



Les coulisses de la recherche scientifique

Une introduction à la
méta-recherche

François Léonard

8 septembre 2023

C'est quoi la méta-recherche ?

La science est le principal moteur du progrès humain, l'amélioration de sa méthode est primordiale à notre développement.

La **méta-recherche** :

- regarde au-delà des études individuelles, regard plus large sur la recherche;
- étudie les pratiques de recherche;
- utilise une approche interdisciplinaire;
- promeut et défend une science de qualité avec des résultats fiables.

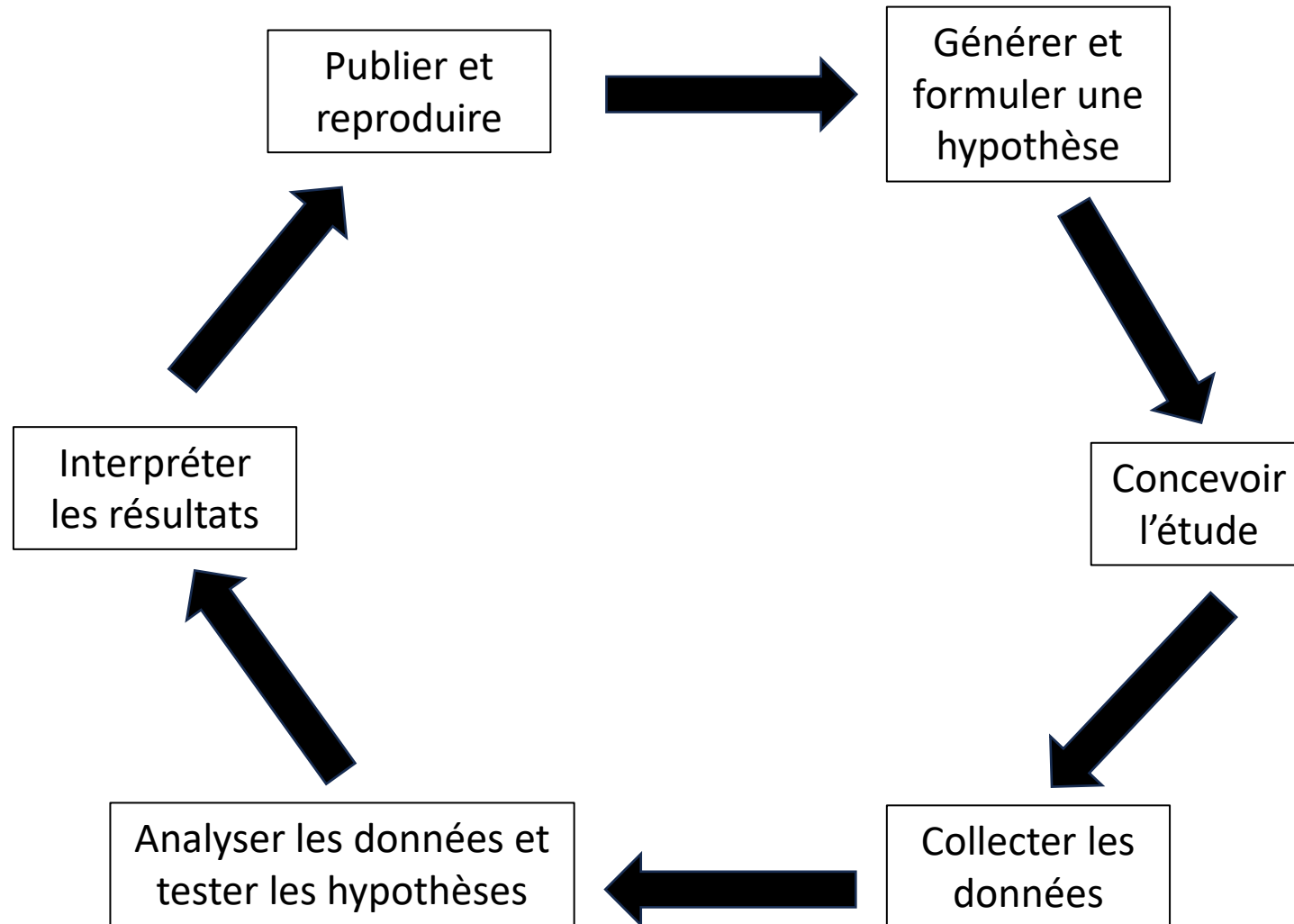
Cinq grandes thématiques

- Pratiques de recherche
- Communication de la recherche
- Reproductibilité des résultats de recherche
- Évaluation de la recherche
- Encouragement des chercheurs

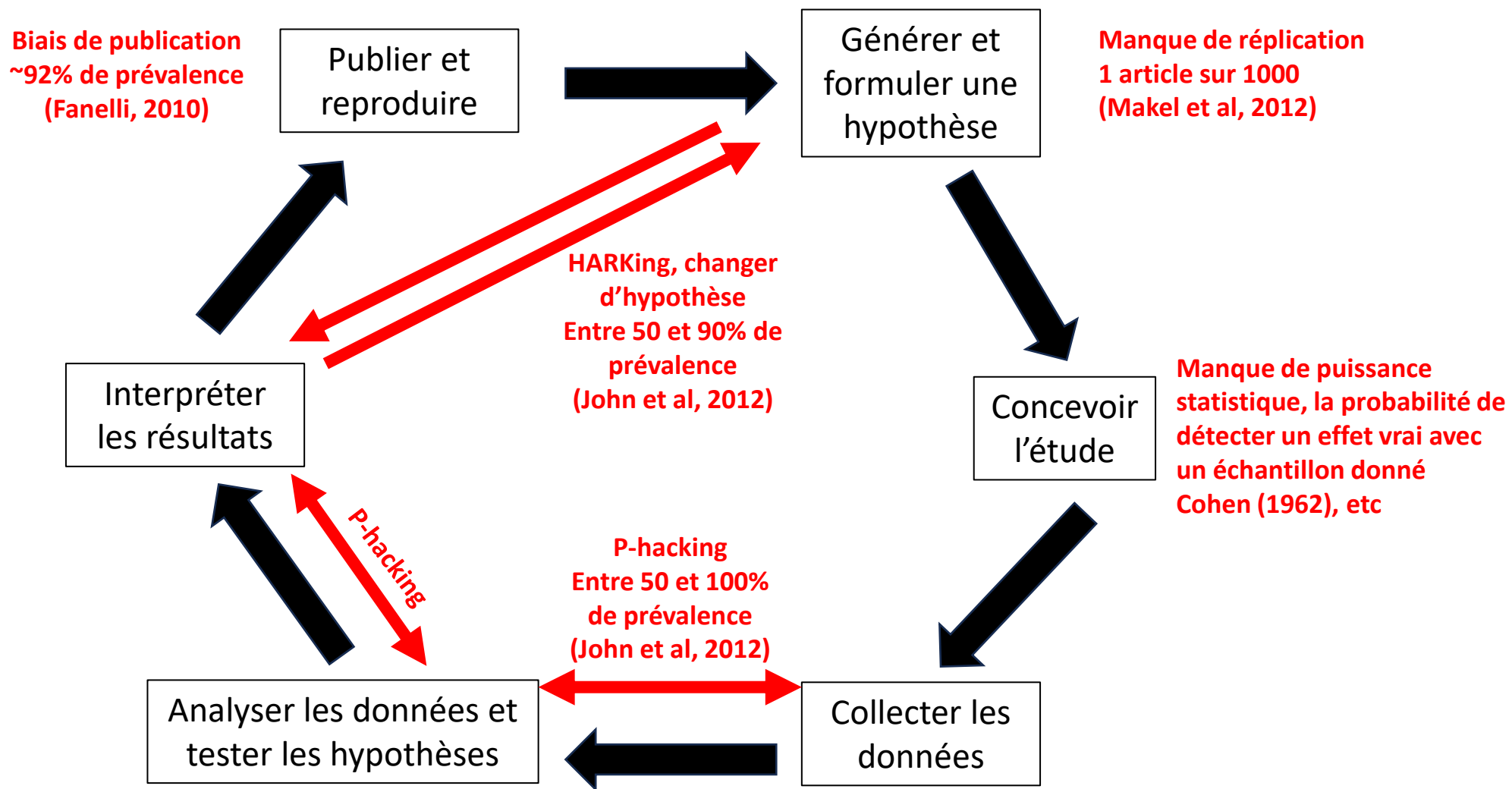
C'est quoi un effet ?

- On veut tester l'efficacité de la potion magique pour augmenter la force
- On constitue deux groupes
 - Un groupe contrôle recevant un placebo
 - Un groupe expérimental recevant la potion magique
- Après l'administration de la potion magique, on mesure la force des sujets
- L'effet de la potion peut être vu comme la différence de force entre les sujets groupe contrôle ceux du groupe expérimental

La méthode hypothético-déductive



La méthode hypothético-déductive





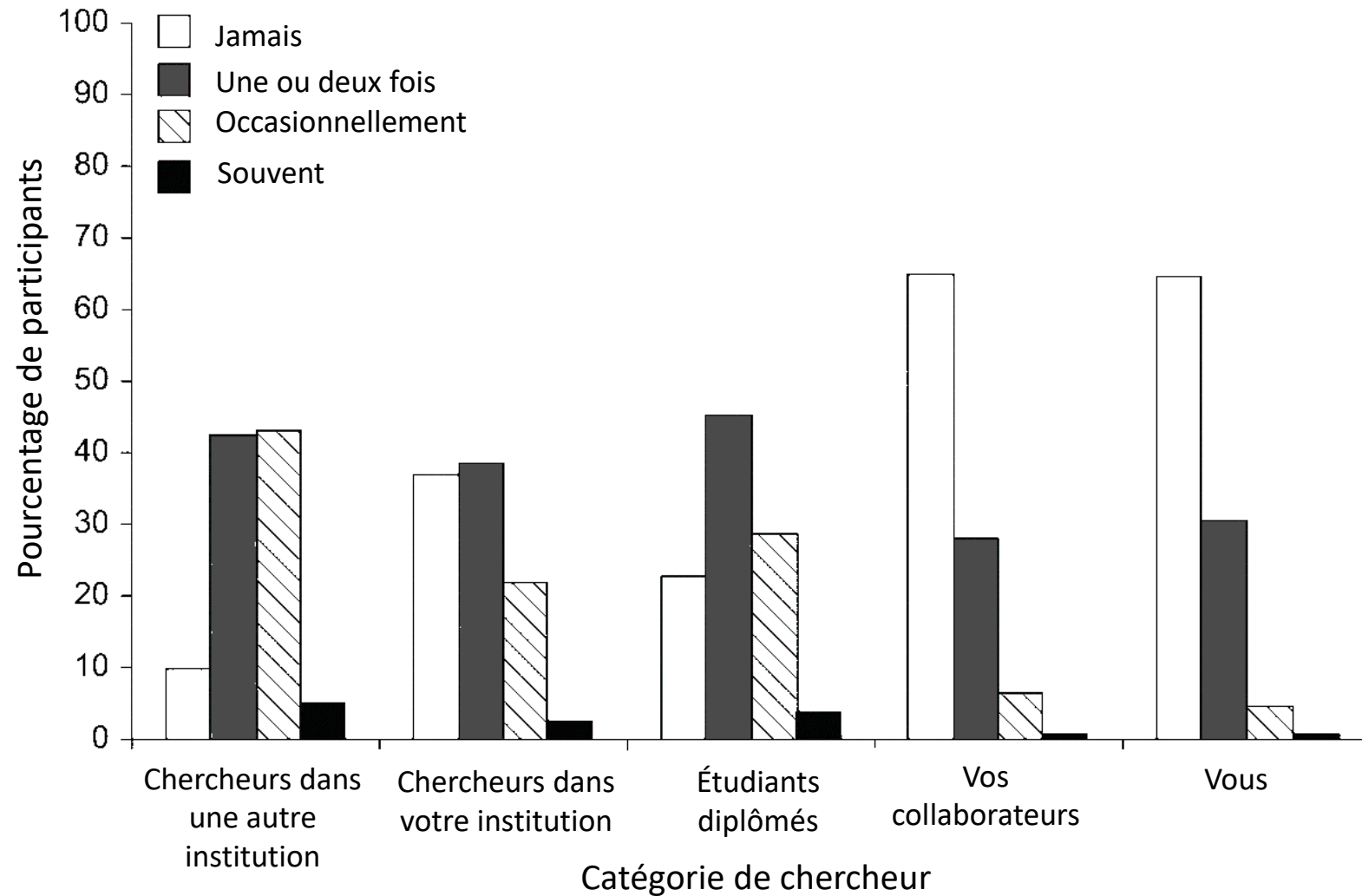
**WHAT IS A FOUNDATION OF
SCIENTIFIC METHOD?**

REPRODUCIBILITY!

Incompétence, pratiques discutables et fraude



« Ce n'est pas moi, ce sont les autres ! »



Prévalence des pratiques discutables et de la fraude

Méta-analyse de Fanelli (2009)

- Synthèse de 21 enquêtes
- 1 à 2% des scientifiques ont falsifié ou fabriqué des résultats une fois dans leur carrière
- 34% admettent recourir à des pratiques discutables
- 74% pensent que leurs collègues recourent à des pratiques de recherches discutables ou fraudes

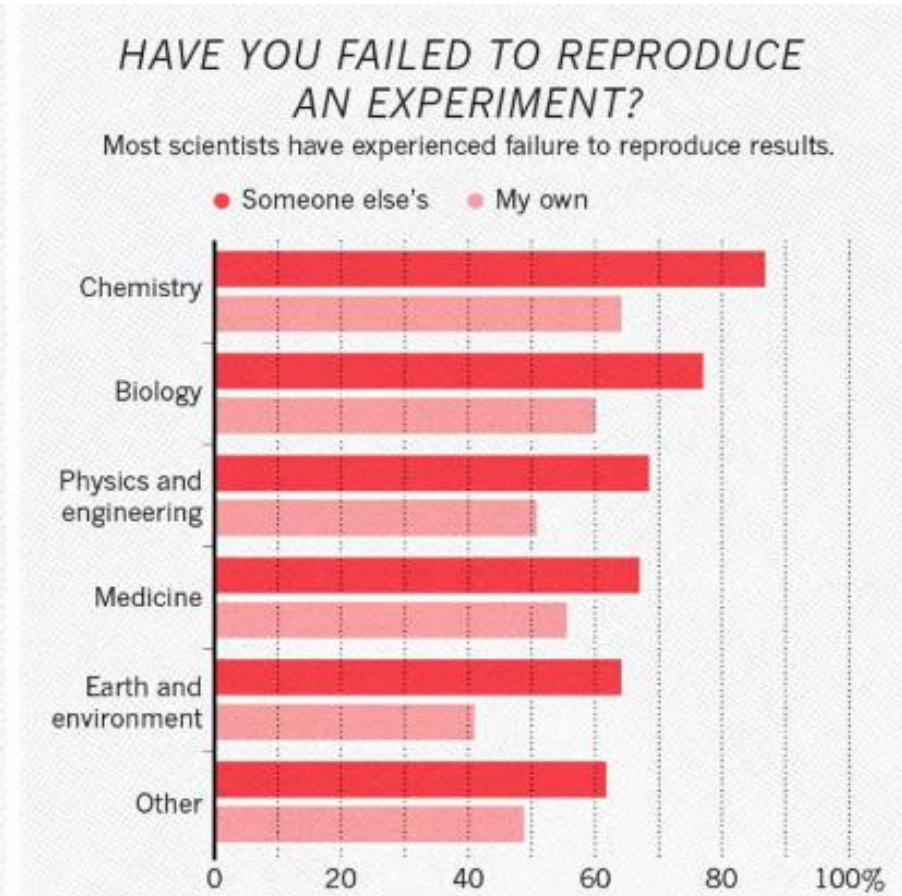
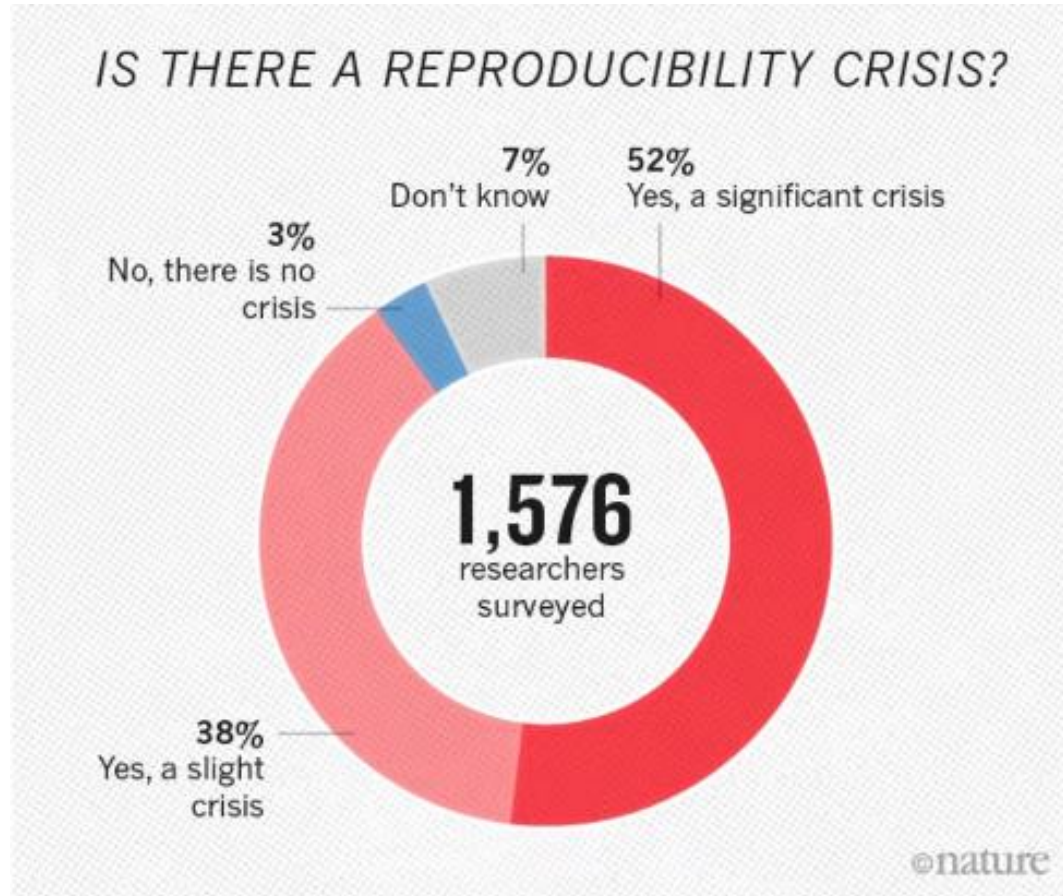
Enquête de Gopalakrishna et al (2022)

- 6 813 chercheurs néerlandais
- Fabrication : 4,3% (IC 95%: 2,9 ; 5,7)
- Falsification : 4,2% (IC 95%: 2,8 ; 5,6)
- 51% des chercheurs admettent s'engager fréquemment dans au moins une pratique discutable

Une crise de la reproductibilité ?

« Y a-t-il une crise de la reproductibilité ? »

« Avez-vous échoué à répliquer une expérience ? »



Une crise de la reproductibilité ?

- 100 études de psychologie répliquées
- 97 comparaisons entre étude originale et répliquée
- Les effets originaux (moyenne = 0,403) sont diminués de moitié dans les études de répliquations (moyenne = 0,197)
- 97% des études originales sont significatives et 36% des répliquations sont significatives

Quatre cavaliers de l'irreproductibilité

HARKing

Échantillon trop petit
(manque de puissance
statistique)

P-Hacking

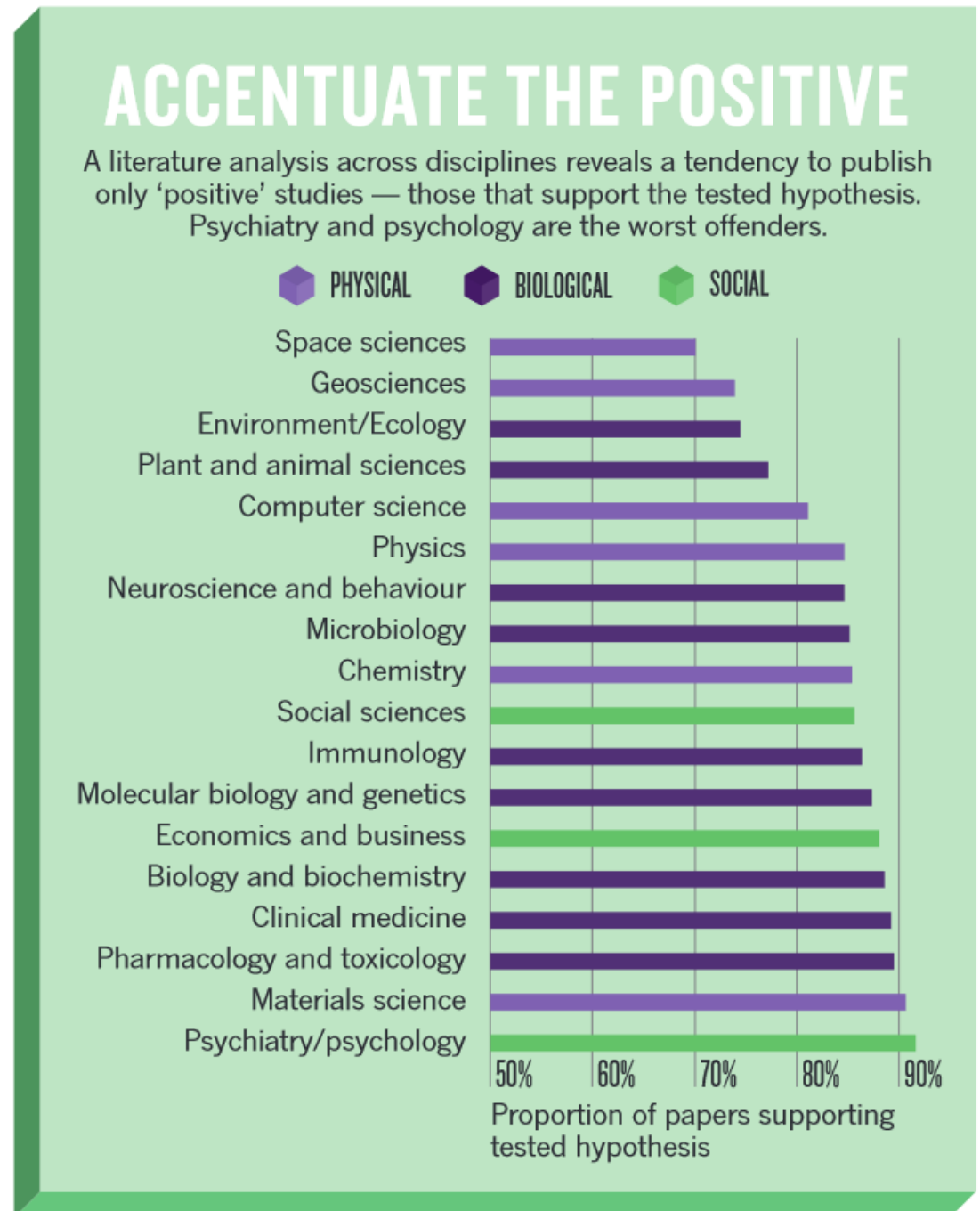
Biais de publication



Peinture de Vasnetsov (1887)

Le biais de publication

- Refus des journaux scientifiques à publier les études 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

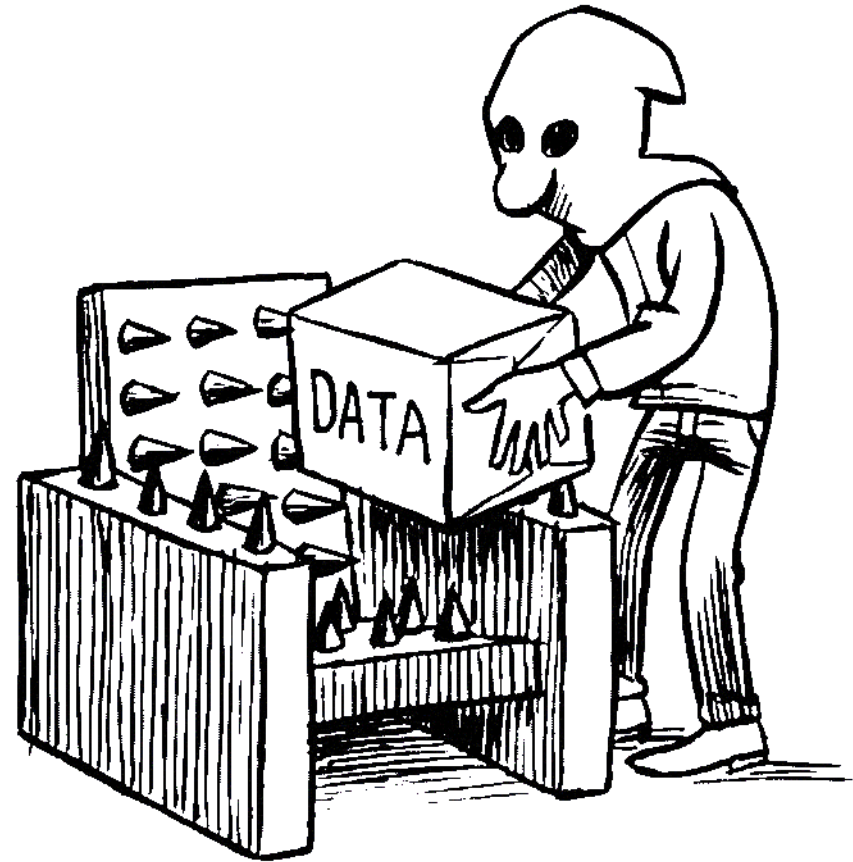


HARKing et P-Hacking

- HARKing (Hypothesis After the Results are Known)

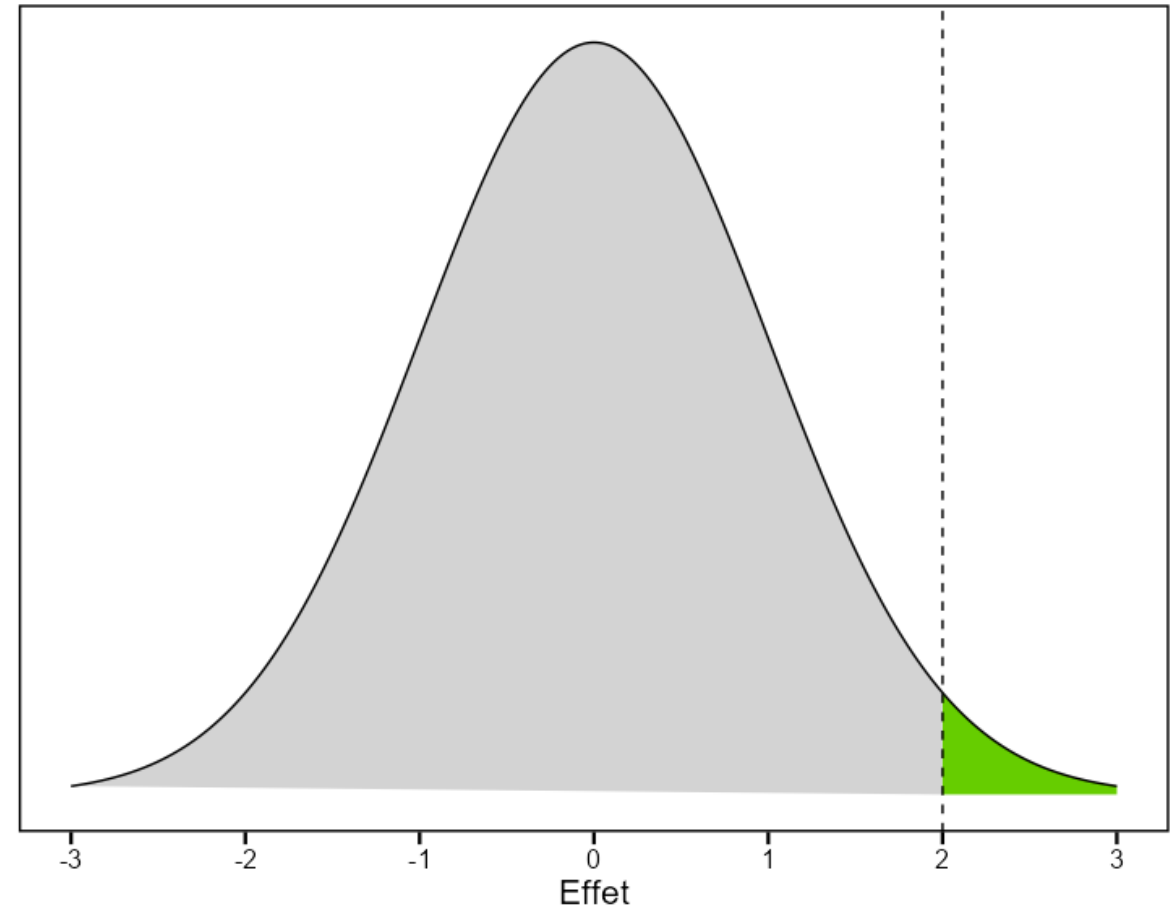


- P-Hacking

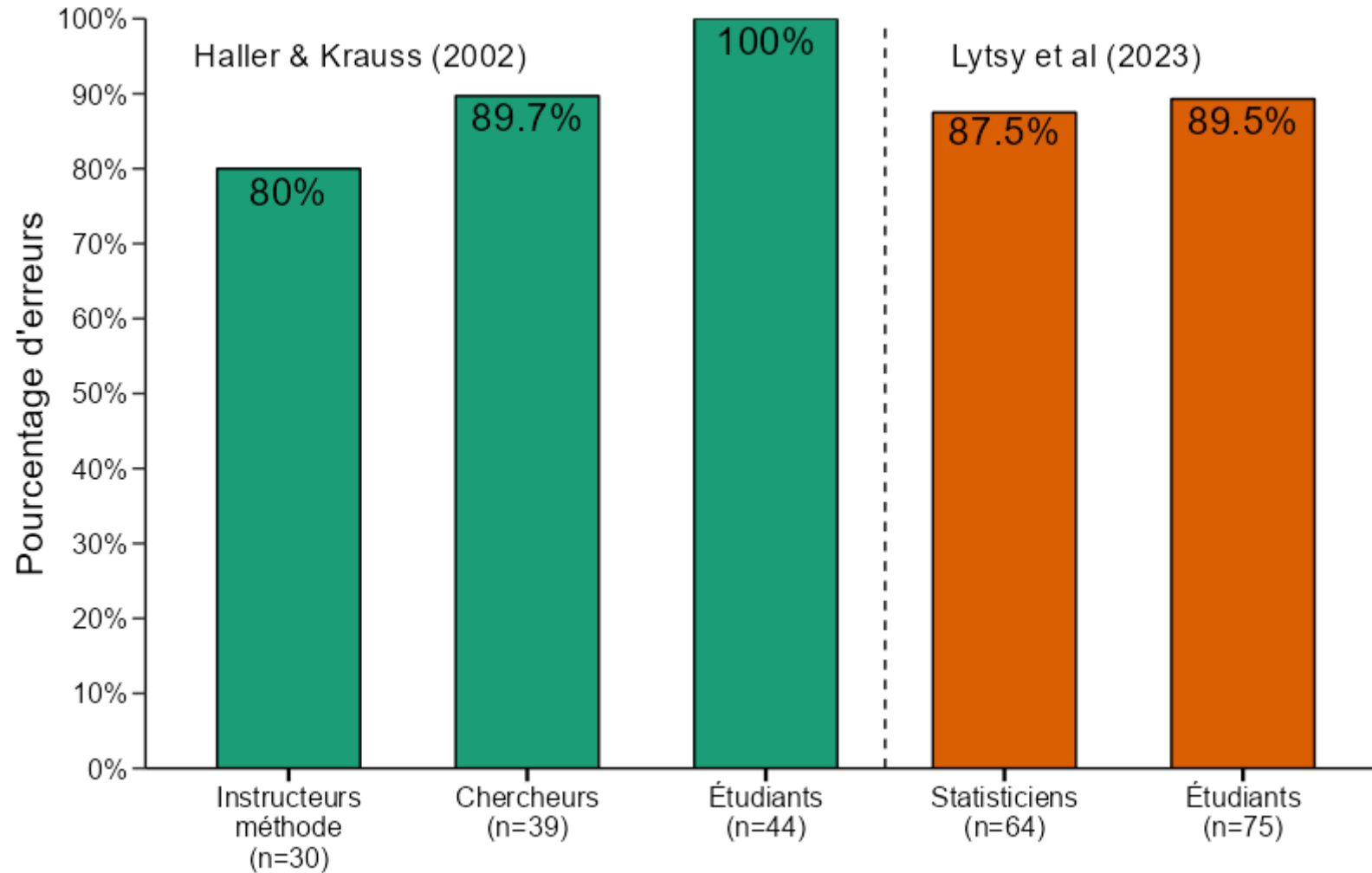


$P < 0,05$: le Graal du chercheur

- La valeur p (« *p-value* ») est la probabilité d'obtenir un résultat (ou un effet) au moins aussi extrême que celui observé, si l'on suppose que l'effet n'existe pas.
 - Représente une sorte degré de surprise
 - statistiquement significative quand p est plus petit que 0,05 (convention)
- Ce n'est pas :
 1. la magnitude de l'effet
 2. La probabilité que notre hypothèse soit fausse
 3. La probabilité d'être face à un faux positif...



Incompréhension des p-valeurs



A photograph of a woman in a blue swimsuit holding a young girl's hands in a swimming pool. The girl is wearing a dark patterned swimsuit and has her hair in pigtails. In the foreground, the face of another child is visible, looking up. The background shows a chain-link fence, palm trees, and a white van. The text 'p-Value' is overlaid on the girl's chest.

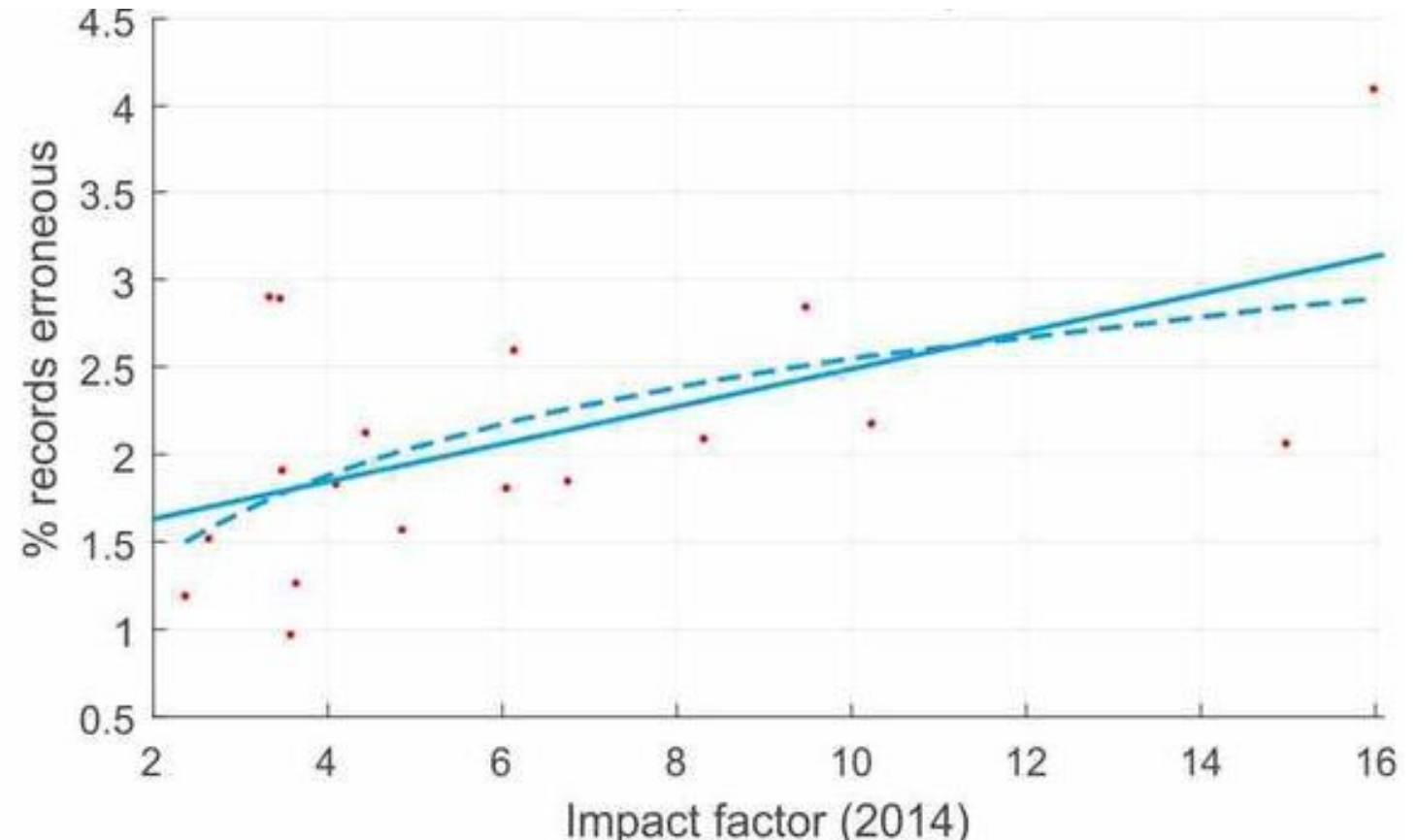
p-Value

**Scientific
Community**

**effect size,
confidence-intervals**

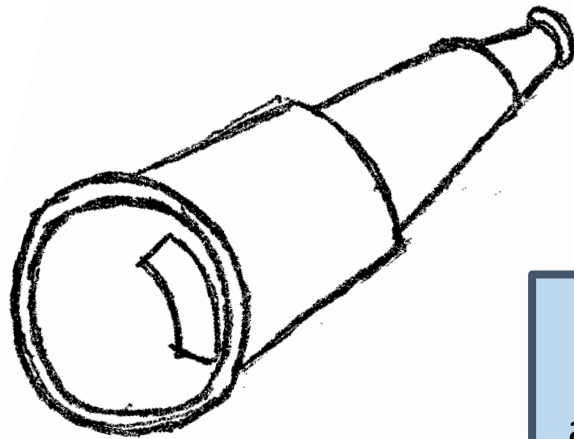
P-valeurs et journaux scientifiques

- 18 journaux fréquemment cités en psychologie et en neurosciences
- 68 762 p-valeurs provenant de 6 273 articles
- 2,3% des p-valeurs significatives sont en réalité non significatives



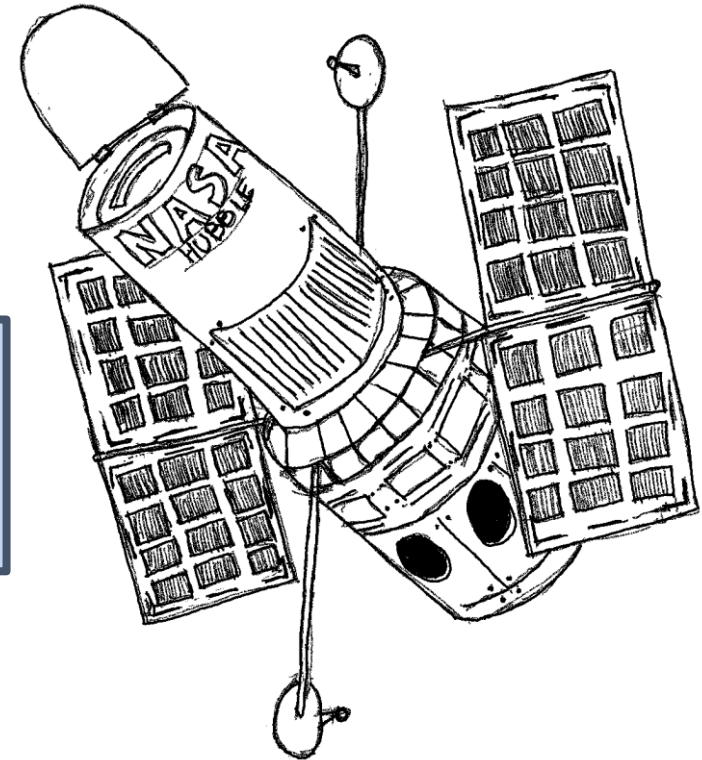
La puissance statistique

Avec une faible puissance (petit effectif), on peut détecter uniquement des grands effets

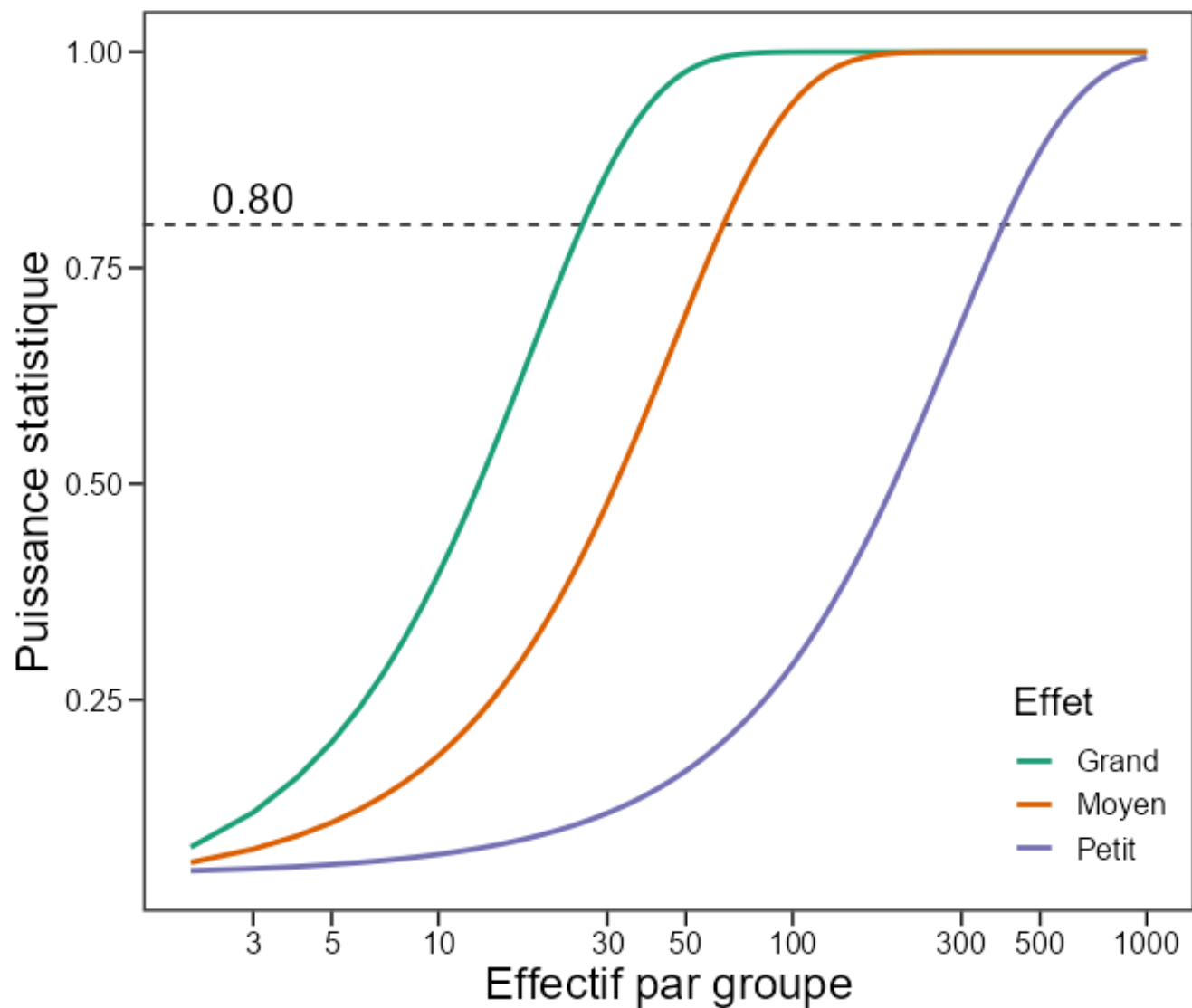


Par convention, on considère qu'un test statistique devrait avoir une puissance supérieure ou égale à 80%.

Avec une puissance importante (grand effectif), on peut détecter de grands comme de petits effets



Détermination des effectifs

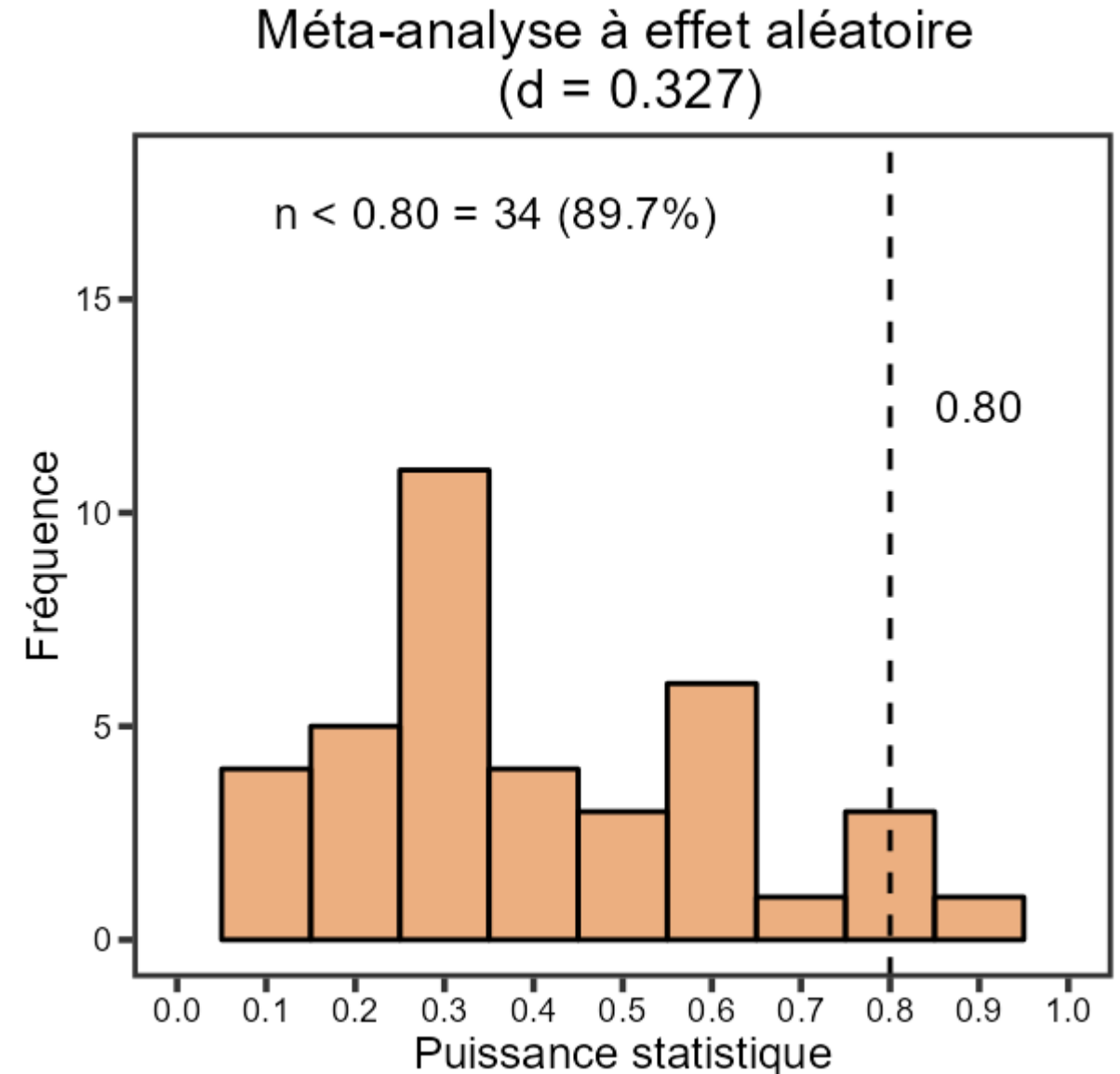


**UNLIMITED STATISTICAL
POWER**

I CAN PROVE ANYTHING

Manque de puissance

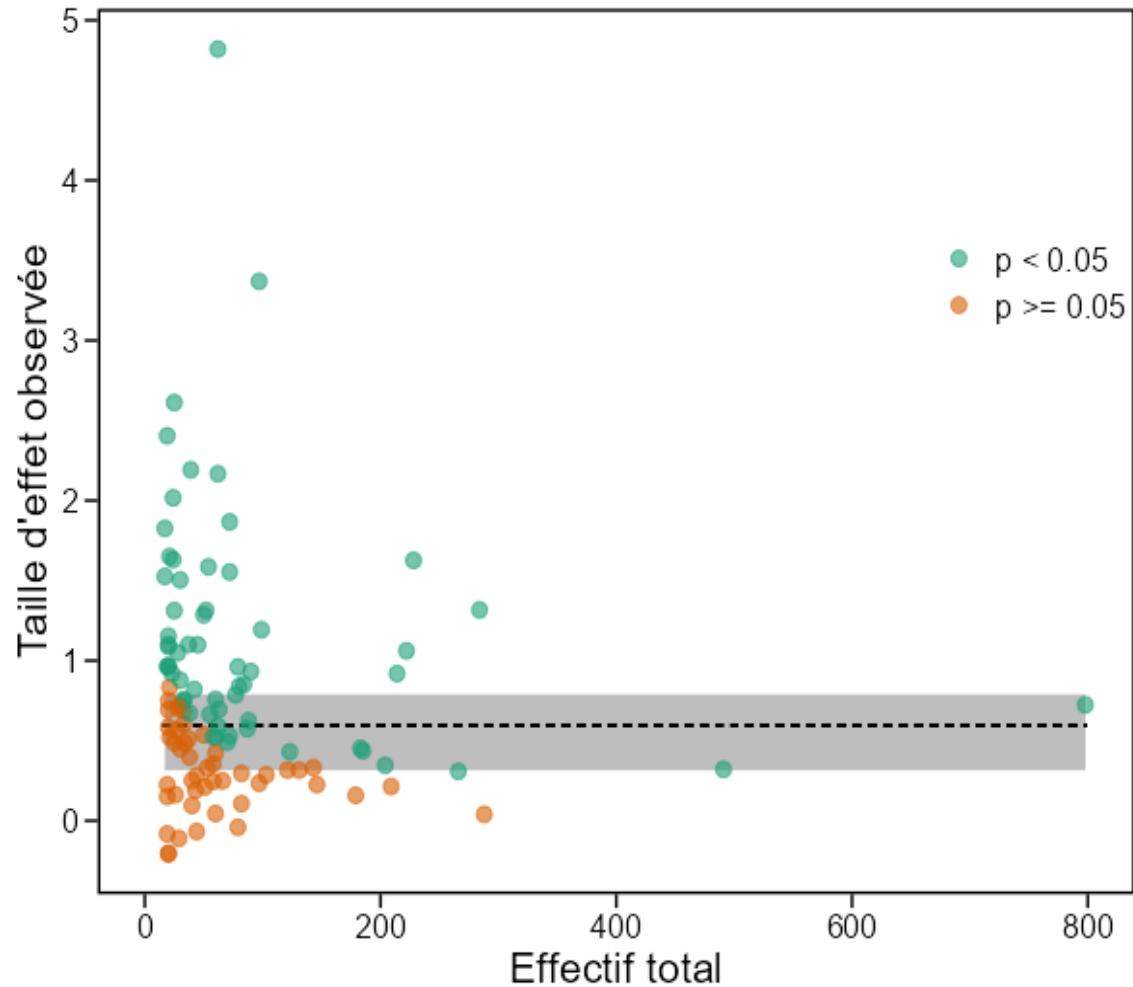
- Première alarme en 1962 par Jacob Cohen.
- Beaucoup d'autres depuis (Bezeau & Graves, 2001; Button et al, 2013; Dumas-Mallet et al, 2017; Fidler et al, 2017; Nord et al, 2017; Szucs & Ioannidis, 2017; Szucs & Ioannidis, 2020; ...)
- Exemple : effet Mozart
 - Analyse de la puissance des 38 effets synthétisés dans la méta-analyse de Pietschnig et al (2010)
 - Effet de 0,327 d de Cohen
 - Puissance médiane de 0.381 [0.311, 0.68]



Conséquences du manque de puissance

- Surestimation des effets détectés
- Augmentation des faux négatifs

Surestimation des effets



- Effets des psychothérapies sur l'intensité de dépressions comparée à des conditions contrôles (n = 107)
- Avec un effectif petit on ne peut détecter que de grands effets de manière « significative »
- Plus l'effectif augmente plus on est capable de détecter de petits effets (ainsi que des grands)
- Phénomènes statistiques proche de la régression à la moyenne

Augmentation des faux négatifs

- Imaginons une maladie nouvelle avec une grande létalité
- Un potentiel traitement existe, avec effet vrai (inconnu)
- Commettre un faux négatif (erreur de type II) nous ferait conclure que le traitement n'est pas efficace
- Que devient le pourcentage de faux négatifs quand on change l'effectif ?

Une étude de petite taille (sous-puissante pour cet effet) nous fera rejeter un bon traitement car elle n'est pas en mesure de détecter ses effets

Effectif par groupe	Faux négatifs (%)
100	5.0
90	8.5
80	10.3
70	16.7
60	23.5
50	31.1
40	40.7
30	50.1
20	63.7
10	82.8

Essay

Why Most Published Research Findings Are False

« Pourquoi la plupart des résultats de recherche publiés sont faux »

John P. A. Ioannidis

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

« Il peut être prouvé que la plupart des résultats de recherche sont faux »

Des solutions ?

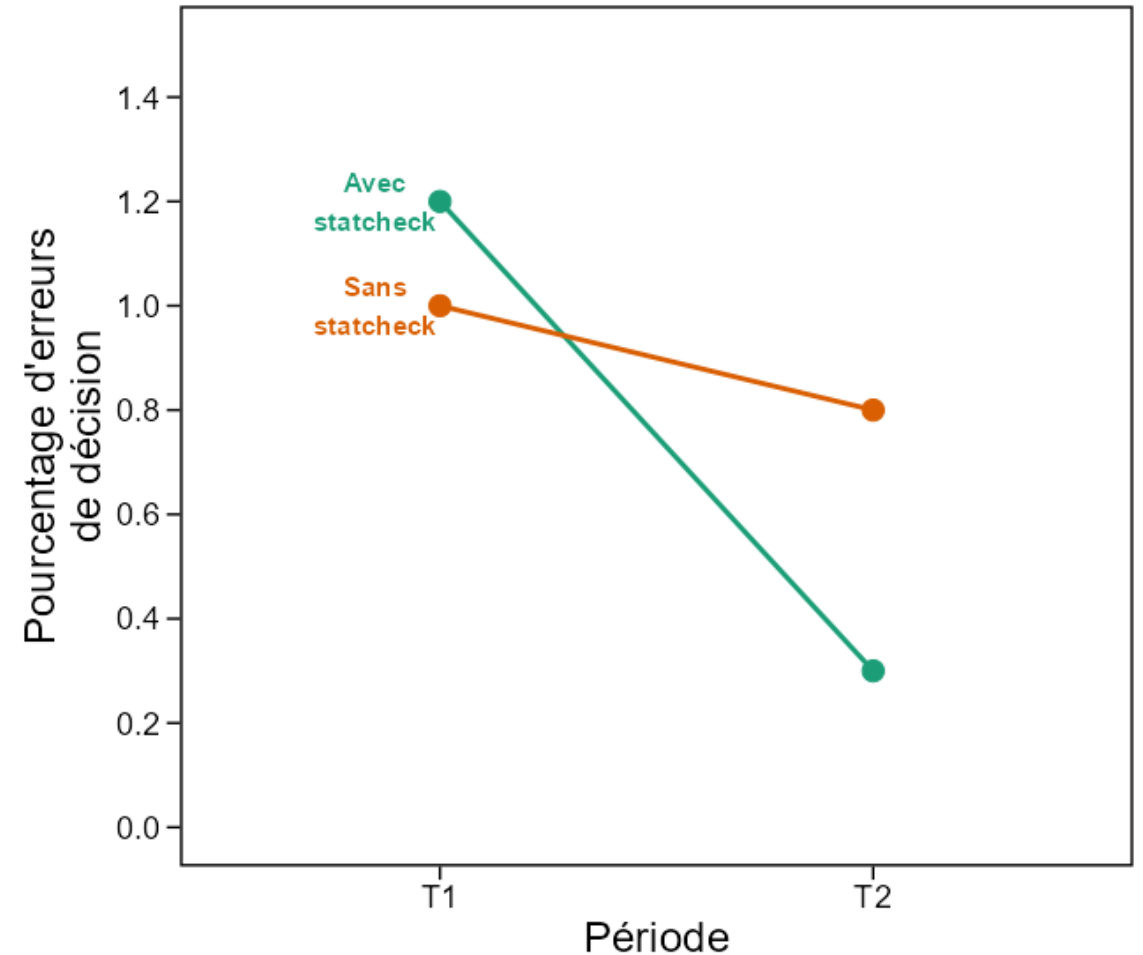
- Pré-enregistrement des études
- Vérification automatique des statistiques
- Open Science
- Évaluation basée sur la qualité
- L'éducation des (futurs) chercheurs
- Protection des « lanceurs d'alertes »
- ...

Pré-enregistrement des études

- Avant de collecter les données !
- Enregistrement sur une plateforme des hypothèses, méthodes de recueil et d'analyse des données
- Accès public ou restreint jusqu'à publication
- Pré-publication parfois possible
- Différence de résultats entre les études pré-enregistrées et les autres
- Effet moyen études pré-enregistrées = 0.21 (sd = 0,19)
- Effet moyen études classiques = 0.40 (sd = 0,21)

Vérification automatique des statistiques

- Environ 50% des articles avec des statistiques présentent des erreurs de p-valeur
- Environ 12.5% ont au moins une erreur rendant une p-valeur significative
- Certains journaux ont implémenté une vérification des p-valeurs avant publication (StatCheck)



Open Science

- Intérêt éthique et moral
- Recherche payée par la société accessible à la société
- Données et matériels
- D'avantage de possibilités de vérifications et de corrections
- Pour la littérature sur l'effet des psychothérapies sur la dépression:
 - L'effectif médian des études publiée en Open Access est de 72 [45; 136]
 - L'effectif médian des études publiée en Open Access est de 43 [28; 90]

Évaluation basée sur la qualité

« *L'université d'Utrecht n'utilisera plus le facteur d'impact dans les décisions d'embauche et de promotion* »

« *Le personnel de l'université d'Utrecht sera évalué à travers son engagement dans des pratiques d'Open Science* »

Woolston (2021) DOI: 10.1038/d41586-021-01759-5

Work / Careers



Utrecht University will no longer use the impact factor in hiring and promotion decisions.

**UNIVERSITY DROPS
IMPACT FACTOR** « *Une université abandonne les facteurs d'impacts* »

Staff at Utrecht University will be assessed through commitment to open science.

By Chris Woolston

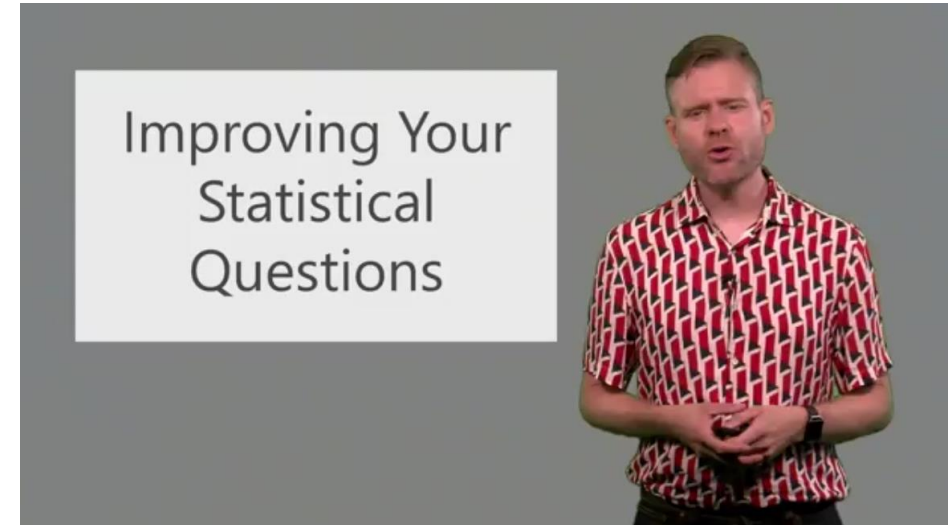
become a very sick model that goes beyond

Éducation des (futurs) chercheurs



« Améliorez vos inférences statistiques »

coursera



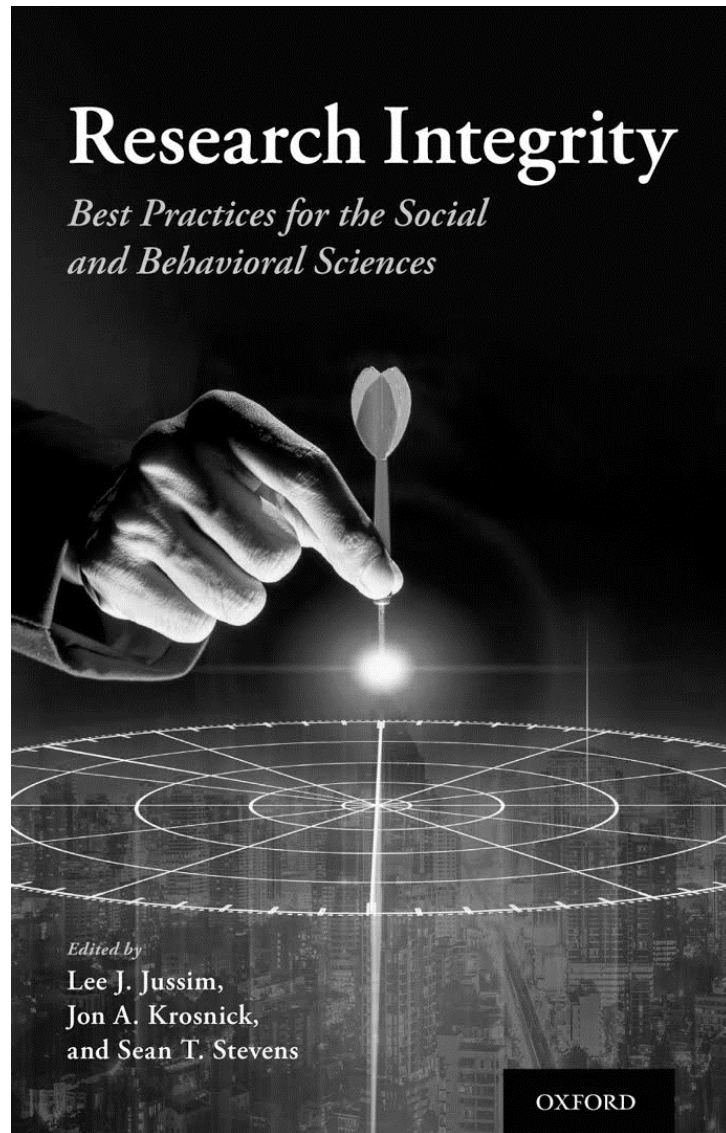
« Améliorez vos questions statistiques »

coursera



« Intégrité scientifique dans les
métiers de la recherche »

FUN Mooc



« Ce livre est dédié à toutes les personnes qui ont aspiré à une carrière dans les domaines académiques, qui se sont engagées dans la recherche de la vérité et qui ont été contraintes d'abandonner cette carrière parce qu'elles ne pouvaient pas rivaliser avec les superstars qui produisaient des recherches très financées, très citées, très influentes mais en réalité non valides. » (dédicace traduite depuis l'anglais)

Remerciements



Pr Ezio Tirelli
Méta-Recherche et Éthique
de la Méthodologie
Quantitative



Pr Christian Monseur
Psychométrie et Édumétrie

Échanges

Questions ?

Impacts sur la société ?

Réserve

Les quatre inférences statistiques

		Réalité	
		L'effet existe L'hypothèse alternative est vraie	L'effet n'existe pas L'hypothèse nulle est vraie
Test statistique	Test statistiquement significatif	Vrai positif Puissance sensibilité	Faux positif Risque α Erreur de type I
	Test statistiquement non significatif	Faux négatif Risque β Erreur de type II	Vrai négatif Confiance spécificité

- La puissance est la probabilité de détecter un effet (significatif) si cet effet existe vraiment
- Elle dépend :
 - Du seuil alpha
 - De la taille d'effet (hypothétique)
 - De l'effectif
 - Du test statistique

$$puissance(1 - \beta) = \frac{VP}{VP + FN} = P(sig|H1)$$