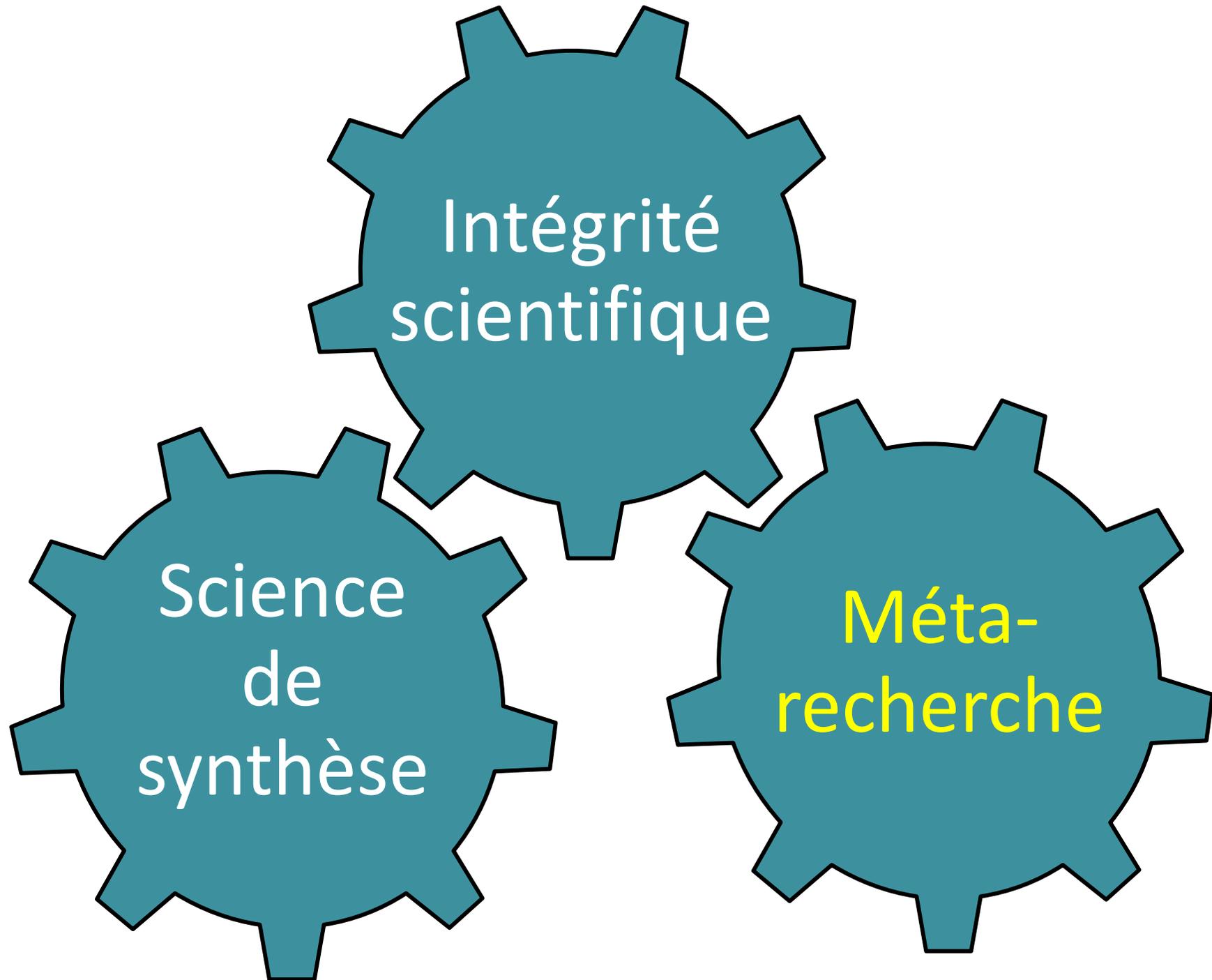


# Méta-recherche, intégrité scientifique et science de synthèse

Nancy Durieux  
François Léonard  
Ezio Tirelli

Liège, le 25 avril 2024





# Reproductibilité

*"Throughout the history of science, leading scientists have endeavoured to theorize and conduct research on fundamental aspects of the scientific method and to identify ways to implement it most efficiently"*



Le point commun entre la recherche et les lapins:  
LA REPRODUCTIBILITÉ





# 1,500 scientists lift the lid on reproducibility

[Monya Baker](#)

[Nature](#) 533, 452–454 (2016)

## IS THERE A REPRODUCIBILITY CRISIS?

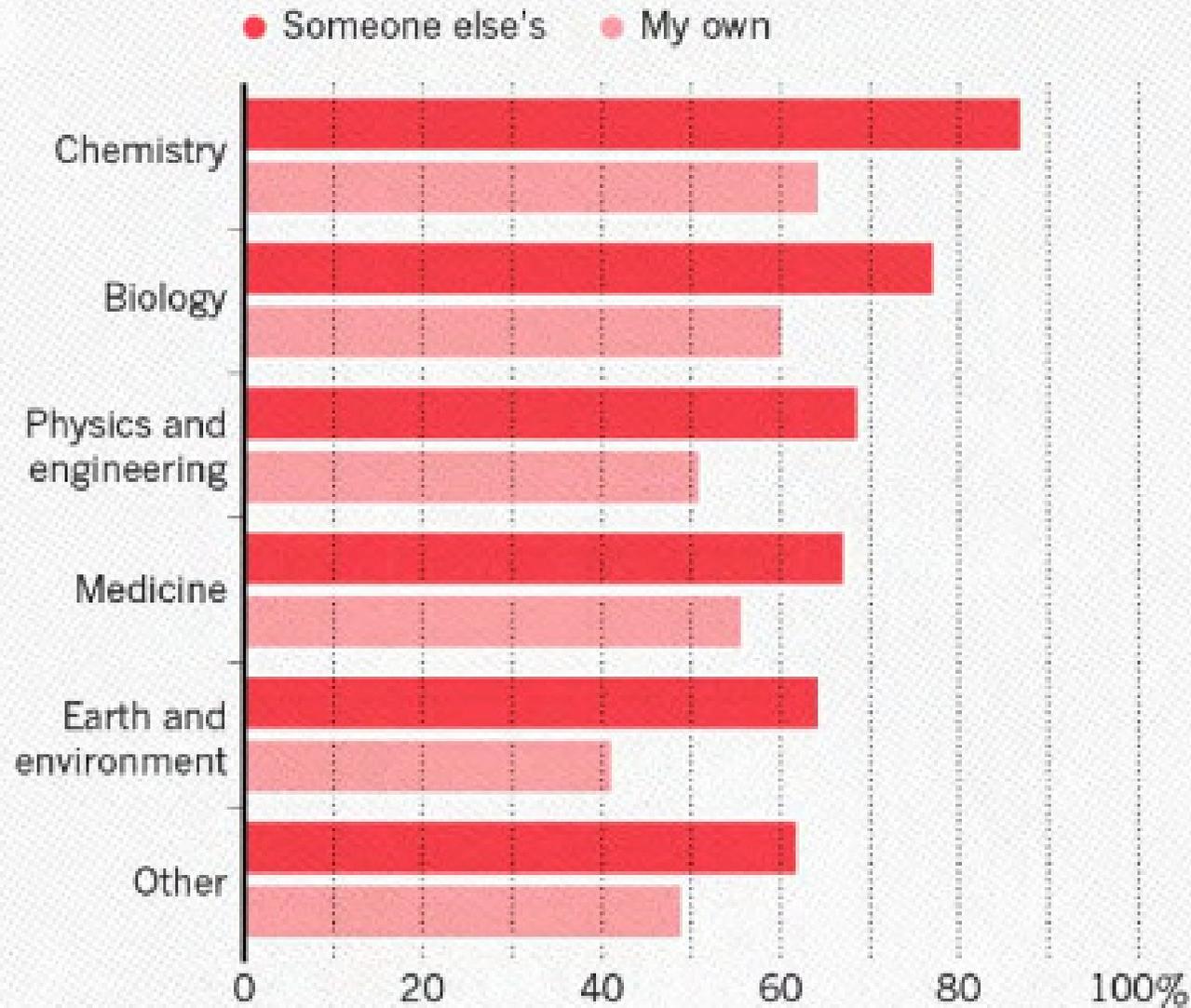


©nature



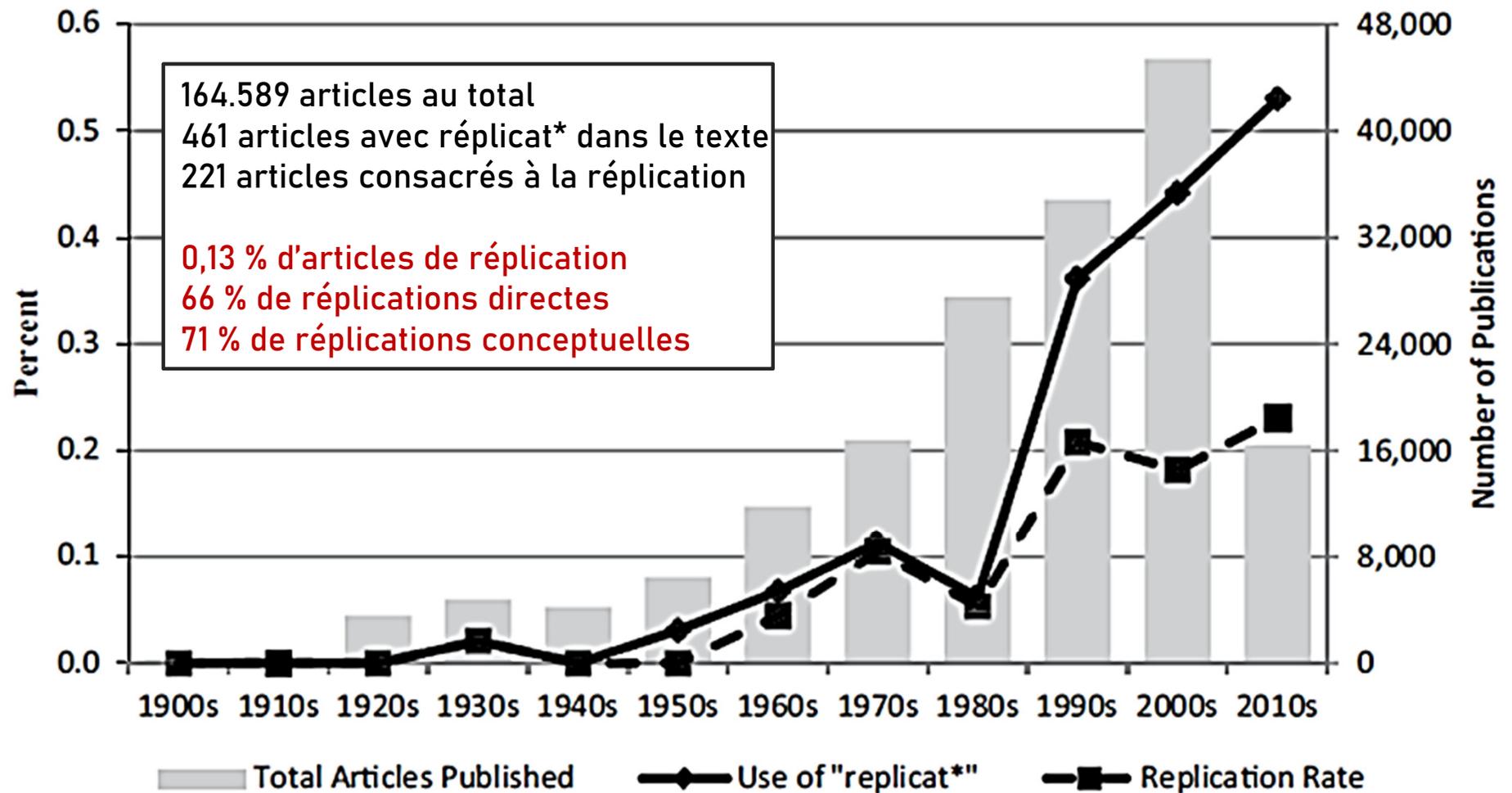
# HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.



## Répliquations des résultats en sciences de l'éducation

Analyse « lexicographique » du taux d'articles consacrés à la répliation dans tous les articles pertinents de 100 revues (depuis leur fondation)





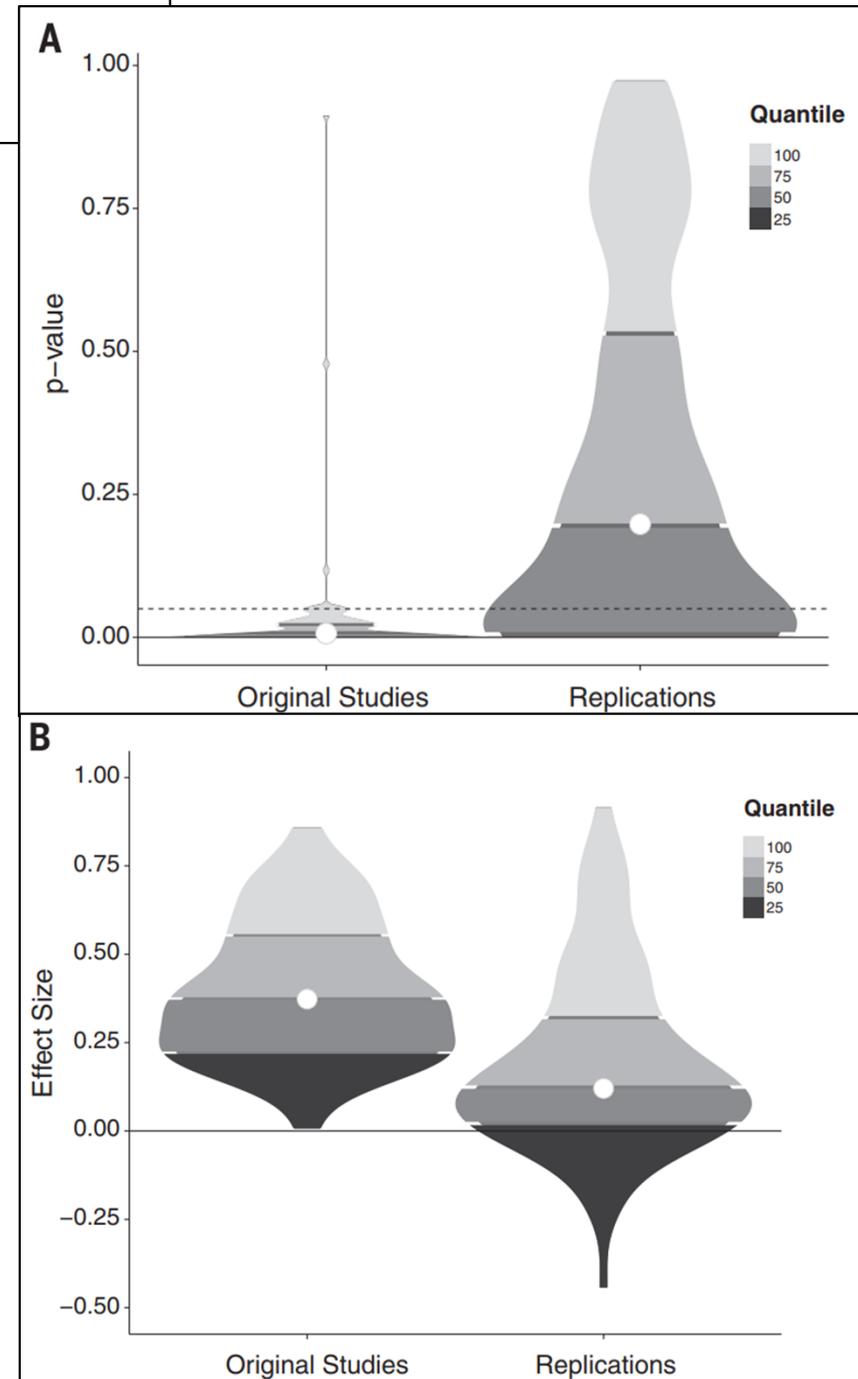
# Estimating the reproducibility of psychological science

OPEN SCIENCE COLLABORATION [Authors Info & Affiliations](#)

SCIENCE • 28 Aug 2015 • Vol 349, Issue 6251 • DOI: 10.1126/science.aac4716

## ► Objectif : répliquer 100 études en psychologie

- Plusieurs indicateurs de reproductibilité utilisés
- 97 études originales statistiquement significatives, contre 35 répliquations statistiquement significatives (36 %)
- 73 répliquations permettaient le calcul d'un IC sur leur taille d'effet
  - 30 IC contenaient la taille d'effet de l'étude original (43 %)
- Les effets originaux (moyenne = 0,403) sont diminués de moitié dans les études de répliquations (moyenne = 0,197)





## Principaux résultats de 11 initiatives majeures d'essais de reproduction/réplication directe de groupes d'études publiées

Etude de réplication	# études	# désaccords	% désaccords
Amgen (Begley & Ellis 2012)	53	47	89
Bayer (Prinz et al 2011)	67	44	66
Preclinical Cancer Biology (Harris 2017)	6	2	33
Preclinical Materials (Vasilevsky et al 2012)	238	109	46
Réplifications auto-rapportées (Hartshorne & Schachner 2012)	257	130	51
<b>Open Science Collaboration (Nozek et al 2015)</b>	<b>100</b>	<b>64</b>	<b>64</b>
Many Labs I (Klein et al 2014)	13	3	23
Many Labs II (Klein et al 2018)	28	13	46
Many Labs III (Ebersole et al 2016)	10	5	50
Experimental Economics (Camerer et al 2016)	18	7	39
Social Science Research (Camerer et al 2018)	21	8	38
<b>TOTAL</b>	<b>811</b>	<b>432</b>	<b>53,3</b>



## Attributs d'un **résultat fiable** (sens de preuve, « evidence »)

Attribut	Question Hypothèse	Chercheur	Données	Méthodes Analyses
Répétable	Identique	Identique	Identique	Identique
Reproductible Reproductibilité de la méthode ou directe ??	Identique	Différent	Identique	Identique
Répliquable Reproductibilité du résultat ou directe ??	Identique	Identique ou différent	Différent (similaire)	Identique
Généralisable	Identique	Identique ou différent	Différent	Différent (analyse identique)
Robuste Reproductibilité conceptuelle ??	Identique	Identique ou différent	Identique ou différent	Différent

**Reproductibilité inférentielle:** cohérence entre les **conclusions et interprétations** d'un ou de plusieurs résultats par divers chercheurs.

# Why Most Published Research Findings Are False

John P. A. Ioannidis

19 ans plus tard:

75,262 Save	7,977 Citation
3,110,679 View	9,987 Share

## Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

## Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a  $p$ -value less than 0.05. Research is not most appropriately represented and summarized by  $p$ -values, but, unfortunately, there is a widespread notion that medical research articles

**It can be proven that most claimed research findings are false.**

should be interpreted based only on  $p$ -values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful. “Negative” is actually a misnomer, and

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is  $R/(R + 1)$ . The probability of a study finding a true relationship reflects the power  $1 - \beta$  (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate,  $\alpha$ . Assuming that  $c$  relationships are being probed in the field, the expected values of the  $2 \times 2$  table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the  $2 \times 2$  table, one gets  $PPV = (1 - \beta)R/(R$



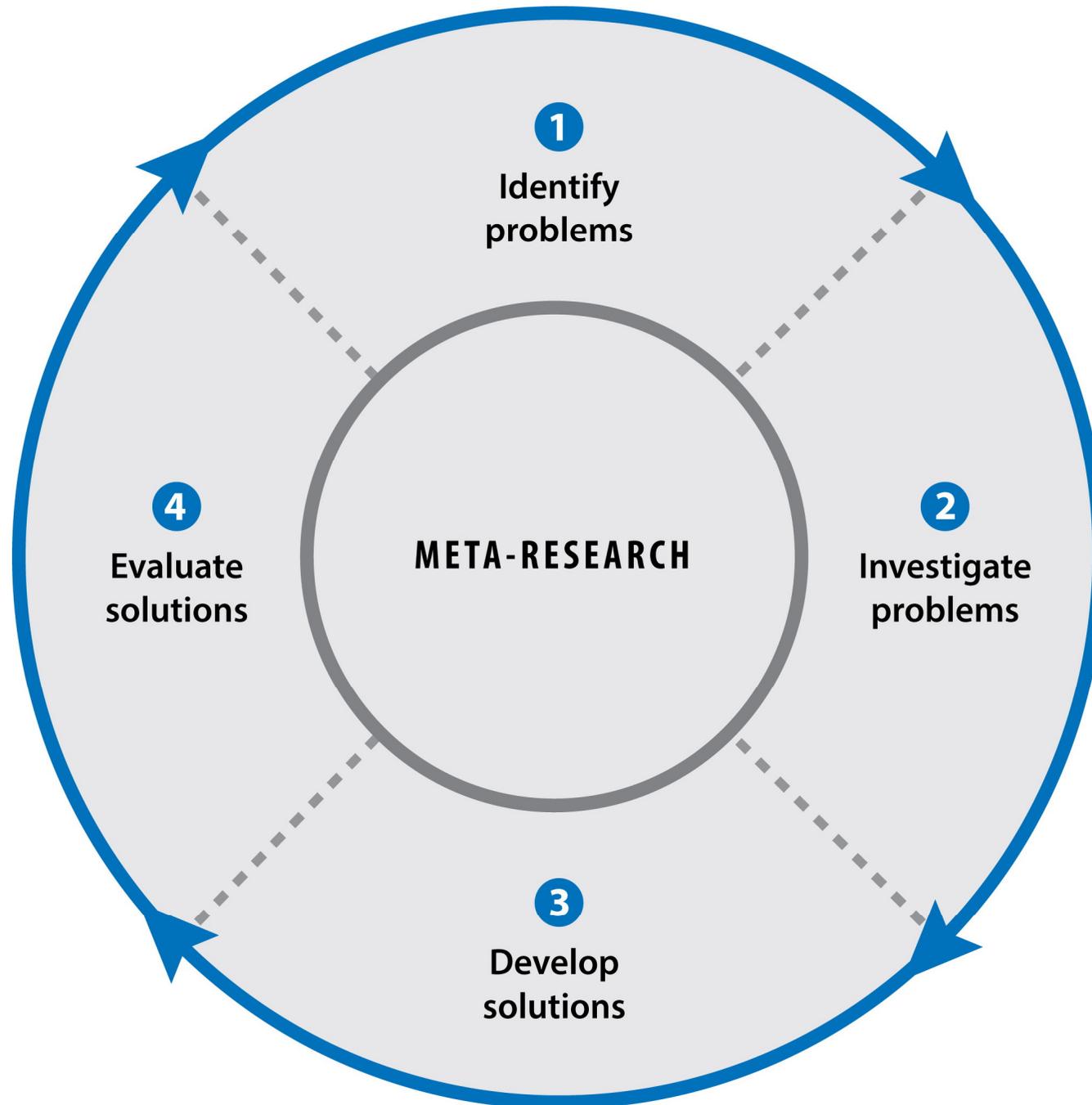
# Méta-recherche

- ▶ Regarde au-delà des études individuelles  
(regard plus large sur la recherche)
- ▶ Etudie les pratiques de recherche
- ▶ Utilise une approche interdisciplinaire
- ▶ Promeut et défend une science de qualité avec des résultats  
fiables



# Méta-recherche : cinq domaines d'intérêt

Meta-research area	Specific interests (nonexhaustive list)
<b>Methods:</b> "performing research"—study design, methods, statistics, research synthesis, collaboration, and ethics	Biases and questionable practices in conducting research, methods to reduce such biases, <u>meta-analysis</u> , <u>research synthesis</u> , <u>integration of evidence</u> , crossdesign synthesis, collaborative team science and consortia, research integrity and ethics
<b>Reporting:</b> "communicating research"—reporting standards, study registration, disclosing conflicts of interest, information to patients, public, and policy-makers	Biases and questionable practices in reporting, explaining, disseminating and popularizing research, conflicts of interest disclosure and management, study registration and other bias-prevention measures, and methods to monitor and reduce such issues
<b>Reproducibility:</b> "verifying research"—sharing data and methods, repeatability, replicability, reproducibility, and self-correction	Obstacles to sharing data and methods, replication studies, replicability and reproducibility of published research, methods to improve them, effectiveness of correction and self-correction of the literature, and methods to improve them
<b>Evaluation:</b> "evaluating research"—prepublication peer review, postpublication peer review, research funding criteria, and other means of evaluating scientific quality	Effectiveness, costs, and benefits of old and new approaches to peer review and other science assessment methods, and methods to improve them
<b>Incentives:</b> "rewarding research": promotion criteria, rewards, and penalties in research evaluation for individuals, teams, and institutions	Accuracy, effectiveness, costs, and benefits of old and new approaches to ranking and evaluating the performance, quality, value of research, individuals, teams, and institutions





# Méta-recherche : méthodes (exemples)

## ▶ Théoriques

- Méthodes analytiques et informatiques

## ▶ Empiriques

- Enquêtes sur les biais dans un domaine donné (-> données descriptives)
- Analyses observationnelles d'association et de corrélation
- Etudes interventionnelles
  - par exemple, des essais randomisés évaluant si une pratique de recherche conduit à de meilleurs résultats qu'une autre
- (Combinaison de ) Méta-analyses

## Diapositive 15

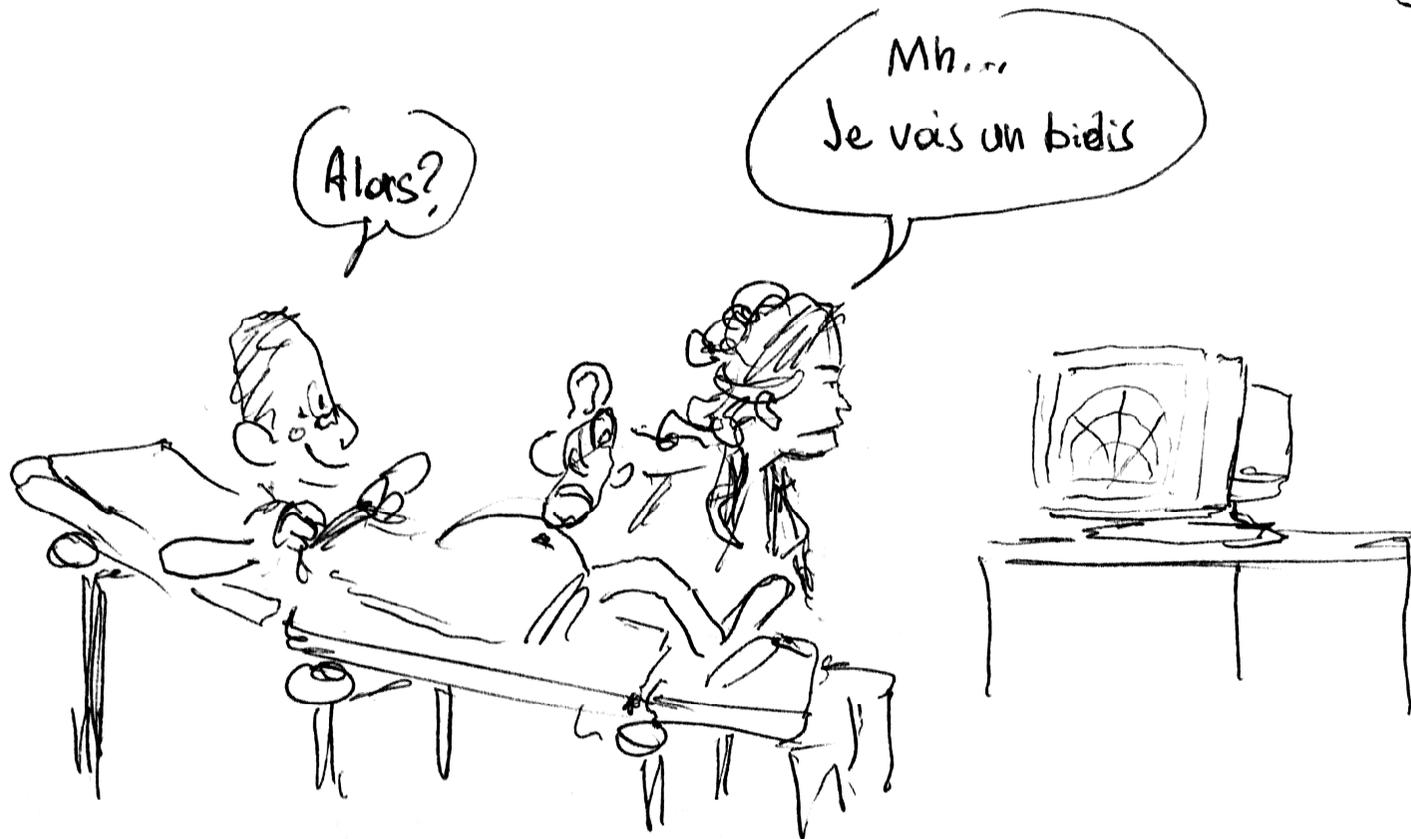
---

- ND0** François, je pense que ça a du sens de parler ici des méthodes pour conduire de telles études  
Durieux Nancy; 2024-04-20T12:37:41.213
- ND0 0** On n'est pas obligé de citer toutes celles reprises par Ioannidis...  
Le titre pourrait être :  
Méta-recherche : méthodes (exemples)  
Durieux Nancy; 2024-04-20T12:40:38.649
- ND0 1** OU on n'en parle pas... tu le diras avec tes deux exemples... et moi avec ma partie  
Durieux Nancy; 2024-04-20T13:24:01.982



## Méta-recherche : quelques enseignements

Ces métiers méconnus : le méthodologue





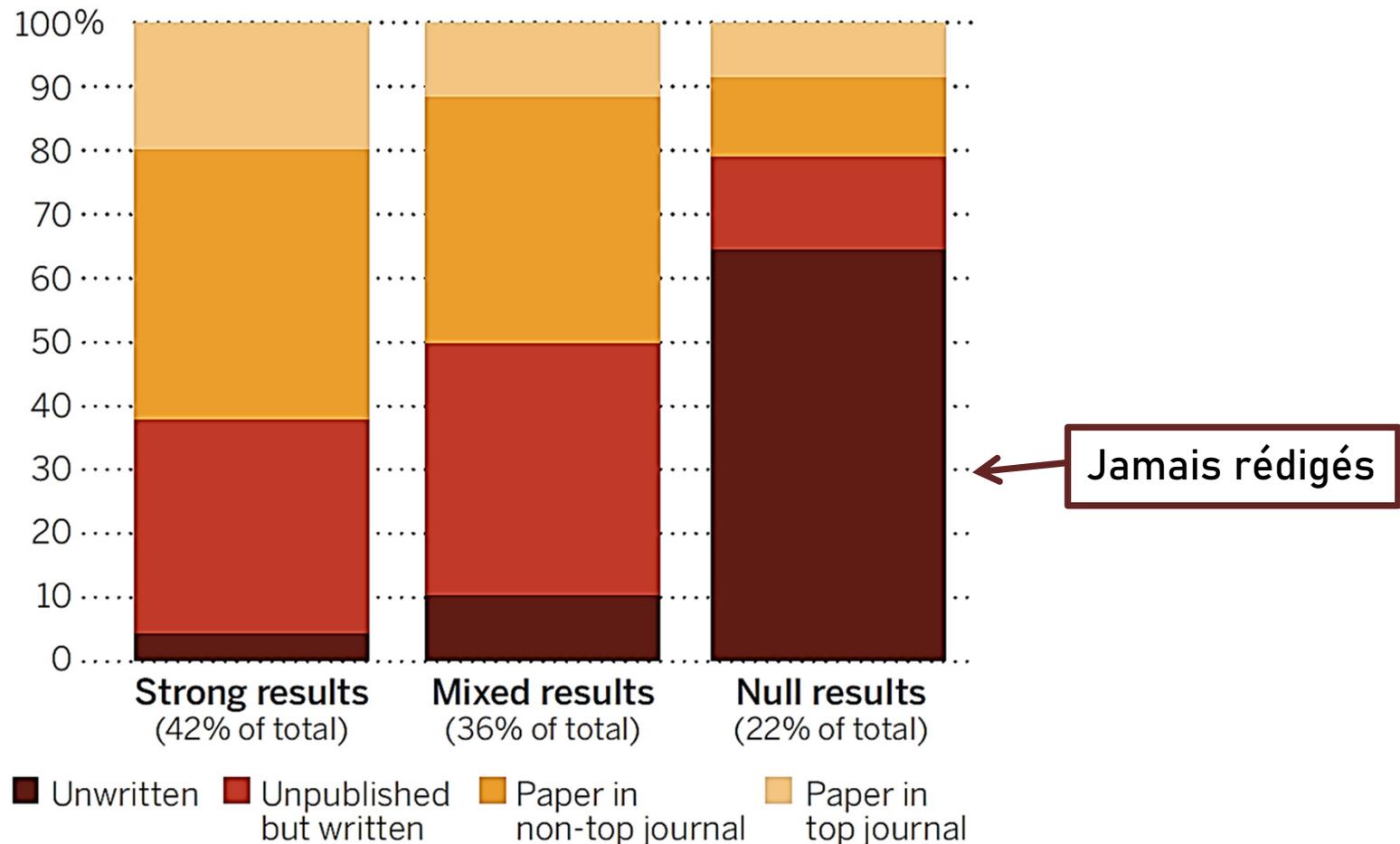
## Biais de publication

- ▶ Non publication des résultats ne montrant pas un effet « significatif » et surprenant
- ▶ Incite les chercheurs à utiliser des pratiques de recherche discutables (parfois fraudes)
- ▶ Promeut le mythe que les chercheurs trouvent toujours la bonne hypothèse



# Biais de publication en sciences sociales (USA)

- ▶ 221 études publiées et non-publiées en sciences sociales faisant appel à des "*survey-based experiments*"
- ▶ Financées et révisées par la NSF dans le cadre du programme "*Time-sharing Experiments in the Social Sciences*" (TESS) entre 2002 et 2012





## Pratiques d'*authorship*

- ▶ Enquête auprès de 800 chercheurs en psychologie
- ▶ Objectifs
  - Déterminer la prévalence du *gift authorship*
  - Déterminer la prévalence du *ghost authorship*
- ▶ Résultats
  - 49 % des répondants ont été témoins de *gift authorship*
  - 17 % des répondants ont été témoins de *ghost authorship*
  - 88,9 % pensent que des *guidelines d'authorship* explicites sont bénéfiques
  - Si les institutions encouragent l'utilisation de *guidelines*, la prévalence de *gift authorship* est plus basse

## Mozart Effect, Effect Size and Small-Study Effect

Léonard, François ; Tirelli, Ezio

2022 • Annual Meeting of the Belgian Association of Psychological  
Sciences

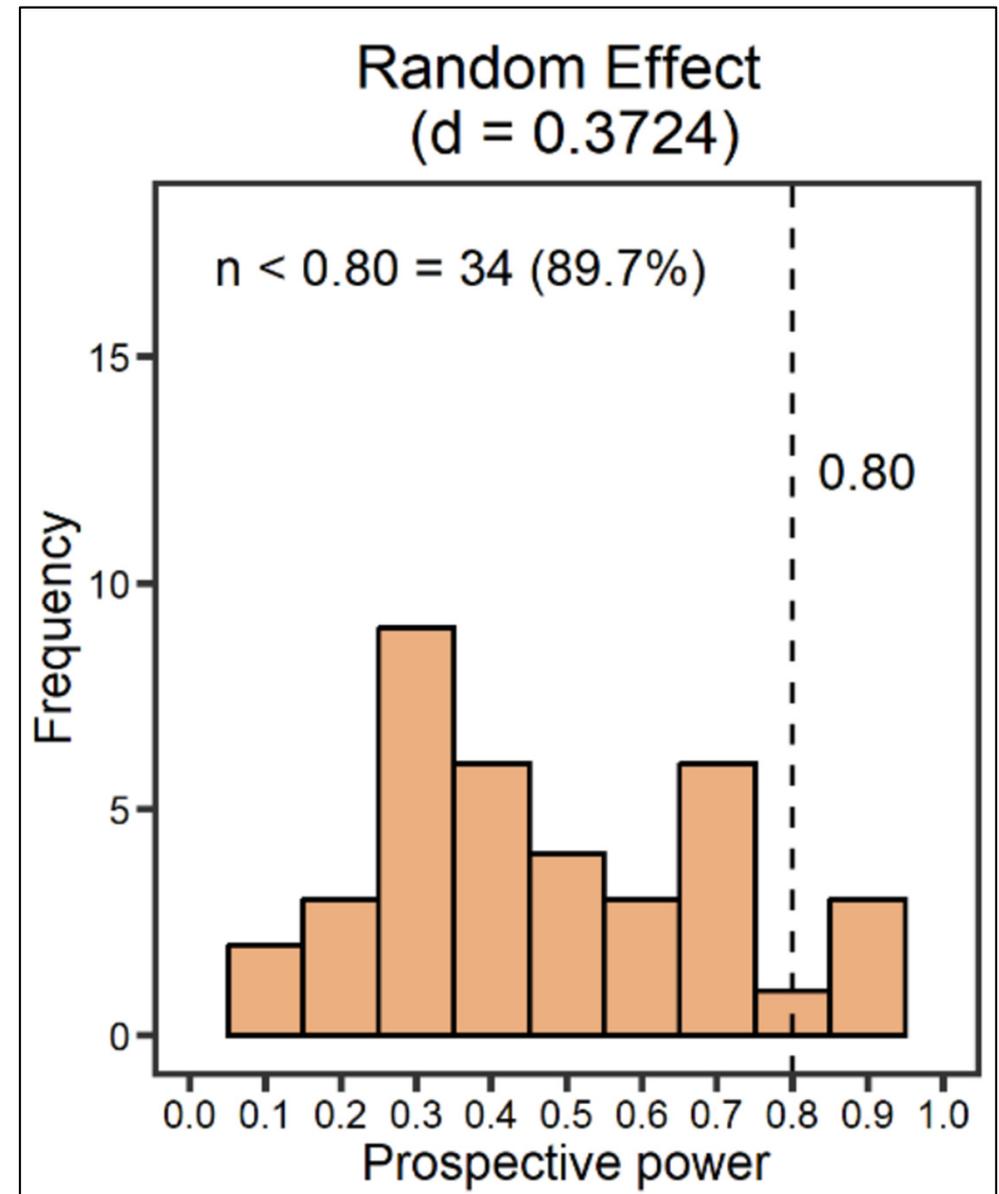
✓ Editorial reviewed

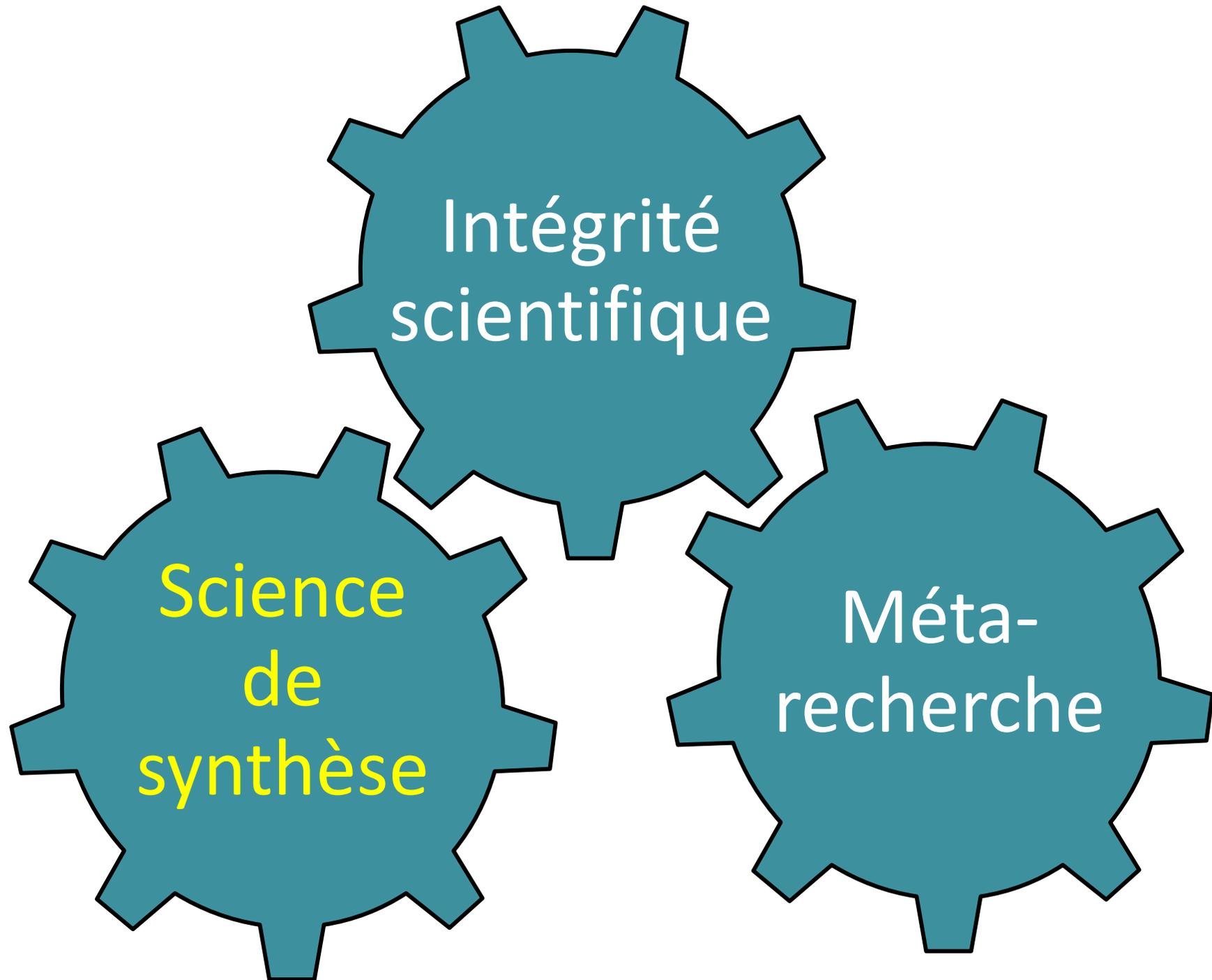


Permalien

<https://hdl.handle.net/2268/303783>

- ▶ Etude observationnelle basée sur les études testant l'effet Mozart
- ▶ Estimation de la puissance statistique pour diverses tailles d'effets méta-analytiques
- ▶ Taille d'effet de l'effet Mozart sans signifiante pratique après l'application d'une correction pour le biais de publication
- ▶ Puissance statistique médiane de 32,7 % (effectif médian de 41)
- ▶ 89,7 % des études sont sous-puissantes
- ▶ Mise à jour de l'étude prévue cette année avec Léonard, F., Durieux, N., & Tirelli, E.







Meta-research area	Specific interests (nonexhaustive list)
<b>Methods:</b> "performing research"—study design, methods, statistics, <u>research synthesis</u> , collaboration, and ethics	Biases and questionable practices in conducting research, methods to reduce such biases, <u>meta-analysis</u> , <u>research synthesis</u> , <u>integration of evidence</u> , crossdesign synthesis, collaborative team science and consortia, research integrity and ethics

Ioannidis et al., 2015 – <https://doi.org/10.1371/journal.pbio.1002264>

## What is a methodological study?

Any study that describes or analyzes methods (design, conduct, analysis or reporting) in published (or unpublished) literature is a methodological study. Consequently, the scope of methodological studies is quite extensive and includes, but is not limited to, topics as diverse as: research question formulation [11]; adherence to reporting guidelines [12,13,14] and consistency in reporting [15]; approaches to study analysis [16]; investigating the credibility of analyses [17]; and studies that synthesize these methodological studies [18]. While the

Mbuagbaw et al., 2020 - <https://doi.org/10.1186/s12874-020-01107-7>

Naming	Methodological study Methodological review Meta epidemiological review Research on research
--------	--

Khalil & Munn, 2023 – <https://doi.org/10.1097/PXH.0000000000000013>



SPECIAL COMMENTARY

## Guidance on conducting methodological studies – an overview

Khalil, Hanan<sup>a</sup>; Munn, Zachary<sup>b</sup>

[Author Information](#) 

*Current Opinion in Epidemiology and Public Health* 2(1):p 2-6, March 2023. | DOI: 10.1097/PXH.0000000000000013

- ▶ Diverses méthodes utilisées, telles que
  - *scoping review methodology*
  - *systematic review methodology*
  
- ▶ Nécessité de développer des directives formelles et complètes pour mener des études méthodologiques
  
- ▶ Contribution importante : tutoriel développé par Mbuagbaw et al., 2020



Fig. 2



A proposed framework for methodological studies



## Research synthesis

- ▶ *Systematic review (avec ou sans méta-analyse)*

*"Systematic reviews seek to collate evidence that fits pre-specified eligibility criteria in order to answer a specific research question. They aim to minimize bias by using explicit, systematic methods documented in advance with a protocol"*

- ▶ *Rapid review*

- ▶ *Umbrella review*

- ▶ *Scoping review*

- ▶ *Etc.*



# Research synthesis : recommandations d'élaboration

## ► Exemples





## *Research synthesis : étapes*

- ▶ Formuler une question
- ▶ Définir des critères d'inclusion / d'exclusion
- ▶ Effectuer une recherche des publications scientifiques
- ▶ Sélectionner les études à inclure
- ▶ [Evaluer la qualité des études]
- ▶ Extraire les données
- ▶ Analyser et synthétiser les données
- ▶ Présenter et interpréter les résultats



## *Research synthesis : matériel / outils*

- ▶ Bases de données bibliographiques et autres ressources
- ▶ Logiciel de support / (semi-)automatisation
- ▶ Grilles d'évaluation critique des études
- ▶ Grilles d'extraction de données
- ▶ Logiciel statistiques (->méta-analyse)
- ▶ ...



# Exemple

COMMUNICATION ORALE NON PUBLIÉE/ABSTRACT (COLLOQUES ET CONGRÈS SCIENTIFIQUES)

## Interventions combinées dans la prise en charge du burnout chez les travailleurs avec un contrat d'emploi : une « scoping review »

Lehaen, Cloé ; Hansez, Isabelle ; Durieux, Nancy

2023 • AIPTLF 2023



## Résultats préliminaires

Intervention individuelle	Intervention organisationnelle	Auteurs
Psychoéducation	<b>Analyse de l'environnement de travail, piste de solutions et implémentation de solutions</b>	Lam et al., (2022)
Formation à la <b>régulation affective</b> basée sur la pleine conscience	La planification avancée des congés, des demi-journées bien-être et la création d'un comité bien-être piloté par les résidents	Lebares et al., (2021)
Evènements et programme bien-être (application d'un modèle d'amélioration de la qualité continue et de cycles en série Planifier, Faire, Étudier, Agir (PDSA))	Interventions qui ont affecté le <b>flux de travail et l'environnement d'apprentissage</b> (changements de leadership, recrutement de professeurs, soutien administratif et innovations technologiques) (application d'un modèle d'amélioration de la qualité continue et de cycles en série Planifier, Faire, Étudier, Agir (PDSA))	Ogunyemi et al., (2022)
Gestion des émotions	<b>Mode de gestion hiérarchique</b> (répartition des tâches en fonction du niveau hiérarchique du personnel)	Liu et al., (2022)
Interventions variées allant de <b>programmes bien-être à des formations professionnelles</b>	<b>Interventions variées</b> : implémentation de postes de travail ergonomiques, amélioration des pratiques de travail..., team-building	Buch et al., (2021)
Conseil individuel, gestion des émotions, groupes d'entraide et de résolution de problèmes, formation axée sur l'expérience personnelle au travail	Travail d'équipe et gestion de conflits	Bui et al., (2021)
<b>Support individuel et en groupe</b> (difficultés liées au travail : suite à un évènement, suite aux aspects émotionnels et pratiques liés à la charge de travail)	<b>Travail d'équipe</b>	Maggio et al., (2021)



## Discussion

- Etudes récentes sur le sujet
- Utilisation du MBI
- Majorité d'études sur le secteur des soins de santé
- Combinaison de plusieurs initiatives

AIPTLF2023

## Discussion

- Pas toujours évident de déterminer ce qui est organisationnel ou individuel
- Informations relatives aux interventions parfois indisponibles
- « Combined » se réfère à plusieurs interventions
- Utilisation d'autres termes tels que multilevel ou multifaceted

AIPTLF2023

L'appel d'un  
temps nouveau :  
**l'humain au cœur de  
la transformation du  
travail**



## Autres exemples



**ARTICLE (PÉRIODIQUES SCIENTIFIQUES)**

**Interventions That Target or Affect Voice or Speech Production During Public Speaking: A Scoping Review.**

Menjot, Pauline \*; Bettahi, Lamia \*; Leclercq, Anne-Lise  *et al.*

2023 • In *Journal of Voice*





**ARTICLE (PÉRIODIQUES SCIENTIFIQUES)**

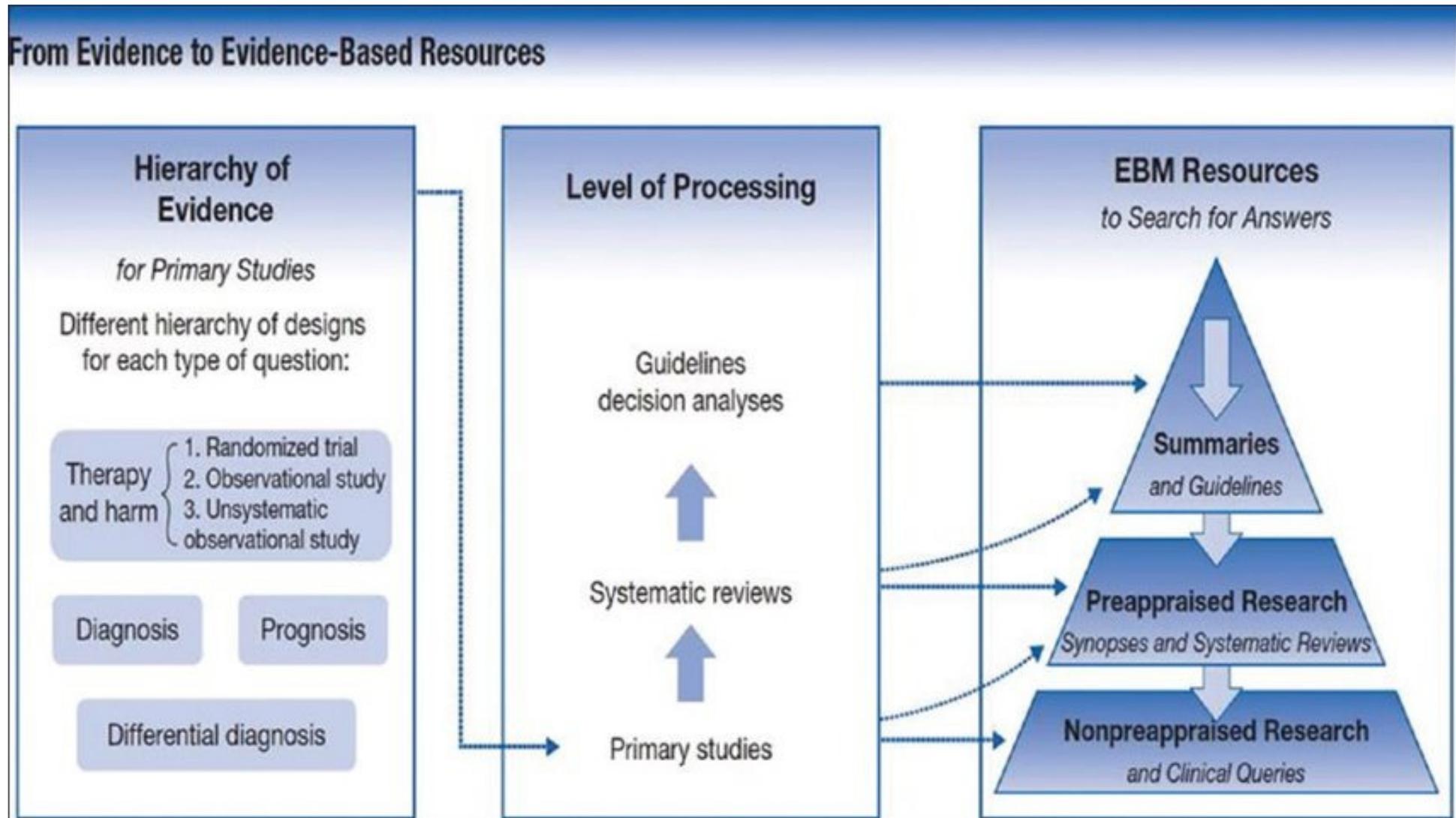
**Patients' perceptions of the mechanisms underlying alcohol use problems after bariatric surgery: A qualitative systematic review.**

Er, Esin ; Durieux, Nancy ; Vander Haegen, Marie  *et al.*

2022 • In *Clinical Obesity*, p. 12551



Grosdent, S., Demoulin, C., Aguilaniu, A., Hidalgo, B., Léonard, F. & Durieux, N. (2024), Effectiveness of manual techniques, exercise therapy or combined treatments in the management of athletes with ankle sprain or chronic ankle instability: a systematic review protocol [Manuscript submitted for publication].





# APA CLINICAL PRACTICE GUIDELINE for the Treatment of Depression Across Three Age Cohorts

GUIDELINE DEVELOPMENT PANEL FOR THE TREATMENT OF DEPRESSIVE DISORDERS

APPROVED BY APA COUNCIL OF REPRESENTATIVES  
FEBRUARY 2019



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION

Exemple : chez les enfants

## SIDEBAR 3

### **EMPIRICALLY SUPPORTED TREATMENT RECOMMENDATIONS**

- **Insufficient evidence** to recommend one treatment over another for children with depression.
- **Recommend** cognitive-behavioral therapy or interpersonal psychotherapy for adolescents with depression.
- If considering pharmacotherapy for adolescents with MDD, **recommend** fluoxetine as a first-line medication.

*To view the full set of recommendations, including conditional recommendations and recommendations against treatments, please refer to [Tables 1](#) and [2](#) of the full guideline document.*



TABLE 5

**Summary of Systematic Reviews and Meta-Analyses Used for Each Age Group**

Age Group	Systematic Reviews and Meta-Analyses Used
Children and Adolescents	<ul style="list-style-type: none"><li>• “Comparative efficacy and acceptability of psychotherapies for depression in children and adolescents: A systematic review and network meta-analysis” (Zhou et al., 2015)</li><li>• “Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: A network meta-analysis” (Cipriani et al., 2016)</li></ul>
General Adult Population	<ul style="list-style-type: none"><li>• “Nonpharmacological versus pharmacological treatments for adult patients with major depressive disorder” (Gartlehner et al., 2015)</li><li>• “Management of major depressive disorder, Evidence synthesis report, Clinical practice guideline” (ECRI Institute, 2015)</li><li>• “The efficacy of non-directive supportive therapy for depression” (Driessen, et al., 2012a)</li><li>• “Psychotherapy for subclinical depression” (Cipriani et al., 2015)</li><li>• “The efficacy of short-term psychodynamic therapy for depression” (Cipriani et al., 2015)</li><li>• “Interpersonal Psychotherapy for Mental Health” (Cipriani et al., 2016)</li></ul>
Older Adults	<ul style="list-style-type: none"><li>• “Managing depression in older age: Psychotherapy” (Cipriani et al., 2016)</li><li>• “Continuation and maintenance treatment of depression in older adults” (Cipriani et al., 2016)</li></ul>

In summary, each review used by the panel has strengths as well as limitations. Importantly, a review is not a final statement and has its own limits, but it allows the field to obtain a better handle on what information is lacking and what needs to be done in the future (e.g., more funding, better quality research, more focus on diverse populations, etc.). Information from these reviews was considered together with information on strength of the evidence base, balance of the benefits vs. harms/burdens of a treatment, patient values and preferences, and applicability to form the panel’s recommendations. In light of these limitations, it needs to be recognized that any recommendations made are based on evidence derived by a method with some drawbacks for psychotherapy research. Psychotherapy research differs in nature from medical research both in the number of variables effecting outcomes and the amount of research that met IOM standards. Results should therefore be read as best evidence according to method used and viewed as a guideline not a prescription for treatment funding and decision-making.



**Implémentation du guide de pratique belge « La dépression chez l'adulte » auprès des médecins généralistes, avec un accent particulier sur l'abandon progressif des antidépresseurs.**

Sur cette thématique, le SPF Santé Publique en collaboration avec le réseau Evikey lancera en 2024 un appel d'offre !

### Objectifs de la mission :

1. AUGMENTER LE CHANGEMENT D'ATTITUDE ET D'EFFICACITÉ PERSONNELLE DES MÉDECINS GÉNÉRALISTES EN CE QUI CONCERNE L'ABANDON PROGRESSIF DES AD

2. DIMINUER LES PRESCRIPTIONS DE AD DE LONGUE DURÉE

3. AMÉLIORER LA COLLABORATION MULTIDISCIPLINAIRE ENTRE LES MÉDECINS GÉNÉRALISTES ET LES PSYCHOLOGUES



# THE JBI MODEL OF EVIDENCE-BASED HEALTHCARE



**Overarching principles**  
Culture - Capacity - Communication - Collaboration



## *Research synthesis : sciences*

- ▶ Groupe de chercheurs qui ...
  - travaillent au développement des méthodes de synthèse
  - évaluent les méthodes en recherche de synthèse
  - développent des outils d'évaluation des études
  - développent des logiciels permettant d'appliquer les méthodes

de manière spécifique (à l'objet d'étude) et autonome



METHODOLOGY

## Moving from consultation to co-creation with knowledge users in scoping reviews: guidance from the JBI Scoping Review Methodology Group

Pollock, Danielle<sup>1</sup>; Alexander, Lyndsay<sup>2,3</sup>; Munn, Zachary<sup>1</sup>; Peters, Micah D.J.<sup>4,5,6</sup>; Khalil, Hanan<sup>7,8</sup>; Godfrey, Christina M.<sup>9</sup>; McInerney, Patricia<sup>10</sup>; Synnot, Anneliese<sup>11,12</sup>; Tricco, Andrea C.<sup>9,13,14</sup>

[Author Information](#) ⓘ

*JBI Evidence Synthesis* 20(4):p 969-979, April 2022. | DOI: 10.11124/JBIES-21-00416



### Research Methods & Reporting

## Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness

*BMJ* 2024 ; 384 doi: <https://doi.org/10.1136/bmj-2023-076335> (Published 06 February 2024)

Cite this as: *BMJ* 2024;384:e076335



*Méthodes développées par la science de synthèse*

*Research synthesis :*  
synthèse des résultats (thématiques)  
(qui peut inclure une analyse critique  
des études)

*Meta-research :*  
synthèse des pratiques de recherche  
(méthodes) et analyse de leur  
influence sur les résultats



## Example

### Knowledge syntheses in medical education: Meta-research examining author gender, geographic location, and institutional affiliation

Lauren A. Maggio , Anton Ninkov, Joseph A. Costello, Erik W. Driessen, Anthony R. Artino Jr.

Published: October 26, 2021 • <https://doi.org/10.1371/journal.pone.0258925>

#### **Introduction**

Authors of knowledge syntheses make many subjective decisions during their review process. Those decisions, which are guided in part by author characteristics, can impact the conduct and conclusions of knowledge syntheses, which assimilate much of the evidence base in medical education. To better understand the evidence base, this study describes the characteristics of knowledge synthesis authors, focusing on gender, geography, and institution.

#### **Methods**

In 2020, the authors conducted meta-research to examine authors of 963 knowledge syntheses published between 1999 and 2019 in 14 core medical education journals.



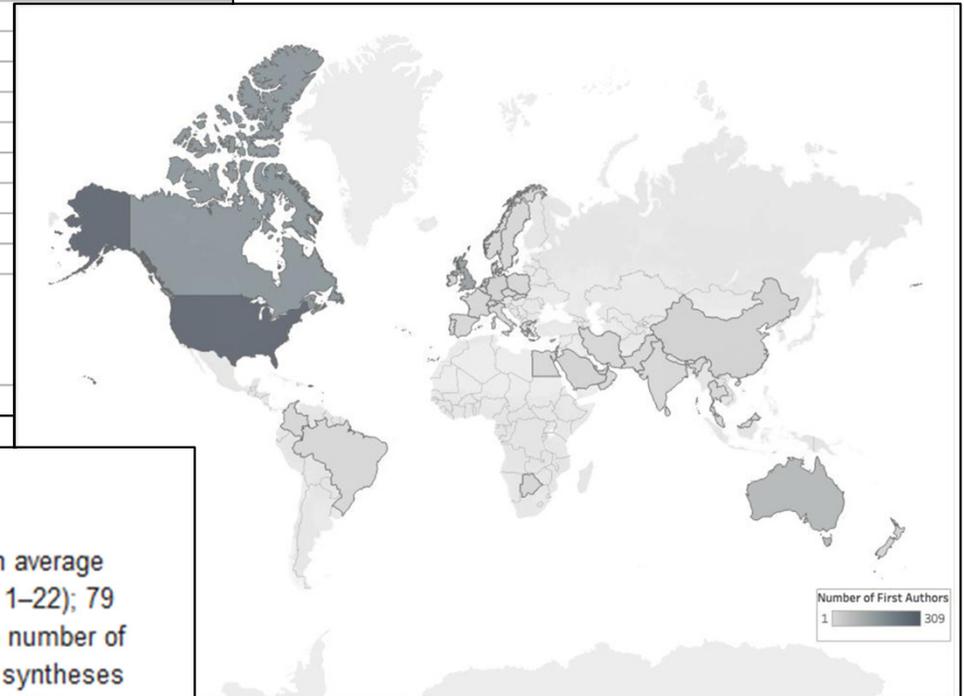
Institution (country)	Count of first authors (%)	Times Higher Education Ranking	Count of authors across all authorship positions (%)
University of Toronto (Canada)	56 (5.7)	18	212 (5.2)
Mayo Clinic (United States)	32 (3.3)	ORG	110 (2.7)
McMaster University (Canada)	18 (1.9)	72	64 (1.6)
Monash University (Australia)	18 (1.9)	75*	72 (1.8)
University of Ottawa (Canada)	18 (1.9)	141*	72 (1.8)
National Health Services (United Kingdom)	17 (1.8)	ORG	70 (1.7)
University of British Columbia (Canada)	15 (1.6)	34	52 (1.3)
University of Utrecht (Netherlands)	14 (1.5)	75*	76 (1.8)
McGill University (Canada)	13 (1.3)	42	52 (1.3)
University of Alberta (Canada)	13 (1.3)	136*	54 (1.3)
University of Calgary (Canada)	13 (1.3)	201–250	47 (1.1)
Maastricht University (Netherlands)	13 (1.3)	127	89 (2.2)

Rankings retrieved from *Times Higher Education* (THE) World University Rankings 2020.

\* indicates a tie in THE rankings.

ORG = organizations (e.g., health system).

<https://doi.org/10.1371/journal.pone.0258925.t001>



## Results

The authors identified 4,110 manuscript authors across all authorship positions. On average there were 4.3 authors per knowledge synthesis (SD = 2.51, Median = 4, Range = 1–22); 79 knowledge syntheses (8%) were single-author publications. Over time, the average number of authors per synthesis increased (M = 1.80 in 1999; M = 5.34 in 2019). Knowledge syntheses were authored by slightly more females (n = 2047; 50.5%) than males (n = 2005; 49.5%) across all author positions. Authors listed affiliations in 58 countries, and 58 knowledge syntheses (6%) included authors from low- or middle-income countries. Authors from the United States (n = 366; 38%), Canada (n = 233; 24%), and the United Kingdom (n = 180; 19%) published the most knowledge syntheses. Authors listed affiliation at 617 unique institutions, and first authors represented 362 unique institutions with greatest representation from University of Toronto (n = 55, 6%). Across all authorship positions, the large majority of knowledge syntheses (n = 753; 78%) included authors from institutions ranked in the top 200 globally.

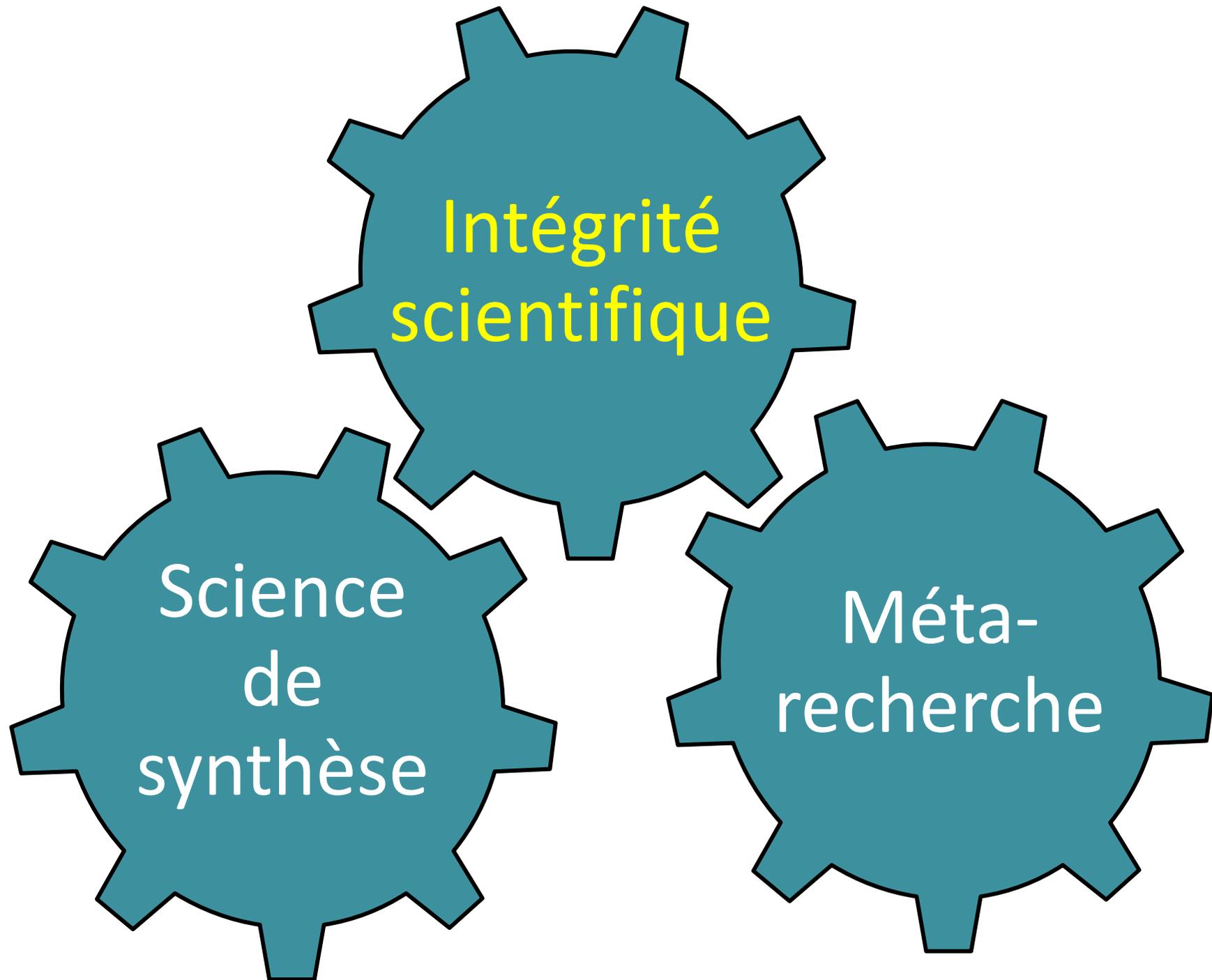
## Conclusion

Knowledge synthesis author teams have grown over the past 20 years, and while there is near gender parity across all author positions, authorship has been dominated by North American researchers located at highly ranked institutions. This suggests a potential overrepresentation of certain authors with particular characteristics, which may impact the conduct and conclusions of medical education knowledge syntheses.



## Autres exemples – projets en cours

- ▶ Effet du genre et du format des questions sur la performance scolaire  
Léonard, F., Durieux, N., (Tirelli, E.), & Monseur, C.
- ▶ Biais de genre dans les pratiques d'*authorship*  
Durieux, N. & Léonard, C.
- ▶ *Collaboration between multilingual teams when conducting a (systematic) review*  
Durieux, N. & Corremans, M. (développement en cours)





**Le système de la science et l'irreproductibilité de beaucoup (trop) de résultats scientifiques (surtout en recherche confirmatoire)**



## The pressure to publish pushes down quality

*Scientists must publish less, says Daniel Sarewitz, or good research will be swamped by the ever-increasing volume of poor work.*

I am pleased to announce that as of the middle of April, my Elsevier publications had received 30,752 page views and 2,025 citations. I got these numbers in a promotional e-mail from Elsevier, and although I'm not sure what they mean, I presume that it would be even better to have even bigger numbers.

turned out to have been a melanoma cell line. The average biomedical research paper gets cited between 10 and 20 times in 5 years, and as many as one-third of all cell lines used in research are thought to be contaminated, so the arithmetic is easy enough to do: by one estimate, 10,000 published papers a year cite work based on contami-



## Our obsession with eminence warps research

*Many decisions about whose work is recognized are at least partially arbitrary, and we should acknowledge that, argues Simine Vazire.*

We can quantify exactly how much faster Usain Bolt is than the next-fastest sprinter. It's much harder to say who is the best scientist, let alone how much better they are than the next-best scientist. Deciding who deserves recognition is, at least in part, a judgement call.

described nearly half a century ago by the sociologist of science Robert Merton (R. Merton *Science* 159, 56–63; 1968).

When eminence begets eminence, noise in the system gets amplified. There's an element of luck to who ends up having the most success, and that luck will build on itself.



# **Les pratiques de recherche problématiques (PRP) ou les manquements à l'intégrité scientifique**



# Intégrité scientifique et bonnes pratiques de recherche (BPR)

= pratiques responsables en recherche

-Concernent **l'ensemble du processus de recherche**,  
**avant tout** (sens restreint) la genèse de l'idée scientifique, la réalisation des travaux de recherche, l'interprétation des résultats et leur communication à la communauté scientifique.

-Constituent le meilleur moyen de garantir l'obtention d'un  
**résultat scientifique fiable**,  
à savoir **répétable, reproductible, répliquable, robuste et généralisable**, et donc  
crédible, y compris auprès de la société.

-Constituent les  
**piliers de L'INTÉGRITÉ SCIENTIFIQUE**,  
**qui est UNE NECESSITE INTELLECTUELLE.**



L'intégrité scientifique implique des **comportements et attitudes responsables** se fondant notamment sur

**l'honnêteté,**  
**l'objectivité,**  
**la complétude,**  
**le désintéressement,**  
**le scepticisme organisé (méthodique),**  
**l'esprit (auto- et allo-)critique,**  
**l'autocorrection,**  
**l'indépendance,**  
**l'acceptation de l'incertitude (des résultats et des conclusions),**  
**l'ouverture recherchée vers d'autres points de vue,**  
**la rigueur,**  
**la transparence.**

Ces deux derniers pouvant (après tout) résumer les autres.





## Deux catégories de PRATIQUES DE RECHERCHE PROBLÉMATIQUES (PRP) ou manquements à l'intégrité scientifique

### -A- Les **PRATIQUES DE RECHERCHE FRAUDULEUSES (PRF)**

Fabrication, falsification et plagiat (FFP).

**Intention délibérée de tromper.**

(aussi: méconduites scientifiques, de l'anglais « misconducts »,  
utilisé dans le code éthique du CNRS).

### -B- Les **PRATIQUES DE RECHERCHE INADÉQUATES /douteuses (PRI)**

ou pratiques détritantes à la recherche (NASEM, USA)  
classiquement « Questionable Research Practices » (QRP)

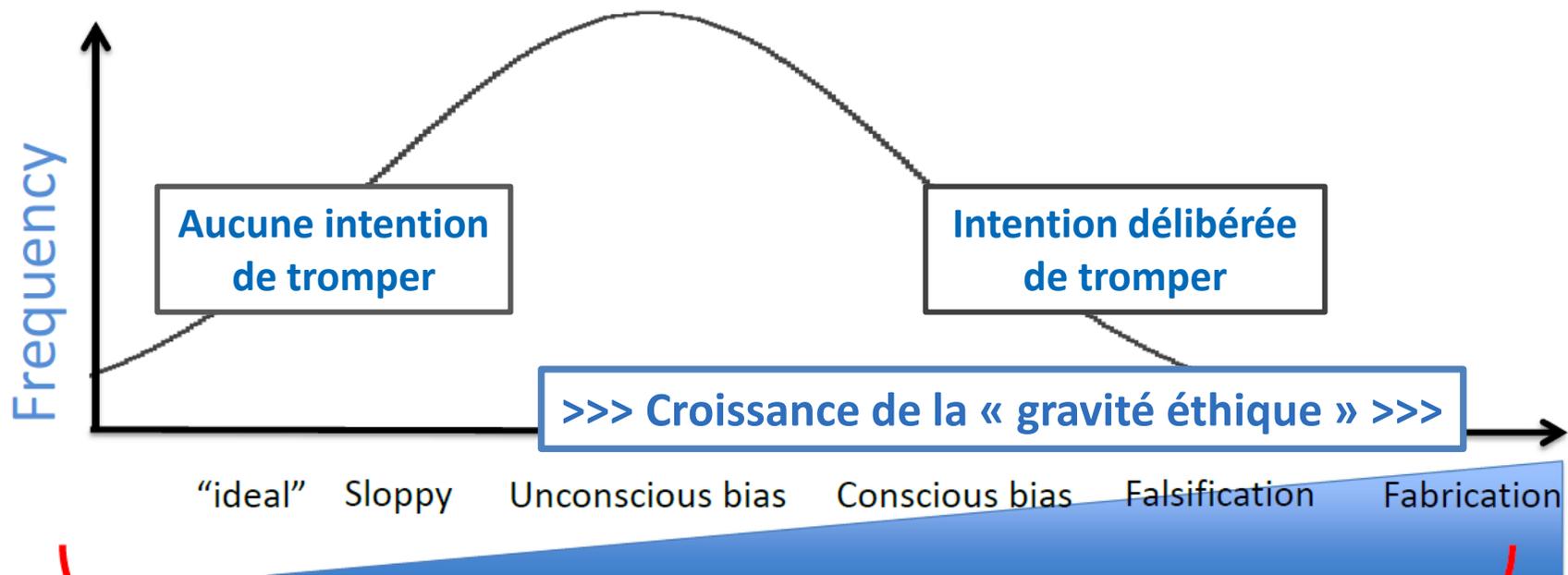
**Pas d'intention de tromper.**

Biais inconscients ou conscients (« les causes »):  
méconnaissances et « fausses croyances »,  
convictions viscérales (« gut feeling ») et « conformisme malhonnête ».  
Laxisme et bâclage.



## Vision graduée des PRP

Les PRP liés aux biais inconscients, aux biais conscients ou au laxisme sont **par leur haute fréquence** les plus détrimentaires à la qualité de la recherche et à la reproductibilité des résultats scientifiques, et donc à leur fiabilité.



**All will increase positive-outcome bias**



Principales grandes catégories de **pratiques de recherches problématiques (manquements à l'intégrité scientifique)** adoptées dans les codes éthiques d'universités, de centres de recherche, d'organes de financement de la recherche ou d'autres organismes internationaux (ex. OCDE).

-La **réalisation de la recherche**, la présentation et l'interprétation des résultats (catégorie parfois scindée en deux: PRI concernant les données ou les méthodes).

-Le **processus de publication** et de dissémination des recherches et de leurs résultats.

-La conservation et la **gestion des données**.

-Les **relations avec les collègues** et tout autre acteur du monde de la recherche ou de l'université.

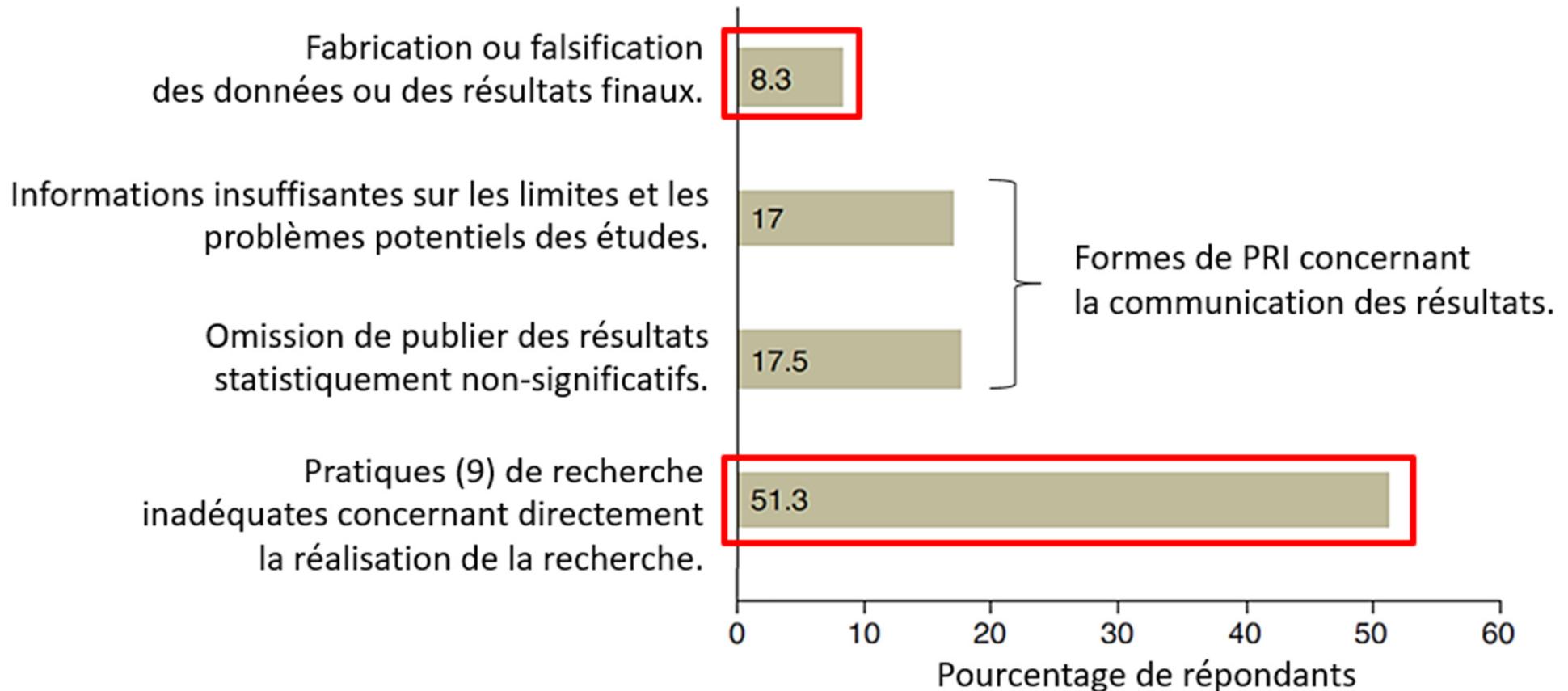
-Les relations avec les **sources de financements** privées (indépendance).

-Le processus de **demande d'un financement** pour la recherche.

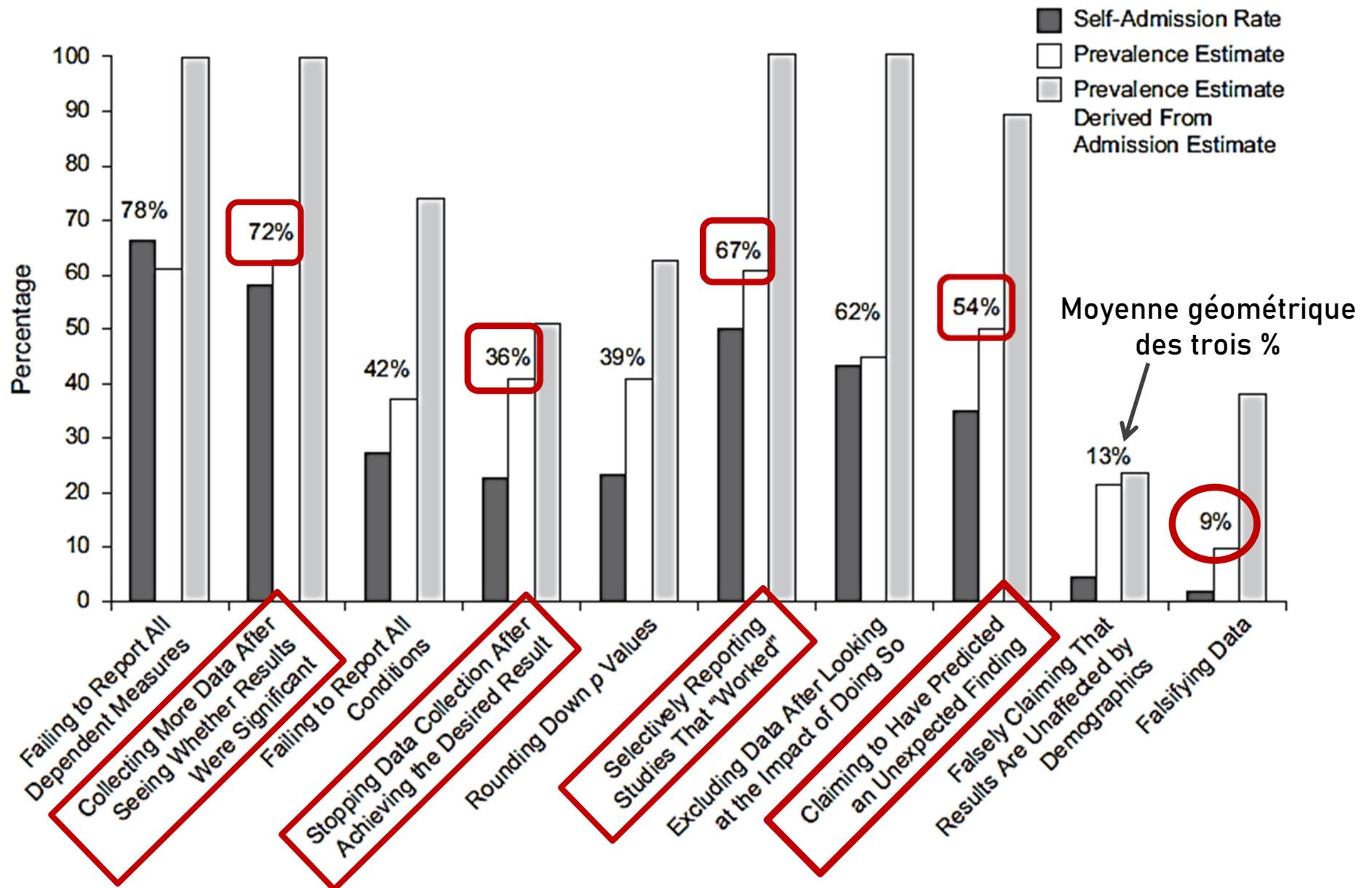
-L'expertise pour des tiers et **évaluations scientifiques**.



Enquête anonyme (questionnaires) auprès de 6.813 chercheurs universitaires néerlandais (60% en recherche biomédicale) concernant 11 PRP effectuées dans les trois dernières années (pourcentages d'admissions).



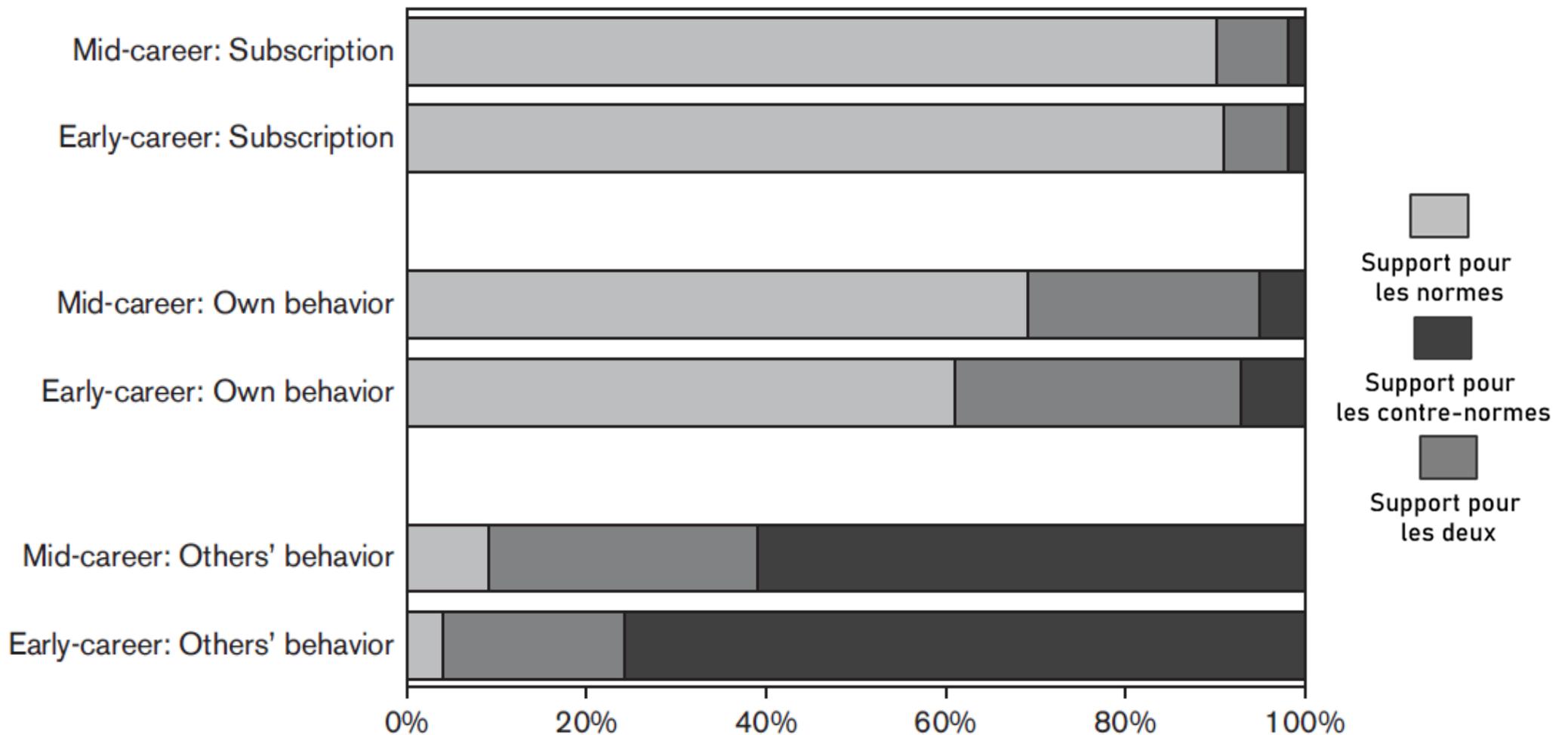
Enquête anonyme (questionnaires) concernant 10 pratiques méthodologiques problématiques (lors des trois dernières années) auprès de 2.155 membres de départements de psychologie US.





Enquête auprès de 3.247 chercheurs US en début ou en milieu de carrière portant sur les attitudes, croyances et pratiques s'inscrivant dans six duos de **normes** et **contre-normes éthiques** (R. Merton) admises pour soi ou pour autrui, et révélant une « **dissonance normative** » chez une bonne partie des chercheurs.

Chaque duo comprend un comportement priorisant les **aspects qualitatifs et transparents** (universalisme, communalisme, désintéressement, scepticisme organisé...) et son « opposé » motivé par les **aspects quantitatifs et compétitifs** (secret, particularisme, auto-promotion, dogmatisme...) de la recherche.





# The Four Horsemen of the Reproducibility Crisis

Dorothy Bishop, Oxford Reproducibility Lecture (2017)

(La mort)

HARKing

(La famine)

Low power

(La guerre)

P-hacking

(La conquête)

Publication bias



(« Les Quatre Cavaliers de l'Apocalypse »  
tableau de Viktor Vasnetsov, 1887.)

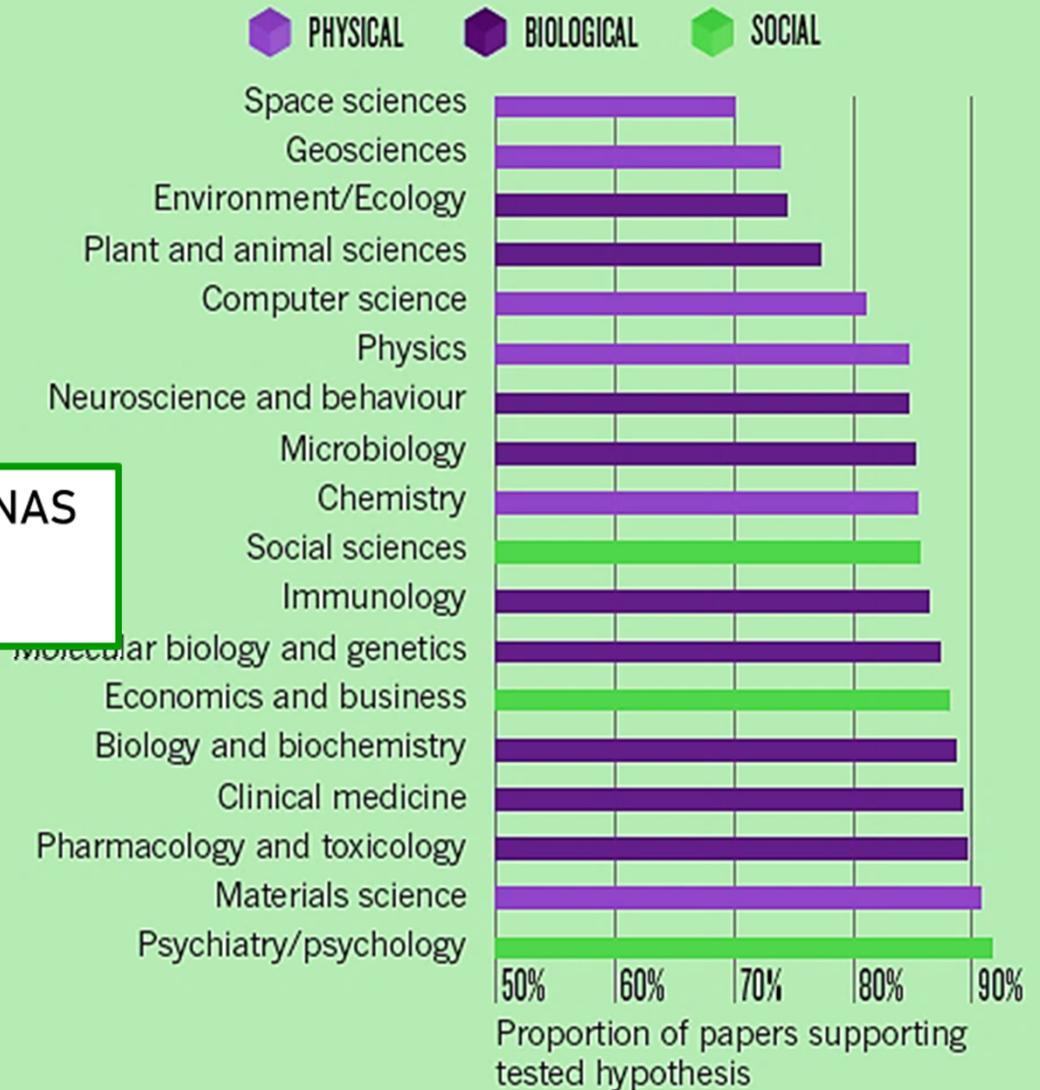
## Biais de publication:

omission de la publication de résultats statistiquement non-significatifs, ce qui produit une prédominance et une surestimation des résultats dans la littérature.

Certaines années, NATURE, SCIENCE ou PNAS arborent 100% d'articles « positifs »  
(Cristea & Ioannidis 2018 PLOS ONE 13: e0197440).

# ACCENTUATE THE POSITIVE

A literature analysis across disciplines reveals a tendency to publish only 'positive' studies — those that support the tested hypothesis. Psychiatry and psychology are the worst offenders.





**HARK-ing**  
**H:** hypothesizing  
**A:** after  
**R:** (the) results  
**K:** (are) known



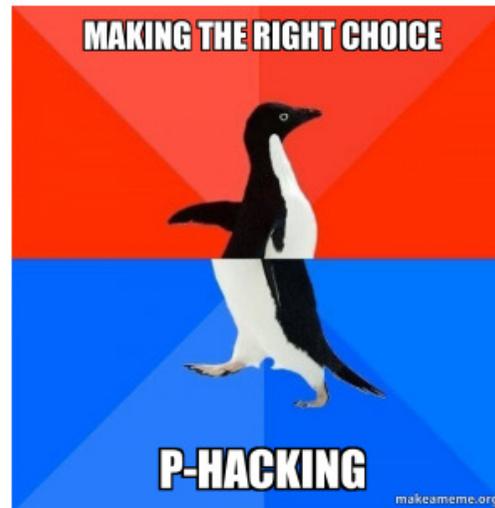
- ▶ = Présenter une proposition comme une hypothèse **APRES** avoir analysé les données de manière plus ou moins exploratoire et examiné les résultats, comme si l'étude était confirmatoire.
- ▶ Le **taux d'admissions auto-déclarées ou relevées chez autrui** par les chercheurs (enquêtes anonymes) peut largement dépasser **50%**

Par exemple

John et al., 2012 - <https://doi.org/10.1177/0956797611430953>;

Fraser et al., 2018 - <https://doi.org/10.1371/journal.pone.0200303>;

Makel et al., 2021 - <https://doi.org/10.3102/0013189X211001356>



= Intervention sur les données et leur analyse statistique **visant à générer un effet statistiquement significatif, une première analyse ne donnant pas lieu à un tel effet.**

La **flexibilité analytique** y est exploitée volontairement ou par ignorance (multiples possibilités inadéquates d'analyse des données).

Le **taux d'admissions auto-déclarées ou relevées chez autrui** par les chercheurs (enquêtes anonymes) dépasse **50 %**

Par exemple

John et al., 2012 - <https://doi.org/10.1177/0956797611430953>

Fraser et al., 2018 - <https://doi.org/10.1371/journal.pone.0200303>

Makel et al., 2021 - <https://doi.org/10.3102/0013189X211001356>



Quelques actes (souvent combinables) de P-hacking (1)  
**Si la p-valeur est plus grande que le seuil de significativité  
(effet statistiquement non-significatif), on procède à:**

- l'ajout successif d'observations au « n » initial ou précédent avec arrêts optionnels  
(aussi: arrêt de la collecte des données en renonçant à l'effectif prévu dès l'obtention d'un effet statistiquement significatif),
- l'élimination des valeurs extrêmes sans règle préalable clairement justifiée,
- l'adoption a posteriori d'un seuil critique « qui convient »,
- l'oubli (suppression) des corrections du risque alpha (peut se faire avant toute collecte de données),
- l'ajout ou au retrait de certaines variables indépendantes ou de covariables,
- au choix de la « meilleure » mesure ou variable dépendante,



Quelques actes (souvent combinables) de P-hacking (2)  
**Si la p-valeur est plus grande que le seuil de significativité (effet statistiquement non-significatif), on procède à:**

- l'ajustement (modifications) de la mesure (variable dépendante),
- l'essais de tous les tests statistiques possibles (même inadéquats),
- au changement ou à l'ajout du ou des groupe(s) contrôle(s),
- l'utilisation récurrente d'un même groupe car « il marche »,
- la fusion ou la division des groupes,
- au changement de variable dépendante primaire ou secondaire,
- l'essais de toutes les transformations de données possibles.



# La sous-puissance

- La puissance statistique prospective est un **pilier majeur de la qualité méthodologique** d'une recherche confirmatoire.
- Elle constitue la probabilité que le test statistique d'une étude donne lieu à **un effet statistiquement significatif, QUAND IL EXISTE** (pour une taille d'effet attendue donnée et un risque alpha donné).
- Une puissance élevée (au moins 0,8 ; et la meilleure des rigueurs dans la réalisation de l'étude), qui implique **un effectif adéquat**, augmente les chances que l'effet soit reproductible.

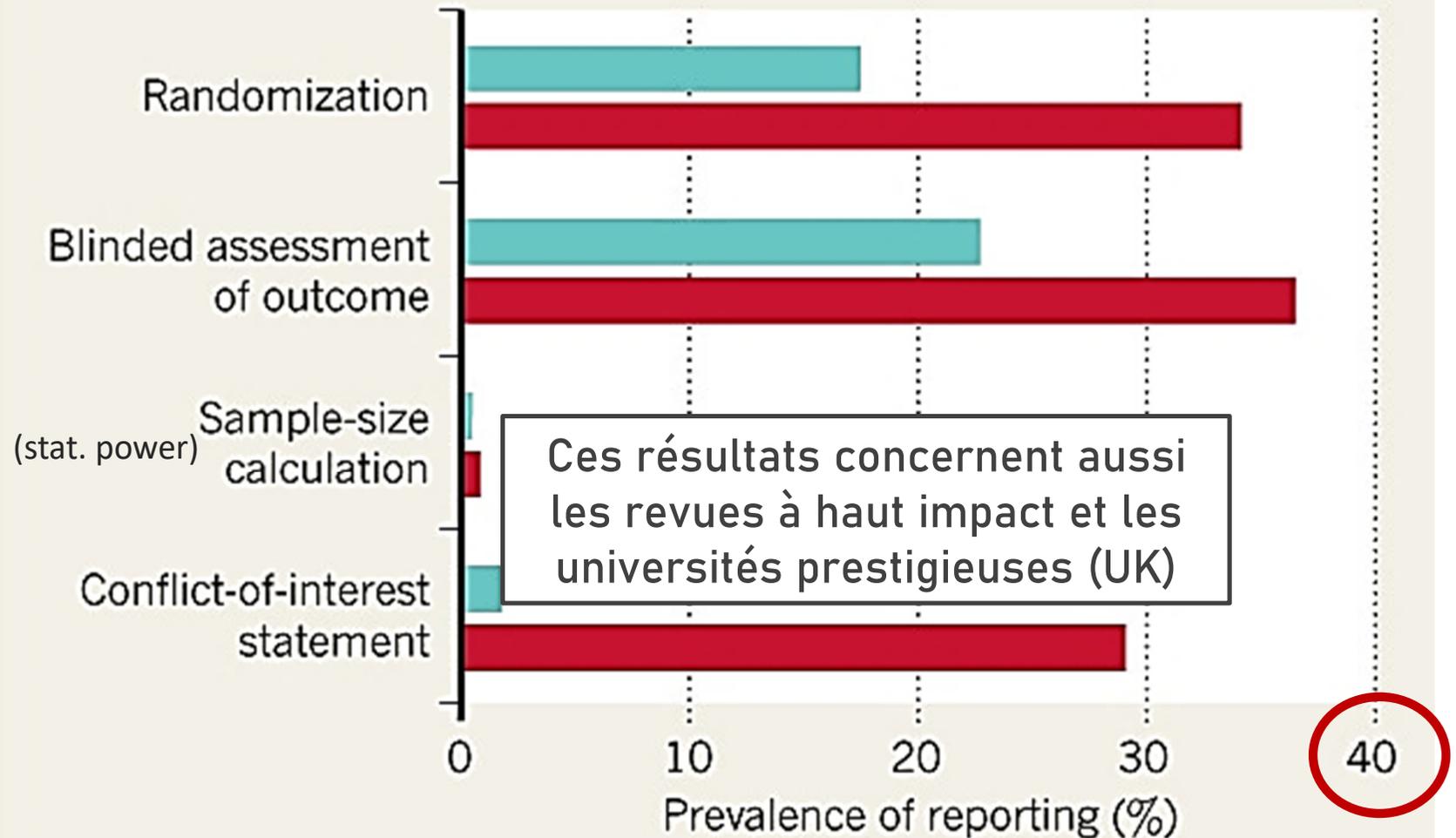
Attention: la puissance n'est pas une mesure de la reproductibilité.

- Une puissance faible (sous-puissance) ne peut que diminuer les chances d'obtention d'un effet fiable (même statistiquement significatif).**

# ANIMAL STUDIES POORLY DESIGNED

Studies testing drugs in animals rarely report the use of basic methods to avoid biased conclusions.

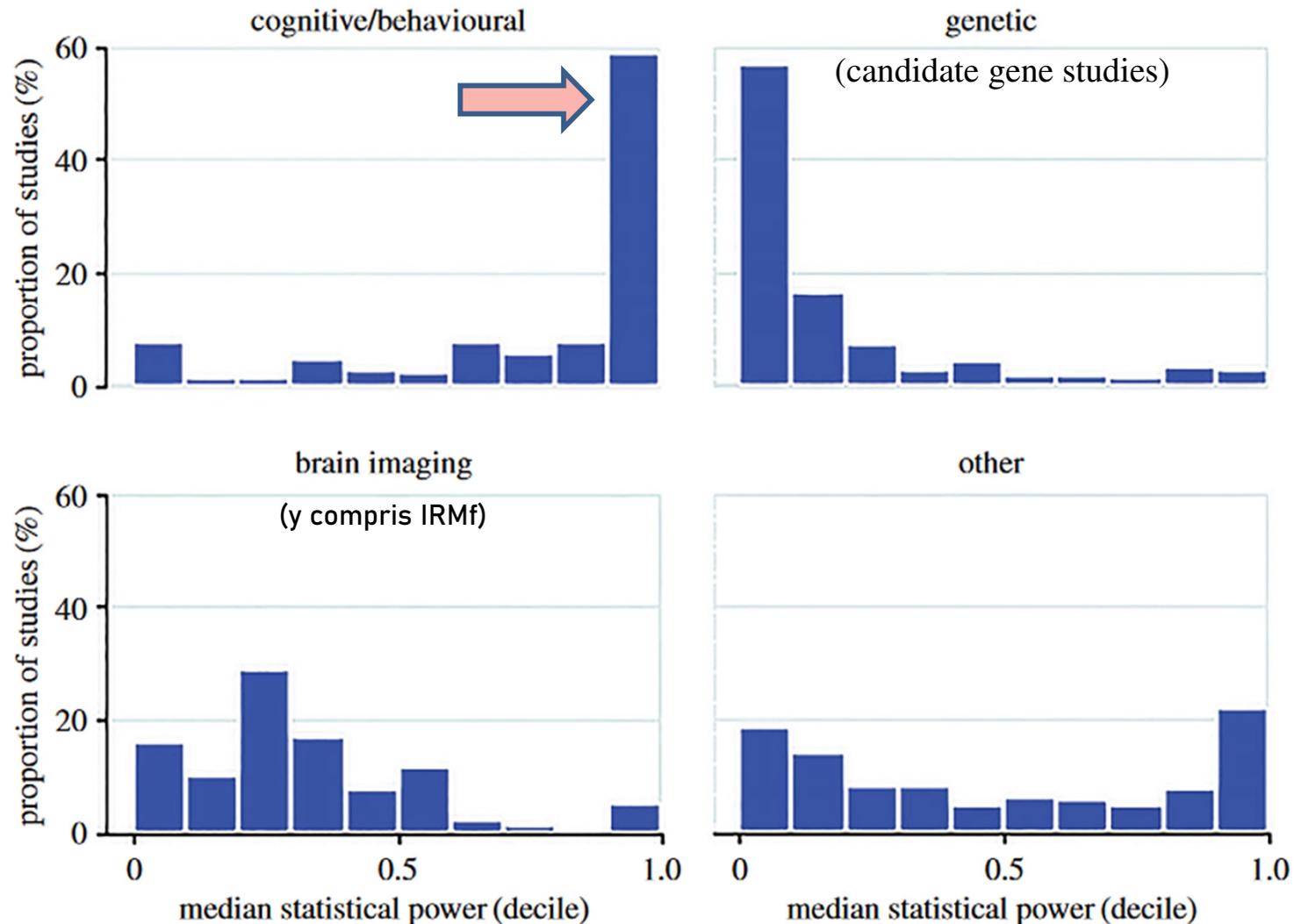
■ 1992–98 ■ 2008–11





## Sous-puissance dans les domaines neuro-psycho-médicaux

Distribution des puissances impliquées en fonction du type d'études.  
Les puissances ont été calculées pour les articles empiriques couverts par plusieurs centaines de méta-analyses dont les tailles d'effet synthétiques ont servi à cette fin (paramètre estimé plus proche de celui de la population cible).





**En guise de conclusion**



## Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition

Marc A. Edwards<sup>\*,†</sup> and Siddhartha Roy<sup>†</sup>

*Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, Virginia.*

*Received: April 25, 2016*

*Accepted in revised form: August 18, 2016*

---

### **Abstract**

Over the last 50 years, we argue that incentives for academic scientists have become increasingly perverse in terms of competition for research funding, development of quantitative metrics to measure performance, and a changing business model for higher education itself. Furthermore, decreased discretionary funding at the federal and state level is creating a hypercompetitive environment between government agencies (e.g., EPA, NIH, CDC), for scientists in these agencies, and for academics seeking funding from all sources—the combination of perverse incentives and decreased funding increases pressures that can lead to unethical behavior. If a critical mass of scientists become untrustworthy, a tipping point is possible in which the scientific enterprise itself becomes inherently corrupt and public trust is lost, risking a new dark age with devastating consequences to humanity. Academia and federal agencies should better support science as a public good, and incentivize altruistic and ethical outcomes, while de-emphasizing output.

**Keywords:** academic research; funding; misconduct; perverse incentives; scientific integrity

“Colleagues, Let’s Admit that We Have a Problem.

[...] we must acknowledge that [...] all of **our careers have benefitted from the research practices that we now realize to be questionable.**”

Russell A. POLDRACK

Stanford Center for Reproducible Neuroscience

Copyrighted Material  
MULTIVARIATE APPLICATIONS SERIES



2011

# Handbook of Ethics in Quantitative Methodology

A. T. Panter • Sonya K. Sterba

Copyrighted Material

Copyrighted Material

2022

# Research Integrity

*Best Practices for the Social  
and Behavioral Sciences*



*Edited by*  
Lee Jussim,  
Jon A. Krosnick,  
and Sean T. Stevens

Copyrighted Material

OXFORD

Merci pour votre attention

**Nancy Durieux**  
**François Léonard**  
**Ezio Tirelli**

Faculté de Psychologie, Logopédie  
et Sciences de l'Education