

Peut-on croire les publications ?
**Biais et embellissements
polluent la science**

Hervé Maisonneuve



Hervé Maisonneuve est médecin de santé publique. Il est consultant en rédaction scientifique et anime le blog www.redactionmedicale.fr. Il a partagé sa carrière entre des postes dans l'industrie pharmaceutique et des institutions publiques françaises. Il a été professeur associé en santé publique (Paris 7 et Paris 11). Il a été membre de comités de rédactions de revues biomédicales (*JAMA*, *La Presse Médicale*, *European Science Editing*), et président de *European Association of Science Editors* (1994-1997).

Ce texte repose partiellement sur des extraits d'une conférence « *Revue scientifique versus institutions : gestion des mauvaises pratiques !* » présentée le 18 juin 2016, lors du colloque international « *Recherche et action sur la fraude scientifique et le plagiat* », Université de Genève.

Depuis les années 2000, des experts attirent l'attention sur les dérives du système des publications. La délégation partielle aux revues de missions institutionnelles (évaluation des carrières, des équipes, allocation de ressources) a orienté la science vers des pratiques peu explicites. Cette délégation, non formalisée, est une habitude prise par les structures d'évaluation. La notoriété d'une revue (et non la qualité d'un article) est devenu un critère d'évaluation. Un dialogue de sourds existe entre revues et institutions, au détriment de la bonne science.

**Les revues évaluent, enregistrent,
diffusent et archivent le savoir**

Depuis la fin du XVII^e siècle jusqu'au début des années 2000, le modèle des revues scientifiques était stable. Le système de l'évaluation par les pairs (*peer-review*) permettait de sélectionner les articles selon des directives éditoriales spécifiques à chaque revue. L'évaluation par les pairs a été confiée à des experts académiques. Un article accepté permettait à son auteur de prendre date et avait ainsi un rôle d'enregistrement des résultats d'une recherche. La diffusion des avancées de la science était une mission des revues. Le savoir était archivé dans les bibliothèques qui aidaient les chercheurs à sélectionner leurs lectures. Les revues étaient la propriété des sociétés savantes, avec parfois le soutien, voire la participation active, des maisons d'éditions. Ces quatre missions de la gestion du savoir (évaluation, enregistrement, diffusion, archivage) confiées aux revues ont attesté d'un équilibre entre les institutions de recherche, les sociétés savantes et les maisons d'édition [1].

Évaluation des carrières et allocations de ressources

Dans les années 1980, pour évaluer les chercheurs et allouer des ressources, les institutions ont utilisé un indicateur conçu pour estimer la notoriété des revues : le facteur d'impact. Cet indicateur a été inventé dans les années 1960 par un documentaliste américain afin d'aider ses confrères à choisir les abonnements et orienter les annonceurs. Les institutions ont fait confiance à cet indicateur, partant du principe que publier dans une revue à haut facteur d'impact était la reconnaissance de la qualité de la recherche. Des revues, avec la complicité des universitaires chargés d'évaluer les carrières et d'allouer des ressources, ont utilisé cet indicateur pour attirer des recherches dites innovantes. Ces revues n'étaient pas intéressées par la validité et la reproductibilité des travaux : l'innovation était le sésame pour publier. Il est devenu plus important de savoir dans quelle revue a été publié un article plutôt que d'évaluer les travaux dont il rend compte. Les dérives de la bibliométrie sont connues [2]. Les missions des revues n'ont jamais été d'évaluer des chercheurs, d'aider à l'allocation de ressources, ni de détecter les mauvaises pratiques [3].

Les conflits d'intérêts omniprésents au sein des revues

Ces conflits sont plutôt intellectuels que financiers. La sélection des articles est l'objet de négociations, de tractations entre les acteurs que sont les auteurs, les rédacteurs des revues, les relecteurs externes et les propriétaires des revues [3]. Ces conflits cachés sont rarement discutés : un rédacteur peut favoriser une école de pensée ; son choix des relecteurs est déjà une orientation de la décision ; le conseil d'administration de la société savante nomme le rédacteur en chef et lui impose quelques règles non écrites ; des mauvais articles pouvant attirer des citations sont acceptés ; des articles allant dans le sens d'idées dominantes sont parfois acceptés alors qu'ils reposent sur de petits échantillons avec des méthodes notoirement inadaptées... Sans parler des cartels de rédacteurs qui s'entendent pour citer des revues amies, des ententes implicites entre auteurs et relecteurs complaisants, etc.



Paul Delaroche (1797 - 1856)

La Renommée distribuant des couronnes

Hémicycle de l'École des
Beaux-Arts de Paris

Fraude	Pratiques discutables en recherche	Méconnaissance méthodologique
<ul style="list-style-type: none"> ▪ Fabrication ▪ Falsification ▪ Plagiat 	<ul style="list-style-type: none"> ▪ « Torture » ou « massage » des données ▪ Changement du critère de jugement ▪ Choix sélectif / omission de données ▪ Références erronées ▪ Changement de tests statistiques ▪ P-Harking (<i>Hypothesizing After Results are Known</i>) ▪ Manipulation d'images ▪ Paternité des articles : conflits d'auteurs ▪ Études animales trompeuses ▪ Non-publication de recherches financées ▪ Résumés, communiqués de presse embellis ▪ ... 	<ul style="list-style-type: none"> ▪ Méthodes « faibles » ▪ Méthodes inappropriées ▪ Échantillon trop faible ▪ Erreurs statistiques ▪ Pas de recherche documentaire avant la recherche ▪ Non respect de réglementations
Intentionnel	Non intentionnel	

Les bonnes pratiques en recherche sont en général explicitement formulées dans des guides adaptés aux domaines de recherche. Il existe des comportements déviants par rapport aux bonnes pratiques : la fraude, les pratiques discutables en recherche, et la méconnaissance méthodologique. La fraude (falsification, fabrication de données et plagiat) est toujours intentionnelle. La méconnaissance méthodologique résulte souvent d'une formation ou d'un encadrement insuffisants pour faire des recherches. Elle est le fait de chercheurs honnêtes.

Les pratiques discutables en recherche sont silencieusement tolérées par la communauté scientifique. Ce sont de petits arrangements avec la méthodologie, le plus souvent pour « embellir » les résultats. Ces « arrangements » sont parfois utiles (des données aberrantes explicitement exclues ; une analyse *a posteriori* explicitement faite). Des chercheurs honnêtes glissent progressivement sous la pression pour publier ou obtenir des crédits de recherche. Ce qui n'est pas acceptable, c'est que les auteurs ne sont pas transparents sur des pratiques cachées intentionnellement. Les biais ne sont pas des pratiques discutables en recherche. La liste des exemples est longue, et une littérature importante existe [4]. Le développement des mesures en faveur de l'intégrité scientifique, de la recherche responsable, et la recherche ouverte (*open access*, *open data*, *open sources*) devraient diminuer les pratiques discutables en recherche. Figure adaptée de Seror et Ravaud. *La Presse Médicale*, 2012, 41:835-840.

Les pratiques discutables en recherche

Des pratiques discutables se situent entre les bonnes pratiques de recherche et la malveillance scientifique (falsification, fabrication de données, plagiat) [4]. De nombreux exemples ont été décrits dans toutes les disciplines.

Ces pratiques sont mises en évidence lors de la publication des recherches (*peer-review*) ou après (post-publication *peer review*).

Qualification des auteurs

Les pratiques discutables les plus fréquentes sont probablement l'attribution de la paternité des œuvres, autrement dit la liste et l'ordre des auteurs des publications [5]. La littérature est abondante, avec des exemples d'auteurs honoraires ou cadeaux (ajoutés volontairement sur un article), et d'auteurs fantômes (oubliés sur un article). La prévalence des auteurs honoraires et fantômes est importante : 21 % des 630 arti-



Le colporteur (Anonyme, XVII^e siècle)

cles des six revues biomédicales les plus prestigieuses avaient en 2008 des auteurs n'ayant aucune qualité pour s'attribuer la paternité d'un travail, ou des auteurs oubliés [6]. La prévalence des auteurs fantômes est en baisse rapide concernant les rédacteurs professionnels des milieux industriels [6]. La bonne pratique est de remercier les rédacteurs professionnels explicitement. Aucune donnée n'existe pour les auteurs fantômes de milieux académiques (ce thésard qui a fait tout le travail et qui est oublié !).

Un quart des références comportent des erreurs

Cet aspect n'attire pas l'attention des chercheurs et des revues, tant la pratique de la citation a été manipulée pour diverses raisons. Les faits sont clairs : une analyse de 7321 citations, compilant des données dans vingt-huit articles sur la qualité des références, a montré que 25,4 % des références avaient des erreurs, soit majeures 11,9 %, soit mineures 11,5 % [7].

La mauvaise qualité des statistiques

Les pratiques des chercheurs dans le domaine de la statistique sont parfois très mauvaises. Trop d'articles ne font pas appel à un statisticien, ressource rare dans certaines équipes. Les chercheurs font leurs propres analyses statistiques, et des erreurs existent. Des études sur de grands échantillons ont montré qu'il y avait environ 10 à 12,5 % d'erreurs graves mettant en cause la conclusion des articles [8,9].

L'effet chrysalide ou embellissement de données

Deux méthodes sont connues pour métamorphoser de piteux résultats en beaux articles : proposer des hypothèses *a posteriori* pour qu'elles correspondent aux données obtenues et torturer les données pour qu'elles répondent aux hypothèses [10]. Des publications ont mis en évidence ces pratiques, en comparant les protocoles de recherche (quand ils existent) avec les publications [11], ou en comparant les thèses avec les publications [10]. Il est admis que tout ne soit pas publié, mais la transparence nécessite de dire quelles hypothèses du protocole ont été publiées, lesquelles n'ont pas été publiées, et lesquelles ont été proposées après l'observation des résultats.

Les manipulations d'images

En biologie, la manipulation d'images est fréquente [12]. Des scandales ont mis en évidence de telles manipulations, concernant le plus souvent des résultats de biologie (chromatographie, *western blots*). Des revues prestigieuses en biologie ont un processus de relecture des images. Pour les quatre

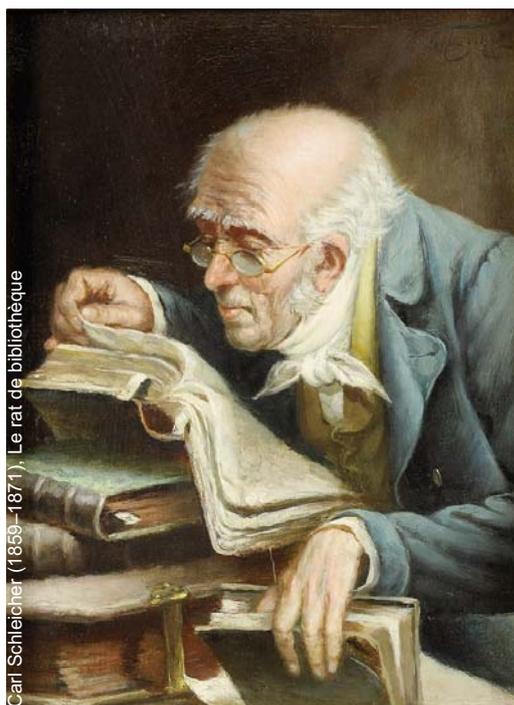
revues EMBO (*European Molecular Biology Organization*), environ 20 % des articles sur le point d'être acceptés ont révélé des manipulations des images [12]. La plupart étaient des embellissements dits naïfs qui ont été rectifiés. Environ 1 % étaient des manipulations conduisant au rejet du manuscrit.

Les dérives du chercheur honnête

Ces pratiques sont admises silencieusement dans la communauté scientifique. Ce sont des chercheurs honnêtes qui emploient ces petits arrangements avec la vérité, sans se cacher des collègues, voire avec complaisance : « *tous les autres font pareil* ». Des sociologues ont expliqué ces glissements progressifs du chercheur [10] : des recherches apportent des résultats dits négatifs néanmoins intéressants, voire allant à l'encontre de théories admises ; les comités de rédaction des revues refusent les articles basés sur ces données jugées peu innovantes ; après plusieurs tentatives, la carrière du chercheur stagne car les publications ne suivent pas ; le chercheur devient progressivement agressif contre les relecteurs anonymes et les rédacteurs des revues ; son niveau de vie ne progresse pas comme prévu ; il essaie alors de légèrement embellir la vérité pour être publié ; progressivement, quelques publications sont acceptées, bien sûr avec des embellissements connus des collègues... Rien de mal, il faut publier ! La frontière entre la réalité de la recherche et la vitrine des publications est floue.

Fiabilité et reproductibilité de la recherche

Les liens entre pratiques discutables en recherche et reproductibilité sont difficiles à explorer. Est-ce que la difficulté à reproduire des résultats est inhérente à la science ? Un séminaire de quatre-vingts experts réunis un jour en Angleterre a examiné la fiabilité et la reproductibilité des recherches biomédicales [13]. Ils ont identifié six problèmes expliquant la mauvaise reproductibilité, toujours à partir de l'évaluation des publications : (1) la recherche d'hypothèse *a posteriori* pour justifier une différence statistiquement significative ; (2) l'omission des résultats « négatifs » ; (3) les études de puissance insuffisante ; (4) les erreurs ; (5) les méthodes insuffisamment décrites ; (6) les méthodes expérimentales faibles. En mai 2016, la revue *Nature* [14] a publié les résultats d'une enquête auprès de 1576 chercheurs (chimie, physique, sciences de la terre, biologie, médecine, autres). Plus de 70 % d'entre eux ont essayé, sans succès, de reproduire les expériences d'un collègue, et plus de 50 % ont échoué pour reproduire leurs propres expé-



Carl Schiecher (1859-1877). Le rat de bibliothèque



Adaptation des oeuvres de Joos van Cleve et Albrecht Dürer, Saint Jerome à Lisbonne

riences. Cela ne veut pas dire que la littérature est fausse, mais les facteurs suggérés par les chercheurs étaient : publications sélectives des données, pression pour publier, faible puissance statistique, absence de reproduction au sein du laboratoire d'origine, supervision insuffisante des jeunes chercheurs, méthodes et codes non disponibles, mauvaises méthodes, données sources non disponibles, fraude, et *peer-review* insuffisant.

De rares rédacteurs osent dire que la qualité de la littérature est mauvaise

Le rédacteur en chef du *Lancet* s'est permis d'écrire un éditorial explicite, dont nous avons traduit quelques phrases [15] : « *La mise en accusation de la science est simple : une grande partie de la littérature scientifique, peut-être la moitié, peut tout simplement être fausse. Gangrénée par des études avec de petits échantillons, des effets minuscules, des analyses exploratoires invalides et des conflits d'intérêts évidents, associée à une obsession de poursuivre les tendances à la mode d'importance douteuse, la science a pris un virage vers l'obscurantisme. Comme l'a dit un participant, "des méthodes nulles donnent des résultats" [...]. L'endémicité apparente du mauvais comportement en recherche est alarmante. Dans leur quête pour raconter une histoire convaincante, les scientifiques sculptent trop souvent leurs données en fonction de leur vision préférée du monde. Ou ils réajustent les hypothèses pour correspondre à leurs données [...]. Les rédacteurs de revues méritent également leur juste part de critiques. Nous aidons et encourageons les pires comportements. Notre soumission aux facteurs d'impact engendre une concurrence malsaine pour accéder au cercle restreint d'un petit nombre de revues privilégiées. Notre amour de "l'importance" polue la littérature avec beaucoup de contes de fée statistiques. [...] Les revues ne sont pas les seules en tort. Les universités sont dans une lutte perpétuelle*

à la recherche de financements et de talents, ce qui pousse à l'utilisation d'indicateurs réducteurs, comme les publications dans les revues avec un facteur d'impact important. Les procédures nationales d'évaluation [...] incitent aux mauvaises pratiques. Et les chercheurs eux-mêmes, y compris la plupart de ceux en charge de responsabilités, font peu pour changer une culture de recherche qui occasionnellement frôle les mauvaises conduites. » Un aveuglement organisationnel : tous coupables !

Toute la communauté scientifique semble aveugle et immobile ; les institutions et les revues s'accusent ; tous sont témoins de faits qui n'appellent pas de réactions ; dans la base de données *Medline*, sur 29 725 résumés d'articles des revues cliniques prestigieuses, 33 % des résumés contenaient des statistiques et la plupart avaient un résultat « *statistiquement significatif* » (est-ce la réalité de la science ?) [16] ; les données dites « positives » sont parfois publiées plusieurs fois alors que les données dites « négatives » ne sont pas toujours publiées ; des enquêtes et consensus d'experts ont montré qu'environ 50 % des articles sont embellis et ne reflèteraient pas les données observées, et qu'environ 50 % des recherches publiées ne sont pas reproductibles [13,14].

Un dialogue de sourds entre institutions et revues scientifiques

Qui sont les responsables de cette situation ? Des universitaires pensent sincèrement que tout est de la faute des revues scientifiques qui travaillent mal. Les comités de rédaction de revues pensent que si les institutions faisaient bien leur travail, ils ne devraient recevoir que des manuscrits de qualité, pré-évalués par les institutions. Il y a du vrai de part et d'autre. Dialogue de sourds à l'évidence. Il est facile de dire que les revues font très mal l'évaluation par les pairs ; c'est vrai, mais ce sont des universitaires qui font ces relectures. Il est facile de dire que les universitaires soumettent des manuscrits de médiocre qualité ; c'est vrai, mais ces revues les publient quand même.

Il faut publier moins, et des articles de qualité avec des données ouvertes

Il existe une pollution de littérature scientifique causée par la course à la publication, et cette situation n'est pas toujours reconnue par la communauté scientifique. Quelques chercheurs osent expliquer qu'il faudrait diminuer la pression pour publier afin d'éviter une dégradation de la qualité des articles [17]. Ils commencent à être entendus. Les articles sont une représentation déformée de la réalité de la recherche. La révolution de la recherche ouverte pourrait apporter des améliorations avec la mise en ligne de toutes les données « *Open access / Open data / Open source* » [18]. L'ouverture de toutes les



données est indispensable et devrait changer la communication scientifique : protocoles en ligne, accès aux données sources, accès aux avis des relecteurs, commentaires rapides des articles (« *post-publication peer-review* »), correction des erreurs, rétractation des articles, etc.

Une excellente nouvelle a été la publication d'un rapport en France sur l'intégrité scientifique [19]. Ce rapport contient seize propositions, avec la suggestion de créer un Office français d'intégrité scientifique (OFIS). Mais restons à la fois sceptiques (qui va mettre en œuvre ces recommandations ?) et confiants (car évoquer les pratiques discutables en recherche est possible). ■

Références

- [1] Ware M, Mabe M. The STM report. "An overview of scientific and scholarly journal publishing". *International Association of Scientific, Technical and Medical Publishers*, 4^e édition, mars 2015. Sur le site www.stm-assoc.org (consulté le 20 août 2016).
- [2] Maisonneuve H. « L'autosuffisance des dictateurs du facteur d'impact cache leur ignorance de la bibliométrie ». *Médecine Thérapeutique*, 2015, 21:77-79.
- [3] Maisonneuve H. « Le management des erreurs et fraudes scientifiques par les revues biomédicales : elles ne peuvent pas se substituer aux institutions ». *Presse Médicale*, 2012, 41:853-860.
- [4] Banks GC, O'Boyle Jr EH, Pollack JM, White CD, Batchelor JH, Whelpley CE, et al. "Questions about questionable research practices in the field of management: a guest commentary". *J Management*, 2016, 42:5-20.
- [5] Marušić A, Bošnjak L, Jerončić A. "A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines". *Plos One*, 2011, 6(9): e23477.
- [6] Wislar JS, Flanagan A, Fontanarosa PB, DeAngelis CD. "Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey". *BMJ*, 2011, 343:d6128.
- [7] Jergas H, Baethge C. "Quotation accuracy in medical journal articles – a systematic review and meta-analysis". *PeerJ*, 2015, 3:e1364.
- [8] Veldkamp CL, Nuijten MB, Dominguez-Alvarez L, van Assen MA, Wicherts JM. "Statistical Reporting Errors and Collaboration on Statistical Analyses in Psychological Science". *Plos One*, 2014, 9(12):e114876. doi :10.1371/journal.pone.0114876
- [9] Nuijten MB, Hartgerink CH, van Assen MA, Epskamp S, Wicherts JM. "The prevalence of statistical reporting errors in psychology (1985-2013)". *Behav Res Methods*, 2015 Oct 23 [Epub ahead of print].
- [10] O'Boyle Jr EH, Banks GC, Gonzalez-Mulé E. "The Chrysalis effect. How ugly initial results metamorphose into beautiful articles". *J Management*, 2014 March 19 [Epub ahead of print]. doi :10.1177/0149206314527133
- [11] The COMPare Trials Project. Goldacre B, Drysdale H, Powell-Smith A, Dale A, Milosevic A, Slade E, et al. Sur le site www.COMPare-trials.org (consulté le 20 août 2016).
- [12] Van Noorden R. "The image detective who roots out manuscript flaws. Jana Christopher discusses her job as a data integrity analyst at EMBO Press". *Nature News*, 2015 June 12. Sur le site www.nature.com (consulté le 20 août 2016).
- [13] Academy of Medical Sciences, BBSRC, MRC, Wellcome Trust. "Reproducibility and reliability of biomedical research: improving research practice". Symposium report, October 2015. Sur le site www.acmedsci.ac.uk (consulté le 20 août 2016).
- [14] Baker M. "Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the 'crisis' rocking science and what they think will help". *Nature*, 2016, 533:452-454.
- [15] Horton R. "Offline: what is medicine's 5 sigma?" *Lancet*, 2015, 385:1380.
- [16] Chavalarias D, Wallach JD, Li AH, Ioannidis JP. "Evolution of reporting P values in the biomedical literature, 1990-2015". *JAMA*, 2016, 315(11):1141-1148.
- [17] Sarewitz D. "The pressure to publish pushes down quality". *Nature*, 2016, 533:147.
- [18] McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, Lin J, et al. "How open science helps researchers succeed". *eLife*, 2016, 10.7554/eLife.16800
- [19] Corvol P. « Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique ». Rapport remis le 29 juin 2016 à Thierry Mandon, Secrétaire d'état, chargé de l'enseignement supérieur et de la recherche. Sur le site www.enseignementsup-recherche.gouv.fr/ (consulté le 20 août 2016).