

# The crisis in health and medical statistics

Adrian Barnett  
@aidybarnett

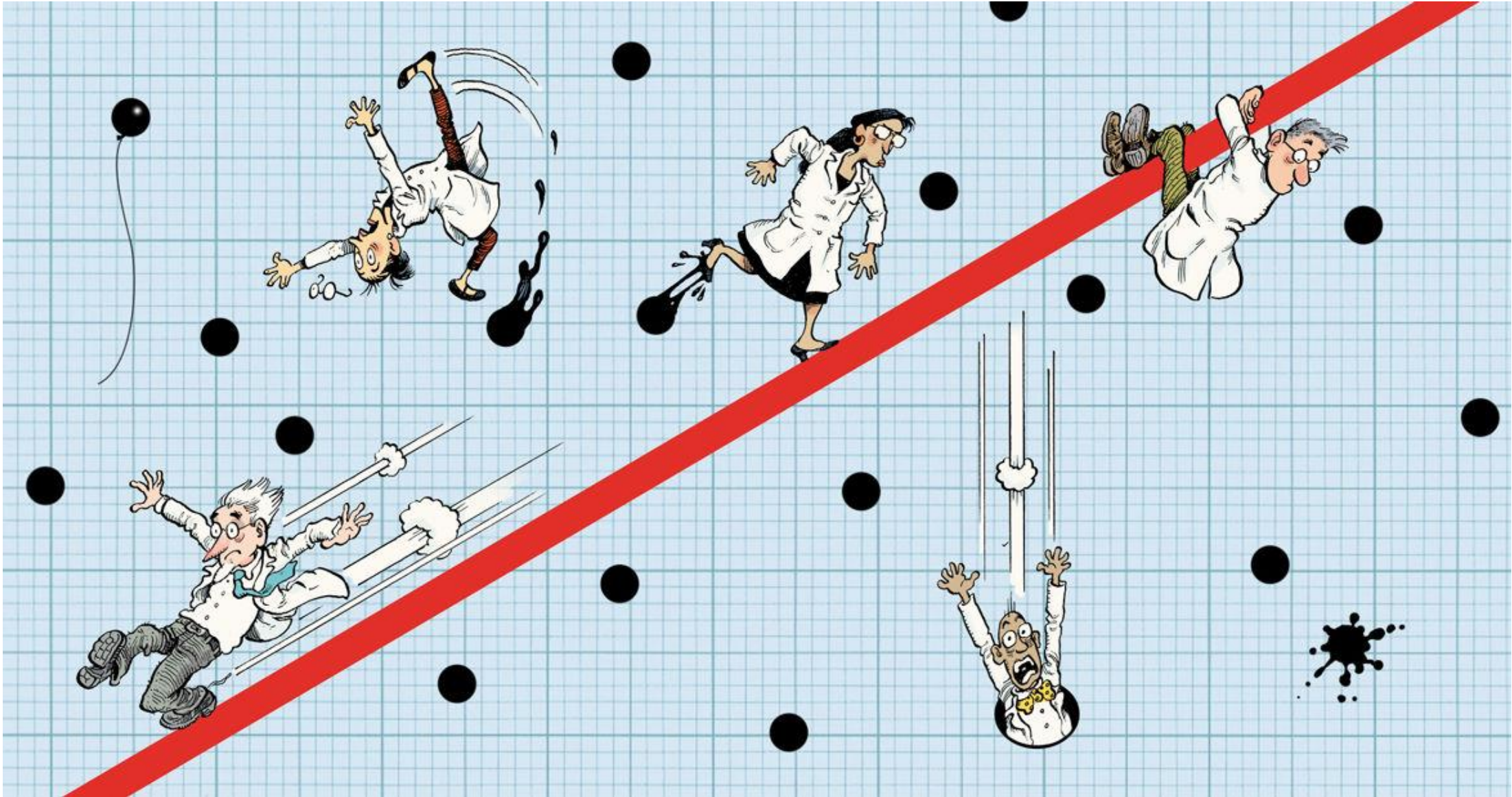


Illustration by David Parks, Nature, 2017



Illustration by Dale Edwin Murray, *Nature*, 2017

“The first principle is that you must not fool yourself — and you are the easiest person to fool”  
Richard Feynman,  
**1964**

Regina Nuzzo “How scientists fool themselves – and how they can stop” [Nature](#) 2015



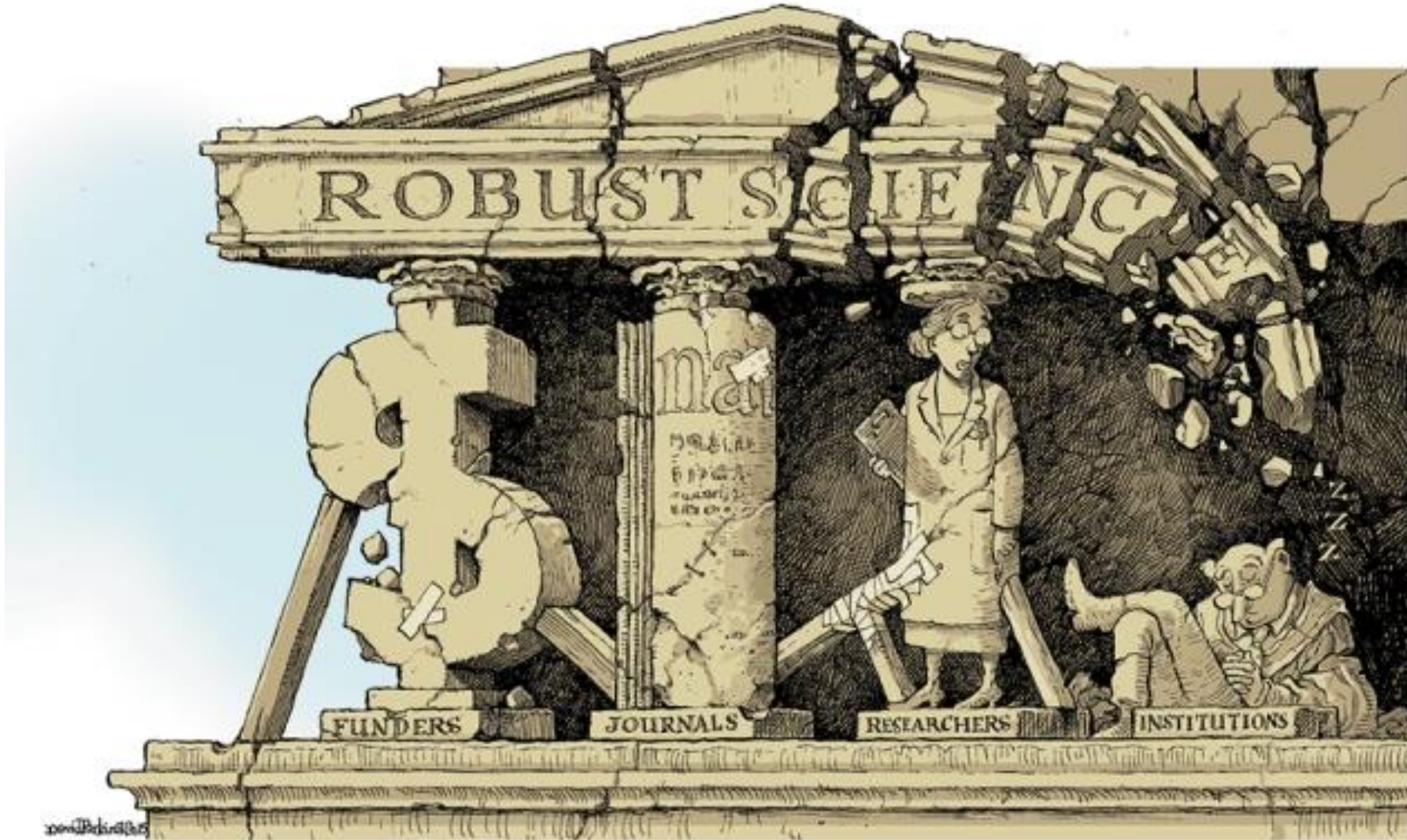
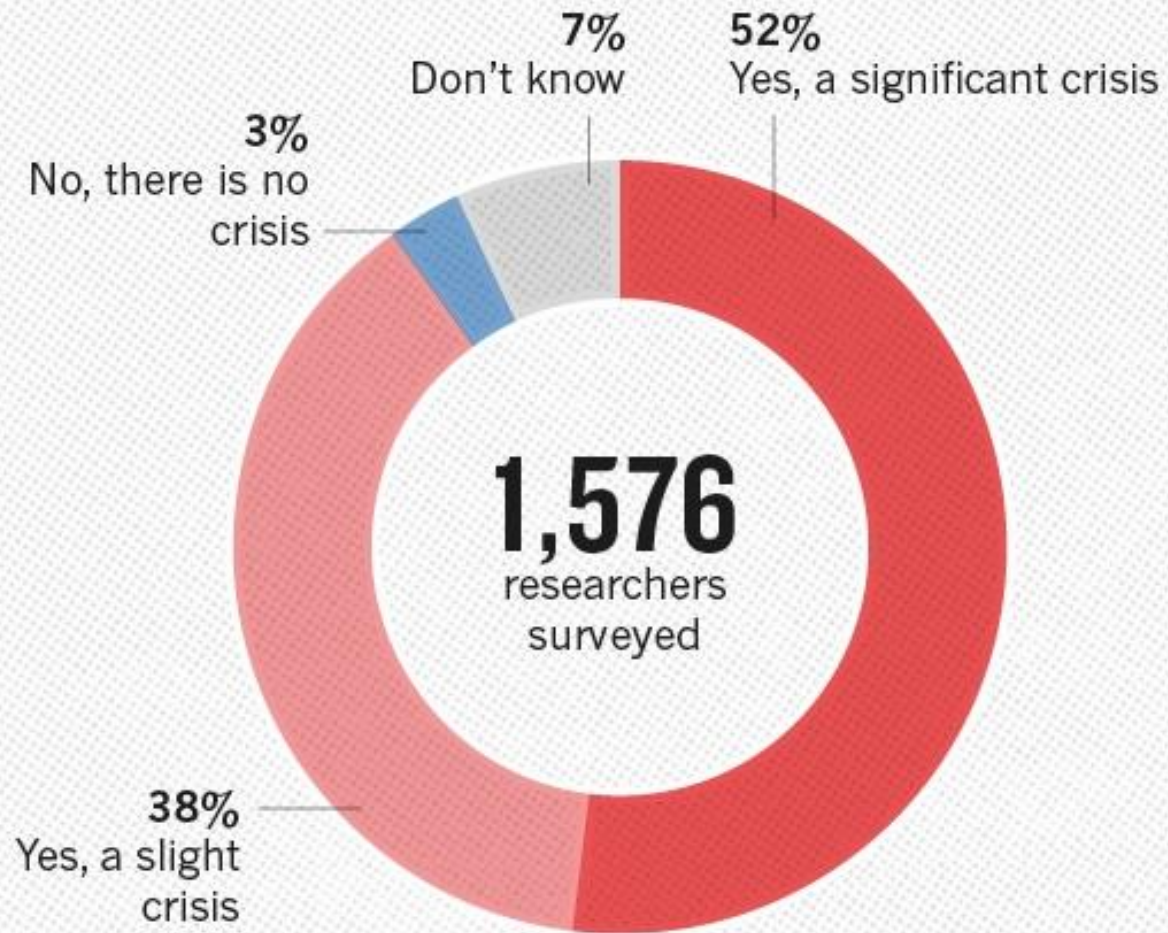


Image source: David Parkins, Nature, 2015

Begley et al “Robust research: Institutions must do their part for reproducibility” Nature 2015

## IS THERE A REPRODUCIBILITY CRISIS?



©nature



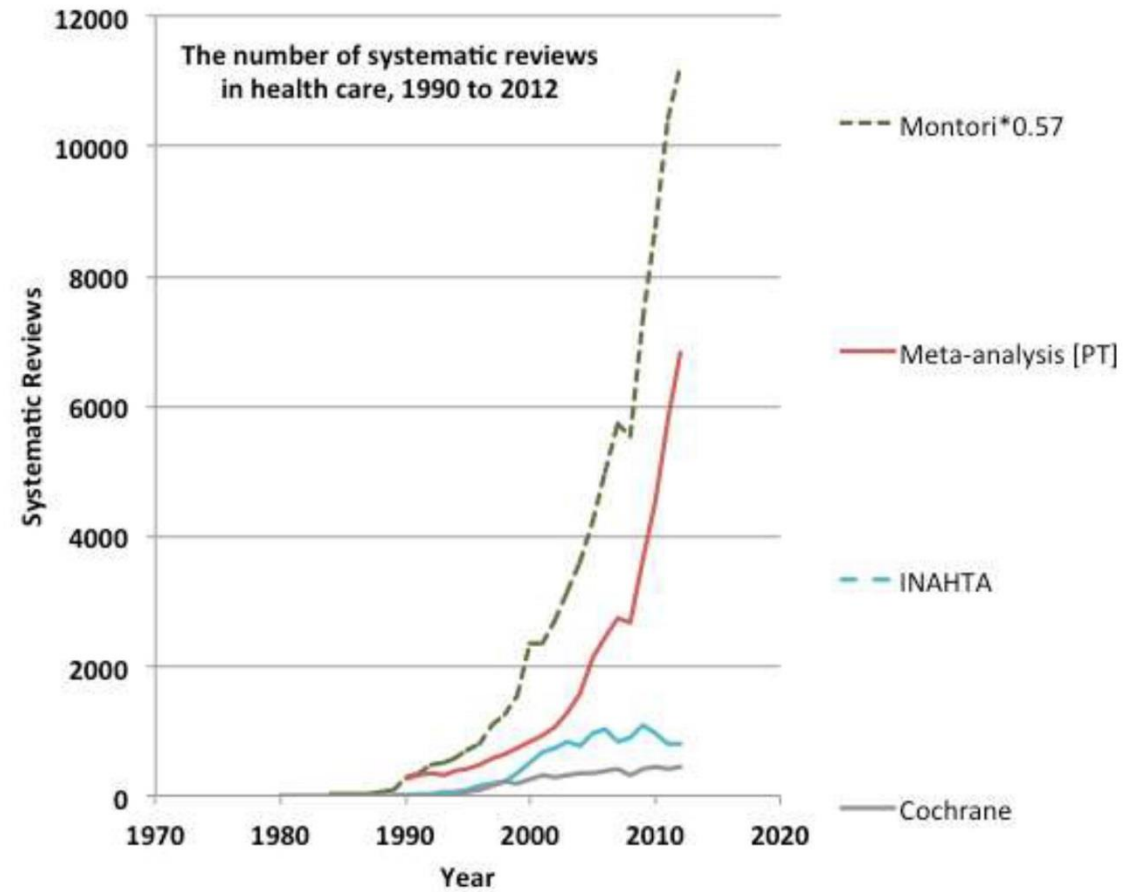
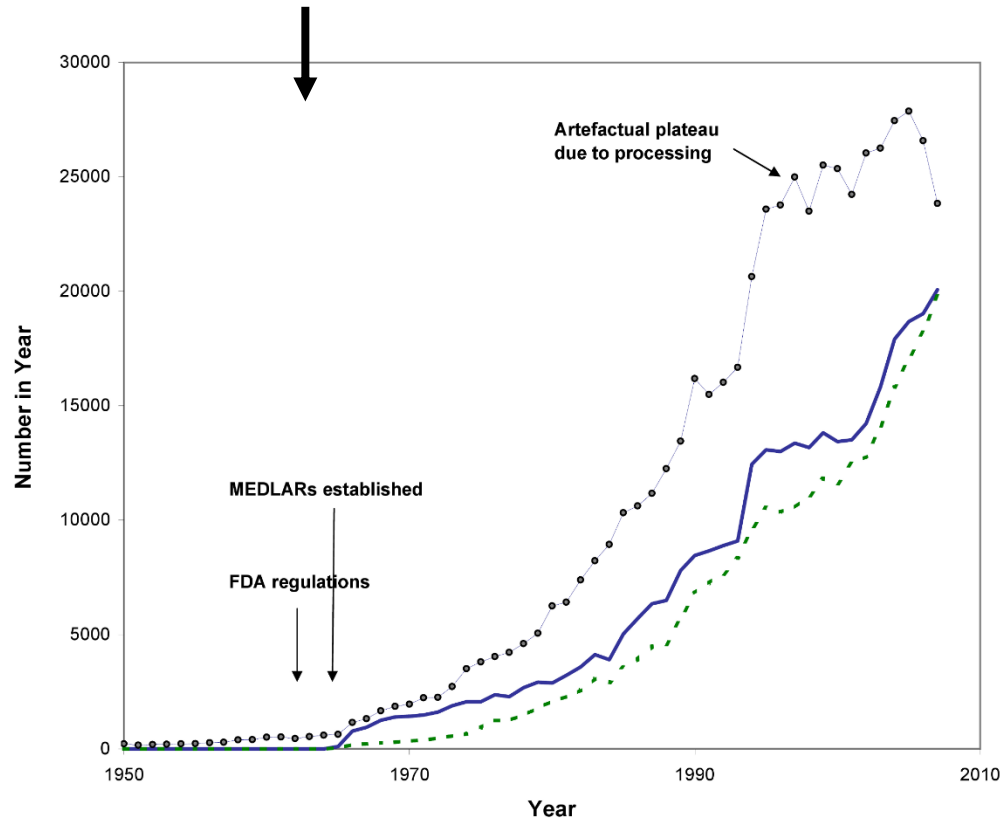
“We need less research,  
better research, and  
research done for the  
right reasons” Doug  
Altman, **1994**



Altman DG. The scandal of poor medical research. BMJ 1994;**308**(6924):283-84

Systematic reviews →

Trials ↓



Twenty-six

# Expert salami slicing, 31 papers!

- ☐ [A Survey on Mental Health Status of Adult Population Aged 15 and above in the Province of Zanjan, Iran.](#)  
1. Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, Akhondzadeh S, Armani Kian A, Nasr S.  
**Arch Iran Med.** 2017 Nov 1;20(11 Suppl. 1):S127-S130.  
PMID: 29481147
- ☐ [A Survey on Mental Health Status of Adult Population Aged 15 and above in the Province of Yazd, Iran.](#)  
2. Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, Akhondzadeh S, Yasini Ardekani SM, Farahzadi MH, Zare F.  
**Arch Iran Med.** 2017 Nov 1;20(11 Suppl. 1):S123-S126.  
PMID: 29481146
- ☐ [A Survey on Mental Health Status of Adult Population Aged 15 and above in the Province of West Azarbaijan, Iran.](#)  
3. Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, Akhondzadeh S, Sedighnia A, Karimi H.  
**Arch Iran Med.** 2017 Nov 1;20(11 Suppl. 1):S119-S122.  
PMID: 29481145
- ☐ [A Survey on Mental Health Status of Adult Population Aged 15 and above in the Province of Tehran, Iran.](#)  
4. Noorbala AA, Bagheri Yazdi SA, Faghihzadeh S, Kamali K, Faghihzadeh E, Hajebi A, Akhondzadeh S, Abbasinejad M, Zarkesh A, Amirloo F, Ghafarzadeh M.  
**Arch Iran Med.** 2017 Nov 1;20(11 Suppl. 1):S115-S118.  
PMID: 29481144



Image source: Kelvin Ma, TuftsNow, 2011

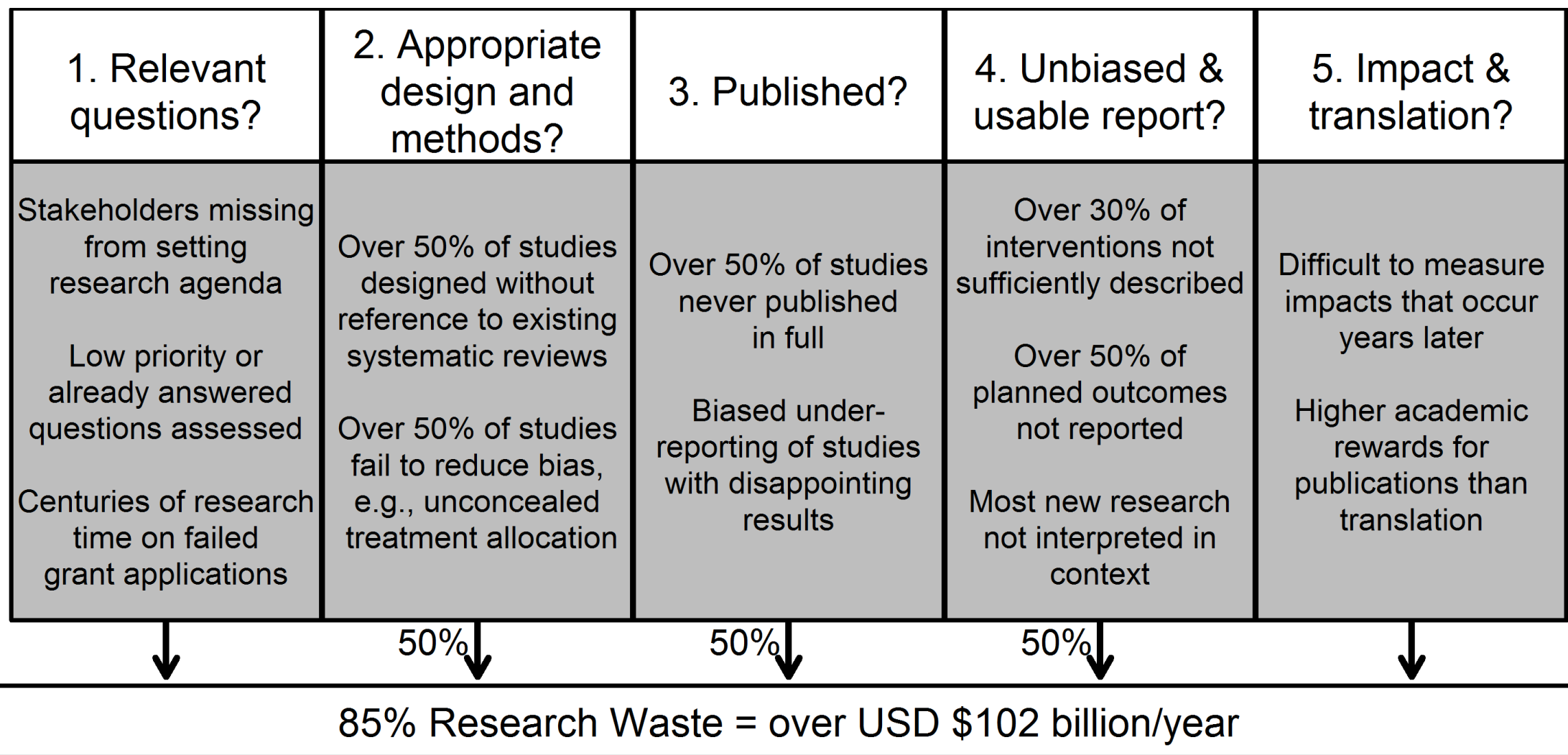
“Most published scientific findings are false” John Ioannidis 2005



# Research waste

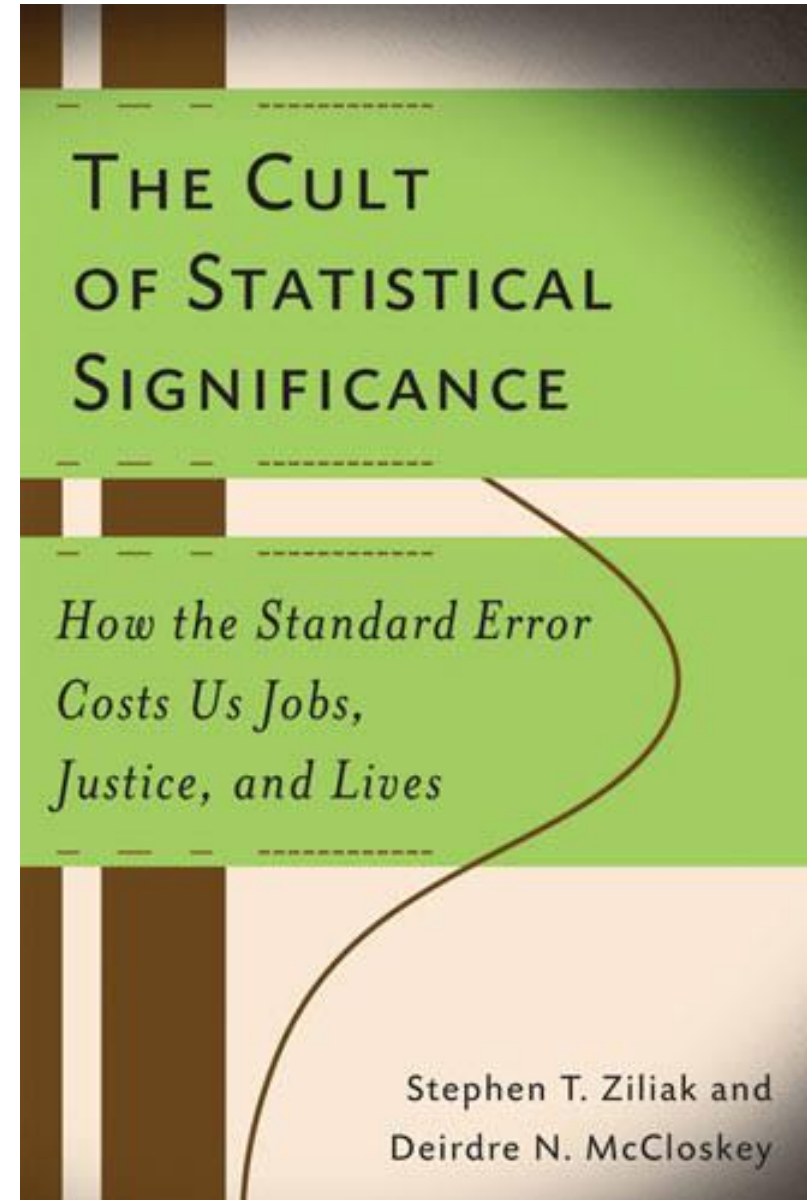


“It pays to be first. It doesn’t necessarily pay to be right.”  
Veronique Kiermer,  
Executive editor PLOS



# P-values

- “Sizeless scientists” Ziliak and McCloskey
- “The confusing fog of statistical significance testing” Ken Rothman





- All possible interactions of 6 predictors: opioid dose and duration, height, tobacco consumption, systolic blood pressure and weight
- 20 three-way, 15 four-way and 6 five-way interactions

- All possible interactions of 6 predictors: opioid dose and duration, height, tobacco consumption, systolic blood pressure and weight
- 20 three-way, 15 four-way and 6 five-way interactions
- “significant interactions including opioid dose–duration terms are found in all patients (p=0.0261) [...] while in women this interaction approaches significance (p=0.0516).”

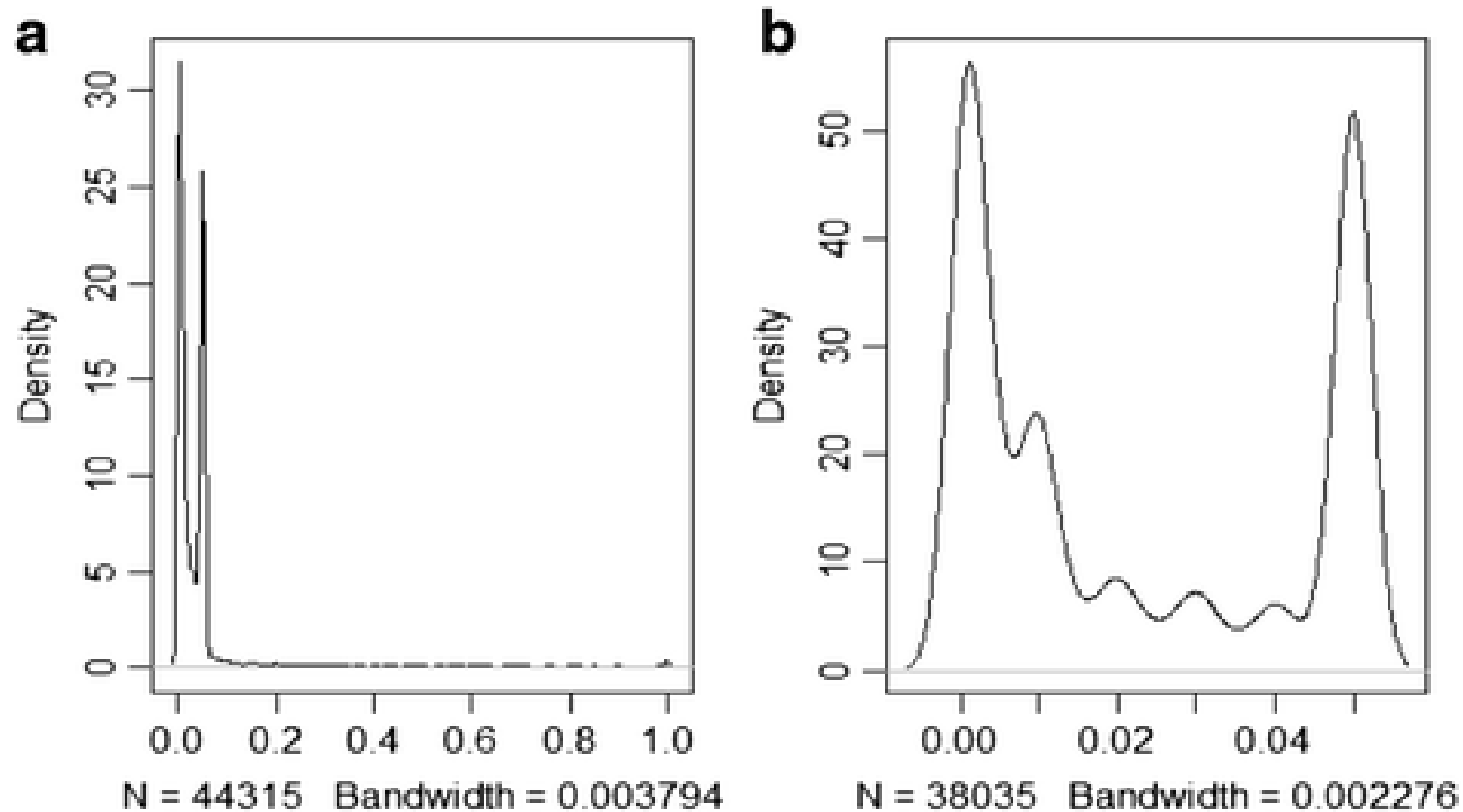
Cross-sectional adjusted regression models				
Group	Variable statistics			
Parameter	Estimate	Upper limit	Lower limit	Pr(> t )
<i>Both sexes</i>				
H.Dura.n: Cigs.d: SP : Weight	0.00023	0.0003	0.0001	0.0001
H.Dura.n: Height: Cigs.d: SP	0.00030	0.0005	0.0001	0.0047
H.Dura.n: Height: Cigs.d	−0.00143	−0.0004	−0.0024	0.0048
H.Dura.n: Height: Cigs.d: SP : Weight	−0.00007	0.0000	−0.0001	0.0051
H.Dura.n: Height: Cigs.d: Weight	0.00032	0.0006	0.0001	0.0063
H.Dose: H.Dura.n: Height: Cigs.d: Weight	−0.00037	−0.0001	−0.0007	0.0261
H.Dose: H.Dura.n: Height: Cigs.d: SP : Weight	0.00008	0.0001	0.0000	0.0262
H.Dose: H.Dura.n: Height: Cigs.d	0.00155	0.0029	0.0002	0.0286
H.Dose: H.Dura.n: Height: Cigs.d: SP	−0.00032	0.0000	−0.0006	0.0290

Image from xkcd  
<https://xkcd.com/1478/>

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001 0.01 0.02 0.03 0.04 0.049	HIGHLY SIGNIFICANT
0.050	
0.051 0.06 0.07 0.08 0.09 0.099	SIGNIFICANT
0.1	
0.101 0.102 0.103 0.104 0.105 0.106 0.107 0.108 0.109 0.11 0.111 0.112 0.113 0.114 0.115 0.116 0.117 0.118 0.119 0.12 0.121 0.122 0.123 0.124 0.125 0.126 0.127 0.128 0.129 0.13 0.131 0.132 0.133 0.134 0.135 0.136 0.137 0.138 0.139 0.14 0.141 0.142 0.143 0.144 0.145 0.146 0.147 0.148 0.149 0.15 0.151 0.152 0.153 0.154 0.155 0.156 0.157 0.158 0.159 0.16 0.161 0.162 0.163 0.164 0.165 0.166 0.167 0.168 0.169 0.17 0.171 0.172 0.173 0.174 0.175 0.176 0.177 0.178 0.179 0.18 0.181 0.182 0.183 0.184 0.185 0.186 0.187 0.188 0.189 0.19 0.191 0.192 0.193 0.194 0.195 0.196 0.197 0.198 0.199	OH CRAP. REDO CALCULATIONS.
0.2	ON THE EDGE OF SIGNIFICANCE
0.201 0.202 0.203 0.204 0.205 0.206 0.207 0.208 0.209 0.21 0.211 0.212 0.213 0.214 0.215 0.216 0.217 0.218 0.219 0.22 0.221 0.222 0.223 0.224 0.225 0.226 0.227 0.228 0.229 0.23 0.231 0.232 0.233 0.234 0.235 0.236 0.237 0.238 0.239 0.24 0.241 0.242 0.243 0.244 0.245 0.246 0.247 0.248 0.249 0.25 0.251 0.252 0.253 0.254 0.255 0.256 0.257 0.258 0.259 0.26 0.261 0.262 0.263 0.264 0.265 0.266 0.267 0.268 0.269 0.27 0.271 0.272 0.273 0.274 0.275 0.276 0.277 0.278 0.279 0.28 0.281 0.282 0.283 0.284 0.285 0.286 0.287 0.288 0.289 0.29 0.291 0.292 0.293 0.294 0.295 0.296 0.297 0.298 0.299	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.3	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS



# P-values from dentistry



Edwin Kagereki, Joseph Gakonyo and Hazel Simila. "Significance bias: an empirical evaluation of the oral health literature". *BMC Oral Health* 2016 **16**:53



**Adrian Barnett**

@aidybarnett



It's a sign of how bad things have got that researchers think it's acceptable to write this in a Nature journal: "we continuously increased the number of animals until statistical significance was reached to support our conclusions." [nature.com/articles/s41466...](https://www.nature.com/articles/s41466-018-0414-6)

9:16 AM - 3 Sep 2018

**989** Retweets **2,190** Likes



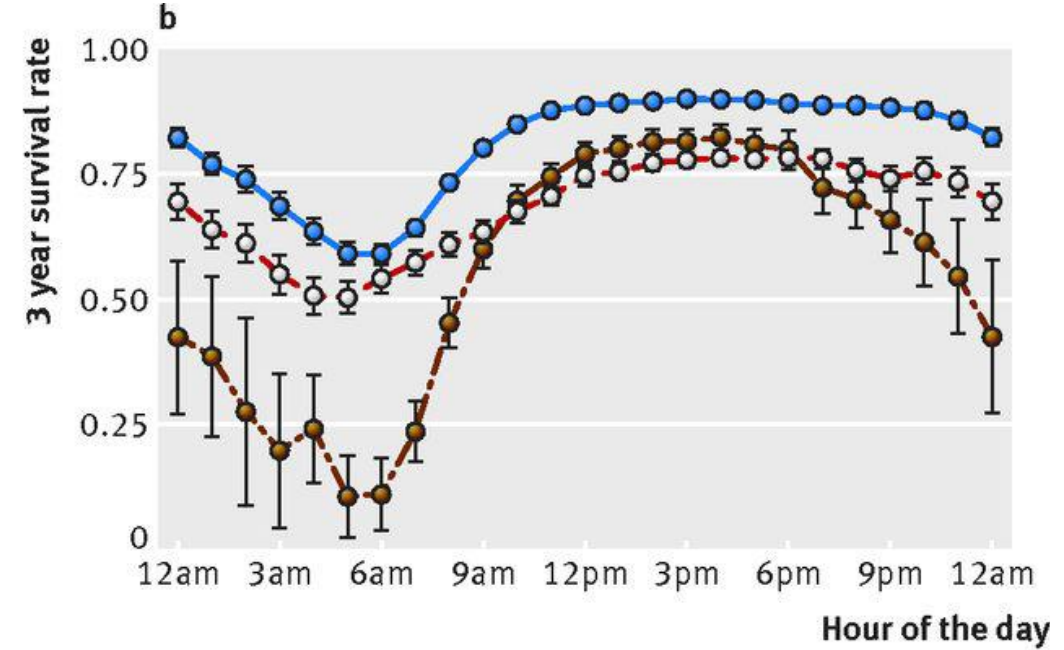
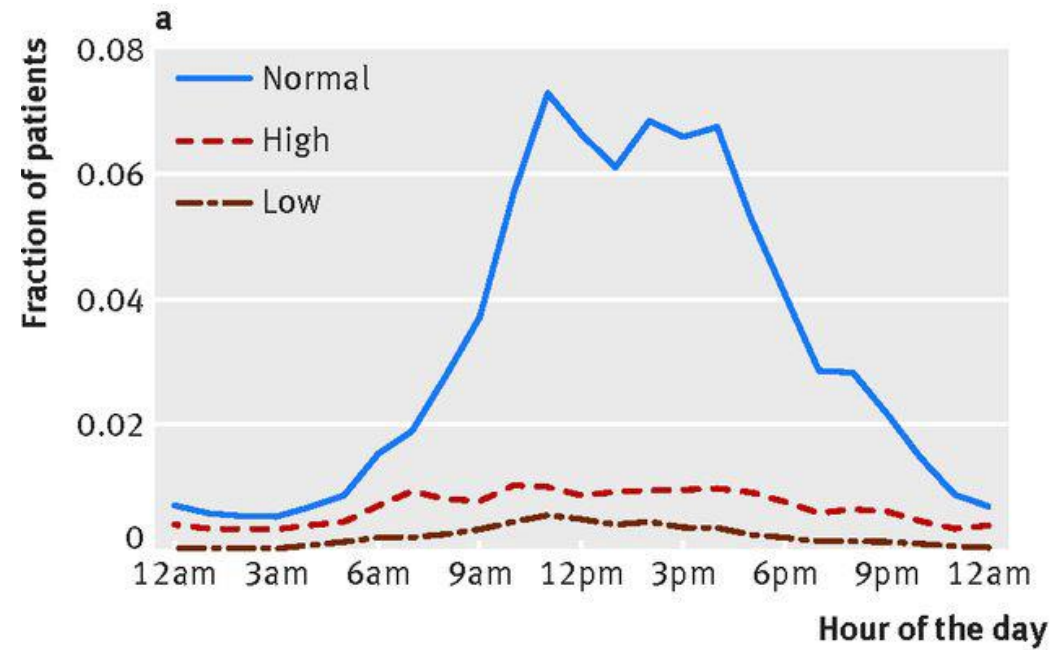
# The law courts are no better

- P-values and statistical significance in US Courts of Appeal cases from 2007 to 2017
- Only 1 in 5 interpretations were correct
- Typical quote: “However, these results were not statistically significant, so AIRE does not teach anything about the correlation between ramipril and stroke risk”
- <https://osf.io/xh zr8/>





Image: Wikimedia Commons



*BMJ* 2018;361:k1479 “Biases in electronic health record data due to processes within the healthcare system: retrospective observational study”

What can we do?

# Whole heap of other problems

Conflicts of interest

Terrible graphs

Programming errors

Data quality

Data reading errors (Excel!)

Multiple comparisons

SD vs SEM

Data dredging

Poorly described methods

Outcome switching

Stepwise selection

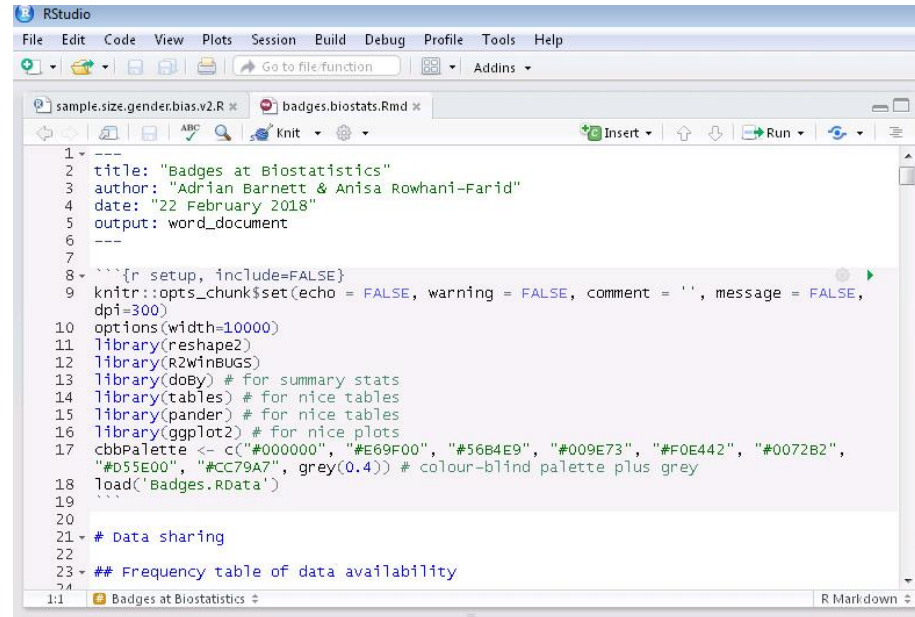
Competing risks

Sample size calculations

Poor image of statistics



# Reproducible research



```
1 ---
2 title: "Badges at Biostatistics"
3 author: "Adrian Barnett & Anisa Rowhani-Farid"
4 date: "22 February 2018"
5 output: word_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = FALSE, warning = FALSE, comment = '', message = FALSE,
10 dpi=300)
11 options(width=10000)
12 library(reshape2)
13 library(R2winBUGS)
14 library(doBy) # for summary stats
15 library(tables) # for nice tables
16 library(pander) # for nice tables
17 library(ggplot2) # for nice plots
18 cbbPalette <- c("#000000", "#56B4E9", "#009E73", "#F0E442", "#0072B2",
19 "#D55E00", "#CC79A7", grey(0.4)) # colour-blind palette plus grey
20 load('Badges.RData')
21
22 # Data sharing
23 ## Frequency table of data availability
```

## Badges at Biostatistics

Adrian Barnett & Anisa Rowhani-Farid

22 February 2018

### Data sharing

#### Frequency table of data availability

	Journal Biostatistics		Statistics in Medicine	
Data available	n	Percent	n	Percent
available	20	8.3	2	0.8
none to share	8	3.3	31	12.9
not available	209	87.1	203	84.6
potentially available	3	1.2	4	1.7
All	240	100.0	240	100.0

- Use *Rmarkdown*
- Share data and code

## Badges for sharing data and code at Biostatistics - an observational study

Version 2 ▾ 28.02.2018, 23:56 by Anisa Rowhani-Farid, Adrian Barnett

This fileset contains the anonymised data and code for the study: Badges for sharing data and code at *Biostatistics* - an observational study. This article is currently under peer-review at F1000Research. It has been accepted as a pre-publication.

The data is saved in rdata format and the code was written in Rmarkdown. WinBUGS (version 1.3.4) was also used in the analysis.

60 views | 9 downloads | 1 citations

#### CATEGORIES

• Science Policy

#### KEYWORD(S)

reproducibility studies | meta-research | incentives

# Use a research checklist



Enhancing the **QUALity** and  
**Transparency Of health Research**



EQUATOR resources in  
[Portuguese](#) | [Spanish](#)

[Home](#)[Library](#)[Toolkits](#)[Courses & events](#)[News](#)[Blog](#)[Librarian Network](#)[About us](#)[Contact](#)

## Your one-stop-shop for writing and publishing high-impact health research

[find reporting guidelines](#) | [improve your writing](#) | [join our courses](#) | [run your own training course](#) | [enhance your peer review](#) | [implement guidelines](#)



### Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.



[Search for reporting  
guidelines](#)



[Not sure which reporting  
guideline to use?](#)



[Reporting guidelines  
under development](#)



[Visit the library for  
more resources](#)



### Reporting guidelines for main study types

[Randomised trials](#)

[Observational studies](#)

[Systematic reviews](#)

[Case reports](#)

[Qualitative research](#)

[Diagnostic / prognostic  
studies](#)

[Quality improvement studies](#)

[Economic evaluations](#)

[Animal pre-clinical studies](#)

[Study protocols](#)

[Clinical practice guidelines](#)

[CONSORT](#)

[Extensions](#)

[Other](#)

[STROBE](#)

[Extensions](#)

[Other](#)

[PRISMA](#)

[Extensions](#)

[Other](#)

[CARE](#)

[Extensions](#)

[Other](#)

[SRQR](#)

[COREQ](#)

[Other](#)

[STARD](#)

[TRIPOD](#)

[Other](#)

[SQUIRE](#)

[Other](#)

[CHEERS](#)

[Other](#)

[ARRIVE](#)

[Other](#)

[SPIRIT](#)

[PRISMA-P](#)

[Other](#)

[AGREE](#)

[RIGHT](#)

[Other](#)

[See all 398 reporting guidelines](#)

### EQUATOR Oncology



Resources supporting  
complete, accurate and  
transparent research

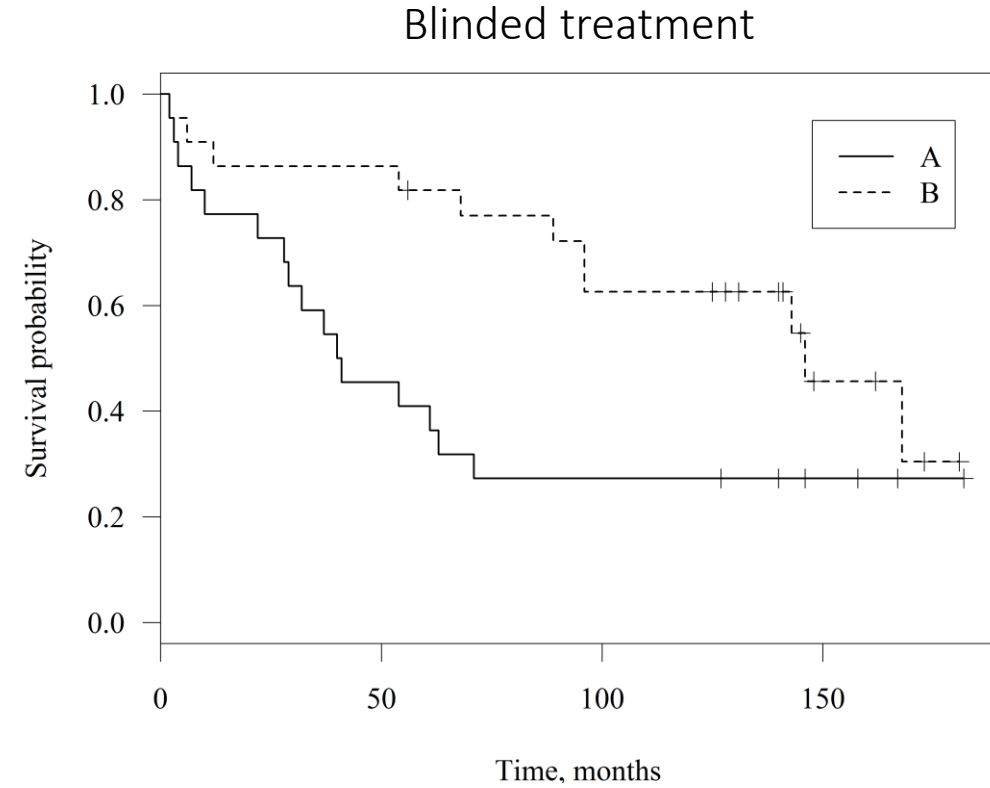
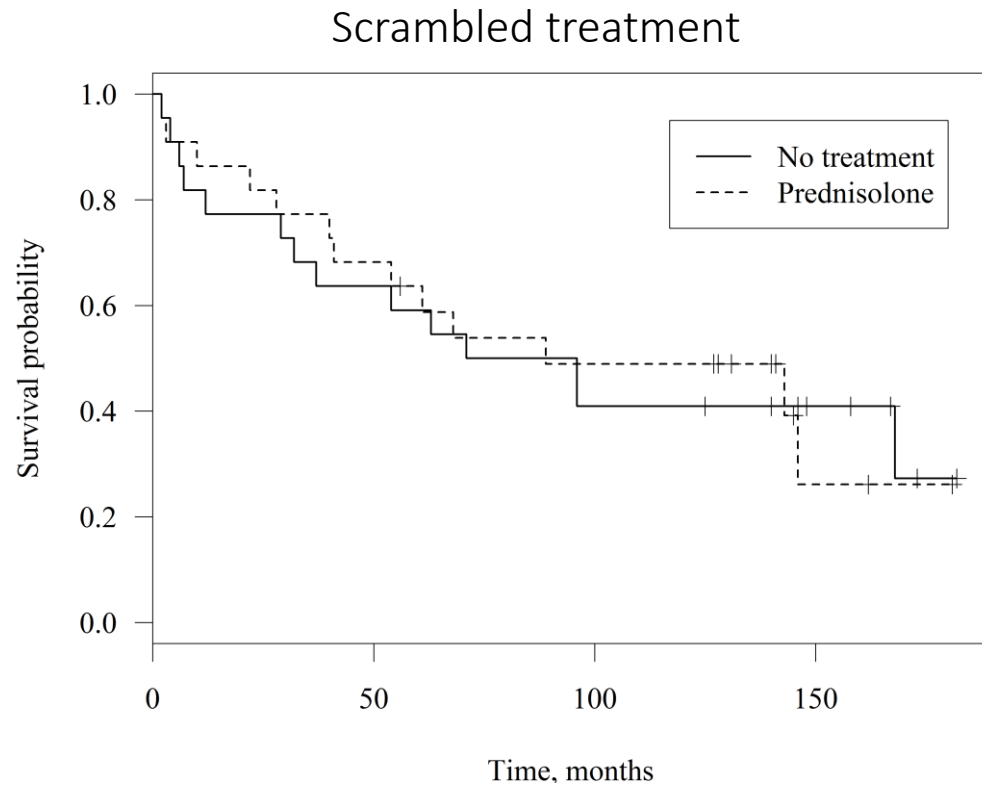


[Visit EQUATOR Oncology](#)

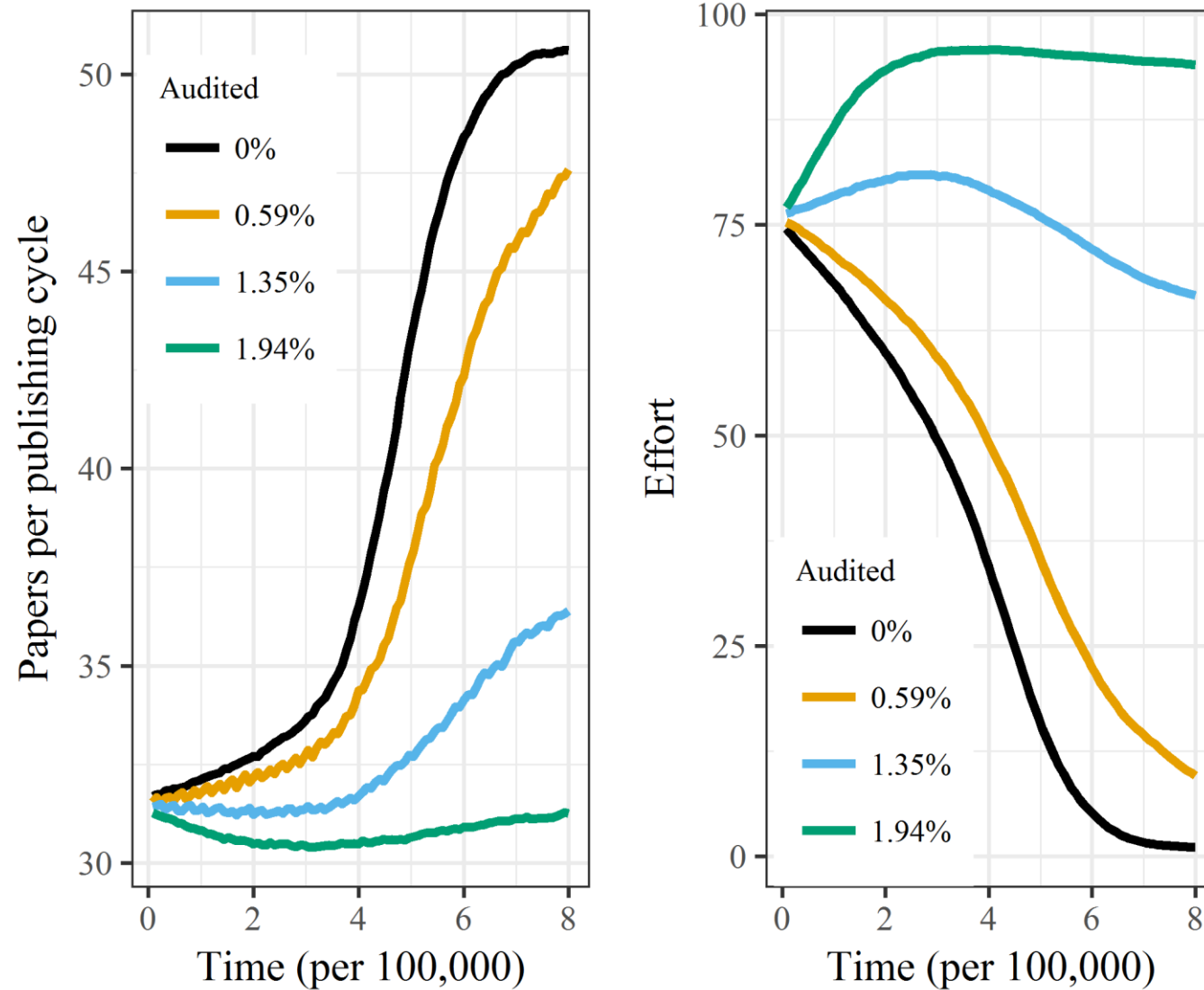
Write protocols and  
analysis plans



# Scrambling and blinding

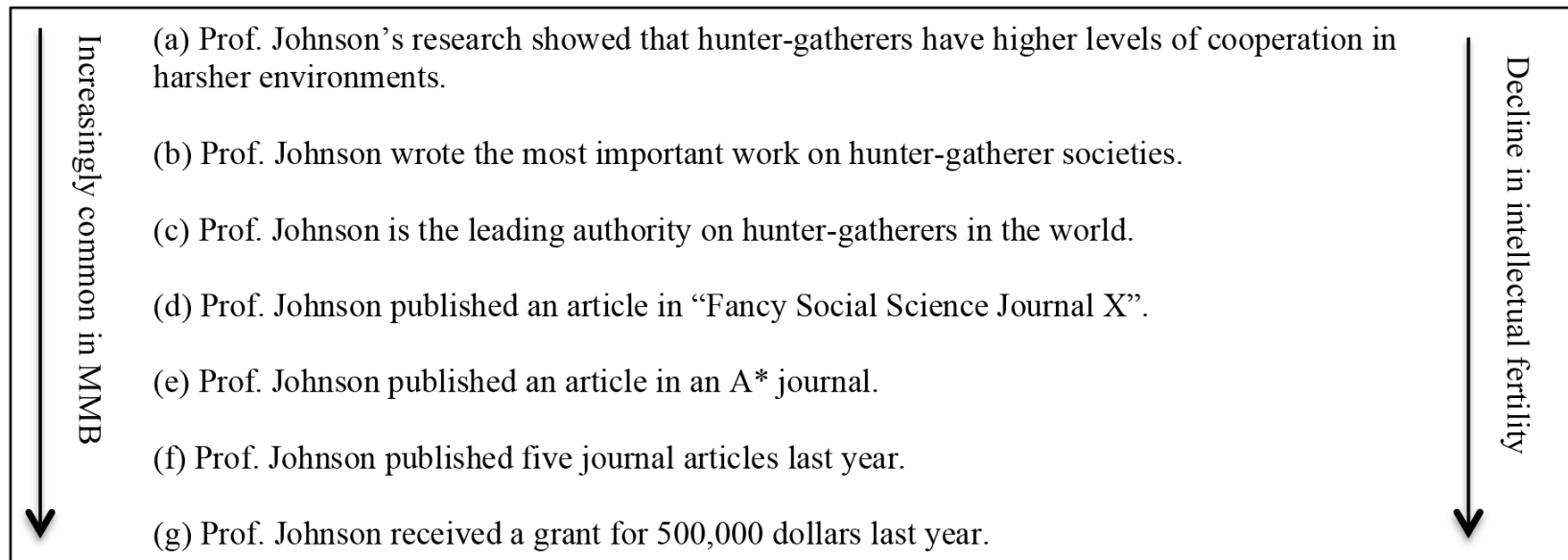


# Random audits of researchers



# Scrap league tables and the competitive culture

“no metric should be taken as a substitute for the actual reading of a paper in determining its quality.”



*Figure 10: Types of descriptions used to attest to the quality of another's work*

# HEALTH STATISTICS REFRESHER

**MONDAY 12 NOVEMBER 2018**

## About the course

This short course is a one-day statistics refresher for people who have had some practical experience with statistics or some undergraduate statistics training. We will discuss simple study designs and some common mistakes, with practical exercises throughout the day.

The course does not use software (e.g. SPSS) as it will focus on the choice of what statistics to use and interpretation of the output, rather than on the practicalities of running analyses.

### Target audience:

Healthcare professionals (all levels of experience and background are welcome), policy makers, academics and students.

## Course Outline

The course will cover the following topics:

- sample size
- statistical significance ( $p < 0.05$ ) and confidence intervals
- non-parametric versus parametric tests
- the Normal assumption
- the basics of causal diagrams
- randomised controlled trials
- intention-to-treat
- pre-post designs (before and after studies)

### ***BYO Laptop***

### *Instructor:*

**Professor Adrian Barnett, QUT**

