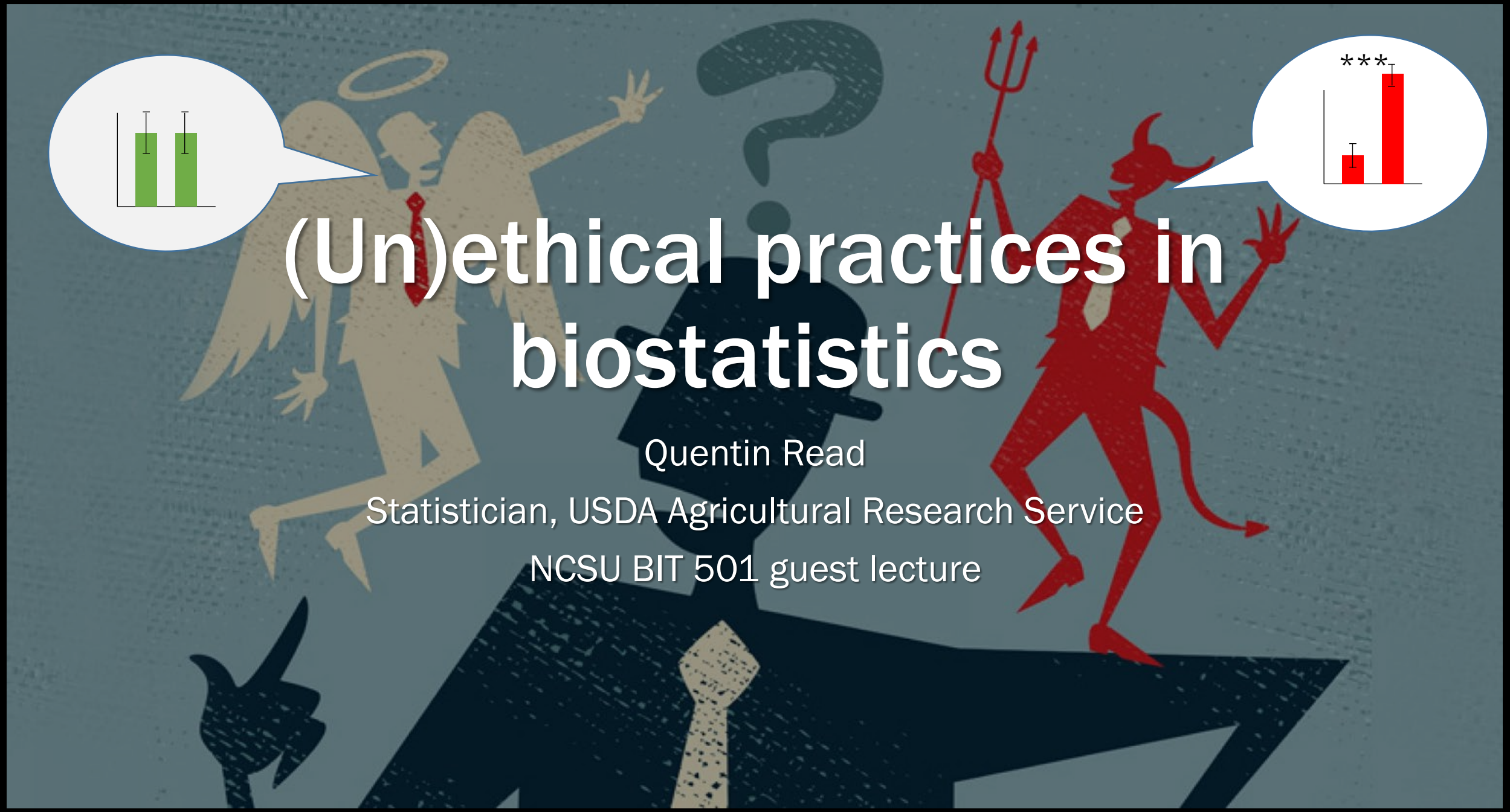



# (Un)ethical practices in biostatistics

Quentin Read

Statistician, USDA Agricultural Research Service

NCSU BIT 501 guest lecture


Statistics' problematic historical legacy

Data fraud: a cautionary tale

Perverse incentives

Questionable research practices

Ways forward

Statistics' problematic historical legacy revisited


Statistics' problematic historical legacy

Data fraud: a cautionary tale

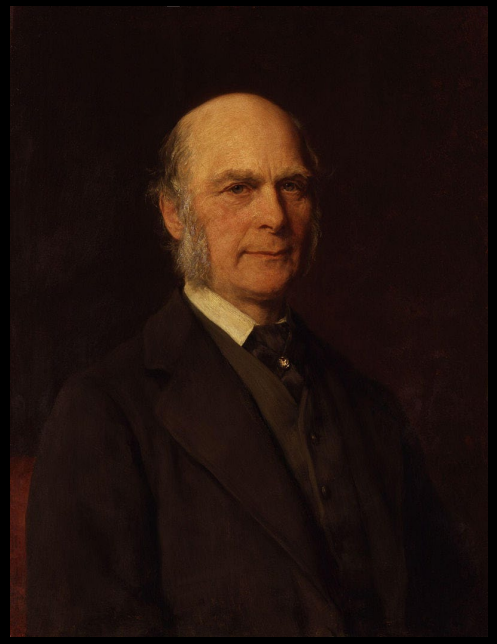
Perverse incentives

Questionable research practices

Ways forward

Statistics' problematic historical legacy revisited

# Problematic historical legacy

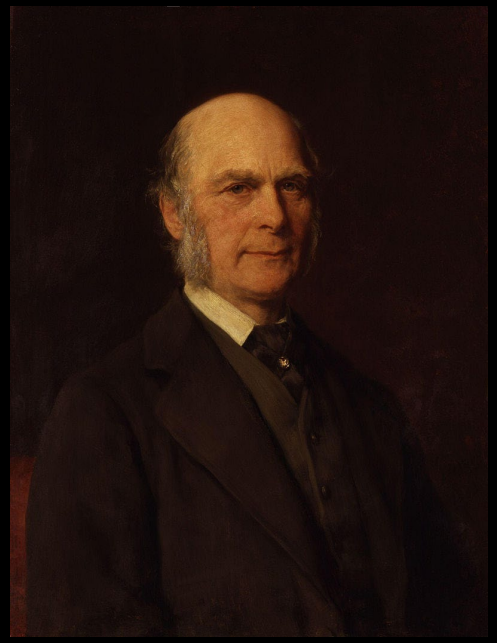


**Francis Galton** (1822-1911)
- Invented correlation

**Ronald A. Fisher** (1890-1962)
- Formalized ANOVA technique
- Proposed statistical significance

**Charles Spearman** (1863-1945)
- Developed factor analysis

# Problematic historical legacy



**Francis Galton** (1822-1911)

- Coined term “eugenics”
- *“give to the more suitable races … a better chance of prevailing speedily over the less suitable.”*

**Ronald A. Fisher** (1890-1962)

- Led eugenics society
- Defended actions of Nazi Germany

**Charles Spearman** (1863-1945)

- Theorized general intelligence quotient *g*
- Promoted theory of hierarchical racial differences in intelligence

Statistics’ problematic historical legacy

Data fraud: a cautionary tale

Perverse incentives

Questionable research practices

Ways forward

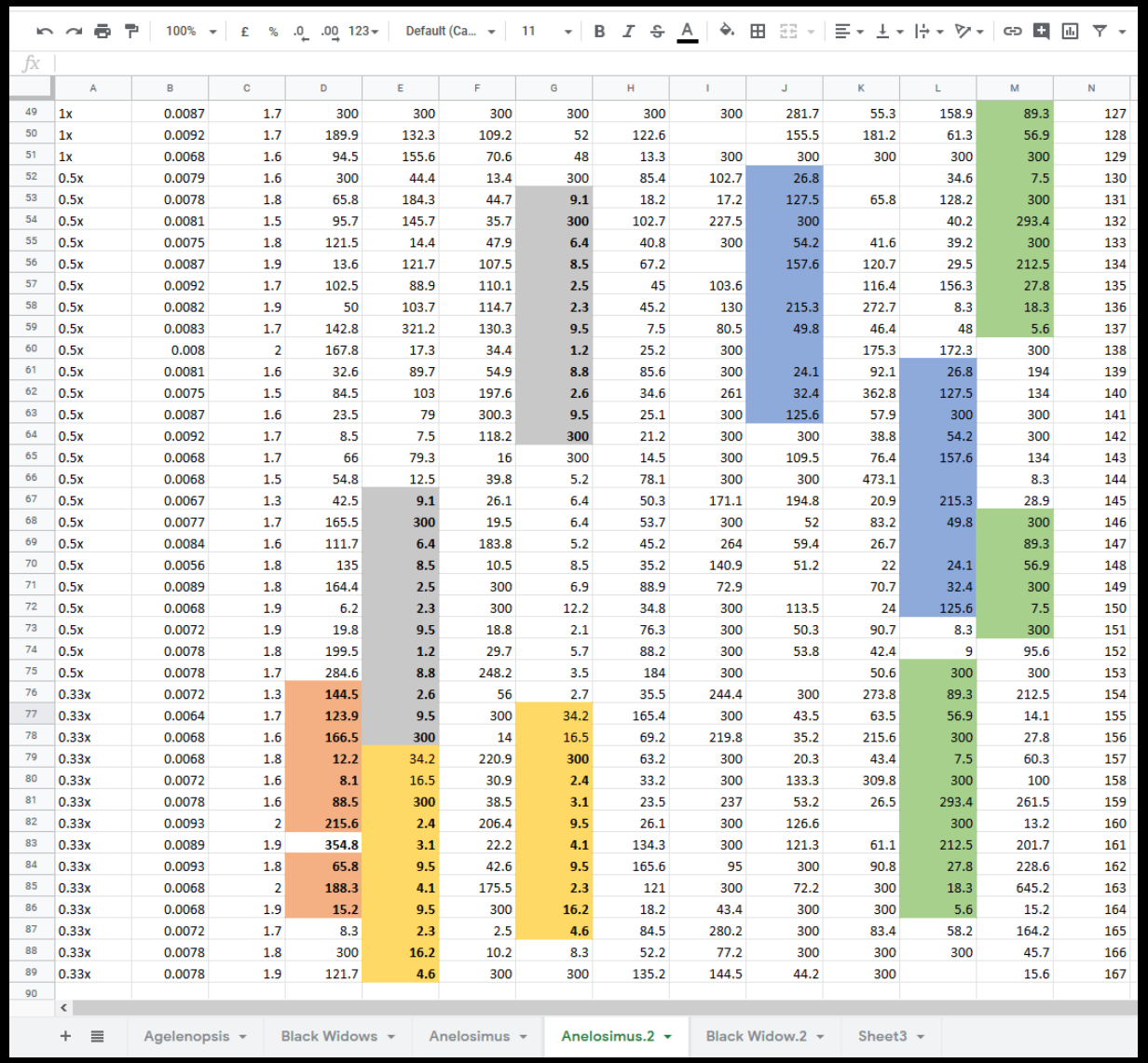
Statistics’ problematic historical legacy revisited

# Case study: Jonathan Pruitt

Behavioral ecologist studying social interactions of spiders

In 2020, researchers interested in his science noticed inconsistencies in published data

This led to a thorough investigation of all his published research



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 1x | 0.0087 | 1.7 | 300 | 300 | 300 | 300 | 300 | 300 | 281.7 | 55.3 | 158.9 | 89.3 | 127 |
| 50 | 1x | 0.0092 | 1.7 | 189.9 | 132.3 | 109.2 | 52 | 122.6 | | 155.5 | 181.2 | 61.3 | 56.9 | 128 |
| 51 | 1x | 0.0068 | 1.6 | 94.5 | 155.6 | 70.6 | 48 | 13.3 | 300 | 300 | 300 | 300 | 300 | 129 |
| 52 | 0.5x | 0.0079 | 1.6 | 300 | 44.4 | 13.4 | 300 | 85.4 | 102.7 | 26.8 | | 34.6 | 7.5 | 130 |
| 53 | 0.5x | 0.0078 | 1.8 | 65.8 | 184.3 | 44.7 | 9.1 | 18.2 | 17.2 | 127.5 | 65.8 | 128.2 | 300 | 131 |
| 54 | 0.5x | 0.0081 | 1.5 | 95.7 | 145.7 | 35.7 | 300 | 102.7 | 227.5 | 300 | | 40.2 | 293.4 | 132 |
| 55 | 0.5x | 0.0075 | 1.8 | 121.5 | 14.4 | 47.9 | 6.4 | 40.8 | 300 | 54.2 | 41.6 | 39.2 | 300 | 133 |
| 56 | 0.5x | 0.0087 | 1.9 | 13.6 | 121.7 | 107.5 | 8.5 | 67.2 | | 157.6 | 120.7 | 29.5 | 212.5 | 134 |
| 57 | 0.5x | 0.0092 | 1.7 | 102.5 | 88.9 | 110.1 | 2.5 | 45 | 103.6 | | 116.4 | 156.3 | 27.8 | 135 |
| 58 | 0.5x | 0.0082 | 1.9 | 50 | 103.7 | 114.7 | 2.3 | 45.2 | 130 | 215.3 | 272.7 | 8.3 | 18.3 | 136 |
| 59 | 0.5x | 0.0083 | 1.7 | 142.8 | 321.2 | 130.3 | 9.5 | 7.5 | 80.5 | 49.8 | 46.4 | 48 | 5.6 | 137 |
| 60 | 0.5x | 0.008 | 2 | 167.8 | 17.3 | 34.4 | 1.2 | 25.2 | 300 | | 175.3 | 172.3 | 300 | 138 |
| 61 | 0.5x | 0.0081 | 1.6 | 32.6 | 89.7 | 54.9 | 8.8 | 85.6 | 300 | 24.1 | 92.1 | 26.8 | 194 | 139 |
| 62 | 0.5x | 0.0075 | 1.5 | 84.5 | 103 | 197.6 | 2.6 | 34.6 | 261 | 32.4 | 362.8 | 127.5 | 134 | 140 |
| 63 | 0.5x | 0.0087 | 1.6 | 23.5 | 79 | 300.3 | 9.5 | 25.1 | 300 | 125.6 | 57.9 | 300 | 300 | 141 |
| 64 | 0.5x | 0.0092 | 1.7 | 8.5 | 7.5 | 118.2 | 300 | 21.2 | 300 | 300 | 38.8 | 54.2 | 300 | 142 |
| 65 | 0.5x | 0.0068 | 1.7 | 66 | 79.3 | 16 | 300 | 14.5 | 300 | 109.5 | 76.4 | 157.6 | 134 | 143 |
| 66 | 0.5x | 0.0068 | 1.5 | 54.8 | 12.5 | 39.8 | 5.2 | 78.1 | 300 | 300 | 473.1 | | 8.3 | 144 |
| 67 | 0.5x | 0.0067 | 1.3 | 42.5 | 9.1 | 26.1 | 6.4 | 50.3 | 171.1 | 194.8 | 20.9 | 215.3 | 28.9 | 145 |
| 68 | 0.5x | 0.0077 | 1.7 | 165.5 | 300 | 19.5 | 6.4 | 53.7 | 300 | 52 | 83.2 | 49.8 | 300 | 146 |
| 69 | 0.5x | 0.0084 | 1.6 | 111.7 | 6.4 | 183.8 | 5.2 | 45.2 | 264 | 59.4 | 26.7 | | 89.3 | 147 |
| 70 | 0.5x | 0.0056 | 1.8 | 135 | 8.5 | 10.5 | 8.5 | 35.2 | 140.9 | 51.2 | 22 | 24.1 | 56.9 | 148 |
| 71 | 0.5x | 0.0089 | 1.8 | 164.4 | 2.5 | 300 | 6.9 | 88.9 | 72.9 | | 70.7 | 32.4 | 300 | 149 |
| 72 | 0.5x | 0.0068 | 1.9 | 6.2 | 2.3 | 300 | 12.2 | 34.8 | 300 | 113.5 | 24 | 125.6 | 7.5 | 150 |
| 73 | 0.5x | 0.0072 | 1.9 | 19.8 | 9.5 | 18.8 | 2.1 | 76.3 | 300 | 50.3 | 90.7 | 8.3 | 300 | 151 |
| 74 | 0.5x | 0.0078 | 1.8 | 199.5 | 1.2 | 29.7 | 5.7 | 88.2 | 300 | 53.8 | 42.4 | 9 | 95.6 | 152 |
| 75 | 0.5x | 0.0078 | 1.7 | 284.6 | 8.8 | 248.2 | 3.5 | 184 | 300 | | 50.6 | 300 | 300 | 153 |
| 76 | 0.33x | 0.0072 | 1.3 | 144.5 | 2.6 | 56 | 2.7 | 35.5 | 244.4 | 300 | 273.8 | 89.3 | 212.5 | 154 |
| 77 | 0.33x | 0.0064 | 1.7 | 123.9 | 9.5 | 300 | 34.2 | 165.4 | 300 | 43.5 | 63.5 | 56.9 | 14.1 | 155 |
| 78 | 0.33x | 0.0068 | 1.6 | 166.5 | 300 | 14 | 16.5 | 69.2 | 219.8 | 35.2 | 215.6 | 300 | 27.8 | 156 |
| 79 | 0.33x | 0.0068 | 1.8 | 12.2 | 34.2 | 220.9 | 300 | 63.2 | 300 | 20.3 | 43.4 | 7.5 | 60.3 | 157 |
| 80 | 0.33x | 0.0072 | 1.6 | 8.1 | 16.5 | 30.9 | 2.4 | 33.2 | 300 | 133.3 | 309.8 | 300 | 100 | 158 |
| 81 | 0.33x | 0.0078 | 1.6 | 88.5 | 300 | 38.5 | 3.1 | 23.5 | 237 | 53.2 | 26.5 | 293.4 | 261.5 | 159 |
| 82 | 0.33x | 0.0093 | 2 | 215.6 | 2.4 | 206.4 | 9.5 | 26.1 | 300 | 126.6 | | 300 | 13.2 | 160 |
| 83 | 0.33x | 0.0089 | 1.9 | 354.8 | 3.1 | 22.2 | 4.1 | 134.3 | 300 | 121.3 | 61.1 | 212.5 | 201.7 | 161 |
| 84 | 0.33x | 0.0093 | 1.8 | 65.8 | 9.5 | 42.6 | 9.5 | 165.6 | 95 | 300 | 90.8 | 27.8 | 228.6 | 162 |
| 85 | 0.33x | 0.0068 | 2 | 188.3 | 4.1 | 175.5 | 2.3 | 121 | 300 | 72.2 | 300 | 18.3 | 645.2 | 163 |
| 86 | 0.33x | 0.0068 | 1.9 | 15.2 | 9.5 | 300 | 16.2 | 18.2 | 43.4 | 300 | 300 | 5.6 | 15.2 | 164 |
| 87 | 0.33x | 0.0072 | 1.7 | 8.3 | 2.3 | 2.5 | 4.6 | 84.5 | 280.2 | 300 | 83.4 | 58.2 | 164.2 | 165 |
| 88 | 0.33x | 0.0078 | 1.8 | 300 | 16.2 | 10.2 | 8.3 | 52.2 | 77.2 | 300 | 300 | 300 | 45.7 | 166 |
| 89 | 0.33x | 0.0078 | 1.9 | 121.7 | 4.6 | 300 | 300 | 135.2 | 144.5 | 44.2 | 300 | | 15.6 | 167 |
| 90 | | | | | | | | | | | | | | |

Agelenopsis | Black Widows | Anelosimus | Anelosimus.2 | Black Widow.2 | Sheet3

Investigations uncovered duplicated values and formulas entered into Excel cells supposed to contain raw data

Many papers were retracted, Pruitt lost his job, the careers of many collaborators were damaged

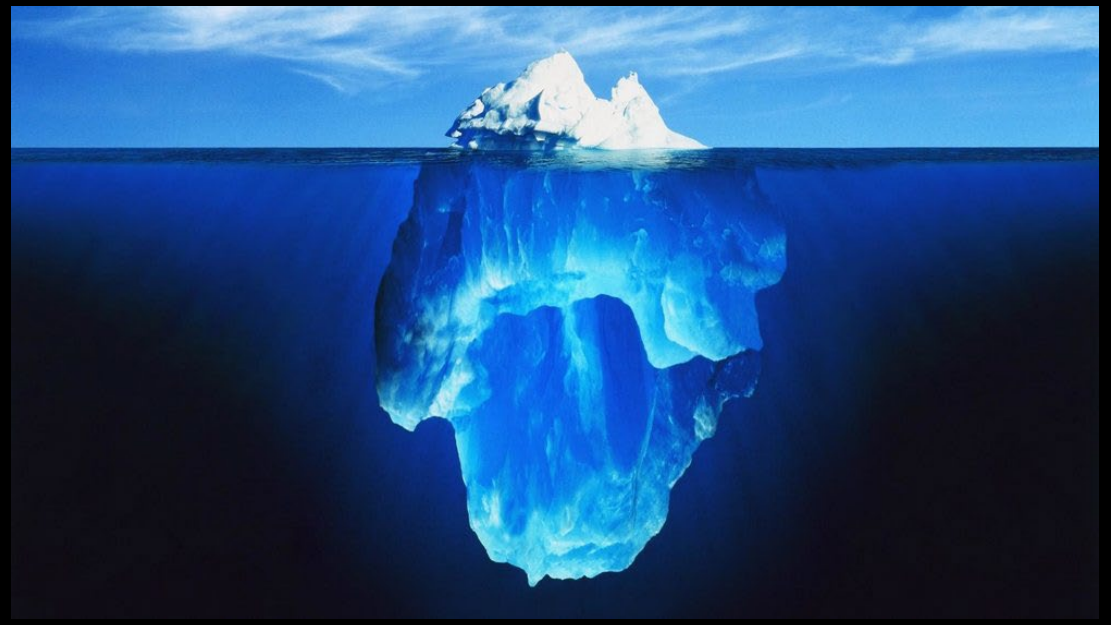
Science is based on trust; he violated that trust

“Open data” enabled detection of the fraud

Image source: https://forbetterscience.com/2020/03/14/spidermans-lawyer-is-having-you-for-dinner-tonight/

# Statistical misconduct is not just fraud

Media narrative: “a few bad apples” commit outright fraud



Or is blatant fraud the “tip of the misconduct iceberg?”

# Meta-analysis of surveys on statistical misconduct

2% of scientists “admitted to have fabricated, falsified, or modified data or results at least once”

34% “admitted other questionable research practices”

- Intentionally not publishing results
- Biased methodology
- Misleading reporting
- "dropping data points based on a gut feeling“
- "changing the design, methodology, or results of a study in response to pressures from a funding source"

Fanelli 2009, *PLoS ONE*

Statistical misconduct is a systemic problem.

Systemic problems have systemic causes.

Statistics’ problematic historical legacy

Data fraud: a cautionary tale

Perverse incentives

Questionable research practices

Ways forward

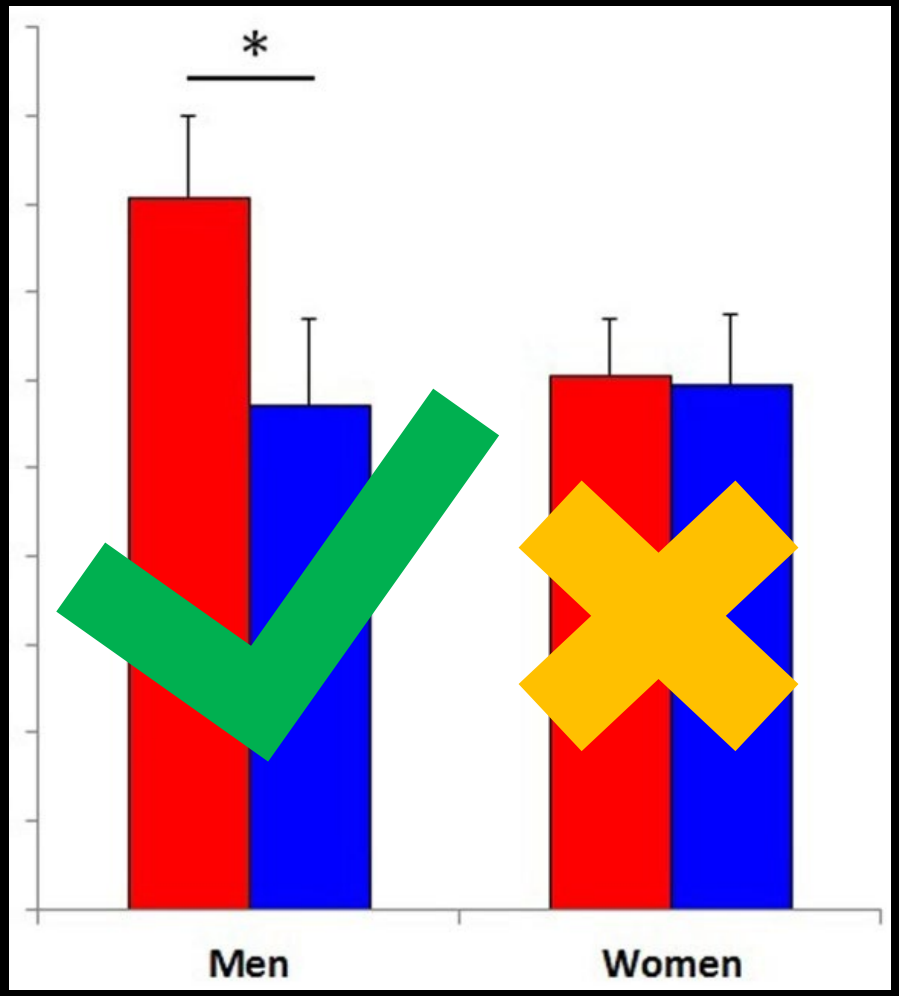
Statistics’ problematic historical legacy revisited

*Think pair share activity:*

Come up with at least three reasons why researchers might engage in statistical misconduct.

# Why do people engage in statistical misconduct?

- “I want definitive answers”
- Statistics is seen as a way to get “yes or no” answers
- Null hypothesis significance testing framework supports this
- $p < 0.05$ is a threshold for “statistical significance”
- It has no basis in biological or any kind of scientific reality
- Much statistical misconduct is based on trying to achieve statistically significant results



# Why do people engage in statistical misconduct?

Financial and professional incentives

- Scientific research is extremely competitive
- Novel and exciting findings have higher impact
- Career advancement is tied to number of high-impact publications
- Non-significant results have a lower chance of publication

Commitment to theory or idea

- Experiments often designed to test a particular hypothesis or theory, with a particular outcome in mind
- Stigma associated with being wrong (reputations at stake)

Statistics’ problematic historical legacy

Data fraud: a cautionary tale

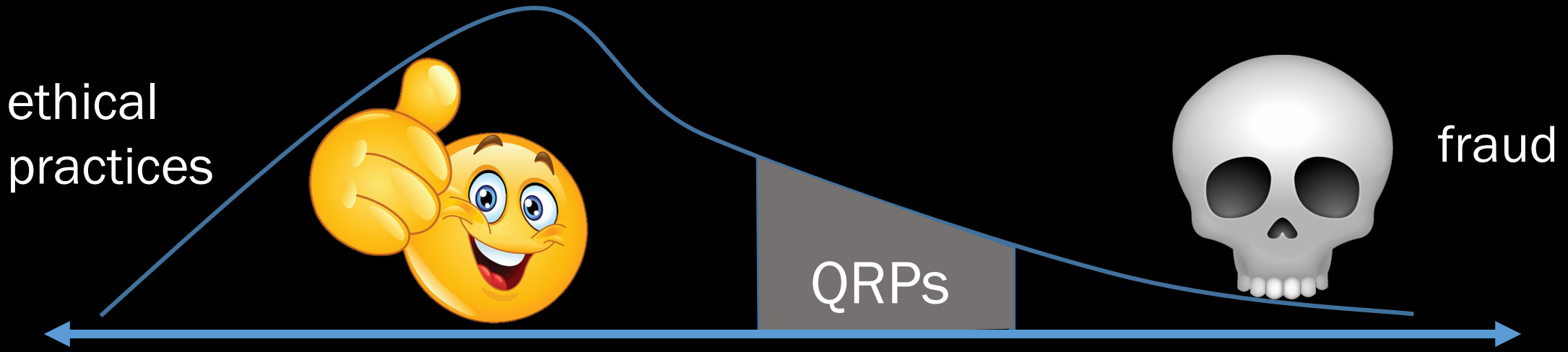
Perverse incentives

Questionable research practices

Ways forward

Statistics’ problematic historical legacy revisited

# Questionable research practices (QRPs)




Statistical misconduct is a continuum

QRPs are potentially unethical practices that fall short of outright fraud

Some QRPs are so subtle that researchers may not even be conscious of engaging in them

# Examples of QRPs

Multiple comparisons

P-hacking

HARKing

File drawer problem

*(these practices overlap to some extent)*

# Multiple comparisons

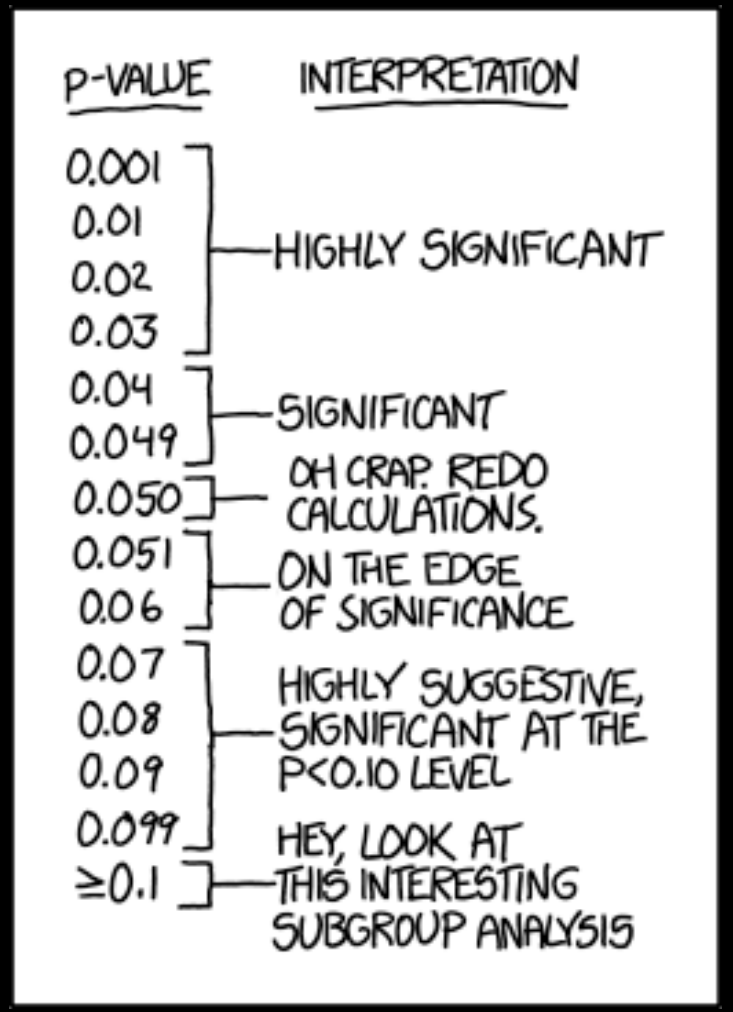
- Also known as “fishing” or “data dredging”
- Testing many possible predictor variables (*x*) or response variables (*y*) – and reporting only statistically significant ones
- It’s possible to correct for this but not often done in practice
- The real problem is the lack of transparency

# P-hacking

Trying many analyses until one yields a significant result

P-hacking refers to p-value

“Researcher degrees of freedom”



# The garden of forking paths

Analyzing subgroups of data may give different results than when the data are pooled

Adding different covariates to the analysis may affect the result

Becomes statistical misconduct (p-hacking) when it's done indiscriminately and without prior planning

Gelman & Loken 2013: http://stat.columbia.edu/~gelman/research/unpublished/forking.pdf

# Many analysts, one dataset

Are soccer (football) referees more likely to give red cards to dark-skinned or light-skinned players?

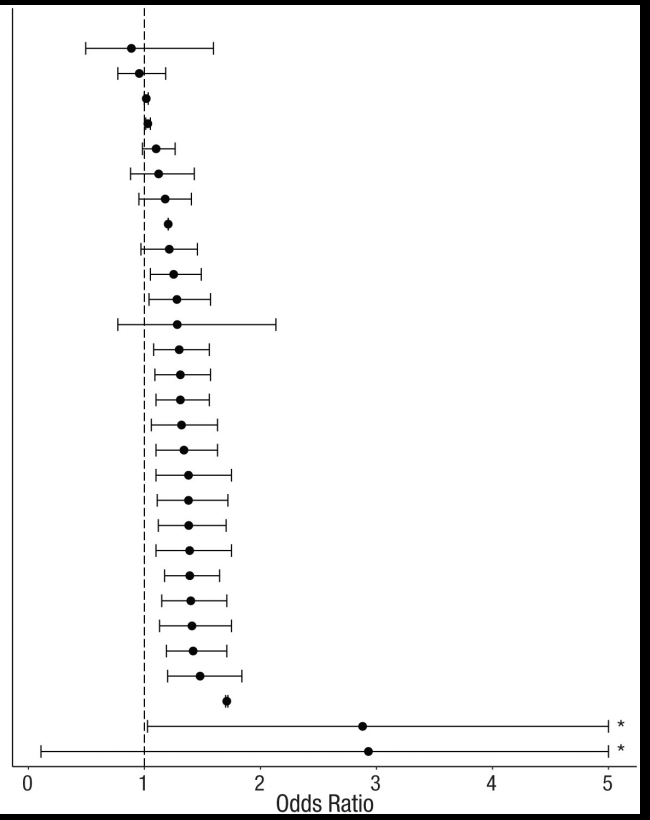
Same dataset was given to 29 teams of analysts

Dataset included many possible variables that could be controlled for

Silberzahn et al. 2018, *Advances in Methods and Practices in Psychological Science*

# Many analysts, one dataset

| Team | Analytic Approach | Odds Ratio |
|---|---|---|
| 12 | Zero-Inflated Poisson Regression | 0.89 |
| 17 | Bayesian Logistic Regression | 0.96 |
| 15 | Hierarchical Log-Linear Modeling | 1.02 |
| 10 | Multilevel Regression and Logistic Regression | 1.03 |
| 18 | Hierarchical Bayes Model | 1.10 |
| 31 | Logistic Regression | 1.12 |
| 1 | OLS Regression With Robust Standard Errors, Logistic Regression | 1.18 |
| 4 | Spearman Correlation | 1.21 |
| 14 | WLS Regression With Clustered Standard Errors | 1.21 |
| 11 | Multiple Linear Regression | 1.25 |
| 30 | Clustered Robust Binomial Logistic Regression | 1.28 |
| 6 | Linear Probability Model | 1.28 |
| 26 | Hierarchical Generalized Linear Modeling With Poisson Sampling | 1.30 |
| 3 | Multilevel Logistic Regression Using Bayesian Inference | 1.31 |
| 23 | Mixed-Model Logistic Regression | 1.31 |
| 16 | Hierarchical Poisson Regression | 1.32 |
| 2 | Linear Probability Model, Logistic Regression | 1.34 |
| 5 | Generalized Linear Mixed Models | 1.38 |
| 24 | Multilevel Logistic Regression | 1.38 |
| 28 | Mixed-Effects Logistic Regression | 1.38 |
| 32 | Generalized Linear Models for Binary Data | 1.39 |
| 8 | Negative Binomial Regression With a Log Link | 1.39 |
| 20 | Cross-Classified Multilevel Negative Binomial Model | 1.40 |
| 13 | Poisson Multilevel Modeling | 1.41 |
| 25 | Multilevel Logistic Binomial Regression | 1.42 |
| 9 | Generalized Linear Mixed-Effects Models With a Logit Link | 1.48 |
| 7 | Dirichlet-Process Bayesian Clustering | 1.71 |
| 21 | Tobit Regression | 2.88 |
| 27 | Poisson Regression | 2.93 |




Effect was measured as odds ratio (OR) with 95% confidence interval

OR = 1 means no effect of skin color; OR > 1 means dark skin color is positively related to red cards

20/29 (69%) found positive effect, 9/29 (31%) found no effect

Silberzahn et al. 2018, *Advances in Methods and Practices in Psychological Science*

# Many analysts, one dataset

Prior beliefs and level of expertise did not affect outcomes of the analysis

All analyses were transparent

No incentive to get a significant or positive result

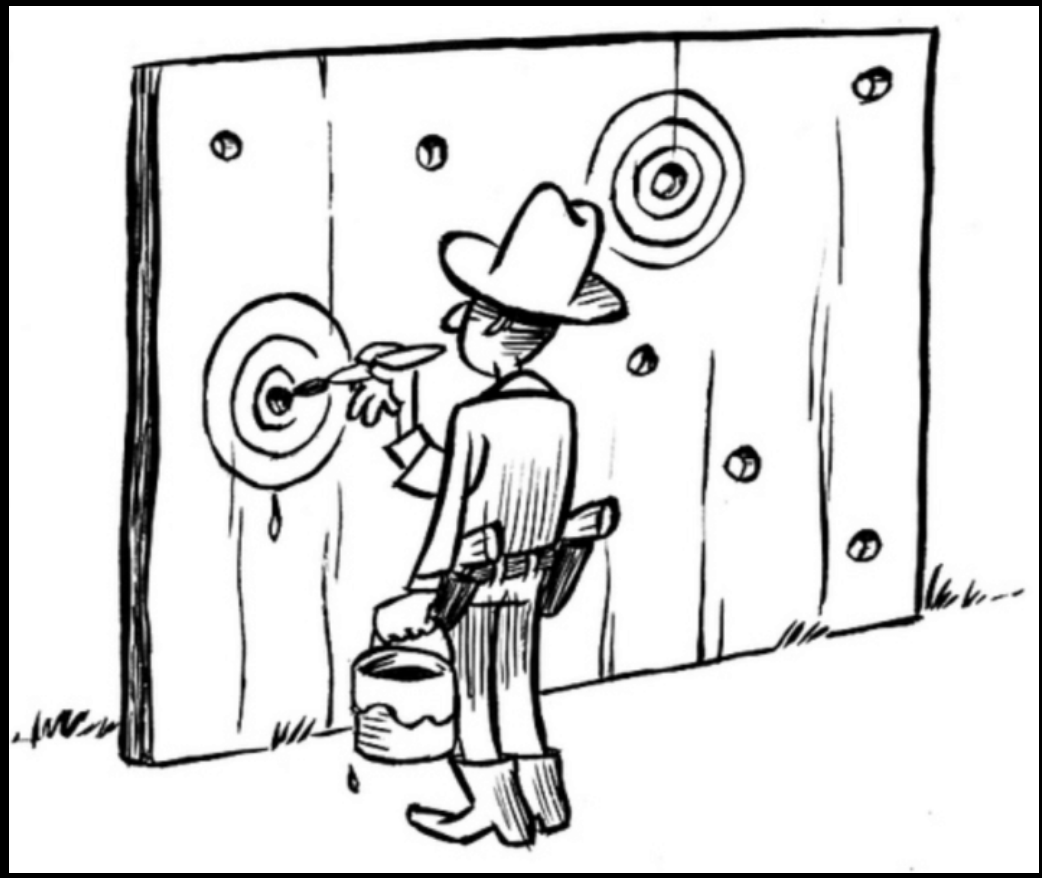
But in the real world there *are* incentives!!!

# “HARK”ing = hypothesizing after results are known

Treating exploratory analysis as if it was confirmatory analysis (testing a preexisting hypothesis)

But p-values assume the data was collected *with that hypothesis in mind*

It’s not wrong to explore data and find new and surprising patterns – but this should be followed up with confirmatory studies



“Texas Sharpshooter” effect

# File drawer problem

- Selective publication
- Selective reporting of dependent variables
- Discarding results with contradicting evidence


# Case study: Brian Wansink

Cornell professor, head of food psychology lab

Media-friendly research on how environmental cues affect how much, and what kind, of food people eat (“nudges”)

- “bottomless bowl”
- Elmo stickers on apples
- all-you-can-eat pizza buffet

# Brian Wansink: lab culture of QRPs

In 2017, blog posts by Wansink bragging about research practices caused people to reexamine many of his papers published over 30 years

Email correspondence revealed a lab culture where graduate students and postdocs were pressured and incentivized to engage in QRPs

*“As Steve Jobs said, ‘Geniuses ship.’”*

https://www.buzzfeednews.com/article/stephaniemlee/brian-wansink-cornell-p-hacking

# “Pizzagate”

Hi Ozge,

Glad you had a chance to take an initial look at the data.

I don't think I've ever done an interesting study where the data "came out" the first time I looked at it. The interesting stories come from seeing when things -- like the 1/2 price buffet -- works and when it doesn't.

I would like you to really dig into this to find a number of situations or people for which this relationship does hold -- that is where the 1/2 price buffet did result in a difference.

Here's some things to do.

First, look to see if there are weird outliers (in terms of how much they ate). If there seems to be a reason they are different, pull them out but specially note why you did so, so that this can be described in the method.

Second, think of all the different ways you can cut the data and analyze subsets of it to see when this relationship holds. For instance, if it works on men but not women, we have a moderator. Here are some groups you'll want to break out separately:

Subgroups: “males, females, lunch goers, dinner goers, people sitting alone, people eating with groups of 2, people eating in groups of 2+, people who order alcohol, people who order soft drinks, people who sit close to buffet, people who sit far away, and so on…”

Response variables: “# pieces of pizza, # trips, fill level of plate, did they get dessert, did they order a drink, and so on…”

# P-hack me Elmo


Hi David,

Here's the Elmo study we are going to spin off and submit.
One sticking point is that although the stickers increase apple selection by 71%, for some reason this is a p value of .06. It seems to me it should be lower. Do you want to take a look at it and see what you think. If you can get the data, and it needs some tweeking, it would be good to get that one value below .05.

Best,

Brian

# Brian Wansink: the fallout

Ultimately ~20 papers were retracted

Wansink resigned his professorship at Cornell in 2019

*“There was no fraud, no intentional misreporting, no plagiarism, or no misappropriation.”*

This is a high-profile case but “probably quite typical of what goes on in a lot of labs.” (Nick Brown)

https://www.vox.com/science-and-health/2018/9/19/17879102/brian-wansink-cornell-food-brand-lab-retractions-jama

# Other QRPs

Stating conclusions not justified by the data

Insufficiently reporting flaws and limitations of study

“Journal shopping”: intentionally submitting a flawed result to many scientific journals until it passes peer review

And more …

# Time for some examples!

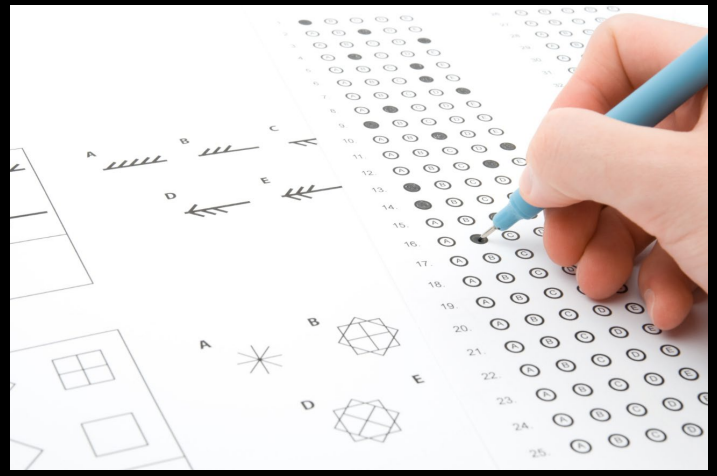
*Think pair share activity:*

For each of the following three (made-up but realistic) examples, think about:

- How, if at all, did the researchers behave unethically?
- Which categories of questionable research practices might it be classified as?
- What would you recommend they have done instead?

# Example 1

A team of researchers is interested in whether smelling essential oils benefits cognitive performance. They set up an experiment where some subjects smell essential oils and some smell a placebo, then take a test. The researchers find no evidence for an effect overall, but they reanalyze the data separately by age group and find a positive effect in subjects aged 18-30. They publish this result in a manuscript titled “Positive effects of essential oils on cognitive function.”



# Example 2

Another team of researchers wants to know whether being exposed to 5G radiation influences risk of disease. They collect data on 10,000 people, including the distance each person lives from a 5G radiation source and each person's disease outcomes for 150 different diseases. They find a positive association between proximity to 5G radiation and probability of developing a certain class of brain tumor. They publish a manuscript titled "5G radiation exposure: potential links to brain cancer."

# Example 3

Yet another team of researchers wants to investigate the possible health benefits of a newly isolated molecule found in coffee beans. They conduct three experiments, with slightly different methodological details each time (different doses, responses measured at different times, different measures of health outcome, etc.). Two of the experiments show no significant differences between control and treatment groups. They write up those results and submit them to journals, but the manuscripts are rejected. The analysis of the data from the third experiment shows significant health benefits associated with the molecule. They write up their results and submit them to a high-profile journal. The manuscript is accepted and published.

Statistics’ problematic historical legacy

Data fraud: a cautionary tale

Perverse incentives

Questionable research practices

Ways forward

Statistics’ problematic historical legacy revisited

# Ways forward

*Think pair share activity:*

What are some ways you can think of that would make researchers in the life sciences more likely to practice good statistical ethics?

# Ways forward (Andrew Gelman)

Communicate uncertainty and variance (not definitive answers)

Cultivate a "culture of respect for data" (just presenting data should be seen as valuable even if it doesn't have innovative data analysis or conclusions, reducing the pressure to hype conclusions)

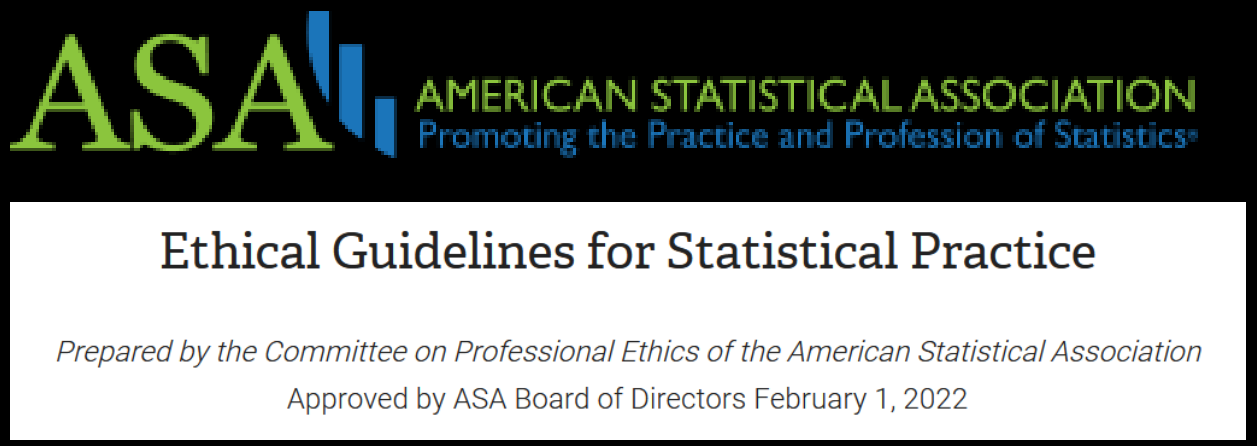
Reduce stigma associated with being corrected, self-correction, and being wrong

Respect the limitations of statistics (if an event is very rare or an effect is very weak, statistical conclusions about it may be unreliable because the data simply can't detect an effect)

View statistics as a way to overcome our natural tendency to see things as yes/no, not as a way to codify that false premise into a formal analysis

Gelman 2018, *Significance*

# Ways forward (AmStat ethical guidelines)




Be transparent and honest about assumptions (not just about results)

Be aware of potential algorithmic biases

Especially when it comes to humans, take variation into account!

https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice

# Specific recommendations

Remove perverse incentives, especially around publication

Do not tie publication to statistical significance

Preregistration of experimental designs

Encourage replication or reproduction of studies

Open and reproducible data and code

Statistical methods not focused on rejecting or accepting a null hypothesis (not yes-or-no answer; Bayesian methods)

Science is fundamentally based on trust, but we have the responsibility to be as transparent as possible about how we got our data and how we analyzed it.

Statistics’ problematic historical legacy

Data fraud: a cautionary tale

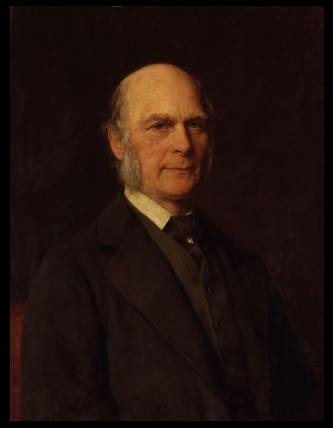
Perverse incentives

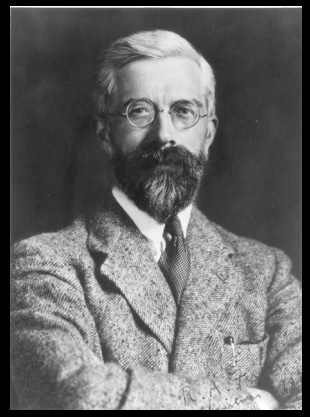
Questionable research practices

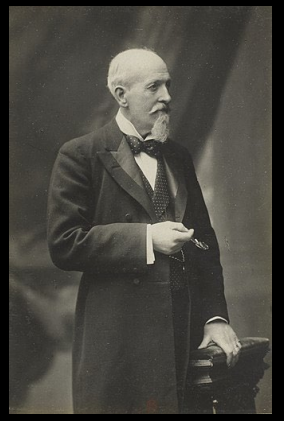
Ways forward

Statistics’ problematic historical legacy revisited

# Problematic historical legacy, revisited







It is not a coincidence that the founders of the “null hypothesis significance testing” framework espoused scientific racism

They developed methods that yield black-and-white definitive answers in part because of their biased, black-and-white, hierarchical worldview

Statistics is a human endeavor and will reflect the bias of the humans that created it

We cannot be completely objective but we can be transparent about our assumptions


# Questions?

quentin.read@usda.gov
https://quentinread.com
@QuentinDRead

# Reading list

## Journal articles

- Gelman 2018 (statistical ethics): https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2018.01193.x
- Fanelli 2009 (meta-analysis of fabrication and falsification in research): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2685008/
- Steneck 2006 (technical definitions of questionable research practices): https://link.springer.com/article/10.1007/PL00022268
- Bouter 2016 (survey about questionable research practices): https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-016-0024-5
- Ioannidis 2005 (why most published research findings are false): https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124
- Silberzahn 2018 (many ways to analyze red card dataset): https://journals.sagepub.com/doi/full/10.1177/2515245917747646

## Blog posts and popular articles

- Blog post summarizing issues in misleading statistics: https://www.datapine.com/blog/misleading-statistics-and-data/
- Article in *Nature* about Jonathan Pruitt: https://www.nature.com/articles/d41586-020-00287-y
- Article in *Science* on whistleblowing on statistical misconduct in ocean acidification studies: https://www.science.org/content/article/does-ocean-acidification-alter-fish-behavior-fraud-allegations-create-sea-doubt
- Article on BuzzFeed about research misconduct in Brian Wansink's food behavior lab: https://www.buzzfeednews.com/article/stephaniemlee/brian-wansink-cornell-p-hacking
- Andrew Gelman's blog with lots and lots of discussion of statistical misconduct: http://andrewgelman.org

## Other resources

- Most recent "ethical guidelines for statistical practice": https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice
- Reading list for a semester-long course on ethics in biostatistics: https://biostatistics.wustl.edu/wp-content/uploads/2019/01/M21-512-Ethics-in-Biostatistics.pdf
- Reading list about eugenics in statistics: https://www.statstree.org/ethics