

Études scientifiques : quelle validation ?

Compte rendu d'un Café de la statistique



Jean-François ROYER

SFds

Comment s'assurer de la validité des résultats d'une étude scientifique ? Aujourd'hui la question est posée, après que de plus en plus d'articles publiés ont fait l'objet de rétractations. Des sites Internet¹ se consacrent à documenter les cas de retrait. Des statisticiens s'interrogent sur leurs procédures traditionnelles, et proposent de nouveaux standards. De façon plus large, le mouvement « pour une recherche reproductible » vise à permettre les validations multiples par des tiers. Ces réponses ne doivent pas dissimuler les causes des dérives, dont certaines sont à chercher dans les modes de financement et d'évaluation de la recherche.

Un Café de la Statistique s'est tenu sur ce thème en février 2015. L'intervenant était Stéphane Gregoir, Insee, ancien directeur de la recherche à l'EDHEC (Ecole des Hautes Etudes Commerciales) et ancien directeur du CREST (Centre de recherche en Economie et en Statistique). Vidéo et compte rendu détaillé sont disponibles sur le site de la SFds

De plus en plus d'études scientifiques sont contestées après avoir été publiées. Le doute atteint beaucoup de disciplines. Les sciences biomédicales sont en première ligne : selon des chercheurs américains (Fang, Steen et Casadevall, 2012), deux mille articles publiés depuis 1975 dans des revues de bio-médecine et de sciences de la vie ont été retirés après leur publication, et le nombre de ces retraits est en forte augmentation depuis le début des années 2000. L'économie est également touchée : des articles d'économistes renommés, publiés dans des revues prestigieuses, ont dû être rectifiés après que des vérifications ont révélé que les conclusions étaient entachées d'erreurs². Des exemples similaires se rencontrent dans d'autres disciplines, qu'elles relèvent des sciences « dures », y compris les mathématiques, ou des sciences humaines et sociales.

Un certain nombre de retraits sont dus à la découverte de plagiat, ou d'une version déjà publiée de la même étude. Les autres cas proviennent soit d'erreurs constatées après publication, soit de fraudes. La fraude ici consiste essentiellement en la construction de données fabriquées ou manipulées pour fournir un résultat désiré. Le plus souvent, c'est donc le comportement de l'auteur de l'article qui est en cause. Depuis l'assertion trop hâtive jusqu'à la fraude caractérisée, on constate toute une gamme de comportements susceptibles d'aboutir à la publication de conclusions fausses.

1. Comme le site « Retraction watch » cité dans « Des faussaires dans les labos », article de David Larousserie, Le Monde Sciences et technologie, 12 mai 2015
2. Un exemple souvent cité est (Reinhart & Rogoff 2011).

Un problème qui concerne la statistique

Les statisticiens sont particulièrement concernés par ce problème. Pourquoi ? Parce que leur discipline joue un rôle pivot dans la construction des preuves des résultats scientifiques. Bien rares sont les études empiriques dont les résultats peuvent être obtenus et présentés sans utiliser des tests statistiques. La validité de la preuve dépend du bon emploi de ces méthodes, standardisées il y a plusieurs décennies³. Trop souvent, certains auteurs appliquent mécaniquement des recettes en utilisant directement les sorties des logiciels sans en maîtriser les conditions d'utilisation : leur formation aux méthodes statistiques devrait être renforcée. Plus en profondeur, certains statisticiens réexaminent l'usage standard : les tailles d'échantillon requises sont-elles suffisantes ? Les niveaux d'erreur admis ne sont-ils pas trop grands ? En utilisant une approche bayésienne, des auteurs montrent comment en utilisant le seuil usuel (le fameux « $\alpha=5\%$ »), on est conduit à publier un pourcentage important d'études qui concluent à tort qu'un résultat est « significatif » (Johnson, 2013). Ces auteurs préconisent d'utiliser des standards beaucoup plus stricts pour établir les résultats scientifiques et restaurer la confiance du public dans la science.

Le fonctionnement de la recherche en question

D'autre part, des pratiques qui relèvent de l'organisation de la recherche « poussent à la faute ». Ainsi, les résultats négatifs sont peu considérés, rarement publiés, peu cités ; au contraire un résultat positif, même à la limite de la significativité, a des chances d'être souvent cité. Dans un contexte où la carrière d'un chercheur, et son accès aux sources de financement, dépendent pour une part du nombre de ses publications et du nombre des citations qui en sont faites, l'effet de ce biais peut être dramatique.

Le système des « referees », qui doivent examiner les articles avant publication dans les revues scientifiques, devrait constituer un garde-fou : un bon referee ne laisse pas passer un article contenant des conclusions fausses ou insuffisamment prouvées. Ce garde-fou ne fonctionne pas toujours correctement. Le « referee » dispose rarement d'une information complète : protocoles d'expérience, bases de données, programmes informatiques, tout cela ne lui est pas toujours communiqué, même si la situation évolue (voir plus loin). Et surtout, le referee ne travaille pas pour lui-même, mais pour la communauté scientifique : c'est un travail astreignant, peu visible et donc peu valorisé. Certains chercheurs confirmés acceptent ce rôle, mais, pris par les délais, sous-traitent les rapports demandés à des chercheurs plus jeunes, quand ce n'est pas à des étudiants insuffisamment expérimentés pour ce travail.

Enfin, lorsqu'un chercheur veut prouver « à tout prix » un résultat qui conforte une position générale qu'il a sur l'économie ou sur la société, il peut devenir moins « regardant » sur la rigueur de ses raisonnements, voire sur l'intégrité de ses données et de ses calculs.

La multiplication des mises en cause d'études scientifiques a provoqué une prise de conscience, et un appel à plus d'exigences dans l'établissement et la publication des résultats.

La sévérité à l'égard des fraudeurs se renforce. Aux Etats-Unis, s'agissant du domaine biomédical, un « Office of research integrity » a été créé pour détecter la fraude et poursuivre les fraudeurs, et a déjà obtenu des condamnations.

La reproductibilité est-elle la solution ?

Mais la fraude n'est qu'une petite partie du problème, on l'a vu. Pour le traiter plus au fond, beaucoup considèrent qu'il faut privilégier l'effort vers une meilleure « reproductibilité » des recherches. L'échec rencontré par une tentative de reproduire un ensemble d'études

3. L'article de Neyman et Pearson sur les tests statistiques qui a fondé une pratique largement répandue date de 1933

fondamentales sur le cancer en 2012 a d'ailleurs beaucoup contribué à la prise de conscience récente (Begley, Ellis 2012). Si toutes les conditions sont remplies pour permettre à d'autres chercheurs de reproduire un travail, celui-ci se doit d'être irréprochable, sans quoi ses lacunes se verront très vite. Le mouvement « pour une recherche reproductible » prend de l'ampleur : de plus en plus de revues demandent aux chercheurs qui leur proposent des articles de fournir aussi leurs jeux de données et leurs programmes. Des logiciels spécifiques sont conçus pour aider les auteurs d'articles à satisfaire ces exigences⁴.

Cela dit, tout le monde n'est pas d'accord sur le concept de reproductibilité, qui ne s'applique pas à toutes les formes de recherche scientifique⁵. Et sa mise en pratique se heurte à de nombreuses difficultés : par exemple, dans certains cas, les données confidentielles ne peuvent être communiquées à des tiers non identifiés, comme le sont les referees (anonymes).

En conclusion

L'enseignement d'une déontologie de la recherche, de plus en plus courante dans les écoles doctorales, permet d'informer les jeunes chercheurs sur l'éthique de leur profession.

Peut-être les incitations qui s'appliquent aux chercheurs en activité seraient-elles à revoir. Si ce sont ces incitations qui engendrent les dérives constatées, elles ne pourraient être contrecarrées que par des réformes de l'organisation même de la recherche, notamment en ce qui concerne le financement et l'évaluation des chercheurs.

Références

- C. Glenn Begley, Lee M. Ellis, 2012 « Drug development: Raise standards for preclinical cancer research » *Nature* n°483
- Ferric C. Fang, R. Grant Steen, Arturo Casadevall, 2012 « Misconduct accounts for the majority of retracted scientific publications » *Proceedings of the National Academy of Sciences - USA* vol 109 n°42
- John P.A. Ioannidis, 2005 « Why most published research findings are false » *PLoS Medicine* vol. 2 n°8
- Valen E. Johnson, 2013 « Revised standards for statistical evidence » *Proceedings of the National Academy of Sciences - USA* n°1313476110
- David Larousserie « Des faussaires dans les labos » *Le Monde Sciences et technologie*, 12 mai 2015
- Jill Nelmark, 2015 « La guerre de la rétractation » *Courrier International* n°1269 du 26 février au 4 mars 2015
- Carmen M. Reinhart and Kenneth S. Rogoff, « From Financial Crash to Debt Crisis », *The American Economic Review*, Vol. 101, No. 5 (Aug. 2011), pp. 1676-1706
- « Unreliable research : Trouble in the lab » *The Economist* 19 octobre 2013
- Compte rendu du Café de la statistique du 3 février 2015 « Comment s'assurer de la validité des arguments statistiques contenus dans les études scientifiques ? ». Sur le site web de la SFdS, ce compte-rendu est accompagné de la vidéo de l'exposé de l'intervenant.
- <http://reproducibleresearch.net/> Site renfermant de nombreux onglets renvoyant eux-mêmes à des informations intéressantes sur la recherche reproductible, la lutte contre le plagiat et la fraude, etc...

4. Voir dans ce même numéro page 35 l'article de Christophe Pouzat, Andrew Davison et Konrad Hinsén

5. En histoire, sciences politiques, sociologie, géographie, la reproductibilité n'a pas de sens et d'autres critères de vérité ont été avancés.