

EBM : "p" et intervalle de confiance. Que nous apportent-ils ?

Mots clés : *Bibliomed* analyse chaque semaine les informations issues des essais contrôlés. Ceux-ci ne nous apportent jamais des certitudes, mais des probabilités plus ou moins grandes concernant l'efficacité de tel traitement ou stratégie. Comment alors apprécier la validité et l'utilité de ces résultats ? A côté des données tenant à la structure même de l'essai (randomisation, puissance, biais...), des calculs statistiques apportent des éléments indispensables à l'analyse. Parmi ceux-ci, le *p* et l'*intervalle de confiance* (IC), reviennent constamment. Plusieurs synthèses récentes nous permettent de « rafraîchir » les idées à ce sujet^{1,2,3,4}.

La valeur de « p »

Elle traduit la significativité statistique, c'est à dire la probabilité que le résultat ne soit pas du au hasard. Le seuil habituellement retenu ($p < 0,05$) signifie qu'il y a 95 chances sur 100 que l'hypothèse soit correcte, 5 sur 100 que le résultat soit dû au simple hasard. Il est fixé arbitrairement, comme le risque maximal que les auteurs sont prêts à prendre pour conclure. La valeur de *p* n'exprime donc pas une **quantité quelconque, mais la puissance statistique de la preuve** vis à vis d'une éventuelle absence d'effet (hypothèse nulle). Par ailleurs, $p < 0,05$ indique que la différence est **statistiquement** significative, mais ne permet pas de dire qu'elle est **cliniquement** significative. La valeur de *p* ne prend pas en compte l'ampleur de l'effet, qui seule permet d'apprécier la pertinence clinique.

L'intervalle de confiance (IC)

L'IC, à la différence du *p*, est calculé à partir de la valeur de l'effet observé dans l'essai et de ses variations possibles si l'étude était faite sur des échantillons différents de patients. Ses marges supérieures et inférieures sont obtenues par un calcul statistique. Calculé avec un seuil arbitraire de 95%, l'IC ajoute donc au résultat ponctuel de l'essai **la fourchette des valeurs à l'intérieur de laquelle nous sommes sûrs à 95% de trouver la valeur réelle.**

L'IC est présenté avec l'association au résultat de l'essai de deux chiffres correspondant aux *marges supérieures et inférieures calculées*. Ainsi, une réduction relative du risque (RRR) de mortalité de 20% observée dans un essai, l'IC sera présentée ainsi :

$$RRR = -20\% , IC_{95\%} = (-35\% ; -5\%)$$

Cela signifie que si l'essai a observé une baisse ponctuelle de 20% de mortalité, il n'est pas possible d'exclure que l'efficacité du traitement soit plus petite (au pire -5%) ou plus grande (au mieux -35%).

Ainsi l'IC **quantifie l'effet étudié et le degré de certitude de cette estimation**. Il permet alors d'en apprécier la pertinence clinique^{2,3,4}. Encore faut-il que le critère étudié dans l'essai soit pertinent pour la pratique.

Que retenir en pratique ? Surtout que la « vérité » clinique n'est jamais binaire – oui ou non – comme pourrait le laisser penser un *p* significatif ou non. *L'intérêt essentiel de l'IC est que l'information qu'il apporte est plus nuancée que celle donnée par un p par définition binaire.* Nous en verrons des exemples concrets au fil des *Bibliomed* à venir.

Associer en pratique p et IC

La largeur de l'intervalle de l'IC, et l'étude de ses bornes sont les éléments utiles pour apprécier l'intérêt réel d'une étude, en complément du *p*.

- **L'IC est d'autant plus large que l'échantillon est petit.** Dans une étude sur 24 tests diagnostiques de l'infection à *H pylori*², la sensibilité du test était élevée, s'exprimant : **95,8% (IC 95% : 75% à 100%)** donc avec une incertitude considérable. Un échantillonnage de 240 sujets avec le même résultat en réduirait beaucoup l'incertitude : **95,8% (IC 95% : 92,5 à 98%)**.

- **les deux bornes de l'IC sont en faveur d'un résultat positif : c'est le cas des résultats statistiquement significatifs ($p < 0,05$).** Mais si la borne « péjorative » de l'IC est très proche de la valeur 0, le résultat peut ne pas être pertinent pour la pratique. **Ce n'est que si ce minimum est « cliniquement significatif » que l'essai peut intéresser le clinicien**³. Ainsi dans une étude de réduction du risque aux résultats suivants : $p = 0,041$, $RRR = -25\%$ (IC 95% [-42% ; -1%]), on ne peut exclure un effet presque nul de 1%, dont il faut apprécier la pertinence clinique ; et la largeur de l'IC rend incertain le résultat malgré sa significativité ($p = 0,041$). Des essais complémentaires ou une méta-analyse sont nécessaires.

- **la borne « péjorative » de l'IC englobe la valeur d'un effet nul. C'est le cas des essais non significatifs ($p > 0,05$).** L'IC apporte des données complémentaires pour l'interprétation.

Exemple 1 : $RRR = -0,5\%$, IC 95% (-4% ; +4%). L'effet observé très proche de l'effet nul, l'IC étroit donc un essai puissant, permettent de conclure que l'intervention testée n'a pas d'intérêt (au mieux 4%).

Exemple 2 : $RRR = -20\%$, IC 95% (-49% ; +17%). L'effet observé est important (20%, peut-être jusqu'à 49%) ; l'IC large est situé en majeure partie du côté favorable : il est possible, malgré la non significativité, que le traitement soit suffisamment efficace pour être intéressant en pratique. Cela doit inciter à mener un nouvel essai de plus grande puissance.

Références :

- 1 - Montori VM et al for the Evidence-Base Medicine Teaching Tips Working Group. : measures of precision (confidence intervals). *CMAJ* 2004 ; 171 : 611-5.
- 2 - Altman DG. Intervalles de confiance : indication du degré de certitude des résultats de recherche. *EBM Journal* (éd française) 1996 ; 4 : 5-6.
- 3 - Boissel JP, Gueyffier F. Principes de l'essai clinique et de la méta-analyse. Accessible sur <http://cri-clrs-wnts.univ-lyon1.fr/Polycopies/Cardiologie/Cardiologie-109.html>
- 4 - Cuherat M. et al : Lecture critique et interprétation des résultats des essais cliniques pour la pratique médicale. Paris : Flammarion ; 2004.