In search of useful clinical research
Αναζητώντας την χρήσιμη κλινική έρευνα

John P.A. Ioannidis

C.F. Rehnborg Professor in Disease Prevention
Professor of Medicine, of Health Research and Policy, of Biomedical Data Science, and of Statistics, Stanford University
Co-Director, Meta-Research Innovation Center at Stanford (METRICS)
Director, PhD program in Epidemiology and Clinical Research
PI, Wellness Living Laboratory
Hierarchies of evidence

MA, SR
RCT
...
...
Experts
Tweets
INVITED COMMENTARY

Evidence-based medicine has been hijacked: a report to David Sackett

John P.A. Ioannidis

Department of Medicine, Stanford Prevention Research Center, Stanford, CA 94305, USA
Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA 94305, USA
Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA 94305, USA
Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA 94305, USA

Accepted 18 February 2016; Published online xxxx

Abstract

This is a confession building on a conversation with David Sackett in 2004 when I shared with him some personal adventures in evidence-based medicine (EBM), the movement that he had spearheaded. The narrative is expanded with what ensued in the subsequent 12 years. EBM has become far more recognized and adopted in many places, but not everywhere, for example, it never acquired much influence in the USA. As EBM became more influential, it was also hijacked to serve agendas different from what it originally aimed for. Influential randomized trials are largely done by and for the benefit of the industry. Meta-analyses and guidelines have become a factory, mostly also serving vested interests. National and federal research funds are funneled almost exclusively to research with little relevance to health outcomes. We have supported the growth of principal investigators who excel primarily as managers absorbing more money. Diagnosis and prognosis research and efforts to individualize treatment have fueled recurrent spurious promises. Risk factor epidemiology has excelled in salami-sliced data-dredged articles with gift authorship and has become adept to dictating policy from spurious evidence. Under market pressure, clinical medicine has been transformed to finance-based medicine. In many places, medicine and health care are wasting societal resources and becoming a threat to human well-being. Science denialism and quacks are also flourishing and leading more people astray in their life choices, including health. EBM still remains an unmet goal, worthy to be attained. © 2016 Elsevier Inc. All rights reserved.
Key tools of non-evidence-based medicine in 1990 (splendidly serving vested interests)

• Ex cathedra pronouncements by prestigious opinion leaders in various conferences
• Editorials
• Non-systematic reviews
• Professional society guidelines done for the glory of the profession
• Pamphlets from drug reps
• Other marketing material disseminated in medical “scientific” meetings
Key tools of non-evidence-based medicine in 2018 (splendidly serving vested interests)

- Practice guidelines based on seemingly rigorous but still partisan processes
- Systematic reviews and meta-analyses
- Randomized trials
- Observational studies and risk factor epidemiology (meet bias: the works!)
- X-omics, predictive medicine, precision medicine, personalized medicine, individualized medicine (brave new worlds for a single person)
Evidence is less than optimal

Destroyed pyramid in Abu Rawash
Bulldozed pyramid by property developers in Peru
(money, money, money!)
Each specialty is trying to maximize its clients

- Disease mongering
- Broader definition of disease, illness, treatable range, e.g.
  - Hypercholesterolemia
  - Hypertension
  - Metabolic syndrome, prediabetes

Experts from different specialties treating the same condition rarely communicate
PDT - Cryo - C&D - L - Rad - SE - MMS

Placebo - Diclofen, Calcitriol, Both

SG - PEP005 - API31510 - LDE - IFN - 5-FU - CA - IMI - MMS

Kim, Tang, Ioannidis, J Clin Epi 2014
A sad realization

EBM is widely tolerated only when it can produce mostly boring evidence reports that can be endorsed by experts and/or serve vested interests of one or more clans.

The very same people who were previously spitting when mentioning “EBM”, are now using the very same term to buttress their eminence-based medicine claims to prestige by (mis)using the tools of EBM
How to survive the medical misinformation mess

John P. A. Ioannidis*,†,‡, Michael E. Stuart§,¶, Shannon Brownlee**,†† and Sheri A. Strite¶

*Departments of Medicine, Health Research and Policy, and Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA, †Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA, ‡Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA, §Department of Family Medicine, University of Washington School of Medicine, Seattle, WA, USA, ¶Delfini Group LLC, Seattle, WA, USA, **Lown Institute, Brookline, MA, USA, ††Department of Health Policy, Harvard T.H. Chan School of Public Health, Cambridge, MA, USA

1. Much published medical research is not reliable or is of uncertain reliability, offers no benefit to patients, or is not useful to decision makers.
2. Most healthcare professionals are not aware of this problem.
3. Even if they are aware of this problem, most healthcare professionals lack the skills necessary to evaluate the reliability and usefulness of medical evidence.
4. Patients and families frequently lack relevant, accurate medical evidence and skilled guidance at the time of medical decision-making.

• EJCI, 2017
How good is the quality of the clinical evidence?

• All 1394 systematic reviews published on the Cochrane Database of Systematic Reviews from January 2013 to June, 2014.
• GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) summary of findings performed in 608 (43.6%).
• Quality of the evidence for the first listed primary outcome: 13.5% high, 30.8% moderate, 31.7% low, 24% very low level.
• Even when all outcomes listed were considered, only 19.1% had at least one outcome with high quality of evidence.
• Of the reviews with high quality of evidence, only 25 had both significant results and a favorable interpretation of the intervention.

Fleming et al, J Clin Epidemiol 2016
Significance of the evidence?

- Almost all scientific papers claim that they have found (statistically and/or conceptually) significant results.
- Obviously all my grant proposals claim that what I am planning to do is highly significant (although I mostly submit mediocre ideas for funding).
- Among abstracts with P-values in Medline (1990-2015), 96% report statistically significant results.
Scientific discovery has become a boring nuisance: 96% of the biomedical literature claims significant results.
Almost any result can be obtained: Vibration of effects and the Janus phenomenon
Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

Context Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

Objectives To understand how frequently highly cited studies are contradicted or find effects that are stronger than in other similar studies and to discern whether specific characteristics are associated with such refutation over time.

Design All original clinical research studies published in 3 major general clinical journals or high-impact-factor specialty journals in 1990-2003 and cited more than 1000 times in the literature were examined.

Main Outcome Measure The results of highly cited articles were compared against subsequent studies of comparable or larger sample size and similar or better controlled designs. The same analysis was also performed comparatively for matched studies that were not so highly cited.

Results Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials (P = .008). Among randomized trials, studies with contradicted or stronger effects were smaller (P = .009) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with “negative” results.

Conclusions Contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes. The extent to which high citations may provoke contradictions and vice versa needs more study. Controversies are most common with highly cited nonrandomized studies, but even the most highly cited randomized trials may be challenged and refuted over time, especially small ones.

JAMA. 2005;294:218-228

www.jama.com
Many treatment effects seem to be large, especially in small, early trials, but few survive scrutiny.
Evolution of very large treatment effects

Group A

Group B

Pereira, Horwitz, Ioannidis, JAMA 2012
Treatment effects of studies with “quality” problems may be exaggerated.
Much clinical research nowadays is done outside the USA and Europe, but are the results unbiased?
Fig 4 Case fatality with calcium antagonists in aneurysmal subarachnoid haemorrhage
**Participating Who Reported Taking Assigned Spironolactone or Placebo**

<table>
<thead>
<tr>
<th>Country</th>
<th>Placebo</th>
<th>Spironolactone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russia (N=160)</td>
<td>91/105</td>
<td>66/70</td>
</tr>
<tr>
<td>United States and Canada (N=206)</td>
<td>82/90</td>
<td>76/101</td>
</tr>
</tbody>
</table>

**Participants Who Reported Taking Spironolactone but Had No Detectable Canrenone Concentration**

<table>
<thead>
<tr>
<th>Country</th>
<th>Participants (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russia (N=66)</td>
<td>30</td>
</tr>
<tr>
<td>United States and Canada (N=76)</td>
<td>3</td>
</tr>
</tbody>
</table>

**Median Canrenone Concentration among Participants Who Reported Taking Spironolactone**

- **Russia**
  - Spearman correlation: 0.09 (P=0.47)
- **United States and Canada**
  - Spearman correlation: 0.43 (P<0.001)

<table>
<thead>
<tr>
<th>Spironolactone Dose Reported by Participant (mg)</th>
<th>Russia (N=3)</th>
<th>United States and Canada (N=25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>N=3</td>
<td>N=16</td>
</tr>
<tr>
<td>30</td>
<td>N=35</td>
<td>N=26</td>
</tr>
<tr>
<td>45</td>
<td>N=26</td>
<td>N=16</td>
</tr>
</tbody>
</table>

**Mean Change in Serum Potassium Level from Baseline to 12 mo**

- Participants with Detectable Canrenone Concentration (N=119): 0.34 (95% CI, 0.24 to 0.43) P<0.001 vs. baseline
- Participants with Undetectable Canrenone Concentration (N=21): 0.03 (95% CI, -0.29 to 0.35) P=0.36 vs. baseline

**Mean Change in Aldosterone Level from Baseline to 12 mo**

- Participants with Detectable Canrenone Concentration (N=116): 66 (95% CI, 48 to 84) P<0.001 vs. baseline
- Participants with Undetectable Canrenone Concentration (N=20): -49 (95% CI, -89 to -9) P=0.02 vs. baseline
Some types of clinical trials almost always favor the sponsor:

• Among trials published in 2011, 55/57 of non-inferiority trials with head to head comparisons sponsored by the industry demonstrated non-inferiority

• Success rate > 96%

• Flacco et al. JCE 2015
Re-analysis: can we trust the data?

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

Joanna Le Noury,1 John M Nardo,2 David Healy,1 Jon Jureidini,3 Melissa Raven,3 Catalin Tufanaru,4 Elia Abi-Jaoude5

ABSTRACT
OBJECTIVES
To reanalyse SmithKline Beecham’s Study 329 (published by Keller and colleagues in 2001), the primary objective of which was to compare the efficacy and safety of paroxetine and imipramine with placebo in the treatment of adolescents with unipolar major depression. The reanalysis under the restoring invisible and abandoned trials (RIAT) initiative was done to see whether access to and reanalysis of a full dataset from a randomised controlled trial would have clinically relevant implications for evidence based medicine.

DESIGN
Double blind randomised placebo controlled trial.

SETTING

PARTICIPANTS
275 adolescents with major depression of at least eight weeks in duration. Exclusion criteria included a range of comorbid psychiatric and medical disorders and suicidality.

INTERVENTIONS
Participants were randomised to eight weeks double blind treatment with paroxetine (20-40 mg), imipramine (200-300 mg), or placebo.

MAIN OUTCOME MEASURES
The prespecified primary efficacy variables were change from baseline to the end of the eight week acute treatment phase in total Hamilton depression scale (HAM-D) score and the proportion of responders (HAM-D score ≤8 or ≥50% reduction in baseline HAM-D) at acute endpoint. Prespecified secondary outcomes were changes from baseline to endpoint in depression items in K-SADS-L, clinical global impression, autonomous functioning checklist, self-perception profile, and sickness impact scale; predictors of response; and number of patients who relapse during the maintenance phase. Adverse experiences were to be compared primarily by using descriptive statistics. No coding dictionary was prespecified.

RESULTS
The efficacy of paroxetine and imipramine was not statistically or clinically significantly different from placebo for any prespecified primary or secondary efficacy outcome. HAM-D scores decreased by 10.7 (least squares mean) (95% confidence interval 9.1 to 12.3), 9.0 (7.4 to 10.5), and 9.1 (7.5 to 10.7) points, respectively, for the paroxetine, imipramine and placebo groups (P=0.20). There were clinically significant increases in harms, including suicidal ideation and behaviour and other serious adverse events in the paroxetine group and cardiovascular problems in the imipramine group.

CONCLUSIONS
Neither paroxetine nor high dose imipramine showed efficacy for major depression in adolescents, and there was an increase in harms with both drugs. Access to primary data from trials has important implications for both clinical practice and research, including that published conclusions about efficacy and safety should not be read as authoritative. The reanalysis of Study 329 illustrates the necessity of making primary trial data and protocols available to increase the rigour of the evidence base.
Reanalyses of Randomized Clinical Trial Data

Shanil Ebrahimi, PhD; Zahra N. Sohani, MSc; Luis Montoya, DDS; Arnav Agarwal, BSc; Kristian Thorlund, PhD; Edward J. Mills, PhD; John P. A. Ioannidis, MD, DSc

**Importance**  Reanalyses of randomized clinical trial (RCT) data may help the scientific community assess the validity of reported trial results.

**Objectives**  To identify published reanalyses of RCT data, to characterize methodological and other differences between the original trial and reanalysis, to evaluate the independence of authors performing the reanalyses, and to assess whether the reanalysis changed interpretations from the original article about the types or numbers of patients who should be treated.

**Design**  We completed an electronic search of MEDLINE from inception to March 9, 2014, to identify all published studies that completed a reanalysis of individual patient data from previously published RCTs addressing the same hypothesis as the original RCT. Four data extractors independently screened articles and extracted data.

**Main Outcomes and Measures**  Changes in direction and magnitude of treatment effect, statistical significance, and interpretation about the types or numbers of patients who should be treated.

**Results**  We identified 37 eligible reanalyses in 36 published articles, 5 of which were performed by entirely independent authors (2 based on publicly available data and 2 on data that were provided on request; data availability was unclear for 1). Reanalyses differed most commonly in statistical or analytical approaches (n = 18) and in definitions or measurements of the outcome of interest (n = 12). Four reanalyses changed the direction and 2 changed the magnitude of treatment effect, whereas 4 led to changes in statistical significance of findings. Thirteen reanalyses (35%) led to interpretations different from that of the original article, 3 (8%) showing that different patients should be treated; 1 (3%), that fewer patients should be treated; and 9 (24%), that more patients should be treated.

**Conclusions and Relevance**  A small number of reanalyses of RCTs have been published to date. Only a few were conducted by entirely independent authors. Thirty-five percent of published reanalyses led to changes in findings that implied conclusions different from those of the original article about the types and number of patients who should be treated.
46% retrieval rate for raw data of randomized trials under full data sharing policy

Naudet et al, BMJ 2018
RCTs versus studies with routinely collected data

Hemkens, Contopoulos-Ioannidis, Ioannidis, BMJ 2015
The systematic review and meta-analysis epidemic
The meta-pie
(see Ioannidis, Milbank Quarterly 2016)

Currently produced meta-analyses

- Unpublished
- Redundant and unnecessary
- Decent, but not useful
- Misleading, abandoned genetics
- Flawed beyond repair
- Decent and clinically useful
Guidelines as a marketing tool and as a potential threat to patients

Ensuring the integrity of clinical practice guidelines: a tool for protecting patients
Jeanne Lenzer, Jerome Hoffman, Curt Furberg, and John Ioannidis pull together a large expert working group to offer a manifesto for clinical guidelines

Box 1: Red flags that should raise substantial skepticism among guideline readers (and medical journals)
• Sponsor(s) is a professional society that receives substantial industry funding;
• Sponsor is a proprietary company, or is undeclared or hidden
• Committee chair(s) have any financial conflict*
• Multiple panel members have any financial conflict*
• Any suggestion of committee stacking that would pre-ordain a recommendation regarding a controversial topic
• No or limited involvement of an expert in methodology in the evaluation of evidence
• No external review
• No inclusion of non-physician experts/patient representative/community stakeholders
*Includes a panelist with either or both a financial relationship with a proprietary healthcare company and/or whose clinical practice/specialty depends on tests or interventions covered by the guideline.
8 of the 15 most-cited papers across medicine in 2016 are (cardiology) guidelines.

19 of the 20 most-cited papers of European Heart Journal in the last decade are guidelines and 1 is a disease definition (causing 10-fold increase in impact factor).

Industry support: >$200M per year to AHA; 77% of ESC budget.
Classics of the most-cited cardiologists:
164 guidelines
24 industry trials
11 other articles
(3 partially industry funded)
What EBM should be

- “Evidence-based medicine is about integrating individual clinical expertise with the best external evidence”

What it actually is

- Clinicians and clinician researchers (a joke?) are under tremendous market pressure.
- Most discussions in department meetings are about money.
- This is mostly finance-based medicine.
Clinicians are incentivized

• to capture the largest possible market share
• to satisfy customers in their shop
• to get high satisfaction scores
• to charge more for their services
• to perform more procedures
• to tick off more items on charge forms
A logical consequence

- How likely is it that physicians will design studies that threaten their jobs, decreasing the procedures, testing and interventions they do?
- Is EBM doomed to be heartily accepted only when it leads to more medicine, even if this means less health?
- Would cardiology professional societies urge their members to quit cardiology and change jobs if it were proven that their services are a waste and a threat to public health?
Concluding comments

- EBM has acquired tremendous power, but it has been hijacked by expert-based medicine masquerading as EBM.
- There is very little useful clinical research and tons of noise or prestigious nonsense (e.g. guidelines).
- Most of the medical evidence is either problematic/spurious/false or has no utility for medical and shared decision making.
- The main utility of systematic reviews and meta-analyses has been to reveal problems with the biomedical evidence.
- Expectations of replacing experimental (randomized) evidence with non-randomized data need to be tempered.
- Conflicts of interpretation need to be minimized for primary trials, meta-analyses and guidelines.
- Professional societies should be radically reformed or dismantled.
Special thanks

Daniele Fanelli
Steve Goodman
Shanil Ebrahim
Despina Contopoulos-Ioannidis
Georgia Salanti
Chirag Patel
Lars Hemkens
Ann Hsing
Lamberto Manzoli
Maria Elena Flacco
George Siontis
Denes Szucs
Kostas Siontis
Vangelis Evangelou
Kristin Sainani
Muin Khoury
Orestis Panagiotou
Florence Bourgeois
Special thanks

Joseph Lau
Malcolm MacLeod
Marcus Munafo
David Allison
Josh Wallach
Fotini Karassa
Athina Tatsioni
Evi Ntzani
Ioanna Tzoulaki
Demos Katritsis
Nikos Patsopoulos
Fainia Kavvoura
Brian Nosek
Victoria Stodden
Ele Zeggini
Belinda Burford
Kostas Tsilidis
Jodi Prochaska
Special thanks

Charitini Stavropoulou
Evropi Theodoratou
Nikos Pandis
Huseyin Naci
Vanesa Bellou
Antony Doufas
Lazaros Belbasis
Chris Doucouliagos
Stelios Serghiou
Anna Chaimani
Fotini Chatzinasio
Stephania Papatheodorou
Florian Naudet
Tom Hardwicke
Perrine Janiaud
Ioana-Alina Cristea
Shannon Brownlee
Vikas Saini
Matthias Egger
Patrick Bossuyt
Andre Uitterlinden
Doug Altman
Deb Zarin
Katherine Flegal
Special thanks

Shanthi Kapaggoda
Ewoud Schuit
Stefania Boccia
David Chavalarias
Jennifer Ware
Viswam Nair
Stephan Bruns
Dorothy Bishop
Tom Trikalinos
Kristina Sundquist
Johanna Int’hout
Kevin Boyack
Brett Thombs
Raj Manrai
Nazmus Saquib
Elizabeth Iorns
Abraham Verghese
Euan Ashley