



### Trop beau pour être vrai ? Quand un effet largement positif n'est pas nécessairement conclusif

Gilles Henrard, Département de Médecine Générale, Université de Liège

Alors que la majorité des études randomisées contrôlées qui évaluent des interventions médicales ne démontrent, quand elles sont positives, que des bénéfices légers, certaines accouchent occasionnellement de **larges effets**. Il serait tentant dans ces situations d'adopter l'intervention en pratique et, en ce qui concerne la recherche, de s'épargner de fastidieuses études complémentaires. D'autant qu'il est parfois difficile de justifier éthiquement une étude complémentaire dans ces situations, au nom du **principe d'équipoise** (1), qui statue qu'il faut une « incertitude suffisante » quant à l'efficacité d'une intervention avant de construire une étude clinique comparative dans laquelle une cohorte de malades n'en bénéficiera pas. Ne pas gaspiller des ressources en recherches superflues est éminemment louable (2,3) et la question de définir « à partir de quand » un clinicien peut adopter une intervention est fondamentale.

Un article du BMJ de 2016 (4) a cherché à savoir si des résultats montrant de larges effets étaient en général confirmés ou pas dans des études complémentaires. Les auteurs ont examiné tous les « forest plot » de la Cochrane Database of Systematic Review comprenant au moins une **RCT « princeps »** avec un effet considéré comme très large (c.-à-d. avec un risque relatif  $\leq 0,2$  ou  $\geq 5$ ), concernant la comparaison d'un critère d'évaluation primaire dichotomique avec une différence statistiquement significative ( $p < 0,05$ ) qui n'était ni une analyse de sous-groupe ni une analyse de sensibilité, et qui était suivie d'au moins une autre étude « de validation » considérée comme large (c.-à-d. comprenant au minimum 200 événements et non événements sur les mêmes critères). Seulement 44 « forest plot » répondaient à ces critères sur les 85002 trouvés, soit à peine 0,05%. Les auteurs constatent que ces études princeps sont rarement à faible risque de biais, le plus souvent de petite taille (médiane de 14 événements) et que, dans une petite moitié des cas (43% avec IC à 95% de 29 à 58%), les larges études subséquentes ne démontrent plus d'effets statistiquement significatifs. Même quand l'effet reste statistiquement significatif, il tend quasi systématiquement à diminuer par la suite (43 cas sur 44). Dans cette sélection d'études, des résultats impressionnants sont donc généralement minimisés par des études ultérieures plus larges.

En tant que clinicien et lecteur critique à présent, quelle attitude adopter face à la sortie d'une étude randomisée démontrant un large effet ? En d'autres termes, quelles sont les raisons qui devraient nous pousser à douter de la pertinence clinique de tels résultats ? Gordon Guyatt, dans la remarquable dernière édition du « Users' guides to the medical literature » qu'il coordonne (5), nous rappelle quelques raisons d'être « attentiste » en général face à de bons résultats, à savoir : la présence fréquente de biais, particulièrement de biais de mention des résultats et le risque potentiel de mise en évidence ultérieure d'effets indésirables sérieux mais rares. Il propose d'être particulièrement sceptique face aux résultats d'une étude randomisée avec un large effet (il prend l'exemple lui d'une  $RRR > 50\%$ ), particulièrement si le nombre d'événements est faible (par exemple  $< 100$ ) et à plus forte raison si l'étude a été arrêtée précocement pour résultat positif sans critère préétabli (valeur  $p$  et IC non valides dans ce cas). Pereira et al. avaient déjà montré en 2012 (6) que même issus d'études bien menées, à forte validité interne, des résultats exceptionnellement bons tendaient à être minimisés par la suite en raison d'un effet de « régression vers la moyenne ». Nagendram et al. ajoutent que, dans leur échantillon, il ne semble pas y avoir d'ampleur d'effet suffisamment large pouvant prémunir de conclusions négatives par la suite.

Alors que les conclusions rapportées ci-dessus pousseraient plutôt la communauté scientifique à ne pas arrêter trop tôt les recherches sur une question clinique, le risque inverse, celui de les poursuivre trop longtemps, existe aussi.

Les travaux désormais classiques de Antman et al. (7) dans les années 90, au moyen de méta-analyses cumulatives illustrent bien l'enjeu de les arrêter « suffisamment tôt ». Dans ce type de méta-analyse, chaque nouvelle étude est, a posteriori et successivement en fonction de leur année de publication, ajoutée à une nouvelle sommation. On peut donc suivre rétrospectivement l'évolution de l'effet cumulé des preuves disponibles sur une question *comme si* chaque nouvelle étude avait été sommée aux précédentes au fur et à mesure de leurs sorties. Historiquement, ces travaux soulignaient l'importance en soi de réaliser des méta-analyses afin d'avoir une vue synthétique des preuves disponibles sur une question. Par exemple, dans l'étude d'Antman, un effet statistiquement significatif des thrombolytiques dans l'infarctus aigu du myocarde était atteint dès la 10<sup>ème</sup> étude, dans le milieu des années 70. Or, ce n'est que 20 ans et 60 études plus tard, que les recherches sur cette question clinique ont effectivement pris fin... et que les Guides de Pratique Clinique l'ont adopté ! Si « statistiquement significatif » ne veut pas dire « cliniquement pertinent » et, comme nous l'avons souligné, qu'un délai d'adoption est compréhensible et parfois raisonnable, il est probable qu'une bonne partie des études subséquentes aient été inutiles et que le délai d'adoption ait représenté « un manque à gagner » pour les patients.

La force d'un niveau de preuve pour une intervention n'est pas figée. Elle peut évoluer au fil de l'accumulation des données. Si l'exemple des méta-analyses cumulatives d'Antman nous montrent que nous avons, en tant que soignants, parfois trop tardé à adopter des interventions utiles, l'article de Nagendran (4) nous donne quelques arguments EBM pour tenir une position de « scepticisme ouvert » face aux sirènes de l'innovation, quand elles chantent la musique de larges effets.

## Références

1. Kurzrock R, Stewart DJ. Equipoise abandoned? Randomization and clinical trials. *Ann Oncol* 2013;24:2471-4. DOI: 10.1093/annonc/mdt358
2. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86-9. DOI: 10.1016/S0140-6736(09)60329-9
3. Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomized trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349-51. DOI: 10.1136/bmj.39070.527986.68
4. Nagendran M, Pereira TV, Kiew G, et al. Very large treatment effects in randomised trials as an empirical marker to indicate whether subsequent trials are necessary: meta-epidemiological assessment. *BMJ* 2016; 355:i5432. DOI: 10.1136/bmj.i5432
5. Guyatt GH, Rennie D, Meade MO, Cook DJ. *Users' guides to the medical literature: a manual for evidence-based clinical practice*, 3rd ed. New York: McGraw-Hill, 2015.
6. Pereira TV, Horwitz RJ, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA* 2012;308:1676-84. DOI: 10.1001/jama.2012.13444
7. Antman EM, Lau J, Kupelnick B, et al. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Treatments for myocardial infarction. JAMA* 1992;268:240-8. DOI: 10.1001/jama.1992.03490020088036