McMaster University
HEALTH RESEARCH METHODS, EVIDENCE, AND IMPACT

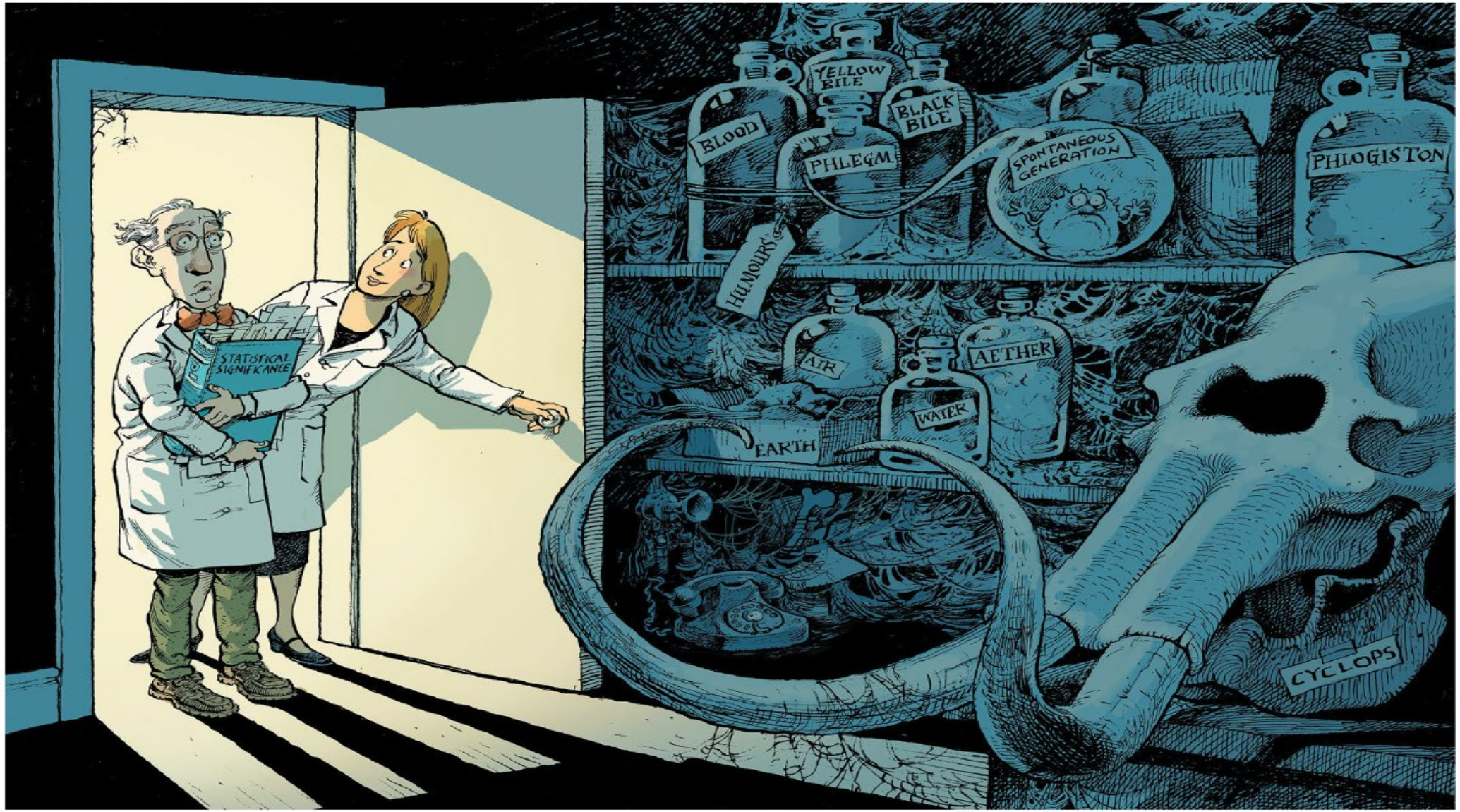
From Statistical Significance to Clinical Relevance:

Time to dethrone or get rid of "significance"?

Lehana Thabane

St. Joseph's Healthcare Hamilton

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

What I plan to share with you

1. Connection between "statistical thinking" and EBP
2. Some history of statistical significance (p-value)
3. Some practical challenges with "statistical significance" or "p-value thresholds"
4. On-going debate on what to do with "statistical significance"
5. Suggestions to put "significance" on "clinical relevance"
6. Some concluding remarks: Why should you care?

Hebert George (HG) Wells

Father of Science

Who was HG Wells?

Born	Herbert George Wells 21 September 1866 in Kent, England
Died	13 August 1946 (aged 79) London, England
Occupation	Novelist, teacher, historian, journalist
Alma mater	Royal College of Science (Imperial College London)
Subjects	World history, progress
Notable work(s)	✓ *The Outline of History* ✓ *The Time Machine* ✓ *The Island of Doctor Moreau* ✓ *The War of the Worlds* ✓ *The First Men in the Moon* ✓ *The Shape of Things to Come*

HG Wells' prediction

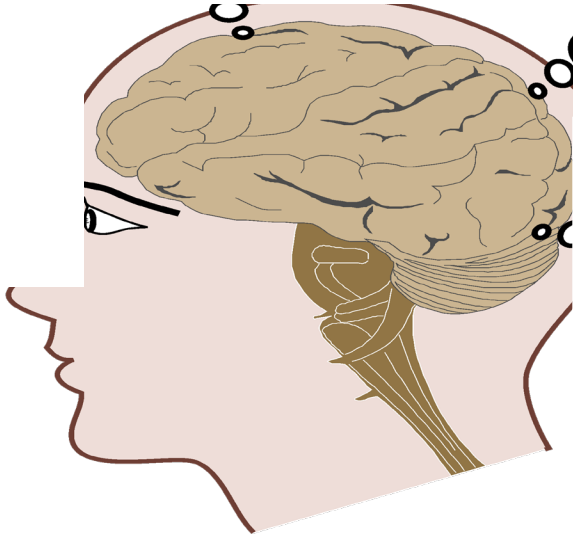
"Statistical thinking will some day be as necessary for efficient citizenship as the ability to read and write"

Statistical thinking

DESCRIPTIVE

I NFERENTIAL

CONTEXTUAL



Statistical thinking

=

Evidence-based thinking

We cannot fully appreciate the **clinical relevance** of evidence unless we adopt **statistical thinking** in how we collect, synthesize, and interpret evidence

First:

We need pragmatic trial evidence

English *et al. BMC Medicine* (2016) 14:5
DOI 10.1186/s12916-016-0556-z

BMC Medicine

Medicine for Global Health

DEBATE Open Access

The need for pragmatic clinical trials in low and middle income settings – taking essential neonatal interventions delivered as part of inpatient care as an illustrative example

Mike English[1,2*], Jamlick Karumbi[1,3], Michuki Maina[1], Jalemba Aluvaala[1,4], Archna Gupta[5], Merrick Zwarenstein[5] and Newton Opiyo[1]

Many interventions are introduced in practice without any evidence of their effectiveness in routine care settings

Second:

We need address the barriers of translating evidence into practice

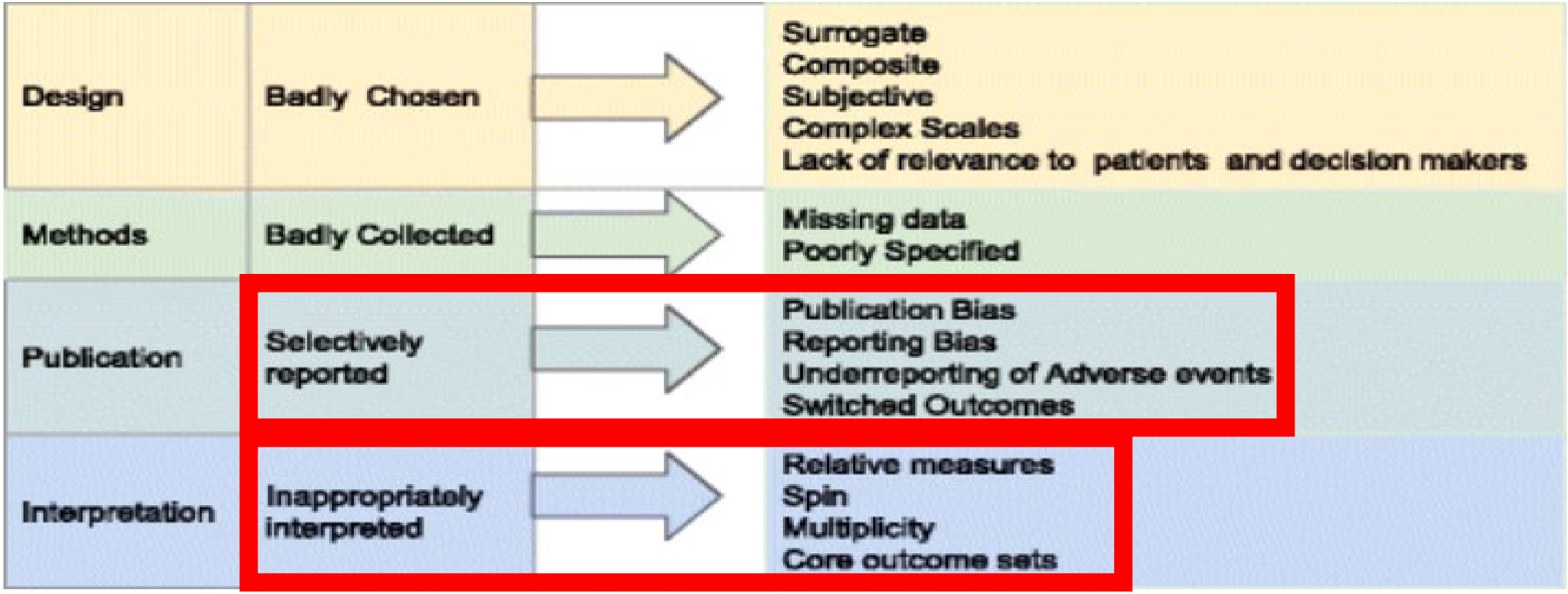
Heneghan et al. Trials (2017) 18:122
DOI 10.1186/s13063-017-1870-2

Trials

COMMENTARY Open Access

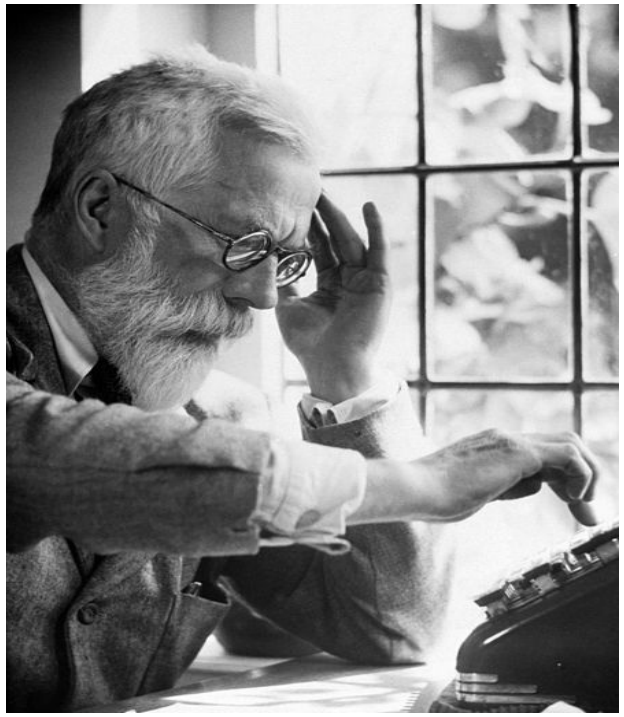
Why clinical trial outcomes fail to translate into benefits for patients

Carl Heneghan*, Ben Goldacre and Kamal R. Mahtani



The beginning of p-values and statistical significance!

RA Fisher

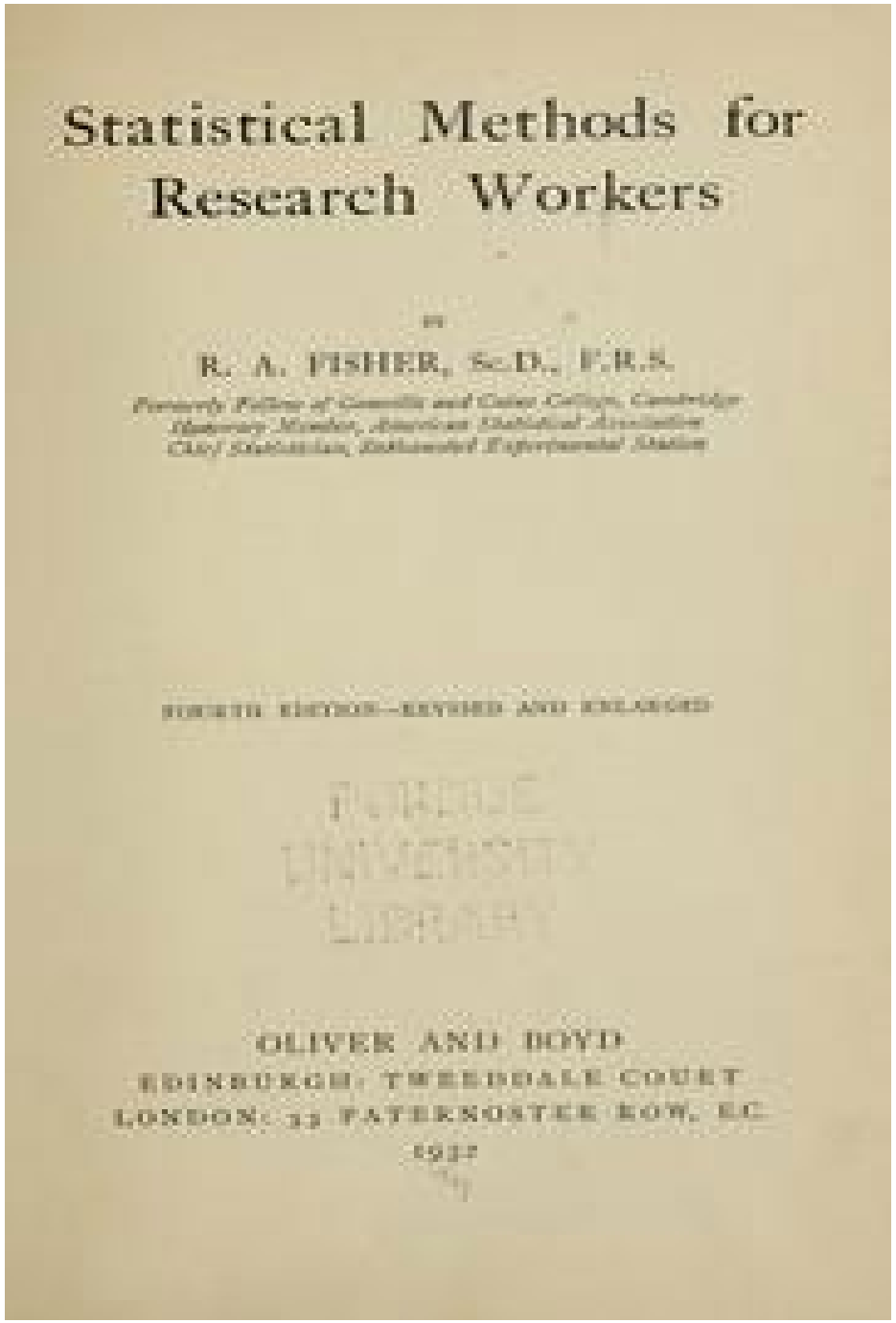


"The Father of Modern Statistics

Often described as "a genius who almost single-handedly created the foundations for modern statistical science"

Who was RA Fisher?

Born	February 17, 1890, East Finchley, London, United Kingdom
Died	July 29, 1962, Adelaide, Australia
Occupation	Statistician and geneticist
Alma mater	Gonville & Caius College, University of Cambridge, Harrow School
Subjects	World history, progress
Influenced by	✓ Charles Darwin, ✓ Karl Pearson, ✓ William Sealy Gosset


Statistical Methods for Research Workers

BY

R. A. FISHER, Sc.D., F.R.S.

Formerly Fellow of Gonville and Caius College, Cambridge
Honorary Member, American Statistical Association
Chief Statistician, Rothamsted Experimental Station

FOURTH EDITION—REVISED AND ENLARGED

OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 33 PATERNOSTER ROW, E.C.
1932

"The value for which P=0.05, or 1 in 20, is 1.96 or nearly 2 ; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not."

BMJ 2001;322:226–31

Sifting the evidence–what's wrong with significance tests?

Jonathan A C Sterne, George Davey Smith

Fisher advocated *"… P < 0.05 (5% significance) as a standard level for concluding that there is evidence against the hypothesis tested, though not as an absolute rule"*

He proposed *"if P is between 0.1 and 0.9 there is certainly no reason to suspect the hypothesis tested. If it's below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at 0.05"*

Alpha = 0.05 was then introduced in statistical science as a criterion for judging "statistical significance"

And this has pretty much become ritualistic in medical research!

"…the accept/reject philosophy of significance testing based on the "magical" p=0.05 barrier ***remains*** dominant in the minds of many non-statisticians."

(Stuart Pocock. BMJ. 1985)

I suggest that the blame goes to both statisticians and non-statisticians alike

Non-statisticians are taught by and collaborate with statisticians

So, what are the practical challenges with "statistical significance" or "p-value thresholds"?

① An overwhelming number (over 90%) of published articles report *P* values of 0.05 or less

(*JAMA* 2016;315(11):1141-8)

However, many of the claims that these reports highlight are likely false

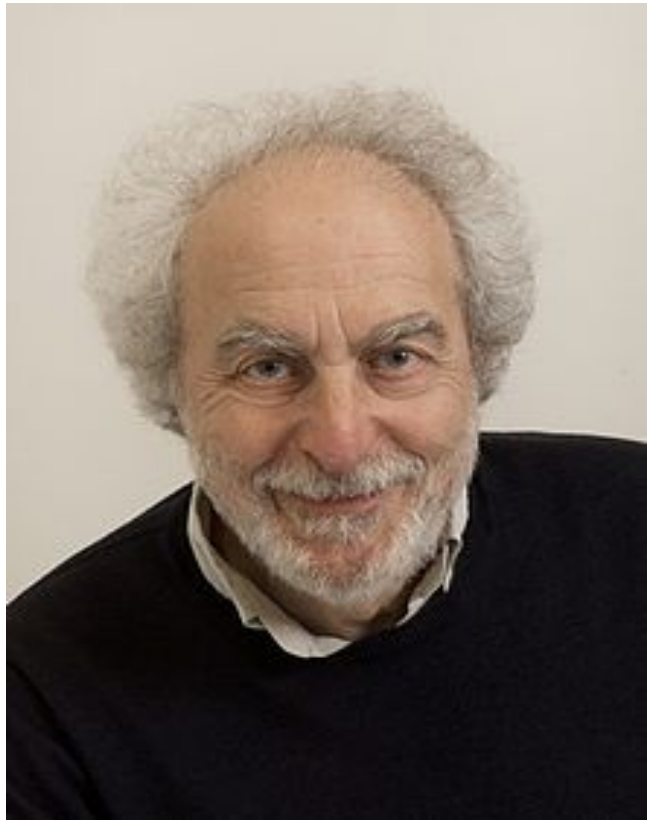
Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Med. 2005;2(8):e124.

② Lack of statistical significance is often interpreted as absence of evidence



"Absence of evidence is not evidence of absence"

(BMJ 1995;331;445)

Why?

Because underpowered studies will often yield results that are not statistically significant

Am Heart J. 2015 Jul;170(1):110-6. doi: 10.1016/j.ahj.2015.04.019. Epub 2015 Apr 18.

Effects of withdrawing vs continuing renin-angiotensin blockers on incidence of acute kidney injury in patients with renal insufficiency undergoing cardiac catheterization: Results from the Angiotensin Converting Enzyme Inhibitor/Angiotensin Receptor Blocker and Contrast Induced Nephropathy in Patients Receiving Cardiac Catheterization (CAPTAIN) trial.

Bainey KR[1], Rahim S[2], Etherington K[2], Rokoss ML[2], Natarajan MK[2], Velianou JL[2], Brons S[2], Mehta SR[3]; CAPTAIN Investigators.

Outcomes of the CAPTAIN trial—a study into withholding ACEi/ARB therapy during coronary angiography in order to prevent CIN in patients with renal insufficiency

Outcome	Discontinued ACEi/ARB therapy (*n* = 106)	Continued ACEi/ARB therapy (*n* = 102)	Risk ratio (95% CI, P-value)
CIN (rise in serum creatinine ≥0.5mg/dL or >25%)	10.9%	18.4%	0.59 (0.30–1.19, P = 0.16)

The P-value was calculated using Fisher's exact test.

- The results show about **41% RRR (RD: 7.5%)** in events—which may be viewed as clinically relevant

- But, indeed, the problem here is small sample sizes leading to fewer events and hence limited statistical power

Based on these results, the authors recommend that *"[withholding ACEi/ARB treatment] could be considered when referring a patient for angiography. However, larger adequately powered clinical trials are required to prove definitive benefits in reducing contrast-induced AKI"*

How common are clinically relevant or meaningful results that are **NOT** statistically significant?

Thus, having findings that are NSS, but clinically relevant is very common in practice

All reporting guidelines recommend justifying the sample size in the methods (during the design of studies) to avoid this problem

SPIRIT

STANDARD PROTOCOL ITEMS: RECOMMENDATIONS FOR INTERVENTIONAL TRIALS

SPIRIT 2013 Checklist: Recommended items to address in a clinical trial protocol and related documents*

Section/item	Item No	Description
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations

CONSORT statement

The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials

*David Moher, Kenneth F Schulz, Douglas G Altman, for the CONSORT Group**

Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement

Developing an evidence base for making public health deci- | Don C. Des Jarlais, PhD, Cynthia Lyles, PhD, Nicole Crepaz, PhD, and the TREND Group

OPEN ACCESS Freely available online PLOS MEDICINE

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies

Erik von Elm[1*], Douglas G. Altman[2], Matthias Egger[1,3], Stuart J. Pocock[4], Peter C. Gøtzsche[5], Jan P. Vandenbroucke[6] for the STROBE Initiative

Reporting Guideline	Sample size item
CONSORT	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.
TREND	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules
STROBE	Describe rationale for study size, including practical and statistical considerations.

In short, p-values from underpowered trials are difficult to interpret

Tom Fleming reminds us that "The important insight is that a p-value is interpretable only when you understand the sampling context from which it is derived"

(Ann Intern Med 2010; 153(6): 400–406)

③ Overpowered studies will always lead to statistically significant results that have no clinical relevance

The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial

Lancet 2011; 377: 2181–92

Outcomes of the SHARP trial—a study of lipid-lowering drugs in patients with CKD

Outcome	Simvastin plus ezetimibe (*n* = 4650) (%)	Placebo (*n* = 4620) (%)	Risk ratio (95% CI, P-value)
Any major atherosclerotic event	11.3	13.4	0.83 (0.74–0.94, P = 0.0021)

ARR=0.021;
NNT =48

- The results show about 17% RRR (2.1% RD) in events—which may NOT be viewed as clinically relevant,

- But, the problem here is large sample sizes leading to over-powered trial

- Based on these results, the authors concluded that "*lowering LDL cholesterol with the combination of simvastatin plus ezetimibe safely reduces the risk of major atherosclerotic events in a wide range of patients with CKD*"
- Furthermore, they suggested that "*widespread use of LDL-cholesterol-lowering therapy in patients with CKD would result in a worthwhile reduction in CVD complications*"

Relative measures can exaggerate findings of modest clinical benefit

Evid Based Med December 2016 | volume 21 | number 6 | 201

CrossMark

'Spin' in reports of clinical research

Kamal R Mahtani

Heneghan *et al. Trials* (2017) 18:122
DOI 10.1186/s13063-017-1870-2

Trials

COMMENTARY Open Access

Why clinical trial outcomes fail to translate into benefits for patients

Carl Heneghan, Ben Goldacre and Kamal R. Mahtani

Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes

Isabelle Boutron, MD, PhD; Susan Dutton, MSc; Philippe Ravaud, MD, PhD; et al

Article Information

JAMA. 2010;303(20):2058-2064. doi:10.1001/jama.2010.651

This can lead "spin" in how results are interpreted

④ Reliance on statistical significance show that trials in many areas are fragile

Journal of Clinical Epidemiology 67 (2014) 622–628

Journal of Clinical Epidemiology

The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index

Michael Walsh[a,b,c,*], Sadeesh K. Srinathan[d], Daniel F. McAuley[e,f], Marko Mrkobrada[g], Oren Levine[b], Christine Ribic[a,b], Amber O. Molnar[h], Neil D. Dattani[i], Andrew Burke[g], Gordon Guyatt[a,b], Lehana Thabane[a], Stephen D. Walter[a,b], Janice Pogue[a,c], P.J. Devereaux[a,b,c]

Trial Result		
	Event	No Event
Treatment A	a	b
Treatment B	c	d
Fisher's Exact Test $p<0.05$		

Calculated Fragility		
	Event	No Event
Treatmnt A	a+f	b-f
Treatment B	c	d
Fisher's Exact Test $p\geq0.05$		

Calculation of the fragility index

- Fragility index = The smallest value of "f" that causes the Fisher's exact P-value to meet or exceed the 0.05 level
- Higher values indicate less fragile results.

There have many reviews or survey of fragility index in different areas

BJA

British Journal of Anaesthesia, 120 (5): 935–941 (2018)

doi: 10.1016/j.bja.2018.01.012
Advance Access Publication Date: 13 February 2018
Review Article

CLINICAL PRACTICE

Th
ra
sy
G.
L. I

Fragility Index in Randomized Controlled Trials of Ischemic Stroke

Kenichiro Sato, M

Neurosurg Rev (2019) 42:9–14
DOI 10.1007/s10143-017-0870-8

CrossMark

REVIEW

All show that many trials are very fragile

RESEARCH METHODS/ORIGINAL RESEARCH

The Results of Randomized Controlled Trials in Emergency Medicine Are Frequently Fragile

Jamin Brown, DO*; Aaron Lane, DO; Craig Cooper, BS; Matt Vassar, PhD

European Heart Journal (2017) **38**, 338–345
doi:10.1093/eurheartj/ehw427

EUROPEAN SOCIETY OF CARDIOLOGY®

CLINICAL RESEARCH
Heart failure/cardiomyopathy

How robust are clinical trials in heart failure?

Kieran F. Docherty[1], Ross T. Campbell[2], Pardeep S. Jhund[2], Mark C. Petrie[1], and John J.V. McMurray[2]*

⑤ P-values are routinely misinterpreted

(*BMJ* 2003; 326:475)

- Examples of statements claiming "no effect or difference":
  - ✓"had no effect"
  - ✓ "the effectiveness [of intervention A] did not differ from that of [intervention B]"
  - ✓ "there was no difference [in outcomes]"
- Examples of *poorly worded* statements claiming "no effect or difference":
  - ✓ "appeared to have equivalent efficacy"
  - ✓ "may be as effective"
  - ✓"did not appear to be effective"
  - ✓"was found to be no more effective"
  - ✓"[the risk of outcome] was similar [between the two groups]"
  - ✓"is not associated with clear benefit"

Examples of *appropriately worded statements:*

- "there was no statistically significant effect or difference"
- "there was insufficient evidence to support or refute"

⑥ Reliance on statistical significance promotes unethical practices in the conduct of science through "p-hacking"

PLOS | BIOLOGY

PERSPECTIVE

The Extent and Consequences of P-Hacking in Science

Megan L. Head[1]*, Luke Holman[1], Rob Lanfear[1,2], Andrew T. Kahn[1], Michael D. Jennions[1]

Provides empirical evidence that p-hacking is widespread in the scientific literature across many disciplines including medical and health sciences

What is p-hacking?

P-hacking = Collect, select data or perform statistical analyses until a result that is not significant becomes significant

- Collect more data p<0.05
- Collect or analyse lots of data, but report only those with p<0.05
- Select covariates until p<0.05
- Exclude participants until p<0.05
- Transform data until p<0.05

The unfortunate consequence of p-hacking is that it makes scientific research harder to replicate or reproduce

⑦ The desire for significant findings is partly responsible research misconduct leading to **retractions due to data fabrication and falsification**

Ethics and Policy

OPEN ACCESS

ORIGINAL ARTICLE *J Med Genet* 2019;**0**:1–7.

Reasons for and time to retraction of genetics articles published between 1970 and 2018

Rafael Dal-Ré,[1] Carmen Ayuso[2,3]

Data fabrication and falsification (research misconduct) are among the top reasons for retractions

There's empirical evidence that retractions are increasing and research misconduct remains the primary driver

OPEN ACCESS Freely available online

PLOS ONE

Why Has the Number of Scientific Retractions Increased?

R. Grant Steen[1*], Arturo Casadevall[2], Ferric C. Fang[3]

1 MediCC!, Medical Communications Consultants, LLC Chapel Hill, North Carolina, United States of America, 2 Albert Einstein College of Medicine of Yeshiva University Bronx, New York, United States of America, 3 University of Washington School of Medicine Seattle, Washington, United States of America

Journal of Multidisciplinary Healthcare

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Exploring the characteristics, global distribution and reasons for retraction of published articles involving human research participants: a literature survey

This article was published in the following Dove Press journal:
Journal of Multidisciplinary Healthcare

Guowei Li[1–3]
Mariam Kamel[1]
Yanling Jin[1]
Michael Kuan Xu[1]
Lawrence Mbuagbaw[1,2]
Zainab Samaan[1,4]
Mitchell AH Levine[1–4]
Lehana Thabane[1,2]

Aim: Article retraction is a measure taken by journals or authors where there is evidence of research misconduct or error, redundancy, plagiarism or unethical research. Recently, the retraction of scientific publications has been on the rise. In this survey, we aimed to describe the characteristics and distribution of retracted articles and the reasons for retractions.

Methods: We searched retracted articles on the PubMed database and Retraction Watch website from 1980 to February 2016. The primary outcomes were the characteristics and distribution of retracted articles and the reasons for retractions. The secondary outcomes included how article retractions were handled by journals and how to improve the journal practices toward article retractions.

Retractions raise a serious challenge of whether we can trust institutions to effectively deal with research misconduct—they all have COIs

⑧ P-values are the main cause of publication bias

RESEARCH

The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey

CMAJ 2007;176(8):1091-6

John P.A. Ioannidis, Thomas A. Trikalinos

Publication bias: Publication of research that only shows a "statistically significant" finding

Research Article

Statistics in Medicine

Received 9 October 2014, Accepted 20 April 2015 Published online 18 May 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6525

Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews

Michal Kicinski,[a*†] David A. Springate[b,c] and Evangelos Kontopantelis[b,d]

The authors conclude: "In the largest study on publication bias in meta-analyses to date, we found evidence of publication bias in Cochrane systematic reviews".

If we cannot trust Cochrane reviews because of "publication bias" or failure to assess it, what hope do we have for guidelines based on these reviews?

⑨ Significant findings lead to over-representation of p-values in abstracts and press releases

Contemporary Clinical Trials Communications 7 (2017) 194–199

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

The over-representation of significant p values in abstracts compared to corresponding full texts: A systematic review of surgical randomized trials

Yusuf Assem[a,c,*], Sam Adie[a,b,c], Jason Tang[a,c], Ian A. Harris[a,b,c]

[a] University of New South Wales, South Western Sydney Clinical School, Liverpool Hospital, Liverpool, Australia
[b] Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, Liverpool, Australia
[c] South West Sydney Local Health District, Liverpool Hospital, Liverpool, Australia

Abstracts are biased towards "statistically significant" findings

The consequence is exaggeration of scientific findings

thebmj

BMJ 2014;349:g7015 doi: 10.1136/bmj.g7015 (Published 9 December 2014) Page 1 of 8

The association between exaggeration in health related science news and academic press releases: retrospective observational study

Petroc Sumner *professor*[1 2], Solveiga Vivian-Griffiths *research assistant*[1 2], Jacky Boivin *professor*[2], Andy Williams *lecturer*[3], Christos A Venetis *senior lecturer*[4], Aimée Davies *research assistant*[2], Jack Ogden *research assistant*[2], Leanne Whelan *research assistant*[2], Bethan Hughes *research assistant*[2], Bethan Dalton *research assistant*[2], Fred Boy *senior lecturer*[5], Christopher D Chambers *professor*[1 2]

"Exaggeration in news is strongly associated with exaggeration in press releases…".

OPEN ACCESS Freely available online

PLOS | MEDICINE

Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study

Amélie Yavchitz[1,2,3], Isabelle Boutron[1,2,3*], Aida Bafeta[1,2,3], Ibrahim Marroun[4], Pierre Charles[4], Jean Mantz[5], Philippe Ravaud[1,2,3]

"…the main factor associated with "spin" in press releases was the presence of "spin" in the article abstract conclusion…".

The problem of "spin" is prevalent in many clinical areas, but "statistical significance" is at the root of the problem

Open Access Research

BMJ Open

Conflicts of interest and spin in reviews of psychological therapies: a systematic review

Klaus Lieb,[1] Jan von der Osten-Sacken,[1] Jutta Stoffers-Winterling,[1] Neele Reiss,[2] Jürgen Barth[3]

Perspective

'Spin' in reports of clinical research

Kamal R Mahtani

Evid Based Med December 2016 | volume 21 | number 6 |

VOLUME 32 · NUMBER 36 · DECEMBER 20 2014

JOURNAL OF CLINICAL ONCOLOGY ORIGINAL REPORT

Impact of Spin in the Abstracts of Articles Reporting Results of Randomized Controlled Trials in the Field of Cancer: The SPIIN Randomized Controlled Trial

Isabelle Boutron, Sally Hopewell, and Philippe Ravaud, Methods of Therapeutic Evaluation of Chronic Diseases

Isabelle Boutron, Douglas G. Altman, Sally Hopewell, Francisco Vera-Badillo, Ian Tannock and Philippe Ravaud

⑩ P-values have several conceptual, computation and interpretation limitations

https://www.fharrell.com/post/pval-litany/

STATISTICAL THINKING

Home Posts Talks Datamethods Links Publications Teaching

A Litany of Problems With p-values

Last updated on 2019-08-04 · 10 min read · 82 Comments

Problems include: conceptual, indirectness, computation, multiplicity mess, inability to incorporate context, distortion of scientific conclusions, etc

Frankly, reliance on p-values has more challenges than I am able to discuss in a short presentation…

On-going debate on what to do with "statistical significance"

The American Statistician

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://amstat.tandfonline.com/loi/utas20

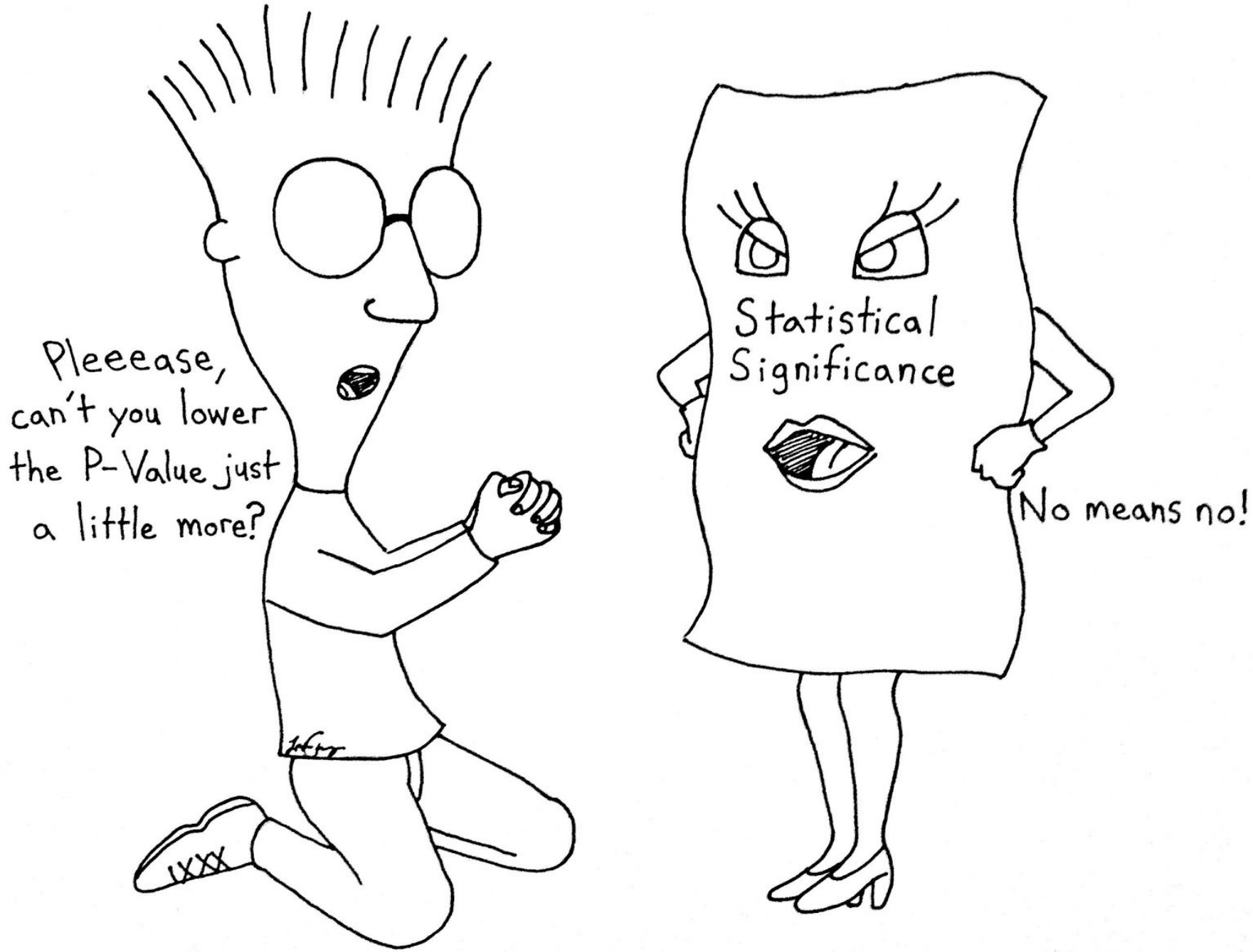
The ASA's Statement on *p*-Values: Context, Process, and Purpose

Ronald L. Wasserstein & Nicole A. Lazar

In 2016, the ASA released a warning against the misuse of statistical significance and *P* values.

The ASA statement's six principles, many of which address misconceptions and misuse of the p-value, are the following:

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.


Pleeease, can't you lower the P-Value just a little more?
Statistical Significance
No means no!

Another group of practicing statisticians and methodologists propose moving the p-value threshold from 0.05 to 0.005

(Nature Human Behaviour 2018; 2:6–10)

comment

Redefine statistical significance

We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

JAMA Editorial by John Ioannidis advocated for the adoption of the proposal to move the p-value threshold from 0.05 to 0.005

(*JAMA*. 2018;319(14):1429-1430)

A note of caution…

A simulation study shows that "this reduced false-positive conclusions but strongly increased the overestimation of significant effects (up to 320%)"

Received: 8 September 2017 | Accepted: 8 February 2018

DOI: 10.1111/eci.12912

METHODS

Eur J Clin Invest. 2018;48:e12912.

WILEY

Consequences of relying on statistical significance: Some illustrations

Ben Van Calster[1,2] | Ewout W. Steyerberg[2] | Gary S. Collins[3] | Tim Smits[4]

In January 2019…

A proposal by over 40 statisticians to disallow the term "statistically significant" and all its cognates and symbolic adjuncts

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 352–357: Statistical Inference in the 21st Century
https://doi.org/10.1080/00031305.2019.1543616

Coup de Grâce for a Tough Old Bull: "Statistically Significant" Expires

Stuart H. Hurlbert[a], Richard A. Levine[b], and Jessica Utts[c]

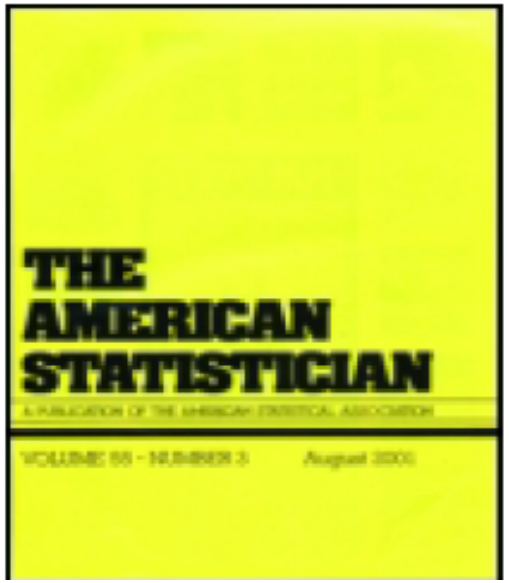
[a]Department of Biology, San Diego State University, San Diego, CA; [b]Department of Statistics, San Diego State University, San Diego, CA; [c]Department of Statistics, University of California, Irvine, CA

ABSTRACT

Many controversies in statistics are due primarily or solely to poor quality control in journals, bad statistical textbooks, bad teaching, unclear writing, and lack of knowledge of the historical literature. One way to improve the practice of statistics and resolve these issues is to do what initiators of the 2016 ASA statement did: take one issue at a time, have extensive discussions about the issue among statisticians of diverse backgrounds and perspectives and eventually develop and publish a broadly supported consensus on that issue. Upon completion of this task, we then move on to deal with another core issue in the same way. We propose as the next project a process that might lead quickly to a strong consensus that the term "statistically significant" and all its cognates and symbolic adjuncts be disallowed in the scientific literature except where focus is on the history of statistics and its philosophies and methodologies. Calculation and presentation of accurate *p*-values will often remain highly desirable though not obligatory. Supplementary materials for this article are available online in the form of an appendix listing the names and institutions of 48 other statisticians and scientists who endorse the principal propositions put forward here..

ARTICLE HISTORY
Received March 2018
Revised October 2018

KEYWORDS
NeoFisherian significance assessment; Statistical significance; Type I error; *p*-values; Dichotomized language; Teaching of statistics



ASA dedicated the whole edition to more than 40 papers on '**Statistical inference in the 21st century: a world beyond *P* < 0.05**'

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 1–19: Editorial
https://doi.org/10.1080/00031305.2019.1583913

Taylor & Francis
Taylor & Francis Group

EDITORIAL

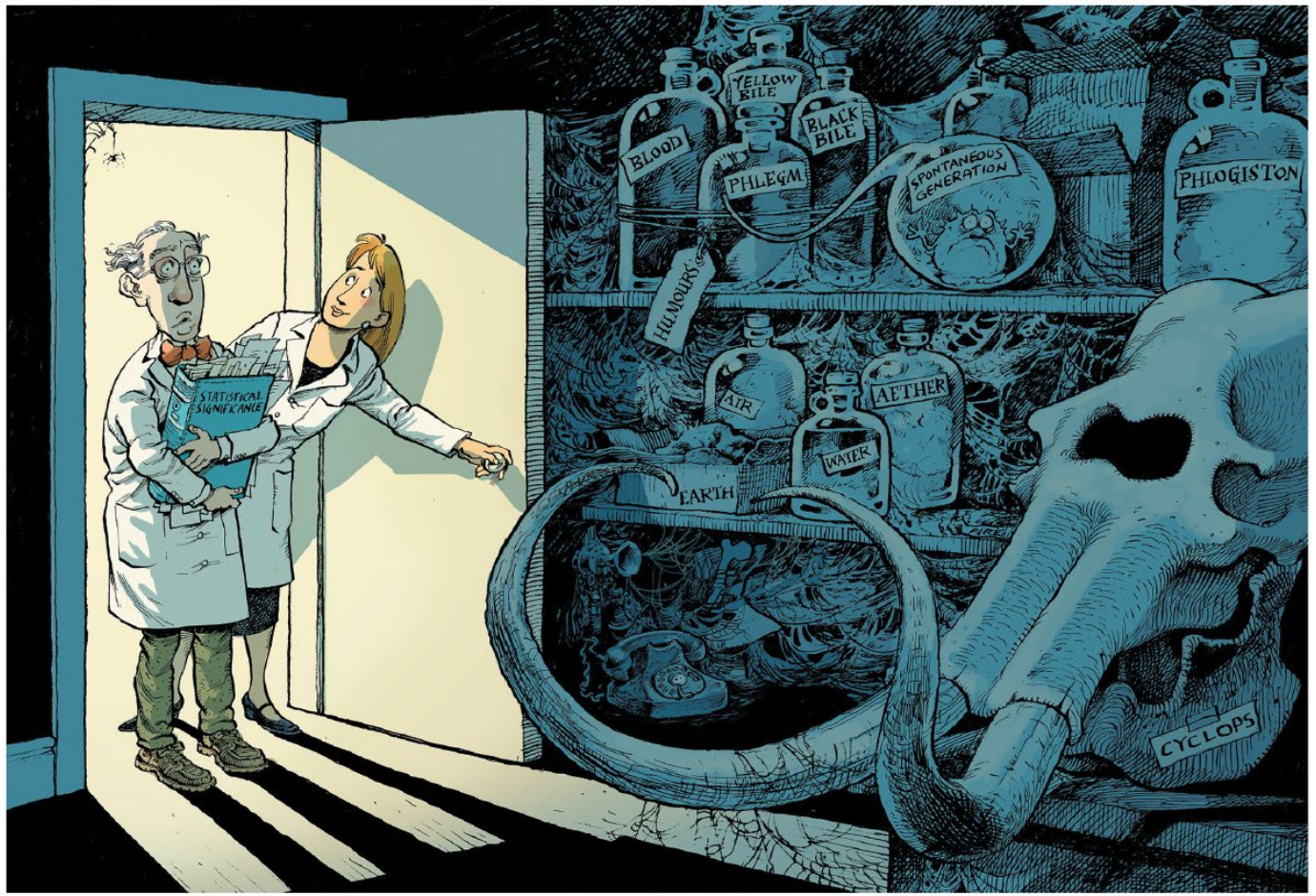
Moving to a World Beyond "$p < 0.05$"

The editors give a caution:

"don't say ~~"~~statistically significant~~"~~"

There was also a call in Nature with more than 800 signatories, calling for the abandonment of "statistical significance"

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

Valentin Amrhein, **Sander Greenland**, **Blake McShane** and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

This generated a lot of debate on both sides

Others hail is as a sensible and practical thing to do

Others push back against it!

CORRESPONDENCE · 22 MARCH 2019

Retiring statistical significance would give bias a free pass

It is quite unlikely that practice will change anytime soon!

The *New England Journal of Medicine* has become the first major journal to change its guidelines for statistical reporting in response to the ASA Statement on P-values and Statistical Significance and subsequent *The American Statistician* special issue on statistical inference

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIAL

New Guidelines for Statistical Reporting in the *Journal*

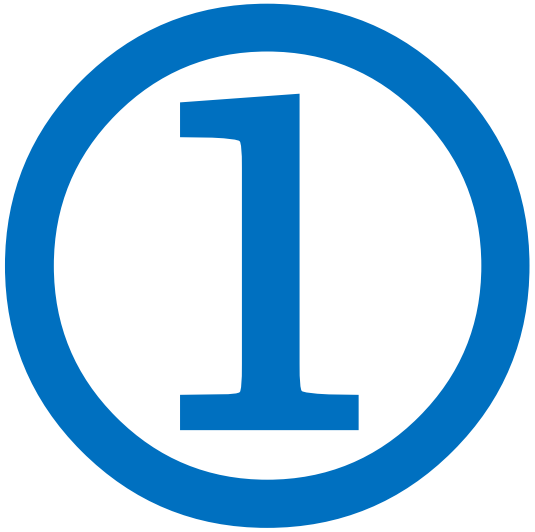
David Harrington, Ph.D., Ralph B. D'Agostino, Sr., Ph.D., Constantine Gatsonis, Ph.D., Joseph W. Hogan, Sc.D., David J. Hunter, M.B., B.S., M.P.H., Sc.D., Sharon-Lise T. Normand, Ph.D., Jeffrey M. Drazen, M.D., and Mary Beth Hamel, M.D., M.P.H

- NEJM has expanded and refined new statistical reporting guidelines for authors to cover both RCTs and observational studies
- Notably, NEJM now requires authors to **replace p-values with both point estimates and 95% CIs** when neither the study protocol nor the data analysis plan included adjustments for multiple testing.

Real change will require a complete overhaul of our educational systems and academic culture

Putting "significance" on "clinical relevance"

For the time being, there has been suggestions and recommendations of how to improve reporting and interpretation



#1: Adopt the CONSORT Statement and its extensions in reporting and interpreting the results

CONSORT TRANSPARENT REPORTING of TRIALS

Search

Home | CONSORT 2010 | Extensions | Downloads | Examples | Resources | About CONSORT

First published in 1996, the CONSORT Statement aimed to improve the reporting of RCTs to facilitate EBM

Welcome to the CONSORT Website

CONSORT stands for Consolidated Standards of Reporting Trials and encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials.

The CONSORT Statement

The main product of CONSORT is the CONSORT Statement, which is an evidence-based, minimum set of recommendations for reporting randomized trials. It offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and

CONSORT 2010 Key Documents

- CONSORT 2010 Checklist
- CONSORT 2010 Flow Diagram
- CONSORT 2010 Statement
- CONSORT 2010 Explanation and Elaboration Document

Recent Tweets

Follow Us

EQUATOR Network @EQUATORNetwork 3 Mar
RT @trished: Declaration of transparency Now adopted by several journals as well as @bmj_latest & @BMJ_Open 86
equator-network.org/2014/08/12/dec...
Retweeted by CONSORT Statement

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	

There is no mention of reporting p-values in the recommendation

Many CONSORT extensions have been published since its inception

ACADEMIA AND CLINIC | Annals of Internal Medicine

CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials

Kenneth F. Schulz, PhD, MBA; Douglas G. Altman, DSc; and David Moher, PhD, for the CONSORT Group*

The CONSORT (Consolidated Standards of Reporting Trials) statement is used worldwide to improve the reporting of randomized, controlled trials. Schulz and colleagues describe the latest version, CONSORT 2010, which updates the reporting guideline based on new methodological evidence and accumulating experience.

Ann Intern Med. 2010;152:726-732. www.annals.org

For author affiliations, see end of text.

* For the CONSORT Group contributors to CONSORT 2010, see the Appendix (available at www.annals.org).

This article was published at www.annals.org on 24 March 2010.

Editor's Note: In order t
CONSORT 2010 Stateme
www.annals.org and will
Lancet, Obstetrics & Gy
Medicine, Journal of Clit
cine, and Trials. *The auth*
article. For details on furth
(www.consort-statement.or

Randomized, controlle
signed, conducted, a
standard in evaluating he
randomized trials can yiel
odological rigor (1). To a
a published report need
information on its meth
nately, attempted assessme
of many trial reports negl
descriptions of that critical

That lack of adequa
ment of the original CO
of Reporting Trials) state
5 years later (6–8). Whil
reporting quality for some
10), many trial reports stil
more, new methodologica
ence has accumulated sin
sequently, we organized a
update the 2001 statemei
result of that process, CC

Intent of CONSORT

The CONSORT 20
cluding the 25-item check

See also:

Web-Only
Appendix
Conversion of graphics i

BMJ

BMJ 2012;345:e5661 doi: 10.1136/bmj.e5661 (Published 4 September 2012)

RESEARCH METHODS & REPORTING

Consort 2010
randomised t

The Consolidated Stanc
the reporting of random
reporting of parallel gro
further update in 2010. A
in 2008. In earlier paper
statement for the reportir
guidance, based on the 2
for the reporting of absti

Marion K Campbell *dire*
healthcare evaluation[2],

[1]Health Services Research Unit, Ur
and Tropical Medicine, London, UK

Many journals now require that
guidelines in the Consolidated
(CONSORT) statement, first pu
2001,[2] and revised most recentl
includes a checklist of items tha
report. These items are evidenc
are regularly reviewed.[3] The sta
including a flow diagram to show
from group assignment through
explanation and elaboration of t
items is provided in an accompa
The standard CONSORT statem
group randomised controlled tri
participants are randomly assig
in some situations it is preferabl
people (such as communities, fa
rather than individuals. Reason
"contamination" (the unintentio
effects from one treatment grou
interventions if individual rand
certain settings, randomisation b
method of conducting a trial.[4] T
variously known as field trials,
randomised trials, place based t
randomised trials.[5] Although we
use of the term "cluster random
those searching electronically f
expand their search strategy to

Reporting of Patient-Reported Outcomes in Randomized Trials

The CONSORT PRO Extension

Melanie Calvert, PhD
Jane Blazeby, MD
Douglas G. Altman, DSc
Dennis A. Revicki, PhD
David Moher, PhD
Michael D. Brundage, MD
for the CONSORT PRO G

RESEARCH METHODS & REPORTING

THE CONSORT
dated Standards o
Trials) Statemen
lished in 1996 a
cently revised in 2010,[1,2]
dence-based recommen
improve the completene
ing of randomized cont
(RCTs). The statement fo
allel-group trials, but a n
tensions for reporting o
signs (cluster, noninfe
equivalence), intervention
macologic and herbal th
for specific data, such as
been developed.[3] The CO
ment is endorsed by major
editorial groups, such as
tional Committee of Me
Editors, and its use has
ated with improved repor

Rationale for a Conse
Extension Focused on
Patient-Reported Ou

Patient reported outcomes
include health-related q
(HRQL), symptoms, utili
isfaction ratings and are
sessments that are patie

[2]Health Services Sciences,
Sunnybrook Hospital, Toronto,
Toronto, Canada
[3]Institute for Clinical Evaluative
Sciences, Toronto
[4]Department of
Health Policy, Management and
Evaluation, University of Toronto,
Toronto
[5]Division of International Health
(IHCAS), Karolinska Institutet,
Stockholm, Sweden
[6]Clinical and Population Sciences
and Education, University of
London, Dundee
[7]Norwegian Knowledge Centre for
the Health Services, Oslo, Norway
[8]Faculty of Medicine, University
of Toronto
[9]Centre for Statistics in Medicine,
University of Oxford, Oxford
[10]Center for Medical Technology
Policy, Baltimore, MD, USA
[11]Division of General Internal
Medicine, Johns Hopkins School of
Medicine, Baltimore, MD
[12]Center for Healthcare Policy,
Stanford University School of
Medicine, Palo Alto, CA, USA
[13]Department of Clinical
Epidemiology and Biostatistics
and Department of Medicine,
McMaster University Faculty of
Health Sciences, Hamilton, ON,
Canada
[14]Clinical Epidemiology Program,
Ottawa Health Research Institute,
Ottawa, Canada
[15]Department of Epidemiology and
Community Medicine, Faculty of
Medicine, University of Ottawa,
Ottawa
Correspondence to: M
lwarenstein@mericx.
lwarenstein@sjhsres.on.ca

Cite this as: *BMJ*
2008;337:a2390
doi: 10.1136/bmj.a2390

Improving the reporting of pragmatic trials: an extension of the CONSORT statement

Merrick Zwarenstein,[1,2,3] Shaun Treweek,[4,5] Joel J Gagnier[6,7] Douglas G Altman,[7] Sean Tunis,[8,9,10] Brian Haynes,[11] Andrew D Oxman,[12] David Moher,[12,13] for the CONSORT and Pragmatic Trials in Healthcare (Practihc) groups

Pragmatic trials are designed to inform decisions about practice, but poor reporting can reduce their usefulness. The **CONSORT and Practihc groups** describe modifications to the CONSORT guidelines to help readers assess the applicability of the results

Abstract

Background
The CONSORT statement is intended to improve reporting of randomised controlled trials and focuses on minimising the risk of bias (internal validity). The applicability of a trial's results (generalisability or external validity) is also important, particularly for pragmatic trials. A pragmatic trial (a term first used in 1967 by Schwartz and Lellouch) can be broadly defined as a randomised controlled trial whose purpose is to inform decisions about practice. This extension of the CONSORT statement is intended to improve the reporting of such trials and focuses on applicability.

Methods
At two, two-day meetings held in Toronto in 2005 and 2008, we reviewed the CONSORT statement and its extensions, the literature on pragmatic trials and applicability, and our experiences in conducting pragmatic trials.

Recommendations
We recommend extending eight CONSORT checklist items for reporting of pragmatic trials: the background, participants, interventions, outcomes, sample size, blinding, participant flow, and generalisability of the findings. These six extensions are presented, along with illustrative examples of reporting, and an explanation of each extension. Adherence to these reporting criteria will make it easier for decision makers to judge how applicable the results of randomised controlled trials are to their own conditions. Empirical studies are needed to ascertain the usefulness and comprehensiveness of these CONSORT checklist item extensions. In the meantime we recommend that those who support, conduct, and report pragmatic trials should use this extension of the CONSORT statement to facilitate the use of trial results in decisions about health care.

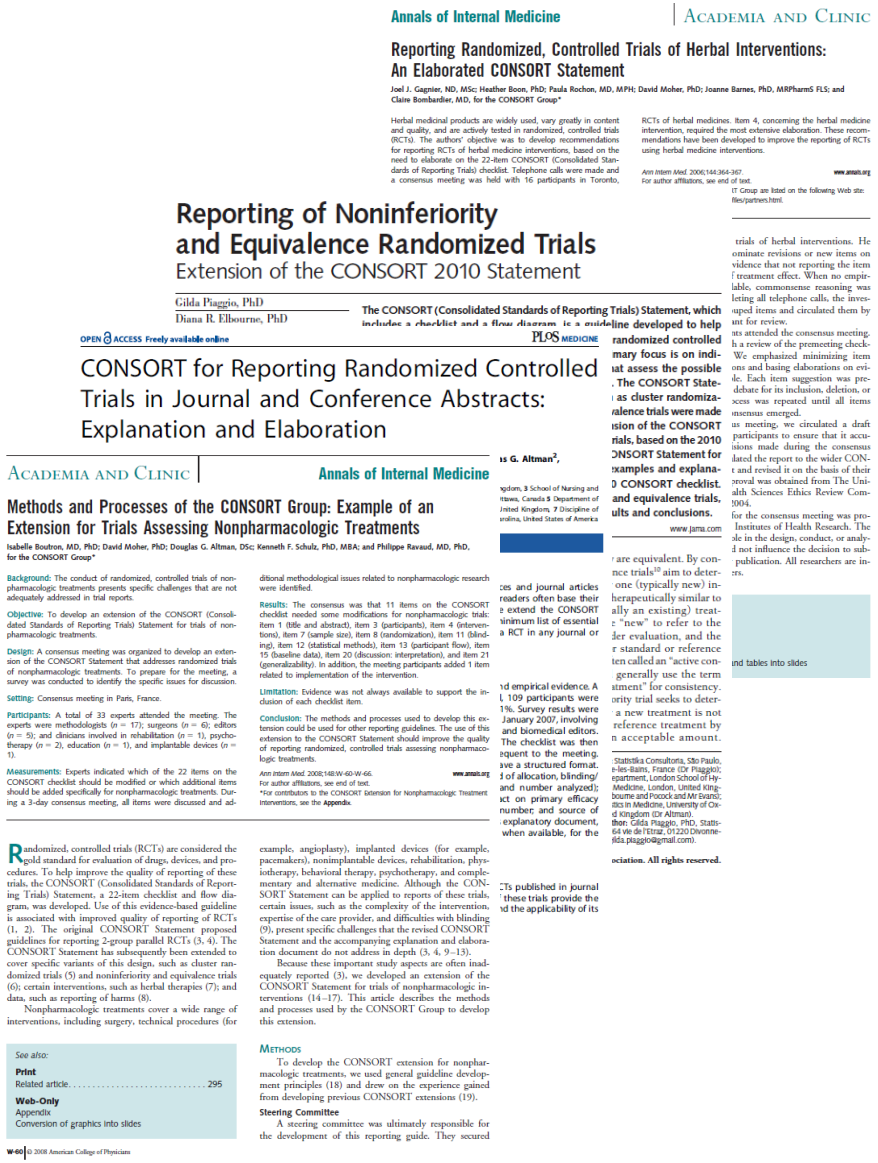
Randomised controlled trials are used to assess the benefits and harms of interventions in health care. If conducted properly, they minimise the risk of bias (threats to internal validity), particularly selection bias.[1,2] There is, however, considerable evidence that trials are not always well reported,[3,4] and this can be associated with bias, such as selective reporting of outcomes.[5]

The usefulness of a trial report also depends on the clarity with which it details the relevance of its interventions, participants, outcomes, and design to the clinical, health service, or policy question it examines. Furthermore, a trial may be valid and useful in the healthcare setting in which it was conducted but have limited applicability (also known as generalisability or external validity) beyond this because of differences between the trial setting and other settings to which its results are to be extrapolated.

Schwartz and Lellouch[6] coined the terms "pragmatic" to describe trials designed to help choose between options for care, and "explanatory" to describe trials designed to test causal research hypotheses—for example, that an intervention causes a particular biological

Table 1: Key differences between trials with explanatory and pragmatic attitudes, adapted from a table presented at the 2008 Society for Clinical Trials meeting by Marion Campbell, University of Aberdeen

Question	Efficacy—can the intervention work?	Effectiveness—does the intervention work when used in normal practice?
Setting	Well resourced, "ideal" setting	Normal practice
Participants	Highly selected. Poorly adherent participants and those with conditions which might dilute the effect are often excluded	Little or no selection beyond the clinical indication of interest
Intervention	Strictly enforced and adherence is monitored closely	Applied flexibly as it would be in normal practice
Outcomes	Often short term surrogates or process measures	Directly relevant to participants, funders, communities, and healthcare practitioners
Relevance to practice	Indirect—little effort made to match design of trial to decision making needs of those in usual setting in which intervention will be implemented	Direct—trial is designed to meet needs of those making decisions about treatment options in setting in which intervention will be implemented



Annals of Internal Medicine | ACADEMIA AND CLINIC

Reporting Randomized, Controlled Trials of Herbal Interventions: An Elaborated CONSORT Statement

Joel J. Gagnier, ND, MSc; Heather Boon, PhD; Paula Rochon, MD, MPH; David Moher, PhD; Joanne Barnes, PhD, MRPharmS FLS; and Claire Bombardier, MD, for the CONSORT Group*

Herbal medicinal products are widely used, vary greatly in content and quality, and are actively tested in randomized, controlled trials (RCTs). The authors' objective was to develop recommendations for reporting RCTs of herbal medicine interventions, based on the need to elaborate on the 22-item CONSORT (Consolidated Standards of Reporting Trials) checklist. Telephone calls were made and a consensus meeting was held with 16 participants in Toronto, RCTs of herbal medicines. Item 4, concerning the herbal medicine intervention, required the most extensive elaboration. These recommendations have been developed to improve the reporting of RCTs using herbal medicine interventions.

Ann Intern Med. 2006;144:364-367. www.annals.org

For author affiliations, see end of text.

IT Group are listed on the following Web site: files/partners.html.

Reporting of Noninferiority and Equivalence Randomized Trials

Extension of the CONSORT 2010 Statement

Gilda Piaggio, PhD
Diana R. Elbourne, PhD

The CONSORT (Consolidated Standards of Reporting Trials) Statement, which includes a checklist and a flow diagram, is a guideline developed to help randomized controlled mary focus is on indi- at assess the possible . The CONSORT State- as cluster randomiza- valence trials were made sion of the CONSORT rials, based on the 2010 ONSORT Statement for xamples and explana- D CONSORT checklist, and equivalence trials, ults and conclusions.

OPEN ACCESS Freely available online PLOS MEDICINE

CONSORT for Reporting Randomized Controlled Trials in Journal and Conference Abstracts: Explanation and Elaboration

s G. Altman[2],

ngdom, 3 School of Nursing and ttawa, Canada 5 Department of United Kingdom, 7 Discipline of arolina, United States of America

www.jama.com

ACADEMIA AND CLINIC | Annals of Internal Medicine

Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments

Isabelle Boutron, MD, PhD; David Moher, PhD; Douglas G. Altman, DSc; Kenneth F. Schulz, PhD, MBA; and Philippe Ravaud, MD, PhD, for the CONSORT Group*

Background: The conduct of randomized, controlled trials of nonpharmacologic treatments presents specific challenges that are not adequately addressed in trial reports.

Objective: To develop an extension of the CONSORT (Consolidated Standards of Reporting Trials) Statement for trials of nonpharmacologic treatments.

Design: A consensus meeting was organized to develop an extension of the CONSORT Statement that addresses randomized trials of nonpharmacologic treatments. To prepare for the meeting, a survey was conducted to identify the specific issues for discussion.

Setting: Consensus meeting in Paris, France.

Participants: A total of 33 experts attended the meeting. The experts were methodologists ($n = 17$); surgeons ($n = 6$); editors ($n = 5$); and clinicians involved in rehabilitation ($n = 1$), psychotherapy ($n = 2$), education ($n = 1$), and implantable devices ($n = 1$).

Measurements: Experts indicated which of the 22 items on the CONSORT checklist should be modified or which additional items should be added specifically for nonpharmacologic treatments. During a 3-day consensus meeting, all items were discussed and additional methodological issues related to nonpharmacologic research were identified.

Results: The consensus was that 11 items on the CONSORT checklist needed some modifications for nonpharmacologic trials: item 1 (title and abstract), item 3 (participants), item 4 (interventions), item 7 (sample size), item 8 (randomization), item 11 (blinding), item 12 (statistical methods), item 13 (participant flow), item 15 (baseline data), item 20 (discussion: interpretation), and item 21 (generalizability). In addition, the meeting participants added 1 item related to implementation of the intervention.

Limitation: Evidence was not always available to support the inclusion of each checklist item.

Conclusion: The methods and processes used to develop this extension could be used for other reporting guidelines. The use of this extension to the CONSORT Statement should improve the quality of reporting randomized, controlled trials assessing nonpharmacologic treatments.

Ann Intern Med. 2008;148:W-60-W-66. www.annals.org

For author affiliations, see end of text.

*For contributors to the CONSORT Extension for Nonpharmacologic Treatment Interventions, see the Appendix.

Randomized, controlled trials (RCTs) are considered the gold standard for evaluation of drugs, devices, and procedures. To help improve the quality of reporting of these trials, the CONSORT (Consolidated Standards of Reporting Trials) Statement, a 22-item checklist and flow diagram, was developed. Use of this evidence-based guideline is associated with improved quality of reporting of RCTs (1, 2). The original CONSORT Statement proposed guidelines for reporting 2-group parallel RCTs (3, 4). The CONSORT Statement has subsequently been extended to cover specific variants of this design, such as cluster randomized trials (5) and noninferiority and equivalence trials (6); certain interventions, such as herbal therapies (7); and data, such as reporting of harms (8).

Nonpharmacologic treatments cover a wide range of interventions, including surgery, technical procedures (for example, angioplasty), implanted devices (for example, pacemakers), nonimplantable devices, rehabilitation, physiotherapy, behavioral therapy, psychotherapy, and complementary and alternative medicine. Although the CONSORT Statement can be applied to reports of these trials, certain issues, such as the complexity of the intervention, expertise of the care provider, and difficulties with blinding (9), present specific challenges that the revised CONSORT Statement and the accompanying explanation and elaboration document do not address in depth (3, 4, 9–13).

Because these important study aspects are often inadequately reported (3), we developed an extension of the CONSORT Statement for trials of nonpharmacologic interventions (14–17). This article describes the methods and processes used by the CONSORT Group to develop this extension.

See also:

Print
Related article 295

Web-Only
Appendix
Conversion of graphics into slides

Methods

To develop the CONSORT extension for nonpharmacologic treatments, we used general guideline development principles (18) and drew on the experience gained from developing previous CONSORT extensions (19).

Steering Committee

A steering committee was ultimately responsible for the development of this reporting guide. They secured

ES published in journal these trials provide the d the applicability of its

ociation. All rights reserved.

Finding the Appropriate Extension

The table below lists the current official extensions of the CONSORT statement. You can click on each extension to learn more about it or to explore that extension in the checklist viewer application.

Designs	Interventions	Data
Cluster Trials	Herbal Medicinal Interventions	CONSORT-PRO
Non-Inferiority and Equivalence Trials	Non-Pharmacologic Treatment Interventions	Harms
Pragmatic Trials	Acupuncture Interventions	Abstracts
N-of-1 Trials		
Pilot and Feasibility Trials		

CONSORT Extension to Pilot Trials Working Group

Sandra Eldridge

Queen Mary University of London

Claire Chan

Michael Campbell

The University Of Sheffield.

Sally Hopewell

UNIVERSITY OF OXFORD

Lehana Thabane

McMaster University
HEALTH RESEARCH METHODS, EVIDENCE, AND IMPACT

Christine Bond

1495 UNIVERSITY OF ABERDEEN

Gillian Lancaster

Keele University

The CONSORT extension to pilot RCT paper

Eldridge *et al. Pilot and Feasibility Studies* (2016) 2:64
DOI 10.1186/s40814-016-0105-8

Pilot and Feasibility Studies

RESEARCH Open Access

CONSORT 2010 statement: extension to randomised pilot and feasibility trials

Sandra M. Eldridge[1*], Claire L. Chan[1], Michael J. Campbell[2], Christine M. Bond[3], Sally Hopewell[4], Lehana Thabane[5], Gillian A. Lancaster[6] and on behalf of the PAFS consensus group

RESEARCH METHODS AND REPORTING

CONSORT 2010 statement: extension to randomised pilot and feasibility trials

Sandra M Eldridge,[1] Claire L Chan,[1] Michael J Campbell,[2] Christine M Bond,[3] Sally Hopewell,[4] Lehana Thabane,[5] Gillian A Lancaster[6] on behalf of the PAFS consensus group

thebmj | *BMJ* 2016;355:i5239 | doi: 10.1136/bmj.i5239

Emphasis on reporting estimates of effects or comparisons (95% CIs) remains a core message across along CONSORT extensions

Focus on determining clinical relevance, instead of p-values

REVIEW

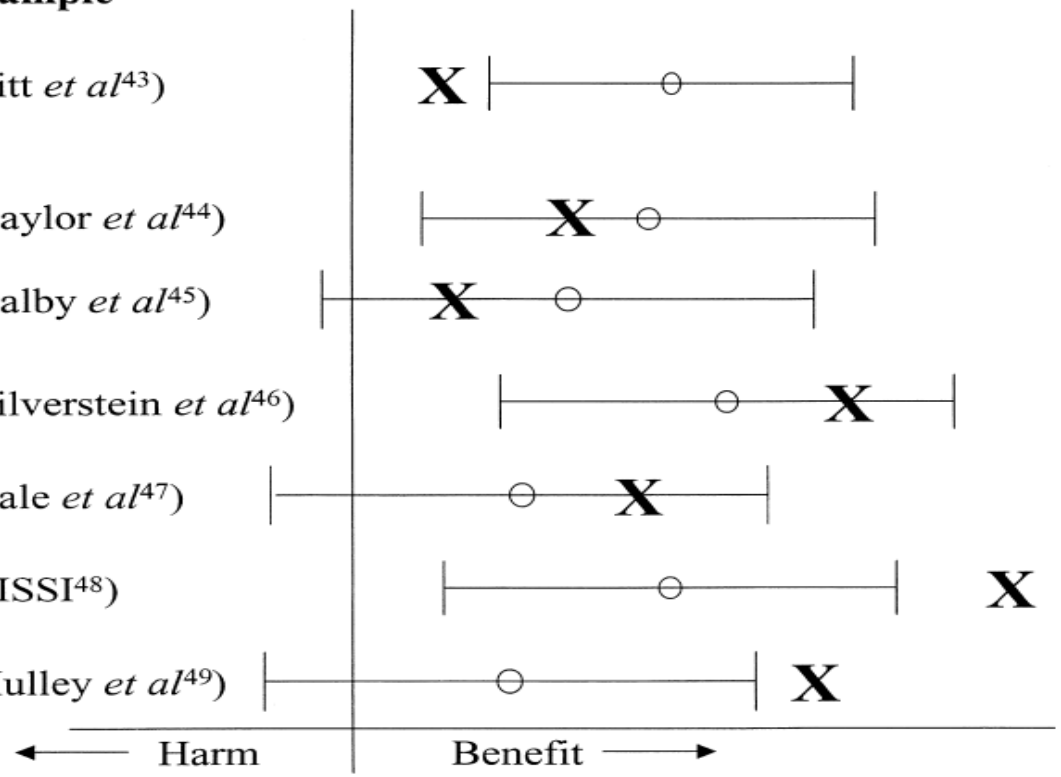
Determination of the Clinical Importance of Study Results

A Review

Malcolm Man-Son-Hing, MD, MSc, Andreas Laupacis, MD, MSc, Keith O'Rourke, MBA, Frank J. Molnar, MD, MSc, Jeffery Mahon, MD, MSc, Karen B. Y. Chan, MD, George Wells, PhD

The present a standardized approach with examples to show how to determine clinical relevance, and its connection to statistical significance

Clinical Importance	Statistical Significance	Example
1) Definite	Yes	A (Pitt *et al*[43])
2) Probable	Yes	B (Naylor *et al*[44])
	No	C (Dalby *et al*[45])
3) Possible	Yes	D (Silverstein *et al*[46])
	No	E (Dale *et al*[47])
4) Definitely Not	Yes	F (GISSI[48])
	No	G (Hulley *et al*[49])



X - minimal clinically important difference

⊢o⊣ - point estimate and 95% confidence interval

Table 1. Benefits of a Simple, Standardized Approach to Determining Clinical Importance

1. Clarifies the relationship between statistical significance and clinical importance.
2. Promotes balanced interpretation of trial results from both a statistical and a clinical perspective.
3. Forces designers of prospective clinical trials to more carefully choose Δ values for their sample size calculations.
4. Can be combined with cumulative meta-analysis to decide when no further studies are needed.
5. May help reduce publication bias.
6. Can provide guidance to trialists regarding stopping rules.
7. Promotes further methodologic work to determine clinically important differences.

For studies with results that are not statistically significant, focus on confidence intervals

Open Access Research

BMJ Open

Interpretation of CIs in clinical trials with non-significant results: systematic review and recommendations

Jennifer S Gewandter,[1] Michael P McDermott,[2] Rachel A Kitt,[1] Jenna Chaudari,[1] James G Koch,[1] Scott R Evans,[3] Robert A Gross,[4,5] John D Markman,[6] Dennis C Turk,[7] Robert H Dworkin[1]

Box 1 CI reporting recommendations for RCTs with statistically non-significant results

- Report CIs for the treatment effect.
- Discuss interpretation of the CI regarding the magnitude of effects that can be ruled out with reasonable confidence.
- Discuss whether the results suggest a 'negative' or 'inconclusive' result.
- Acknowledge any uncertainty regarding what is considered a clinically meaningful treatment effect on the outcome measure used in the trial.

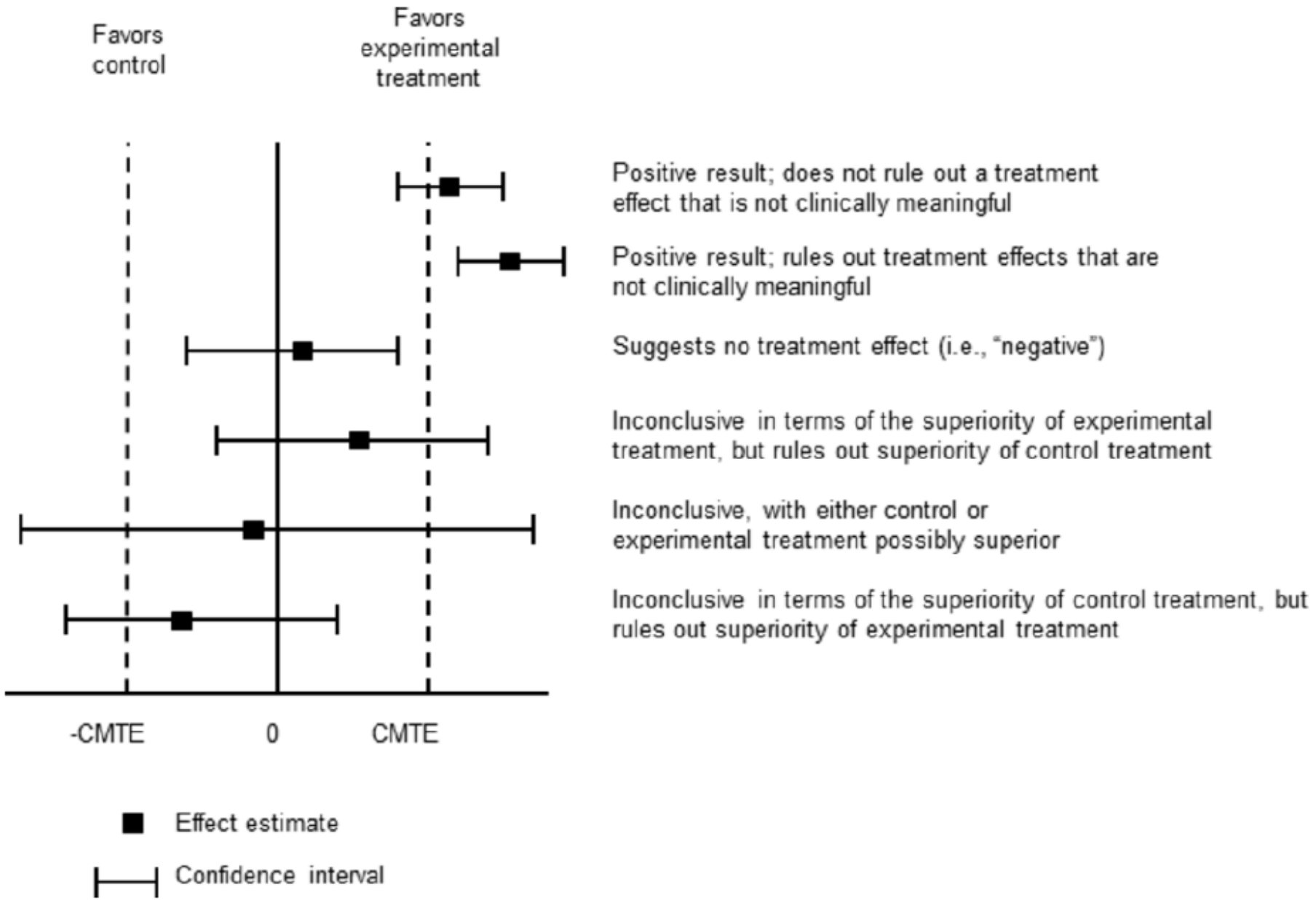


Figure 1 Using CIs to interpret results of randomised clinical trials. Note that a value of zero indicates no treatment effect in this case; in other cases such as when the treatment effect is quantified using, for example, an OR, HR or relative risk, a value of 1 would indicate no treatment effect. Adapted from Senn.[23] CMTE, clinically meaningful treatment effect.

Consider abandoning "p-values thresholds" and "use of statistical significance"

In the JAMA Editorial by John Ioannidis proposed provided a summary of proposals to address the situations

(*JAMA*. 2018;319(14):1429-1430)

Table. Various Proposed Solutions for Improving Statistical Inference on a Large Scale

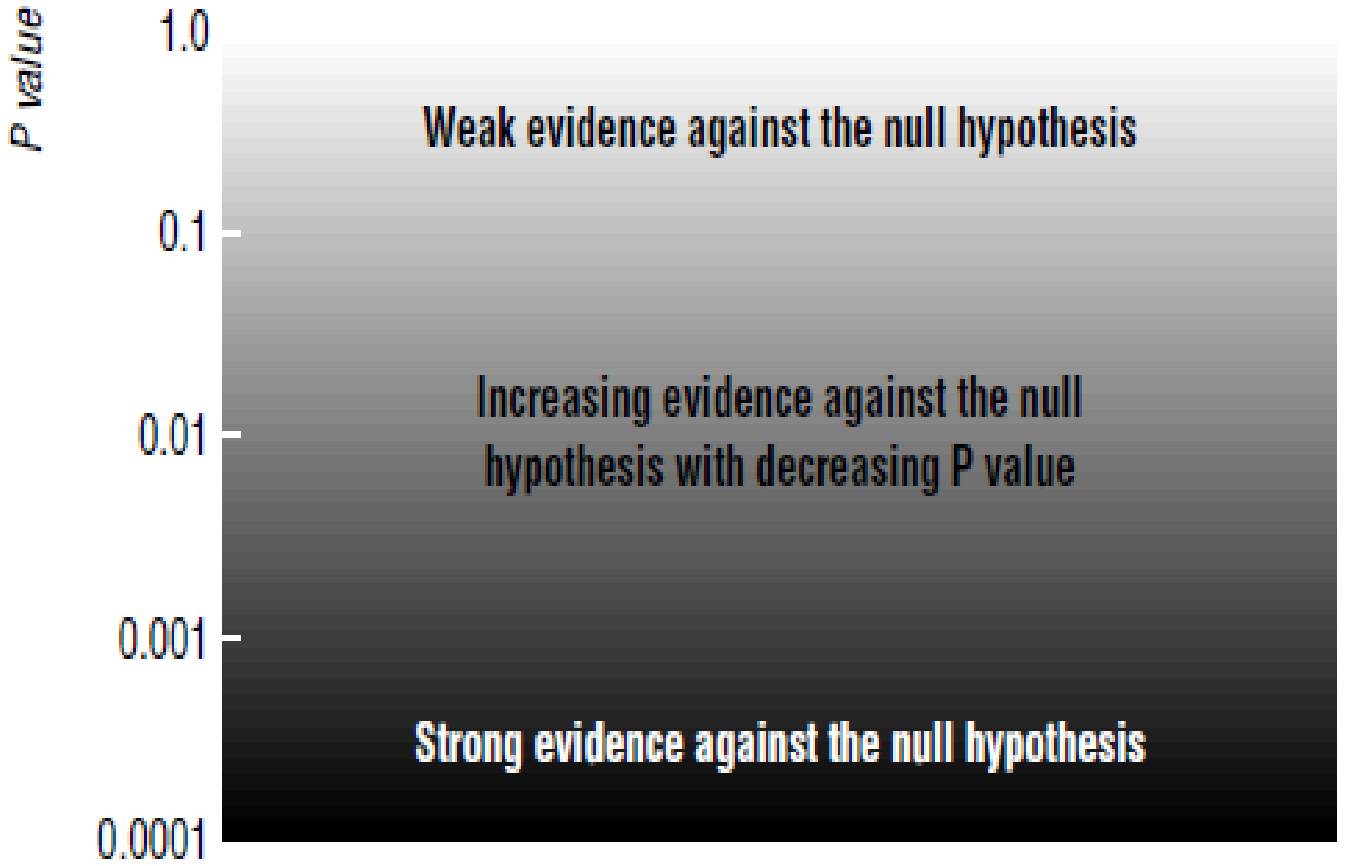
	Apply to Past Literature: Easy or Fast Solution?	Apply to Future Research and Publications: Easy or Fast Solution?
Lower *P* value thresholds	A rather simple temporizing solution	Has potential collateral harms (see text) and success depends on adoption or enforcement by stakeholders (eg, journals, funders, societies)
Abandon *P* value thresholds and instead use exact *P* value	Many published *P* values have only been reported with thresholds	Success depends on extent of adoption or enforcement by stakeholders
Abandon *P* values entirely	Not easy because often nothing or little else has been provided; many articles did not report effect sizes and most lacked confidence intervals *P* values are still a good choice for some research applications	Previous pleas have not been successful to gain traction May succeed more easily in some fields (eg, assessment of diagnostic performance or choosing of predictors for prognostic models in which use of *P* values makes little or no sense)
Use alternative inference methods (eg, Bayesian statistics)	Partly doable (eg, one may convert *P* values to Bayes factors, but needs sophisticated training)	Would be suitable for most studies; increase in use of Bayesian methods (and other inferential approaches such as false-discovery rates) has been substantial recently, but would need to accelerate in the future
Focus on effect sizes and their uncertainty	Often not reported at all, but has become more common in more recent literature, particularly in clinical trials and meta-analyses	Relevant to the vast majority of the clinical literature, should be heavily endorsed as more directly linked to decision making, and it may be easier to promote than more sophisticated solutions
Train the scientific workforce	Takes time and major commitment to achieve sufficient statistical literacy.	Can lead to a more definitive solution, choosing fit for purpose statistics and inference tools, but may require major recasting of training priorities in curricula
Address biases that lead to inflated results	Requires major training; biases are often impossible to detect from published reports	Preemptively dealing with biases is ideal, but needs concerted commitment of multiple stakeholders to promote and incentivize better research practices

Provide researchers with specific guidance on how to report results and interpret p-values

BMJ 2001;322:226–31

Sifting the evidence—what's wrong with significance tests?

Jonathan A C Sterne, George Davey Smith



Suggested interpretation of P values from published medical research

The cartoonists will probably chuckle that we are adopting their advice!

Perhaps, instead of clinical relevance, consider focusing on other dimensions

Patient-importance

EDITORIALS

Patients at the center: In our practice, and in our use of language

Gordon Guyatt; Victor Montori; P J Devereaux; Holger Schunemann; Mohit Bhandari
ACP Journal Club; Jan/Feb 2004; 140, 1; ProQuest Nursing Journals
pg. A11

Person significance

THE LANCET

ESSAY | VOLUME 351, ISSUE 9096, P134-136, JANUARY 10, 1998

Personal significance: the third dimension

Dr Kieran G Sweeney, MRCGP • Domhnall MacAuley, MRCGP • Prof Denis Pereira Gray, FRCGP

Published: January 10, 1998 • DOI: https://doi.org/10.1016/S0140-6736(97)06316-2

Patient centred

BLOG NEWSROOM HELP CENTER SUBSCRIBE

pcori Patient-Centered Outcomes Research Institute

ABOUT US | RESEARCH & RESULTS | TOPICS | ENGAGEMENT | FUNDING OPPORTUNITIES | MEETINGS & EVENTS

1. "Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?"

2. "What are my options, and what are the potential benefits and harms of those options?"

3. "What can I do to improve the outcomes that are most important to me?"

4. "How can clinicians and the care delivery systems they work in help me make the best decisions about my health and health care?"

Patient-oriented

Government of Canada | Gouvernement du Canada

Canada.ca | Services | Departments | Français

Canadian Institutes of Health Research

Canada

Search

Funding | Institutes | College of Reviewers | Initiatives | Collaboration | Discoveries for life

Home → Strategies → Strategy for Patient-Oriented Research

Strategy for Patient-Oriented Research

- About us
- Facilitating research
- Funding research
- Foundations of SPOR
- Patient engagement

Strategy for Patient-Oriented Research

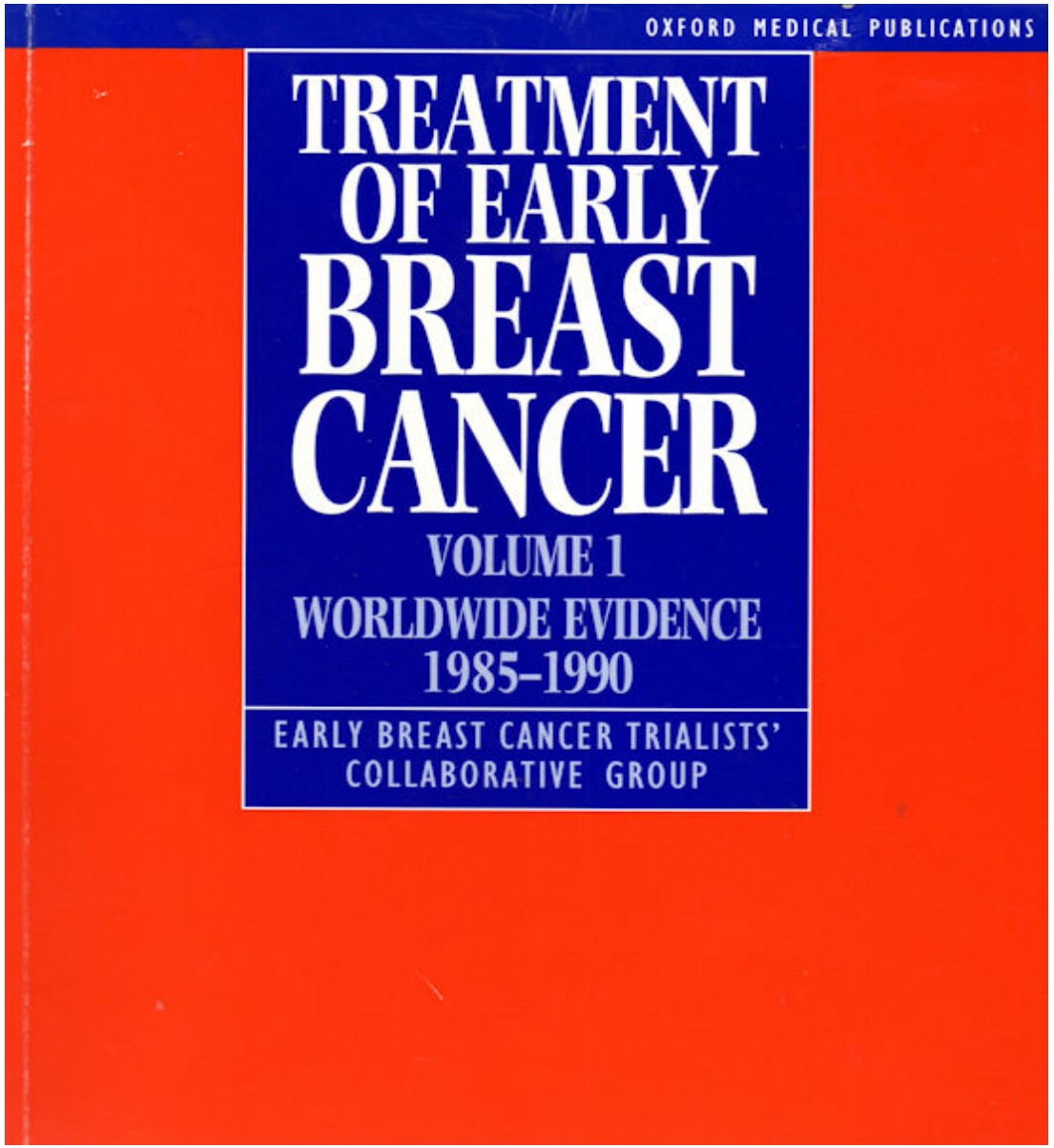
Stay connected

Subscribe to our mailing list – email us at spor-srap@cihr-irsc.gc.ca.

Contact Information

Questions regarding the Strategy for Patient-

Humanly important



Some final thoughts

Statistical thinking

=

Evidence-based thinking

We don't need

"statistical significance"

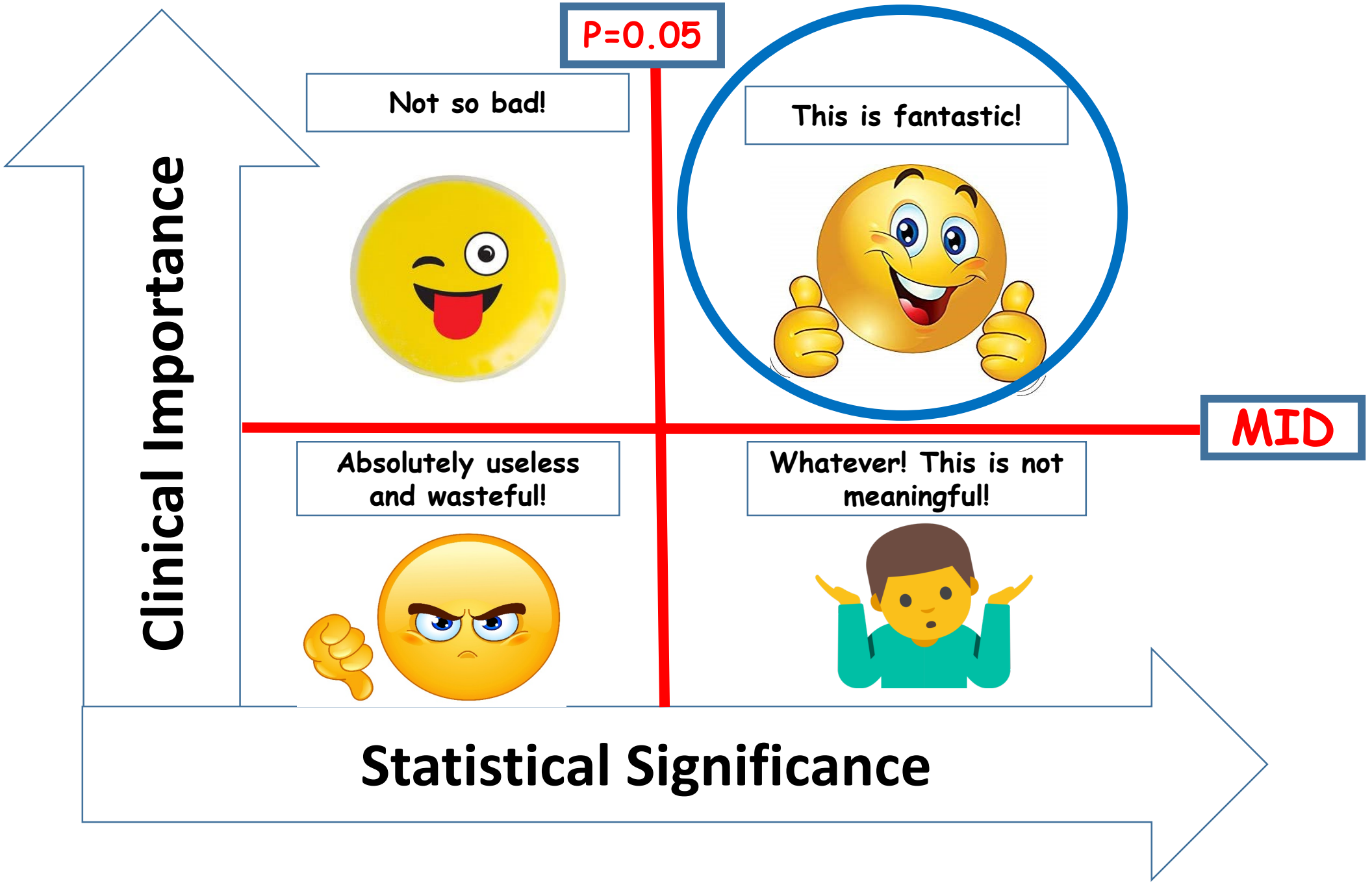
to think

"statistically"

However, …

We need
"clinical relevance"
to be
"evidence-based"

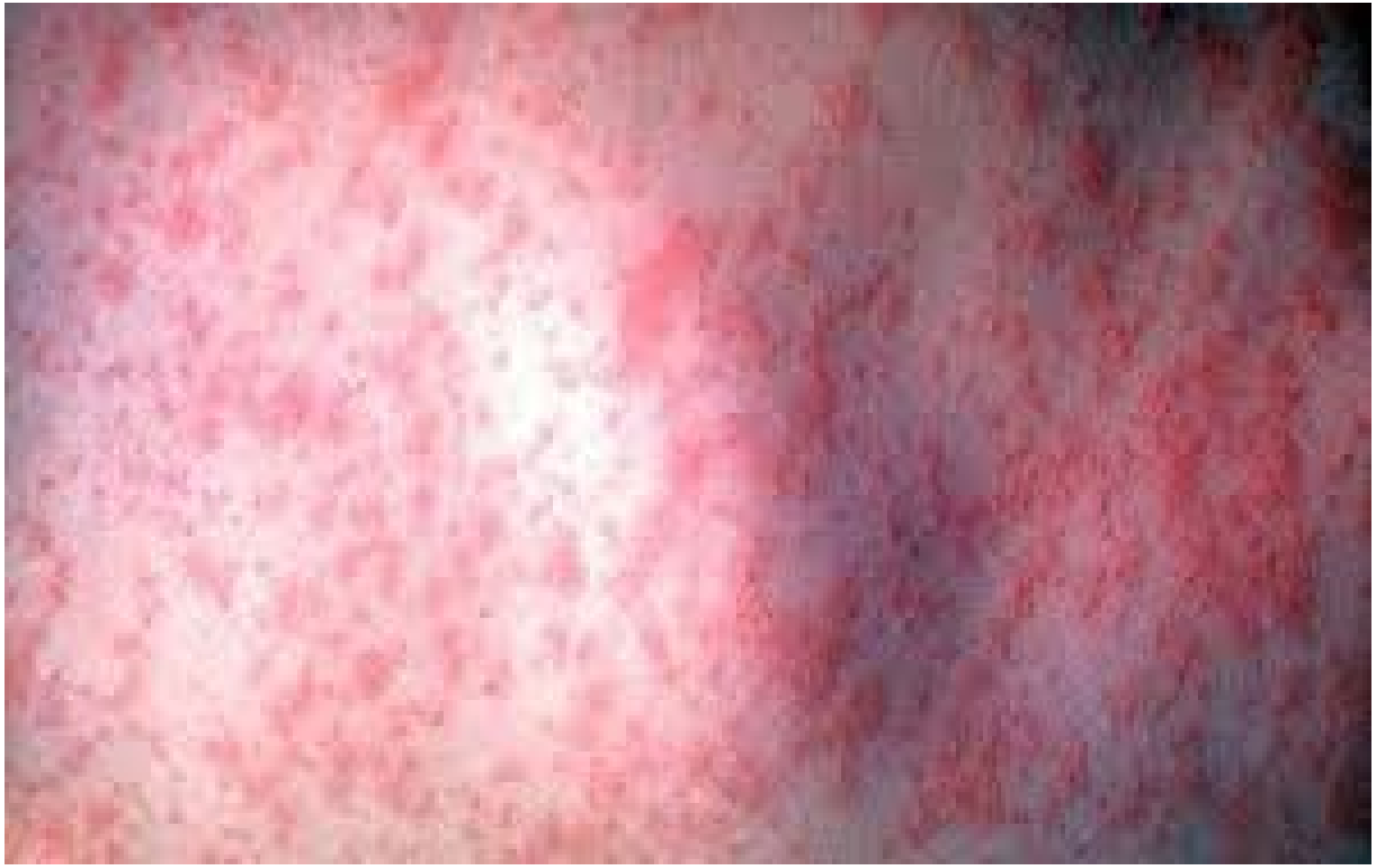
Ideally, …


P=0.05
Not so bad!
This is fantastic!
MID
Absolutely useless and wasteful!
Whatever! This is not meaningful!
Clinical Importance
Statistical Significance

The tragedy of poor research, and misguided focus on p-values can have serious, deleterious public health consequences

Measles and MMR Vaccine:

The controversy



CBCnews | British Columbia

LIVE Vancouver More Streams
CBC Radio One
Listen Live

Home | World | Canada | Politics | Business | Health | Arts & Entertainment | Technology & Science | Community | Weather | Video

Canada | BC | News | Events | Weather | Programs | Video | Audio | Contact

Measles outbreak provokes huge demand for vaccine in B.C.

Disneyland outbreak prompts concern in run up to Spring Break

Global NEWS

Toronto
Change Location

TV News Programs
Newscasts & Videos

World | Canada | Local | Politics | Money | Entertainment | Health | Lifestyle | Tech | Science | Sports | Vi

QUEBEC MEASLES OUTBREAK

February 21, 2015 9:36 pm

Updated: February 23, 2015 9:50

Officials confirm 19th case of measles in Quebec, 18th case confirmed in Ontario

The Controversy started in 1998 by Wakefield study published in Lancet, and later retracted

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associated by the parents, with measles, mumps, and rubella vaccination in eight of the 12 children, with measles infection in one child, and otitis media in another. All 12 children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation in the colon in 11 children and reactive ileal lymphoid hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls ($p=0.003$), low haemoglobin in four children, and low serum IgA in four children.

Interpretation We identified associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

Lancet 1998; **351**: 637–41
See Commentary page

RETRACTED

Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and bloating and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features of these children.

Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (diarrhoea, abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for 1 week, accompanied by their parents.

Clinical investigations

We took histories including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental histories included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and cerebrospinal-fluid lactate were measured to exclude known

The study claimed there was a causal link between MMR vaccine and development of autism

Autism problems

- ✓ Communicating (non)verbally
- ✓ Relating to others and the world around them
- ✓ Thinking and behaving flexibly

However, there was overwhelming evidence to the contrary

Systematic reviews of observational studies show no significant association between MMR an autism

Arch Pediatric Adolesc Med 2003;157:628-34

Safety of Vaccines Used for Routine Immunization of US Children: A Systematic Review

Pediatrics 2014;134:325–337

AUTHORS: Margaret A. Maglione, MPP,[a] Lopamudra Das, MPH,[a] Laura Raaen, MPH,[a] Alexandria Smith, MPH,[a] Ramya Chari, PhD,[a] Sydne Newberry, PhD,[a] Roberta Shanman, MLS,[a] Tanja Perry, BHM,[a] Matthew Bidwell Goetz, MD,[b] and Courtney Gidengil, MD, MPH[a,c]

abstract

FREE

BACKGROUND: Concerns about vaccine safety have led some parents to decline recommended vaccination of their children, leading to the

"Exposure to the MMR vaccine was unlikely to be associated with autism, ..."

CHRANE
ORATION®

What was wrong with the Wakefield study?

- ✓ **Small** sample size (n=12)
- ✓ **Uncontrolled** nature of the design
- ✓ **Speculative** nature of the conclusions—not supported by evidence
- ✓ **Serious ethical** violations and scientific misrepresentation
  - ❑ No consent or ethical clearance for conducting the research
  - ❑ Sampling was described as consecutive, while it was selective

By far, the most serious damage was the **p-value** of the test of association between MMR vaccine exposure (vs matched "selected" controls), and urinary methylmalonic acid

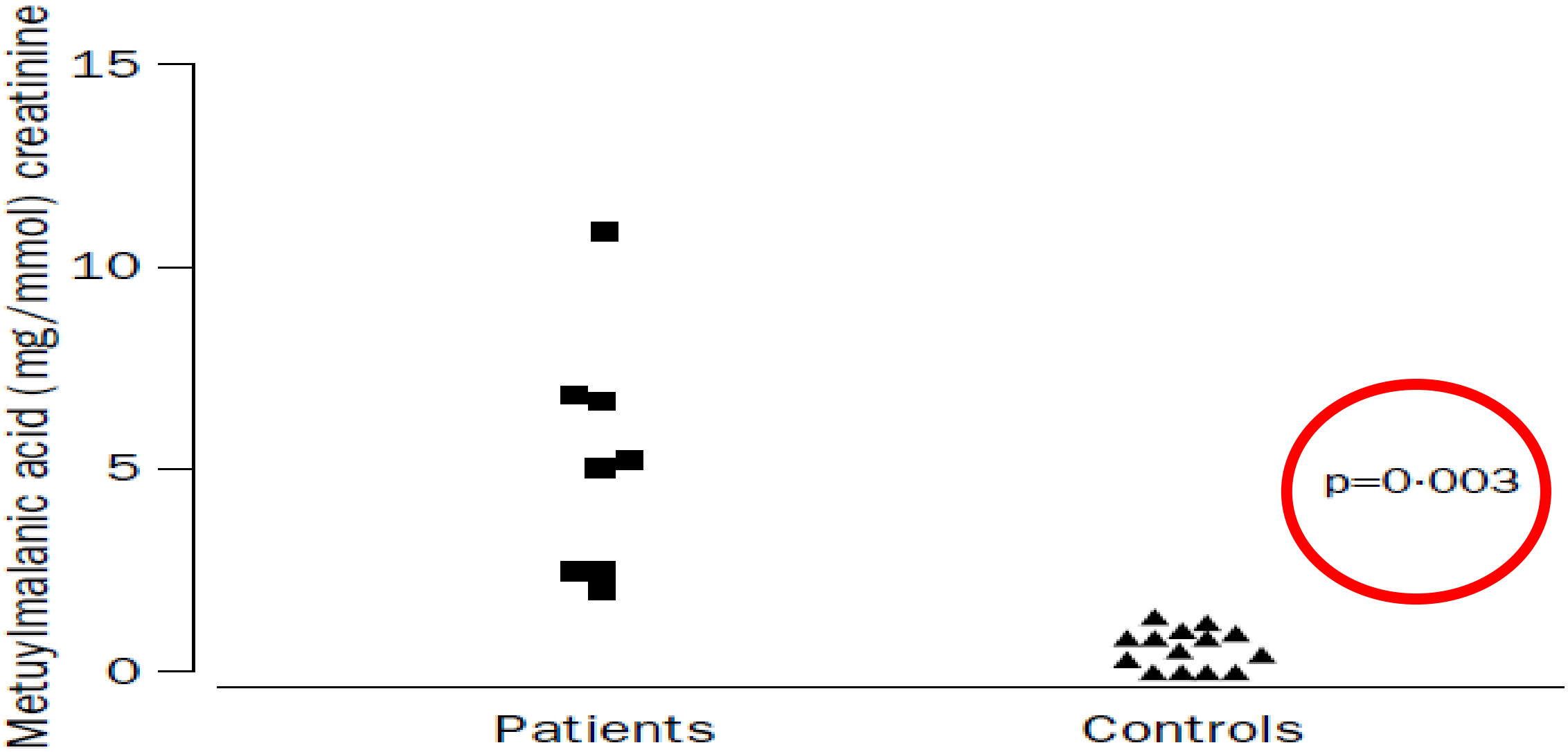


Figure 1: **Urinary methylmalonic-acid excretion in patients and controls**

p=Significance of mean excretion in patients compared with controls.

This was part of key findings in the abstract!

Findings Onset of behavioural symptoms was associa by the parents, with measles, mumps, and rub a vaccination in eight of the 12 children, with measl infection in one child, and otitis media in an r. All 12 children had intestinal abnormalities rangi from lymphoid nodular hyperplasia to a thoid u ration. Histology showed patchy chronic inflam tion i in 11 children and reactive ileal ympho perplasia in seven, but no granulomas. Be vioural diso rs included autism (nine), disintegrativ sy sis (one), a possible postviral or vaccinal encephalitis o). There were no focal neurological ab malities and I and EEG tests were normal. Abno al laboratory results re significantly raised urinary ethylmal c id compared with age matched contro (p= 03) children, a low s m IgA in r children.

Inter tation e iden associated gastrointestinal dis se and evelopmental regression in a group of previ sly mal chi n, which was generally associated in time possible environmental triggers.

Lancet 199 **351:** 637–41

RETRACTED

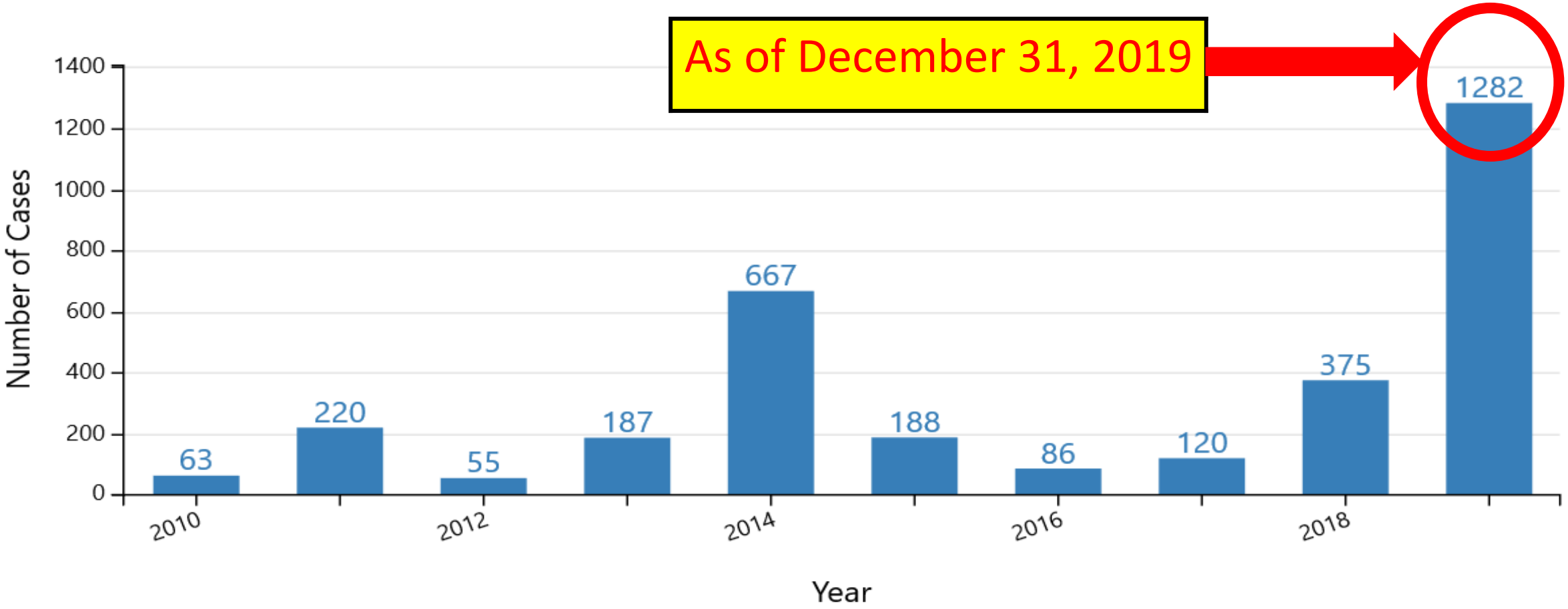
BUT, the seeds of

doubt and confusion

have already been planted—with serious implications for public health

The number of measles cases is on the rise in US because of social media fake news and mis-information—mostly likely spurred by the Wakefield's flawed study!!!

2010-2019**(as of December 31, 2019)



If you think the situation in USA is bad….

…across the Atlantic in the DRC, it is catastrophic!

Democratic Republic of Congo

A deadly measles outbreak is spreading like wildfire

The disease has spread to every part of DR Congo

More than 800,000 children are to be targeted for vaccination in the Democratic Republic of Congo, after a measles outbreak killed more than 3,500 people this year.

"As of 17 September, a total of **183,837 suspected measles cases (5,989 confirmed) had been reported in 192 of the 519 health zones nationwide, including 3,667 deaths** - which exceed the number of deaths due to Ebola. **Nearly all the deaths have been children**," **the WHO said in a statement**.

In the country's east, **Ebola has claimed more than 2,100 lives** since erupting in August last year, and is the second largest outbreak of the disease on record.

On January 7, 2020

https://reliefweb.int/report/democratic-republic-congo/deaths-democratic-republic-congo-measles-outbreak-top-

07 Jan 2020

Deaths from Democratic Republic of the Congo measles outbreak top 6000

REPORT from World Health Organization

Published on 07 Jan 2020

Kinshasa, 7 January 2020 - With the death toll from the world's worst measles epidemic in the Democratic Republic of the Congo (DRC) surpassing 6000, the World Health Organization (WHO) is calling for more funding to stop the outbreak.

The outbreak is driven mostly by lack of resources and limited access to care

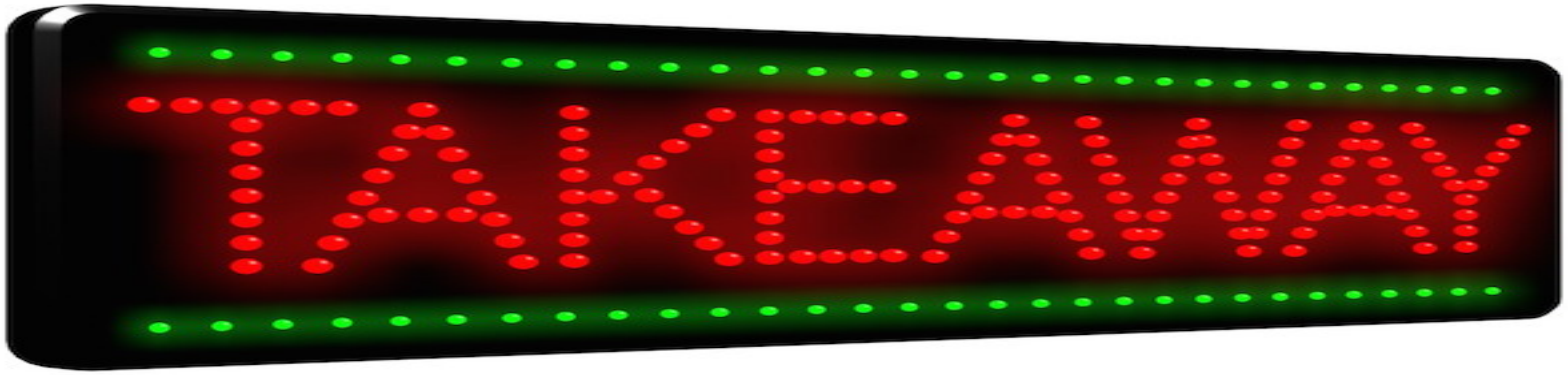
The world could be redirecting their focus on the DRC outbreak instead of fighting the ones in developed countries driven by fake news—
based on flawed p-values!!!

Check the CBC Marketplace for a story on the anti-vaccination movement

CBC | MENU

news | Top Stories | Local | The National | Opinion | World | Canada | Politics | Indigenous

Inside the anti-vaccination movement: CBC's Marketplace consumer cheat sheet



- Statistical significance
  - ✓ ≠ Clinical relevance
  - ✓ ≠ Personal significance
  - ✓ ≠ Patient-oriented
  - ✓ ≠ Patient importance
  - ✓ ≠ Humanly important
- There is a strong push to dethrone or get rid of "statistical significance" in clinical research
- More focus on CI in reporting and interpreting results—CIs give better information about other dimensions


thank you
merci
gracias
danke
謝謝
شكراً جزيلاً
ngiyabonga
teşekkür ederim
спасибо
рахмат
Баярлалаа
obrigado
dziękuję
bedankt
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
terima kasih
감사합니다
ευχαριστώ
merci
dank je
tapadh leat
mauruuru
hvala
sagolun
takk
dakujem
arigatô
grazie
tack
welalin
barka
mersi
spas
vinaka
blagodaram
misaotra
matondo
paldies
grazzi
mahalo
kiitos
dankie
faafetai lava
nandri
nanni
dhanyavad
köszönöm
akun
dankon
ačiū
gràcie
bayarlalaa
enkosi
djiere dieuf
tau
дякую
mamnun
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
sobodi
dēkuji
mèsi
didi madloba
najis tuke
ありがとう
tanemirt
rahmet
diolch
dhanyavadagalu
shukriya
mercè
мерси
trugarez
kam sah hamnida
rahmat
তোমাকে ধন্যবাদ
xièxie
хвала
asante
manana
obrigada
murakoze
tenki
chokrane