

## Essais de non-infériorité et d'équivalence : points clés de méthodologie

Marie Herr, Alexis Descatha, Philippe Aegerter

### ► To cite this version:

Marie Herr, Alexis Descatha, Philippe Aegerter. Essais de non-infériorité et d'équivalence : points clés de méthodologie. La Revue De Médecine Interne / Fondée Par La Société Nationale Française De Médecine Interne, Publisher : Paris : Baillière [1980-. Latest Publisher : Paris : Elsevier, 2017, Epub ahead of print. <10.1016/j.revmed.2017.06.010>. <inserm-01576223>

**HAL Id: inserm-01576223**

**<http://www.hal.inserm.fr/inserm-01576223>**

Submitted on 22 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Essais de non-infériorité et d'équivalence : points clés de méthodologie

### Non-inferiority and equivalence trials: methodological key points

Marie HERR<sup>1,2,3\*</sup>, Alexis DESCATHA<sup>1,2,4,5</sup>, Philippe AEGERTER<sup>1,2,3,6</sup>

<sup>1</sup> INSERM, U1168, VIMA : Vieillesse et Maladies chroniques : approches épidémiologique et de santé publique, Villejuif, France

<sup>2</sup> Univ Versailles St-Quentin-en-Yvelines, UMR-S 1168, Montigny le Bretonneux, France

<sup>3</sup> Département Hospitalier d'Epidémiologie et de Santé Publique, Hôpitaux Universitaires de Paris Ile-de-France Ouest, AP-HP, France

<sup>4</sup> INSERM, UMS011, Cohortes épidémiologiques en population, Villejuif, France

<sup>5</sup> Unité Hospitalo-Universitaire de Santé professionnelle, Hôpitaux Universitaires de Paris Ile-de-France Ouest, AP-HP, France

<sup>6</sup> Unité de Recherche Clinique, Hôpitaux Universitaires de Paris Ile-de-France Ouest, AP-HP, France

\*Auteur correspondant : UMR 1168, UFR des Sciences de la Santé Simone Veil, 2 avenue de la source de la Bièvre, 78180 Montigny-le-Bretonneux ; marie.herr@uvsq.fr

Conflits d'intérêt : aucun.

### Résumé

Les essais de non-infériorité et d'équivalence sont indiqués pour rendre disponible un nouveau traitement dont on n'attend pas a priori qu'il soit plus efficace que le traitement de référence dans une indication donnée. En contrepartie d'une possible perte d'efficacité définie a priori, on attend en revanche du nouveau traitement qu'il apporte un bénéfice par rapport au traitement de référence sur un autre plan, en matière de sécurité d'utilisation par exemple. Les essais de non-infériorité et d'équivalence présentent des différences méthodologiques par rapport aux essais de supériorité, souvent mieux connus de la communauté médicale. Dans cet article, nous présentons les points clés de la méthodologie des essais de non-infériorité et d'équivalence afin d'éclairer la lecture de ce type d'essai. La méthodologie générale (hypothèses, règles de décision, nombre de sujets nécessaire, stratégie d'analyse) est présentée à l'aide d'exemples et d'illustrations graphiques. Les enjeux et points critiques sont identifiés et discutés, en particulier le choix du comparateur et de la marge de non-infériorité.

**Mots clé** : essais cliniques, non-infériorité, équivalence, méthodologie

## **Abstract**

Non-inferiority and equivalence trials aim to promote new treatments that are not expected to be superior to existing ones in a given indication. In order to compensate for a possible loss of efficacy, the new treatment should offer other advantages compared to the reference treatment, a better safety of use for example. Their methods somewhat differ from those of superiority trials, often better known to the medical community. This article presents the key points of the methodology of non-inferiority and equivalence trials in order to inform the readers of such trials about the issues and critical points. The general methodology (hypotheses, decision rules, number of subjects required, and strategy of analysis) is presented using examples and graphic illustrations. The issues and critical points are identified and discussed, in particular the choice of the comparator and of the margin of non-inferiority.

**Key words:** clinical trials, non-inferiority, equivalence, methodology

## Introduction

En matière de démonstration d'efficacité thérapeutique, l'essai de supériorité reste la référence. Néanmoins, les conditions de réalisation d'un tel essai ne sont pas toujours réunies. La comparaison versus placebo n'est en effet pas toujours possible pour des raisons éthiques. On peut alors envisager un essai contrôlé versus traitement de référence, mais encore faut-il qu'il y ait une raison pour espérer une plus grande efficacité du nouveau traitement par rapport au traitement de référence. Un nouveau traitement ne prétend pas toujours à une meilleure efficacité mais peut apporter un bénéfice par ailleurs, une meilleure sécurité d'utilisation par exemple. Les essais de non-infériorité et d'équivalence sont alors tout à fait indiqués et valorisables auprès des autorités de régulation [1, 2]. Alors pourquoi ce type d'essais suscite-il régulièrement des réactions de scepticisme, voire d'inquiétude, dans la littérature ? En témoignent les publications remettant en cause la non-infériorité des nouveaux anticoagulants par rapport à l'héparine ou aux antivitamines K [3, 4] ou la tribune de Garattini et Bertele dans le Lancet en 2007, intitulée « Non-inferiority trials are unethical because they disregard patients' interests » [5]. Les essais de non-infériorité et les essais d'équivalence reposent sur des hypothèses différentes de celles des essais de supériorité et nécessitent une stratégie d'analyse particulière. Ils posent également des questions méthodologiques spécifiques, comme le choix du traitement comparateur et de la marge de non-infériorité ou d'équivalence. Il n'est donc pas possible d'avoir la même grille de lecture pour un essai de non-infériorité ou d'équivalence que pour un essai de supériorité. A ce titre, les critères CONSORT (Consolidated Standards of Reporting Trials) d'évaluation de la qualité des rapports d'essais cliniques font l'objet d'une extension aux essais de non-infériorité et d'équivalence [6]. Dans cet article, nous présentons les points clés de la méthodologie des essais de non-infériorité et d'équivalence afin d'éclairer la lecture de ce type d'essai.

## Principe général

Un essai de non-infériorité cherche à démontrer que l'efficacité d'un nouveau traitement A n'est pas inférieure à celle du traitement de référence B, en d'autres termes que A « ne fait pas moins bien » que B. Une telle démonstration nécessite de définir a priori la perte d'efficacité maximale autorisée pour conclure à la non-infériorité de A par rapport à B. En contrepartie de la perte d'efficacité consentie, on attend de A qu'il apporte un bénéfice par ailleurs : une meilleure sécurité d'utilisation, une voie d'administration moins invasive, une réduction de la durée de traitement, une simplification des modalités de dispensation ou de conservation, ou une réduction du coût de traitement [7, 8]. Le tableau 1 illustre le bénéfice attendu du nouveau traitement et la perte d'efficacité consentie dans différents essais de non-infériorité récents dans le domaine de l'épilepsie, de l'analgésie et des anti-infectieux.

Les essais d'équivalence sont très similaires aux essais de non-infériorité, à la différence que l'hypothèse de départ est bilatérale : l'efficacité du traitement A ne doit pas être trop inférieure ni trop supérieure à celle du traitement B. Ce type d'essai est indiqué lorsqu'un surcroît d'efficacité pourrait être délétère, par exemple dans le cas des traitements sédatifs en anesthésiologie. Un autre exemple d'utilisation est la démonstration de l'équivalence thérapeutique de médicaments antiasthmatiques inhalés suite à un changement de formulation (changement de gaz et de dispositif d'administration) [9]. Néanmoins, la démonstration de la non-infériorité reste beaucoup plus courante que la démonstration de l'équivalence dans les essais thérapeutiques ou prophylactiques. Les essais d'équivalence doivent être distingués des essais de bioéquivalence qui utilisent des critères pharmacocinétiques pour montrer que deux traitements, souvent un médicament princeps et son générique, ont la même biodisponibilité [6, 10, 11].

### **Hypothèses et règles de décision**

Les hypothèses posées dans les essais de supériorité, de non-infériorité et d'équivalence sont schématisées dans les Figures 1 à 3. Dans un essai de supériorité, on définit a priori une différence  $\Delta$  correspondant à la plus petite différence d'effet que l'on juge intéressante entre le traitement étudié et le placebo (ou traitement de référence). Ce  $\Delta$  permet de calculer le nombre de sujets nécessaire pour être en capacité de montrer cette différence si elle existe mais il n'intervient pas formellement dans la règle de décision de l'essai. Dès lors que la différence d'effet observée entre les groupes est significative (autrement dit que l'intervalle de confiance à 95% -en cas de risque de première espèce  $\alpha$  à 5%- de la différence est strictement supérieur à 0), la supériorité statistique est établie. Le  $\Delta$  peut également jouer un rôle au moment de l'interprétation du résultat et de sa pertinence clinique [19]. Dans le cas des essais de non-infériorité et d'équivalence, le  $\Delta$  correspond à la plus grande différence d'effet tolérée pour conclure à la non-infériorité ou à l'équivalence. Son rôle est central dans la règle de décision : il définit la zone, dans laquelle l'intervalle de confiance à  $1-\alpha$  de la différence entre groupes doit être entièrement compris pour conclure à la non-infériorité ou à l'équivalence (Figures 2 et 3). Ainsi, deux traitements peuvent être jugés équivalents malgré une différence d'effet moyen statistiquement significative si cette différence reste inférieure au  $\Delta$  fixé a priori (cas n°2 sur les Figures 2 et 3) [7, 11]. On notera que les figures présentées dans cet article illustrent le cas d'un critère de jugement qui augmente avec l'efficacité du traitement ; dans le cas d'un critère de jugement qui diminuerait avec l'efficacité du traitement (comme l'évaluation de la douleur sur une échelle visuelle analogique), les figures seraient inversées (avec un  $\Delta$  de non-infériorité supérieur à 0 et une zone de non-infériorité sur la gauche de l'axe).

## Choix du comparateur

Dans un essai de non-infériorité ou d'équivalence, il convient de s'assurer que le comparateur est bien le traitement présentant la meilleure efficacité dans l'indication considérée et qu'il est utilisé de façon optimale, à la bonne posologie notamment [20].

L'utilisation comme traitement comparateur d'un médicament ayant lui-même été mis sur le marché sur la base d'un essai de non-infériorité peut être problématique. Imaginons qu'un traitement A ait fait la preuve de son efficacité versus placebo dans un premier essai, puis que l'on démontre la non-infériorité du traitement B par rapport à A, le traitement B peut-il être utilisé comme comparateur dans le développement d'un nouveau traitement C ? Cela n'est pas souhaitable car c'est prendre le risque que C ne soit pas beaucoup plus efficace que le placebo, compte tenu de la perte d'efficacité consentie pour B par rapport à A puis pour C par rapport à B (Figure 4). Néanmoins, ce genre de situation semble difficilement évitable lorsque les standards de traitement évoluent et que des médicaments commercialisés sur la base d'essais de non-infériorité deviennent la référence. C'est par exemple ce que préfigurent les recommandations de l'*International League Against Epilepsy* qui admettent désormais l'utilisation du lévétiracétam, antiépileptique de 2<sup>ème</sup> génération, comme comparateur dans les essais de non-infériorité, au même titre que la carbamazépine commercialisée en 1964 [21].

Lorsque cela est possible d'un point de vue éthique, il est conseillé d'inclure un troisième bras placebo dans l'essai afin de vérifier l'efficacité des deux traitements par rapport au placebo [19, 22, 23]. A minima, l'essai doit reproduire le plus fidèlement possible les conditions ayant permis de prouver l'efficacité du traitement de référence, en matière de critères de sélection, de modalités de randomisation, de respect de l'aveugle, de schéma de traitement, de choix du critère de jugement principal et de méthode de mesure. La mise en évidence de similarités avec l'essai princeps, notamment en ce qui concerne l'observance et les abandons, est un critère de fiabilité de l'essai [8, 22].

Une méthodologie alternative lorsqu'il n'est pas possible d'inclure un bras placebo dans l'essai est d'évaluer le nouveau traitement par rapport à un placebo historique (lorsque disponible). Cette méthode repose sur une méta-analyse des essais antérieurs (comparateur versus placebo) qui permet d'une part de définir l'efficacité minimale historique du comparateur pour le choix de la marge de non-infériorité (et s'assurer que le comparateur présente bien un bénéfice vis-à-vis du placebo) et d'autre part d'évaluer si le nouveau traitement apporte un bénéfice par rapport à un placebo hypothétique (ici historique). Cette méthode présente des limites (évolution des pratiques, hétérogénéité des études, populations différentes) qu'il convient d'évaluer et de minimiser dans le choix du design de l'étude [20, 24].

## Choix de la marge de non-infériorité ou d'équivalence

La marge de non-infériorité ou d'équivalence  $\Delta$  représente la plus grande différence d'efficacité que l'on tolère entre les deux traitements. Il s'agit d'un point critique de la méthodologie des essais de non-infériorité ou d'équivalence. Un  $\Delta$  réduit permet une approche conservatrice par rapport au traitement de référence mais la non-infériorité ou l'équivalence est plus coûteuse (en nombre de sujets) à démontrer. A l'inverse, un  $\Delta$  plus large facilite la mise en évidence de la non-infériorité ou de l'équivalence, au prix d'une incertitude plus grande sur la réelle efficacité du traitement [23, 25].

Dans tous les cas, le  $\Delta$  doit être inférieur à la différence d'efficacité minimale du traitement de référence par rapport au placebo, initialement démontrée dans un essai de supériorité [24]. Le terme « minimale » renvoie à la borne inférieure de l'intervalle de confiance de la différence d'efficacité montrée dans l'essai versus placebo. Les essais cliniques étant réalisés dans des échantillons de patients, la mesure de l'efficacité est assortie d'un intervalle de confiance, très souvent à 95%, qui représente l'intervalle dans lequel la valeur vraie (fixe mais inconnue) de l'efficacité a 95% de chances de se trouver. Considérer la borne inférieure de l'intervalle de confiance à 95% revient à considérer l'estimation basse de l'efficacité du traitement de référence. Or, nombre d'essais utilisent l'estimation moyenne et non l'estimation basse de l'efficacité du traitement de référence. Le  $\Delta$  s'en trouve augmenté et il est alors plus facile de conclure à la non-infériorité ou à l'équivalence. Ainsi, Pérard et al [3] montraient en 2007 que la perte d'efficacité consentie dans la maladie thromboembolique veineuse était supérieure au gain d'efficacité minimal du traitement de référence dans 85% des cas. Il faut donc être vigilant sur ce point pour repérer les cas où la marge de non-infériorité est définie de façon trop tolérante pour faciliter la démonstration de la non-infériorité, réduire le nombre de sujets nécessaire et in fine la durée et le coût de l'essai, mais au risque d'approuver un traitement dont on n'est finalement pas sûr qu'il fasse mieux que le placebo [26].

Une règle communément admise est de considérer le  $\Delta$  comme acceptable s'il est inférieur à la moitié (voire au tiers) de l'efficacité minimale démontrée du traitement de référence. Lorsque les données d'efficacité du traitement de référence sont nombreuses, il est recommandé de réaliser au préalable une méta-analyse de ces données et d'utiliser l'estimation globale pour décider du  $\Delta$  [19]. Si cette règle mathématique donne une limite à ne pas franchir, le  $\Delta$  doit nécessairement faire l'objet d'un consensus d'experts qui le définissent comme la plus grande différence considérée comme cliniquement négligeable ou comme la plus grande perte d'efficacité n'entraînant pas une perte de chance substantielle pour le patient [23, 27]. Le nombre d'experts ayant participé à la décision, le nombre d'institutions représentées et le niveau d'accord sur le  $\Delta$  choisi sont des informations utiles mais rarement données dans les publications [28]. Pour des critères de jugement « durs », comme la mortalité, la marge de non-infériorité doit être d'autant plus réduite que la gravité de la perte d'efficacité est grande ; en revanche, pour des critères de

jugement plus « mous » et lorsque la perte d'efficacité est compensée par un bénéfice substantiel, une marge de non-infériorité plus large peut être admise [4]. Schumi et Wittes [19] suggèrent de compléter l'avis des experts cliniciens sur la perte d'efficacité consentie au regard des bénéfices potentiels par celui de patients experts, plus à même de juger de l'acceptabilité du  $\Delta$ .

En pratique, la détermination de la marge de non-infériorité fait très souvent appel au jugement clinique ou à des recommandations générales de sociétés savantes plutôt qu'à une approche formelle.

### Nombre de sujets nécessaire

Comme dans les essais de supériorité, le calcul du nombre de sujets nécessaire repose sur le choix des risques de première et deuxième espèce ( $\alpha$  et  $\beta$ ), le  $\Delta$  fixé a priori, une estimation de la variance de la réponse sur la base des études antérieures ( $\sigma^2$ ), l'éventuelle différence d'efficacité attendue entre les deux traitements ( $\delta$ ) et le caractère unilatéral ou bilatéral de l'hypothèse à tester [25, 29].

Pour un essai avec deux bras parallèles de taille identique, le nombre de sujet par groupe est calculé selon les formules suivantes :

$$N = (Z_{1-\alpha} + Z_{1-\beta})^2 \times 2\sigma^2 / (\Delta - \delta)^2 \text{ pour une hypothèse unilatérale (non-infériorité),}$$

$$N = (Z_{1-\alpha} + Z_{1-\beta/2})^2 \times 2\sigma^2 / (\Delta - \delta)^2 \text{ pour une hypothèse bilatérale (équivalence),}$$

avec  $Z_{1-\alpha} = 1,64$  si  $\alpha = 0,05$  et  $Z_{1-\beta} = 0,84$  et  $Z_{1-\beta/2} = 1,28$  si  $\beta = 0,20$ .

Pour un critère de jugement qualitatif :  $2\sigma^2 = p_A (1 - p_A) + p_B (1 - p_B)$  avec  $p_A$  = le taux de réponse attendu avec le nouveau traitement et  $p_B$  = le taux de réponse attendu avec le traitement de référence.

Il convient ensuite de multiplier l'effectif obtenu par un facteur d'inflation pour tenir compte de l'attrition attendue, c'est-à-dire la réduction d'effectif entre le début et la fin de l'essai en raison des sorties d'essais pour diverses raisons (dans des proportions raisonnables). La présence du  $\Delta$  au carré au dénominateur de la formule explique l'influence considérable du choix de la marge de non-infériorité sur le nombre de sujets à inclure. Cela peut être illustré à partir de l'essai visant à comparer le co-trimoxazole per os à la benzathine benzylpénicilline dans le traitement de l'impétigo de l'enfant cité dans le tableau 1 [18]. Le calcul du nombre de sujets nécessaire proposé par les auteurs est basé sur une marge de non-infériorité fixée à 10% (scenario 1, tableau 2). Si l'on réduit la marge de non-infériorité de 10% à 5%, le nombre de sujets nécessaire dans chaque bras est multiplié par un facteur 4 (scenario 2). Au contraire, si l'on relâche la contrainte sur la marge de non-infériorité jusqu'à 15%, le nombre de sujets nécessaire dans chaque bras est divisé par 2,25 (scenario 3). Le calcul du nombre de sujets nécessaire est également très sensible à



l'efficacité attendue dans chaque groupe. Lorsque l'efficacité attendue du nouveau traitement est considérée moindre que celle du traitement de référence, la démonstration de la non-infériorité devient bien plus exigeante pour ce qui est de l'effectif (scénario 4).

### **Stratégie d'analyse**

L'analyse en intention de traiter (ITT) porte sur tous les patients randomisés, qu'ils aient ou non reçu le traitement selon les modalités prévues dans le protocole. C'est l'approche recommandée dans les essais de supériorité car elle est conservatrice. En effet, les abandons en cours d'essai et les cas d'inobservance sont souvent liés à des problèmes d'effets indésirables ou de manque d'efficacité. Leur prise en compte dans l'analyse permet d'approcher au mieux l'efficacité du traitement en conditions réelles mais diminue les chances de montrer une différence entre les groupes. L'analyse en per protocole (PP) ne porte que sur les patients ayant reçu le traitement selon les modalités prévues par le protocole. En excluant les abandons et les cas d'inobservance, l'analyse en PP produit une estimation maximale de l'efficacité du traitement et augmente les chances de montrer une différence entre les groupes. Dans un contexte où l'on cherche à montrer l'absence de différence entre deux traitements, l'analyse en PP est donc la stratégie conservatrice. C'est pourquoi l'analyse en PP est l'analyse recommandée dans le cadre des essais de non-infériorité et d'équivalence [8].

Cependant, il existe des cas où l'analyse en PP peut être biaisée. Par exemple, si les patients qui répondent mal au traitement quittent l'essai prématurément, la population analysée sera enrichie en patients « bons répondeurs » ; l'équivalence par rapport au traitement de référence sera alors plus facile à montrer mais ne correspondra pas à l'effet attendu en population. Pour cette raison, il est essentiel d'avoir une description des caractéristiques des patients ayant interrompu leur participation. Si les deux traitements à l'étude n'ont pas les mêmes profils pour ce qui est de l'observance, ceci est déjà une preuve qu'ils ne sont pas entièrement équivalents [22].

Au final, et conformément aux exigences des autorités de régulation européenne et américaine, les deux types d'analyse doivent être présentés [8, 19]. Un essai ne pourra conclure à la non-infériorité ou à l'équivalence que si les résultats sont concluants en ITT et en PP. Or, les résultats d'une revue de 162 essais de non-infériorité et d'équivalence publiés en 2003 et 2004 montraient que moins de la moitié d'entre eux présentaient effectivement les résultats des analyses en ITT et en PP [30].

### **Le poids des écarts au protocole**

Dans un essai de supériorité, les écarts au protocole (arrêts ou échanges de traitements, perdus de vue...) tendent à minimiser l'écart entre le traitement étudié et le comparateur (placebo ou traitement de référence). Il y a donc une incitation forte à suivre de près la qualité de la mise en œuvre du protocole pour le promoteur et l'investigateur. Dans un essai de non-infériorité, les écarts au protocole vont également réduire la probabilité de mettre en évidence une différence entre les deux traitements comparés. Mais au lieu de pénaliser l'essai, ces problèmes augmentent les chances de conclure à la non-infériorité ou à l'équivalence. En cas d'écarts au protocole, le risque est donc de conclure à tort à la non-infériorité. C'est pourquoi, il faut être particulièrement vigilant sur la qualité des essais de non-infériorité et d'équivalence [22, 23].

### **Passage de la non-infériorité à la supériorité et inversement**

Il est possible qu'en plus d'être non-inférieur, le nouveau traitement A fasse mieux que le traitement de référence B (cas n°1 sur la figure 2). On peut alors se demander si finalement A n'est pas supérieur à B. Les essais de non-infériorité nécessitent en général un plus grand nombre de sujets que les essais de supériorité (parce que le  $\Delta$  de non-infériorité est plus petit que le  $\Delta$  de supériorité). La question de la puissance est donc rarement limitante pour passer de la non-infériorité à la supériorité. L'analyse de supériorité devra néanmoins avoir été prévue au protocole, en ITT, avec une évaluation du bénéfice clinique supplémentaire au regard des risques du traitement. Un  $\Delta$  minimal d'efficacité pour conclure à la supériorité peut être défini a priori même s'il ne participe pas au calcul du nombre de sujets nécessaire [25, 31].

Lorsqu'un essai de supériorité versus traitement de référence n'est pas concluant, la démonstration de la non-infériorité pourrait être envisagée comme une solution de repli pour « sauver » le traitement à l'étude. Cela nécessite qu'un  $\Delta$  de non-infériorité, plus petit que le  $\Delta$  de supériorité, ait été défini a priori selon un consensus d'experts dans le protocole. Il faut par ailleurs s'assurer que l'étude soit suffisamment puissante pour permettre de passer de la supériorité à la non-infériorité. Une telle transformation s'avère plus hasardeuse que la précédente [23, 25, 31].

### **Polémiques, dérives et extension**

En 2007, Garattini et Bertele [5] cristallisent les inquiétudes quant aux essais de non-infériorité dans une tribune du Lancet. Les auteurs dénoncent l'utilisation des essais de non-infériorité pour commercialiser de nouvelles molécules sans réel intérêt par rapport aux traitements existants (médicaments qualifiés de

« me-too »). Les auteurs rappellent que le laxisme sur les marges de non-infériorité revient à commercialiser des médicaments qui sont potentiellement beaucoup moins efficaces que le traitement de référence, entraînant une perte de chance substantielle pour les patients. De nombreux auteurs ont réagi [32-36], rappelant l'intérêt des essais de non-infériorité lorsque les essais versus placebo ne sont plus éthiques et appelant à la rigueur méthodologique pour ne pas décrédibiliser les essais de non-infériorité.

La valorisation des essais de non-infériorité et d'équivalence passe par une méthodologie rigoureuse et une présentation conforme aux recommandations CONSORT [6]. Une réelle marge de progression existe : parmi les 168 essais de non-infériorité publiés entre 2010 et 2015 dans huit revues médicales majeures, moins de la moitié justifiaient le choix de la marge de non-infériorité et 39% ne donnaient les résultats que d'un seul type d'analyse, en ITT le plus souvent [28]. Des extensions peuvent également être imaginées pour promouvoir le bénéfice apporté par les nouveaux traitements. D'Agostino et al [24] suggèrent de coupler la démonstration de la non-infériorité pour ce qui est de l'efficacité à la démonstration de la supériorité du point de vue de la sécurité d'utilisation. Pour cela, il convient de réaliser deux calculs du nombre de sujets nécessaire, un pour chaque objectif, et de choisir le nombre plus élevé afin d'être en mesure de répondre aux deux objectifs.

## **Conclusion**

De façon pratique, cette synthèse invite le lecteur d'essais de non-infériorité et d'équivalence à être vigilant sur les points suivants :

1. Titre et résumé : identification du type d'essai ;
2. Introduction : justification du schéma de non-infériorité ou d'équivalence ;
3. Méthode : justification du choix du comparateur et des critères de jugement ;
4. Méthode : justification de la marge de non-infériorité ou d'équivalence ;
5. Résultats : présentation des résultats de l'analyse en PP et en ITT, avec les intervalles de confiance ;
6. Résultats : analyse des écarts entre la population en ITT et la population PP, globalement et par bras de traitement, à partir du diagramme de flux ;
7. Discussion : interprétation des résultats en fonction des hypothèses de départ, notamment s'il est décidé de passer en supériorité.

## Glossaire

- $\alpha$  risque de première espèce, i.e. risque de conclure à une différence au vu des données de l'échantillon alors qu'elle n'existe pas dans la population
- $\beta$  risque de deuxième espèce, i.e. risque de conclure à l'absence de différence au vu des données de l'échantillon alors que la différence existe dans la population
- $1-\beta$  puissance
- $\Delta$  plus petite différence cliniquement pertinente définie a priori dans le cas d'un essai de supériorité ( $\Delta_s$ ) ; plus grande perte d'efficacité tolérée pour conclure à la non-infériorité ou à l'équivalence ( $\Delta_{ni}$ ) ; plus grande différence d'efficacité tolérée pour conclure à l'équivalence ( $\Delta_{eq}$ )
- $\delta$  différence d'efficacité attendue entre les deux traitements (souvent  $\delta=0$ )
- $\sigma^2$  estimation de la variance de la réponse sur la base des études antérieures
- $IC_{1-\alpha}$  intervalle de confiance à  $1-\alpha$ ,  $IC_{95\%}$  quand  $\alpha=5\%$
- ITT** intention de traiter
- PP** per protocole
- $p_x$  taux de réponse attendu dans le groupe traité par X
- $z_q$  quantile de rang q de la fonction de répartition de la loi normale centrée réduite

## Références

1. International Conference on Harmonisation. ICH Harmonised Tripartite Guideline : Statistical principles for clinical trials - E 9 .  
[http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Efficacy/E9/Step4/E9\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf). Accès au site le 6 avril 2017.
2. Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness : Guidance for Industry .

<https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf>. Accès au site le 6 avril 2017.

3. Perard L, Hot A, Cucherat M, Simon M, Desmurs H, Coppere B, et al. [Non-inferiority trial used in venous thromboembolic disease. A warily interpretation is necessary!]. *La Revue de medecine interne*. 2007;28(11):731-6.
4. Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. *Journal of the American College of Cardiology*. 2005;46(11):1986-95.
5. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370(9602):1875-7.
6. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *Jama*. 2012;308(24):2594-604.
7. Sedgwick P. What is a non-inferiority trial? *BMJ*. 2013;347:f6853.
8. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *Journal of hepatology*. 2007;46(5):947-54.
9. Ebbutt AF, Frith L. Practical issues in equivalence trials. *Statistics in medicine*. 1998;17(15-16):1691-701.
10. Guideline on the Investigation of bioequivalence, CPMP, 2010, CPMP/EWP/QWP/1401/98 Rev. 1/Corr \*\*.  
[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2010/01/WC500070039.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500070039.pdf). Accès au site le 6 avril 2017.
11. Sedgwick P. Equivalence trial. *BMJ*. 2013;346:f184.
12. Baulac M, Rosenow F, Toledo M, Terada K, Li T, De Backer M, et al. Efficacy, safety, and tolerability of lacosamide monotherapy versus controlled-release carbamazepine in patients with newly diagnosed epilepsy: a phase 3, randomised, double-blind, non-inferiority trial. *The Lancet Neurology*. 2016.
13. Stessel B, Boon M, Joosten EA, Ory JP, Evers S, van Kuijk SM, et al. Metamizole versus ibuprofen at home after day surgery: study protocol for a randomised controlled trial. *Trials*. 2016;17(1):471.

14. Pickering G, Moustafa F, Macian N, Schmidt J, Pereira B, Dubray C. A New Transmucous-Buccal Formulation of Acetaminophen for Acute Traumatic Pain: A Non-inferiority, Randomized, Double-Blind, Clinical Trial. *Pain physician*. 2015;18(3):249-57.
15. Marks D, Bisset L, Thomas M, O'Leary S, Comans T, Ng SK, et al. An experienced physiotherapist prescribing and administering corticosteroid and local anaesthetic injections to the shoulder in an Australian orthopaedic service, a non-inferiority randomised controlled trial and economic analysis: study protocol for a randomised controlled trial. *Trials*. 2014;15:503.
16. Paul M, Bishara J, Yahav D, Goldberg E, Neuberger A, Ghanem-Zoubi N, et al. Trimethoprim-sulfamethoxazole versus vancomycin for severe infections caused by meticillin resistant *Staphylococcus aureus*: randomised controlled trial. *BMJ*. 2015;350:h2219.
17. Bernard L, Dinh A, Ghout I, Simo D, Zeller V, Issartel B, et al. Antibiotic treatment for 6 weeks versus 12 weeks in patients with pyogenic vertebral osteomyelitis: an open-label, non-inferiority, randomised, controlled trial. *Lancet*. 2015;385(9971):875-82.
18. Bowen AC, Tong SY, Andrews RM, O'Meara IM, McDonald MI, Chatfield MD, et al. Short-course oral co-trimoxazole versus intramuscular benzathine benzylpenicillin for impetigo in a highly endemic region: an open-label, randomised, controlled, non-inferiority trial. *Lancet*. 2014;384(9960):2132-40.
19. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials*. 2011;12:106.
20. International Conference on Harmonisation. ICH Harmonised Tripartite Guideline : Guidance on choice of control group and related design and conduct issues in clinical trials - E 10. [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Efficacy/E10/Step4/E10\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf). Accès au site le 6 avril 2017.
21. Glauser T, Ben-Menachem E, Bourgeois B, Cnaan A, Guerreiro C, Kalviainen R, et al. Updated ILAE evidence review of antiepileptic drug efficacy and effectiveness as initial monotherapy for epileptic seizures and syndromes. *Epilepsia*. 2013;54(3):551-63.
22. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ*. 1996;313(7048):36-9.

23. Vicaut E, Cucherat M. [Noninferiority trials: several simple principles]. *Presse medicale*. 2007;36(3 Pt 2):531-5.
24. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Statistics in medicine*. 2003;22(2):169-86.
25. Elie C, De Rycke Y, Jais JP, Marion-Gallois R, Landais P. [Methodological and statistical aspects of equivalence and non inferiority trials]. *Revue d'epidemiologie et de sante publique*. 2008;56(4):267-77.
26. Powers JH, Cooper CK, Lin D, Ross DB. Sample size and the ethics of non-inferiority trials. *Lancet*. 2005;366(9479):24-5.
27. Committee for Medicinal Products for Human U, Efficacy Working P, Committee for Release for C. Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Statistics in medicine*. 2006;25(10):1628-38.
28. Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PP. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ open*. 2016;6(10):e012594.
29. Julious SA, Campbell MJ. Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data. *Statistics in medicine*. 2012;31(24):2904-36.
30. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *Jama*. 2006;295(10):1147-51.
31. The European Agency for the Evaluation of Medicinal Products. Committee on Proprietary Medical Products : Points to consider on switching between superiority and non-inferiority. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003658.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003658.pdf). Accès au site le 6 avril 2017.
32. Nunn AJ, Meredith SK, Spiegelman MK, Ginsberg AM, Gillespie SH. The ethics of non-inferiority trials. *Lancet*. 2008;371(9616):895; author reply 6-7.
33. Gandjour A. The ethics of non-inferiority trials. *Lancet*. 2008;371(9616):895; author reply 6-7.
34. Soliman EZ. The ethics of non-inferiority trials. *Lancet*. 2008;371(9616):895; author reply 6-7.

35. Chuang-Stein C, Beltangady M, Dunne M, Morrison B. The ethics of non-inferiority trials. *Lancet*. 2008;371(9616):895-6; author reply 6-7.
36. Menten J, Boelaert M. The ethics of non-inferiority trials. *Lancet*. 2008;371(9616):896; author reply -7.



**Tableau 1. Exemples d'essais de non-infériorité**

<b>Nouveau traitement</b>	<b>Traitement de référence</b>	<b>Indication</b>	<b>Bénéfice apporté par le nouveau traitement</b>	<b>Critère de jugement principal</b>	<b>Efficacité du traitement de référence ou déviation standard du critère de jugement principal</b>	<b>Différence maximale pour conclure à la non-infériorité (<math>\Delta</math>)</b>	<b>Référence de l'essai</b>
Lacosamide	Carbamazépine	Epilepsie partielle	Moins inducteur enzymatique	% de sujets indemnes de crise à 6 mois après équilibrage de la dose	60%	-12%	Baulac et al, Lancet Neurol 2016 [12]
Métamizole + paracétamol	Ibuprofène + paracétamol	Douleurs post-opératoires après une chirurgie ambulatoire douloureuse	Antalgique non opioïde ayant un meilleur profil de sécurité que les Anti-Inflammatoires Non Stéroïdiens (AINS)	Score de douleur entre 0 et 10 à J1	2,5	+1 point	Stessel et al, Trials 2016 [13]
Paracétamol par voie trans-muqueuse	Paracétamol par voie intraveineuse	Douleurs modérées liées à un traumatisme du bras ou de la jambe aux urgences	Moins invasif	Score de douleur entre 0 et 10 à 30 minutes	1,2	+1 point	Pickering et al, Pain Physician 2015 [14]

<b>Nouveau traitement</b>	<b>Traitement de référence</b>	<b>Indication</b>	<b>Bénéfice apporté par le nouveau traitement</b>	<b>Critère de jugement principal</b>	<b>Efficacité du traitement de référence ou déviation standard du critère de jugement principal</b>	<b>Différence maximale pour conclure à la non-infériorité (<math>\Delta</math>)</b>	<b>Référence de l'essai</b>
Injection sous-acromiale de corticoïde et d'un anesthésique local par un kinésithérapeute	Injection sous-acromiale de corticoïde et d'un anesthésique local par un chirurgien orthopédique	Douleurs de l'épaule nécessitant une prise en charge orthopédique	Réduction du coût et amélioration de l'accessibilité de la procédure	Score Shoulder Pain and Disability Index de 0 à 100, à baseline, 6 et 12 semaines	21,7	+15 points	Marks et al, Trials 2014 [15]

<b>Nouveau traitement</b>	<b>Traitement de référence</b>	<b>Indication</b>	<b>Bénéfice apporté par le nouveau traitement</b>	<b>Critère de jugement principal</b>	<b>Efficacité du traitement de référence ou déviation standard du critère de jugement principal</b>	<b>Différence maximale pour conclure à la non-infériorité (<math>\Delta</math>)</b>	<b>Référence de l'essai</b>
Triméthoprim-sulfaméthoxazole	Vancomycine	Infections sévères à Staphylocoque aureus Résistant à la Méricilline (SARM)	Option thérapeutique supplémentaire en cas d'antibio-résistance + relais per os	% d'échec du traitement à 7 jours, défini par un critère composite incluant le décès, la persistance de l'hypotension, de la fièvre ou de la bactériémie et l'absence d'amélioration du score Sequential Organ Failure Assessment (SOFA)	30%	+15%	Paul et al, BMJ 2015 [16]

<b>Nouveau traitement</b>	<b>Traitement de référence</b>	<b>Indication</b>	<b>Bénéfice apporté par le nouveau traitement</b>	<b>Critère de jugement principal</b>	<b>Efficacité du traitement de référence ou déviation standard du critère de jugement principal</b>	<b>Différence maximale pour conclure à la non-infériorité (<math>\Delta</math>)</b>	<b>Référence de l'essai</b>
Antibiothérapie 6 semaines	Antibiothérapie 12 semaines	Ostéomyélite vertébrale pyogène	Réduction du risque d'effets indésirables, du risque d'infection nosocomiale, du coût de traitement et du risque de résistance bactérienne	% de sujets guéris à 1 an	85%	-10%	Bernard et al, Lancet 2015 [17]
Co-trimoxazole per os	Benzathine benzylpénicilline en intra-musculaire	Impétigo extensif chez l'enfant	Moins invasif + option thérapeutique supplémentaire en cas d'antibio-résistance	% de succès thérapeutique (guérison ou réduction des lésions) à 7 jours	80%	-10%	Bowen et al, Lancet 2014 [18]

**Tableau 2. Calcul du nombre de sujets nécessaire d'un essai de non-infériorité selon différentes hypothèses de départ**

<b>Scenario</b>	<b>Puissance (1-<math>\beta</math>)</b>	<b>Risque de première espèce (<math>\alpha</math>)</b>	<b>Efficacité attendue du traitement de référence</b>	<b>Efficacité attendue du traitement expérimental</b>	<b>Marge de non-infériorité</b>	<b>Effectif à inclure dans chaque groupe</b>
1 (original)	80%	5%	80%	80%	10%	198
2	80%	5%	80%	80%	5%	792
3	80%	5%	80%	80%	15%	88
4	80%	5%	80%	75%	10%	860

*Calculs d'effectif réalisés avec SampSize, Epi Genesys Limited, d'après Julious et Campbell, 2012 [29].*

**Figure 1. Hypothèses et règles de décision d'un essai de supériorité**

*Lecture :  $\Delta_s$  = plus petite différence cliniquement pertinente définie a priori*

**Figure 2. Hypothèses et règles de décision d'un essai de non-infériorité**

*Lecture :  $\Delta_{ni}$  = plus grande perte d'efficacité tolérée pour conclure à la non-infériorité ; \* : La supériorité est admise si la borne inférieure de l'intervalle de confiance à 95% est strictement supérieure à 0, sous réserve que le passage à la supériorité ait été prévu au protocole et que l'analyse soit réalisée en ITT.*

**Figure 3. Hypothèses et règles de décision d'un essai d'équivalence**

*Lecture :  $\Delta_{eq}$  = plus grande différence d'efficacité tolérée pour conclure à l'équivalence.*

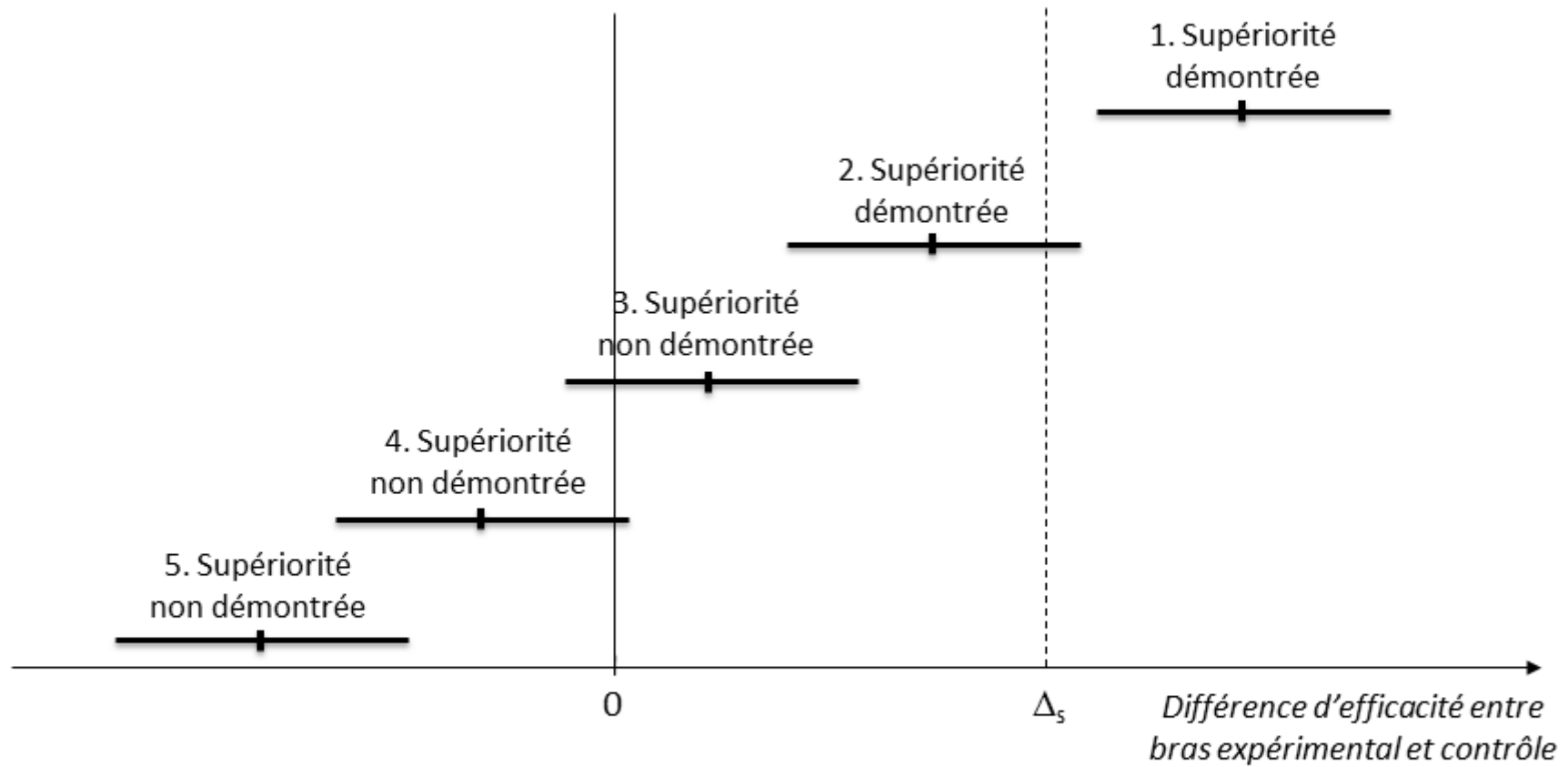
**Figure 4. Illustration du risque de perte d'efficacité en cascade lorsque le traitement comparateur d'un essai de non-infériorité a lui-même fait la démonstration de son efficacité dans un essai de non-infériorité**

*Dans l'hypothèse d'une perte d'efficacité consentie correspondant à la moitié de l'efficacité du traitement de référence dans chaque essai.*

### Supériorité :

$H_0$  : L'IC<sub>95%</sub> de la différence d'efficacité contient 0 ou est entièrement  $<0$

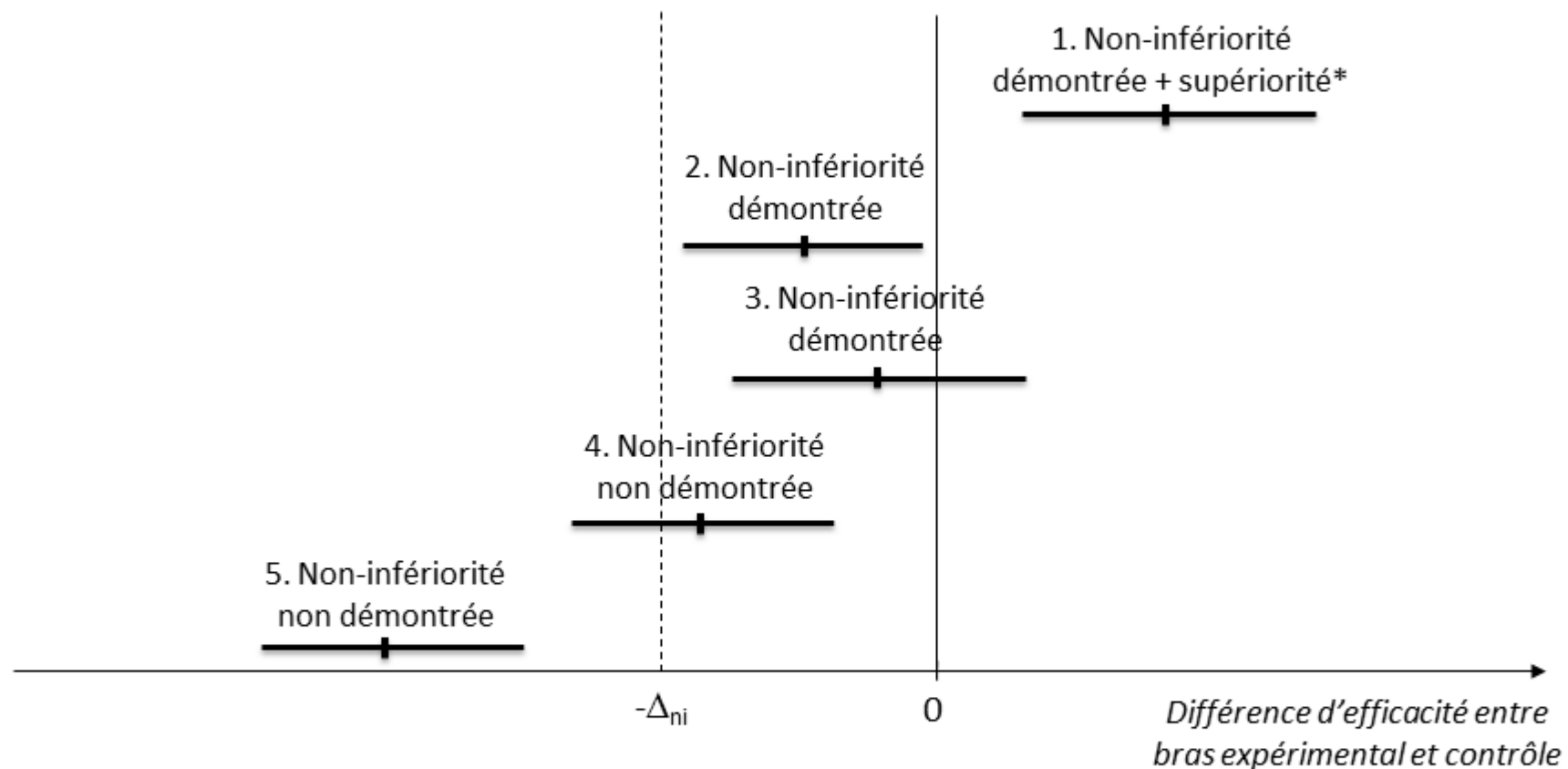
$H_1$  : L'IC<sub>95%</sub> de la différence d'efficacité est entièrement  $>0$



### Non-infériorité :

$H_0$  : La borne inférieure de l'IC<sub>95%</sub> de la différence d'efficacité est  $\leq -\Delta_{ni}$

$H_1$  : La borne inférieure de l'IC<sub>95%</sub> de la différence d'efficacité est  $> -\Delta_{ni}$





### Equivalence :

$H_0$  : L'IC<sub>95%</sub> de la différence d'efficacité contient  $-\Delta_{eq}$  ou  $\Delta_{eq}$  ou l'IC<sub>95%</sub> de la différence est entièrement  $<-\Delta_{eq}$  ou  $>\Delta_{eq}$

$H_1$  : L'IC<sub>95%</sub> de la différence d'efficacité est entièrement compris entre  $-\Delta_{eq}$  et  $\Delta_{eq}$

