Canadian Psychology/Psychologie canadienne

De grâce, soyez un peu négatif!

--Manuscript Draft--

Manuscript Number:	CAP-2019-0082R1
Full Title:	De grâce, soyez un peu négatif!
Abstract:	Cet article porte sur un phénomène récurrent dans les publications scientifiques : la tendance de plusieurs éditeurs à privilégier les résultats dont les hypothèses sont confirmées et ce, au détriment des résultats dits négatifs. L'article comprend cinq parties. Nous présentons d'abord un court historique du phénomène suivi de l'état des lieux. Par la suite, nous évoquons la cause du peu de publications présentant des résultats négatifs avant d'identifier quatre conséquences. Dans la dernière partie nous présentons trois pistes de solutions. This article focuses on a recurring phenomenon in scientific publications: the tendency of publishers to favor results whose hypotheses are confirmed to the detriment of articles with so-called negative results. The article consists of five parts. We first present a short history of the phenomenon followed by the state of situation. We then discuss the cause of the few publications with negative results before identifying four consequences. Finally, we present three solutions to mitigate this situation.
Article Type:	Masked Article / Article anonyme
Keywords:	
Corresponding Author:	Carl Bouchard, Ph.D. Universite du Quebec en Outaouais Gatineau, Quebec CANADA
Corresponding Author E-Mail:	carl.bouchard@uqo.ca
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universite du Quebec en Outaouais
Other Authors:	Serge Larivée, Ph.D.
Author Comments:	
Corresponding Author's Secondary Institution:	
First Author:	Carl Bouchard, Ph.D.
Order of Authors Secondary Information:	
Manuscript Region of Origin:	CANADA
Opposed Reviewers:	
Order of Authors:	Carl Bouchard, Ph.D.
	Serge Larivée, Ph.D.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

De grâce, soyez un peu négatif![1]

Carl Bouchard
Département de psychoéducation et de psychologie
Université du Québec en Outaouais
courriel : carl.bouchard@uqo.ca

Et

Serge Larivée
École de Psychoéducation
Université de Montréal
courriel : serge.larivee@umontreal.ca

[1] Nous remercions L. Brunet, J. Canac-Marquis, J. Désy, J. Dubé, F. Filiatrault, A. Imbeault, M. Hiel et D. Morin pour leurs critiques d'une première version du manuscrit. Ils ont chacun à leur manière permis d'améliorer sensiblement la forme et le contenu de ce texte.

Masked Manuscript without Author Information / Manuscrit anonyme, c.-à-d. n'ayant aucun renseignement sur l'auteur

De grâce, soyez un peu négatif!

Résumé

Cet article porte sur un phénomène récurrent dans les publications scientifiques : la tendance de plusieurs éditeurs à privilégier les résultats dont les hypothèses sont confirmées et ce, au détriment des résultats dits négatifs. L'article comprend cinq parties. Nous présentons d'abord un court historique du phénomène suivi de l'état des lieux. Par la suite, nous évoquons la cause du peu de publications présentant des résultats négatifs avant d'identifier quatre conséquences. Dans la dernière partie nous présentons trois pistes de solutions.

Mots clés : publications scientifiques, résultats négatifs, «publish or perish», biais de publication

Abstract

This article focuses on a recurring phenomenon in scientific publications: the tendency of publishers to favor results whose hypotheses are confirmed to the detriment of articles with so-called negative results. The article consists of five parts. We first present a short history of the phenomenon followed by the state of situation. We then discuss the cause of the few publications with negative results before identifying four consequences. Finally, we present three solutions to mitigate this situation.

Keywords: scientific publications, negative results, publish or perish, publication bias

Intérêt public

Les revues scientifiques ont tendance à privilégier les études dont les hypothèses sont confirmées au détriment des études dont les données ne supportent pas l'hypothèse initiale. Pourtant, tout résultat d'une recherche menée selon les règles de l'art est en soi digne d'être publié.

L'une des solutions à la publication d'articles dont les résultats sont négatifs se trouve dans le préenregistrement des études sans en connaître les résultats. Cette solution découle du principe selon lequel la qualité des articles doit être évaluée en fonction de l'importance de la question de recherche et de la rigueur de la méthodologie et non pas en fonction du fait que l'hypothèse de recherche soit acceptée ou rejetée.

Public interest

Scientific journals tend to favor studies whose hypotheses are confirmed to the detriment of studies whose data do not support the initial hypothesis. However, any result of a research carried out according to the rules of the art is in itself worthy of being published. One of the solutions to publishing articles with negative results is to preregister studies without knowing the results.

De grâce, soyez un peu négatif!

On s'inquiète de plus en plus de ce que les décisions menant à la publication des articles soumis aux revues scientifiques dépendent largement du caractère positif de leurs résultats. En fait, les revues ont tendance à privilégier les études dont les hypothèses sont confirmées au détriment des études dont les données ne supportent pas l'hypothèse initiale. Pourtant, tout résultat d'une recherche menée selon les règles de l'art est en soi digne d'être publié. Cela est d'autant plus vrai pour les travaux en sciences humaines et sociales dont les hypothèses, pour diverses raisons, ne sont pas souvent confirmées (Larivée, 2017).

L'objectif de cet article est de faire le point sur un aspect important du fonctionnement de la science : la publication des résultats négatifs. Nous dresserons d'abord un court historique de la situation, suivi de l'état actuel des lieux. Par la suite, nous évoquerons brièvement la cause qui conduit à ignorer les résultats négatifs avant d'identifier quatre de ses conséquences. En terminant, nous suggérons trois pistes de solutions.

Un bref historique

L'expression « résultats négatifs » est utilisée dans la littérature scientifique depuis la moitié du 20e siècle et réfère généralement aux résultats statistiquement non-significatifs qui ne supportent pas les hypothèses de recherche et entraînent l'acceptation de l'hypothèse nulle (H0) Dès la fin des années 1950, Smith (1956), alors rédacteur en chef du *Journal of Abnormal Social Psychology,* constatait que les études avec des résultats négatifs étaient moins susceptibles d'être publiées. En 1959, Sterling faisait un constat similaire : très peu de résultats négatifs étaient rapportés dans quatre revues scientifiques, suggérant ainsi un biais de publication. Plus précisément, Sterling trouva que 97,3% des articles publiés dans ces revues rapportaient des résultats statistiquement significatifs et que seuls 8 des 294 articles (soit 2.7%) ne parvenaient pas à confirmer les hypothèses. Cette même année, Goldfried et Walters (1959) suggéraient

que la valeur scientifique des résultats négatifs était sous-estimée. À leurs yeux, les résultats négatifs pouvaient tout aussi bien enrichir nos connaissances ne serait-ce qu'en indiquant les maladresses méthodologiques et les aspects infructueux d'une étude, permettant ainsi d'éviter perte de temps et d'efforts en reproduisant une recherche inefficace. Les fondements empiriques de la science exigent en effet de savoir autant ce qui fonctionne que ce qui ne fonctionne pas. Le détachement émotif face à nos théories doit également se traduire pour les résultats obtenus. Goldfried et Walters (1959) déploraient également le fait que seules les études susceptibles de figurer dans les revues scientifiques étaient celles donnant des résultats publiables, c'est-à-dire des résultats statistiquement significatifs. L'étude transversale de Smart (1964) soutient les affirmations de Goldfried et Walters. Les résultats de cette étude montrent qu'une importante majorité des articles publiés en psychologie et en éducation comportent principalement des résultats statistiquement significatifs. En fait, un examen minutieux des revues scientifiques en sciences humaines et sociales montre que les études avec des résultats négatifs ne représente qu'environ 9% du volume total des articles publiés.

D'autres études réalisées au cours des décennies suivantes allaient permettre d'observer le même biais de publication en faveur des études obtenant des résultats positifs. Wilson, Smoke et Martin (1973) réalisèrent une analyse de contenu de trois revues scientifiques publiées entre juillet 1969 et juin 1970 : l'*American Journal of Sociology*, l'*American Sociological Review* et le *Social Forces*. Les résultats montrent que 61 articles sur 76 (80,3%) obtenaient des résultats statistiquement positifs. De son côté, Lindsay (1994) a analysé tous les articles empiriques publiés dans *Accounting, Organizations and Society, The Accounting Review,* et *Journal of Accounting Research* de 1970 à 1987. Son analyse a permis d'observer que 84,2% des études présentaient des résultats positifs. Pour leur part, Hubbard et Armstrong (1992) ont analysé 32 numéros choisis au hasard du *Journal of Marketing,* du *Journal of Marketing Research* et du *Journal of Consumer Research* publiés entre 1974 et 1989 pour constater que des 692 articles examinés, 638 (92,2%) avaient obtenus des résultats positifs.

Ce biais favorisant les études avec des résultats positifs a par la suite été confirmé à plusieurs reprises dans des revues spécialisées en sciences sociales (Glass, McGaw, & Smith, 1981; Smith, 1980a, 1980b) et en médecine (Ernst & Pittler, 1997; Moscati, Jehle, Ellis, Fiorello, & Landi, 1994; Sterling, Rosenbaum, & Weinkam 1995; Vickers, Goyal, Harland, & Rees 1998). Egger et Davey Smith (1995) montrèrent même que les études avec des échantillons réduits qui présentaient des résultats positifs étaient davantage publiées que les recherches avec un plus large échantillon mais qui ne démontraient aucun résultat significatif. Depuis, de nombreuses études ont documenté l'excès de résultats positifs dans plusieurs autres domaines, dont la biologie, l'écologie et l'économie (Doucouliagos, Laroche, & Stanley, 2005; Kyzas, Denaxa-Kyza, & Ioannidis, 2007; Song et al., 2010; Sterling. Rosenbaum, & Weinkam, 1995).

État des lieux

Bien que certaines études rapportant des résultats négatifs puissent légitimement être rejetées en raison d'une piètre qualité méthodologique (Dickersin, 1990; Smith, 1956), les études qui rapportent des résultats négatifs sont tout de même moins susceptibles d'être publiées en dépit de méthodologies de qualité comparable à celles qui obtiennent des résultats positifs (Koren, Shear, Graham, & Einarson, 1989). De plus, lorsque des résultats positifs de recherches reproduites par d'autres chercheurs se changent en résultats négatifs, plusieurs éditeurs de revues scientifiques hésitent à les publier (Charlton, 1987; Smart, 1964).

Plusieurs recherches réalisées dans différents domaines scientifiques au cours des 30 dernières années illustrent l'ampleur du phénomène. Notamment, celle de Fanelli (2010) qui a recensé 1 316 articles de diverses disciplines publiés aux États-Unis entre les années 2000 et 2007. Selon l'auteur, la culture du *publish or perish* omniprésente dans les universités force en quelque sorte les chercheurs à produire des résultats publiables à tout prix, c'est-à-dire des résultats positifs qui confirment l'hypothèse de recherche. Selon Fanelli, cette pression à publier devrait se traduire par une augmentation d'articles dont les résultats sont positifs, particulièrement dans les milieux universitaires plus compétitifs. Pour vérifier son hypothèse, il a analysé un échantillon aléatoire d'articles publiés entre 2000 et 2007 dont l'auteur était attaché à une

université américaine. Fanelli a pris en considération trois variables : la méthode utilisée, le nombre d'hypothèses testées et la discipline. Les effets potentiels de ces facteurs ont été contrôlés avec des modèles de régression. Les résultats montrent que le pourcentage moyen d'articles présentant des résultats positifs dans les universités de tous les États américains était de 82,38%. En tenant compte des trois variables contrôlées dans les analyses de régression, les résultats de cette étude mettent en évidence une corrélation positive entre le nombre d'articles publiés dans un état particulier et le pourcentage d'articles présentant des résultats positifs.

Dans une seconde recherche, Fanelli (2012) élargit son échantillonnage aux études publiées dans les années 1990. Cette fois, il analyse 4 656 articles publiés dans 70 pays provenant de la base de données *Essentials Sciences Indicators.* Son objectif était double : observer d'une part la présence éventuelle d'une évolution au fil des ans et, d'autre part, d'une éventuelle différence entre les pays. En 1990 et 1991, la proportion d'articles avec des résultats positifs ou des hypothèses partiellement vérifiées était de 70,2% pour se situer à 88,6% en 2005 et à 85,9% en 2007. Les résultats montrent qu'on assiste à des augmentations similaires dans les 70 pays.

Dans le domaine biomédical, le phénomène est encore plus patent. Les compagnies pharmaceutiques semblent plus favorables à la publication d'essais cliniques témoignant de l'efficacité du médicament évalué. De nombreuses preuves empiriques montrent en effet que les conclusions tirées d'essais cliniques randomisés sont plus positives à l'égard d'interventions expérimentales financées par des organisations à but lucratif (Als-Nielsen, Chen, Gluud, & KIjaergard, 2003; Rochon et al., 1994). Ainsi, Als-Nielsen et al. (2003) ont mené une étude afin de déterminer si le lien entre le financement et les résultats obtenus dans les essais randomisés sur les médicaments reflétait bien les effets du traitement, tant sur le plan de leurs bienfaits que sur celui de leurs effets indésirables. Pour ce faire, ils ont sélectionné un échantillon aléatoire de 167 revues de la base des données Cochrane, dont 25 contenaient des méta-analyses éligibles selon leurs critères de sélection. L'association entre le type de financement (lucratif vs non lucratif)

et les résultats obtenus dans les études a été analysée à l'aide de régressions logistiques permettant de tenir compte de l'effet du traitement, dont les effets secondaires, tout en contrôlant les effets potentiels d'autres facteurs tels que la qualité méthodologique des essais cliniques, la taille de l'échantillon et l'année de publication. Leurs résultats montrent que le médicament expérimental était recommandé comme traitement de choix dans 16% des essais financés par des organisations à but non lucratif, dans 30% des essais qui ne rapportaient aucun financement, dans 35% des essais qui étaient financés à la fois par des organisations à but non lucratif et à but lucratif et dans 51% des essais financés par des organisations à but lucratif. De plus, des analyses plus poussées ont montré que les essais financés par des organisations à but lucratif étaient nettement plus susceptibles de recommander le médicament expérimental comme traitement de choix. En conclusion, les auteurs suggèrent que les essais financés par des organisations à but lucratif peuvent être considérés plus positifs en raison d'une interprétation biaisée des résultats des essais cliniques, les lecteurs devant soigneusement évaluer si les conclusions tirées d'essais randomisés sont bel et bien étayées par les données obtenues.

Dans une autre étude visant à étudier la relation entre l'efficacité déclarée des médicaments dans les essais cliniques publiés et la prise en charge des essais par le fabricant du médicament en cours d'évaluation, Rochon et al. (1994) ont étudié un échantillon d'essais cliniques sur les anti-inflammatoires non stéroïdiens (AINS) utilisés dans le traitement de l'arthrite en examinant 56 essais cliniques associés à un fabricant. Les résultats de leur enquête montrent que l'efficacité du médicament associé au fabricant a été rapporté comparable (71,4%) ou supérieur (28,6%) à un médicament de comparaison dans les essais cliniques. Parmi les essais ayant identifié un médicament moins toxique (n = 22), l'innocuité du médicament associé au fabricant a été signalée supérieure à un médicament générique dans 86,4% des cas. Ainsi, les AINS associés au fabricant auraient presque toujours une efficacité égale ou supérieure à celles du médicament de comparaison. Toutefois, ces prétentions à la supériorité ne sont souvent pas étayées par les résultats des essais cliniques. Selon les auteurs, les conclusions de leur étude soulèvent des préoccupations quant à la

publication sélective ou à une interprétation biaisée des résultats dans les essais cliniques associés aux fabricants qui subventionnent ces mêmes essais.

Concernant les publications d'essais cliniques sur des antidépresseurs recensés par la Food and Drug Administration (FDA) aux États-Unis, pour bien saisir les résultats d'études telles que celle de Turner, Matthews, Linardatos, Tell et Rosenthal (2008), il importe de bien comprendre les différentes phases des essais cliniques dans les sciences biomédicales (voir Encadré 1).

Encadré 1 – Phases des essais cliniques dans les sciences biomédicales (Umscheid, Margolis, & Grossman, 2011).

Lors de la ***première phase***, les chercheurs passent plusieurs mois à examiner les effets du médicament sur environ 20 à 80 personnes ne présentant aucun problème de santé. Cette phase vise à déterminer la dose la plus élevée qu'un individu puisse prendre sans ressentir d'effets secondaires graves. La ***seconde phase*** implique plusieurs centaines de participants atteints de la maladie que le nouveau médicament est censé traiter. Ils reçoivent généralement la même dose que celle qui s'est avérée sûre au cours de la phase précédente. Les enquêteurs surveillent alors les participants pendant plusieurs mois, voire plusieurs années, pour vérifier l'efficacité du médicament et recueillir davantage d'informations sur d'éventuels effets secondaires. Bien que cette phase implique plus de participants que les phases précédentes, elle n'est toujours pas suffisante pour démontrer la sécurité globale d'un médicament. La ***troisième phase*** implique généralement jusqu'à 3 000 participants atteints de la maladie que le nouveau médicament est censé traiter. Cette phase peut durer plusieurs années. Son objectif est d'évaluer le fonctionnement du nouveau médicament par rapport aux médicaments existants pour le même problème. Les essais de cette phase sont généralement à double insu, ce qui signifie que ni le chercheur ni le participant ne savent quel médicament prend ce dernier. S'il est démontré que le médicament est au moins aussi sûr et efficace que les autres déjà sur le marché, la FDA approuve habituellement le médicament. La ***quatrième phase*** a lieu une fois que la FDA a approuvé le traitement. Elle implique des milliers de participants et peut durer de nombreuses années. Les chercheurs utilisent alors cette phase pour obtenir plus d'informations sur l'innocuité, l'efficacité et les autres avantages à long terme du médicament.

Dans le cadre de l'étude de Turner et al. (2008), les phases deux et trois des programmes d'essais cliniques pour 12 antidépresseurs approuvés par la FDA entre 1987 et 2004 ont été recensées par les

chercheurs. La méthodologie de tous les essais sélectionnés comprend une répartition aléatoire des participants à double insu, avec un groupe contrôle recevant un placebo. Seules les données d'efficacité du traitement de la dépression à court terme ont été analysées et seuls les essais cliniques présentant les premiers résultats d'efficacité ont été sélectionnés. Sur 74 essais cliniques enregistrés à la FDA, 23 (31%) d'entre eux n'ont pas donné lieu à une publication. Les auteurs indiquent qu'il n'y a pas de différence significative quant à la taille de l'échantillon entre les articles publiés et non publiés. Les résultats montrent aussi que sur les 74 études retenues, les 38 (51%) ayant des effets positifs sont publiées, à l'exception d'une seule. De plus, sur les 36 études ayant des résultats négatifs, seules trois ont été publiées, 22 n'ont pas été publiées et 11 l'ont été en annonçant des résultats positifs. Aussi, 48 articles publiés sur 51 témoignent de l'efficacité du traitement tandis que, selon la FDA, seuls 38 articles présentaient des résultats considérés positifs. Les auteurs concluent à la présence d'un biais en faveur de la publication d'articles ayant des résultats positifs et s'inquiètent d'une forme de manipulation relative à l'interprétation des résultats concernant la véritable efficacité des traitements étudiés ainsi que des véritables raisons permettant d'expliquer la non-publication de certains articles inclus dans leur échantillon. Ceux-ci sont d'autant plus inquiets que les articles choisis pour leur étude se limitent aux compagnies pharmaceutiques inscrites auprès de la FDA. Un tel résultat montre que les professionnels de la santé qui prescrivent ces médicaments peuvent être maintenus dans l'ignorance.

En pharmacie, procéder à des essais cliniques randomisés est indispensable pour valider la supériorité d'un médicament sur un autre ou sur un placebo. Étant donné la longue et coûteuse procédure obligatoire avant de conclure, les compagnies pharmaceutiques qui financent ces recherches ont tout intérêt à déboucher sur des résultats positifs. Or, comme ce n'est pas toujours le cas, certaines compagnies cherchent alors à présenter leurs résultats sans donner l'impression d'un échec. Hartley et al. (2018) ont présenté les stratégies utilisées par les leaders d'opinion (*key opinion leaders*) du domaine de la cardiologie quand ils expriment leurs avis à propos des résultats obtenus lors d'essais cliniques pour tester des

médicaments. L'échantillon qui comportait 127 articles ont fait l'objet de 438 avis par les leaders d'opinion dont 108 (45%) faisaient appel à des prétextes afin d'excuser des résultats négatifs, prétextes que les auteurs ont classé en 17 catégories dont les deux plus utilisés sont des classiques largement connus par la communauté scientifique : l'échantillon était trop petit pour 31 essais cliniques (31%) et d'autres études sont nécessaires (21%). On aura compris que l'attachement émotif aux résultats positifs contribue à mettre le tort sur la méthode – ce qui peut bien sûr être le cas – plutôt que de se plier au constat que le médicament testé ne fonctionne pas – ce qui peut bien sûr être le cas.

La cause : *publish or perish*

Nous ne considérons ici intentionnellement une seule cause parce que tout semble tourner autour du *publish or perish*. La liste des publications d'un chercheur est probablement le principal critère de promotion dans le cadre universitaire, en plus d'être un indice de productivité (De Rond & Miller, 2005), même si cela n'a pas toujours été le cas (Klatersky, 1987). Comme nous l'avons vu plus tôt, il ne suffit pas de produire des articles, encore faut-il privilégier les articles que les revues scientifiques préfèrent, ceux dont les résultats confirment les hypothèses de recherche, principalement lorsque celles-ci sont séduisantes (Thornton & Lee, 2000). Compte tenu du taux élevé de rejet des articles soumis et de la compétition féroce pour l'obtention de fonds de recherche (Joober, Schmitz, Annable, & Boksa, 2012; Young, Ioannidis, & Al-Ubaydli, 2008), les chercheurs sont conscients que soumettre des résultats de recherche novateurs augmente la probabilité d'être publiés. Les organismes de financement ont également tendance à financer à nouveau les chercheurs dont les travaux débouchent sur des résultats positifs (Djulbegovic et al., 2000).

Conséquences

La valorisation, voire la nécessité de publier exclusivement des résultats positifs, entraîne au moins quatre conséquences relatives à la nature des conclusions publiées, à l'impact sur la reproductibilité, aux véritables effets des programmes d'intervention et à l'état des connaissances scientifiques.

La nature des résultats publiés

Devant le peu de valeur accordée par un grand nombre de revues scientifiques aux études de reproductibilité, quel chercheur voudra tenter de reproduire des résultats négatifs? Deux comportements, qui sont en quelque sorte les deux côtés de la même médaille, ont alors émergé (Chevassus-au-Louis, 2016). Le premier, appelé HARKing (Hypothesing After the Results are Known), consiste à proposer des hypothèses *a posteriori* de façon à ce que celles-ci correspondent aux résultats de la recherche. Le second se manifeste par le tripotage des résultats pour que ceux-ci correspondent aux hypothèses. Cette manipulation des données peut prendre différentes formes : arrondir ici et là les résultats, modifier les procédés afin que les résultats montrent ce que l'expérimentateur veut bien qu'ils montrent, utiliser abusivement ou de manière non appropriée des tests statistiques afin d'arriver à confirmer une hypothèse, etc. L'ensemble de ces comportements découle évidemment, comme on a vu plus haut, de la tendance des revues scientifiques à privilégier les résultats positifs.

Impact sur la reproductibilité

Le peu de résultats négatifs dans la littérature scientifique est étroitement lié au faible taux de reproductibilité des résultats de recherche, un problème qui touche particulièrement les sciences humaines et sociales (Makel, Plucker, & Hegarty, 2012). Selon Popper (1973), le noyau dur du consensus déterminant le caractère scientifique d'une recherche réside en effet dans la nature reproductible des résultats qui constitue en quelque sorte le test décisif pour établir la validité d'un fait. Autrement dit, contrairement aux autres formes de connaissance, les affirmations scientifiques exigent d'être vérifiées par d'autres chercheurs. Cette reproductibilité constitue en fait un mécanisme de régulation de la science. Cependant, au moins deux éléments font en sorte qu'en pratique les scientifiques reproduisent rarement les expériences de leurs collègues. D'une part, ils font d'emblée confiance aux résultats publiés, postulant du coup qu'ils sont reproductibles et, par conséquent, fiables. D'autre part, vu le nombre élevé de publications, seules les recherches de pointe sont reproduites, d'autant plus que les éditeurs privilégient souvent la « nouveauté ».

En 2011, une étude de reproduction réalisée par trois chercheurs de la firme pharmaceutique Bayer a rappelé la communauté scientifique à son devoir (Prinz, Schlange, & Asadullah, 2011). Ceux-ci ont tenté de reproduire 67 projets de recherche en biologie fondamentale, en amont donc du processus menant à l'invention d'un nouveau médicament. Les spécialistes du domaine pharmaceutique appellent cette phase de la recherche fondamentale « l'identification de la cible », c'est-à-dire la description d'un mécanisme cellulaire qui semble suffisamment crucial dans la pathologie pour que l'on envisage de rechercher des molécules susceptibles de le modifier. Les résultats sont clairs : les deux tiers des résultats des études étaient impossibles à reproduire.

Le problème de la reproductibilité peut s'expliquer par les manipulations statistiques parfois effectuées par les chercheurs, comme le montre l'étude de Wang, Yan et Katz (2018). Leur objectif est de déterminer la fréquence et la « sévérité » des demandes formulées aux biostatisticiens par les chercheurs afin d'examiner les pratiques statistiques inappropriées lors de consultations statistiques. À partir d'un échantillon de 522 biostatisticiens membres de l'*American Statistical Association*, 390 (74,7%) ont répondu au questionnaire en indiquant combien de fois au cours des cinq dernières années ils avaient été sollicités pour répondre à dix-huit demandes inacceptables en matière d'analyses de données. Les résultats montrent que les quatre demandes indues les plus fréquemment rencontrées, qualifiées de «plus sévères» par au moins 20% des répondants, ont été par ordre de fréquence : (1) supprimer ou modifier certaines données qui permettraient de confirmer l'hypothèse de recherche, (2) interpréter les résultats sur la base des attentes et non des résultats obtenus, (3) ne pas mentionner les problèmes liés aux données manquantes susceptibles de biaiser les résultats et (4) ignorer les violations aux prérequis statistiques qui changeraient les résultats de positifs en négatifs. Bien que les auteurs n'aient pas demandé aux biostatisticiens s'ils avaient effectivement obéi aux demandes formulées par les chercheurs qui les emploient, les pratiques statistiques inappropriées décrites dans l'étude sont utilisées afin de publier à tout prix et contribuent au problème de la reproductibilité – ou de la non-reproductibilité devrait-on dire – de la recherche scientifique. Les

conclusions de cette étude nous amènent à une réflexion : ce ne sont évidemment pas tous les chercheurs qui ont la possibilité de faire appel à un biostatisticien pour mener leurs analyses. S'ils doivent faire le travail eux-mêmes, comme c'est souvent le cas, cela suppose de leur part une constante mise à jour de leurs connaissances en statistique, ce qui demande beaucoup de temps et d'énergie. Il ne serait donc pas surprenant que ceux-ci aillent à l'essentiel lorsque vient le temps de mener des analyses statistiques et qu'ils s'auto-censurent lorsqu'ils obtiennent des résultats négatifs, parce qu'il faut bien publier.

La situation dans le domaine des sciences humaines et sociales semble être encore plus complexe compte tenu de la forte prévalence de résultats positifs, et également, compte tenu du fait que certaines données se prêtent peu à la reproduction (Ferguson, 2015). Selon Larivée (2017), quatre raisons permettent d'expliquer la difficulté d'évaluer l'impact du biais de reproduction en sciences humaines et sociales. Premièrement, il y a plus de chercheurs en sciences biomédicales qu'en psychologie, ce qui entraîne plus de financement, et du coup, plus de vérifications, particulièrement dans les études cliniques. Deuxièmement, il est plus facile de détecter les irrégularités méthodologiques dans le domaine biomédical qu'en psychologie. Par exemple, pour les études biomédicales, les dossiers médicaux des patients sont conservés et les essais cliniques doivent être enregistrés auprès de certains organismes de vérification comme la Food and Drug Administration (FDA) aux États-Unis. Troisièmement, la recherche biomédicale reçoit plus de prix prestigieux que la recherche en psychologie. Enfin, il y a une enquête uniquement en cas de soupçons de fraude, ce qui, compte tenu de la nature des données colligées en sciences humaines et sociales, se prête moins bien à la vérification (Ferguson, 2015).

Cette difficulté s'observe dans le travail d'une équipe internationale de plus de 250 chercheurs, qui rassemblait en novembre 2011 un échantillon de 100 études de psychologie sociale et cognitive, sélectionnées parmi trois prestigieuses revues du domaine, soit *Psychological Science, Journal of Personality and Social Psychology* et *Journal of Experimental Psychology : Learning, Memory, and Cognition*, puis tenta de reproduire les résultats de ces études, toutes publiées en 2008 (Open Science

Collaboration, 2015). On comprendra dès lors que la conclusion de cette recherche publiée dans la revue *Science* ait eu l'effet d'une bombe dans le monde de la psychologie : après un travail de quatre ans, elle révélait l'incapacité de cette équipe à reproduire les résultats de près des deux tiers d'entre elles, soulevant du coup un sérieux doute sur leur rigueur scientifique.

Par ailleurs, au plan pratique, la communauté scientifique ne saura probablement jamais que des résultats n'ont jamais pu être reproduits (Pashler & Harris, 2012). En tenant à l'écart les publications concernant ces tentatives infructueuses, d'autres chercheurs pourront tenter de montrer ces mêmes effets qui n'existent malheureusement pas. Des ressources humaines et financières provenant de fonds publics seront alors inutilement dépensées. La reproductibilité de résultats non publiés peut également avoir des implications plus importantes pour les sciences du comportement, particulièrement au plan des thérapies et des interventions privilégiées par les intervenants psychosociaux. À cet égard, Makel et al., (2012) avancent que la publication des résultats reproductibles dans le domaine de la psychologie ne représenterait qu'à peine 1% de toutes les publications.

Les véritables effets des programmes d'intervention

D'un point de vue éthique, les résultats dits négatifs, particulièrement en sciences humaines et sociales, méritent d'être connus par les intervenants psychosociaux. Il importe en effet que ceux-ci basent leurs interventions sur des programmes dont l'efficacité est avérée. Toutefois, l'évaluation d'un programme peut produire des résultats négatifs et donner une mauvaise impression de son efficacité et, conséquemment, en réduire les possibilités de financement (U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth, and Families, 2010). On aura compris que les propositions de nouvelles interventions doivent être évaluées de manière adéquate, ce qui est impossible sans une attention particulière au processus d'implantation. Ainsi, des résultats négatifs peuvent être obtenus, particulièrement si l'implantation du programme n'est pas correctement effectuée. Cette évaluation devient essentielle afin d'établir la validité interne et externe des interventions. Comme le précisent Durlak

et Dupre (2008), lors de la vérification des résultats relatifs à l'efficacité d'un programme, l'implantation parfaite ou presque parfaite est pratiquement impossible. Les auteurs ont réalisé une revue systématique des études qui évaluent l'impact de la fidélité d'implantation sur les résultats de différents programmes et qui permet d'identifier les facteurs qui affectent le processus. L'examen de plus de 500 études quantitatives suggère que la qualité d'implantation est la variable qui affecte le plus les résultats obtenus dans l'évaluation de l'efficacité des programmes de promotion et de prévention. Plus spécifiquement, les résultats montrent qu'un grand nombre de programmes recensés atteignent des niveaux d'implantation proches de 60% et que très peu arrivent à des niveaux supérieurs à 80%. Aucune étude n'a permis d'atteindre un niveau d'implantation de 100%. Pour ainsi dire, l'obtention de résultats négatifs dans les études d'implantation de programme serait plus la norme que l'exception.

Derzon, Sale, Springer et Brounstein (2005) constatent que parmi les facteurs de nature méthodologique ayant les effets les plus importants sur les résultats, trois sont liés à l'implantation. Deux d'entre eux concernent le degré auquel les objectifs et les procédures du programme sont intégrés à la pratique quotidienne ainsi que l'intensité de la prestation du programme. Le troisième facteur touche la nature des activités auxquelles sont exposées les participants constituant les groupes de contrôle dans les études d'évaluation de programme. Derzon et al. (2005) ont réestimé les résultats des évaluations de programmes recensés en ajustant les données originales à l'influence de ces trois facteurs. Les objectifs de l'étude étaient de déterminer quels seraient les résultats ajustés si les participants des groupes contrôle ne bénéficiaient d'aucun service et si les programmes étaient implantés de manière cohérente (intégrés à la pratique quotidienne) et avec une intensité de prestation suffisante. Les résultats montrent que les tailles d'effets moyens (d de Cohen) pour les 46 programmes étudiés sont passées de 0,02 à 0,24, ces effets étant statistiquement significatifs. Les auteurs concluent en précisant que si les problèmes liés à l'implantation des programmes et la nature des activités auxquelles sont soumis les groupes de contrôle avait été davantage planifiée, les programmes étudiés auraient été 12 fois plus efficaces.

La mise en évidence de résultats négatifs ne devrait pas être considérée comme une preuve d'échec du programme, mais plutôt comme une occasion d'améliorer sa mise en œuvre. Les résultats d'une évaluation de programme méthodologiquement bien conçue doivent tout simplement être informatifs, qu'ils soient positifs ou négatifs (Glasgow & Steiner, 2018). Dans un processus d'évaluation de programme, les efforts visant à modifier les connaissances, les attitudes et les comportements des parties prenantes ne vont pas toujours dans le sens des résultats attendus (The W.K. Kellogg Foundation, 2017). Il devient alors important de présenter les résultats qui ne sont pas concluants mais qui peuvent se révéler prometteurs et qui justifient des études supplémentaires visant à établir plus clairement l'efficacité du programme à l'étude.

La W.K. Kellogg Foundation (2017) offre quelques conseils afin de bien gérer la communication des résultats négatifs auprès des mandataires ayant commandé l'évaluation d'un programme:

1. Insister dès le départ sur le fait que le processus d'évaluation est un acte d'apprentissage.
2. Impliquer les parties prenantes dans la conception de l'implantation du programme et communiquer tout au long du processus d'évaluation pour ajuster les attentes relatives aux résultats.
3. Partagez les résultats négatifs lors des discussions afin de faciliter le processus d'apprentissage et de réduire les risques de malentendus quant aux résultats obtenus.
4. Modifier le langage utilisé afin de transformer les résultats négatifs en aspects positifs : « il y a des réalisations et des défis », « comment nous pouvons faire mieux à partir des résultats obtenus » et « l'évaluation de programme est un travail en cours ».

L'état des connaissances scientifiques

La publication des études dont les résultats sont négatifs est importante pour la crédibilité des recensions systématiques et des méta-analyses, mais surtout, pour éviter les biais de sélection menant ainsi à une autre forme de biais, celui de la publication (Dickersin, 1990). Les méta-analyses et les recensions

systématiques sont utilisées pour éclairer d'importantes questions empiriques et pratiques, dont l'efficacité des programmes d'intervention. Toutefois, afin que la communauté scientifique et les praticiens aient confiance en leurs résultats, celles-ci doivent montrer qu'elles sont exemptes de biais (Banks, Kepes, & Banks, 2012).

Afin de bien comprendre les biais pouvant affecter les résultats des méta-analyses, regardons d'abord quelles sont les étapes de leur réalisation (voir Encadré 2). Il existe plusieurs explications quant à des biais éventuels lors de la réalisation des méta-analyses. Par exemple, certains auteurs choisissent volontairement d'exclure des études, arguant que les résultats d'une méta-analyse ne sont significatifs que si le même protocole de recherche a été utilisé dans chacune des études sélectionnées, et seulement si celles-ci ont eu recours à la répartition aléatoire des participants dans chaque condition. Selon ce point de vue, les méta-analyses devraient être effectuées uniquement à partir d'essais cliniques randomisés. Cependant, et particulièrement dans les sciences sociales et humaines, de nombreuses études sont de nature observationnelle et leur exclusion pourrait sérieusement biaiser les résultats de la méta-analyse (Thornton & Lee, 2000).

Encadré 2 – Étapes de réalisation des méta-analyses (Egger, Davey-Smith & Altman 2001).

La première étape consiste à définir les critères d'éligibilité des articles à sélectionner afin de s'assurer d'évaluer des résultats comparables, et surtout, fiables. Cette étape est basée sur les critères suivants: (1) la qualité des études cliniques; (2) le type d'études cliniques; (3) les participants; (4) les résultats obtenus; et (5) la durée des suivis. La bonne évaluation du premier critère, la qualité des études cliniques, dépendra fortement du jugement du chercheur. *La seconde étape* consiste à décrire la meilleure stratégie de sélection des études à inclure dans l'analyse. C'est à cette étape que le chercheur devrait envisager l'inclusion d'études non publiées en raison des effets de biais souvent associés aux études publiées comportant des résultats positifs. *La troisième étape* consiste à extraire les données contenues dans les articles sélectionnés lors de la première étape. Pour ce faire, on utilise des échelles conçues pour évaluer la qualité des études cliniques (Moher et al. 1995). Les individus responsables de la sélection des articles pour les méta-analyses devraient ignorer: (1) les auteurs et leurs institutions, (2) les noms des revues, (3) les sources de financement, et (4) les remerciements. Lors de *la quatrième étape*, les résultats bruts des études sélectionnées lors de

l'étape précédente doivent être standardisées de manière à les rendre comparables. *La cinquième étape* vise à rendre les études sélectionnées comparables entre elles sur différents critères méthodologiques tels que la taille des échantillons et les procédures statistiques. Les techniques statistiques pour se faire peuvent être classées en deux grands modèles : le modèle des « effets fixes » et celui des « effets aléatoires ».

Certaines études peuvent également être exclues des méta-analyses pour d'autres raisons. Par exemple, la langue de publication de près de 80% des revues scientifiques est l'anglais (Moher et al., 1996) et les méta-analyses n'incluent de ce fait que des articles publiés en anglais (Egger, Davey Smith, Schneider et Minder, 1997; Grégoire, Derderian, & Le Lorier, 1995). Même si la plupart des revues scientifiques de haut niveau sont publiées en anglais, une grande partie de la littérature scientifique est publiée dans une autre langue. Par exemple, plus de 2 500 revues biomédicales sont publiées en chinois, dont moins de 6% sont indexées dans Medline (Xia & Adams, 2008). Les publications dans des langues autres que l'anglais sont souvent exclues des revues systématiques en raison de leur inaccessibilité ou d'un manque de ressources pour leur traduction. Vérifier si la qualité et les résultats de la recherche diffèrent systématiquement entre les études rapportées en anglais par rapport à d'autres langues est toutefois à établir (Morrison et al., 2012).

Les raisons expliquant cette contrainte relèvent, entre autres éléments, de la difficulté à identifier les documents pertinents publiés dans d'autres langues et la présumée supériorité des articles publiés en anglais. Ainsi, la reproduction de certaines méta-analyses originalement constituées uniquement d'articles en anglais a donné lieu à des résultats différents lorsque des articles pertinents publiés dans d'autres langues ont été inclus. Par exemple, Grégoire et al. (1995) ont réalisé une étude dont l'objectif était de refaire des méta-analyses en y incluant des articles publiés dans une langue autre que l'anglais. Leurs conclusions montrent que trois méta-analyses ont obtenu des résultats statistiquement significatifs comparables même en y ajoutant des articles supplémentaires d'une autre langue alors qu'une quatrième dont les résultats initiaux n'avaient pas atteint le niveau de signification statistique le devenait après l'inclusion d'un article supplémentaire de langue allemande.

Ces biais de sélection des études ont été systématiquement examinés, en particulier dans des essais cliniques (Liebeskind, Kidwell, Sayre, & Saver, 2006). Le recours aux recensions systématiques et aux méta-analyses a permis d'exposer le problème encore plus clairement et de mieux comprendre les possibles conséquences de tels biais (Sutton, Duval, Tweedie, Abrams, & Jones, 2000). En utilisant un nombre suffisant d'études originales, la méta-analyse peut même quantifier le biais attribuable aux données non publiées. Entre autres méthodes, l'utilisation des courbes d'Egger et l'analyse de type « trimand-fill » (Duval & Tweedie, 2000) permet d'estimer les résultats probables des études non publiées en examinant la quantité de données manquantes et également d'estimer les tailles d'effet résultant de l'inclusion des données manquantes. L'inclusion des données manquantes modélisées dans la méta-analyse entraîne parfois une perte complète de l'effet publié d'une intervention ou même de l'existence d'un phénomène.

Les biais de sélection peuvent avoir d'importantes conséquences. Ainsi, comme les études rapportant des résultats positifs sont plus fréquemment publiées que celles qui présentent des résultats négatifs, elles seront évidemment plus facilement accessibles dans le cadre des méta-analyses et des revues systématiques (Thornton & Lee, 2000). Du coup, cette façon de faire peut donner la fausse illusion que leurs conclusions découlent de données provenant de l'ensemble des études, alors qu'en fait, elles sont biaisées en raison du nombre excessif de résultats positifs inclus lors de la sélection des études (Rothstein, Sutton, & Borenstein, 2005). Cela a d'importantes implications pratiques car les méta-analyses constituent une source d'information très importante pour les praticiens.

Une autre forme de biais, celui de citation - qui consiste à porter plus attention aux études fréquemment citées-, a été documentée dans plusieurs domaines et disciplines. Afin de mieux comprendre le phénomène, Duyx, Urlings, Swaen, Bouter et Zeegers (2017) ont réalisé une recension systématique de la littérature et une méta-analyse portant sur le biais de citation. Les auteurs visaient deux objectifs : d'une part, ils voulaient identifier et évaluer les études publiées à ce sujet, indépendamment de la discipline scientifique et, d'autre part, ils souhaitaient quantifier l'impact du type de résultat (positif ou négatif) de

l'article sur la probabilité d'être cité. Les auteurs ont effectué une recherche dans les bases de données du *Web of Science Core Collection* et de *Medline* ; 52 études ont été classées selon la discipline scientifique. Des méta-analyses à effets aléatoires ont été effectuées pour mettre en relief l'effet des résultats positifs et négatifs sur la probabilité d'être citée. Les résultats montrent que le biais de citation est plus fréquemment rencontré dans les sciences biomédicales que dans les sciences naturelles. De plus, les articles dont les résultats sont positifs sont cités 1,6 fois plus souvent que les articles avec des résultats négatifs. Enfin, les résultats des articles (positifs **c**. négatifs) et le facteur d'impact des revues sont plus fortement associés à la probabilité d'être cité que tout autre facteur. Les auteurs concluent que les conséquences négatives du biais de citation peuvent être similaires à celles d'autres pratiques de recherche discutables comme le biais de publication. Ils peuvent se produire avec les meilleures intentions et leurs effets individuels peuvent être faibles, mais ils conduisent tout de même à une surreprésentation des résultats positifs dans la littérature scientifique.

La tendance à porter plus attention au facteur d'impact des revues scientifiques et aux études ne rapportant que des résultats positifs est considéré comme une pratique douteuse (Steneck, 2006). Plus souvent qu'autrement, ces pratiques ne sont pas délibérées et leurs effets sont habituellement moins graves que les fraudes habituelles. Néanmoins, ces pratiques sont fréquentes et peuvent avoir un fort impact négatif sur le développement des connaissances. À cet égard, Fanelli (2009) rapporte que les chercheurs déclarent s'engager dans ces pratiques beaucoup plus souvent (environ 34%) que dans l'inconduite en recherche (environ 2%). La plupart des études montrant des biais de citation sont réalisées en sciences biomédicales. Cela pourrait refléter une plus grande conscience de ce type de biais par rapport à d'autres disciplines plutôt qu'une prévalence plus élevée. En fait, le domaine biomédical semble généralement plus avancé dans l'utilisation d'initiatives visant à contrer ces biais comme en témoignent la préinscription des essais cliniques (Kim & Lee, 2015).

Dans une enquête menée par Bouter, Tijdink, Axelsen, Martinson et ter Riet (2016) auprès de 1 353 participants à des conférences internationales sur l'intégrité de la recherche, le biais de citation a été classé comme l'un des comportements fautifs les plus fréquemment rencontrés en recherche. Pour faire leur enquête, les auteurs ont demandé aux participants de noter sur des échelles Likert de 5 points 60 comportements répréhensibles en recherche selon leurs points de vue et leurs perceptions sur la fréquence de ces comportements ainsi que leur impact sur la confiance envers les scientifiques. Les résultats suggèrent que les biais de sélection, les biais de citation et la qualité du mentorat des futurs chercheurs sont considérés comme les problèmes majeurs de la recherche moderne. Les « péchés capitaux » concernent la fabrication, la falsification et le plagiat de résultats de recherche. Quant au plagiat, les participants croient qu'il est courant en recherche, mais que parmi les comportements répréhensibles, il affecte peu leur confiance globale envers la recherche scientifique. Les auteurs suggèrent que les fautes graves attirent l'attention mais qu'il faudrait davantage prévenir les comportements issus de ce qu'ils appellent la *sloppy science*. Ceux-ci correspondent aux différentes formes de violation des principes méthodologiques qui menacent la pertinence, la validité, la fiabilité ou l'efficacité de l'étude en cause et pouvant être commis intentionnellement ou non. Toutefois, étant donné leur fréquence et la gravité de leurs impacts, ces pratiques douteuses se situeraient au cœur du problème de la recherche scientifique plutôt qu'à sa périphérie (Bouter et al., 2016).

Solutions

N'est-il pas surprenant de parler de résultats positifs et négatifs en science? « Tout résultat d'une recherche menée selon une méthodologie correcte n'est-il pas, en soi, digne d'être communiqué aux pairs? » (Chevassus-au-Louis, 2016, p.58). Une étude qui infirme une hypothèse fait avancer la connaissance autant, sinon plus, que celle qui la confirme. Cette assertion est particulièrement vraie pour les sciences humaines et sociales, dont les hypothèses sont souvent infirmées. Il existe déjà deux pistes de solution pour favoriser la publication de résultats négatifs : des journaux centrés uniquement sur les

résultats négatifs et le pré-enregistrement de la méthode et des analyses prévues. Nous en suggérons une troisième : un prix pour le meilleur article dont les résultats sont négatifs.

Les résultats négatifs à l'honneur

Dès 1959, Goldfried et Walters proposaient qu'une façon de faire face au problème des résultats négatifs serait d'avoir un endroit ou publier ces résultats. Cela pourrait être fait en créant un journal de résultats négatifs. De toute évidence, depuis le début des années 2000, la communauté scientifique a compris le problème (Teixeira Silva, 2015) puisque pas moins de 30 revues dont la mission est la publication de résultats négatifs ont été mises sur pied. Le tableau 1 présente la liste de ces revues avec leur première année de parution et leur statut actuel.

Tableau 1 – Journaux qui publient uniquement des résultats négatifs (n=30)

Nom de la revue	ISSN	Début	Fin / statut actuel
All Results Journals Biol	2172-4784	2010	Actif
All Results Journals Chem	2172-4563	2010	Actif
All Results Journals Nano	2444-0035	2010	Actif
All Results Journals Phys	2174-1417	2010	Actif
Contingent Negative Variation	0191-393X	1973	1980
Cortex	0010-9452	2013	Actif
Journal of Articles in Support of the Null Hypothesis	1539-8714	2002	Actif
Journal of Errology		2011	Interrompu
Journal of Interesting Negative Results		2008	2008
Journal of Negative and No Positive Results	2529-850X	2016	Actif
Journal of Negative Observation in Genetic Oncology		1997	2008
Journal of Negative Results – Ecology and Evolutionary Biology	1459-4625	2004	Actif
Journal of Negative Results in BioMedicine	1477-5751	2002	2017
Journal of Negative Results in Speech and Audio Sciences	1547-0407	2004	2004
Journal of Null Results		2010	2011
Journal of Pharmaceutical Negative Results	0976-9234	2010	Actif
Journal of Unsolved Questions		2011	Actif
Nature Negative Results section		2010	Actif
Negat		?	?
Negations	1949-3371	1996	Non résolu
Negative Capability	0277-5166	1981	1994
Negative Results		2016	Actif

Negative Results Scientific Journal		2017	Actif
New Negatives in Plant Science	2352-0264	2014	Actif
Rejecta mathematica	1948-8351	2009	Actif
The International Journal of Negative And Null Results		2014	Interrompu
The Journal of Spurious Correlations		2005	Interrompu
The Null Journal		2009	Interrompu
University of Colorado Database of Negative Results		2011	Interrompu
Yixue Zhengming	1974-8913	1980	Actif

Douze d'entre elles (38.7%) ont cessé leur activité et on ignore le statut de deux autres revues (6.5%), ce qui laisse dix-sept revues (54.8%) encore actives. Un tel résultat n'est guère surprenant puisque Chevassus-au-Louis (2016) rappelle que depuis sa création en 2002, le *Journal of Negative Results in Biomedicine* qui a cessé de paraître en 2017 n'a publié qu'une dizaine d'articles par an et son homologue de *The Journal of Negative Results – Ecology and Evolutionary Biology* n'a publié qu'un ou deux articles par année dans un domaine où les conflits d'intérêts sont pourtant bien moins prégnants.

Méthodes et analyses préenregistrées

La revue *Cortex* « permet depuis 2013 un processus de publication en deux temps. Avant la conduite de l'étude, les méthodes et les analyses prévues sont soumises à une évaluation par les pairs pour acceptation. Une fois l'étude menée et les résultats obtenus, cela garantit leur publication dans la revue, même s'ils s'avèrent négatifs » (Moris, 2017, p.9). Cette méthode en deux temps exige que l'étude soit d'abord préenregistrée, c'est-à-dire que la méthodologie et les analyses prévues soient initialement acceptées par les pairs à la suite de leur évaluation.

Si les revues scientifiques acceptaient de préenregistrer les études sans en connaître les résultats, cela permettrait d'éviter les biais de publication et éliminerait la tentation des chercheurs à recourir à des pratiques douteuses. Ces études préenregistrées découlent du principe selon lequel la qualité des articles doit être évaluée en fonction de l'importance de la question de recherche et de la rigueur de la méthodologie et non pas en fonction du fait que l'hypothèse de recherche soit acceptée ou rejetée. À la fin des années

soixante, Rosenthal (1966) proposait une telle solution qui, selon ses dires, permettrait d'évaluer les études uniquement sur les procédures utilisées. Si celles-ci sont jugées pertinentes, raisonnables et suffisamment rigoureuses pour permettre de tirer des conclusions, indépendamment du résultat (positif ou négatif), les études devraient être pré-acceptées pour publication ultérieure. Quelques années plus tard, Walster et Cleary (1970) proposaient d'appliquer au processus de publication la même règle qui existe en méthodologie expérimentale, soit que toute décision concernant le traitement des données soit prise avant l'inspection des données. De cette façon, la décision de publier ne serait pas liée aux résultats obtenus mais bien à la qualité de la procédure méthodologique.

Dans son ouvrage *The 7 deadly sins of psychology*, Chambers (2017) explique en détail la procédure par laquelle les études seraient préenregistrées préalablement à leur publication. Cette procédure comporte deux phases. Lors de la première phase, six critères sont examinés : (1) l'importance de la (des) question (s) de recherche; (2) la logique, la raison d'être et la vraisemblance des hypothèses proposées; (3) la pertinence et la faisabilité de la méthodologie et des analyses statistiques; (4) la clarté et le degré de ce qu'il faut de détails méthodologiques pour reproduire exactement les procédures expérimentales proposées; (5) une description suffisamment claire et détaillée des méthodes permettant d'empêcher d'éventuelles dérives méthodologiques non dévoilées aux plans expérimentaux et analytiques dans la réalisation ultérieure de l'étude et (6) une vérification pour voir si les auteurs de l'étude ont prévu de contrôler les variables pouvant affecter la validité interne de l'étude, s'assurant ainsi que les résultats obtenus permettent de vérifier les hypothèses énoncées.

À la suite d'une évaluation favorable de ces critères, les éditeurs de la revue peuvent proposer aux auteurs une acceptation de principe. Ceux-ci peuvent alors réaliser la recherche préenregistrée. Une fois l'étude réalisée, la seconde phase consiste à vérifier cinq critères afin de déterminer si celle-ci peut être publiée : (1) on s'assure que les données colligées sont en mesure de vérifier les hypothèses proposées tout en satisfaisant aux critères de contrôle des variables susceptibles d'affecter la validité de l'étude; (2) on

vérifie que l'introduction, le problème et les hypothèses énoncées sont identiques à la soumission initialement approuvée lors de la première phase; (3) on s'assure que les auteurs ont utilisé les procédures expérimentales annoncées; (4) on examine si les analyses post-hoc ajoutées par les auteurs sont justifiées et méthodologiquement valables et (5) on analyse les conclusions de l'article afin de s'assurer qu'elles soient justifiées compte tenu des résultats obtenus.

Selon Chambers, cette procédure de soumission des articles, même si elle n'est pas parfaite, éviterait que la décision de publier ou non repose seulement sur l'évaluation de la pertinence, de la nouveauté ou de l'originalité des résultats. On s'assure également de ne pas tenir compte de la nature positive ou négative des résultats obtenus. Enfin, demander aux auteurs de spécifier leurs hypothèses et leurs plans d'analyse à l'avance permettrait de réduire le recours à des pratiques de recherche douteuses telle que le HARKing. Toutefois, des incitatifs sont nécessaires pour encourager les chercheurs à préenregistrer leurs études. Plutôt que de se concentrer sur le nombre total d'articles publiés, le rendement académique des chercheurs devrait explicitement prendre en compte la proportion d'études amorcées (par exemple, celles qui reçoivent une approbation en matière d'éthique ou un financement) pour lesquelles des protocoles ont été partagés et dont les données de recherche ont été réutilisées par d'autres chercheurs (Chan et al. 2014).

Que le meilleur gagne!

Les solutions pour contrer l'embargo qui est en quelque sorte mis sur les résultats négatifs ne sont guère nombreuses. À ce jour, comme nous l'avons vu, la mise sur pied de journaux qui publient uniquement des résultats négatifs est la seule initiative originale mise en place. Deux autres incitatifs peuvent être envisagés : publier moins d'articles et créer un prix pour le meilleur article dont les résultats sont négatifs. Avec la deuxième solution, nous voudrions désamorcer le *Publish or perish*. Cette proposition rejoint celle de Chavassus-au-Louis (2016), qui est d'avis qu'il y a toutes les raisons de penser que le niveau de publication affiché par certains chercheurs ne peut probablement pas être atteint sans quelque manquement à l'éthique. « Face à la malscience qui ressemble tant à la malbouffe des fast-foods, il faut ralentir. Prendre

le temps. Celui de penser » (p.188). En définitive, il s'agit de prioriser la qualité plutôt que la quantité. Cette solution peut paraitre à première vue surprenante, voire farfelue, mais nous pensons qu'elle pourrait déboucher sur une certaine valorisation des recherches dont les hypothèses sont infirmées et éviter du coup les manœuvres visant à transformer des résultats négatifs en résultats positifs. Un tel prix pourrait également inciter les éditeurs de revues scientifiques à publier des articles dont les résultats sont négatifs lorsque la collecte des données respecte bien les règles de l'art.

Au total, les résultats négatifs devraient avoir un statut…positif. Ils sont précieux puisque la vérité scientifique n'est jamais absolue et demande à être toujours reconquise. D'où la nécessité d'être réfutable, c'est-à-dire susceptible d'être contredite par une expérience empirique, comme le rappelle le critère de réfutabilité de Popper (1973) qui constitue encore, jusqu'à nouvel ordre, une puissante barrière contre la malscience.

Références

Als-Nielsen, B., Chen, W., Gluud, C., & Kjaergard, L. L. (2003). Association of funding and conclusions in randomized drug trials: reflection of treatment effect or adverse events? *JAMA, 290,* 921-928. doi:10.1001/jama.290.7.921

Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment, 20,* 182-196. http://dx.doi.org/10.1111/j.1468-2389.2012.00591.x.

Bouter, L. M., Tijdink, J. K., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review, 1,* 1-8. doi: 10.1186/s41073-016-0024-5.

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.

Chan, AW., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P. C., … van der Worp, H. B. (2014). Increasing value and reducing waste: Addressing inaccessible research. *Lancet, 383,* 257–266. https://doi.org/10.1016/S0140-6736(13)62296-5.

Charlton, B. (1987). Think negative. *New Scientist, 116,* 72.

Chevassus-Au-Louis, N. (2016). *Malscience : de la fraude dans les labos.* Éditions du Seuil.

De Rond, M., & Miller, A. (2005). Publish or perish: Bane or boon of academic life? *Journal of Management Inquiry, 14,* 321-329. Doi : 10.1177/1056492605276850.

Derzon, J. H., Sale, E., Springer, J. F., & Brounstein, P. (2005). Estimating intervention effectiveness: Synthetic projection of field evaluation results. *The Journal of Primary Prevention, 26,* 321–343. DOI: 10.1007/s10935-005-5391-5.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA, 263,* 1385–1389.

Djulbegovic, B., Lacevic, M., Cantor, A., Fields, K. K., Bennett, C. L., Adams, J. R., Kuderer, N M., & Lyman, G. H. (2000). The uncertainty principle and industry-sponsored research. *Lancet, 356,* 635-638.

Doucouliagos, C., Laroche, P., & Stanley, T. D. (2005). Publication bias in union-productivity research. *Relations Industrielles/Industrial Relations, 60,* 320–346. doi :10.7202/011724ar

Durlak, J.A., & Dupre, E.P. (2008). Implementation Matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327-350. https://doi.org/10.1007/s10464-008-9165-0.

Duval, S., & Tweedie, R. (2000). A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89-98. doi :11.1080/01621459.2000.10473905.

Duyx, B., Urlings, M. J. E., Swaen, G. M. H., Bouter, L. M., & Zeegers, M. P. (2017). Scientific citations favor positive results: A systematic review and meta-analysis. *Journal of Clinical Epidemiology, 88,* 92-101. https://doi.org/10.1016/j.jclinepi.2017.06.002.

Egger, M., & Davey Smith G. (1995). Misleading meta-analysis. *British Medical Journal, 310,* 752–754. doi: https://doi.org/10.1136/bmj.310.6982.752

Egger, M., Davey Smith, G., Schneider, M., & Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634. doi: https://doi.org/10.1136/bmj.315.7109.629.

Egger, M., Davey Smith, G., & Altman D (2001). *Systematic reviews in health care: Meta-analysis in context. 2nd Edition of Systematic Reviews.* BMJ Books.

Ernst, E., & Pittler, M.H. (1997). Alternative therapy bias. *Nature, 385,* 480.

Fanelli D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *Plos One, 4(5).* doi:ARTN e5738 10.1371/journal.pone.0005738.

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States data. *PLoS One, c10271.* doi:10. 1371/journal.ponc.0010271.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90,* 891–904.

Ferguson, C. J. (2015). "Everybody knows psychology is not a real science": Public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *American Psychologist, 70,* 527–542.

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in Social Research.* Sage.

Glasgow, R. E., & Steiner, J. F. (2018). Comparative effectiveness research to accelerate translation: recommendations for an emerging field of science. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice (Second edition)*. Oxford University Press.

Goldfried, M. R., & Walters, G. C. (1959). Needed: Publication of negative results. *American Psychologist, 14,* 598. http://dx.doi.org/10.1037/h0038357

Grégoire, G., Derderian, F., & Le Lorier, J. (1995). Selecting the language of the publications included in a meta-analysis: Is there a Tower of Babel bias? *Journal of Clinical Epidemiology, 48,* 159-163. DOI:10.1016/0895-4356(94)00098-b.

Hartley, A., Shah, M., Nowbar, A., Rajkumar, C., Howard, J., & Francis, D. (2018). Key opinion leaders' guide to spinning a disappointing clinical trial result. *British Medical Journal,* 363. https://doi.org/10.1136/bmj.k5207

Hubbard, R., & Strong J. S. (1992). Are null results becoming an endangered species in marketing? *Marketing Letters, 3,* 127-136.

Joober R., Schmitz N., Annable L., & Boksa P. (2012). Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry and Neuroscience, 37,* 149-152. doi: 10.1503/jpn.120065

Klatersky, J. (1987). La publication scientifique. *Revue médicale de Bruxelles,8*(8), 413-419.

Kim, M. A., & Lee, H. S. (2015). Duo-trio difference-preference test with two replications: Use of psychological biases for measuring meaningful preference. *Journal of Sensory Studies, 30,* 211-24. https://doi.org/10.1111/joss.12152.

Koren, G., Shear, H., Graham, K., & Einarson, T. (1989). Bias against the null hypothesis: The reproductive hazards of cocaine. *The Lancet, 2,* 1440–1442. doi:10.1016/s0140-6736(89)92044-8.

Kyzas P. A., Denaxa-Kyza D., & Ioannidis, J. P. (2007). Almost all articles on cancer prognostic markers report statistically significant results. *European Journal of Cancer, 43,* 2559-2579. https://doi.org/10.1016/j.ejca.2007.08.030.

Larivée, S. (2017). Le côté sombre de la science. *Revue de psychoéducation, 46* (2), 421-452. Doi : https://doi.org/10.7202/1042258ar

Liebeskind, D. S., Kidwell, C. S., Sayre, J. W., & Saver, J. L. (2006). Evidence of publication bias in reporting acute stroke clinical trials. *Neurology, 67,* 973–979.

Lindsay, R. M. (1994). Publication system biases associated with the statistical testing paradigm. *Contemporary Accounting Research, 11,* 33-57. https://doi.org/10.1111/j.1911-3846.1994.tb00435.x.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7,* 537–542. doi: 10.1177/1745691612460688.

Moher, D., Forin, P., Jadad, A. R., Juni, P., Klassen, T., Lelorier, J., Liberati, A., Linde, K., & Penna, A. (1996). Completeness of reporting of trials published in languages other than English: Implications for conduct and reporting of systematic reviews. *Lancet, 347,* 363–366.

Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P. & Walsh, S. (1995) Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials, 16,* 62-73. https://doi.org/10.1016/0197-2456(94)00031-W.

Moris, K. (2017). La publication des résultats de recherche négatifs. *Science & pseudosciences, 319,* 8-9.

Morrison, A., Polisena, J., Husereau, D., Moulton, K., Clark, M., Fiander, M., Mierzwinski-Urban, M., Clifford, T., Hutton, B., & Rabb D. (2012). The effect of English-language restriction on systematic review-based meta-analyses: A systematic review of empirical studies. *International Journal of Technology Assessment in Health Care, 28,* 138–44. doi: https://doi.org/10.1017/S0266462312000086.

Moscati, R., Jehle, D., Ellis, D., Fiorello, A., & Landi, M. (1994). Positive-outcome bias: Comparison of emergency medicine and general medicine literatures. *Academic Emergency Medicine, 1,* 267–271. https://doi.org/10.1111/j.1553-2712.1994.tb02443.x.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* 1-8. doi: 10.1126/science.aac4716.

Pashler, H. & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531-534. https://doi.org/10.1177/1745691612463401.

Popper, K. (1973). *La logique de la découverte scientifique*. Payot.

Prinz, F., Schlange, T. & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery, 10,* 712.

Rochon, P., Gurwitz, J. H., Simms, R. W., Fortin, P. R., Felson, D. T., Minaker, K. L., & Chalmers, T. C. (1994). A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Archives of Internal Medicine, 154,* 157–163. doi:10.1001/archinte.1994.00420020059007.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. Appleton-Century-Crofts.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. John Wiley & Sons.

Smart, R. G. (1964). The importance of negative results in psychological research. *Canadian Psychologist/Psychologie canadienne, 5,* 225-232.

Smith, M. B. (1956). Editorial. *The Journal of Abnormal and Social Psychology, 52,* 1-4.

Smith, M.L. (1980a). Publication bias and meta-analysis. *Evaluating Education, 4,* 22–24.

Smith, M.L. (1980b). Sex bias in counseling and psychotherapy. *Psychological Bulletin, 87,* 392–407. http://dx.doi.org/10.1037/0033-2909.87.2.392.

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J. J., Sutton, A. J., Hing, C., Kwok, C. S., Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment, 14* (8). doi: 10.3310/hta14080.

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12,* 53-74.

Sterling, T.D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association, 54,* 30–34. https://doi.org/10.1080/01621459.1959.10501497.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49,* 108–112. doi: 10.2307/2684823.

Sutton, A. J., Duval, S., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal, 320,* 1574-1577.

Teixeira da Silva, J. A. (2015). Negative results: Negative perceptions limit their potential for increasing reproducibility. *Journal of Negative Results in Biomedicine, 14*(12),1-4. doi : 10.1186/s12952-015-0033-9.

The W.K. Kellogg Foundation. (2017). The Step-By-Step Guide to Evaluation: How to Become Savvy Evaluation Consumers. Retrouvé https://www.wkkf.org/resource-directory/resource/2017/11/wk-kellogg-foundation-step-by-step-guide-to-evaluation.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology, 53,* 207-16. https://doi.org/10.1016/S0895-4356(99)00161-4

Turner, E.H., Matthews, A.M., Linardatos, E., Tell, R.A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *The New England Journal of Medicine, 358,* 252-260. doi: 10.1056/NEJMsa065779.

Umscheid, C. A., Margolis, D. J., & Grossman, C. E. (2011). Key concepts of clinical trials: A narrative review. *Postgraduate Medicine, 123,* 194-204. https://doi.org/10.3810/pgm.2011.09.2475.

U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth, and Families (2010). The Program Manager's Guide to Evaluation, Second Edition. Washington, DC

Vickers, A., Goyal, N., Harland, R., & Rees, R. (1998). Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clinical Trials, 19,* 159–166.

Walster, G. W., & Cleary, T. A. (1970). A Proposal for a new editorial policy in the social sciences. *The American Statistician*, 24, 16-19. doi:10.1080/00031305.1970.10478884.

Wang, M. Q., Yan, A. F., & Katz, R. V. (2018). Researcher requests for inappropriate analysis and reporting: A U.S. survey of consulting biostatisticians. *Annals of Internal Medecine, 169* (8): 554-558. doi: 10.7326/M18-1230. doi: 10.7326/M18-1230.

Wilson, F. D., Smoke, G. L., & Martin, J. D. (1973). The replication problem in sociology: A report and a suggestion. *Sociological Inquiry, 43,* 141-149. https://doi.org/10.1111/j.1475-682X.1973.tb00711.x.

Xia, J., Wright, J., & Adams, C. E. (2008). Five large Chinese biomedical bibliographic databases: Accessibility and coverage. *Health Information and Libraries Journal, 25,* 55–61. https://doi.org/10.1111/j.1471-1842.2007.00734.x.

Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine, 5* (10). doi:10.1371/journal.pmed.0050201.