

Les pièges des analyses en sous-groupes

En tant que cliniciens, nous nous interrogeons surtout sur l'inférence des résultats des études cliniques pour les patients que nous soignons : peuvent-ils espérer des résultats semblables à ceux obtenus par les sujets inclus dans l'étude, pour autant qu'ils aient les mêmes caractéristiques ? Une distinction est-elle également possible au point de vue efficacité pour certains sous-groupes de patients ? Des informations précieuses pourraient ainsi permettre de cibler les personnes pouvant tirer un profit maximal de l'intervention. Dans la plupart des études incluant un nombre important de sujets, les auteurs se livrent donc à des analyses en sous-groupes. Dans l'étude CHARISMA¹, précédemment analysée dans Minerva², aucun bénéfice n'est montré pour un traitement clopidogrel plus aspirine versus aspirine seule chez l'ensemble des patients inclus. En analyse en sous-groupes, une bénéfice serait montré pour les personnes en prévention cardio-vasculaire secondaire mais non pour les personnes présentant des facteurs de risque sans avoir fait d'incident cardio-vasculaire. Dans l'étude GAIT³, également précédemment commentée dans Minerva⁴, un bénéfice en termes de diminution de la douleur ne peut être montré pour l'ensemble des patients souffrant d'arthrose du genou. En analyse par sous-groupes définis en fonction de l'importance de la douleur, un avantage serait montré en cas de douleur modérée à sévère.

Les résultats de ces études peuvent-ils être traduits dans notre pratique quotidienne ? La réalisation et l'interprétation d'analyses en sous-groupes réclament une grande circonspection ; de nombreuses conditions sont liées à ces opérations. Les analyses en sous-groupes doivent, en premier lieu, être prévues dans le protocole initial, ce qui permet d'en tenir compte lors de la randomisation, avec des groupes suffisamment importants en nombre de participants. Le rationnel d'une analyse en sous-groupes doit également être clairement exposé : pourquoi estimer qu'une différence de réponse peut être observée entre différents sous-groupes⁵⁻⁶ ? L'observation, dans l'étude ISIS incluant des sujets post infarctus du myocarde, d'une réponse favorable à l'aspirine réservée aux patients nés sous un signe zodiacal particulier, est

l'illustration classique de l'importance potentielle du hasard. Il est important, ensuite, que les analyses en sous-groupes soient prises en compte dans le calcul de la puissance de l'étude⁶. Comme le font remarquer des éditorialistes ou commentateurs dans les numéros de revues publiant les résultats des études précitées⁷⁻⁹, une analyse en sous-groupes doit toujours être interprétée avec beaucoup de prudence. Pour atteindre une signification statistique, des résultats observés doivent être associés à une valeur $p < 0,05$, indiquant ainsi que le hasard peut expliquer de tels résultats dans moins de 5% des cas. En analyse en sous-groupes, cette valeur p doit être plus petite pour qu'une signification statistique soit reconnue, l'analyse en sous-groupes augmentant le risque que le hasard soit responsable des résultats observés. Par exemple, dans l'étude CHARISMA¹, dans le sous-groupe de patients présentant une pathologie cardio-vasculaire, un risque relatif de survenue d'un nouvel incident cardio-vasculaire grave est de 0,88 (IC à 95% de 0,77 à 0,998 ; $p=0,046$), ce qui peut sembler statistiquement significatif... mais ne l'est cependant pas ! Dans une analyse en sous-groupes, la valeur p doit, en effet, être divisée par le nombre de « tests » d'analyse. Dans cette étude, le nombre de « tests » est de 20. Par nombre de « tests », il faut entendre multiplication du nombre de sous-groupes caractérisés par le nombre de critères d'évaluation analysés^{7,10,11}. Dans l'étude CHARISMA, pour être significative la valeur p doit donc être $< 0,0025$, ce qui n'est pas le cas. Ce résultat en sous-groupe ne peut donc être qualifié de statistiquement significatif.

Dans de nombreux cas, les sous-groupes ne sont déterminés que lors de l'analyse des résultats. Cette analyse dite « post hoc » doit être considérée avec de plus amples réserves encore. Des analyses statistiques ne sont, dans ce cas, pas fiables et leur « signification » éventuelle n'a aucune base¹¹. Des analyses en sous-groupes ne manquent pas d'attrait, peuvent apporter des informations cliniques intéressantes, mais peuvent aussi être trompeuses. Accordons leur la valeur qu'elles peuvent apporter ; considérons les comme d'intéressantes hypothèses.

P. Chevalier et M. van Driel

Références

1. Bhatt DL, Fox KA, Hacke W et al. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med* 2006;354:1706-17.
2. Chevalier P. Pas d'intérêt de rajouter du clopidogrel à l'aspirine en prévention cardiovasculaire? *Minerva F* 2006;5(6):88-91.
3. Clegg DO, Reda DJ, Harris CL et al. Glucosamine, chondroitin sulfate, and the two in combination for painful knee osteoarthritis. *N Engl J Med* 2006;354:795-808.
4. Chevalier P. Glucosamine et/ou chondroïtine pour la gonarthrose ? *MinervaF* 2006;5(10):156-8.
5. Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. *Am Heart J* 2000;139:952-61.
6. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.
7. Lagakos SW. The challenge of subgroup analyses - reporting without distorting. *N Engl J Med* 2006;354:1667-9.
8. Pfeffer MA, Jarcho JA. The charisma of subgroups and the subgroups of CHARISMA. *N Engl J Med* 2006;354:1744-5.
9. Hochberg MC. Nutritional supplements for knee osteoarthritis - still no resolution. *N Engl J Med* 2006;354:858-60.
10. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000;355:1064-9.
11. Yusuf S, Wittes J, Probstfeld J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93-8.