

Crédibilité et inflation de l'efficacité dans les méta-analyses

P. Chevalier

NST et RAR

Nous avons déjà insisté dans la revue *Minerva*, sur la nécessité d'une interprétation prudente des chiffres donnés pour un Nombre de Sujets à Traiter¹. Le NST (Number Needed to Treat NNT en anglais) est le nombre de personnes à traiter pendant une période déterminée (celle de l'étude) pour guérir ou pour prévenir un cas supplémentaire de la pathologie considérée. Comme nous l'avons souligné, des NST issus de différentes études ne peuvent être valablement comparés que s'ils évaluent les mêmes traitements, avec les mêmes comparateurs, sur le même critère, dans des populations au même stade de la pathologie (avec le même risque initial) et suivies pendant la même durée. Les NST mentionnés dans les méta-analyses doivent être analysés avec la même prudence et la comparaison des NST entre différentes méta-analyses est, pour les raisons énumérées ci-dessus, très aléatoire².

Le NST est calculé à partir de la modification absolue de risque (souvent Réduction Absolue de Risque, RAR) pour le critère d'évaluation choisi, RAR qui reflète bien l'ampleur de l'efficacité du traitement. Cette ampleur (taille) de l'effet d'un traitement donnée dans une méta-analyse est-elle fiable ?

Inflation de l'ampleur d'effet et crédibilité

Deux auteurs ont récemment exploré la crédibilité et la taille de l'effet rapportée dans des méta-analyses de la Cochrane Collaboration³. Ils ont sélectionné 461 méta-analyses/synthèses méthodiques de la Cochrane disponibles en 2005, basée chacune sur au moins 4 études, avec un critère de jugement binaire et donnant un Odds Ratio (OR) statistiquement significatif ($p < 0,05$) pour le résultat. Ils ont retrouvé en 2010 dans cette même base de données Cochrane Database of Systematic Reviews, 80 des 461 méta-analyses initiales avec une mise à jour.

Sur ce matériel, les auteurs ont analysé la crédibilité des 461 méta-analyses initiales, l'évolution de l'ampleur d'effet rapportée et l'évolution de la crédibilité en 2010.

La crédibilité est un concept issu de la statistique bayésienne des probabilités conditionnelles ou hypothèses (assumptions). La crédibilité est la probabilité a posteriori de résultats vrais positifs, reposant sur le calcul d'un facteur bayésien basé sur les preuves fournies par la méta-analyse, calcul fort complexe. Selon cette analyse, 63 à 84% des méta-analyses initiales présenteraient, selon différentes hypothèses choisies, des résultats vraiment positifs mais les 16 à 37% d'autres ont des résultats faussement positifs.

Pour l'ampleur d'effet, une diminution significative est observée lors de la mise à jour : la modification médiane de l'OR est $\times 0,85$ (IQR de 0,66 à 1,06) pour la mise à jour 2010 versus versions de 2005. La modification d'ampleur d'effet est inversement corrélée avec le poids de la méta-analyse, le nombre

total d'études incluses, le nombre de patients et le nombre d'événements rapportés. La modification médiane d'OR est moins importante ($\times 0,88$) dans les 40 méta-analyses avec le poids le plus élevé que dans les 40 avec poids plus faible ($\times 0,65$). Ceci est une nouvelle illustration du phénomène bien connu de l'inflation de la taille de l'effet dans des petits échantillons. Les auteurs expliquent cette observation, entre autres, par la « malédiction du gagnant » : les résultats sont choisis du fait qu'ils sont supérieurs à un certain seuil et en même temps une évaluation de l'ampleur de l'effet est faite, avec la conséquence mathématique d'une inflation des résultats, en moyenne⁴. D'autres biais peuvent être impliqués : des résultats souvent plus prometteurs initialement que par la suite, des études initiales également moins robustes au point de vue méthodologique, des populations plus à risque dans les études initiales avec de meilleurs résultats, des résultats initiaux encourageants invitant à réaliser des études plus importantes ... qui ne montrent pas d'effet. C'est pourquoi les auteurs plaident pour des méta-analyses ajoutant séquentiellement les résultats d'études en fonction de leur parution. L'intérêt de telles méta-analyses séquentielles (ou cumulatives) avait déjà bien été illustré par Juni et coll.⁵ : les risques cardiovasculaires liés au rofécoxib apparaissaient nettement dans une méta-analyse de ce type longtemps avant leur reconnaissance entraînant le retrait du marché (*voir graphiques sur le site web*).

Pour l'évolution de la crédibilité, le facteur bayésien suggère, dans cet article de Pereira³, une meilleure crédibilité pour 56 des 80 méta-analyses mises à jour.

Des techniques d'analyse bayésiennes montrent que les méta-analyses peuvent présenter des résultats faussement positifs et/ou avec une inflation de la taille de l'effet, particulièrement quand elles sont de petite taille. Une mise à jour régulière est d'autant plus nécessaire.

Références

1. Chevalier P. Nombre de sujets à traiter. *MinervaF* 2009;8(2):24.
2. Chevalier P. Résultats différents de méta-analyses. *MinervaF* 2009;8(1):12.
3. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol* 2011;64:1060-9.
4. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640-8.
5. Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofécoxib: cumulative meta-analysis. *Lancet* 2004;364:2021-9.