Overcoming Challenges to Ensure the Truthworthiness and Impact of Evidence Synthesis

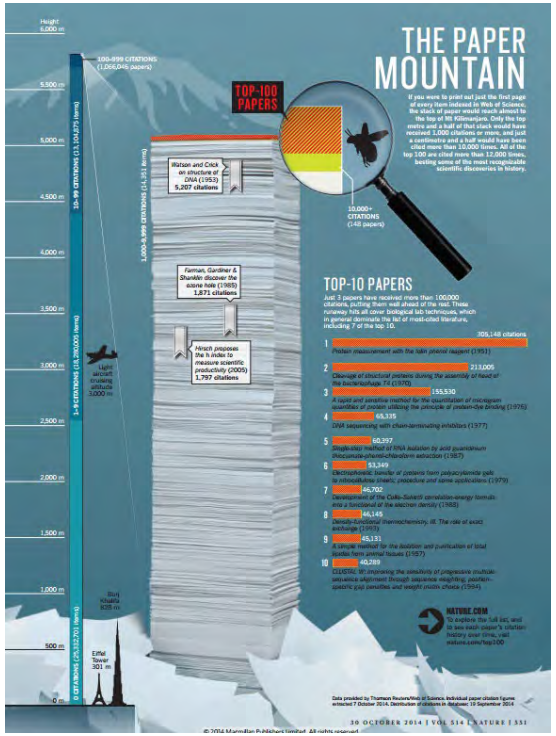
Isabelle Boutron

Cochrane France
Centre for Research in Epidemiology and Statistics (CRESS)
Université Paris Cité

Trustworthy evidence synthesis to support decision making

Medical doctors, health care providers, public health decision makers are overwhelmed with the amount of new information produced

- 37 millions articles in the field of life science and medicine
- The amount of information doubles every 3 years
- About 900 000 clinical trials indexed in PubMed
- 35 000 new clinical trials published every year



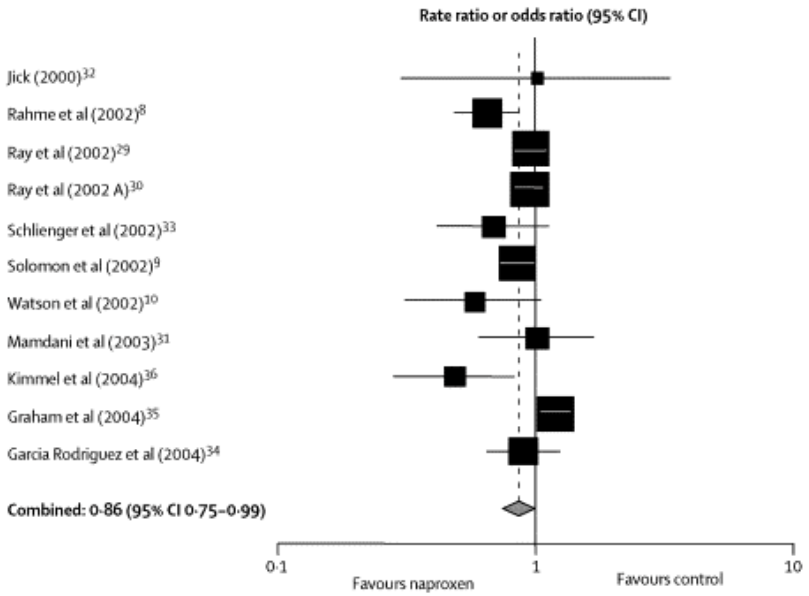
If you were to print out just the first page of every item indexed in Web of Science, **the stack of paper would reach almost to the top of Mt Kilimanjaro**

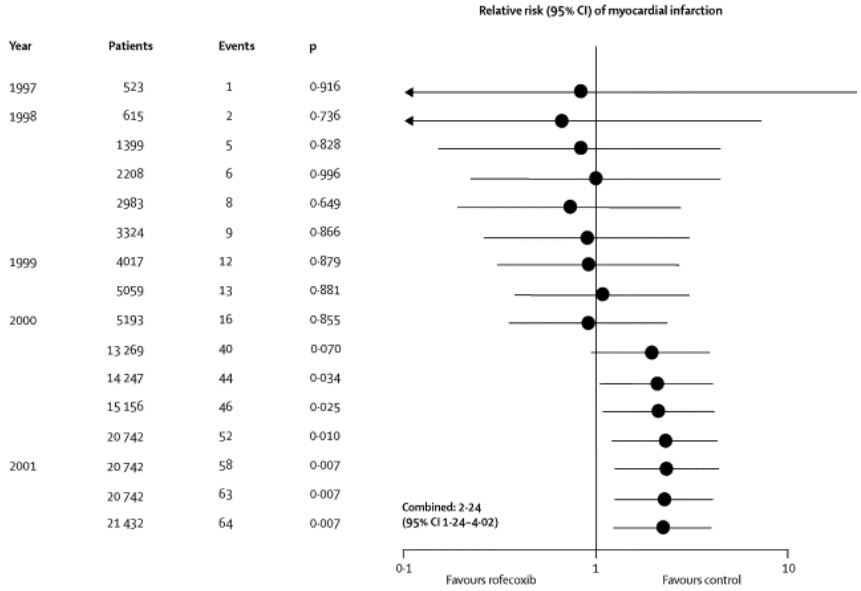
There is a crucial need for a reliable, independent, accessible and up-to-date synthesis of the knowledge produced by research

Trustworthy evidence synthesis to support decision making

An accurate, concise and unbiased synthesis of the available evidence is arguably one of the most valuable contributions a research community can offer decision-makers (Donnelly, Nature 2021)

Rofecoxib example
2004 drug withdrawal





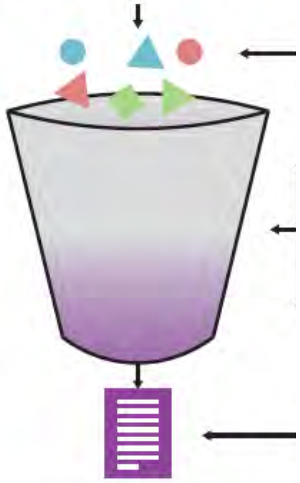
Jüni P, et al. Lancet. 2004
Boutron I, Créquit P, (…), Ravaud P. **J Clin Epidemiol**. 2020
Créquit P, Boutron I, (…), Ravaud P. **J Clin Epidemiol**. 2020
Ravaud P, Créquit P, (…) Boutron I. **J Clin Epidemiol**. 2020

Evidence synthesis methods to increase trustworthiness

- **A strong process** is implemented to improve untrustworthiness
  - Protocol, registration
  - Identification of all information related to primary studies (registry data, protocol, SAP, publication, Clinical study reports, etc)
  - Standardized extraction of all information

Evidence synthesis

Definition of the research question



Searching studies

Data extraction
Risk of bias assessment
Meta-analysis
Grading

Report of a systematic review

- **Specific tools** have been developed to ensure trustworthiness
  - Risk of bias tool (RoB 2)
  - Tool for addressing reporting bias (RoB-ME)
  - Tool for addressing conflicts of interest in trials (TACIT)
  - Tool to identify problematic trials
    - INSPECT-SR to identify problematic trials
    - Research Integrity Assessment (RIA) tool
    - Trustworthiness in Randomized Controlled Trials (TRACT) screening tool
    - The IPD Integrity tool

Evidence synthesis methods to increase trustworthiness

- Sources of information: clinical trial registries

PLOS | MEDICINE **Timing and Completeness of Trial Results Posted at ClinicalTrials.gov and Published in Journals**

202 published RCTs with posted results

Completeness of reporting	ClinicalTrials.gov N=202	Published article N=202	P-value
Flow of participants	64%	48%	<0.001
Efficacy results	79%	69%	0.02
Adverse events	73%	45%	<0.001
Serious adverse events	99%	63%	<0.001

Riveros et al. Plos Med 2013

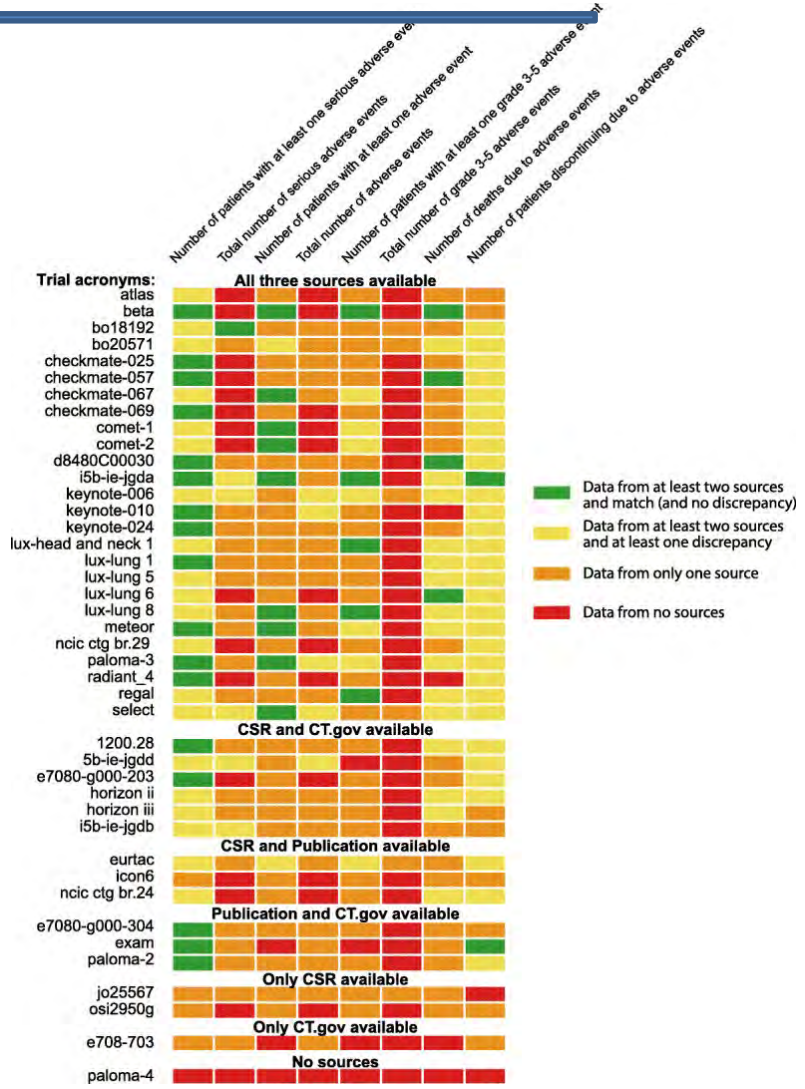
Evidence synthesis methods to increase trustworthiness

Other sources of information: Clinical study reports (CSR)

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

Clinical data publication

- 42 trials (2015-18)
- Harms reporting was more complete in CSRs than other sources.
- Marked discrepancies in harms data between sources



Paludan-Müller AS, Créquit P, Boutron I. BMC Med. 2021

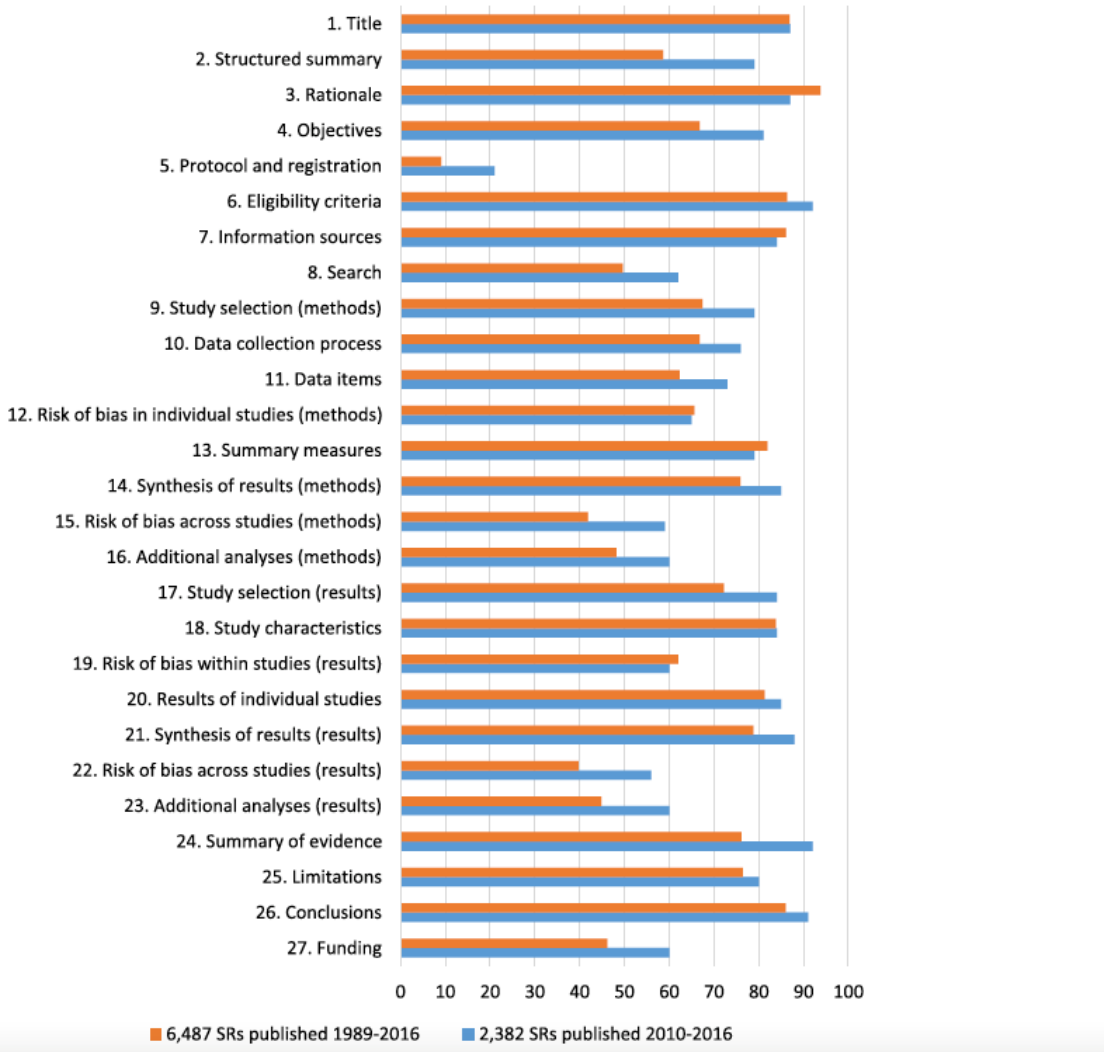
Transparency of evidence synthesis

Reporting guidelines

1999: QUOROM (QUality Of Reporting Of Meta-analyses) Statement

2009: PRISMA Statement
7 extensions to the PRISMA

- 2012 Equity
- 2013 Abstracts
- 2015 Network Meta-analyses
- 2015 Individual Participant Data
- 2015 Protocols
- 2016 Harms
- 2017 Complex interventions



Completeness of reporting improved but remained insufficient

Transparency of evidence synthesis

PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews

Matthew J Page,[1] David Moher,[2] Patrick M Bossuyt,[3] Isabelle Boutron,[4] Tammy C Hoffmann,[5] Cynthia D Mulrow,[6] Larissa Shamseer,[7] Jennifer M Tetzlaff,[8] Elie A Akl,[9] Sue E Brennan,[1] Roger Chou,[10] Julie Glanville,[11] Jeremy M Grimshaw,[12] Asbjørn Hróbjartsson,[13] Manoj M Lalu,[14] Tianjing Li,[15] Elizabeth W Loder,[16] Evan Mayo-Wilson,[17] Steve McDonald,[1] Luke A McGuinness,[18] Lesley A Stewart,[19] James Thomas,[20] Andrea C Tricco,[21] Vivian A Welch,[22] Penny Whiting,[18] Joanne E McKenzie[1]

For numbered affiliations see end of the article.

Correspondence to: M Page matthew.page@monash.edu (ORCID 0000-0002-4242-7526)

The methods and results of systematic reviews should be reported in sufficient detail to allow users to assess the trustworthiness and applicability of the

makers, who would otherwise be confronted by an overwhelming volume of research on which to base their decisions. To allow decision makers to assess the trustworthiness and applicability of review findings, reports of systematic reviews should be transparent

PRISMATIC Project

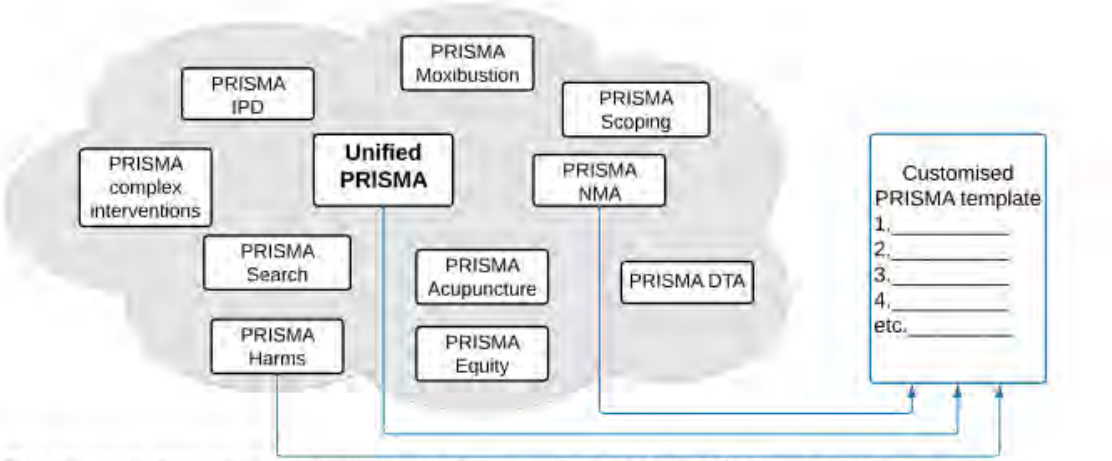


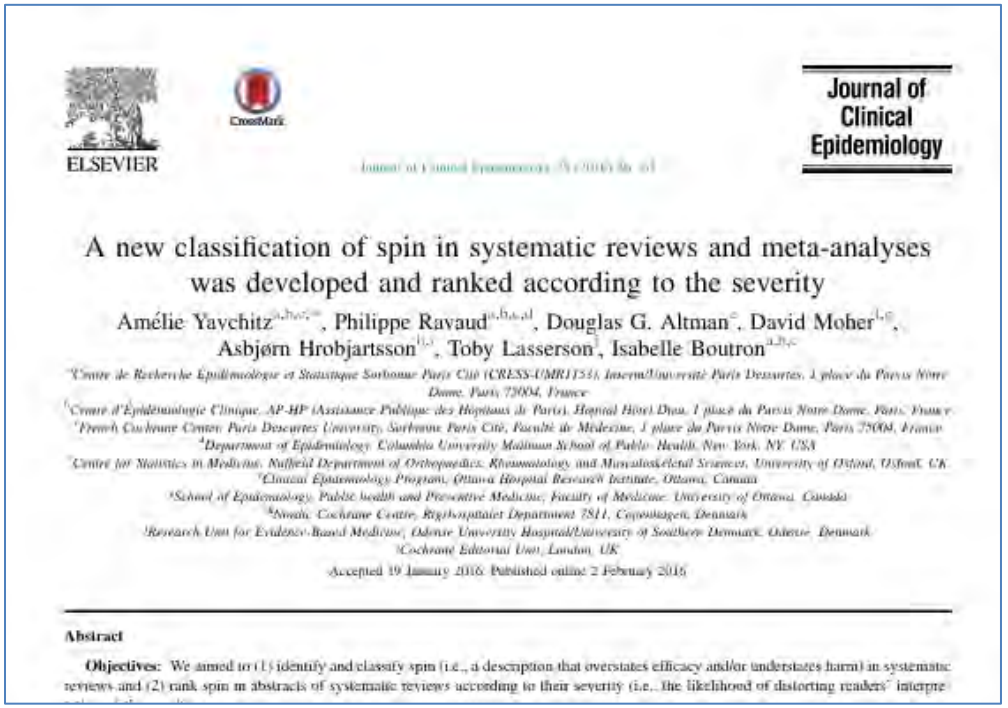
Fig. 1 Model of the PRISMA Web app. In the example depicted, a customised checklist and template for a systematic review with network meta-analysis of adverse events associated with COVID-19 vaccines are created. The items are drawn from three statements—the unified PRISMA statement, PRISMA Harms extension [15] and PRISMA NMA extension [7]. Adapted from Hopewell et al. [28]

Accuracy and spin in evidence synthesis

Spin is defined as specific intentional or unintentional reporting that fails to faithfully reflect the findings and could affect the impression the results produce in readers

Distortion of results interpretation

25% SR reported with spin



ELSEVIER

Journal of Clinical Epidemiology

A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity

Amélie Yavchitz, Philippe Ravaud, Douglas G. Altman, David Moher, Asbjørn Hrobjartsson, Toby Lasserson, Isabelle Boutron

Accepted 19 January 2016; Published online 2 February 2016

Abstract

Objectives: We aimed to (1) identify and classify spin (i.e., a description that overstates efficacy and/or understates harm) in systematic reviews and (2) rank spin in abstracts of systematic reviews according to their severity (i.e., the likelihood of distorting readers' interpre

META-RESEARCH ARTICLE

'Spin' in published biomedical literature: A methodological systematic review

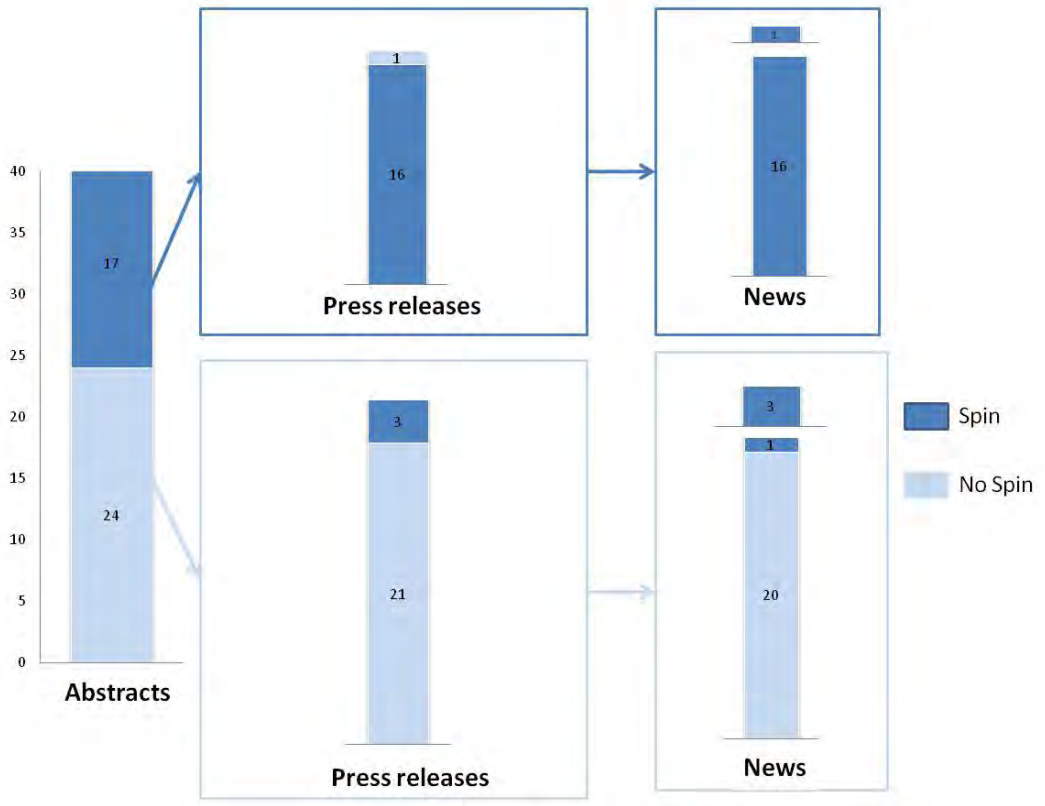
Kellia Chiu, Quinn Grundy, Lisa Bero*

Charles Perkins Centre, Faculty of Pharmacy, The University of Sydney, Sydney, New South Wales, Australia

* lisa.bero@sydney.edu.au

Accuracy and spin in evidence synthesis

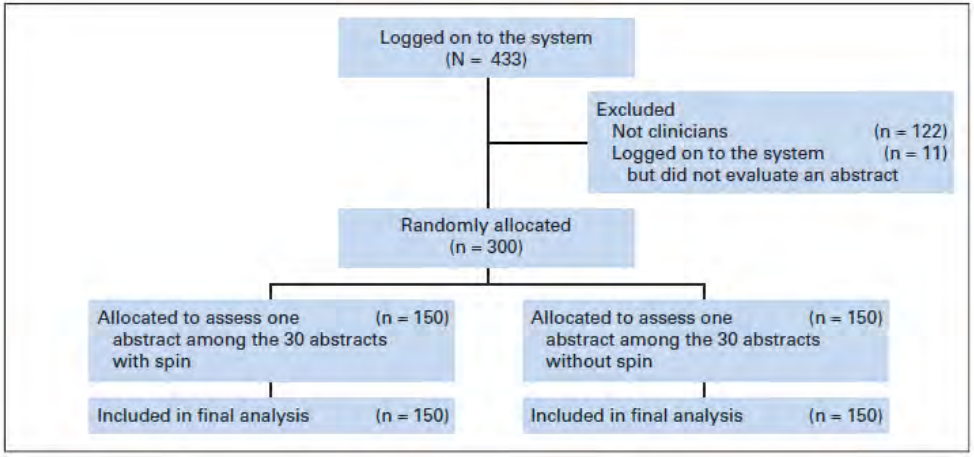
Spin disseminates from the publication to the press release and the news



Yavchitz A, Boutron I, Bafeta A, Marroun I, Charles P, Mantz J, Ravaud P. PLoS Med. 2012

Accuracy and spin in evidence synthesis

Spin impacts readers' interpretation



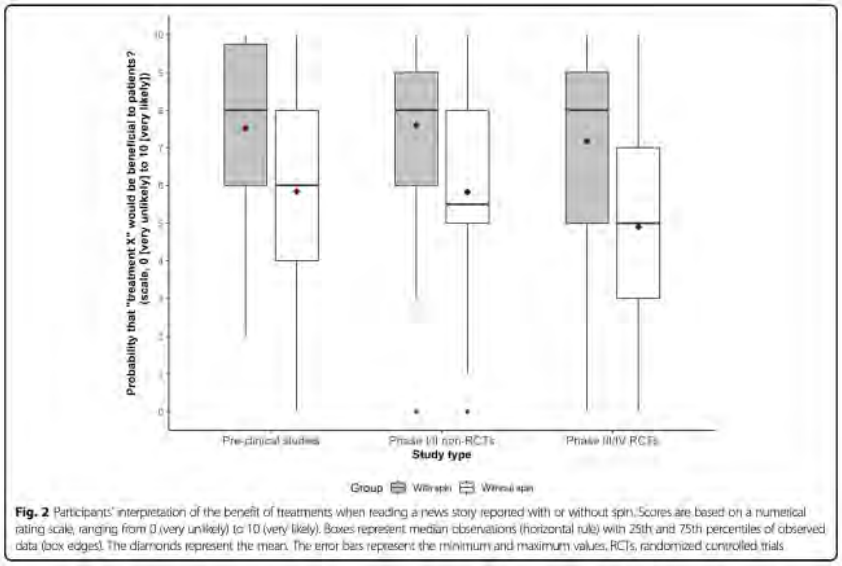


Fig. 2 Participants' interpretation of the benefit of treatments when reading a news story reported with or without spin. Scores are based on a numerical rating scale, ranging from 0 (very unlikely) to 10 (very likely). Boxes represent median observations (horizontal rule) with 25th and 75th percentiles of observed data (box edges). The diamonds represent the mean. The error bars represent the minimum and maximum values. RCTs, randomized controlled trials

Based on this abstract, do you think treatment A would be beneficial to patients?

Scale, 0 [very unlikely] to 10 [very likely)

Mean difference = 0.71 (95% CI, 0.07-1.35);
P .03

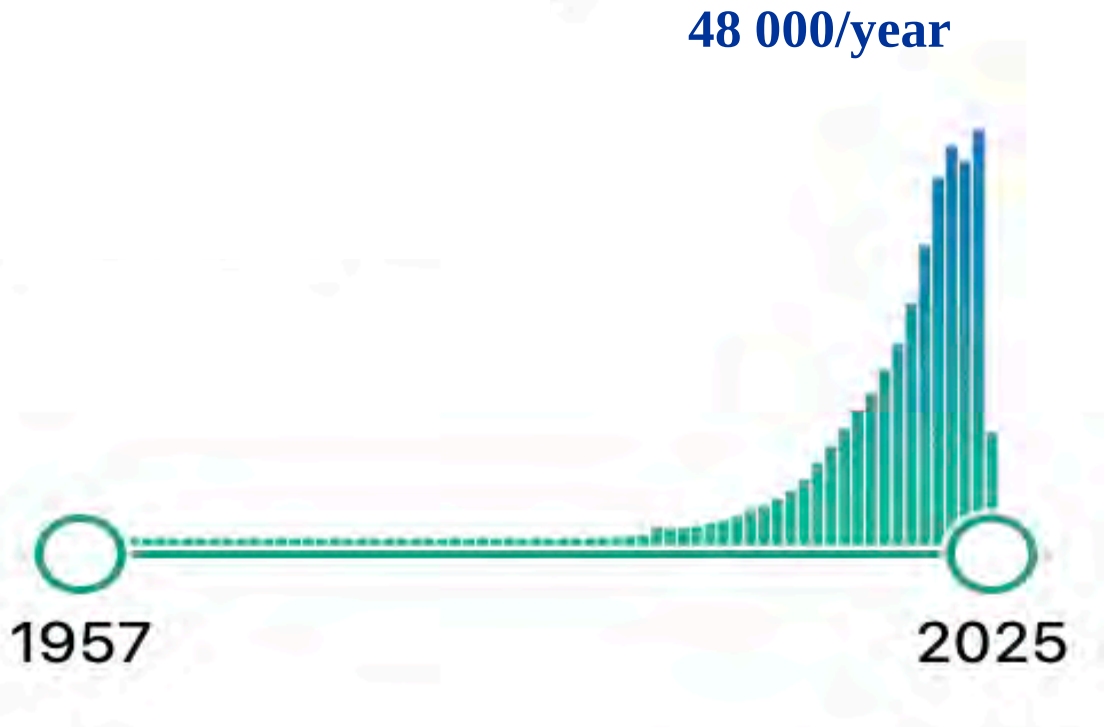
"What do you think is the probability that 'treatment X' would be beneficial to patients?" (scale, 0 [very unlikely] to 10 [very likely])

Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. J Clin Oncol. 2014
Boutron I, Haneef R, Yavchitz A, Baron G, Novack J, Oransky I, Schwitzer G, Ravaud P. BMC Med. 2019

Trustworthy evidence synthesis to support decision making

Rapid growth in the publication of systematic reviews

PubMed Search for
systematic reviews



Trustworthy evidence synthesis to support decision making

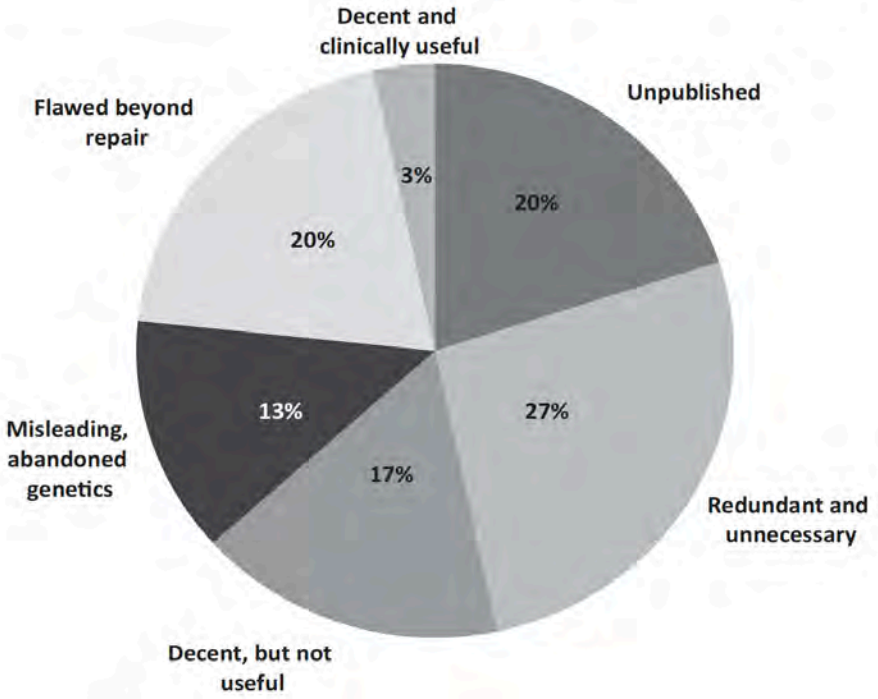
THE MILBANK QUARTERLY
A MULTIDISCIPLINARY JOURNAL OF POPULATION HEALTH AND HEALTH POLICY

Original Investigation

The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses

JOHN P.A. IOANNIDIS

Figure 4. A Summary Overview of Currently Produced Meta-analyses



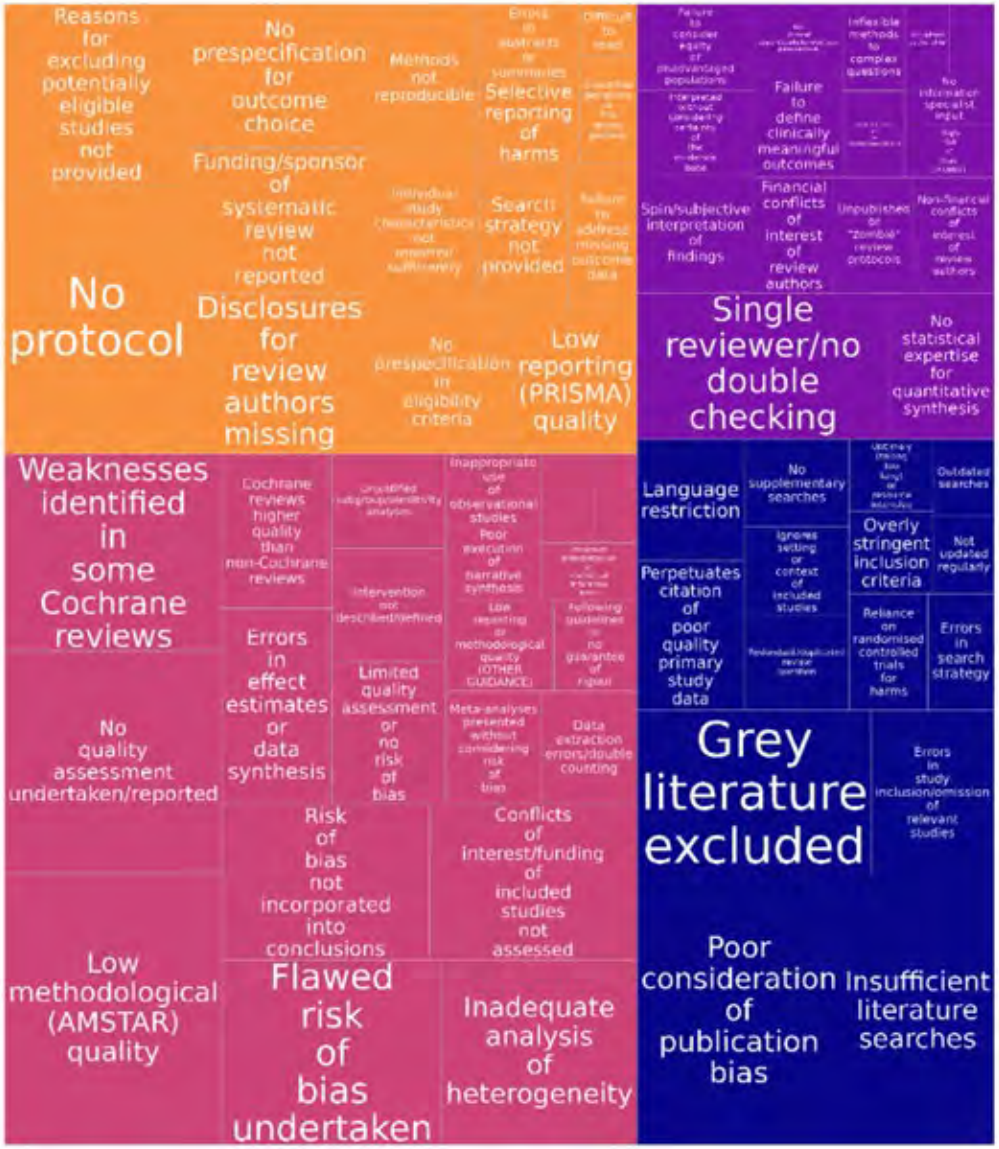
Trustworthy evidence synthesis to support decision making

The problems with systematic reviews: a living systematic review

Lesley Uttley[a,*], Daniel S. Quintana[b,c,d], Paul Montgomery[e], Christopher Carroll[a], Matthew J. Page[f], Louise Falzon[a], Anthea Sutton[a], David Moher[g]

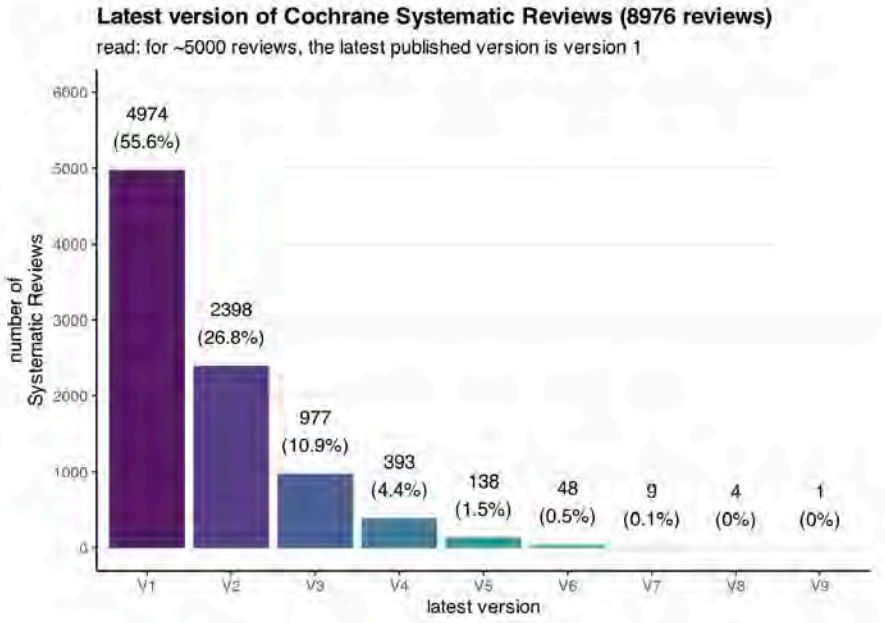
Journal of Clinical Epidemiology

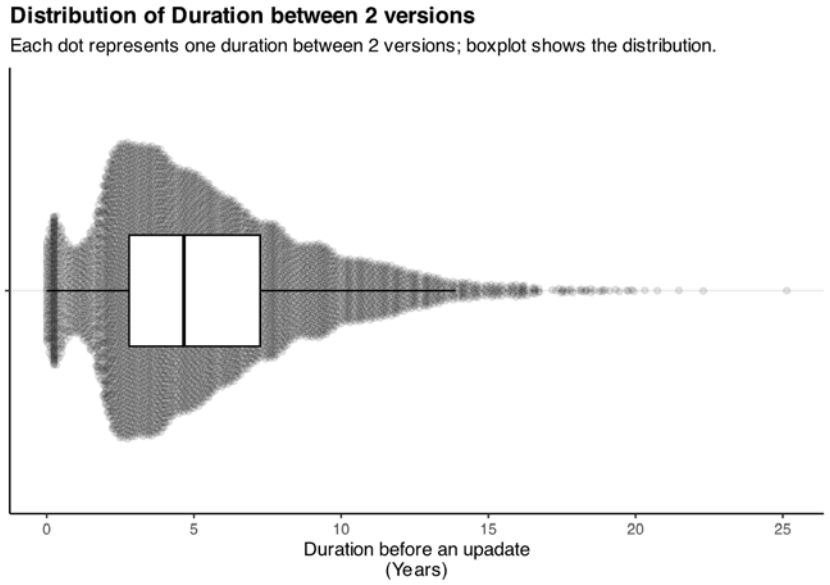
2023



Trustworthy evidence synthesis to support decision making

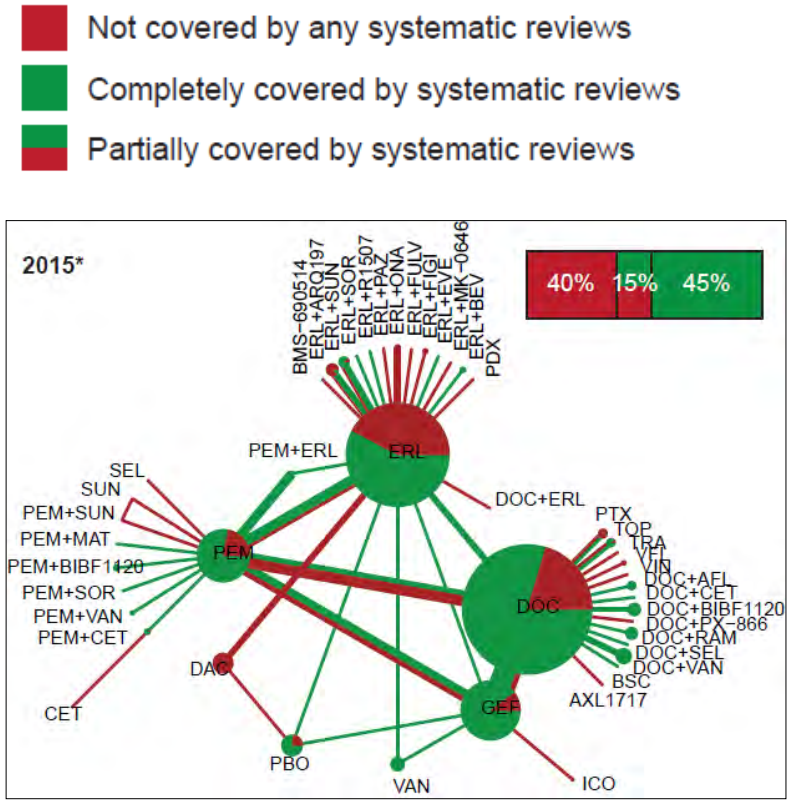
Systematic reviews are rarely updated





Evidence syntheses are incomplete and fragmented

- What proportion of randomized evidence is included in systematic reviews. The example of lung cancer?



- From 2009 to 2015 the evidence covered by existing systematic reviews was consistently incomplete and did not consider
  - 45 % to 70 % of trials;
  - 30 % to 58 % of patients;
  - 40 % to 66 % of treatments;
  - 38 % to 71 % of comparisons

Créquit,..,Ravaud **BMC Med**. 2017
Créquit,..,Ravaud **BMC Med**. 2016
Crequit..,Ravaud **BMJ Open** 2016

From meta-analysis to living meta-analysis to living network meta-analysis

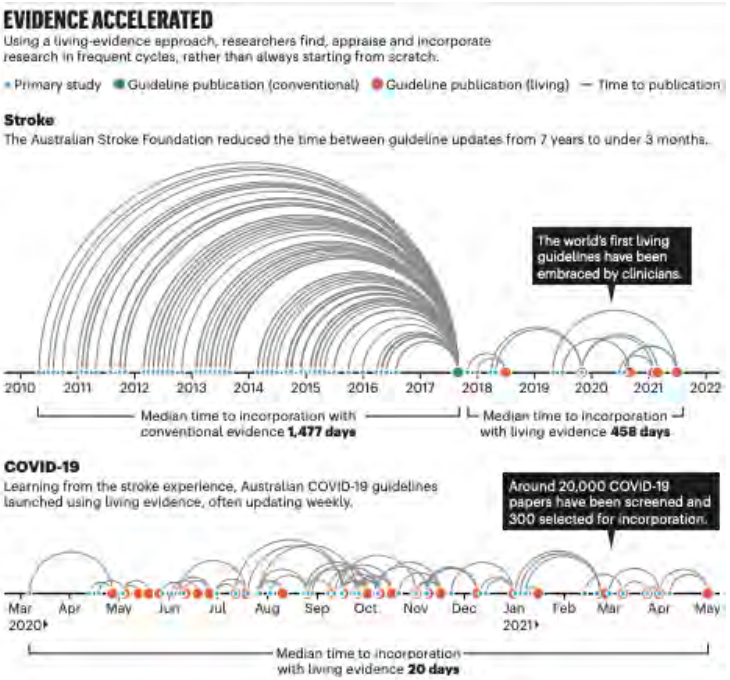
Decision makers need 'living' evidence synthesis

Julian H. Elliott, Rebecca Lawrence, Jan C. Minx, Olufemi T. Oladapo, Philippe Ravaud, Britta Tendal Jeppesen, James Thomas, Tari Turner, Per Olav Vandvik & Jeremy M. Grimshaw

Fund and use dynamic evidence summaries of the latest data to steer research, practice and policy.

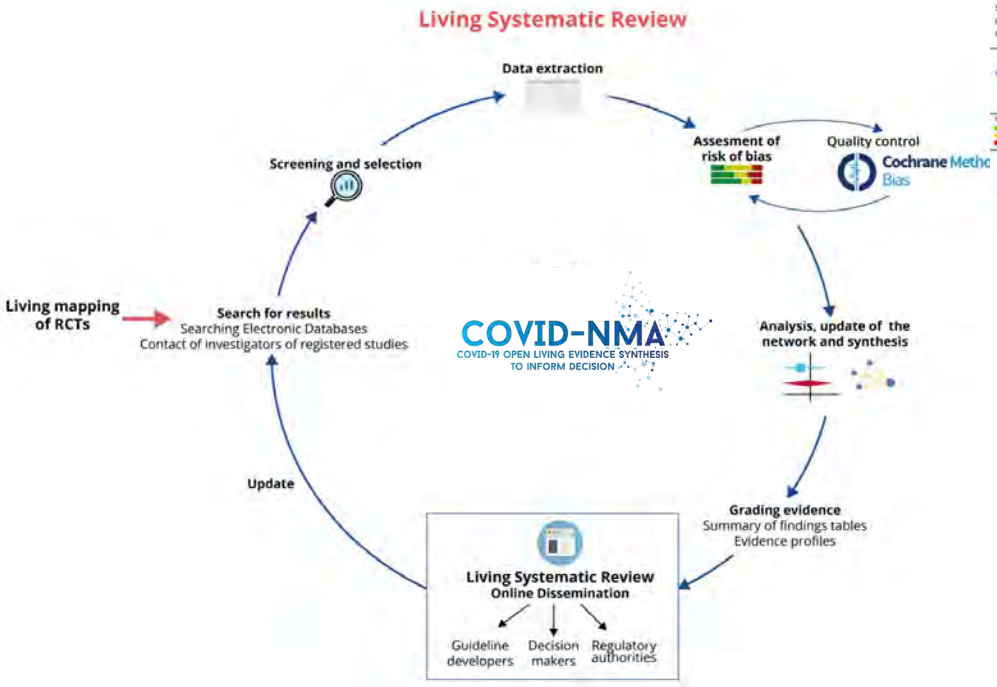
Council, worried that the cacophony would create confusion and anxiety among already-stressed clinicians. We argued for key bodies to come together quickly and use robust, evidence-based processes to find signals in the

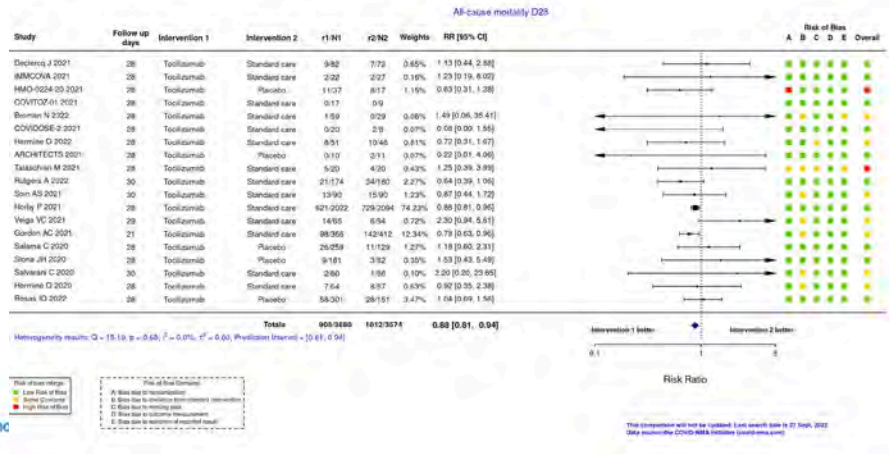
the evidence pipeline[1]. Take the example of remdesivir, an intravenous treatment originally developed for Ebola virus. In May 2020, weak but promising data suggested it could be used to treat COVID-19. Over the next 18 months,



Trustworthy evidence synthesis to support decision making

Paradigm change
Living cumulative network meta-analyses





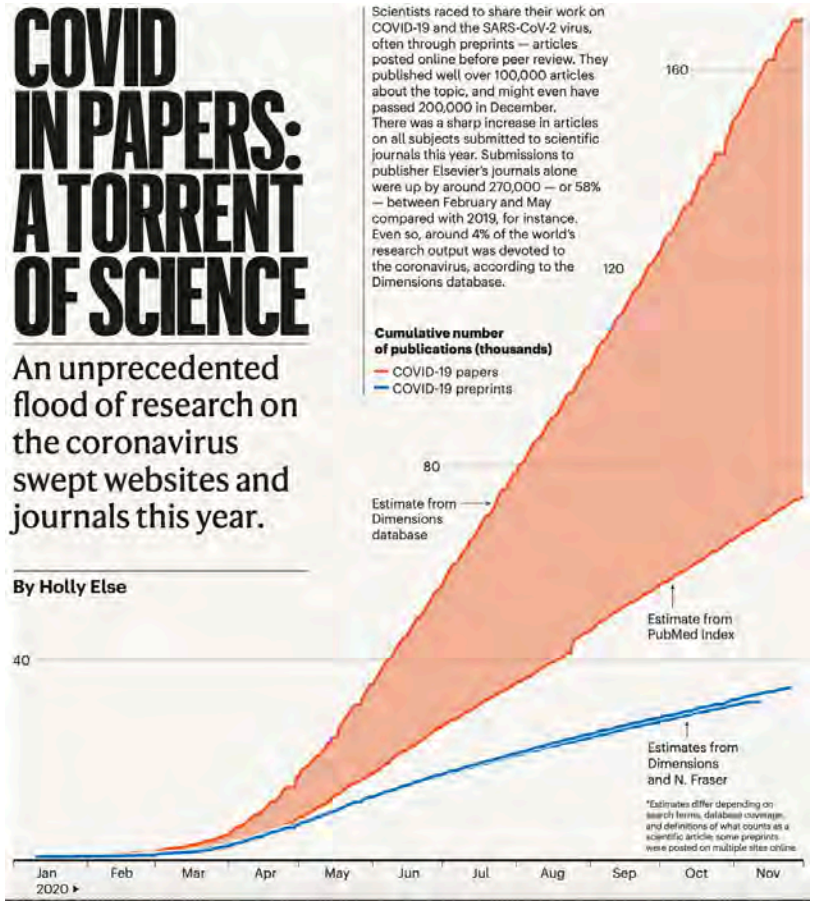
> 900 studies included

World Health Organization

The *Lancet* COVID-19 Commission

Chaimani A, Porcher R, Sbidian E, Mavridis D. **Stat in Med** 2021
Boutron,..,Ravaud **J Clin Epidemiol**. 2020
Ravaud..,Boutron **J Clin Epidemiol**. 2020
Boutron, Chaimani…, Ravaud. **Ann Internal Med**, 2020
Oikonomidi, Boutron, …, Ravaud, **BMC Med**, 2020

The ecosystem is evolving rapidly

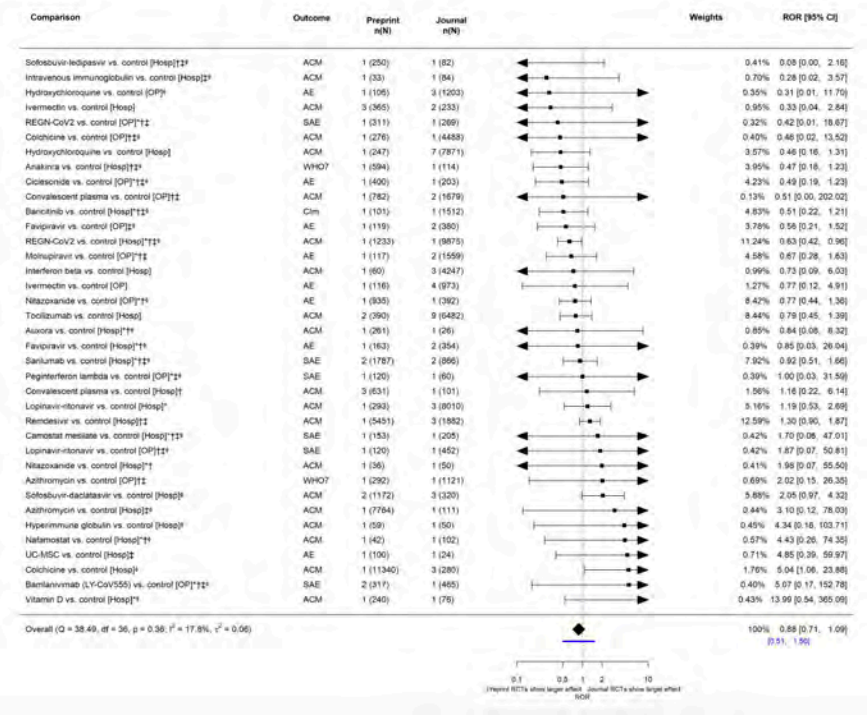


Holly Else, Nature 2020

Davidson et al JCE 2023

Davidson et al BMC Med Res Methodol. 2024

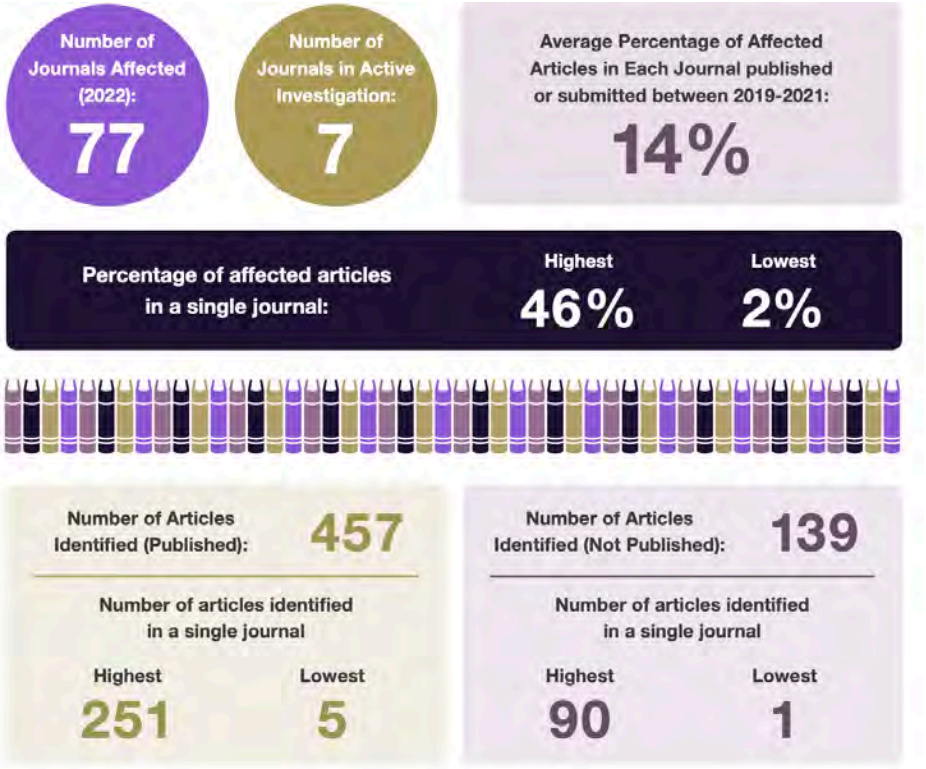
Can we trust results published as preprint?

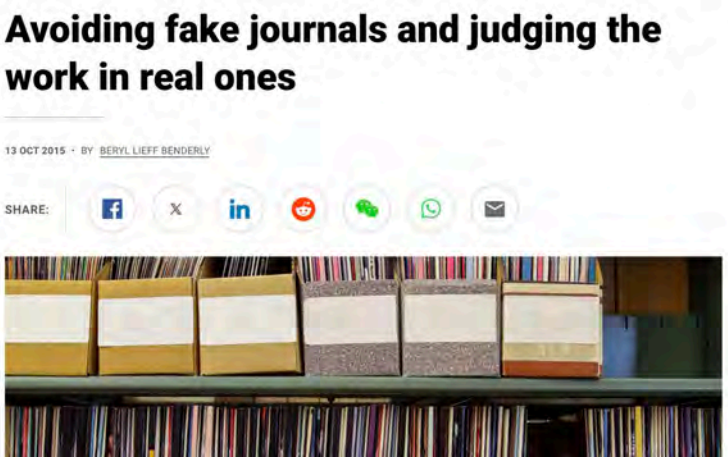


Preprint are living documents -> Iterative assessment

Paper mills and predatory journals

53,000 papers analysed (6 publishers)





Paper Mills – Research report from COPE & STM 2022

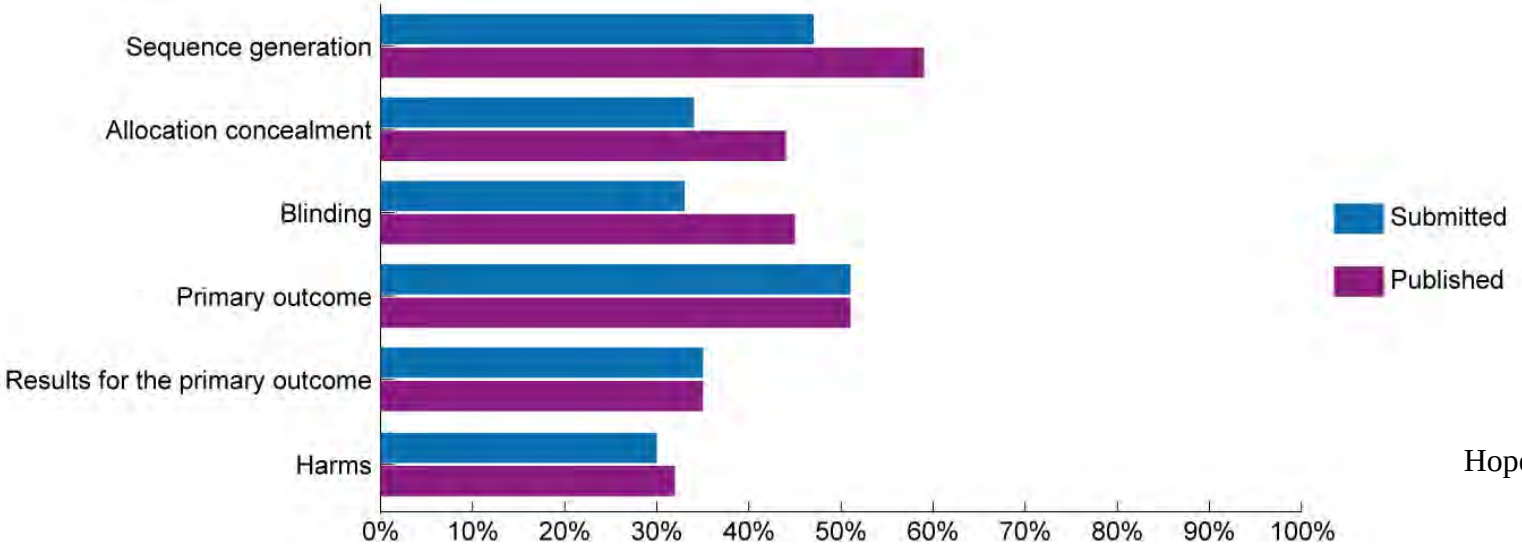
The role of the peer-review process is questioned

- Costly and not sustainable
  - Estimation that 63,4 million hours devoted to peer-review in 2015[1]
  - US$ 6 billion in 2020[3]
- Delay access to results
  - COVID-19 pandemic: a 1-month delay in the results on the effect of steroids (RECOVERY trial) could cost about 200 lives
- Subject to biases
  - Peer-reviewers randomized to assess a study report differing only in the direction of the findings for the primary outcome
    - were **more likely to recommend the positive** version of the manuscript for publication (97% vs 80%, P<0.001)
    - detected **more errors** and **lower methods scores** in the no-difference version than in the positive version

[1] Kovanis et al. Plos One 2016
[2] Knowlson, Torgerson F1000 2020
[3] LeBlanc et al. Res Integr Peer Rev 2023

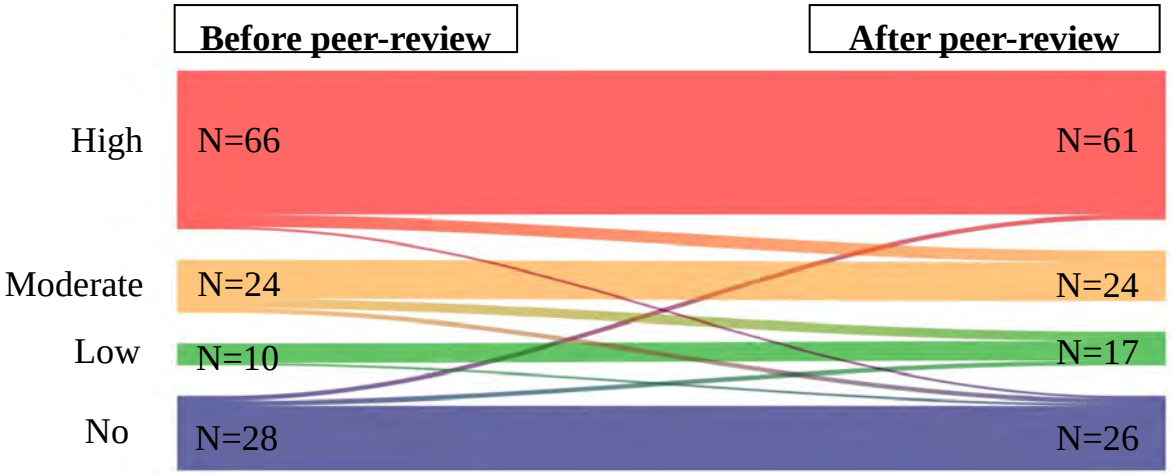
The role of the peer-review process is questioned

Impact on completeness of reporting (adherence to CONSORT)



Hopewell S et al. . BMJ. 2014

Impact on spin (distorted reporting and interpretation of results)



Lazarus C et al Clin Epidemiol. 2016

Generative AI is transforming research practices

HOW GENERATIVE AI COULD DISRUPT SCIENTIFIC PUBLISHING

A world of AI-assisted writing and reviewing might transform the nature of the scientific paper. **By Gemma Conroy**

234 | Nature | Vol 622 | 12 October 2023

Is ChatGPT making scientists hyper-productive? The highs and lows of using AI

Large language models are transforming scientific writing and publishing. But the productivity boost that these tools bring could have a downside.

ChatGPT continues to steal the spotlight, more than a year after its public debut.

The artificial intelligence (AI) chatbot was released as a free-to-use tool in November 2022 by technology company OpenAI in San Francisco, California. Two months later, ChatGPT had already been listed as an author on a handful of research papers.

use of LLMs in research writing. In a 2023 *Nature* survey of more than 1,600 scientists, almost 30% said that they had used generative AI tools to help write manuscripts, and about 15% said they had used them to help write grant applications.

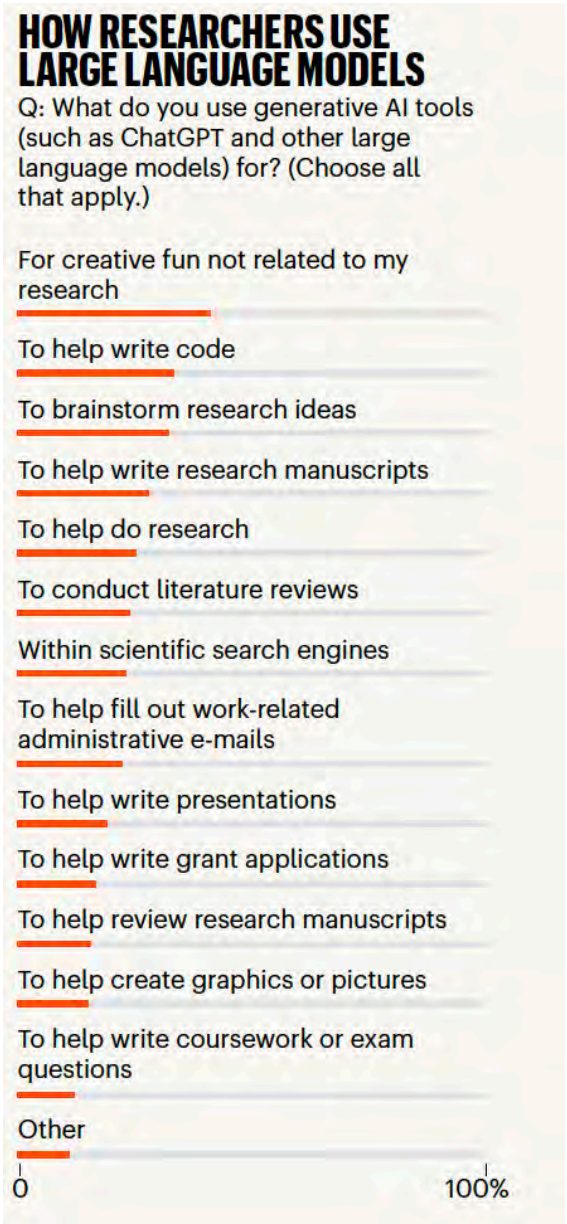
And LLMs have many other uses. They can help scientists to write code, brainstorm research ideas and conduct literature reviews. LLMs from other developers are improving, as well — Google's Gemini, for example, and Claude 2 by Anthropic, an AI company in San Francisco. Researchers with the right skills

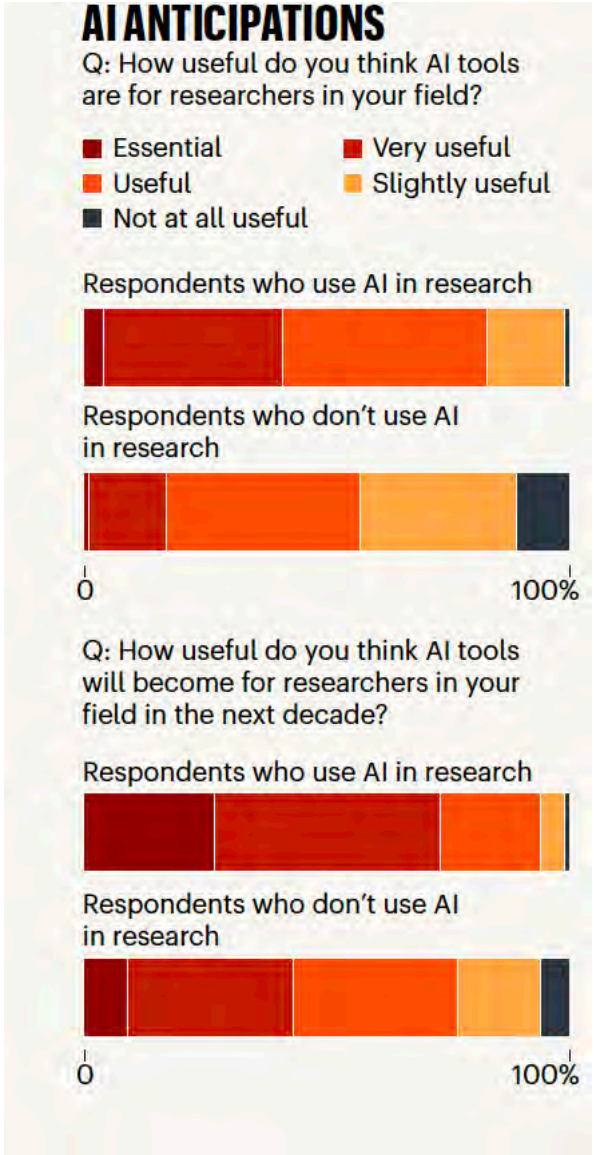
WILL AI TAKE OVER PEER REVIEW?

Artificial intelligence software is increasingly involved in reviewing papers – provoking interest and unease. **By Miryam Naddaf**

852 | Nature | Vol 639 | 27 March 2025

Nature's survey of 1,600 respondents on artificial intelligence (AI)





ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

4. Artificial Intelligence (AI)–Assisted Technology

At submission, the journal should require authors to disclose whether they used Artificial Intelligence (AI)-assisted technologies (such as Large Language Models [LLMs], chatbots, or image creators) in the production of submitted work. Authors who use such technology sh… nit- tec… ow the… sis- tar… see Se… sis, or… in the… as Ch… ney car… nd originality of the work, and these responsibilities are required for authorship (see Section II.A.1). Therefore, humans are responsible for any submitted material that included the use of AI-assisted technologies. Authors should carefully review and edit the result because AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased. Authors should not list AI and AI-assisted technologies as an author or co-author, nor cite AI as an author. Authors should be able to assert that there is no plagiarism in their paper, including in text and images produced by the AI. Humans must ensure there is appropriate attribution of all quoted material, including full citations.

(..) the journal should require authors to **disclose whether they used Artificial Intelligence ((…) in the production of submitted work.**

3. Peer Reviewers

(…) Reviewers must maintain the **confidentiality** of the manuscript as outlined above, which may **prohibit the uploading of the manuscript to software or other AI technologies** (…)

Reviewers must **request permission from the journal prior to using AI technology to facilitate their review.**

Reviewers should be **aware** that AI can generate authoritative-sounding output that **can be incorrect, incomplete, or biased**.

MORE THAN 10,000 RESEARCH PAPERS WERE RETRACTED IN 2023 — A NEW RECORD

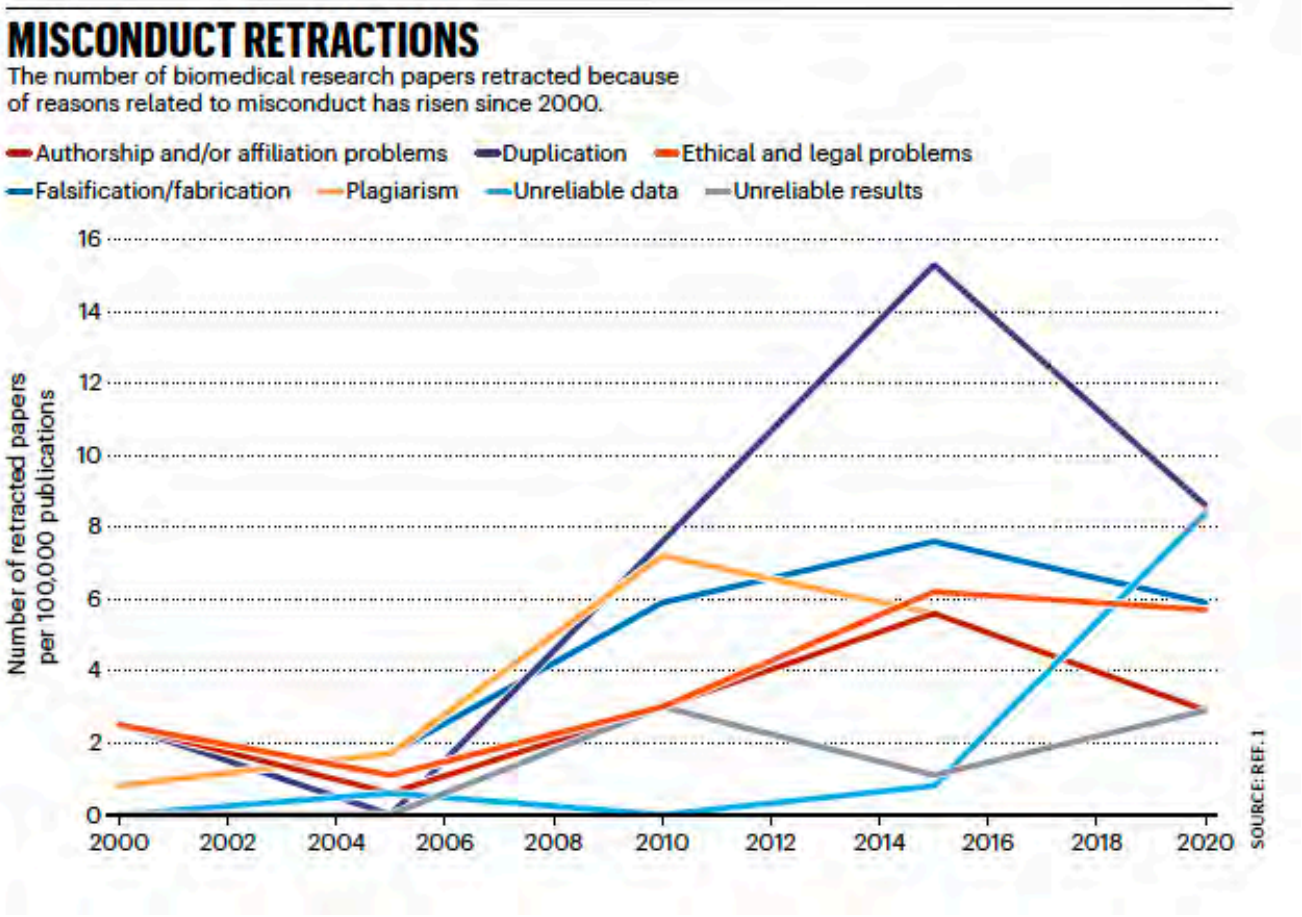
The number of articles being retracted rose sharply this year. Integrity experts say that this is only the tip of the iceberg.

Richard Van Noorden
Nature 2023

BIOMED RETRACTIONS HAVE QUADRUPLED IN 20 YEARS — WHY?

Data falsification and other forms of misconduct are driving a growing proportion of retractions.

Holly Else, Nature 2024



Development of new tools to detect problematic articles

Problematic Paper Screener

Est. February 27th, 2021

- Automated tools are being developed to detect AI generated articles (Guillaume Cabanac, Cyril Labbé)

- Relies on a network of researchers ("research integrity sleuth") checking the papers and commenting on PubPeer and on social media

➢ Hundreds of retraction
  - Springer Nature retracted >300 articles

https://www.irit.fr/~Guillaume.Cabanac/problematic-paper-screener

Development of new tools to detect studies citating retracted studies

Goal: Stop bad research propagating through the scientific literature

G Cabanac, University of Toulouse

Search by journal allows identification of articles citing retracted studies

=> Detection of retracted studies included in meta-analysis

N DeVito, B Goldacre University of Oxford

More than 100,000 researchers infomed by email they cite retracted studies
A randomized trial is ongoing

Grana, … Boutron Jama Intern Med March 31 2025

AI is also an opportunity for systematic reviews

Annals of Internal Medicine®

Research and Reporting Methods | 21 May 2024

Sensitivity and Specificity of Using GPT-3.5 Turbo Models for Title and Abstract Screening in Systematic Reviews and Meta-analyses

Authors: Viet-Thi Tran, MD, PhD, Gerald Gartlehner, MD, MPH, Sally Yaacoub, PhD, Isabelle Boutron, MD, PhD, Lukas Schwingshackl, PhD, MSc, Julia Stadelmaier, MSc, Isolde Sommer, PhD, Farzaneh Alebouyeh, MSc, Sivem Afach, PhD, Joerg Meerpohl, MD, PhD, and Philippe Ravaud, MD, PhD AUTHOR, ARTICLE, & DISCLOSURE INFORMATION

Evidence synthesis

« *The ultimate goal of systematic reviews and meta-analysis is to create an effective marketplace for synthesis in which policy-makers (..)always seek the best evidence because they know it will be available, and researchers synthesize evidence because they know it will make a difference.* » (C A. Donnelly, Nature 2018)

The current evidence synthesis ecosystem does not fulfil this goal.

Mass production of systematic reviews and meta-analyses

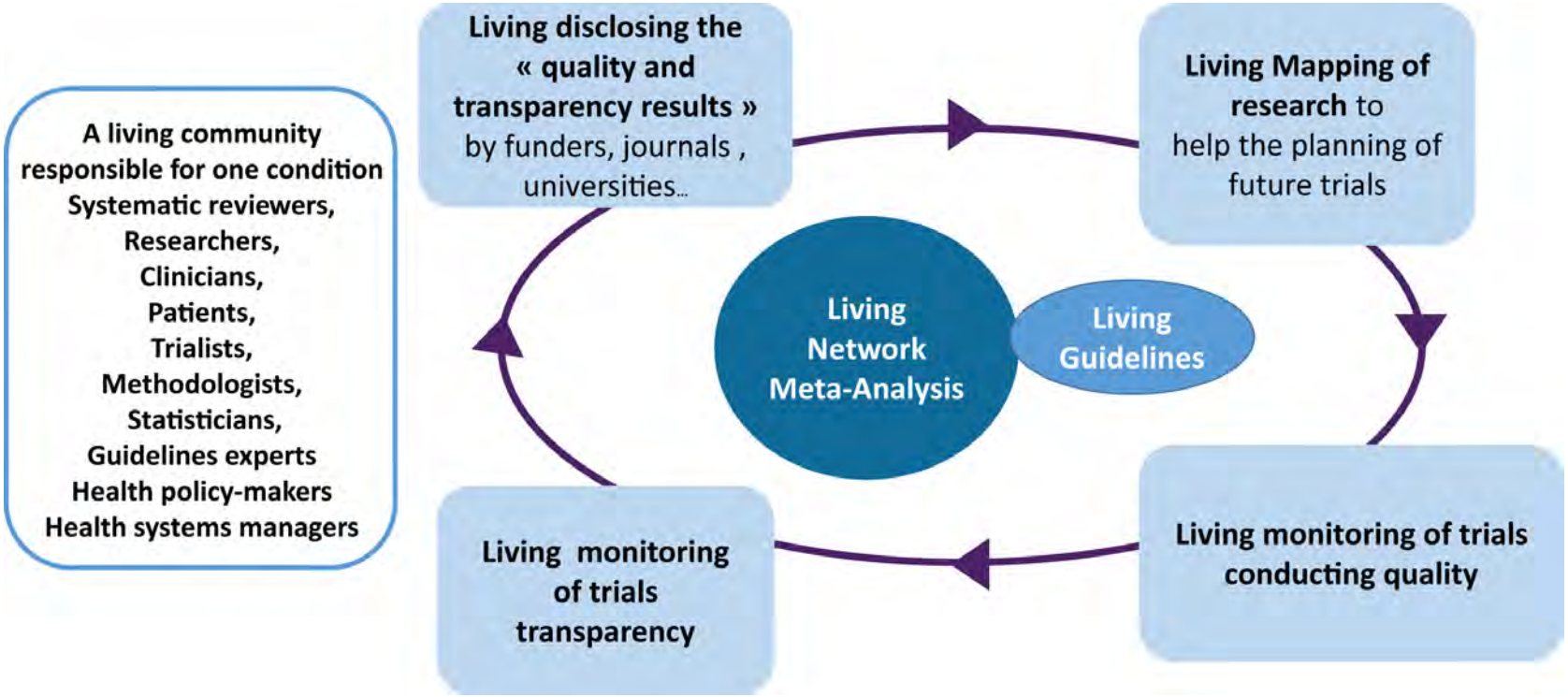
- Low quality
- Redundant
- Not covering all evidence
- Delay in producing the review (2-3 years)
- Rarely updated
- Rely on primary evidence of low quality

Opportunities and challenges

- Access to new source of data
  - Preprint, clinical trial registries, protocols, and clinical study reports from regulatory agencies or pharmaceutical companies
- Access to new types of data
  - IPD
  - Non randomized studies of routinely collected data (rich data, new design such as emulated trials, new statistical methods)
- New technology
  - AI tools, Large language models

Rethinking the role of evidence synthesis

Toward a new research ecosystem relying on a culture of continuous improvement



Collaboration with IT experts to develop the relevant tools to accelerate the process

Ravaud et al. JCE 2020

Conclusion

- Trustworthy and up-to-date evidence synthesis is essential to support decision making
- The landscape is changing rapidly with new challenges and opportunities arising
- Novel approaches are needed to fulfil stakeholders' needs