Peut-on faire confiance aux publications scientifiques ?

Claude Ben Slama

Connaissez-vous John P. Ioannidis ? C'est à cause de lui que cet éditorial commence par une question qui peut paraitre particulièrement "déplacée" dans cette revue ! Tout est parti d'un article très célèbre de Ioannidis dans PLOS Medicine : "Why most published research findings are false **?**" (*pourquoi la plupart des resultats scientifiques publiés sont-ils faux?*) [1].

Quand vous saurez que cet article a dépassé les 700 000 téléchargements sur le net (il est l'article le plus téléchargé sur le site du journal PLoS Medicine) vous vous direz que la question n'est peut-être pas si extravaguante qu'il n'y parait.
Et quand vous saurez que Ioannidis est professeur à la Stanford School of Medicine, directeur du Centre de Recherche en Médecine Préventive de l'Université de Rehnborg et qu'il travaille sur les méthodologies des études de recherche, vous vous direz qu'il doit avoir quelques arguments pour faire cette affirmation. Même les statisticiens les plus sceptiques à son égard reconnaissent que beaucoup de résultats d'études scientifiques sont moins "affirmatifs" que les lecteurs ne le croient.

Les « preuves » que les résultats sont souvent faux

Un biais est défini comme la combinaison de facteurs provenant de programmes d'étude et/ou de données et/ou d'analyse qui produisent des résultats qui n'auraient pas dûs l'être.
Un biais ne doit pas être confondu avec le hasard qui est une variabilité spontanée qui aboutit à certaines conclusions fausses par hasard, alors que la conception de l'étude, les données, l'analyse et la présentation sont corrects [1]. Des biais peuvent entraîner une manipulation dans l'analyse des résultats et les effets peuvent paraître différents selon la façon dont ils sont présentés.

Quelques astuces classiques sont utilisées comme celle de présenter un résultat sous forme de risque relatif plutôt que de risque absolu ; le premier est en général plus impressionnant que le second : par exemple, une réduction du risque relatif de 30 % peut correspondre à une réduction du risque absolu de 0,03 % (quand le risque dans le groupe témoin est de 0,1%) !
Les conflits d'intérêts doivent aussi être pris en considération. Pour un même profil d'effets thérapeutiques et d'effets indésirables, les essais parrainés par l'industrie pharmaceutique sont susceptibles d'être présentés comme plus favorables que les essais indépendants.

Les méta-analyses utilisent parfois des données qui ne sont plus à jour : sur 157 méta-analyses publiées en 2005, les deux tiers ne comprenaient pas d'essais publiés après l'an 2000 [5].

Extrapoler les résultats d'un essai clinique est toujours un "acte de foi" qui devient encore plus périlleux pour certains sous groupes de patients. Les critères d'inclusion dans une étude peuvent réduire voire éliminer la présence de patients à très haut risque ou qui ont des comorbidités importantes ou qui prennent certains médicaments "perturbants". Donc, décider si une intervention apparemment bénéfique pour l'ensemble des patients sera aussi efficace sur ces sous groupes, est une généralisation hasardeuse mais largement encouragée par l'industrie pharmaceutique [6].

Les résultats d'une recherche peuvent être contredits par des études ultérieures. Ceci est vrai aussi bien pour des études épidémiologiques traditionnelles que pour la recherche moléculaire la plus moderne [1]. Une évaluation des 49 articles les plus cités sur l'efficacité des interventions médicales entre 1990 et 2004 a montré que, 5 ans plus tard, un quart des essais randomisés et cinq sur six études non randomisées avait déjà été soit contredits soit réévalués "à la baisse" [7].

Les revues prestigieuses attirent la présentation des résultats les “sensationnels”. Si les éditeurs privilégient la publication des résultats significatifs, cela peut entraîner des biais de sélection complexes. Les grandes revues peuvent ne pas publier de petits essais, sauf s'ils sont les premiers publiés sur une question importante ou s’ils montrent des résultats avec des implications majeures. Quant aux essais “négatifs”, ils sont moins susceptibles d'être publiés dans de grandes revues et ils peuvent être publiés avec retard voire même pas du tout ! [3].

Il y a des “facteurs de risque” pour des résultats faux

Ionnadis identifie 6 facteurs de risque majeurs pour aboutir à des résultats faux [1] :

- les études sont de petite taille
- l’effet produit est faible
- le nombre des hypothèses testées est grand et la sélectivité des relations testées est faible
- la flexibilité dans la conception, les résultats et les modes d'analyse est grande
- les intérêts financiers (ou autres intérêts) sont importants
- le thème de recherche est très sensible (avec de nombreuses équipes travaillant sur ce thème)

Un petit “p” qui ne grandit pas toujours la recherche !

Le taux élevé de non replication (absence de confirmation) des résultats d’étude de recherche est une conséquence de la pratique d’évaluer par la signification statistique d’une valeur de p inférieure à 0.05. Selon Ionnadis, la valeur de p n'est pas la façon la plus appropriée de présenter le résultat d’une recherche. Malheureusement, il y a une idée très répandue que les résultats doivent être interprétés uniquement en fonction des valeurs de p et la valeur à 0.05 se charge d’un poids quasi mythique [1] !

Etant donné le prestige des études randomisées, il est inévitable que, parfois, les essais “positifs” constituent un obstacle à la conduite d'essais supplémentaires. Si leurs résultats ne sont pas corrects, nous nous heurtons à la situation où une fausse conclusion empêche de redoubler d'efforts pour la contredire. Ainsi, discerner le vrai du faux n'est manifestement pas chose facile [6].

Quand on analyse les articles les plus cités au cours de la dernière décennie dans 21 domaines scientifiques, on constate que 6 revues rassemblent 53 à 94 des 100 articles les plus cités dans leur domaine. Ainsi, malgré une pléthore de revues disponibles, les documents les plus influents sont très concentrés dans quelques revues [2].
La rareté des revues disponibles est artificielle, basée sur les coûts d'impression élevés (alors que nous sommes à l’ère du numérique) et basée aussi sur la conviction (fausse) que sélectivité est synonyme de qualité.
Les facteurs d'impact sont largement adoptés comme critères de réussite. Le "jeu" des facteurs d'impact est explicite : les éditeurs font des estimations concernant le nombre probable de citations pour les articles soumis à leur évaluation. Le jeu de la citation a créé des relations hiérarchiques distinctes parmi les revues dans différents domaines [7].

Un effet placebo qui est loin d’être neutre

Les nouvelles interventions thérapeutiques peuvent être comparées à un placebo ou même à l’absence de traitement, alors que la vraie question est de savoir comment elles se comportent par rapport aux autres traitements déjà disponibles et reconnus comme efficaces.
Certains comparateurs actifs peuvent avoir été sélectionnés alors qu’ils avaient une efficacité reconnue faible, de manière à démontrer un bénéfice supérieur pour le nouveau traitement.
Des comparaisons entre des agents efficaces “head to head” peuvent être évitées tandis que dans d'autres situations, ces comparaisons “head to head” peuvent être utiles pour la promotion d’un médicament qui n'a pas démontré son efficacité contre placebo.
A l'opposé, certains essais peuvent être sur dimensionnés (“over powered”) pour détecter des avantages statistiquement significatifs mais qui n’ont en fait aucune importance clinique [6].

Comment s'en sortir quand on est simple endocrinologue ?

Devant un essai randomisé, Ionnadis [6] recommande de se poser quelques questions sur sa crédibilité et sa pertinence :
1. L'essai a-t-il une puissance adéquate pour un résultat cliniquement pertinent ?
2. Y a-t-il un biais dans l'étude portant sur sa conception, sa réalisation ou son analyse ?
4. L'étude randomisée est-elle pertinente pour répondre à la question posée ?
5. les résultats obtenus sont-ils vraiment généralisables ?
6. L'interprétation qui est faite des résultats est-elle vraiment appropriée ?

La sélection de la réalisation d'études à grande échelle sur la base de critères tels que la promotion d'un médicament spécifique, est du gaspillage pour la recherche. De plus, de très grandes études sont plus susceptibles de trouver des différences statistiques significatives pour un effet qui ne diffère pas vraiment en réalité de l'hypothèse nulle [1].

Certes, on peut accepter le système de publication actuel comme un aboutissement idéal mais l'environnement global a changé au fil du temps. La publication numérique peut soulager les efforts inutiles de soumissions répétées. On a aussi suggéré de publier préferentiellement des résultats négatifs plutôt que positifs (comme étant plus susceptibles d'être vrais) [1]**.**

Conclusion

Alors ? "Evidence based medicine"… hallucination collective… voire même formatage intellectuel?

La confiance dans le secteur de la recherche risque d'être compromise quand nous prétendons que les découvertes sont plus sûres qu'elles ne le sont vraiment [1]. Nous avons besoin de réplications et pas seulement de découvertes.

Le grand problème est qu'il est impossible de savoir avec 100% de certitude quelle est la vérité dans une question de recherche. Le vrai "gold standard" est inaccessible. Cependant, il y a des moyens pour s'en approcher et éviter les pièges; simplement, il faut le vouloir....

donc à suivre…

Bibliographie

1. *Why Most Published Research Findings Are False. Ioannidis JPA. PLoS Med 2005, 2(8): e124. doi:10.1371/journal.pmed.0020124*
2. *Concentration of the most-cited papers in the scientific literature: analysis of journal ecosystems. Ioannidis JP. PLoS One. 2006 Dec 20;1:e5.*
3. *Magnitude of effects in clinical trials published in high impact general medical journals. Siontis K, Evangelou E, Ioannidis J. International Journal of Epidemiology 2011;40:1280–1291*
4. *Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. Saquib N1, Saquib J, Ioannidis JP. BMJ. 2013;347:f4313. doi: 0.1136/bmj.f4313.*
5. *The use of older studies in meta-analyses of medical interventions: a survey. Patsopoulos NA, Ioannidis JP. Open Med. 2009;3(2):e62-8.*
6. *Some main problems eroding the credibility and relevance of randomized trials. Ioannidis JPA. Bulletin of the NYU Hospital for Joint Diseases 2008;66(2):135-9*
7. *Why current publication practices may distort science. Young NS, Ioannidis JP, Al-Ubaydli O. PLoS Med. 2008 Oct 7;5(10):e201. doi: 10.1371/journal.pmed.0050201.*