

Astuces pour l'analyse critique d'article scientifique

Comment analyser les résultats de l'analyse en sous-groupes d'une étude ?



T. PEZEL

Service de Cardiologie, CHU Lariboisière, PARIS ;
Unité Inserm-UMR 942, PARIS.

Les analyses en sous-groupes deviennent de plus en plus fréquentes dans les études cliniques, mais comment doit-on les analyser en pratique ? En effet, on commence à connaître le principe : lorsqu'une étude est négative, les auteurs nous proposent toujours des résultats d'analyse en sous-groupes avec un résultat qui devient significatif dans certains sous-groupes d'intérêt clinique de l'étude, comme très récemment à l'ESC 2019 avec l'étude PARAGON-HF. Mais alors que peut-on en retenir ? Quelles sont les limites de ce type de résultat ?

pas seulement dans un sous-groupe particulier qui expliquerait que le résultat global soit statistiquement significatif.

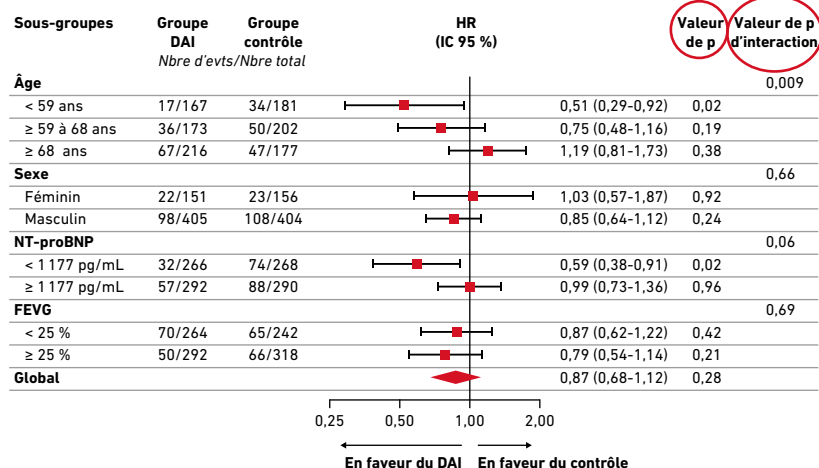
Dans ce but, les auteurs utilisent souvent un **test d'homogénéité sur une analyse en sous-groupes** afin de tester l'interaction

Principe de l'analyse en sous-groupes

Dans les études thérapeutiques, l'analyse en sous-groupes consiste à **comparer l'efficacité de la nouvelle intervention dans une partie limitée de l'effectif total** (par exemple, seulement chez les patients en FA). Elle permet ainsi de mettre en évidence des résultats qui se rapportent à cette sous-population d'intérêt clinique sélectionnée par les auteurs.

Intérêt de l'analyse en sous-groupes

L'intérêt d'avoir accès aux résultats d'analyse en sous-groupes est avant tout de **pouvoir s'assurer de l'homogénéité du résultat dans les différents sous-groupes**. En effet, plus le résultat sur le critère de jugement est homogène dans les différents sous-groupes évalués (âge, sexe, comorbidités...), plus cela traduit un effet clinique global chez l'ensemble des patients de la population étudiée, et



Analyse détaillée de cette figure :

- sur l'axe des abscisses : on note l'effet du traitement (rapport de cotes pour le décès ou l'invalidité) ;
- sur l'axe des ordonnées : on place les sous-groupes selon une échelle arbitraire ;
- sur chaque ligne = l'effet du traitement observé dans un sous-groupe est représenté par un symbole (rond ou carré) entouré d'un trait horizontal qui représente l'intervalle de confiance associé ;
- en regard de chaque sous-groupe :
 - on fait figurer les statistiques résumées avec le nombre d'événements et effectifs pour chacun des 2 groupes de traitement (**à gauche de la figure**), ainsi que l'effet du traitement (HR) avec son intervalle de confiance et son degré de significativité p (**à droite de la figure**) ;
 - on rapporte la valeur de p d'interaction associée à chaque test d'homogénéité réalisé (**tout à droite de la figure**).

Encadré 1 : Diagramme de Forest d'analyse en sous-groupes d'une étude thérapeutique évaluant l'efficacité du DAI (défibrillateur automatique implantable) sur cardiopathie non ischémique (étude DANISH, NEJM 2016).

entre telle ou telle caractéristique définissant un sous-groupe (p. ex : patients en FA, patients diabétiques, patients insuffisants rénaux) et le résultat sur le critère de jugement principal de l'étude. Nous pouvons retenir que si **p d'interaction < alpha pour le test d'homogénéité**, alors on peut évoquer que le résultat sur le critère de jugement principal est différent d'un sous-groupe à l'autre (**encadré 1**).

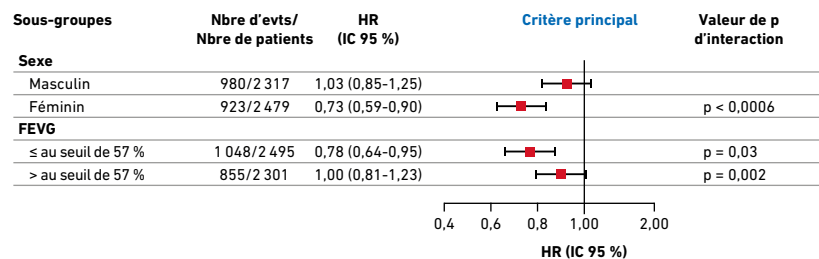
Pourquoi ne peut-on pas conclure sur les résultats d'une analyse en sous-groupes ?

Il est important de noter qu'il existe systématiquement un **défaut de la puissance a priori** en raison du faible effectif du sous-groupe. En effet, le nombre de sujets nécessaires (NSN) a été calculé pour obtenir une puissance *a priori* satisfaisante sur le critère de jugement principal avec l'ensemble de l'effectif de la population étudiée, et non pas sur une partie de l'effectif. Ainsi, en cas de résultat statistiquement significatif sur un sous-groupe, il faudra refaire une étude spécifique dans le sous-groupe considéré.

Attention, nous devons rester vigilants face à certains essais thérapeutiques qui l'utilisent à tort, pour pallier l'absence de significativité du critère de jugement principal car il est souvent aisé de trouver *a posteriori* un sous-groupe particulier où la différence est significative. En effet, la conclusion sur un sous-groupe peut être fausse car simplement liée au hasard, si elle n'a pas été prévue *a priori* (**encadré 2**).

Quel est l'intérêt d'utiliser une randomisation stratifiée pour réaliser ensuite une analyse en sous-groupes ?

Pour rappel, la **randomisation stratifiée** correspond à un cas particulier de randomisation qui consiste à forcer une répartition équilibrée du ou de plusieurs facteur(s) de stratification entre les



Lors du congrès de l'ESC 2019, les auteurs de cette étude ont insisté sur l'importance des résultats significatifs de l'analyse de 2 sous-groupes, avec une efficacité significative du sacubitril/valsartan comparé au valsartan pour les patients **avec FEVG entre 45 et 57 %** et pour les **femmes** de l'étude (car les intervalles de confiance ne comprennent pas 1 — voir diagramme de Forest ci-dessus).

→ Mais attention, il est important de se rappeler que **l'on ne peut pas conclure sur les résultats d'une analyse en sous-groupe réalisée a posteriori**, surtout lorsque le résultat principal de l'étude est négatif. Et cela, même si la tentation est grande de suggérer une efficacité du sacubitril/valsartan chez les patients avec une FEVG < 57 %, et donc une potentielle autorisation de mise sur le marché comme certains pourraient le proposer. Restons vigilants sur ce type de résultats en sous-groupes analysés *a posteriori* avec des seuils choisis *a posteriori* pour aller "chercher la significativité du résultat" (ici le seuil de 57 %).

→ Concernant le **résultat significatif dans le groupe des femmes de cette étude**, différentes hypothèses peuvent être évoquées, et notamment celle de l'amylose cardiaque. En effet, comme l'a rappelé le Pr Thibaud Damy pendant le congrès, les données de la littérature suggèrent que 13 % des IC à FEVG préservée seraient des amyloses cardiaques (essentiellement TTR). Or, cette pathologie touche plus volontiers les hommes que les femmes.

Encadré 2 : Analyse en sous-groupes de l'étude négative évaluant l'efficacité du sacubitril/valsartan versus placebo chez des insuffisants cardiaque à FEVG préservée (étude PARAGON-HF, NEJM, 2019).

2 groupes afin de limiter le risque de biais de confusion dans l'analyse. Ainsi, il est important de savoir que la **randomisation stratifiée permet d'augmenter la validité interne (fiabilité) des analyses en sous-groupes**. En effet, en réalisant une randomisation stratifiée *a priori* sur les sous-groupes les plus importants pour notre étude, on augmente la fiabilité des résultats de l'analyse en sous-groupes qui sera réalisée *a posteriori* (car en stratifiant *a priori* sur ces critères, on prouve que le choix d'analyser le critère de jugement principal dans ces sous-groupes était prévu avant même que l'étude ne commence, et donc que l'on n'a pas choisi les sous-groupes significatifs *a posteriori*, pour nous arranger !). Attention, cela ne signifie pas pour autant que les auteurs ont le droit d'en tirer une conclusion : **on ne peut jamais conclure sur une analyse en sous-groupes !**

Conclusion

D'un point de vue méthodologique, il est important de retenir que l'on n'est pas autorisé à conclure sur les résultats des analyses en sous-groupes, *a fortiori* lorsque le résultat principal de l'étude est négatif.

Les deux rôles essentiels des analyses en sous-groupes sont donc :

- à **visée complémentaire** : pour venir densifier le message porté par le résultat principal de l'étude, lorsque celui-ci est bien significatif, permettant de s'assurer de l'homogénéité du résultat dans les différents sous-groupes ;
- à **visée exploratoire** : afin de guider le *design* de futures études au sein d'un sous-groupe d'intérêt (l'objectif sera alors de confirmer le résultat évoqué en sous-groupe).