

Bulletin d'information trimestriel sur l'actualité du **Centre de Recherche Clinique (CRC)**

Thème

Les essais de non-infériorité et d'équivalence, le nouveau design « à la mode » ?

Dre Angèle Gayet-Ageron

En recherche clinique, l'essai randomisé contrôlé représente le « gold standard » des designs d'étude pour évaluer un nouveau programme de santé et pour répondre à un jugement de causalité. Au milieu du XX^{ème} siècle, la pharmacopée étant encore assez limitée, les essais cliniques randomisés avaient pour objectif de montrer la supériorité d'un nouveau traitement comparé à un placebo puis plus tard comparé aux traitements de référence de l'époque. Depuis une dizaine d'années, du fait notamment des nombreux traitements disponibles dans des domaines médicaux variés, les essais de supériorité ont été remplacés par les essais d'équivalence et surtout de non-infériorité.

Qu'est-ce qui distingue ces essais d'équivalence et de non-infériorité de l'essai de supériorité ?

Dans les essais de supériorité, le but est de montrer que le nouveau traitement est supérieur au placebo (ou au traitement standard) d'au minimum une différence fixée *a priori*, nommée « différence cliniquement pertinente » ou Δ (hypothèse alterne). Si le critère de jugement étudié est par exemple le pourcentage de sujets avec une charge virale indétectable du virus de l'immunodéficience humaine (VIH), il faudra qu'au moment de la conception de l'étude les chercheurs responsables décident à quelle différence minimale Δ ils jugeront le nouveau traitement comme cliniquement supérieur au traitement de comparaison. Ils doivent par ailleurs également choisir le pourcentage de charge virale indétectable de base attendu avec le traitement de référence.

Dans les essais d'équivalence, le but sera de montrer que le nouveau traitement ne diffère pas plus d'une certaine valeur fixée *a priori*, appelée « marge d'équivalence » ou ϵ . Le « vrai » effet du nouveau traitement devrait être

compris entre $-\epsilon$ et $+\epsilon$. En contrepartie ce nouveau traitement présentera certains avantages implicites, non testés par l'essai clinique, comparé à la référence (coût, effets secondaires, galénique, etc.). Ce type d'essai est rarement utilisé en recherche clinique.

En revanche les essais de non-infériorité sont plus souvent utilisés et vont chercher à montrer que le nouveau traitement n'est pas inférieur à l'ancien de plus que la marge d'équivalence (hypothèse alterne), appelée dès lors « marge de non-infériorité » ou θ . Dans un essai de non-infériorité, la démonstration d'une supériorité du nouveau traitement serait un bénéfice inattendu, ce qui n'est pas gênant ; en revanche toute évidence d'infériorité conduirait au rejet du nouveau traitement. La marge de non-infériorité θ doit être plus petite que la différence cliniquement pertinente utilisée dans les essais de supériorité sinon on considérerait comme équivalents des traitements qui sont en fait différents.

Finalement dans les essais d'équivalence ou de non-infériorité, les hypothèses nulle et alterne sont inverses à celles des essais de supériorité : on souhaite montrer que le nouveau traitement n'est pas différent de l'ancien au-delà d'une certaine valeur (différence maximale tolérée).

Comment interpréter de manière correcte les résultats d'un essai d'équivalence ou de non-infériorité ?

Une fois l'essai terminé, la différence d'effet observée dans la population d'étude devra être comparée aux hypothèses de recherche afin de conclure si le nouveau traitement n'est pas moins efficace que l'ancien dans les limites fixées *a priori* et qu'il pourra donc faire partie de l'arsenal thérapeutique de la maladie étudiée. Il s'agira donc de rapporter la différence observée assortie de son intervalle de confiance à 95% (IC95%). Il existe plusieurs méthodes statistiques pour calculer l'IC95% qu'il faudra choisir au moment de la conception de l'étude.

Prenons comme critère de jugement la guérison. Si les résultats de l'étude montrent que la borne inférieure de l'IC95% autour de la différence observée est plus grande que la marge de non-infériorité fixée *a priori* (IC95% inférieure $> -\theta$) alors le nouveau traitement sera jugé non inférieur à l'ancien. Si par contre la

CRC INFO

borne inférieure de l'IC95% chevauche la marge de non-infériorité fixée *a priori* (IC95% inférieure $< -\theta$) alors il n'y aura pas d'évidence que ce nouveau traitement n'est pas moins efficace que l'ancien et il ne représentera pas une alternative thérapeutique recommandable.

La figure ci-dessous présente différents résultats possibles d'un essai de supériorité et d'un essai de non-infériorité ainsi que leurs interprétations. L'ancien traitement sera jugé meilleur que le nouveau (respectivement le nouveau sera jugé inférieur à l'ancien dans les essais de non-infériorité) si l'IC95% de la différence d'efficacité est totalement situé en-dessous de la valeur « 0 ».

Comment choisir la marge d'équivalence ou de non-infériorité?

Autant le choix de l'erreur de type 1 ou α (probabilité de rejeter l'hypothèse nulle lorsque celle-ci est vraie) en général fixée à 5% et plus rarement 1%, ou de l'erreur de type 2 ou β (probabilité de ne pas rejeter l'hypothèse nulle lorsque l'hypothèse alterne est vraie) fixée au maximum à 20%, est plus codifié, autant celui de la marge d'équivalence est plus obscure et n'a pas fait l'objet d'un consensus.

En 1988, Jacob Cohen a proposé que pour une variable continue une différence entre 2 moyennes de 0.2 écart-type peut être considérée comme petite, de 0.5 écart-type comme moyenne et de 0.8 écart-type ou plus comme grande. Dans un essai de non-infériorité la marge de non-infériorité sera préférentiellement petite, soit 0.2 écart-type. D'autres disent que la marge d'équivalence doit avoir une étendue deux à trois fois plus petite que la différence cliniquement pertinente Δ . Finalement en l'absence d'un réel consensus, le choix de cette marge devrait être lié au domaine de recherche étudié, au type de critère de jugement évalué (mortalité, guérison, rechute, etc.) et au niveau de risque de base. Ceci peut difficilement faire l'objet d'un consensus général même si l'étude des pratiques en recherche clinique montre que le choix de cette marge est rarement argumenté.

Mettons-nous à la place d'un chercheur dans le domaine du VIH qui souhaite montrer qu'un nouveau traitement antirétroviral n'est pas inférieur à un ancien concernant le pourcentage d'échappement virologique mais que celui-ci présente l'avantage d'une prise unique favorisant ainsi l'adhérence thérapeutique et évitant ainsi

l'échec virologique. Si avec l'ancien traitement antirétroviral, le pourcentage d'échappement virologique est de 10%, quel pourcentage d'échec virologique sera jugé acceptable pour conclure que le nouveau traitement antirétroviral n'est pas moins efficace que l'ancien ? Est-ce que 20% d'échec virologique avec le nouveau traitement (soit une différence absolue de 10%, ou un risque relatif de 2), est encore acceptable pour porter une telle conclusion ou au contraire est-ce qu'une différence de 2%, soit 12% avec le nouveau traitement, paraît beaucoup plus raisonnable ? Le choix de la différence clinique acceptable est finalement assez arbitraire et nécessite d'être argumenté pour être recevable. Car finalement ce qui est le plus « manipulable » par les chercheurs, c'est bien le choix de cette marge d'équivalence, ou plus généralement de la différence cliniquement pertinente, qui conditionne directement la taille d'échantillon et donc la faisabilité de l'étude. Ce processus a été joliment dénommé « Sample size samba » par Schulz et Grimes.

En conclusion, il faut rappeler l'existence des recommandations CONSORT adaptées en 2006 aux essais d'équivalence et de non-infériorité qui restent une « référence » pour la présentation des résultats de ce type d'études (<http://www.consort-statement.org/extensions/designs/non-inferiority-and-equivalence-trials/>)

Figure. Interprétation de différents résultats d'essais de supériorité (haut) et de non-infériorité (bas).

