

Pratique de recherche discutables et crise de la réplication

Elisa Sarda

12/12/2018

Pratiques de recherche discutables

Quelques éléments d'introduction

Objectif de la recherche : découvrir des vérités

Pratiques de recherche discutables

Quelques éléments d'introduction

Objectif de la recherche : découvrir des vérités

Utilisation d'une méthode scientifique pour tester des hypothèses

Pratiques de recherche discutables

Quelques éléments d'introduction

Objectif de la recherche : découvrir des vérités

Utilisation d'une méthode scientifique pour tester des hypothèses

Recherche de qualité :

- Conclusions valides
- Contributions au savoir
- Publication dans les meilleurs journaux

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

-Participer à une recherche de qualité

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

-Participer à une recherche de qualité

-Provoquer l'adoption des pratiques de recherche discutables (Giner & Sorolla 2012)

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

-Participer à une recherche de qualité

-Provoquer l'adoption des pratiques de recherche discutables (Giner & Sorolla 2012)

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

-Participer à une recherche de qualité

-Provoquer l'adoption des pratiques de recherche discutables (Giner & Sorolla 2012)

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

- Rendre ses données plus conformes aux hypothèses
- Modifier des graphiques des résultats pour les rendre plus intéressants

Pratiques de recherche discutables

Quelques éléments d'introduction

Publication dans les meilleurs périodiques

-> Compétition

-Participer à une recherche de qualité

-Provoquer l'adoption des pratiques de recherche discutables (Giner & Sorolla 2012)

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

- Rendre ses données plus conformes aux hypothèses
- Modifier des graphiques des résultats pour les rendre plus intéressants

> **Crise de la répliquabilité des résultats de recherche (Maxwell et al., 2015)**

Pratiques de recherche discutables

Quelques éléments d'introduction

Cela est vrai pour plusieurs disciplines (biologie, médecine, psychologie, neuroscience...)

Pratiques de recherche discutables

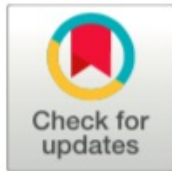
META-RESEARCH ARTICLE

Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature

Denes Szucs^{1*}, John P. A. Ioannidis²

1 Department of Psychology, University of Cambridge, Cambridge, United Kingdom, **2** Meta-Research Innovation Center at Stanford (METRICS) and Department of Medicine, Department of Health Research and Policy, and Department of Statistics, Stanford University, Stanford, California, United States of America

* ds377@cam.ac.uk



Abstract

We have empirically assessed the distribution of published effect sizes and estimated power by analyzing 26,841 statistical records from 3,801 cognitive neuroscience and psychology papers published recently. The reported median effect size was $D = 0.93$ (interquartile range: 0.64–1.46) for nominally statistically significant results and $D = 0.24$ (0.11–0.42) for nonsignificant results. Median power to detect small, medium, and large effects was 0.12, 0.44, and 0.73, reflecting no improvement through the past half-century. This is so because sample sizes have remained small. Assuming similar true effect sizes in both disciplines, power was lower in cognitive neuroscience than in psychology. Journal impact fac-

OPEN ACCESS

Citation: Szucs D, Ioannidis JPA (2017) Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol 15(3): e2000797. doi:10.1371/

Pratiques de recherche discutables

CANCER BIOLOGY

Replication Study: Intestinal inflammation targets cancer-inducing activity of the microbiota

Kathryn Eaton et al.

Editors' Summary: This Replication Study has reproduced some parts of the original paper but it also contains results that are not consistent with other parts of the original paper.

REPLICATION STUDY Oct 3, 2018

[HTML + PDF](#)

CANCER BIOLOGY

Replication Study: *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma

John Repass, Reproducibility Project: Cancer Biology

Editors' Summary: This Replication Study did not reproduce those experiments in the original paper that it attempted to reproduce.

REPLICATION STUDY Mar 13, 2018

[HTML + PDF](#)

Pratiques de recherche discutables

PERSPECTIVE

The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman^{1*}, Iain M. Cockburn², Timothy S. Simcoe^{2,3}

1 Global Biological Standards Institute, Washington, D.C., United States of America, **2** Boston University School of Management, Boston, Massachusetts, United States of America, **3** Council of Economic Advisers, Washington, D.C., United States of America

* lfreedman@gbsi.org

Abstract

Low reproducibility rates within life science research undermine cumulative knowledge production and contribute to both delays and costs of therapeutic drug development. An analysis of past studies indicates that the cumulative (total) prevalence of irreproducible preclinical research exceeds 50%, resulting in approximately US\$28,000,000,000 (US \$28B)/year spent on preclinical research that is not reproducible—in the United States alone. We outline a framework for solutions and a plan for long-term improvements in reproducibility rates that will help to accelerate the discovery of life-saving therapies and cures.



CrossMark
click for updates

OPEN ACCESS

Citation: Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. doi:10.1371/journal.pbio.1002165

Published: June 9, 2015

Pratiques de recherche discutables

Quelques éléments d'introduction

Cela est vrai pour plusieurs disciplines (biologie, médecine, psychologie, neuroscience...)

Pratiques de recherche discutables

Quelques éléments d'introduction

Cela est vrai pour plusieurs disciplines (biologie, médecine, psychologie, neuroscience...)

Nous on est psychologues, **on illustrera surtout des exemples en psycho**

Pratiques de recherche discutables

Crise de la réplication en psychologie

RESEARCH

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

INTRODUCTION: Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. Scientific claims should not gain credence because of the status or authority of their originator but by the replicability of their supporting evidence. Even research of exemplary quality may have irreproducible empirical findings because of random or systematic error.

viously observed finding and is the means of establishing reproducibility of a finding with new data. We conducted a large-scale, collaborative effort to obtain an initial estimate of the reproducibility of psychological science.

RESULTS: We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when

substantial decline. Ninety-seven percent of original studies had significant results ($P < .05$). Thirty-six percent of replications had significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

CONCLUSION: No single indicator sufficiently describes replication success, and the five indicators examined here are not the only ways to evaluate reproducibility. Nonetheless, collectively these results offer a clear conclu-

ON OUR WEB SITE
Read the full article at <http://dx.doi.org/10.1126/science.aac4716>

September 3, 2015

Pratiques de recherche discutables

Crise de la réplication en psychologie RP:P

Pratiques de recherche discutables

Crise de la réplication en psychologie RP:P

Réplication de 100 études expérimentales et corrélationnelles en psychologie

Pratiques de recherche discutables

Crise de la réplication en psychologie RP:P

Réplication de 100 études expérimentales et corrélationnelles en psychologie

- 270 chercheurs, 64 universités, 11 pays

Pratiques de recherche discutables

Crise de la réplication en psychologie RP:P

Réplication de 100 études expérimentales et corrélationnelles en psychologie

- 270 chercheurs, 64 universités, 11 pays

-Commencé en 2011 et publié en 2015 dans Science

Pratiques de recherche discutables

Crise de la réplication en psychologie RP:P

Réplication de 100 études expérimentales et corrélationnelles en psychologie

- 270 chercheurs, 64 universités, 11 pays

-Commencé en 2011 et publié en 2015 dans Science

Résultats :

- 36 % des études répliqués ont des résultats significatifs

Pratiques de recherche discutables

Crise de la réplication en psychologie

Replication

Investigating Variation in Replicability

A “Many Labs” Replication Project

Richard A. Klein,¹ Kate A. Ratliff,¹ Michelangelo Vianello,² Reginald B. Adams Jr.,³
Štěpán Bahník,⁴ Michael J. Bernstein,⁵ Konrad Bocian,⁶ Mark J. Brandt,⁷ Beach Brooks,¹
Claudia Chloe Brumbaugh,⁸ Zeynep Cemalcilar,⁹ Jesse Chandler,^{10,36} Winnee Cheong,¹¹
William E. Davis,¹² Thierry Devos,¹³ Matthew Eisner,¹⁰ Natalia Frankowska,⁶ David Furrow,¹⁵
Elisa Maria Galliani,² Fred Hasselman,^{16,37} Joshua A. Hicks,¹² James F. Hovermale,¹⁷
S. Jane Hunt,¹⁸ Jeffrey R. Huntsinger,¹⁹ Hans IJzerman,⁷ Melissa-Sue John,²⁰
Jennifer A. Joy-Gaba,¹⁷ Heather Barry Kappes,²¹ Lacy E. Krueger,¹⁸ Jaime Kurtz,²²
Carmel A. Levitan,²³ Robyn K. Mallett,¹⁹ Wendy L. Morris,²⁴ Anthony J. Nelson,³
Jason A. Nier,²⁵ Grant Packard,²⁶ Ronaldo Pilati,²⁷ Abraham M. Rutchick,²⁸
Kathleen Schmidt,²⁹ Jeanine L. Skorinko,²⁰ Robert Smith,¹⁴ Troy G. Steiner,³ Justin Storbeck,⁸
Lyn M. Van Swol,³⁰ Donna Thompson,¹⁵ A. E. van ‘t Veer,⁷ Leigh Ann Vaughn,³¹
Marek Vranka,³² Aaron L. Wichman,³³ Julie A. Woodzicka,³⁴ and Brian A. Nosek^{29,35}

¹University of Florida, Gainesville, FL, USA, ²University of Padua, Italy, ³The Pennsylvania State University, University Park, PA, USA, ⁴University of Würzburg, Germany, ⁵Pennsylvania State University Abington, PA, USA, ⁶University of Social Sciences and Humanities Campus Sopot, Poland, ⁷Tilburg University, The Netherlands, ⁸City University of New York, USA.

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

- Réplication de 13 effets en psychologie

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

- Réplication de 13 effets en psychologie
- 36 Laboratoires

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

- Réplication de 13 effets en psychologie
- 36 Laboratoires
- 6 344 participants

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

- Réplication de 13 effets en psychologie
- 36 Laboratoires
- 6 344 participants
- Utilisation du même matériel de l'étude original

Pratiques de recherche discutables

Crise de la réplication en psychologie Many Labs 1

- Réplication de 13 effets en psychologie
- 36 Laboratoires
- 6 344 participants
- Utilisation du même matériel de l'étude original

Résultats :

- Le têt de réplication des études est moindre de ce que l'on voudrait

Rick en parlera plus !

L'année 2011 en psycho

Cas d'une fraude : Diederik Stapel

L'année 2011 en psycho

Cas d'une fraude : Diederik Stapel

- Chercheur très productif en psychologie sociale

L'année 2011 en psycho

Cas d'une fraude : Diederik Stapel

- Chercheur très productif en psychologie sociale
- Il a travaillé à l'Université de Tilburg aux Pays-Bas

L'année 2011 en psycho

Cas d'une fraude : Diederik Stapel

- Chercheur très productif en psychologie sociale
- Il a travaillé à l'Université de Tilburg aux Pays-Bas

Il a falsifié et fabriqué ses données

L'année 2011 en psycho

Cas d'une fraude : Diederik Stapel

- Chercheur très productif en psychologie sociale
- Il a travaillé à l'Université de Tilburg aux Pays-Bas

Il a falsifié et fabriqué ses données

- 53 articles rétractés et 10 thèses de doctorat avec des données inventées ou douteuses



L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

General Article



False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
<http://pss.sagepub.com>
The SAGE logo consists of a circular icon containing a stylized 'S' followed by the word 'SAGE' in a bold, sans-serif font.

Joseph P. Simmons¹, Leif D. Nelson², and Uri Simonsohn¹

¹The Wharton School, University of Pennsylvania, and ²Haas School of Business, University of California, Berkeley

Abstract

In this article, we accomplish two things. First, we show that despite empirical psychologists' nominal endorsement of a low rate of false-positive findings ($\leq .05$), flexibility in data collection, analysis, and reporting dramatically increases actual false-positive rates. In many cases, a researcher is more likely to falsely find evidence that an effect exists than to correctly find evidence that it does not. We present computer simulations and a pair of actual experiments that demonstrate how unacceptably easy it is to accumulate (and report) statistically significant evidence for a false hypothesis. Second, we suggest a simple, low-cost, and straightforwardly effective disclosure-based solution to this problem. The solution involves six concrete requirements for authors and four guidelines for reviewers, all of which impose a minimal burden on the publication process.

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

- Est-ce qu'il faut récolter plus des données ?

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

- Est-ce qu'il faut récolter plus des données ?
- Faut-il exclure certaines observations ?

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

- Est-ce qu'il faut récolter plus des données ?
- Faut-il exclure certaines observations ?
- Faut-il tester l'interaction avec des variables non pertinentes ?

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

- Est-ce qu'il faut récolter plus des données ?
- Faut-il exclure certaines observations ?
- Faut-il tester l'interaction avec des variables non pertinentes ?
- Est-ce qu'il faut reporter toutes les conditions expérimentales dans le papier ?

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Dans cet article les chercheurs ont mené 2 expériences pour **démontrer quelque chose de faux** : écouter certaines chansons peut changer notre âge

Pourquoi ?

Les chercheurs ont de nombreuses décisions à prendre (**degré de liberté du chercheur**) :

- Est-ce qu'il faut récolter plus des données ?
- Faut-il exclure certaines observations ?
- Faut-il tester l'interaction avec des variables non pertinentes ?
- Est-ce qu'il faut reporter toutes les conditions expérimentales dans le papier ?

> **Le problème est que toute cette flexibilité laisse aux chercheurs, augmente la probabilité de trouver des résultats significatifs (alors que à la base ils ne le sont pas)**

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

Researcher degrees of freedom	Significance level		
	$p < .1$	$p < .05$	$p < .01$
Situation A: two dependent variables ($r = .50$)	17.8%	9.5%	2.2%
Situation B: addition of 10 more observations per cell	14.5%	7.7%	1.6%
Situation C: controlling for gender or interaction of gender with treatment	21.6%	11.7%	2.7%
Situation D: dropping (or not dropping) one of three conditions	23.2%	12.6%	2.8%
Combine Situations A and B	26.0%	14.4%	3.3%
Combine Situations A, B, and C	50.9%	30.9%	8.4%
Combine Situations A, B, C, and D	81.5%	60.7%	21.5%

Faux positifs et p-hacking (Simmons et al., 2011)

- Simmons et al démontrent qu'il est possible de trouver des résultats qui confirment les hypothèses des chercheurs et qui ont peu de chance d'être correctes

L'année 2011 en psycho

Faux positifs et p-hacking (Simmons et al., 2011)

- Simmons et al démontrent qu'il est possible de trouver des résultats qui confirment les hypothèses des chercheurs et qui ont peu de chance d'être correctes

-Leurs simulations montrent que **combiner différents « trucs » statistiques augmente la possibilité de trouver des effets significatifs qui n'existent pas**

Cela amorce l'idée des **pratiques questionnables de recherche et du p-hacking** (manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables)

L'année 2011 en psycho

Cette année nous a aidé à réaliser à quel point notre pratique de recherche (parfois questionnables) ont eu une influence sur la qualité de notre science

Cela permet de souligner des **problèmes**, mais également des **solutions**

■ **Utilisation des meilleures pratiques de recherche**

Pour terminer :

Les pratiques de recherche discutables :

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

-Stopper la collecte des données lorsque $p < .05$

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$
- Enlever des participants pour obtenir $p < .05$

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$
- Enlever des participants pour obtenir $p < .05$
- Ne pas reporter toutes les conditions de l'étude

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$
- Enlever des participants pour obtenir $p < .05$
- Ne pas reporter toutes les conditions de l'étude
- Ne pas reporter toutes les mesures de l'étude

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$
- Enlever des participants pour obtenir $p < .05$
- Ne pas reporter toutes les conditions de l'étude
- Ne pas reporter toutes les mesures de l'étude

HARKing Hypothesizing after results are known

■ Raconter l'histoire qui va mieux avec les résultats

Pour terminer :

Les pratiques de recherche discutables :

Ensemble des pratiques qui visent à rendre nos résultats plus jolis

P-hacking : manipulation et ou sélection des données, de telle sorte à obtenir des résultats favorables

- Stopper la collecte des données lorsque $p < .05$
- Utilisation (ou pas) des modérateurs pour obtenir $p < .05$
- Enlever des participants pour obtenir $p < .05$
- Ne pas reporter toutes les conditions de l'étude
- Ne pas reporter toutes les mesures de l'étude

HARKING Hypothesizing after results are known

■ Raconter l'histoire qui va mieux avec les résultats

Il faut faire une différence entre les hypothèses post-hoc et les hypothèses à priori

Pratiques de recherche discutables:

Problème : Les pratiques de recherche discutables augmentent la probabilité de ne pas repliquer les études

Pratiques de recherche discutables:

Problème : Les pratiques de recherche discutables augmentent la probabilité de ne pas repliquer les études

2 autres problèmes pour la réplication des études :

Pratiques de recherche discutables:

Problème : Les pratiques de recherche discutables augmentent la probabilité de ne pas repliquer les études

2 autres problèmes pour la réplication des études :

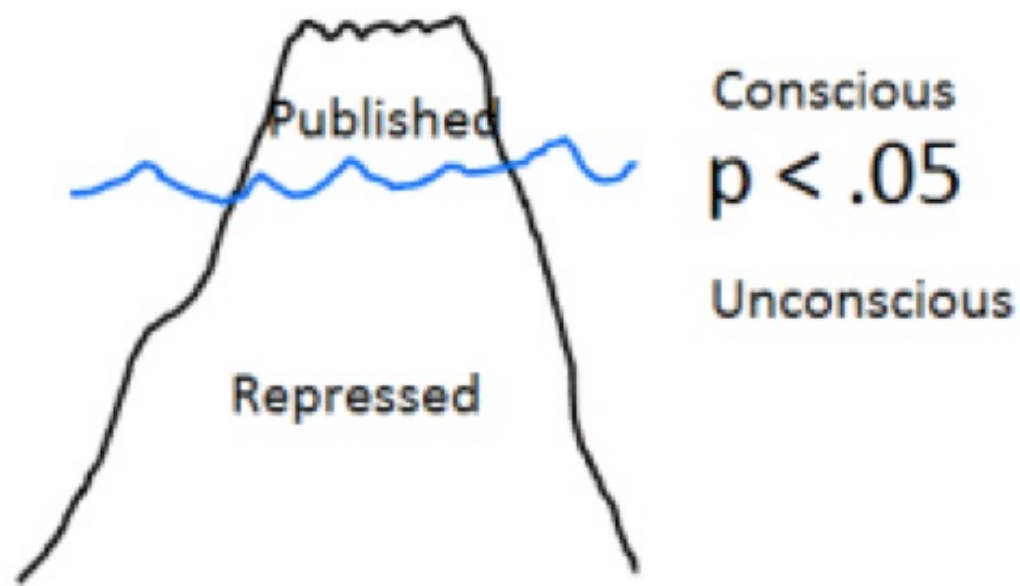
- Les études en psychologie ont des échantillons trop petits

Pratiques de recherche discutables:

Problème : Les pratiques de recherche discutables augmentent la probabilité de ne pas répliquer les études

2 autres problèmes pour la réplification des études :

- Les études en psychologie ont des échantillons trop petits
- Publication Bias



D'après Schimmack (2014)

