of being a murderer—much less a serial murderer—in the United Kingdom is quite small. Any probability that Meadow used should have been put in the context of that very small probability, which would have made it look a lot less impressive to a jury. Indeed, a mathematician estimated the real probability to be closer to nine in ten that Sally Clark was not guilty based solely upon the deaths of her two children—a far, far cry from the 1 in 73 million chance that Meadow claimed.

The prosecutor's fallacy is powerful because it appeals to our innate misunderstanding of probabilities. It's counterintuitive that a tiny probability (a one in a million chance of a test's going wrong, for example) can, in certain contexts, wind up being extremely large. As a result, few people notice when a prosecutor ignores the context and, through a little numerical hanky-panky, makes a shaky case seem like a rock-solid one.



Prosecutor's Fallacy

APPENDIX: BAYES' THEOREM

For 2 events or dichotomous conditions A and B, Bayes' theorem is usually stated as

$$\Pr(B|A) = \frac{\Pr(A|B) \times \Pr(B)}{\Pr(A)},$$

where Pr (•) indicates the probability of •, and | indicates the "given" operator. So Bayes' theorem can be stated as "the probability of B given A is equal to the probability of A given B times the probability of B, all divided by the probability of A." As a starting point for intuition building, rewrite the equation as

$$\Pr(B|A) = \Pr(A|B) \times \frac{\Pr(B)}{\Pr(A)}$$

and note that it is obvious that $\Pr(B|A)$ and $\Pr(A|B)$ will be equal only if the unconditional probability of A is equal to the unconditional probability of B. Assuming that $\Pr(B|A) = \Pr(A|B)$ when either $\Pr(A) \mathrel{!=} \Pr(B)$ or the equality of $\Pr(A)$ and $\Pr(B)$ has not been established is precisely the prosecutor's fallacy. However, again, symbolic representations of this issue may not be intuitive for some students, especially those new to statistical thinking. Visualizations in main text may help build intuition.

Prosecutor's Fallacy

APPENDIX: BAYES' THEOREM

For 2 events or dichotomous conditions A and B, Bayes' theorem is usually stated as

$$\Pr(B|A) = \frac{\Pr(A|B) \times \Pr(B)}{\Pr(A)},$$

where Pr (•) indicates the probability of •, and | indicates the "given" operator. So Bayes' theorem can be stated as "the probability of B given A is equal to the probability of A given B times the probability of B, all divided by the probability of A." As a starting point for intuition building, rewrite the equation as

$$\Pr(B|A) = \Pr(A|B) \times \frac{\Pr(B)}{\Pr(A)}$$

and note that it is obvious that Pr (B|A) and Pr (A|B) will be equal only if the unconditional probability of A is equal to the unconditional probability of B. Assuming that Pr (B|A) = Pr (A|B) when either Pr (A) != Pr (B) or the equality of Pr (A) and Pr (B) has not been established is precisely the prosecutor's fallacy. However, again, symbolic representations of this issue may not be intuitive for some students, especially those new to statistical thinking. Visualizations in main text may help build intuition.

Quick Bayes Theorem Calculator

The Calculation - Stage 1

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Kingston University London

Prosecutor's Fallacy

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|not\ H)P(not\ H)}$$

In mathematical language, we need to find the *conditional probabilities* of the various possible causes of death, *given* the fact that the children died. If H is some hypothesis - for example, that both of Sally Clark's children died of cot death - and D is some data - that both children are dead, - we want to find the probability of the hypothesis given the data, which is written $P(H|D)$. Let's write A for the alternate hypothesis - that the children did not die of cot death. We will discount all other possibilities, for example that someone else murdered both children, or that Sally Clark murdered only one of them, or that they died of natural causes other than cot death.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D|H)P(H) + P(D|A)P(A)}. \qquad (1)$$

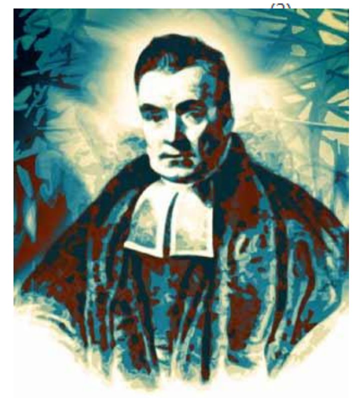
This is not as complicated as it looks. We already have an estimate for $P(H)$ of 1/100,000. This is an absolute, rather than a conditional, probability - the probability that a random pair of siblings die a cot death, not the probability that a random pair of *dead* siblings died a cot death. Trivially, $P(D|H)$ is equal to 1. It is the probability that two of the children are dead, given that that two of the children have died of natural causes. $P(A)$, being the alternate hypothesis to $P(H)$, is equal to $1 - P(H)$

A completely accurate version of Bayes Theorem would take into account all sorts of factors - for example, the fact that social services had not been involved with the Clark family, their income level, and so on - but there isn't a sufficient amount of data available to do this. However, if we are only looking to analyse the case against Sally Clark, it is sufficient to make a ball-park estimate, in order to decide whether or not the deaths of two babies provide compelling evidence of guilt. In British trials, there is a presumption of innocence - it is for prosecutions to prove "beyond reasonable doubt" that defendants are guilty, not for defendants to prove their innocence. So all we need to do is to make reasonable estimates and show that these lead to reasonable doubt.

$P(D|A)$ is the probability that the children died given that they did not die of natural causes. In other words, it is the probability that a randomly chosen pair of siblings will both be murdered. This is the most difficult figure to estimate. Statistics on such double murders are pretty much nonexistent, because child murders are so rare (far, far more rare than cot deaths) and because in most cases, someone known to have murdered once is not free to murder again. So we fall back on the Home Office statistic that fewer than 30 children are known to be murdered by their mother each year in England and Wales. Since 650,000 are born each year, and murders of pairs of siblings are clearly rarer than single murders, we should use a figure much smaller than 30/650,000=0.000046. We will put a number ten times as small here - 0.0000046 - which is almost certainly overestimating the incidence rate of double murder.

Now we get a rough and ready estimate of Sally Clark's innocence:

$$P(H|D) = \frac{P(H)}{P(H) + P(D|A)(1 - P(H))} = \frac{0.00001}{0.00001 + 0.0000046 \times (1 - 0.00001)} > \frac{2}{3}.$$



Very possibly, as you're reading this, you are making the same mistake. Are you thinking "okay, so the odds aren't as extreme as 1 in 73 million, but they're still astronomically high. There's not that much difference between odds of 1 in 73 million and 1 in 100,000, so Sally Clark must still be guilty." If so, you're committing the "Prosecutor's Fallacy".

Simply put, this is the incorrect belief that the chance of a rare event happening is the same as the chance that the defendant is innocent. Even with the more accurate figure of 1 in 100,000 for the chance that a randomly chosen pair of siblings will both die of cot death, this is not the chance that Sally Clark is innocent. It is the chance that an arbitrary family will lose two children in cot deaths. It's not the most scary statistic you will ever read, but in a big country like England, even such improbable events will happen often enough. Are we to believe, with no evidence, that every mother bereaved in this way is a murderer, just because such an event will only happen a few times a year? (https://plus.maths.org/content/beyond-reasonable-doubt)

Prosecutor's Fallacy

Peter Donnelly: How stats fool juries

TED

232K views • 16 years ago

SIGNIFICANCE ROYAL STATISTICAL SOCIETY ASA AMERICAN STATISTICAL ASSOCIATION Statistical Society of Australia

Feature | Free Access

Misleading statistics within criminal trials

The Sally Clark case

Richard Nobles, David Schiff

First published: 24 February 2005 | https://doi.org/10.1111/j.1740-9713.2005.00078.x | Citations: 10

plus bringing mathematics to life

Home Articles Collections Podcasts Puzzles Discover Videos Login

Maths in a minute: The prosecutor's fallacy

plus bringing mathematics to life

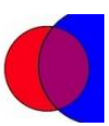
Home Articles Collections Podcasts Puzzles Discover Videos

Beyond reasonable doubt

Helen Joyce

plus bringing mathematics to life

Home Articles Collections Podcasts Puzzles Discover Videos Login



Maths in a minute: Bayes' theorem

Rachel Thomas

plus bringing mathematics to life

Home Articles Collections Podcasts Puzzles Discover Videos

It's a match!

Philip Dawid and Rachel Thomas

Kingston University London

Threats to Validity of Research

Interpreting epidemiological findings

In the third article in this series **Mona Okasha** gives a step by step guide to understanding an epidemiological study

Last month's article considered important features of epidemiological study design.[1] The focus of this article is how to interpret the study's results. This guide is equally applicable to a study of your own as to published journal articles. Using this same structure you will be able to evaluate an epidemiological study that you are faced with in an exam situation.

Between study design and interpreting results is a wide gap—the statistical analysis. Statistics are used to come up with the numbers which form the results. That is a whole subject area in itself, and I refer you to Kirkwood's book (see Further reading) for a clear overview of the subject.

Considerations for evaluating epidemiological results

- Chance—random variation. ⇒ **Assessed by the *P* value**
- Bias—deviation from the truth; selection bias, when participants are not chosen at random; information bias, when the accuracy of data differs according to exposure or outcome.

No quantitative assessments for the rest

- Confounding—an alternative explanation.
- Reverse causality—perhaps the outcome caused the exposure and not vice versa.
- Causal association—this is a very strong assumption, beware.

Hanin *BMC Medical Research Methodology* (2017) 17:127
DOI 10.1186/s12874-017-0399-0

BMC Medical Research Methodology

DEBATE Open Access

Why statistical inference from clinical trials is likely to generate false and irreproducible results

Leonid Hanin[1,2]

Kingston University London

Statistical Noise

Statistical Noise

CENGAGE Views **2,051,598** Updated

Statistical Noise

BIBLIOGRAPHY

Statistical noise refers to variability within a sample, stochastic disturbance in a regression equation, or estimation error. This noise is often represented as a random variable.

In the case of a regression equation

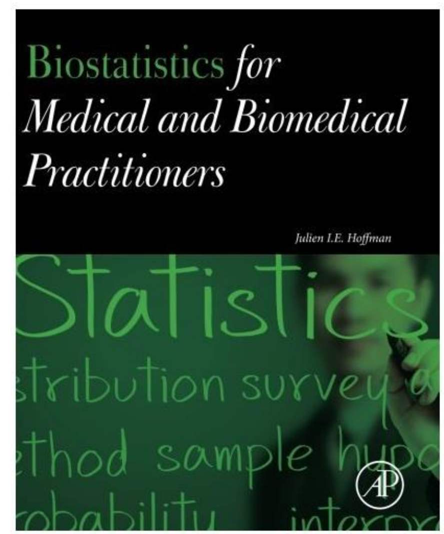
$$Y = m(X \mid \theta) + \epsilon$$

with $E(\epsilon) = 0$, the random variable ϵ is called a disturbance or error term and reflects statistical noise. Noise in this context is usually viewed as arising from omitted explanatory variables; as such, the error term is a proxy for variables not included in the regression. Variables may be omitted from the regression for several reasons. The theory determining the behavior of the dependent variable Y may be incomplete, or perhaps some variables known to influence Y are unavailable to the researcher. Variables that have only slight influence on Y might be eliminated from the regression in order to maintain a parsimonious

Statistical Noise

Measured value = True value ± Bias ± Noise.

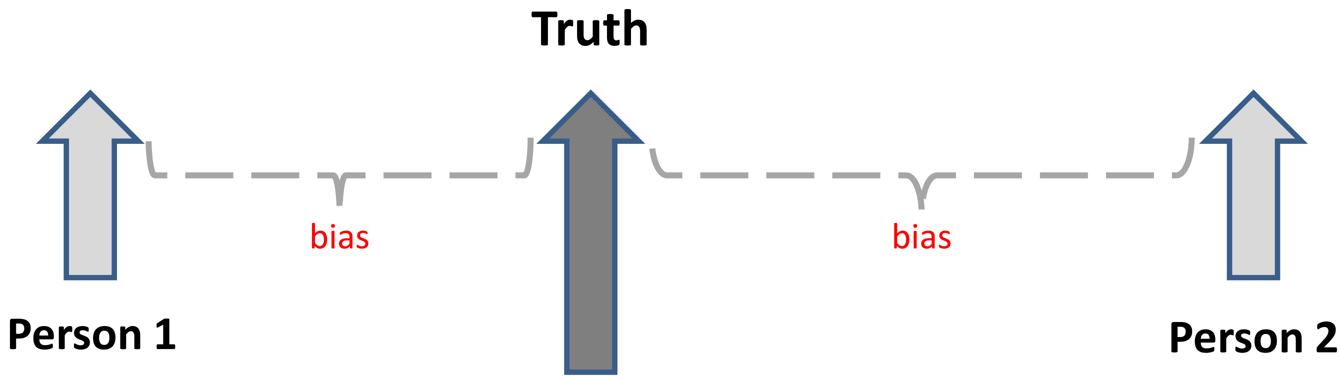
We get this We want this We do not want these



Kingston University London

Statistical Noise

Two colleagues have a dispute and they file complaints about one another.
Their line manager listens to both of them and hear their versions of the story.



The importance of listening to both sides of any story; or the argument and its counter-argument!

Bias in Research

Bias

Kingston University London

Boston University School of Public Health

Bias in Research

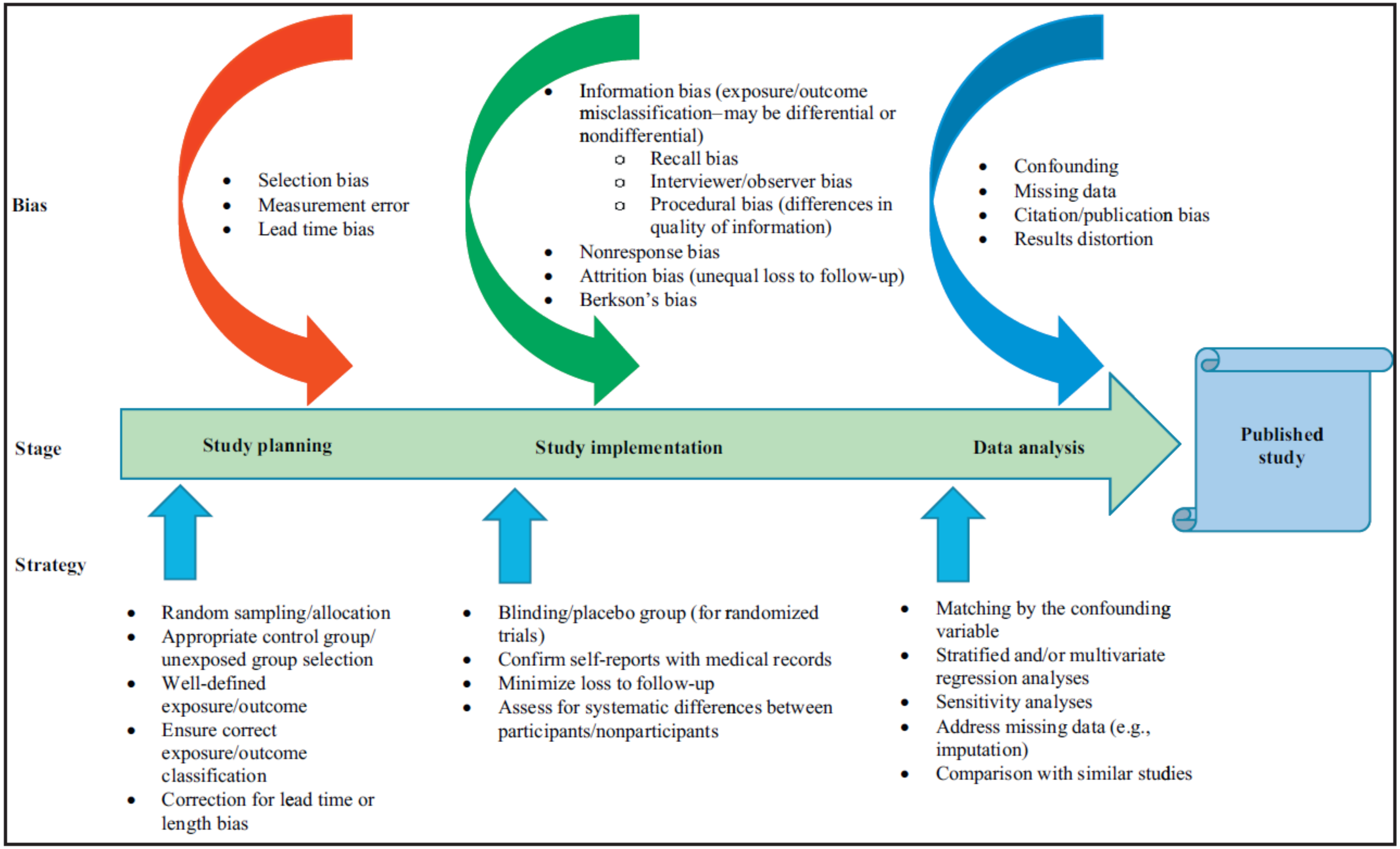


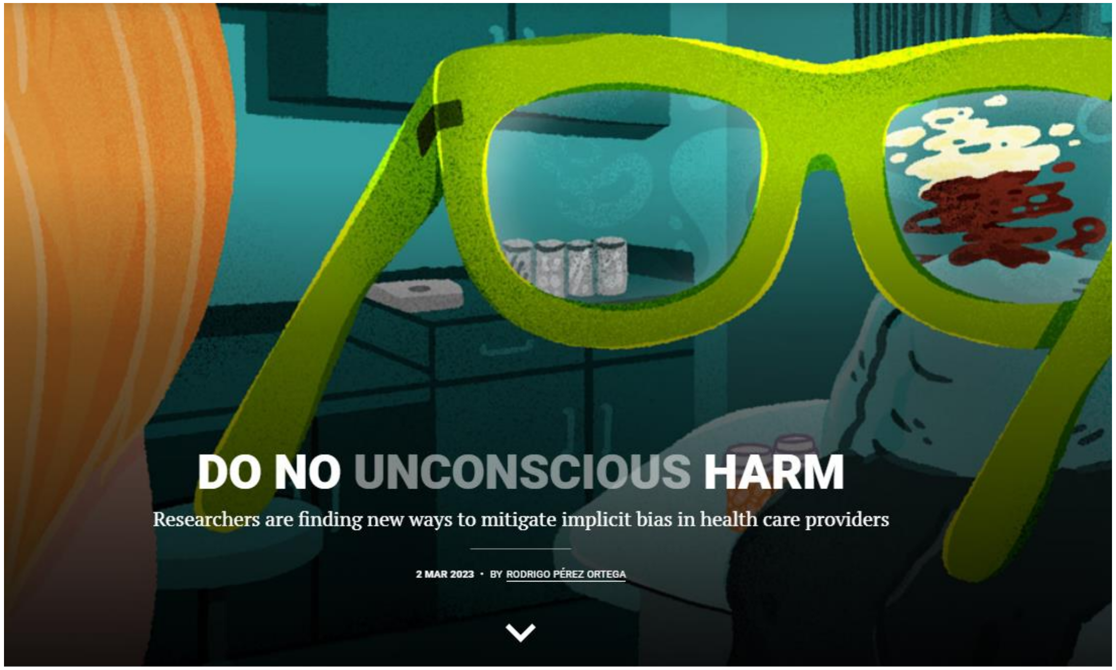
Figure 3. Strategies to minimize biases common to observational research. Methods for addressing various biases in epidemiologic research are shown, although this list in not exhaustive. Readers are referred to several excellent reviews, including Choi and Pak (2005), Delgado-Rodríguez and Llorca (2004), and Sackett (1979).

Kingston University London

Research Techniques Made Simple: Interpreting Measures of Association in Clinical Research

Michelle R. Roberts[1,2], Sepideh Ashrafzadeh[1,2] and Maryam M. Asgari[1,2]

Bias in Medicine/Health Care



JACC: BASIC TO TRANSLATIONAL SCIENCE

STATE-OF-THE-ART REVIEW

Bias in Medicine

Lessons Learned and Mitigation Strategies

M. Elizabeth H. Hammond, MD,[a,b,c] Josef Stehlik, MD, MPH,[a,b] Stavros G. Drakos, MD, PhD,[a,b] Abdallah G. Kfoury, MD[a,b,c]

The American Journal of Medicine
Official Journal of the Alliance for Academic Internal Medicine

COMMENTARY | VOLUME 132, ISSUE 8, P895-896, AUGUST 2019

Bias in Medicine

Joseph S. Alpert, MD

SHARE:

A version of this story appeared in Science, Vol 379, Issue 6635.

UnBIASED
A study by UW and UCSD

Understanding Biased patient-provider Interaction And Supporting Enhanced Discourse

Project Implicit

Kingston University London

Survivorship Bias

Survivorship Bias

Statistics By Jim

Making statistics intuitive

Graphs | Basics | Hypothesis Testing | Regression | ANOVA | Probability | Time Series | Fun

Survivorship Bias: Definition, Examples & Avoiding

By Jim Frost — 5 Comments

Bias in Medicine/Health Care

Survivorship Bias ~ Incident-prevalent Case Bias

In Genetic Association Studies

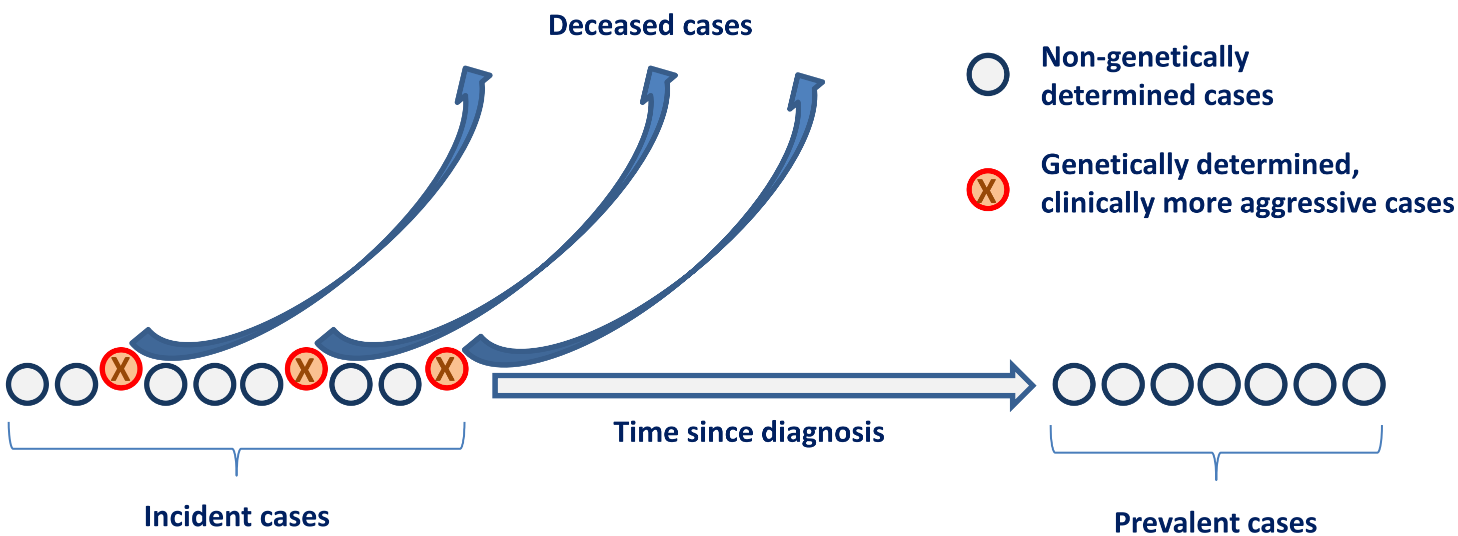
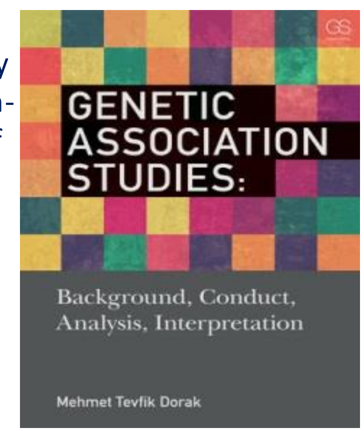


Figure 4.2. The difference between an incident case group and a prevalent case group. If all consecutively diagnosed cases are included in the study, this incident case group contains all genetically and non-genetically determined cases. Genetically determined cases tend to be clinically more aggressive and die earlier than non-genetically determined cases. If prevalent cases are recruited for a study, there will be a relative deficiency of genetically-determined cases.



Kingston University London

Bias in Medicine/Health Care

Pai Global TB Group

teach epi

a website for learning and teaching epidemiology

COURSES | TEACHING RESOURCES | FOR MENTORS | FOR STUDENTS

The B-Files (Bias Case Studies)

Case studies of bias in real life epidemiologic studies
By: Madhukar Pai & Jay S. Kaufman

THE B FILES

Bias File 1. The Rise and Fall of Hormone Replacement Therapy
Bias File 2. Should we stop drinking coffee? The story of coffee and pancreatic cancer
Bias File 3. Émile Durkheim and the ecological fallacy
Bias File 4. The early controversy over estrogen and endometrial cancer
Bias File 5. How blind are the blind? The story of Vitamin C for common cold
Bias File 6. Double whammy: recall and selection bias in case-control studies of congenital malformations
Bias File 7. Confounding by indication: a most stubborn bias?
Bias File 8. Don't call my number, anymore! Bias in surveys of sexual behavior
Bias File 9. Circumcision and HIV

Causality Assessment

Causal Inference

Kingston University London

BU Boston University School of Public Health

Causality Assessment

Required Elements for Causation

Element	Explanation
Association	Do the variables covary empirically? Strong associations are more likely to be causal than are weak associations.
Precedence	Does the independent variable vary before the effect exhibited in the dependent variable?
Nonspuriousness	Can the empirical correlation between two variables be explained away by the influence of a third variable?
Plausibility	Is the expected outcome biologically plausible and consistent with theory, prior knowledge, and results of other studies?

Statistical Concepts Series

Kimberly E. Applegate, MD, MS
Philip E. Crewson, PhD

Index terms:
Radiology and radiologists, research
Statistical analysis

Published online before print
10.1148/radiol.2252010933
Radiology 2002; 225:318–322

[1] From the Department of Radiology, Riley Hospital for Children, 702 Barnhill Rd, Indianapolis, IN 46202 (K.E.A.); and the Department of Research, American College of Radiology, Reston, Va

An Introduction to Biostatistics[1]

This introduction to biostatistics and measurement is the first in a series of articles designed to provide *Radiology* readers with a basic understanding of statistical concepts. Although most readers of the radiology literature know that application of study results to their practice requires an understanding of statistical issues, many may not be fully conversant with how to interpret statistics. The goal of this series is to enhance the ability of radiologists to evaluate the literature competently and critically, not make them into statisticians.

Hierarchy of Evidence

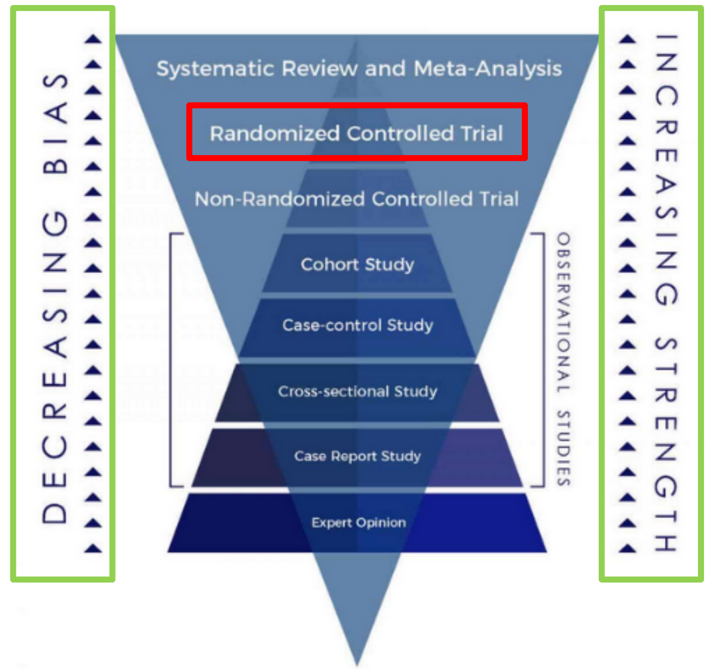
Evaluating Evidence

The important thing is not to stop questioning. Curiosity has its own reason for existing. -Albert Einstein

Not all research is created equal. Many times, scientific studies conclude results that contradict each other, and scientists express opposing viewpoints on subject matter. This can lead to confusion for consumers and increased perceptions of risk and hazard. This brochure provides a greater understanding of what can seem like a mysterious scientific process and context for critical evaluation of scientific literature.

HIERARCHY OF SCIENTIFIC EVIDENCE

When examining the strength of scientific evidence, a number of factors comes into play. Of the most important factors, however, is study design. In the hierarchy of evidence, the strongest evidence results from randomized controlled trials (RCT) and intervention studies. By comparison, weaker evidence results from case reports and expert opinion.



TYPES OF SCIENTIFIC JOURNAL ARTICLES

Full-length research paper: the majority of articles published in scientific journals. These include presentation of relevance for a specific study, methods used to perform the study, presentation of results gathered, and a discussion of conclusions about the study.

Industry interest piece: summary of the work conducted in a study and discusses recommendations (resulting from the research work) that may be helpful to a specific scientific industry. It does not include a detailed account of the study.

Hot topic: report on a study that is not complete at time of publication. Because the study topic is so ground-breaking or important to the field, the journal will allow publication of the preliminary results.

Technical note: report on a new or improved method in the specific scientific field. Invited review: written by authors suggested by journal editors and usually summarizes a specific scientific topic or recent symposia.

Letter to the Editor: usually has a short word limit (~300 words) and reflects topics of concern for the readers. It may include corrections made to articles after publication or even rebuttals from disagreeing scientists.

In the 20th century, randomised controlled trials leapt to the top of the hierarchy of evidence-based medicine because of their ability to minimise confounding & isolate the therapeutic effect of interventions.

Kingston University London

Observational Studies

Panel 1: **What to look for in observational studies**

Is selection bias present?
In a cohort study, are participants in the exposed and unexposed groups similar in all important respects except for the exposure?

In a case-control study, are cases and controls similar in all important respects except for the disease in question?

Is information bias present?
In a cohort study, is information about outcome obtained in the same way for those exposed and unexposed?

In a case-control study, is information about exposure gathered in the same way for cases and controls?

Is confounding present?
Could the results be accounted for by the presence of a factor—eg, age, smoking, sexual behaviour, diet—associated with both the exposure and the outcome but not directly involved in the causal pathway?

If the results cannot be explained by these three biases, could they be the result of chance?
What are the relative risk or odds ratio and 95% CI?[11,12]

Is the difference statistically significant, and, if not, did the study have adequate power to find a clinically important difference?[13,14]

If the results still cannot be explained away, then (and only then) might the findings be real and worthy of note.

Interventional Studies

Randomised Controlled Trials (RCTs)

<u>Causal methods</u> can be divided into randomized clinical trials (RCTs), natural experiments, and statistical models. The first two approaches can potentially control for both known and unknown confounders, while statistical methods control only for known and measured confounders. The criterion standard, RCTs, can have important limitations, especially regarding generalizability.

Review

December 11, 2019

Applying Causal Inference Methods in Psychiatric Epidemiology

A Review

Henrik Ohlsson, PhD[1]; Kenneth S. Kendler, MD[2,3]

Author Affiliations

JAMA Psychiatry. 2020;77(6):637-644. doi:10.1001/jamapsychiatry.2019.3758

Abstract

Importance Associations between putative risk factors and psychiatric and substance use disorders are widespread in the literature. Basing prevention efforts on such findings is hazardous. Applying causal inference methods, while challenging, is central to developing realistic and potentially actionable etiologic models for psychopathology.

Observations Causal methods can be divided into randomized clinical trials (RCTs), natural experiments, and statistical models. The first 2 approaches can potentially control for both known and unknown confounders, while statistical methods control only for known and measured confounders. The criterion standard, RCTs, can have important limitations, especially regarding generalizability. Furthermore, for ethical reasons, many critical questions in psychiatric epidemiology cannot be addressed by RCTs. We review, with examples, methods that try to meet as-if randomization assumptions, use instrumental variables, or use pre-post designs, regression discontinuity designs, or co-relative designs. Each method has strengths and limitations, especially the plausibility of as-if randomization and generalizability. Of the large family of statistical methods for causal inference, we examine propensity scoring and marginal models, which are best applied to samples with strong predictors of risk factor exposure.

Statistical Fallacies

Simpson's Paradox

Simpson's paradox is a phenomenon in which a trend appears in different groups of data but disappears or reverses when these groups are combined.

- Overview
- Exploration
- References

Exploring Simpson's Paradox

For this app, we are going to look at a data set containing data from 2010 for 12 selected states. We will look at these states' average teachers' salaries (US$ 2010), total SAT scores, and SAT participation rates (percent of students in each state taking the SAT).

We've divided the 12 selected states into two groups based upon how their SAT participation compares to the National level of 27%:

- **Low:** States with SAT participation less than 27%
- **High:** States with SAT participation higher than 27%

View Data +

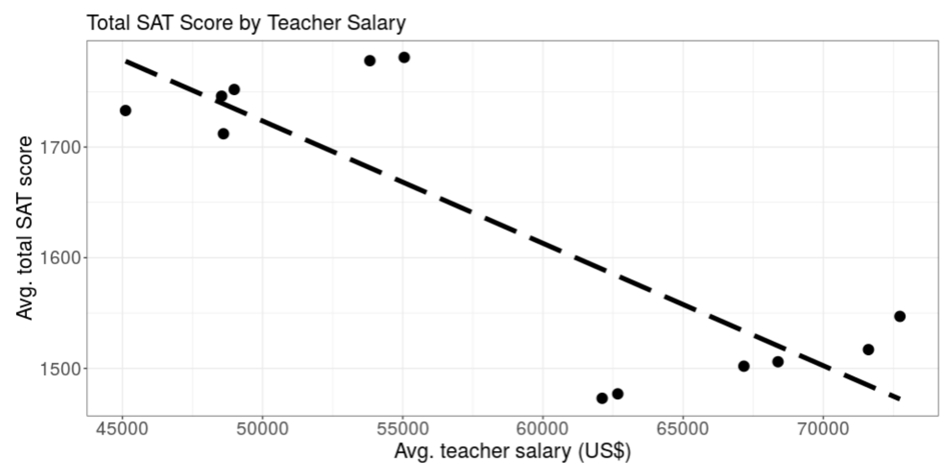
☐ Show Grouping

Vary participation

0 — 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

- Set the slider to 1 to see the actual effect of the paradox.
- Set the slider to 0 to see no paradox effect at all.
- Varying the slider in between will make the SAT participation more (towards 0) or less (towards 1) similar.



Challenge

How does the slope of the overall regression line change as the SAT participation rate become more similar from State to State?

PennState
Eberly College of Science

Kingston University London

Statistical Fallacies

Simpson's Paradox

Simpson's paradox

Article Talk

From Wikipedia, the free encyclopedia

Simpson's paradox is a phenomenon in probability and statistics in which a trend appears in several groups of data but disappears or reverses when the groups are combined. This result is often encountered in social-science and medical-science statistics,[1][2][3] and is particularly problematic when frequency data are unduly given causal interpretations.[4] The paradox can be resolved when confounding variables and causal relations are appropriately addressed in the statistical modeling[4][5] (e.g. through cluster analysis[6]).

Simpson's paradox has been used to illustrate the kind of misleading results that the misuse of statistics can generate.[7][8]

Edward H. Simpson first described this phenomenon in a technical paper in 1951,[9] but the statisticians Karl Pearson (in 1899[10]) and Udny Yule (in 1903[11]) had mentioned similar effects earlier. The name *Simpson's paradox* was introduced by Colin R. Blyth in 1972.[12] It is also referred to as **Simpson's reversal**, the **Yule–Simpson effect**, the **amalgamation paradox**, or the **reversal paradox**.[13]

Mathematician Jordan Ellenberg argues that Simpson's paradox is misnamed as "there's no constriction involved, just two different ways to think about the same data" and suggests that its lesson "isn't really to tell us which viewpoint to take but to insist that we keep both the parts and the whole in mind at once."[14]

Statistical Fallacies

Simpson's Paradox

	Hospital A	Hospital B	Survivors from A/%	Survivors from B/%
Fair condition	700	100	600/86%	90/90%
Serious condition	200	200	100/50%	150/75%
Critical condition	100	700	10/10%	300/43%
Total	1000	1000	710/71%	540/54%

Total is lower for Hospital B, but in each subcategory, it has a higher rate

See: https://plus.maths.org/content/maths-minute-simpsons-paradox

Kingston University London

Statistical Fallacies

Simpson's Paradox

Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.6.1)

Danielle Navarro (bookdown translation: Emily Kothe)

1.2 The cautionary tale of Simpson's paradox

The following is a true story (I think...). In 1973, the University of California, Berkeley had some worries about the admissions of students into their postgraduate courses. Specifically, the thing that caused the problem was that the gender breakdown of their admissions, which looked like this...

	Number of applicants	Percent admitted
Males	8442	44%
Females	4321	35%

...and the were worried about being sued.[4] Given that there were nearly 13,000 applicants, a difference of 9% in admission rates between males and females is just way too big to be a coincidence. Pretty compelling data, right? And if I were to say to you that these data *actually* reflect a weak bias in favour of women (sort of!), you'd probably think that I was either crazy or sexist.

Oddly, it's actually sort of true ... when people started looking more carefully at the admissions data (Bickel, Hammel, & O'Connell, 1975) they told a rather different story. Specifically, when they looked at it on a department by department basis, it turned out that most of the departments actually had a slightly *higher* success rate for female applicants than for male applicants. The table below shows the admission figures for the six largest departments (with the names of the departments removed for privacy reasons):

	Males		Females	
Department	Applicants	Percent admitted	Applicants	Percent admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	272	6%	341	7%

Remarkably, most departments had a *higher* rate of admissions for females than for males! Yet the overall rate of admission across the university for females was *lower* than for males. How can this be? How can both of these statements be true at the same time?

Here's what's going on. Firstly, notice that the departments are *not* equal to one another in terms of their admission percentages: some departments (e.g., engineering, chemistry) tended to admit a high percentage of the qualified applicants, whereas others (e.g., English) tended to reject most of the candidates, even if they were high quality. So, among the six departments shown above, notice that department A is the most generous, followed by B, C, D, E and F in that order. Next, notice that males and females tended to apply to different departments. If we rank the departments in terms of the total number of male applicants, we get **A**>**B**>D>C>F>E (the "easy" departments are in bold). On the whole, males

[4]Earlier versions of these notes incorrectly suggested that they actually were sued – apparently that's not true. There's a nice commentary on this here: https://www.refsmmat.com/posts/2016-05-08-simpsons-paradox-berkeley.html. A big thank you to Wilfried Van Hirtum for pointing this out to me!

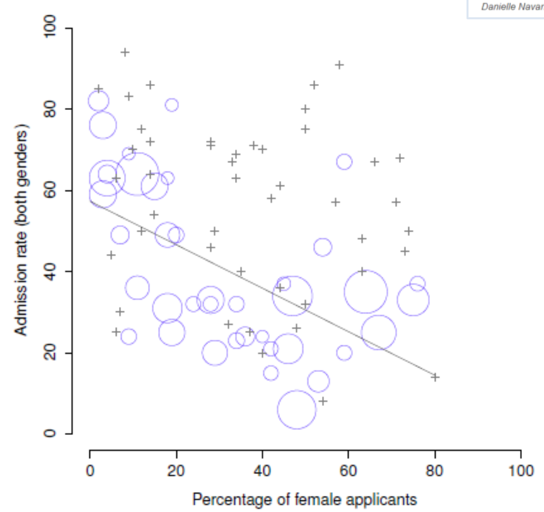


Figure 1.1: The Berkeley 1973 college admissions data. This figure plots the admission rate for the 85 departments that had at least one female applicant, as a function of the percentage of applicants that were female. The plot is a redrawing of Figure 1 from Bickel et al. (1975). Circles plot departments with more than 40 applicants; the area of the circle is proportional to the total number of applicants. The crosses plot department with fewer than 40 applicants.

tended to apply to the departments that had high admission rates. Now compare this to how the female applicants distributed themselves. Ranking the departments in terms of the total number of female applicants produces a quite different ordering C>E>D>F>**A**>**B**. In other words, what these data seem to be suggesting is that the female applicants tended to apply to "harder" departments. And in fact, if we look at all Figure 1.1 we see that this trend is systematic, and quite striking. This effect is known as **Simpson's paradox**. It's not common, but it does happen in real life, and most people are very surprised by it when they first encounter it, and many people refuse to even believe that it's real. It is very real. And while there are lots of very subtle statistical lessons buried in there, I want to use it to make a much more important point ... doing research is hard, and there are *lots* of subtle, counterintuitive traps lying in wait for the unwary. That's reason #2 why scientists love statistics, and why we teach research methods. Because science is hard, and the truth is sometimes cunningly hidden in the nooks and crannies of complicated data.

Before leaving this topic entirely, I want to point out something else really critical that is often overlooked in a research methods class. Statistics only solves *part* of the problem. Remember that we started

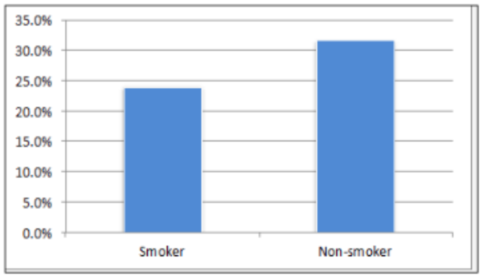
Kingston University London

Statistical Fallacies

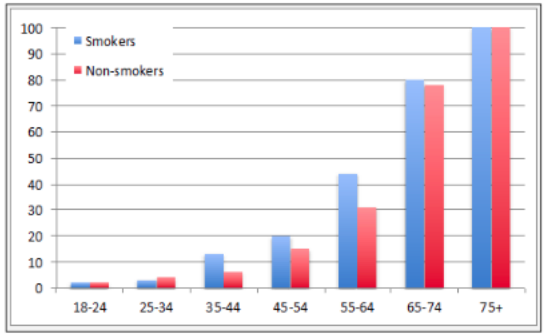
Simpson's Paradox

A 1996 study on the effects of smoking on women revealed higher mortality rates for non-smokers. This is obviously counter-intuitive and extremely surprising. However, upon segregating the data into different age-groups, the results revealed that smokers in all-but-one categories had higher mortality rates. The graphs are shown below.

Mortality Rates – Total



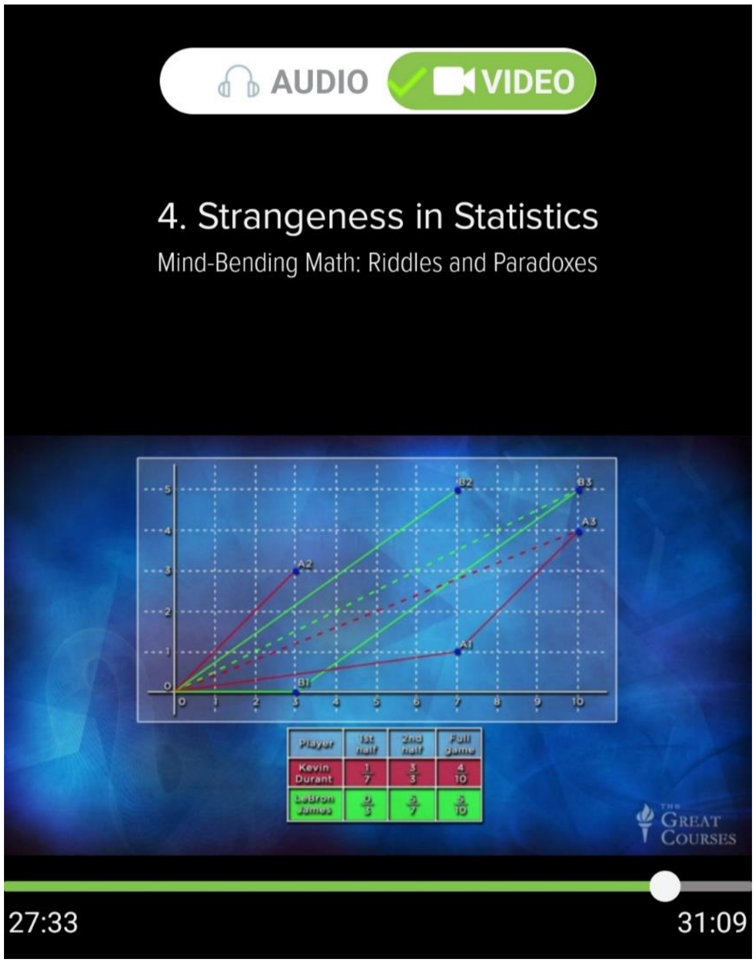
Mortality Rates – By Age Group



The above study is a classic case of a phenomenon called the ***Simpson's Paradox*** where the trends in different groups of data are reversed after the data is aggregated. An important implication of this paradox is that causal inferences from observational data can potentially lead to flawed analysis and insight generation. Hence, effective validation mechanisms must be enforced to rule out

Statistical Fallacies

Simpson's Paradox



It should really be Yule's paradox.

Trends in subgroups do not guarantee a trend in the same direction but may be reversed.

Screenshot is from Great Courses > Mind Bending Mathematics > Lesson: Strangeness in Statistics.

Other examples are given with good explanations including a medical example: a procedure B may be better overall to remove kidney stones, but procedure A may be better for both small and large stones.

Statistical Fallacies

STATISTICAL FALLACIES

Statistical fallacies are misleading statements or arguments expressed with numbers. Statistics can present us with good evidence for a claim or be part of a plausible chain of reasoning. But they are frequently used to deceive us, to get us to accept a conclusion that we should reject or question. Here are a few examples:

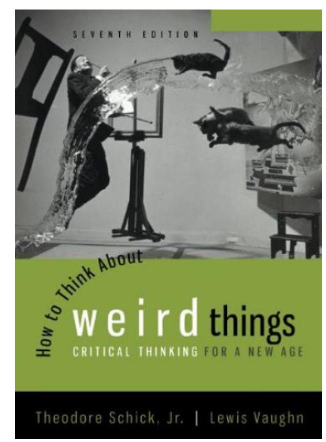
Misleading Averages

In statistics, there are three kinds of averages—mean, median, and mode. A *mean* is what most people refer to as an average. The mean of the five numbers 2, 3, 5, 8, and 12 is 6 (2 + 3 + 5 + 8 + 12 = 30 divided by 5 = 6). The *median* is the middle value in a sequence of numbers (the median of our five numbers is 5). The *mode* is the most frequently appearing value in a series.

Trouble comes when people don't specify which kind of average they are using, or they employ the kind that will make their weak case look strong. Imagine that the president promises a huge tax cut for the whole country, amounting to a mean tax savings of $10,000. But the mean has been driven upward by a few very rich people whose tax savings will be $1,000,000 or more. Ninety-five percent of taxpayers (who make less than $20,000 a year) will see a tax savings of less than $400. The president's boast of a mean tax savings of $10,000 is technically accurate—but deceiving. More truth can be told to the taxpayers by the median, which is $300, or even the mode of $250.

Missing Values

Much mischief can occur when people fail to distinguish between relative and absolute statistical values. Suppose you read that in the last year there has been a 75 percent increase in the number of muggings in your town. This sounds serious. But the 75 percent is the *relative*



Statistical Fallacies

Will Rogers' Phenomenon

Assume that you are tabulating survival for patients with a certain type of tumour. You separately track survival of patients whose cancer has metastasized and survival of patients whose cancer remains localized. As you would expect, average survival is longer for the patients without metastases. Now a fancier scanner becomes available, making it possible to detect metastases earlier. What happens to the survival of patients in the two groups?

The group of patients without metastases is now smaller. The patients who are removed from the group are those with small metastases that could not have been detected without the new technology. These patients tend to die sooner than the patients without detectable metastases. By taking away these patients, the average survival of the patients remaining in the "no metastases" group will improve.

What about the other group? The group of patients with metastases is now larger. The additional patients, however, are those with small metastases. These patients tend to live longer than patients with larger metastases. Thus the average survival of all patients in the "with-metastases" group will improve.

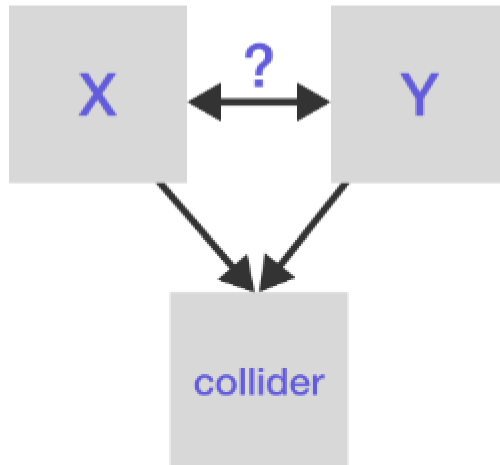
Changing the diagnostic method paradoxically increased the average survival of both groups! This paradox is called the Will Rogers' phenomenon after a quote from the humorist Will Rogers ("When the Okies left California and went to Oklahoma, they raised the average intelligence in both states"). (www)

See also Festenstein, 1985 (www)

Kingston University London

Statistical Fallacies

Berkson's Paradox ~ Collider Bias



Berkson's Paradox

Adam Strandberg, **Christopher Williams**, **Ken Jennison**, and 1 other contributed

Berkson's paradox is a result in statistics, very closely related to Simpson's paradox, that demonstrates that two values can statistically be negatively correlated even when they appear positively correlated in the population. In a high school, being taller may appear to be positively correlated with being good at math. However, statistically, a student's height and math skills are not correlated, and being taller doesn't make someone better or worse at math. It is simply the case that taller students tend to be older and have studied more math. Many applications of Berkson's paradox are less obvious.

Berkson's paradox is a particular kind of selection bias, or statistical result, caused by systematically observing some events more than others. In this paradox, observations are restricted to those where two variables sum together. If you know that $A + B$ must be within a certain range, then having a high A results in a lower B, and vice versa. Since observations are often more or less likely given a combination of variables, Berkson's paradox is ubiquitous. Berkson's paradox, and selection bias in general, appear in almost every field of research, particularly epidemiology and economics.

Kingston University London

Regression to the Mean (RTM)

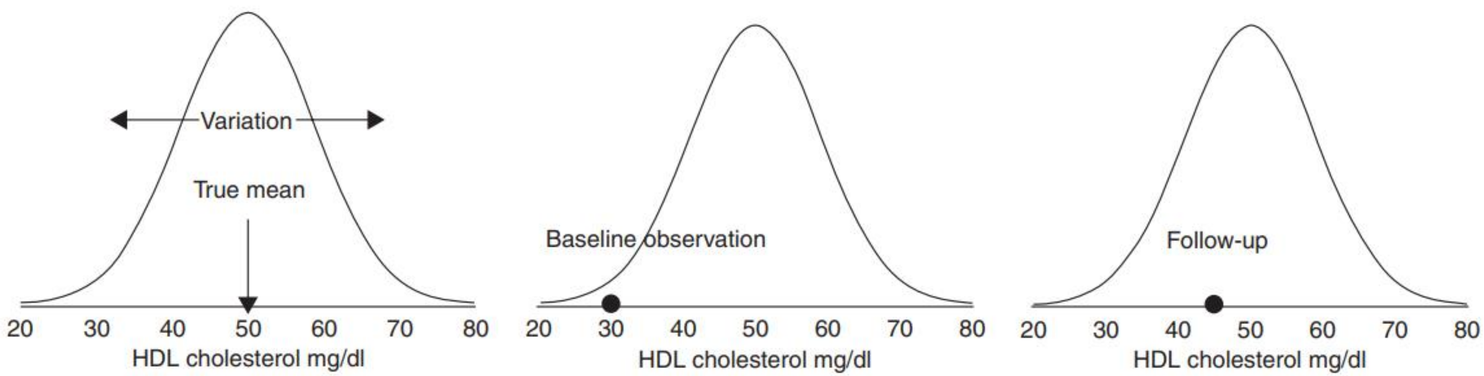


Figure 1 Graphical example of true mean and variation, and of regression to the mean using a Normal distribution. The distribution represents high density lipoprotein (HDL) cholesterol in a single subject with a true mean of 50 mg/dl and standard deviation of 9 mg/dl

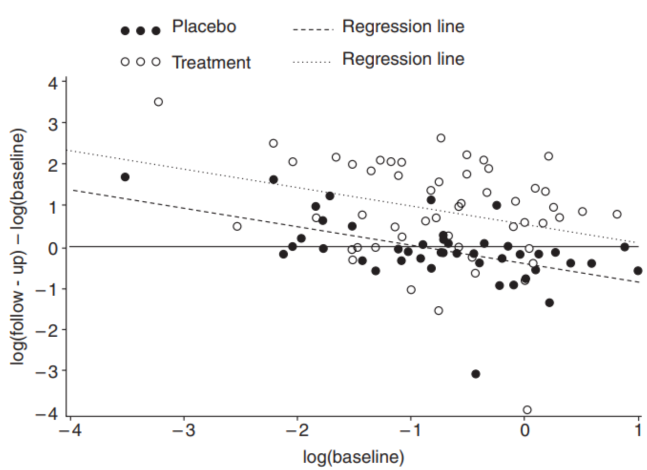


Figure 3 Scatter-plot of n = 96 paired and log-transformed betacarotene measurements showing change (log(follow-up) minus log(baseline)) against log(baseline) from the Nambour Skin Cancer Prevention Trial. The solid line represents perfect agreement (no change) and the dotted lines are fitted regression lines for the treatment and placebo groups

RTM is a ubiquitous phenomenon in repeated data and should always be considered as a possible cause of an observed change. Its effect can be alleviated through better study design and use of suitable statistical methods.

IJE vol.34 no.1 © International Epidemiological Association 2004; all rights reserved. *International Journal of Epidemiology* 2005;34:215–220
Advance Access publication 27 August 2004 doi:10.1093/ije/dyh299

EDUCATION

Regression to the mean: what it is and how to deal with it

Adrian G Barnett,[1] Jolieke C van der Pols[1] and Annette J Dobson[1]

Kingston University London

Regression Fallacy

The Regression Fallacy

The regression fallacy fails to account for natural fluctuations and rather ascribes cause where none exists.

Ignoring the statistical nature of regression towards the mean and attributing a causal or deterministic role to it.

KEY POINTS

- Things such as golf scores, the earth's temperature, and chronic back pain fluctuate naturally and usually regress towards the mean. The logical flaw is to make predictions that expect exceptional results to continue as if they were average.
- People are most likely to take action when variance is at its peak. Then, after results become more normal, they believe that their action was the cause of the change, when in fact, it was not causal.
- In essence, misapplication of regression to the mean can reduce all events to a "just so" story, without cause or effect. Such misapplication takes as a premise that all events are random, as they must be for the concept of regression to the mean to be validly applied.

See also:

https://plus.maths.org/content/maths-minute-regression-mean

Examples of the Regression Fallacy

- *When his pain got worse, he went to a doctor, after which the pain subsided a little. Therefore, he benefited from the doctor's treatment.*The pain subsiding a little after it has gotten worse is more easily explained by regression towards the mean. Assuming the pain relief was caused by the doctor is fallacious.
- *The student did exceptionally poorly last semester, so I punished him. He did much better this semester. Clearly, punishment is effective in improving students' grades.* Often, exceptional performances are followed by more normal performances, so the change in performance might better be explained by regression towards the mean. Incidentally, some experiments have shown that people may develop a systematic bias for punishment and against reward because of reasoning analogous to this example of the regression fallacy.
- *The frequency of accidents on a road fell after a speed camera was installed. Therefore, the speed camera has improved road safety*. Speed cameras are often installed after a road incurs an exceptionally high number of accidents, and this value usually falls (regression to mean) immediately afterwards. Many speed camera proponents attribute this fall in accidents to the speed camera, without observing the overall trend.
- Some authors have claimed that the alleged "Sports Illustrated Cover Jinx" is a good example of a regression effect: extremely good performances are likely to be followed by less extreme ones, and athletes are chosen to appear on the cover of Sports Illustrated only after extreme performances. Assuming athletic careers are partly based on random factors, attributing this to a "jinx" rather than regression, as some athletes reportedly believed, would be an example of committing the regression fallacy.

Kingston University London

Statistical Fallacies

Winner's Curse

Statistical Thinking for the 21st Century

Draft: 2022-12-22

Another kind of error can also occur when statistical power is low: Our estimates of the effect size will be inflated. This phenomenon often goes by the term “winner’s curse”, which comes from economics, where it refers to the fact that for certain types of auctions (where the value is the same for everyone, like a jar of quarters, and the bids are private), the winner is guaranteed to pay more than the good is worth. In science, the winner’s curse refers to the fact that the effect size estimated from a significant result (i.e. a winner) is almost always an overestimate of the true effect size.

We can simulate this in order to see how the estimated effect size for significant results is related to the actual underlying effect size. Let’s generate data for which there is a true effect size of d = 0.2, and estimate the effect size for those results where there is a significant effect detected. The left panel of Figure 18.2 shows that when power is low, the estimated effect size for significant results can be highly inflated compared to the actual effect size.

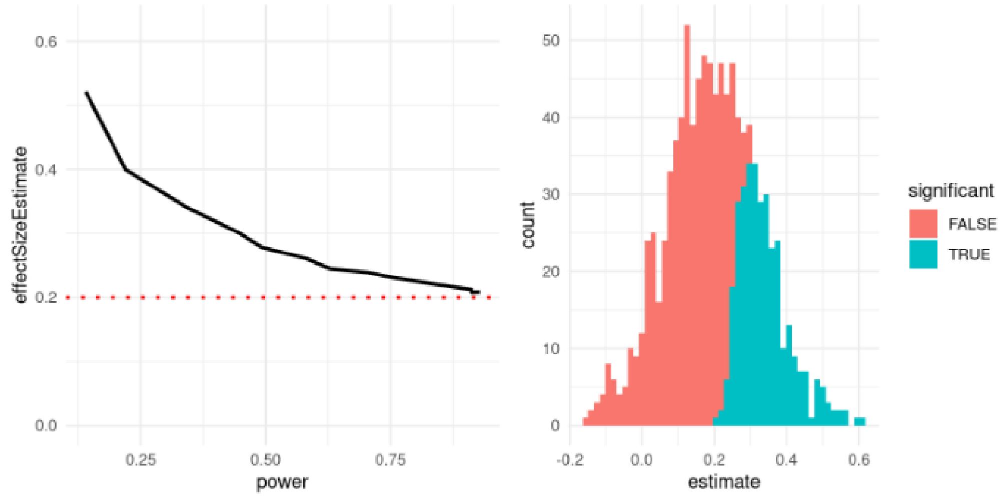


Figure 18.2: Left: A simulation of the winner’s curse as a function of statistical power (x axis). The solid line shows the estimated effect size, and the dotted line shows the actual effect size. Right: A histogram showing effect size estimates for a number of samples from a dataset, with significant results shown in blue and non-significant results in red.

Statistical Fallacies

Other fallacies [edit]

Pseudoreplication is a technical error associated with analysis of variance. Complexity hides the fact that statistical analysis is being attempted on a single sample (N=1). For this degenerate case the variance cannot be calculated (division by zero). An (N=1) will always give the researcher the highest statistical correlation between intent bias and actual findings.

The gambler's fallacy assumes that an event for which a future likelihood can be measured had the same likelihood of happening once it has already occurred. Thus, if someone had already tossed 9 coins and each has come up heads, people tend to assume that the likelihood of a tenth toss also being heads is 1023 to 1 against (which it was before the first coin was tossed) when in fact the chance of the tenth head is 50% (assuming the coin is unbiased).

The prosecutor's fallacy[29] assumes that the probability of an apparently criminal event being random chance is equal to the chance that the suspect is innocent. A prominent example in the UK is the wrongful conviction of Sally Clark for killing her two sons who appeared to have died of Sudden Infant Death Syndrome (SIDS). In his expert testimony, now discredited Professor Sir Roy Meadow claimed that due to the rarity of SIDS, the probability of Clark being innocent was 1 in 73 million. This was later questioned by the Royal Statistical Society;[30] assuming Meadows figure was accurate, one has to weigh up all the possible explanations against each other to make a conclusion on which most likely caused the unexplained death of the two children. Available data suggest that the odds would be in favour of double SIDS compared to double homicide by a factor of nine.[31] The 1 in 73 million figure was also misleading as it was reached by finding the probability of a baby from an affluent, non-smoking family dying from SIDS and squaring it: this erroneously treats each death as statistically independent, assuming that there is no factor, such as genetics, that would make it more likely for two siblings to die from SIDS.[32][33] This is also an example of the ecological fallacy as it assumes the probability of SIDS in Clark's family was the same as the average of all affluent, non-smoking families; social class is a highly complex and multifaceted concept, with numerous other variables such as education, line of work, and many more. Assuming that an individual will have the same attributes as the rest of a given group fails to account for the effects of other variables which in turn can be misleading.[33] The conviction of Sally Clark was eventually overturned and Meadow was struck from the medical register.[34]

The ludic fallacy. Probabilities are based on simple models that ignore real (if remote) possibilities. Poker players do not consider that an opponent may draw a gun rather than a card. The insured (and governments) assume that insurers will remain solvent, but see AIG and systemic risk.

Other types of misuse [edit]

Other misuses include comparing apples and oranges, using the wrong average,[35] regression toward the mean,[36] and the umbrella phrase garbage in, garbage out.[37] Some statistics are simply irrelevant to an issue.[38]

Anscombe's quartet is a made-up dataset that exemplifies the shortcomings of simple descriptive statistics (and the value of data plotting before numerical analysis).

See also [edit]

- Deception
- Ecological fallacy
- Ethics in mathematics
- Metascience
- Misuse of p-values
- Misleading graph
- Post hoc analysis
- Simpson's paradox
- Statcheck

Fallacy Files

Kingston University London

Fallacies

- What are *The Fallacy Files*?
- Main Page
- Weblog (RSS/XML)
- What is a fallacy?
- How to use *The Fallacy Files*
- Quote...Unquote
- **Examples:** Fallacies, Arguments
- Book Shelf
- Sources & Resources
- Familiar Misquotations
- How to be a Prophet
- How to Read a Poll
- Funny Fallacies
- Humorous Headlines
- Glossary
- Taxonomy
- How to Use the Taxonomy
- About the Author

Fallacy Files

Texas Sharpshooter

> The Texas Sharpshooter Fallacy is a logical fallacy based on the metaphor of a gunman shooting the side of a barn, then drawing targets around the bullet-hole clusters to make it look like he hit the target.

Related concepts:

- Anthropic principle – Philosophical principle about the occurrence of sapient life in the Universe
- Availability heuristic – Heuristic bias that if something can be recalled, it must be important
- Confirmation bias – Bias confirming existing attitudes
- HARKing – Acronym for "Hypothesizing after the results are known"
- Look-elsewhere effect
- Overfitting – Flaw in machine learning computer model
- Postdiction – Explanations given after the fact
- Ramsey theory – Branch of mathematical combinatorics
- Scan statistic

Human Reproduction, Vol.32, No.7 p. 1363, 2017
Advanced Access publication on May 24, 2017 doi:10.1093/humrep/dex103

human reproduction EDITORIAL

The Texas sharpshooter fallacy

J.L.H. (Hans) Evers*

Kingston University London

Pseudoreplication

Pseudoreplication: Don't Fall For This Simple Statistical Mistake

Published January 23, 2012
Posted in: Lab Statistics & Math

A doctor is measuring cholesterol levels in the blood of his male patients. Twenty men are each subjected to two blood tests each.

Pop Quiz: What are the degrees of freedom?

Despite there being 40 blood tests, the sample size is 20 (the number of men tested) so df = 19. Two blood tests taken from the same man cannot be **independent** of each other, a necessity for true replication. Though they can provide internal consistency, these multiple tests cannot be used as replicates.

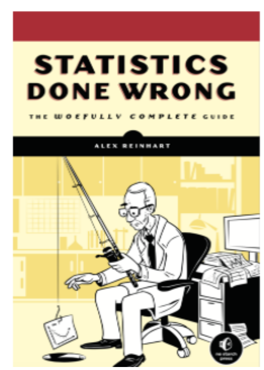
BiteSize Bio
LET'S DO SCIENCE BETTER

Do not use your replicates as independent samples; they are NOT!

Pseudoreplication: choose your data wisely

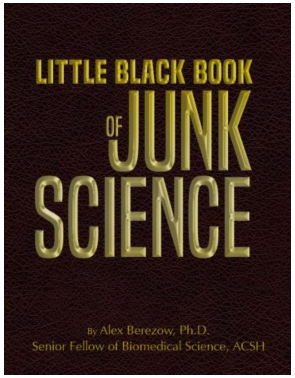
Many studies strive to collect more data through replication: by repeating their measurements with additional patients or samples, they can be more certain of their numbers and discover subtle relationships that aren't obvious at first glance. We've seen the value of additional data for improving statistical power and detecting small differences. But what exactly counts as a replication?

Let's return to a medical example. I have two groups of 100 patients taking different medications, and I seek to establish which medication lowers blood pressure more. I have each group take the medication for a month to allow it to take effect, and then I follow each group for ten days, each day testing their blood pressure. I now have ten data points per patient and 1,000 data points per group.



Kingston University London

Pseudoscience / Junk Science



ACRYLAMIDE

One of those evil-sounding chemicals that chemophobes worry about, acrylamide is just a molecule formed in a process called the "Maillard reaction," which is responsible for the brown color of cooked foods. Humans have been cooking food far longer than environmentalists have existed, so we've been eating acrylamide for hundreds of thousands of years. It's safe to eat toast.

ACUPUNCTURE

Sticking tiny needles into your face or all over your body is not effective for treating pain or any other health problem. If there are any benefits from acupuncture, they are attributable to the placebo effect.

ALKALINE DIET

The "alkaline diet" is premised upon several myths: (1) We eat too much processed food, sugar, and GMOs; (2) Those foods cause our bodies to become "acidic"; (3) Acidic bodies generate cancer; (4) Therefore, cancer can be prevented by eating an alkaline diet[2]. There are no health problems associated with eating processed food or GMOs. Furthermore, the food we eat does not change our blood pH, which our bodies tightly regulate to remain within the range of 7.35 to 7.45.

ALTERNATIVE MEDICINE

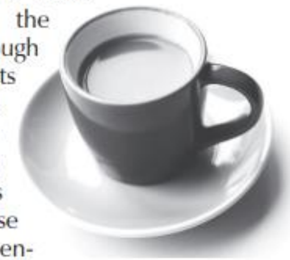
Alternative medicine includes everything from acupuncture and Traditional Chinese Medicine to herbal remedies and homeopathy. If alternative medicine worked it would just be called medicine. Instead, the vast majority of practices that fall under the alternative label lack scientific evidence.

CARCINOGENS

Judging by environmentalists and sensationalist media, you are very lucky that you haven't caught cancer (yet) from coffee, toast, bacon, or even that new car smell. Because many are natural, carcinogens are everywhere. Coffee, for example, has dozens of carcinogens but they are present in very tiny amounts. The dose makes the poison.

COFFEE

Coffee is probably the most studied beverage on the planet. Ironically, though it contains tiny amounts of many carcinogens, there appear to be some health benefits of drinking it. Always remember that "the dose makes the poison." So, enjoy a daily cup of joe or two. Just not 7,000.

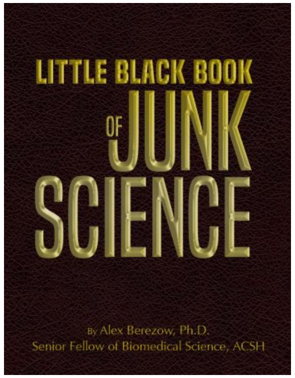


ESSENTIAL OILS

Essential oils acquired their name because they contain the "essence" of the plant from which they are derived, not because they are essential to human nutrition. Like many plant extracts, essential oils may contain useful compounds, such as natural food preservatives and antimicrobials[30]. Depending on the plant, they also might smell nice. Beware (essential?) snake oil salesmen offering it as a cure for anything.

Kingston University London

Pseudoscience / Junk Science



FOOD ADDITIVES

Anything that is added to our food must be approved (either expressly or tacitly) by the FDA. Food colors or dyes and preservatives are perfectly safe.

FOOD PRESERVATIVES

Processed food often contains preservatives, which maintain freshness, increase shelf-life, prevent foodborne illness, and reduce food waste. Preservatives are safe, and many come from natural sources, particularly fruits[37]. For unknown reasons, BHT (found in algae but produced commercially) and BHA (a synthetic compound) scare people, but they shouldn't.

GILLES-ÉRIC SÉRALINI

A hero of the anti-GMO movement, Gilles-Éric Séralini famously conducted an unethical animal experiment in which he grew giant tumors on rats and then blamed GMOs and the herbicide glyphosate. His methods were severely flawed, and his claims are consistently in defiance of the scientific literature.

GMOS

Genetically modified organisms represent one of the biggest advances in the history of science. (Pro tip: Scientists don't use the term GMOs because that is a legal term; instead, they call them "transgenic crops.") With surgical precision, molecular biologists place unique genes into a plant's genome, conferring new properties that hybridizing could never achieve. It has allowed the creation of insect-resistant crops (which require fewer insecticides), fast-growing salmon (that will reduce overfishing), non-browning apples, and rice that provides extra vitamin A to malnourished people. Everyone who cares about feeding people with the smallest possible environmental footprint embraces this technology.

HIGH-FRUCTOSE CORN SYRUP

Table sugar (sucrose) consists of two sugars linked together: glucose and fructose. As its name implies, high-fructose corn syrup contains more fructose than glucose. But so does honey, which has roughly the same ratio of fructose to glucose[47].

HOMEOPATHY

Homeopathy is based upon two crazy beliefs: That water molecules can "remember" the presence of other molecules and that extremely dilute solutions of poisons can cure people of various illnesses. Chemically, the solutions are often so dilute that the original substance is no longer present. Homeopathic products are a scam.

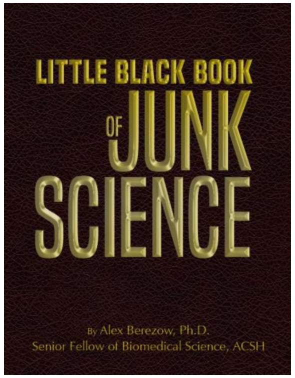
INTERNATIONAL AGENCY FOR RESEARCH ON CANCER (IARC)

The International Agency for Research on Cancer (IARC) was created to identify possible carcinogens for further study, studies that would include risk. Today, IARC spends its time blaming perfectly normal things – like hot water and bacon – for causing cancer. Any proclamation made by IARC is now met with serious skepticism.

"NATURAL IS BETTER"

The widespread myth that "natural is better" underlies everything from alternative medicine to the organic food movement. It's the poster child for junk science. Smallpox, HIV, arsenic, poison ivy, rattlesnakes, and scorpions are all natural, while many beneficial medicines are not.

Pseudoscience / Junk Science



NOCEBO EFFECT

Imagine participating in a study, and you are told that the pill you take will make you woozy. A few minutes later, you are feeling a bit woozy, but then you find out you just took a sugar pill. Your "side effects" were completely psychosomatic because you expected them to occur. Such is the power of the nocebo effect. It explains why some people claim that Wi-Fi makes them sick. Essentially, the nocebo effect is the opposite of the placebo effect.

ORGANIC FOOD

Organic food is a gigantic scam. Despite marketing claims, $12 bananas aren't healthier, tastier, more nutritious, or better for the environment. Organic farmers also use pesticides, though they are quite content letting the public believe otherwise. Because organic farming is inefficient, we could not feed the world using it alone[57].

OSTEOPATHY

In the United States, osteopaths are nearly identical to medical doctors, but in other parts of the world, they are basically chiropractors[58]. If an osteopath offers to cure your baby's sniffles by shaking it, find an M.D.

PHARMACEUTICAL INDUSTRY (BIG PHARMA)

Big Pharma has produced medicines and vaccines that have saved literally hundreds of millions of lives. They have also conducted ethically outrageous clinical trials, gouged patients on drug prices, downplayed dangerous side effects, and manipulated the scientific literature. Demeaning an entire industry due to the actions of a few players throughout history is not constructive. In modern times, generic companies have had similar pricing and ethical issues.

PHTHALATES

In use for nearly 100 years, phthalates are necessary to make tubing flexible. Epidemiological studies of possible adverse effects of phthalates have been contradictory while toxicology reports show no risk because animal doses were thousands of times higher than possible human exposure. Ignore the chemophobic hype.

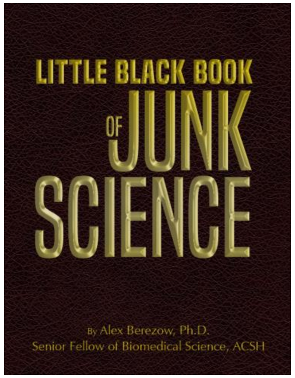
PLACEBO EFFECT

Imagine you are in a clinical trial for a pill that will make you feel stronger, smarter, and more energetic. It will even increase your libido! After a few days of taking the pill, that's exactly how you feel. A miracle, right? Not if the drug was a sugar pill – a placebo – which is inert. You felt amazing because you expected to feel amazing. This is a well-documented phenomenon called the placebo effect, and it's precisely why new drugs are tested against placebos to see if they actually work. The placebo effect explains why some people claim alternative medicine "works."

PREDIABETES

A slightly elevated blood glucose level now has its own diagnosis: Prediabetes. But is it real? Not at the A1C level the CDC chose. Instead, only about 5% of those who score a 5.8 on a glycohemoglobin test go on to develop diabetes[68]. Using their arbitrary number, the CDC claims 80 million Americans are prediabetic[69], which creates an unnecessary panic. Their reasoning is dubious[70].

Pseudoscience / Junk Science



PUBLICATION BIAS

Like newspapers, scientific journals want readers. So, they are biased toward publishing flashy new results and biased against "boring" research, such as negative results or replications. Unfortunately, boring research is often more important and better supported.

RECALL BIAS

If you get diagnosed with a scary or rare disease, you are likely to remember your life differently than people without the disease. This is a well-known phenomenon called recall bias, and it explains why some epidemiological studies are inaccurate. As a general rule, people's memories are notoriously unreliable.

STATISTICAL SIGNIFICANCE

Statistical significance is a fancy way to say, "We think these results aren't due to a random fluke." In other words, the result is real with a certain degree of confidence. However, just because a result is not a random fluke does not mean it is important or relevant. If a chemical raises your risk of cancer from one in a million to two in a million, it may be statistically significant but it's irrelevant to your life. A lot of junk science claims result from misuse of statistical significance.

SODIUM NITRATE/NITRITE

Sodium nitrate and sodium nitrite are added to food as preservatives. Most of the nitrates in our diet (perhaps 85%) come from vegetables, while nitrites come mainly from processed meats, which are added to prevent botulism[84]. Junk science foodies claim nitrates and nitrites cause cancer, but they provide no believable explanation. Both nitrates and nitrites have beneficial health effects, and nitrate/nitrite conversions occur within our bodies.

VITAMIN SUPPLEMENTS

Healthy people who eat a relatively balanced diet (i.e., one that does not consist exclusively of hot dogs and Twinkies) do not need to take multivitamins or other dietary supplements. If you think you are deficient in a particular vitamin (such as vitamin D if you live in a cloudy place), consult your doctor. Despite being debunked for decades, junk science claims mega-doses of vitamin C will cure or prevent colds. At the very most, daily supplementation with vitamin C might reduce the duration and severity of a cold[96].

Pseudoscience / Junk Science

Skeptical Inquirer
THE MAGAZINE FOR SCIENCE AND REASON

SUBSCRIBE RENEW MENU

Category: Medical Pseudoscience around the World

Medical Pseudoscience around the World

Homeopathy Research Hits New Low

Skeptical Inquirer Volume 47, No. 3
May/June 2023
Norbert Aust and Viktor Weisshäupl

"Homeopathy in Cancer Patients: Almost Too Good to Be True." That was the headline of an article in the October 23, 2022, issue of the Austrian weekly news magazine Profil reporting on an investigation by the Austrian Agency for Scientific Integrity (OeAWI) (Schönberger 2022). The subject of the investigation was a study on the use …

This article is available for free to all.

Medical Pseudoscience around the World

Rise of Ayurveda: A Dangerous Trend to Decolonize the Scientific Method

Skeptical Inquirer Volume 47, No. 3
May/June 2023
Samit Ghosal

Progress made in India in the past decade in digitization, banking reforms, and economic structuring has been phenomenal. However, there seems to be an inversely proportional relationship between economic/technological advancement and health/education in the present-day political India. It seems that the concept of going back to one's roots has been exclusively reserved for health and …

This article is available for free to all.

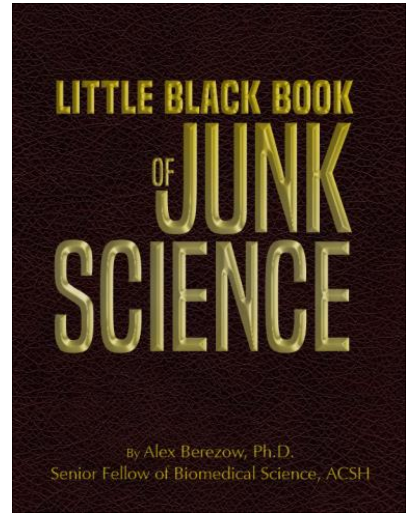
Kingston University London

Media Hoaxes

Bad Science

The media's MMR hoax

August 30th, 2008 by Ben Goldacre in bad science, badscience, MMR | 57 Comments »



AMERICAN COUNCIL ON SCIENCE AND HEALTH

Promoting science and debunking junk since 1978.

This website is for educational purposes.

Home About Donate Publications Media/Contact Subscribe Write For Us

Health Hoaxes And Health Hoaxes, Revisited

FactCheck.org®

SciCheck

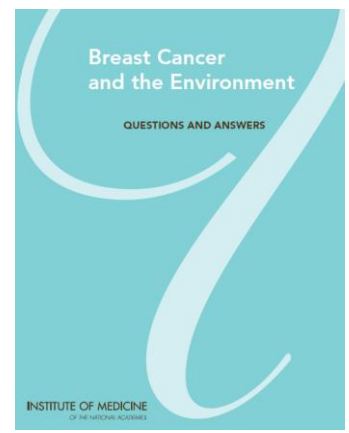
FactCheck.org's SciCheck feature focuses exclusively on false and misleading scientific claims that are made by partisans to influence public policy. It was launched in January 2015 with a grant from the Stanton Foundation. The foundation was founded by the late Frank Stanton, president of CBS for 25 years, from 1946 to 1971.

Media Hoaxes

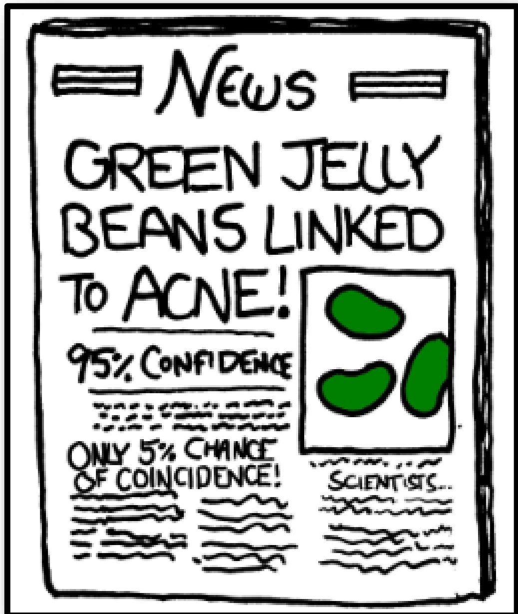
QUESTION What does it mean to say 1 in 8 women will get breast cancer?

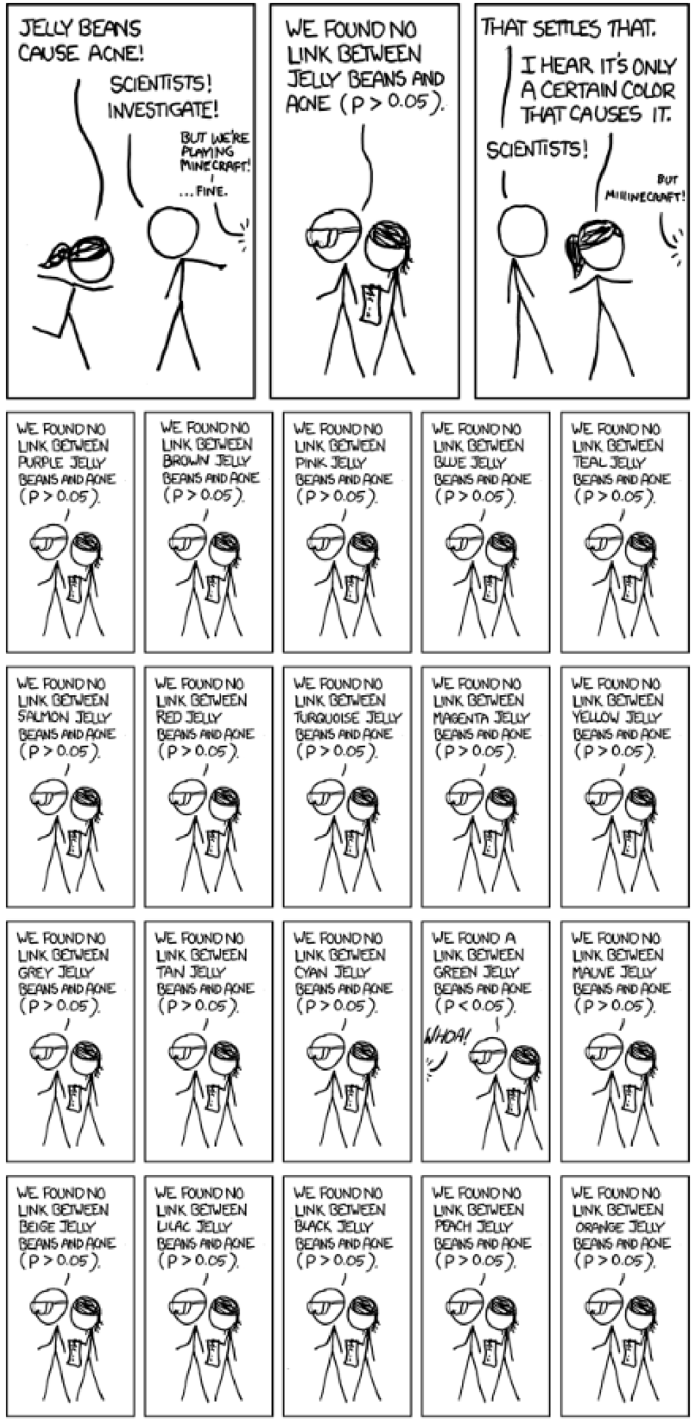
ANSWER The 1 in 8 number we often hear is the risk of getting breast cancer during a woman's lifetime. But it does not mean that 1 in every 8 women is diagnosed with breast cancer each year. A breast cancer diagnosis is never good news, and thankfully most women will never get one. In fact, the National Cancer Institute estimates that if a group of 1,000 women were followed for 10 years from their 50th to their 60th birthdays, about 20 to 30 of them would be diagnosed with breast cancer by their 60th birthday. The other 970 to 980 women in this group would not develop breast cancer during these 10 years – although some of them might develop it later in life.

Age alone is a big factor in who develops breast cancer. Until women reach their thirties, the chance of being diagnosed with breast cancer is very low, and after that the risk begins to rise. The risk of breast cancer is at its highest when women are in their sixties and seventies. Women who begin menstruating later, have a first child at a younger age, or enter menopause earlier will tend to have a relatively lower risk of breast cancer.



Kingston University London





SIGNIFICANT

Randall Munroe's cartoon on statistical significance

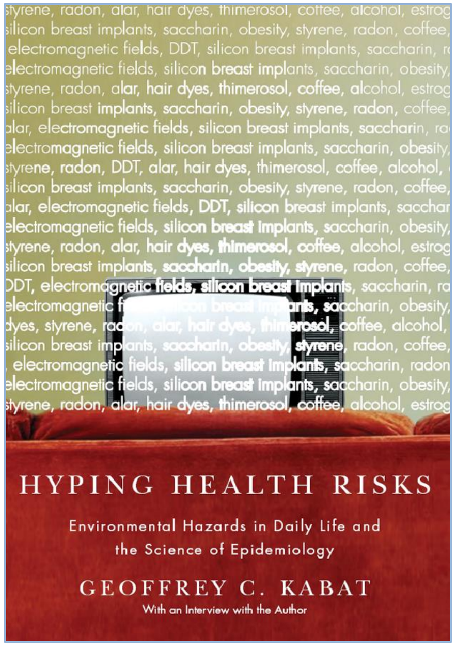
Explanation

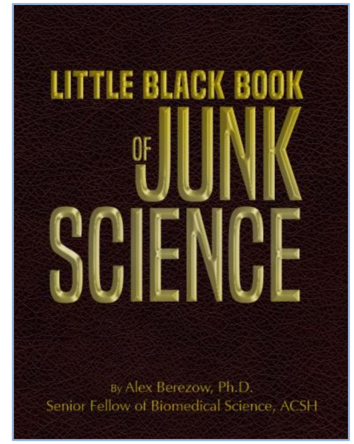
Kingston University London

(Social) Media Hoaxes

Detox, superfoods, fat, sugar and natural sugar, alkali water etc.

Hyping Health Risks

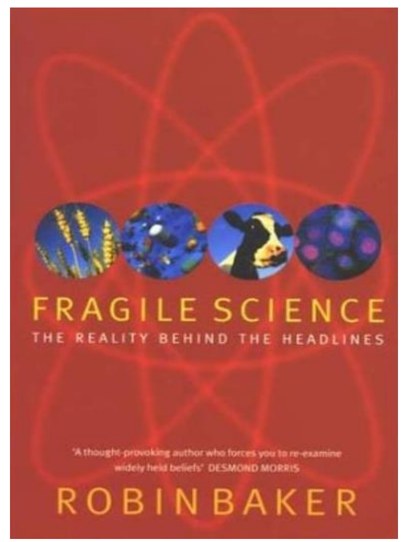


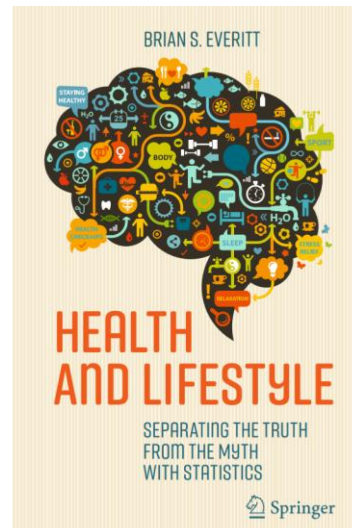


EDITORIALS

Promoting Healthy Skepticism in the News: Helping Journalists Get It Right

Steven Woloshin, Lisa M. Schwartz, Barnett S. Kramer





Hyping Health Risks

Panel 2
Public health estimates in the headlines

Hunger kills 6 million children a year.[1] **BBC news story, 2005**

Few children actually die of starvation alone. An estimated 11 million children younger than 5 years died in the year 2000. Undernutrition was an underlying cause of more than 50% of these deaths, as were several infectious diseases such as diarrhoea, pneumonia, and malaria.

AIDS, without a doubt, is the greatest epidemic in the history of mankind.[56] **Peter Piot, Executive Director of UNAIDS, 2004**

An epidemic is a time-limited simultaneous occurrence of disease in a population, often caused by a new infectious agent introduced from the outside. AIDS is therefore no longer an epidemic—it has existed for more than 25 years. It is true that more than 20 million people have died of AIDS during this time. But the 1918 influenza epidemic, which lasted for 2 years, is estimated to have killed up to 100 million.[57]

"I will tell you today unequivocally that the risk of another pandemic influenza is one. It's a one."[58] **Michael Osterholm, director of the Center for Infectious Disease Research and Policy, 2005**

As pandemic influenza does occur routinely, technically the probability of it occurring again in an unspecified time period is 1. However, we do not know whether the pandemic will occur next year, in 10 years, or in 50 years, or even whether if the virus mutates it will occur in a highly virulent and deadly form.

Panel 3
How to use public-health estimates

1 Ignore any estimate that is not accompanied by a clear description of input data, assumptions, and methods

2 Examine the quality of the estimate: has it been reviewed by independent technical experts who are identified by name? Are the tools and input data available for review? Have country-level scientists participated in the development and validation of the estimate?

3 What metric is being used in the estimate? Think about how different metrics for the same disease or condition might lead to different interpretations. When comparing across diseases, the comparisons should be done across a range of metrics, not just one.

4 Examine the measure being estimated—is it meaningful and comparable to the other conditions being considered in the priority-setting process?

5 Be sceptical about "examplar-based estimates" and insist they be put in an appropriate context.

6 Question every change (whether point or trend) to determine how much might be due to changes in methods or data inputs.

7 Demand information on uncertainty, and use it to determine whether reported differences are meaningful in public-health terms.

Interpreting health statistics for policymaking: the story behind the headlines

Dr Neff Walker PhD [a], Jennifer Bryce EdD [b], Robert E Black MD [b]

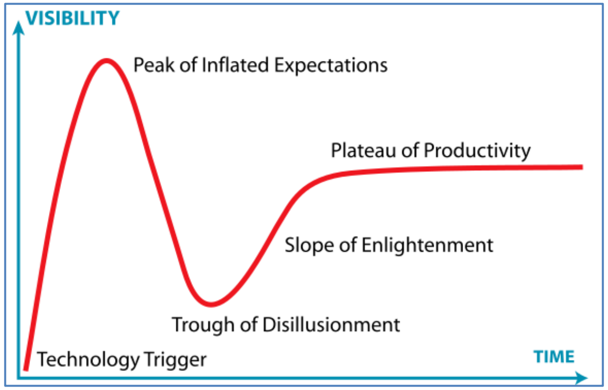
THE LANCET

Volume 369, Issue 9565, 17–23 March 2007, Pages 956-963

Hype Curve (in Technology)

Insights / Information Technology / Article

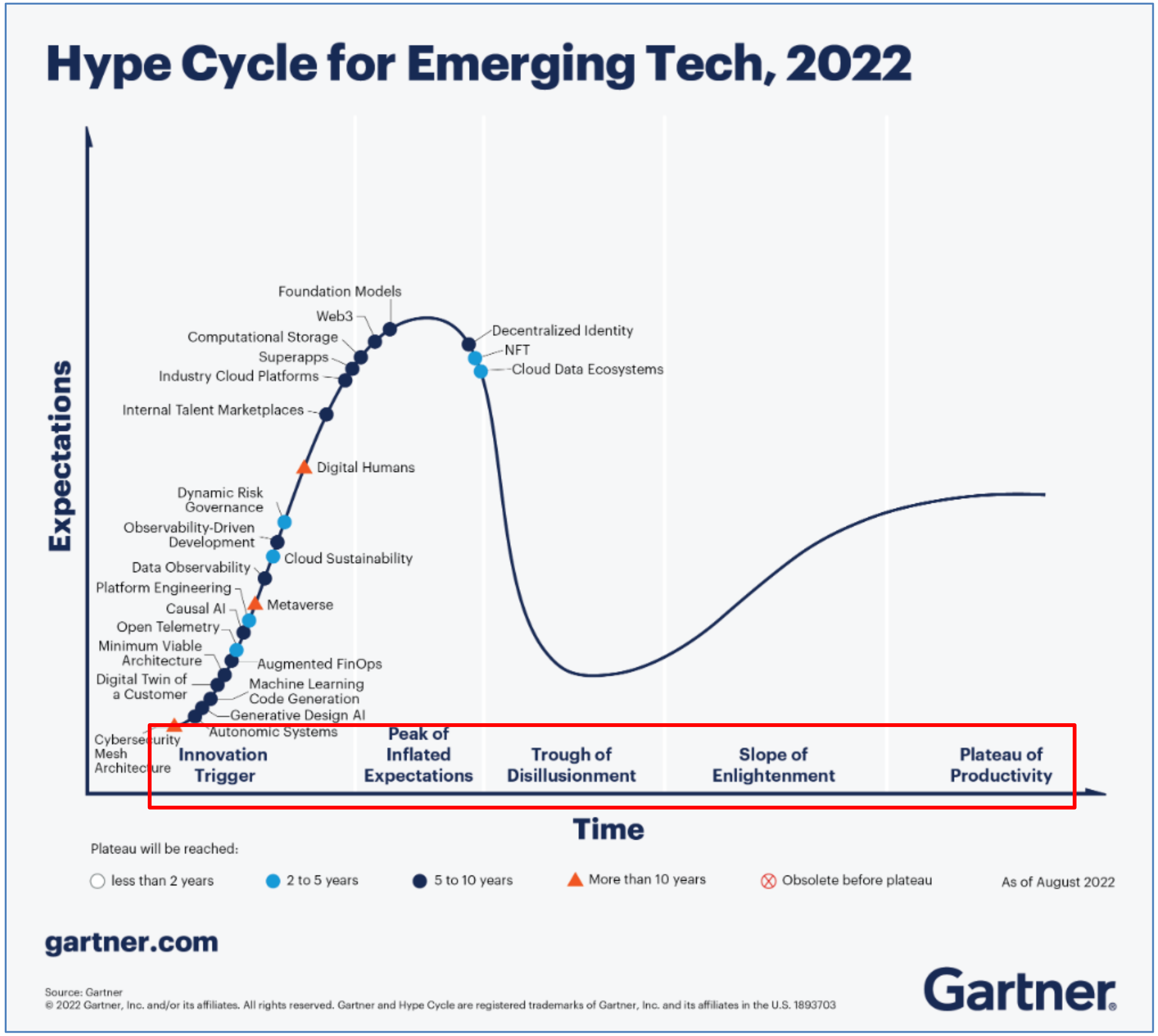
What is New in the 2022 Gartner Hype Cycle for Emerging Technologies



Gartner hype cycle

Article Talk

From Wikipedia, the free encyclopedia



Kingston University London

Reported Risks are Inflated

The New York Times

TheUpshot

THE NEW HEALTH CARE

A Link Between Alcohol and Cancer? It's Not Nearly as Scary as It Seems

Typically, risks associated with exposures are reported as exposed vs unexposed; or exposed at highest degree vs unexposed. Either has its pros and cons. It is best to get to the bottom of it and explore the data properly.

Kingston University London

Statistical Literacy

90% of Medical Research is False !

Why Most Published Research Findings Are False

John P. A. Ioannidis

It can be proven that most claimed research findings are false.

What's the probability that a significant p-value indicates a true effect?

nicebread / 2015-11-03

ShinyApps: Experience Statistics | About | Links | nicebread.de | Contact

When does a significant *p*-value indicate a true effect?

Understanding the Positive Predictive Value (PPV) of a *p*-value

Kingston University London

Statistical Literacy

ShinyApps: Experience Statistics | About | Links | nicebread.de | Contact

When does a significant *p*-value indicate a true effect?

Understanding the Positive Predictive Value (PPV) of a *p*-value

This app is based on

Ioannidis JPA (2005) Why most published research findings are false

Across all investigated hypotheses: What % of them is actually true?

% of a priori true hypotheses:

0 | 9 | 100

0 10 20 30 40 50 60 70 80 90 100

What is your Type I error (α; typically 5%)?

α level

0.005 | 0.05 | 0.1

0.005 0.015 0.025 0.035 0.045 0.055 0.065 0.075 0.085 0.095 0.1

% of p-hacked studies

0 | 30 | 100

0 10 20 30 40 50 60 70 80 90 100

Do you want to specify power directly or indirectly by specifying sample size per group and effect size? (Assuming a two-group t-test)

- (selected) specify power (1-β) directly
- specify power indirectly through sample size (n) and effect size (d)

Presets by Ioannidis (2005)

6: Adequately powered exploratory epidemiological study

Power (1-β)

0.01 | 0.8 | 0.99

0.01 0.11 0.21 0.31 0.41 0.51 0.61 0.71 0.81 0.91 0.99

What's the probability that a significant p-value indicates a true effect?

nicebread / 2015-11-03

Statistical Literacy

ShinyApps: Experience Statistics About Links nicebread.de Contact

When does a significant *p*-value indicate a true effect?

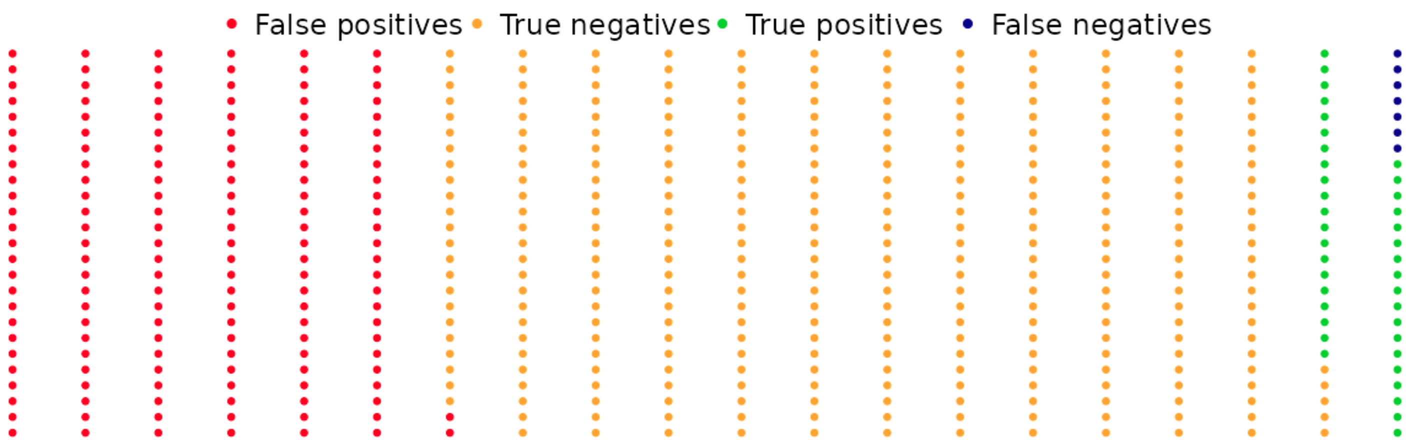
Understanding the Positive Predictive Value (PPV) of a *p*-value

This app is based on
Ioannidis JPA (2005) Why most published research findings are false

true positives: 7.7%; false negatives: 1.3%; true negatives: 60.5%; false positives: 30.5%

Positive predictive value (PPV): 20.2% of claimed findings are true
False discovery rate (FDR): 79.8% of claimed findings are false

If we consider all findings, it looks like this (each point is one study):



What's the probability that a significant p-value indicates a true effect?

nicebread / 2015-11-03

Things Cheaters Do

Comparing graphics with different scales or using scales that are unjustifiable

Graphics with scales not including the zero to augment small differences

Presenting relative risks without mentioning absolute risks

Emphasis on statistical significance (*P* value) without any mention of clinical significance (effect size and 95% CI or margin of error)

Survey with small sample sizes, biased samples or low response rates

Using percentages without giving the absolute numbers (denominators)

Using averages without any indication of the spread (dispersion) and outliers

Using oddly specific intervals or follow-up periods

Not taking into account subgroups in the data

Proving anything with statistics!

Presenting an observed correlation as causal

Things Cheaters Do

P Hacking

What is P Hacking: Methods & Best Practices

By Jim Frost — Leave a Comment

You've probably heard of p-hacking, but what is it? And is it really a concern? You might be surprised to learn that it has already affected the scientific literature negatively!

P-hacking is a set of statistical decisions and methodology choices during research that artificially produces statistically significant results. These decisions increase the probability of false positives—where the study indicates an effect exists when it actually does not.

Learn how p-hacking has already affected the scientific literature negatively, how it occurs, and the best practices for avoiding it.

[What is P-Hacking: Methods and Best Practices]

Things Cheaters Do

Misleading Graphs & Lying with Statistics

How to spot a misleading graph - Lea Gaslowitz

TED-Ed
18.5M subscribers
Subscribe
81K Share Download Clip

Identifying Misleading Graphs - Konst Math
Bill Konst
114K views • 12 years ago

Misleading Graphs Real Life Examples
Prof. Essa
104K views • 8 years ago

This is How Easy It Is to Lie With Statistics
Zach Star
5.4M views • 4 years ago

How to defend yourself against misleading statistics in the...
TEDx Talks
167K views • 6 years ago

Peter Donnelly: How stats fool juries
TED
232K views • 16 years ago

Data Visualization and Misrepresentation

Kingston University London

Rules of Thumb in Statistics

- *P* value does not prove anything, only indicates the probability of true-positivity (1 - *P*)
- *P* value is meaningless without the effect size (and its 95% confidence interval)
- *P* value (> 0.05) is meaningless if the study does not have sufficient statistical power
- There is no statistics for bias (with a few exceptions like publication bias testing in a meta-analysis) or confounding (other than statistical adjustment or controlling for a potential confounder if data is available); a small *P* value (< 0.05) does not rule out bias/confounding
- Correlation does not mean causation
- Extrapolation beyond the limits of data is dangerous
- Most standard statistics assumes linearity (a highly strong non-linear correlation would be statistically significant with tests for linearity)
- Using mean to represent complex data with high dispersion is dangerous
- Trends are more important than snap shots
- The absence of evidence is not evidence of absence

Rules of Thumb in Statistics

A Dozen Rules of Thumb for Journalists (RSS)

https://rss.org.uk/RSS/media/File-library/News/2020/rss-number-hygiene-list-2014.pdf

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

This document was prepared for the RSS Science Journalism Programme with assistance from Professor David Spiegelhalter and David Walker. It is not intended to be a prescriptive statement on what journalism students ought to know about statistics, rather, it is a guide to what those visiting media colleges might like to cover.

A dozen rules of thumb for journalists

Numbers are compelling, but treacherous. They can make a story, but they are also open to misunderstanding and manipulation. Caution and a preparedness to check and check again should be a journalist's watchwords

1. You come across a number in a story or press release. Buyer beware. Before making it your own, ask who cooked it up. What are their credentials? What is their pitch? Do we have alternative evidence; what numbers are they not showing us; why this number, now? If the number comes from a study or research, has anyone reputable said the work is any good?

2. Sniff around. Do the numbers refer to a whole group of people or things or just a sample of them? If it's a sample, are the people being questioned or the things being referred fairly representative of the wider group? Say a company is claiming something applies to the population at large. If they mean it is a sample of the population, beware. A panel of internet users, say, that the company goes back to time and again may not be representative–not everyone uses the internet. Organisations use samples based on their own mailing lists, or on people who have received a free sample of their product, and the samples may be biased.

3. More probing. What questions were the sample asked? Wording can hugely influence the answers you get. In a jobs survey, our understanding of what it means to 'be employed' may differ; likewise in a crime survey our sense of what is 'violent'. The public's understanding may not correspond with the survey researcher's. Might a pollster's choice of words have led people into giving a particular and slanted response?

4. A single number is often used to sum up a group, the *average*. But different averages measure different things. Here are some definitions. The *mean* is extremely sensitive to highs and lows: Bill Gates coming to live in the UK would push up mean wealth. The *median* tells us, for example, the income of a person at the midpoint –half the population get less, half more. Comparing earnings, the *mode* tells us the salary most people earn.

5. Editors like a sure thing, but with numbers can be uncertain. We need to be sure the number on offer is not just due to chance. With a sample, check the *margin of error*, usually plus or minus 3 per cent. A poll saying 52 per cent of people are in favour of something is not a definitive statement: it could be 49 per cent. Uncertainty is inevitable when you are using a sample, so reputable polling companies state the margin of error around their confidence a sample does represent the wider population. Remember the margin of error tells us only about the sampling, not whether the right questions were asked appropriately.

6. Beware league tables, except in sports reports. Manchester United is higher than Chelsea for a simple and genuine reason: the side has collected more points. With hospitals or schools, a single score is unlikely to be a valid basis for comparing one with another. A teaching hospital may have a worse score, but only because sicker patients are referred to it. Comparisons between universities or police forces are unreliable if the scores fall within margins of error. Midshires scores 650 on the ranking and Wessex 669: they could be performing at the same level or their respective positions could be reversed.

7. The numbers show a big increase or sharp decrease. Yet a single change does not mean a trend. Blips happen often. Peaks and troughs go away, so we have to ask whether a change in the numbers is just a recovery or return to normal after a one-off rise or fall. This is what statisticians refer to as 'regression to the mean'. The numbers may come from a survey, such as ONS figures for household spending or migration. Is the change being recorded bigger than the margin of error?

8. After a controlled experiment (such as a trial of a new drug, based on a randomly chosen group, some of whom don't know they are getting a placebo), researchers are more confident in saying that a causes b. The numbers may show an association between two things, say obesity and cancer. But a correlation is not the same as saying obesity causes cancer. The connection may be spurious and explicable by a third or background factor. If children's use of mobile phones is associated with later behavioural disorders, the connexion could be the parents, and the way *their* behaviour affects both things. If the numbers do suggest an association, we have to try to assess whether it is plausible, on the back of other evidence. Finding a link may stimulate further study, but ought not itself to be the basis for action, let alone some new government policy. Recommendations for changing daily behaviour such as eating should not be based on speculative associations between particular food and medical conditions.

9. A question to pose of any number is 'out of how many?' Some events are rare --such as the death of a British child of junior school age. That's why they are news, but that's also why they have to be put in context. Noting scarcity value is part of good reporting, which tells us about an event's significance. The meaning of an event for an individual or family has to be distinguished from its public importance.

10. Billions and millionths are hard to grasp. We take in figures better if they are human scale. One comparison is between a number and the whole UK. Another is to capture the effect of an event or behaviour on an individual. Colourful comparisons can make risk intelligible: the risk of dying while being operated on under a general anaesthetic is on average the same as the risk of being killed while travelling 60 miles on a motorbike.

11. Good reporting gives a balanced view of the size of the numbers being reported. Better to focus on the most likely number rather than the most extreme, for example in stories about the effects of a flu pandemic. 'Could be as high as' points to an extreme; better to say 'unlikely to be greater than'. Think about how your audience will perceive a number.

12. Risk is risky. 'Eating bacon daily increases an individual's lifetime risk of bowel cancer by 20 per cent.' Another way of saying that is: out of 100 people eating a bacon sandwich every day one extra person will get bowel cancer. Using the first without noting the second tells a story that is both alarmist and inaccurate. If the information is available, express changes in risk in terms of the risks experienced by 100 or 100,000 people.

Kingston University London

Rules of Thumb in Statistics

Statistical Rules of Thumb

1 The Basics

1. Any statistical treatment must address the questions

- What is the question?
- Can it be measured?
- When, where, and how will you get the data?
- What do you think the data are telling you?

2. Observation is selection
3. Replicate to characterize random variation
4. Variability occurs at multiple levels
5. Invalid selection is the primary threat to valid inference
6. Compared with experimental studies, observational studies provide less robust information
7. Make a sharp distinction between observational and experimental studies
8. Always look for a physical model underlying the data being analyzed. Assume that a statistical model, such as a linear model, is a good first start only
9. Keep models as simple as possible but no more simple
10. Be sure to understand the components and purpose of an omnibus quantity
11. Do not multiply probabilities more than necessary. Probabilities are bounded by 1; multiplication of enough probabilities will always lead to a small number
12. The use of one sided p-values is discouraged. Ordinarily, use 2-sided p-values
13. When designing experiments or observational studies, focus on p-values to calculate sample size; when representing results, focus on sample size
14. Use atleast 12 observations in constructing a confidence interval
15. For samples ≥ 20, a point estimate +/- 2 standard errors has a 95% coverage for a wide variety of distributions
16. Always know what the unit of a variable is
17. Do not let scale of measurement rigidly determine method of analysis
18. The practical applied statistician uses methods by all three schools (Neyman-Pearson, Likelihood, Bayesian) as appropriate

Kingston University London

Common Mistakes in Research

Fifteen common mistakes encountered in clinical research

Glenn T. Clark DDS, MS[a,1,*], Roseann Mulligan DDS, MS[b,1]

1. Failure to carefully examine the literature for similar, prior research . . . .
2. Failure to critically assess the prior literature . . . .
3. Failure to specify the inclusion and exclusion criteria for your subjects . . . .
4. Failure to determine and report the error of your measurement methods. . . . .
5. Failure to specify the exact statistical assumptions made in the analysis. . . . .
6. Failure to perform sample size analysis before the study begins . . . .
7. Failure to implement adequate bias control measures . . . .
8. Failure to write and stick to a detailed time line . . . .
9. Failure to vigorously recruit and retain subjects. . . . .
10. Failure to have a detailed, written and vetted protocol . . . .
11. Failure to examine for normality of the data . . . .
12. Failure to report missing data, dropped subjects and use of an intention to treat analysis . .
13. Failure to perform and report power calculations. . . . .
14. Failure to point out the weaknesses of your own study. . . . .
15. Failure to understand and use correct scientific language . . . .

Kingston University London

Statistical Errors

Insufficient statistical power

Misinterpretation of negative results (not considering insufficient statistical power)

Using the wrong statistical test (overlooking trend tests);
overlooking assumptions of a test

Treating replicates as independent observations

Assuming linearity when it does not exist

Too much emphasis on the *P* value (statistical significance) as opposed to
effect size and its confidence intervals (clinical/biological/practical significance)

Relying on the *P* value to validate the results, and mixing up statistical significance
with practical significance

Unnecessarily categorizing continuous data
Using wrong adjustments
Not considering interactions (effect modification)
& many more…

Statistical Errors

STATISTICAL ERRORS

P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume.

BY REGINA NUZZO

Kingston University London

Statistical Errors

eLIFE
elifesciences.org

FEATURE ARTICLE

SCIENCE FORUM

Ten common statistical mistakes to watch out for when writing or reviewing a manuscript

Abstract Inspired by broader efforts to make the conclusions of scientific research more robust, we have compiled a list of some of the most common statistical mistakes that appear in the scientific literature. The mistakes have their origins in ineffective experimental designs, inappropriate analyses and/or flawed reasoning. We provide advice on how authors, reviewers and readers can identify and resolve these mistakes and, we hope, avoid them in the future.

TAMAR R MAKIN* AND JEAN-JACQUES ORBAN DE XIVRY

PRiMER
peer-reviewed reports in medical education research
special article

A Very Short List of Common Pitfalls in Research Design, Data Analysis, and Reporting

Maarten van Smeden, PhD

PRiMER. 2022;6:26.

On the 12th Day of Christmas, a Statistician Sent to Me . . .

The BMJ's statistical editors relish a quiet Christmas, so make their wish come true and pay attention to the list of common statistical faux pas presented here by Riley and colleagues

Richard D Riley, [1] Tim J Cole, [2] Jon Deeks, [1] Jamie J Kirkham, [3] Julie Morris, [4] Rafael Perera, [5] Angie Wade, [6] Gary S Collins[7]

PM&R
WILEY

Statistically Speaking | Free Access

Avoiding Careless Errors: Know Your Data

Kristin L. Sainani PhD

Kingston University London

Statistical Errors

Fifteen common mistakes encountered in clinical research

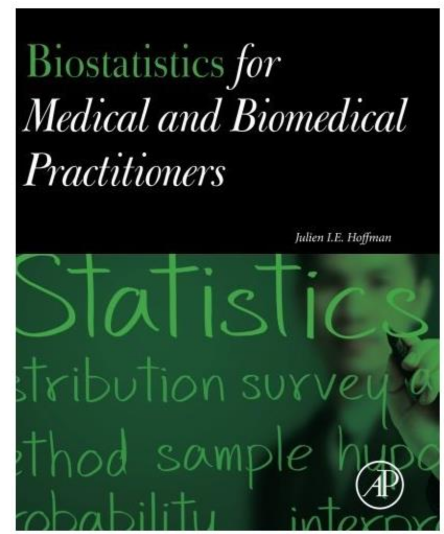
Glenn T. Clark DDS, MS[a,1,*], Roseann Mulligan DDS, MS[b,1]

1. Failure to carefully examine the literature for similar, prior research . . . .
2. Failure to critically assess the prior literature . . . .
3. Failure to specify the inclusion and exclusion criteria for your subjects . . . .
4. Failure to determine and report the error of your measurement methods. . . . .
5. Failure to specify the exact statistical assumptions made in the analysis. . . . .
6. Failure to perform sample size analysis before the study begins . . . .
7. Failure to implement adequate bias control measures . . . .
8. Failure to write and stick to a detailed time line . . . .
9. Failure to vigorously recruit and retain subjects. . . . .
10. Failure to have a detailed, written and vetted protocol . . . .
11. Failure to examine for normality of the data . . . .
12. Failure to report missing data, dropped subjects and use of an intention to treat analysis . .
13. Failure to perform and report power calculations. . . . .
14. Failure to point out the weaknesses of your own study. . . . .
15. Failure to understand and use correct scientific language . . . .

Statistical Errors

Table 2.2 List of some important statistical errors in biomedical publications

1. Failure to state clearly the hypothesis to be tested (Drummond et al., 2010; Harris et al., 2009; Ludbrook, 2008)
2. Failure to check the accuracy of data used for analysis
3. Failure to describe the statistical tests and software used (innumerable)
4. Failure to understand the prerequisites of statistical tests, leading frequently to serious misinterpretation of the results (Badgley, 1961; Glantz, 1980; Gore et al., 1977; Hayden, 1983; Pocock et al., 1987; Schoolman et al., 1968; Schor and Karten, 1966; Sheehan, 1980; Sheps and Schechter, 1984); failure to use control groups, or adequate control groups (Badgley, 1961; Ross, 1951; Schor and Karten, 1966); and failure to indicate whether the data are normally distributed or not, with consequent complications of analysis and interpretation (Gore et al., 1977; Kurichi and Sonnad, 2006)
5. Failure to assess effect size or to use a large enough sample size to give adequate power (Freiman et al., 1978; George, 1985; Hokanson et al., 1986; Huang et al., 2002; Kurichi and Sonnad, 2006; Murphy, 1979; Sackett, 1981b; Sheps and Schechter, 1984; Williams et al., 1997; Yates, 1983)
6. Confusion between standard deviation and standard error (Bunce et al., 1980; Gardner, 1975; Glantz, 1980; Oliver and Hall, 1989; Reed et al., 2003; Weiss and Bunce, 1980) and absence or misuse of confidence limits (Harris et al., 2009; Hayden, 1983; Hokanson et al., 1986; Huang et al., 2002)
7. Use of multiple *t*-tests without appropriate correction or failure to use techniques such as analysis of variance designed for comparisons of more than two groups (Glantz, 1980; Kurichi and Sonnad, 2006; Kusuoka and Hoffman, 2002; Pocock et al., 1987; Schor and Karten, 1966; Williams et al., 1997)
8. Incorrect use or definition of sensitivity and specificity (Schor and Karten, 1966; Sheps and Schechter, 1984) and failure to understand when the odds ratio is an unreliable guide to relative risk (Feinstein, 1986; Holcomb et al., 2001; Katz, 2006; Schwartz et al., 1999)
9. Failure to understand how P values should be interpreted (Dar et al., 1994; Oliver and Hall, 1989)
10. A number of the above errors are common in clinical trials, which may also show failure of or inadequate randomization, failure to describe how patients are included in the trial, failure to use double blind procedures, failure to define when a trial should be stopped early (Harris et al., 2009; Hayden, 1983; Hokanson et al., 1986; Huang et al., 2002)



Kingston University London

Statistical Errors

1. **Write a Protocol for Checking Data as it is Collected.** For example, in a pain study, we had research assistants examine each questionnaire as it was received to check that every question on the questionnaire had been answered, that all pain scores were between 0 and 10, and that the name of each analgesic in the medication diary was legible. Patients were contacted in cases of missing, illogical, or illegible data.

2. **Use "Sign-Offs."** We asked research assistants to sign and date questionnaires that they had checked before filing them. The signature serves the same function as the courtroom oath to tell the whole truth: "I have checked this questionnaire and found it to be complete and correct." It also means that, if a mistake has been made, we can work out who made it, a boon to quality control and a powerful incentive to researcher assistants to be as careful as possible.

1. **The Best way of Preventing Data Entry Errors is to Avoid Data Entry Altogether.** Take Exhibit 4: patients' heights and weights could have been downloaded into a spreadsheet directly from the surgery database and body mass index calculated using a formula. Where information is obtained from patients, you can use forms that can be optically scanned, or Web-based systems. In one study I am planning, we will email patients every 3 months with a link to a secure Web site on which they directly enter their symptom scores.

2. **Use Double-Data Entry.** Data entry from paper forms such as questionnaires is often unavoidable. The best system to avoid data-entry errors is what is known as "double-data entry." In brief, data from paper forms is typed onto a database; a blank copy of the database is then made and the data reentered; the 2 databases are then merged to discover inconsistencies. The last time I did this, there was at least 1 inconsistency on 14 of the 99 patient records in the data set.

3. **Write a Protocol for Data Entry.** This specifies rules for data entry such as how to handle illegible or ambiguous data. As a typical example, we specify that if a patient circles 2 responses to a question, we take the response corresponding to the higher symptom score.

4. **Use Sign-Offs.** Data-entry personnel should sign and date paper forms after they have been entered onto the computer (see point 2 under data collection).

1. **Create a Log File.** Record with dates all the analyses you do, along with their rationale. The log file should also document the names of files and folders you set up to manage your data.

2. **Check the Final Data Set.** Once you have the data set in your statistical software, you should check for missing data. You should also conduct consistency and range checks. A *consistency* check determines whether the value of one variable is unlikely or impossible given the value of a different variable: the biomarkers case in Exhibit 3 is one example; others include checking whether a patient's date of recurrence is before their date of surgery or after their date of death, or whether the level of a variable that is part of the total (eg, days with severe pain) is higher than the variable representing the total (eg, days with any pain). A *range* check determines whether the values of any variable are *prima facie* unlikely, for example, a body mass index of 2000 or a hemoglobin of 1.25.

3. **Program your Analyses.** An introduction to statistical programming is beyond the remit of this article; however, the key point is that analyses should be conducted by writing a program, which can be reproduced, rather than by using pull-down menus, which are not reproducible. The program should also include automatic output suitable for importing into a word processing program: cutting and pasting individual numbers from software output is an important source of error.

1. **Check Every Number on the Manuscript Against the Printout from the Statistics Software.** This offers an additional way of ensuring that the paper says what it is meant to say.

2. **Double-Check the Proof.** Errors often creep in when papers are reformatted by editorial staff.

Medscape Business of Medicine > Stats for the Health Professional

Look at Your Garbage Bin: It May Be the Only Thing You Need to Know About Statistics

Andrew J. Vickers, PhD

Disclosures | November 03, 2006

Statistical Errors

Statistical Thinking Frank Harrell

Contents

- Misinterpretation of P-values and of Main Study Results
- Dichotomania
- Change from Baseline
- Improper Subgrouping
- Serial Data and Response Trajectories
- Cluster Analysis
- Further Reading
- Discussion Archive (2017-2021)

Statistical Errors in the Medical Literature

</> Code

PREDICTION | LOGIC | P-VALUE | VALIDATION | BAYES | EVIDENCE | SUBGROUP | DICHOTOMIZATION | MEDICINE | INFERENCE | CHANGE-SCORES | RCT | PERSONALIZED-MEDICINE | RESPONDER-ANALYSIS | HYPOTHESIS-TESTING | MEDICAL-LITERATURE | 2017

AUTHOR
Frank Harrell

Vanderbilt University
School of Medicine
Department of Biostatistics

PUBLISHED
April 8, 2017

1. Misinterpretation of P-values and Main Study Results
2. Dichotomania
3. Problems With Change Scores
4. Improper Subgrouping
5. Serial Data and Response Trajectories
6. Cluster Analysis

thebmjopinion | Latest | Authors | Topics

Time to assume that health research is fraudulent until proven otherwise?

July 5, 2021

Kingston University London

Statistical Errors

COMMON ~~MISTEAKS~~ MISTAKES IN USING STATISTICS:

Spotting and Avoiding Them

Types of mistakes

Many mistakes in using statistics fall into one of the following categories:

- Expecting too much certainty
- Misunderstandings about probability
- Mistakes in thinking about causation
- Problematical choice of measure
- Errors in sampling
- Over-interpretation
- Mistakes involving limitations of frequentist inference techniques (hypothesis tests and confidence intervals)
- Using an inappropriate method of analysis
- Inadequate attention to communication

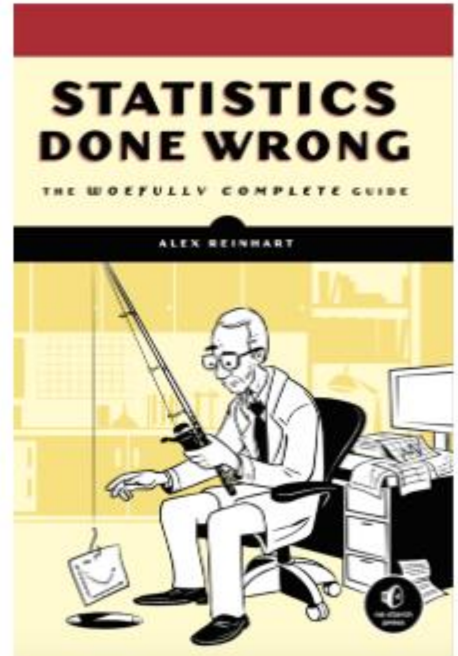
Suggestions for reducing the incidence of mistakes in using statistics

- Suggestions for teachers of statistics
- Suggestions for consumers of research
- Suggestions for researchers
- Suggestions for referees of research articles and editors of journals

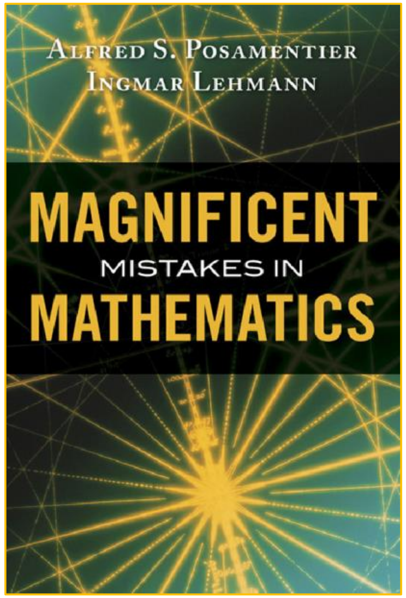
Glossary

Resources

Statistical Errors



Statistical Errors



CHAPTER 5

MISTAKES IN PROBABILITY AND STATISTICS

Including the birthday problem and Monty Hall problem

Kingston University London

Statistical Errors

Matt Parker

HUMBLE Pi

A Comedy of Maths Errors

ELEVEN

Stats The Way I Like It

Including the problem with averages, bias and causality

Assessment of Statistical Literacy

International Statistical Review (2002), **70**, 1, 1–51, Printed in The Netherlands

Adults' Statistical Literacy: Meanings, Components, Responsibilities

Iddo Gal

University of Haifa, Israel

Table 3

Sample "worry questions" about statistical messages.

1. Where did the data (on which this statement is based) come from? What kind of study was it? Is this kind of study reasonable in this context?
2. Was a sample used? How was it sampled? How many people did actually participate? Is the sample large enough? Did the sample include people/units which are representative of the population? Is the sample biased in some way? Overall, could this sample reasonably lead to valid inferences about the target population?
3. How reliable or accurate were the instruments or measures (tests, questionnaires, interviews) used to generate the reported data?
4. What is the shape of the underlying distribution of raw data (on which this summary statistic is based)? Does it matter how it is shaped?
5. Are the reported statistics appropriate for this kind of data, e.g., was an average used to summarize ordinal data; is a mode a reasonable summary? Could outliers cause a summary statistic to misrepresent the true picture?
6. Is a given graph drawn appropriately, or does it distort trends in the data?
7. How was this probabilistic statement derived? Are there enough credible data to justify the estimate of likelihood given?
8. Overall, are the claims made here sensible and supported by the data? e.g., is correlation confused with causation, or a small difference made to loom large?
9. Should additional information or procedures be made available to enable me to evaluate the sensibility of these arguments? Is something missing? e.g., did the writer "conveniently forget" to specify the base of a reported percent-of-change, or the actual sample size?
10. Are there alternative interpretations for the meaning of the findings or different explanations for what caused them, e.g., an intervening or a moderator variable affected the results? Are there additional or different implications that are not mentioned?

Kingston University London

Assessment of Statistical Literacy

Table 1 Quick Risk Test

Question	Possible answers
1. A test's sensitivity is a central criterion for its quality as a diagnostic tool. The sensitivity is	A) the proportion of people with a positive test result among those who are sick. *** B) the proportion of people with a negative test result among those who are sick. C) the proportion of people with a positive test result among those who are healthy. D) the proportion of people with a negative test result among those who are healthy.
2. A test's specificity is a central criterion for its quality as a diagnostic tool. The specificity is	A) the proportion of people with a positive test result among those who are sick. B) the proportion of people with a negative test result among those who are sick. C) the proportion of people with a positive test result among those who are healthy. D) the proportion of people with a negative test result among those who are healthy. ***
3. Which test characteristic quantifies the probability that a person with a positive test result actually has the disease?	A) Positive predictive value *** B) Negative predictive value C) Specificity D) Sensitivity
4. Which test characteristic quantifies the probability that a person with a negative test result does not have the disease?	A) Sensitivity B) Positive predictive value C) Negative predictive value *** D) Sensitivity
5. A medical test's manufacturer tells you the sensitivity and the specificity of its test. You would like to tell your patient the probability that they are sick if they have a positive test result. Which measurement do you need for your calculation?	A) Mortality B) Prevalence *** C) Coherence D) Latency

Open access Research

BMJ Open

Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany

Mirjam Annina Jenny,[1] Niklas Keller,[2] Gerd Gigerenzer[1]

Kingston University London

Assessment of Statistical Literacy

6. Mammography is often used as a screening-test to detect breast cancer early. The probability that a woman has breast cancer is 1%. When a woman has breast cancer her probability of receiving a positive mammogram is 90%. When a woman does not have breast cancer her probability of nevertheless receiving a positive mammogram is 9%. What is the best estimate for the number of women with a positive screening mammogram who actually have breast cancer?	A) 9 in 10 B) 8 in 10 C) 1 in 10 *** D) 1 in 100
7. In a medical publication you read that screening with mammography lowers the probability of dying from breast cancer by 20%. This number is	A) a relative risk reduction. *** B) an absolute risk reduction. C) a specific risk reduction. D) an evident risk reduction.
8. A patient asks you about the benefits of cancer screening. Which criterion should you consider here?	A) 5-year survival rate B) Incidence C) Mortality rate *** D) Prevalence
9. Imagine two groups of people who all die of cancer at age 70. In group A, cancer is detected via screening at the age of 60. In this group, the 5-year survival rate is 100%. Group B is not screened. In this group, cancer is detected at age 68. Everyone dies at age 70. Thus, the 5-year survival rate is 0%. Which bias explains why both groups have different 5-year survival rates?	A) Selection bias B) Overdiagnosis bias C) Lead-time bias *** D) Performance bias
10. A higher screening rate results in more positive diagnoses. In screening, if anomalies are discovered, which because of their extremely slow growth would never cause symptoms or an early death, this is called	A) selection bias. B) attrition bias. C) lead-time bias. D) overdiagnosis bias. ***

Questions and multiple-choice answers of the 10-item Quick Risk Test (*** denotes the correct answer).

Open access Research

BMJ Open

Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany

Mirjam Annina Jenny,[1] Niklas Keller,[2] Gerd Gigerenzer[1]

Statisticians Know How to Avoid Common Pitfalls

Using statistical analyses to produce findings for a study is the culmination of a long process. This process includes constructing the study design, selecting and measuring the variables, devising the sampling technique and sample size, cleaning the data, and determining the analysis methodology among numerous other issues. The overall quality of the results depends on the entire chain of events. A single weak link might produce unreliable results. The following list provides a small taste of potential problems and analytical errors that can affect a study.

The Importance of Statistics

By Jim Frost — 47 Comments

Statistics By Jim

Making statistics intuitive

Accuracy and Precision: Before collecting data, you must ascertain the accuracy and precision of your measurement system. After all, if you can't trust your data, you can't trust the results!

Biased samples: An incorrectly drawn sample can bias the conclusions from the start. For example, if a study uses human subjects, the subjects might be different than non-subjects in a way that affects the results. See: Populations, Parameters, and Samples in Inferential Statistics.

Overgeneralization: Findings from one population might not apply to another population. Unfortunately, it's not necessarily clear what differentiates one population from another. Statistical inferences are always limited, and you must understand the limitations.

Causality: How do you determine when X causes a change in Y? Statisticians need tight standards to assume causality whereas others accept causal relationships more easily. When A precedes B, and A is correlated with B, many mistakenly believe it is a causal connection! However, you'll need to use an experimental design that includes random assignment to assume confidently that the results represent causality. Learn how to determine whether you're observing causation or correlation!

Incorrect analysis: Are you analyzing a multivariate study area with only one variable? Or, using an inadequate set of variables? Perhaps you're assessing the mean when the median might be a better? Or, did you fit a linear relationship to data that are nonlinear? You can use a wide range of analytical tools, but not all of them are correct for a specific situation.

Violating the assumptions for an analysis: Most statistical analyses have assumptions. These assumptions often involve properties of the sample, variables, data, and the model. Adding to the complexity, you can waive some assumptions under specific conditions—sometimes thanks to the central limit theorem. When you violate an important assumption, you risk producing misleading results.

Data mining: Even when analysts do everything else correctly, they can produce falsely significant results by investigating a dataset for too long. When analysts conduct many tests, some will be statistically significant due to chance patterns in the data. Fastidious statisticians track the number of tests performed during a study and place the results in the proper context.

Numerous considerations must be correct to produce trustworthy conclusions. Unfortunately, there are many ways to mess up analyses and produce misleading results. Statisticians can guide others through this swamp!

Kingston University London

Case Study

5

The lady who tasted tea—a bit of statistical history

In the beginning, on a summer afternoon in Cambridge, England (circa 1920), an extraordinary meeting took place between two totally different people: a British lady who bravely claimed she could tell whether tea was prepared by adding milk to tea or tea to milk and Ronald Aylmer Fisher who was destined to become one of the world's most renowned statisticians.

The lady's claim was met with skepticism. Clearly tasting one cup of tea would not provide satisfactory evidence that the "tea-tasting lady" had an unusual ability. In the complete absence of an ability to determine two different methods of preparation her answer would be correct half the time. Fisher proposed an experiment.

History does not record the details but it was decided eight cups of tea would be prepared, four with tea added to milk and four with milk added to tea, producing perhaps the most famous 2 × 2 table in the history of statistics:

Table 5.1 Tea-tasting lady

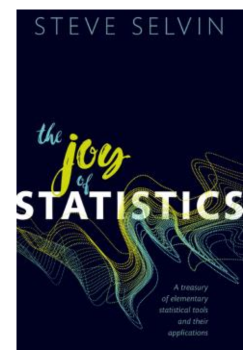
	"tea to milk"	"milk to tea"	total
tea added to milk	4	0	4
milk added to tea	0	4	4

The results from Fisher's experiment produced substantial but intuitive evidence that at least one person could tell the difference between these two tea preparations. This Sunday afternoon experiment remarkably crystallized two concepts fundamental to modern statistics.

First, the probability of correctly identifying all eight cups of tea by chance alone (guessing) was recognized to be a useful number. It is:

$$P(\textit{all eight correct} \mid \textit{guessing}) = \frac{4}{8} \times \frac{3}{7} \times \frac{2}{6} \times \frac{1}{5} = \frac{24}{1680} = 0.014.$$

The Joy of Statistics: A Treasury of Elementary Statistical Tools and their Applications. Steve Selvin. © Steve Selvin 2019. Published in 2019 by Oxford University Press. DOI: 10.1093/oso/9780198833444.001.0001



Case Study

Antioxidant Megavitamins for Brain Health: Puffery vs. Fact

Reynold Spector

From: Volume 43, No. 2
March / April 2019

Skeptical Inquirer
THE MAGAZINE FOR SCIENCE AND REASON

Reasons for Popularity of Worthless Therapies, including Megavitamins/Multivitamins

1. Placebo effect
2. Erroneous studies (see text)
3. Incorrect authoritarian advice (see text)
4. Dissatisfaction with conventional therapy
5. Habit (once started, it is hard to stop)
6. "More the merrier"
7. Prevention of deficiency
8. Demand to "do something"
9. Advertising, often misleading or false
10. Corruption

Kingston University London

Statistical Literacy
Part III

Assessment of Diagnostic Tests
(Sensitivity/Specificity, Positive/Negative Predictive Values,
Accuracy/Precision/Validity/Reliability)
Critical Appraisal of Research Papers

Assessing a Diagnostic Test

You have a new diagnostic or predictive test. Congratulations!

What proportions of truly diseased individuals have a positive result by this test? How much is good enough?

What proportion of truly non-diseased individuals will have a negative result?

How many times out of 100, does the test show a true positive or negative results?

If the test is positive, what is the probability that the individual really has (or will have) the disease?

If the test is negative, what is the probability that the individual do not have (or will not have) the disease?

How will you know it is better or worse than an existing (gold standard) test?

Assessing a Diagnostic Test

You have a new diagnostic or predictive test. Congratulations!

What proportions of truly diseased individuals are identified by this test? How much is good enough? **Sensitivity [and false-negatives]**

What proportion of truly non-diseased individuals will get a negative result?

Specificity [and false-positives]

How many times out of 100, does the test show a true result (true positives/negatives)?

Accuracy [proportion of true positives + negatives]

If the test is positive, what is the probability that the individual really has (or will have) the disease?

Positive predictive value [probability of a positive test being a true positive]

If the test is negative, what is the probability that the individual do not have (or will not have) the disease?

Negative predictive value [probability of a negative test being a true negative]

How will you know it is better or worse than an existing (gold standard) test?

Area under the curve of receiver operating characteristics analysis (AUC-ROC)

Kingston University London

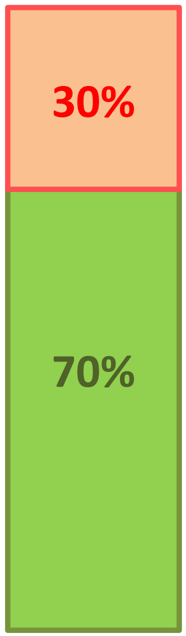
Sensitivity & Specificity

Disease positive

Disease negative

	Disease (+)	Disease (-)
Test (+)	70	30
Test (-)	30	70

False negatives = 0.30

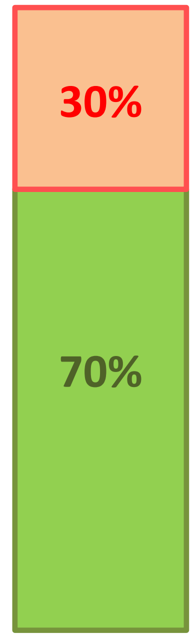


True positives = 0.70

Sensitivity = 0.70

	Disease (+)	Disease (-)
Test (+)	70 True Positive	30 False Positive
Test (-)	30 False Negative	70 True Negative

False positives = 0.70



True negatives = 0.70

Specificity = 0.70

Kingston University London

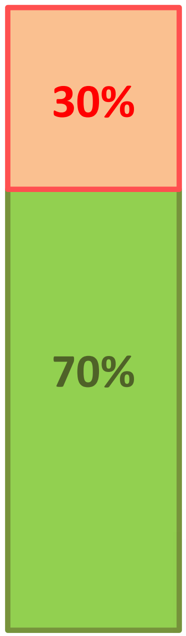
Sensitivity & Specificity

Disease positive

Disease negative

	Disease (+)	Disease (-)
Test (+)	70	30
Test (-)	30	70

False negatives = 0.30



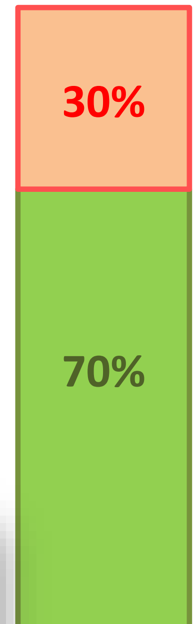
True positives = 0.70

Sensitivity = 0.70

	Disease (+)	Disease (-)
Test (+)	70 True Positive	30 False Positive
Test (-)	30 False Negative	70 True Negative

	Reference standard	Reference standard	
Index test	True positives (TP)	False positives (FP)	Positive predictive value TP/(TP+FP)
Index test	False negatives (FN)	True negatives (TN)	Negative predictive value TN/(FN+TN)
	Sensitivity TP/(TP+FN)	Specificity TN/(FP+TN)	

False positives = 0.70



True negatives = 0.70

Specificity = 0.70

Kingston University London

Sensitivity & Specificity

1. What is Sensitivity (True Positive Rate)?
2. What is Specificity (True Negative Rate)?
3. Positive Predicted Values
4. Negative Predicted Values

What is a Sensitive Test?

The sensitivity of a test (also called the true positive rate) is defined as the proportion of people **with** the disease who will have a **positive** result. In other words, a highly sensitive test is one that **correctly identifies patients with a disease**. A test that is 100% sensitive will identify *all* patients who have the disease. It's extremely rare that any clinical test is 100% sensitive. A test with 90% sensitivity will identify 90% of patients who have the disease, but will miss 10% of patients who have the disease.

A highly sensitive test can be useful for **ruling out a disease** if a person has a negative result. For example, a negative result on a pap smear probably means the person does not have cervical cancer. The acronym widely used is SnNout (high **Sn**ensitivity, **N**egative result = rule **out**).
Back to Top

What is a Specific Test?

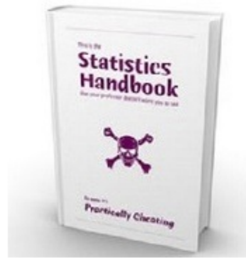
The specificity of a test (also called the True Negative Rate) is the proportion of people **without** the disease who will have a **negative** result. In other words, the **specificity** of a test refers to how well a test identifies patients who do not have a disease. A test that has 100% specificity will identify 100% of patients who do not have the disease. A test that is 90% specific will identify 90% of patients who do not have the disease.

Tests with a high specificity (a high true negative rate) are most useful when the result is positive. A highly specific test can be useful for ruling in patients who have a certain disease. The acronym is SPin (high **Sp**ecificity, rule **in**).

What is a "High" Range?

What qualifies as "high" sensitivity or specificity varies by the test. For example the cut-offs for Deep Vein Thrombosis and Pulmonary Embolism tests range from 200-500 ng/dL (Pregerson, 2016).
Back to Top

High sensitivity/low specificity example



Check the video @
https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/sensitivity-vs-specificity-statistics

Sensitivity & Specificity

STUB – Statistical Thinking in Undergraduate Biology

HOME ABOUT US JOIN US RESOURCES WORKSHOPS

Search

HOME / CONTENT • UNCATEGORIZED / SENSITIVITY AND SPECIFICITY: U ...

SENSITIVITY AND SPECIFICITY: UNDERSTANDING SHORTCOMINGS OF ANTIBODY TESTS WITH THE COVID-19 VIRUS

SENSITIVITY AND SPECIFICITY: UNDERSTANDING SHORTCOMINGS OF ANTIBODY TESTS WITH THE COVID-19 VIRUS

Date: June 29, 2020 Posted by: Jill VanderStoep Content, Uncategorized

Sensitivity and specificity are commonly taught concepts in biostatistics courses. The article linked here gives a nice visual explanation of false positives and false negatives with tests that do have high sensitivity and specificity, but a low infection rate of the disease in the population.

RECENT POSTS

- Resampling-based methods for biologists
- Teaching quantitative ecology online: An evidence-based prescription of best practices

Sensitivity & Specificity

Table 1. Diagnostic test accuracy summary statistics [2]

Summary statistics	Equation	Definition
Sn	TP/(TP+FN)	Proportion of persons who have positive test results to those with disease
Sp	TN/(FP+TN)	Proportion of persons who have negative test result to those without disease
PPV	TP/(TP+FP)	Proportion of persons with disease to those who have positive test result
NPV	TN/(FN+TN)	Proportion of persons without disease to those who have negative test result
LR+	Sn/(1-Sp)	Ratio of the probability of a positive test result among those with disease to that of a positive test result among those without disease
LR-	(1-Sn)/Sp	Ratio of the probability of a negative test result among those with disease to that of a negative test result among those without disease
Accuracy of index test	(TP+TN)/(TP+FP+FN+TN)	The proportion of persons who are true positive and persons who are true negative among all subjects
DOR	(TP*TN)/(FP*FN)	The ratio of the OR for a positive test result among persons with disease to that among persons without disease

Sn, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; LR+, positive likelihood ratio; LR-, negative likelihood ratio; DOR, diagnostic odds ratio; TP, true positive; FP, false positive; FN, false negative; TN, true negative; OR, odds ratio.

Kingston University London

Diagnostic test accuracy: application and practice using R software

Sung Ryul Shim[1,2], Seong-Jang Kim[3,4], Jonghoo Lee[5]

Positive/Negative Predictive Value

Positive Predictive Value (PPV) is the ratio of patients truly diagnosed as positive to all those who had positive test results (including healthy subjects who were incorrectly diagnosed as patient, i.e., false-positives). This characteristic can predict how likely it is for someone to truly be patient, in case of a positive test result.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4608333

Positive/Negative Predictive Value

Prevalence Impact on Positive Predictive Value (PPV) and Negative Predictive Value (NPV)

Prevalence thus impacts the positive predictive value (PPV) and negative predictive value (NPV) of tests. As the prevalence increases, the PPV also increases but the NPV decreases. Similarly, as the prevalence decreases the PPV decreases while the NPV increases.

For a mathematical explanation of this phenomenon, we can calculate the positive predictive value (PPV) as follows:

- PPV = (sensitivity x prevalence) / [(sensitivity x prevalence) + ((1 – specificity) x (1 – prevalence))]

If we hold all values except for the prevalence the same then as prevalence increases the numerator will also increase for PPV. In the denominator note the last term of "1 – prevalence." Thus as prevalence increases towards 100% (a value of one) the term "1 – prevalence" goes towards zero. This drives the second part of the denominator, "(1 – specificity) x (1 – prevalence)", to smaller and smaller values as prevalence increases. Thus at a very high prevalence the value of "1 – prevalence" goes towards zero and the PPV equation reduces to:

- PPV = (sensitivity x prevalence) / [(sensitivity x prevalence) + ((1 – specificity) x (0))] =
- PPV = (sensitivity x prevalence) / [(sensitivity x prevalence) + (0)] =
- PPV = (sensitivity x prevalence) / (sensitivity x prevalence) = 1

Similarly we can write the negative predictive value (NPV) as follows:

- NPV = (specificity x (1 – prevalence)) / [(specificity x (1 – prevalence)) + ((1 – sensitivity) x prevalence)]

For the NPV as the prevalence increases (goes towards one) the term "1 – prevalence" becomes smaller making the numerator smaller. In the denominator NPV has the same first term as the numerator, "specificity x (1 – prevalence)" which will also become smaller as the prevalence increases. The second term in the denominator, "(1 – sensitivity) x prevalence" will increase as the prevalence increases. As the prevalence comes very close to 100% we can write NPV as:

- NPV = (specificity x (1 – 1)) / [(specificity x (1 – 1)) + ((1 – sensitivity) x 1)] =
- (specificity x 0) / [(specificity x 0) + (1 – sensitivity)] =
- 0 / (0 + (1 – sensitivity)) = 0

STATPEARLS
CONTENT IS KING

Kingston University London

Accuracy & Precision

Statistics By Jim

Making statistics intuitive

Graphs | Basics | Hypothesis Testing | Regression | ANOVA | Probability | Time Series | Fun

Accuracy vs Precision: Differences & Examples

By Jim Frost — 8 Comments

Reliability & Validity

Statistics By Jim

Making statistics intuitive

Graphs | Basics | Hypothesis Testing | Regression | ANOVA | Probability | Time Series | Fun

Reliability vs Validity: Differences & Examples

By Jim Frost — Leave a Comment

AN IDEAL NEW ASSAY:

Sensitive

Can detect a lower amount of the analyte than the established method (100% true positivity)

Specific

Only detects what it is supposed to detect (100% true negativity)

Accurate

Results are close to the true/reference values (no deviation/bias, which can be systematic or non-systematic)

Precise

Repeat measurements yield similar results (not necessarily accurate results though)

Reliable

Always yields similar results with accuracy

Critical Appraisal of Research Papers

Critical Appraisal of Scientific Articles

Part 1 of a Series on Evaluation of Scientific Publications

Jean-Baptist du Prel, Bernd Röhrig, Maria Blettner

SUMMARY

Introduction: In the era of evidence-based medicine, one of the most important skills a physician needs is the ability to analyze scientific literature critically. This is necessary to keep medical knowledge up to date and to ensure optimal patient care. The aim of this paper is to present an accessible introduction into critical appraisal of scientific articles.

Methods: Using a selection of international literature, the reader is introduced to the principles of critical reading of scientific articles in medicine. For the sake of conciseness, detailed description of statistical methods is omitted.

Results: Widely accepted principles for critically appraising scientific articles are outlined. Basic knowledge of study design, structuring of an article, the role of different sections, of statistical presentations as well as sources of error and limitation are presented. The reader does not require extensive methodological knowledge. As far as necessary for critical appraisal of scientific articles, differences in research areas like epidemiology, clinical, and basic research are outlined. Further useful references are presented.

Conclusion: Basic methodological knowledge is required to select and interpret scientific articles correctly.

Dtsch Arztebl Int 2009; 106(7): 100–5
DOI: 10.3238/arztebl.2009.0100

DEUTSCHES ÄRZTEBLATT INTERNATIONAL

Kingston University London

Critical Appraisal of Research Papers

Critical Appraisal of Scientific Articles

Part 1 of a Series on Evaluation of Scientific Publications

Jean-Baptist du Prel, Bernd Röhrig, Maria Blettner

Dtsch Arztebl Int 2009; 106(7): 100–5
DOI: 10.3238/arztebl.2009.0100

BOX 2

Critical questions

- Does the study pose scientifically interesting questions?
- Are statements and numerical data supported by literature citations?
- Is the topic of the study medically relevant?
- Is the study innovative?
- Does the study investigate the predefined study goals?
- Is the study design apt to address the aims and/or hypotheses?
- Did practical difficulties (e.g. in recruitment or loss to follow-up) lead to major compromises in study implementation compared with the study protocol?
- Was the number of missing values too large to permit meaningful analysis?
- Was the number of cases too small and thus the statistical power of the study too low?
- Was the course of the study poorly or inadequately monitored (missing values, confounding, time infringements)?
- Do the data support the authors' conclusions?
- Do the authors and/or the sponsor of the study have irreconcilable financial or ideological conflicts of interest?

Kingston University London

Critical Appraisal of Research Papers

Critical Appraisal of Scientific Articles

Part 1 of a Series on Evaluation of Scientific Publications

Jean-Baptist du Prel, Bernd Röhrig, Maria Blettner

Dtsch Arztebl Int 2009; 106(7): 100–5
DOI: 10.3238/arztebl.2009.0100

TABLE 2

Checklist to evaluate the quality of scientific publications

	Yes	Unclear	No
Design			
Is the aim of the study clearly described?	☐	☐	☐
Are the study population(s) and the inclusion and exclusion criteria described in detail?	☐	☐	☐
Were the patients allocated randomly to the different arms of the study?	☐	☐	☐
If yes: Is the method of randomization described?	☐	☐	☐
a) Is the number of cases discussed?	☐	☐	☐
b) Were sufficient cases enrolled (e.g. Power ≥50%)?	☐	☐	☐
Are the methods of measurement (e.g. laboratory examination, questionnaire, diagnostic test) suitable for determination of the target variable (with regard to scale, time of investigation, standardization)?	☐	☐	☐
Is there information regarding data loss (response rates, loss to follow-up, missing values)?	☐	☐	☐
Study inception and implementation			
Are treatment and control groups matched with regard to major relevant characteristics (age, sex, smoking habits etc.)?	☐	☐	☐
Are the drop-outs analyzed for differences between the treatment and control groups?	☐	☐	☐
How many cases were observed over the whole study period?	☐	☐	☐
Are side effects and adverse events during the study period described?	☐	☐	☐
Analysis and evaluation			
Have the correct statistical parameters and methods been selected, and are they clearly described?	☐	☐	☐
Are the statistical analyses clearly described?	☐	☐	☐
Are the important parameters (prognostic factors) included in the analysis or at least discussed?	☐	☐	☐
Is the presentation of the statistical parameters appropriate, comprehensive, and clear?	☐	☐	☐
Are the effect sizes and confidence intervals stated for the principal findings?	☐	☐	☐
Is it apparent why the given study design/statistical methods were chosen?	☐	☐	☐
Are all conclusions supported by the study's findings?	☐	☐	☐

By using a checklist such as this, the statistical and methodological soundness of a study can be assessed and improvements considered.
Not all of the points in this checklist can be used to evaluate all study types; for example, randomization is particularly applicable to clinical studies.

Critical Appraisal of Research Papers

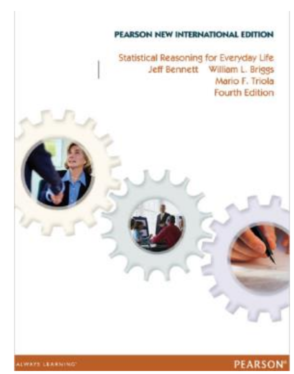
SHOULD YOU BELIEVE A STATISTICAL STUDY?

Already you know enough to achieve one of the major goals of this text: being able to answer the question "Should you believe a statistical study?"

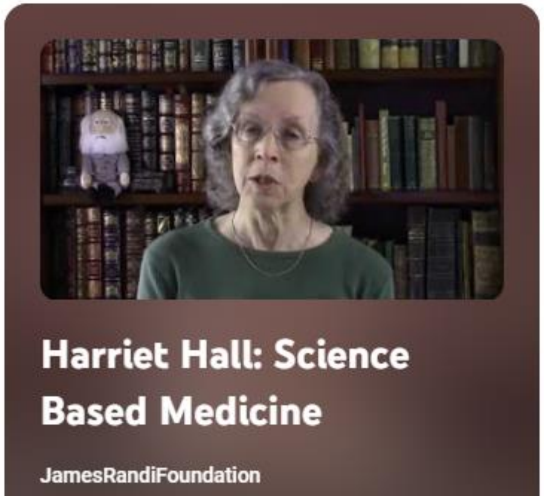
Most researchers conduct their statistical studies with honesty and integrity, and most statistical research is carried out with diligence and care. Nevertheless, statistical research is sufficiently complex that bias can arise in many different ways, making it very important that we always examine reports of statistical research carefully. There is no definitive way to answer the question "Should I believe a statistical study?" However, in this section we'll look at eight guidelines that can be helpful. Along the way, we'll also introduce a few more definitions and concepts that will prepare you for discussions to come later.

Eight Guidelines for Critically Evaluating a Statistical Study

1. *Get a Big Picture View of the Study.* For example, you should understand the goal of the study, the population that was under study, and whether the study was observational or an experiment.
2. *Consider the Source.* In particular, look for any potential biases on the part of the researchers.
3. *Look for Bias in the Sample.* That is, decide whether the sampling method was likely to produce a representative sample.
4. *Look for Problems Defining or Measuring the Variables of Interest.* Ambiguity in the variables can make it difficult to interpret reported results.
5. *Beware of Confounding Variables.* If the study neglected potential confounding variables, its results may not be valid.
6. *Consider the Setting and Wording in Surveys.* In particular, look for anything that might tend to produce inaccurate or dishonest responses.
7. *Check That Results Are Presented Fairly.* For example, check whether the study really supports the conclusions that are presented in the media.
8. *Stand Back and Consider the Conclusions.* For example, evaluate whether study achieved its goals. If so, do the conclusions make sense and have practical significance?



Critical Appraisal of Research Papers

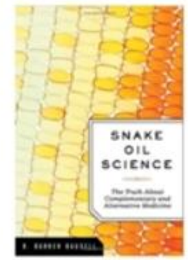


9 

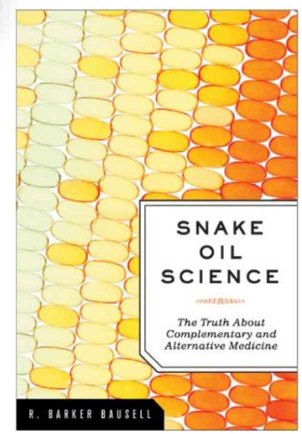
Lecture 9: Pitfalls in Research

JamesRandiFoundation • 14K views • 7 years ago

Bausell's Quick Checklist



1. Is it randomized with a credible control group?
2. Are there at least 50 subjects per group?
3. Is the dropout rate 25% or less?
4. Was it published in a high-quality, prestigious, peer-reviewed journal?



Kingston University London

Critical Appraisal of Research Papers

INTERDISCIPLINARY IDEAS

Why Should We All Embrace Statistical Thinking?

[Data Literacy 101]

BY KRISTIN HUNTER-THOMSON

Way #1: We need to consider what the sample is every time we look at data

How was sampling done? Is the sample random and representative of the population? Is the sample size large enough? Was a sample size calculation performed for sufficient statistical power? Are baseline characteristics comparable in comparison groups?

Way #2: We need to talk about uncertainty

Are 95% confidence intervals provided? Is the sample size as large as possible?

Way #3: We support with evidence, not prove the hypothesis

Is the paper claiming to have proven anything. The results should be used to confirm or refute the (alternative) hypothesis (if there was one to begin with).

Way #4: We can only make claims from the data we have, not what we want to have

Are the results supported by valid data? Is it a randomized, blinded study (or observational)? What is the response/drop-out rate? Are there extrapolations? Is causality implied/assumed or assessed?

Way #5: We need to think about whether a finding is truly meaningful

Is the absolute risk change given (or just relative changes)? Is the effect size given? Is clinical/biological significance discussed? Are bias and confounding minimized and assessed? Is it a peer-reviewed study and published in a reputable journal?

Kingston University London

Critical Appraisal of Research Papers

Overview of Critical Reading

How to Critcally Review an Article

Kingston University London

BU Boston University School of Public Health

Statistical Literacy

Can Case Studies Be Used to Teach Critical Thinking?

Clyde Freeman Herreid

If I had to choose one general characteristic that cuts across smart people it would be scepticism - the ability to ask oneself and others if the conclusions and data are correct. Smart people silently or openly say, "What is the evidence for this or that idea? Why should I believe this? Are there other explanations for the data? Is there another way to explain the data? What do you mean when you say this?" If you routinely ask such questions, even when dealing with subjects out of your own area of expertise, you will be well off. Certainly, this is true in the political arena. We have just had a terrible brouhaha -fiasco, is more like it - over the war in Iraq.

This brings me to case studies. If reading, arguing, and challenging are hallmarks of critical thinking, then case studies are the poster children for the process. Most of them are discipline specific. rtainly. But

Best-Case Scenario

The best case technique that I know is one called the "Interrupted Case Method." Readers can see a version of it on the National Center for Case Study Teaching in Science website, titled "Mom Always Liked You Best." The method begins when the teache gives students (ideally

see model behavior from the experts. I love this method because it is the way real science works—we have to work with incomplete data, make tentative hypotheses, collect more information, refine our hypotheses, make more predictions, get more data, and so on. In fact, this interrupted method is very one that I

Quiz: Critical Thinking

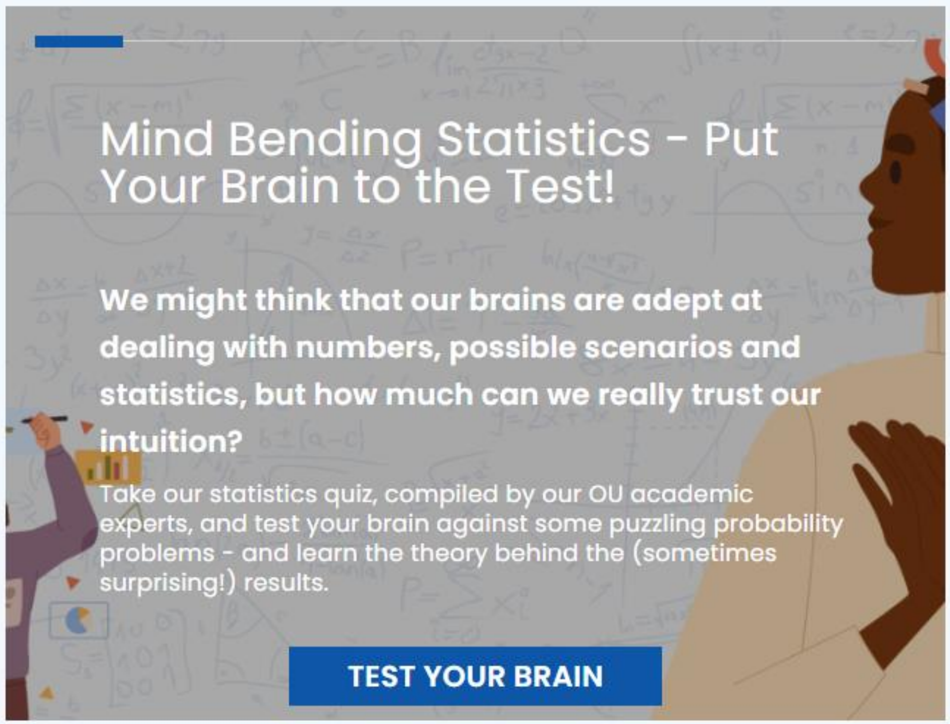
Quiz: Statistical Literacy

RADIO 4 More or Less: Behind the Stats

Home | More or Less on Radio 4 | More or Less on the World Service

More or Less Statistics Quiz

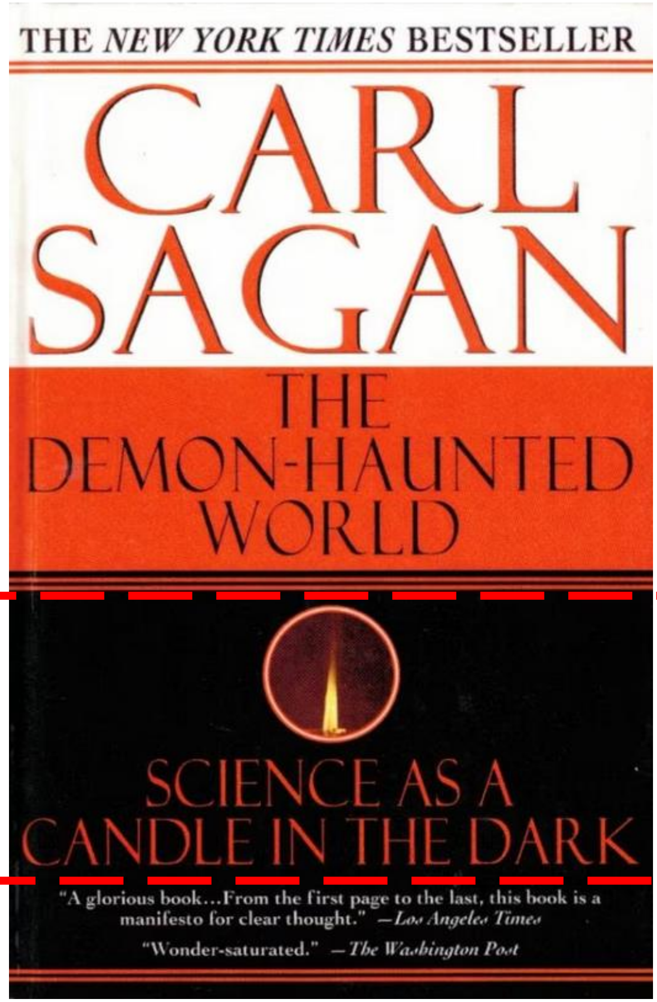
Think your brain is up to the challenge? Take the statistics quiz compiled by our OU academic experts and see whether you're surprised by the results...



Kingston University London

Final Word

Do not despair!
Science will show the way…



Further Reading

UNECE

SUSTAINABLE DEVELOPMENT GOALS

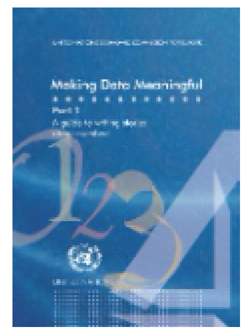
Advanced Search

About UNECE | Our work | Themes | SDGs | Open UNECE | Events | Publications | Media

UNECE › STATISTICS › MAKING DATA MEANINGFUL

Making Data Meaningful

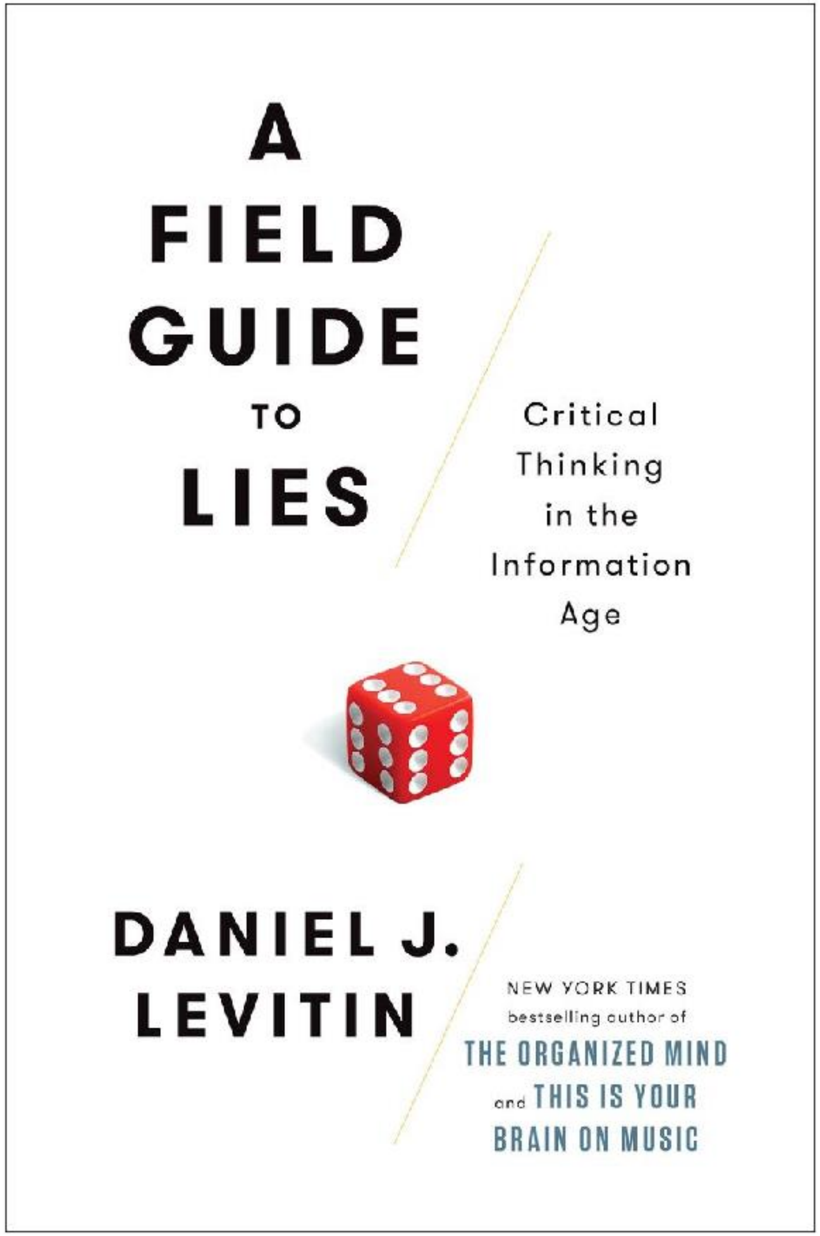
Four practical guides to help managers, statisticians and media relations officers in statistical organizations use text and visualizations to bring statistics to life for non-statisticians; find the best way to get their message across or define strategies for improving statistical literacy.



The guides are available in various languages. Print copies of the **English version** can be obtained free of charge (Part 2 is out of stock) by clicking **here**, indicating the publication title, the language and postal address. For more information about the Russian version, please click **here**

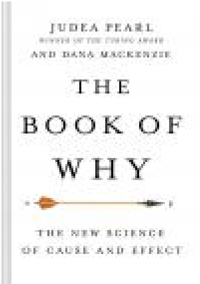
Kingston University London

Books for Further Reading



Kingston University London

Books for Further Reading

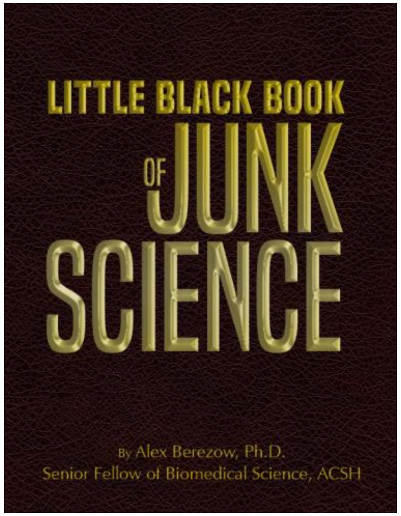


The Book of Why: The New Science of Cause and Effect

Basic Books

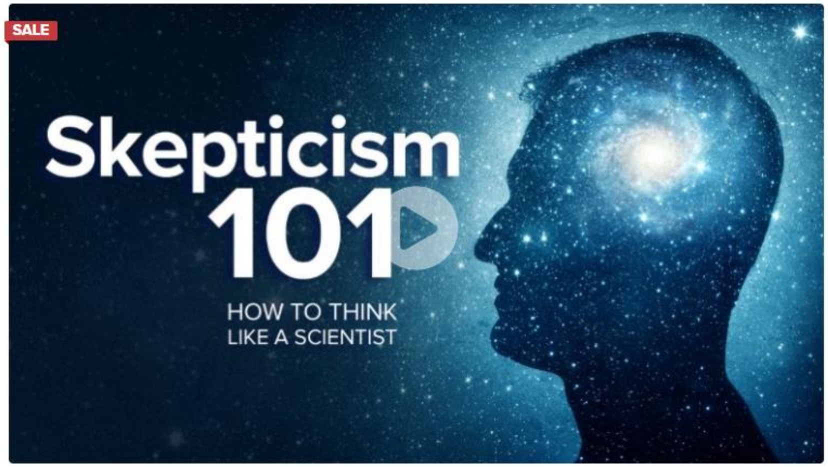
Judea Pearl, Dana Mackenzie

Year: 2018 Language: English



Kingston University London

Further Study



Your Deceptive Mind: A Scientific Guide to Critical Thinking Skills

Steven Novella, M.D. Professor, Yale School of Medicine

Own this Course

Instant Video ~~£39.9~~

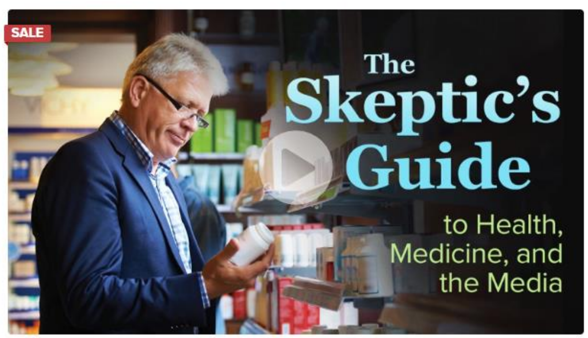
DVD ~~£54.9~~

Add to Cart

e-Gift

The Skeptic's Guide to Health, Medicine, and the Media

Roy Benaroch, M.D. Professor, Emory University



Own this Cour

Instant Video

Instant Audio

DVD

Add to

e-Gift

Medical Myths, Lies, and Half-Truths: What We Think We Know May Be Hurting Us

Steven Novella, M.D. Professor, Yale School of Medicine

Own this Course

Instant Video ~~£39.99~~ £9.9

Instant Audio ~~£24.99~~ £6.9

DVD ~~£54.99~~ £14.9

Add to Cart

e-Gift

Wish List

Further Study

Statistics

A collection of TED Talks (and more) on the topic of Statistics.

Video playlists about Statistics

10 TALKS

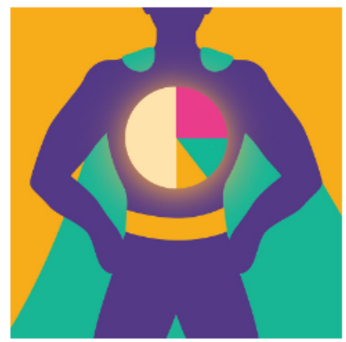
The value of skepticism

These TED Talks push us to question more -- our doctors, our governments and even our own eyes.

8 TALKS

Talks to help you become a better researcher

Strengthen your skills with these informative talks on how to get the most out of your research.



4 TALKS

How data can save lives

An exploration of what we can do to save the world with the massive amounts of data collected every day.

10 TALKS

Statistically speaking ...

Want a different perspective of the world? Get a better, more friendly grasp of statistics -- one of the most understood and misused tools in modern society.

See all playlists on Statistics

Further Study

Statistics

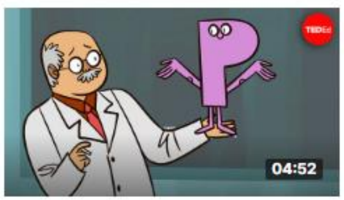
A collection of TED Talks (and more) on the topic of Statistics.

MONA CHALABI
What we miss when we focus on the average

MONA CHALABI
Do 9 out of 10 dentists really recommend that toothpaste?

TED AUDIO COLLECTIVE
Introducing: Am I Normal? with Mona Chalabi

CHARLOTTE DEGOT
A more accurate way to calculate emissions



JAMES A. SMITH
The method that can "prove" almost anything

DENNIS E. SHASHA
Can you solve the fantasy election riddle?

ALEX GENDLER
Can you solve the monster duel riddle?

DAN FINKEL
Can you solve the riddle and escape Hades?

GERD GIGERENZER
Why do people fear the wrong things?

ALEX GENDLER
Can you outsmart this logical fallacy?

ALEX EDMANS
What to trust in a "post-truth" world

MICHAEL GREEN
The global goals we've made progress on -- and the ones we haven't

Further Study

PLAYLIST

Statistically speaking ...

Want a different perspective of the world? Get a better, more friendly grasp of statistics -- one of the most understood and misused tools in modern society.

Watch now   Add to list

ALAN SMITH

Why you should love statistics

Think you're good at guessing stats? Guess again. Whether we consider ourselves math people or not, our ability to understand and work with numbers is terribly limited, says data visualization expert Alan Smith. In this delightful talk, Smith explores the...

HANNAH FRY

The mathematics of love

Finding the right mate is no cakewalk -- but is it even mathematically likely? In a charming talk, mathematician Hannah Fry shows patterns in how we look for love, and gives her top three tips (verified by math!) for finding that special someone.

ARTHUR BENJAMIN

Teach statistics before calculus!

Someone always asks the math teacher, "Am I going to use calculus in real life?" And for most of us, says Arthur Benjamin, the answer is no. He offers a bold proposal on how to make math education relevant in the digital age.

PETER DONNELLY

How juries are fooled by statistics

Oxford mathematician Peter Donnelly reveals the common mistakes humans make in interpreting statistics -- and the devastating impact these errors can have on the outcome of criminal trials.

STEVEN LEVITT

Surprising stats about child carseats

Steven Levitt shares data that shows car seats are no more effective than seatbelts in protecting kids from dying in cars. However, during the question and answer session, he makes one crucial caveat.

GARY WOLF

The quantified self

At TED@Cannes, Gary Wolf gives a 5-min intro to an intriguing new pastime: using mobile apps and always-on gadgets to track and analyze your body, mood, diet, spending -- just about everything in daily life you can measure -- in gloriously geeky...

CHRIS JORDAN

Turning powerful stats into art

Artist Chris Jordan shows us an arresting view of what Western culture looks like. His supersized images picture some almost unimaginable statistics -- like the astonishing number of paper cups we use every single day.

ANNE MILGRAM

Why smart statistics are the key to fighting crime

When she became the attorney general of New Jersey in 2007, Anne Milgram quickly discovered a few startling facts: not only did her team not really know who they were putting in jail, but they had no way of understanding if their decisions were...

HANS AND OLA ROSLING

How not to be ignorant about the world

How much do you know about the world? Hans Rosling, with his famous charts of global population, health and income data (and an extra-extra-long pointer), demonstrates that you have a high statistical chance of being quite wrong about wha...

SEBASTIAN WERNICKE

Lies, damned lies and statistics (about TEDTalks)

In a brilliantly tongue-in-cheek analysis, Sebastian Wernicke turns the tools of statistical analysis on TEDTalks, to come up with a metric for creating "the optimum TEDTalk" based on user ratings. How do you rate it? "Jaw-dropping"? "Unconvincing"? Or ju...

TED Ideas worth spreading

**** Further Study ****

Harriet Hall: Science Based Medicine

JamesRandiFoundation

10 videos 94,775 views Last updated on Nov 26, 2015

Play all Shuffle

This course is presented by Harriet Hall and consists of 10 lectures:

Science-Based Medicine vs. Evidence-Based Medicine
What Is CAM?
Chiropractic
Acupuncture
Homeopathy
Naturopathy and Herbal Medicine
Energy Medicine
Miscellaneous "Alternatives"
Pitfalls in Research
Science-Based Medicine in the Media and Politics

A Course Guide is available at:
http://web.randi.org/uploads/3/7/3/7/37377621/c

1. Lecture 1: Science Based Medicine vs.Evidence Based Medicine — 37:01 — JamesRandiFoundation • 97K views • 7 years ago
2. Lecture 2: What is CAM — 36:09 — JamesRandiFoundation • 36K views • 7 years ago
3. Lecture 3: Chiropractic — 35:47 — JamesRandiFoundation • 78K views • 7 years ago
4. Lecture 4: Acupuncture — 31:49 — JamesRandiFoundation • 45K views • 7 years ago
5. Lecture 5: Homeopathy — 32:33 — JamesRandiFoundation • 42K views • 7 years ago
6. Lecture 6: Naturopathy and Herbal Medicines — 43:58 — JamesRandiFoundation • 37K views • 7 years ago
7. Lecture 7: Energy Medicine — 36:59 — JamesRandiFoundation • 22K views • 7 years ago
8. Lecture 8: Miscellaneous "Alternatives" — 42:39 — JamesRandiFoundation • 18K views • 7 years ago
9. Lecture 9: Pitfalls in Research — 44:14 — JamesRandiFoundation • 14K views • 7 years ago
10. Lecture 10: Science-Based Medicine in the Media and Politics — 44:45 — JamesRandiFoundation • 13K views • 7 years ago

James Randi Educational Foundation

HOME SWIFT ABOUT EDUCATION JREF PRESS

Search: harriet hall

Filter
Everything (50)
Pages (8)
Blog posts (41)

- Ask the SkepDoc: Cellphone antennas and electrodermal screening
- Ask the Skepdoc: Pancreas miracle and Dispenza's thought healing
- Ask the SkepDoc: Hepatitis shots, memory supplement and autism
- Ask the Skepdoc: Qivana diet supplement claims and hyperthermi...
- Last Week in Science-Based Medicine 22 September 2014
- Last week in Science Based Medicine 27-October 2014
- Science-based medicine content from TAM 2013
- Last Week in Science-based Medicine 29 September 2014
- Free Course: " Science Based Medicine" by Harriet Hall, MD - The S...

Kingston University London

Critical Skills

A Guide for University Students

Mehmet Tevfik DORAK, MD PhD

School of Life Sciences, Pharmacy & Chemistry

Kingston University London

www.dorak.info

Kingston University London

Critical Thinking

A Guide for University Students

Mehmet Tevfik DORAK, MD PhD

School of Life Sciences, Pharmacy & Chemistry

Kingston University London

www.dorak.info

Kingston University London

Welcome to Mehmet Tevfik DORAK's Website

Epidemiology
Biostatistics & Statistical Literacy
Statistics & Graphics with R
Critical Skills   Critical Thinking
Leadership & Equality   Gender Equality
Essentials of Genetics
Genome Biology for Genetic Epidemiologists
Immunogenetics
Evolutionary Biology Notes
Music Notes

Mehmet Tevfik DORAK, BA (Hons), MD, PhD

http://www.dorak.info