

Mesure de l'efficacité d'un traitement : pièges et propositions

A propos des antidépresseurs

Pierre Chevalier, Centre Académique de Médecine Générale, Université Catholique de Louvain
Texte sous la responsabilité de la rédaction francophone

Nous avons récemment abordé dans un éditorial de la revue *Minerva* en 2013¹ les pièges liés aux critères d'évaluation de l'efficacité des antidépresseurs, particulièrement dans les méta-analyses quand il s'agit de sommer des résultats d'études utilisant des critères de jugement différents. Le recours à des échelles différentes, Hamilton Depression Rating Scale (HDRS) et Montgomery-Åsberg depression rating scale (MADRS) par exemple, en est une illustration. Dans ce dernier cas, les auteurs d'une méta-analyse ont recours à des Différences Moyennes Standardisées (DMS). Cohen² avait proposé des valeurs repères de DMS pour pouvoir comparer les différentes tailles d'effet de traitements : 0,20 (peu d'effet), 0,50 (effet modéré) et 0,80 (effet important). Ces valeurs repères ont ensuite été utilisées, probablement à tort, comme seuil de pertinence clinique de l'efficacité observée. Dans leur méta-analyse, Kirsch et coll.³ ont suivi les auteurs du guide de pratique de NICE dans leur choix d'un seuil arbitraire de 0,50 pour la DMS pour déclarer que le traitement se montre efficace. Ce choix d'un seuil d'efficacité clinique pertinente prêtant à discussion, ouvrait le doute sur la fiabilité des conclusions des auteurs.

Turner et Rosenthal⁴ expliquent par une métaphore pourquoi cette utilisation de NICE⁵ et de KIRSCH³ est inadéquate. Imaginons une mesure de l'efficacité des antidépresseurs en termes de litres d'un jus d (appelé ci-après d de Cohen, d étant l'ampleur d'effet mesurée). Quand Turner et coll. mesurent une quantité de jus d de 0,41 litre dans le verre des articles de revues, versus 0,31 litre dans le verre de la FDA, ils peuvent conclure que le verre de la FDA est relativement moins rempli que le leur. Ils reconnaissent cependant que les 0,31 litre représentent une quantité mesurable et significative. Ils mentionnent que quand dans leur méta-analyse Kirsch et coll.³ mesurent 0,32 litre de jus d, ils considèrent que le verre n'est pas suffisamment plein (pour un seuil arbitraire $\geq 0,5$) et concluent que le verre ne contient virtuellement pas de jus. Turner et Rosenthal sont d'accord pour reconnaître que le « verre » antidépresseurs est loin d'être rempli mais ils ne sont pas d'accord pour considérer qu'il est entièrement vide.

Comme explicité dans l'éditorial de *Minerva* précité¹, l'utilisation de l'outil HDRS lui-même peut être remis en question, pour certains de ces aspects. Des auteurs ont montré⁶ que la précision de l'évaluation au score HDRS diminue quand le niveau de sévérité de dépression diminue, donc au fur et à mesure que l'état du patient s'améliore, rendant une comparaison entre patients avec des niveaux de sévérité initiale de dépression différents non valide. Ils concluaient qu'une ampleur faible de l'effet peut être liée à la faible précision de l'outil de mesure et à sa faible sensibilité pour évaluer une modification, particulièrement en cas de dépression de sévérité légère ou modérée.

Nous isolons, grâce aux publications susnommées, deux problèmes dans les synthèses évaluant l'efficacité des antidépresseurs : des échelles de scores et/ou critères de jugement différents, avec une sommation difficile des résultats d'une

part, et d'autre part, un score d'évaluation imparfait pour refléter des différences moyennes dans l'évolution des patients sous traitement. Gibbons et coll.⁷ nous proposent peut-être une solution ... sous conditions ! Ils évaluent l'efficacité à court terme (6 semaines) de 2 antidépresseurs, la fluoxétine et la venlafaxine. Ils ne sélectionnent que les RCTs menées par les firmes fabriquant ces médicaments, ce qui est une limite très importante, mais leur permet de disposer des données individuelles des patients. Les patients sont, selon les études, des adultes, des adolescents ou des personnes âgées. Les chercheurs ne prennent en considération que les études avec la même échelle de score d'évaluation (HDRS pour les adultes et personnes âgées, Children's Depression Rating Scale-Revised (CDRS-R) pour les adolescents). Disposant des données individuelles, ils analysent les résultats avec des méthodes statistiques bayésiennes fort complexes, en régression dans 3 niveaux simultanés (délai de mesure, patient, étude) et montrent ainsi, par exemple pour les adultes et personnes âgées, une modification du score HDRS de - 11,82 sous antidépresseur et de - 9,26 sous placebo soit une estimation du maximum de vraisemblance marginale de - 2,55 (erreur standard de 0,20 ; $p < 0,001$), soit une Augmentation Relative de Risque de 27,7%.

Les auteurs concluent, sur base de ce résultat et d'autres, à l'efficacité de ces 2 antidépresseurs chez les adultes et personnes âgées en termes de réponses et de rémissions versus placebo. Si les différences au point de vue répondeurs (réduction de 50% de la sévérité, NST de 5,41) et de rémissions (score HDRS < 8 , NST de 7,3) semblent favorables, une différence de - 2,5 au score HDRS reste inférieure au seuil de 3 points défini comme la différence cliniquement pertinente (correspondant à une DMS de 0,50) par NICE⁵... La sélection orientée des études dans cette synthèse, la durée d'évaluation sur 6 semaines, la limite à 2 antidépresseurs n'invitent pas à se fier aux conclusions. Le mode d'analyse des données est cependant à retenir pour d'autres synthèses concernant l'efficacité des antidépresseurs, en soulignant l'intérêt de pouvoir disposer de données individuelles.

Références

- Chevalier P. Evaluation d'un traitement et fiabilité des outils de mesure. [Éditorial] *Minerva* 2013;12(1):1.
- Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York: Lawrence Erlbaum Associates, 1988.
- Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5:260-7.
- Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336:516-7.
- NICE. *Depression: Management of depression in primary and secondary care*. Clinical Practice Guideline Number 23. National Institute for Health and Care Excellence. London, 2004.
- Isacsson G, Adler M. Randomized clinical trials underestimate the efficacy of antidepressants in less severe depression. *Acta Psychiatr Scand* 2011;125:453-9.
- Gibbons RD, Hur K, Brown CH, et al. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012;69:572-9.