



Société Française de
Pharmacologie et de Thérapeutique

Groupe de Travail Méthodologie

Livre blanc SFPT

De la nécessité de la méthodologie
dans l'évaluation des médicaments

Document compagnon

Dossier 6 – L'essai de non-infériorité

19 février 2022

Comité de rédaction et relecture (par ordre alphabétique)

Jean Luc Cracowski

Michel Cucherat

Dominique Deplanque

Behrouz Kassai

Charles Khouri

Silvy Laporte

Clara Locher

Florian Naudet

Edouard Ollier

Matthieu Roustit



[Licence Creative Commons](https://creativecommons.org/licenses/by/4.0/)

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International

Vous êtes autorisé à :

- Partager — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- Adapter — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

Table des matières

1	Introduction.....	7
2	Justification de la recherche de la non-infériorité	8
3	L'impossible démonstration de l'équivalence.....	11
3.1	Seuil de non-infériorité.....	12
3.2	Intervalle de confiance unilatéral ou bilatéral	12
3.3	Signification clinique du seuil de non-infériorité	13
4	Équivalence clinique ou non-infériorité	15
5	Le choix du seuil de non-infériorité.....	16
5.1	Introduction.....	16
5.2	Positionnement par rapport à l'efficacité du traitement de référence	17
5.3	Calcul du seuil.....	18
5.4	Limite en différence de risque ou en risque ratio	19
6	Situation actuelle.....	21
6.1	Étude de cas 1.....	21
6.2	Étude de cas 2.....	22
7	Enchaînement d'évaluations en non-infériorité	24
8	Méthodologie, le contrôle des biais dans l'essai de non-infériorité	25
8.1	Retour sur le principe de protection contre les biais dans l'essai de supériorité	25
8.2	Les biais de l'essai de non -infériorité et leur control	26
9	Méta-épidémiologie des essais de non-infériorité	29
10	Absence de différence dans un essai de différence (supériorité).....	30

1 Introduction

Les essais de non-infériorité (« *non-inferiority trial* »), parfois appelés par abus de langage essais d'équivalence (« *equivalence trial* »), sont fréquents dans l'évaluation clinique des nouveaux traitements [1]. Ce type d'essais fait appel à une méthodologie et à des techniques statistiques particulières et différentes sur plusieurs points par rapport à l'essai de supériorité [2, 3]. De ce fait, des nouveaux traitements peuvent être acceptés sur la base d'essais d'équivalence discutables par méconnaissance des pièges et des spécificités de ce type d'études [4, 5]. En particulier, le processus décisionnel qui leur est attaché nécessite l'introduction d'une limite de non-infériorité choisie **arbitrairement**. De la valeur de ce seuil dépend entièrement le résultat et la pertinence de l'essai.

Les conclusions de ces essais sont aussi très souvent surinterprétées. Malgré les apparences, l'essai de non-infériorité, dit « d'équivalence », ne permet pas de conclure que le traitement étudié a une efficacité identique à celle du traitement de référence, mais simplement qu'il n'est pas nettement inférieur au traitement de référence utilisé comme comparateur. Comme nous le verrons par la suite, les méthodes disponibles permettent seulement de raisonnablement éliminer la possibilité que le traitement étudié soit nettement moins efficace que le traitement de référence. Ces techniques permettent d'exclure que le nouveau traitement entraîne une perte d'efficacité supérieure à une certaine limite, fixée a priori et qui est devrait être la plus grande perte d'efficacité cliniquement négligeable.

Ainsi, à l'issue d'un essai de non-infériorité concluant, rien ne permet d'exclure que le nouveau traitement soit en réalité moins efficace que le traitement de référence. La seule chose qui soit acquise (avec un risque alpha d'erreur de 5%) est que cette perte d'efficacité est inférieure à la limite que les investigateurs sont prêts à perdre compte tenu des avantages qu'offre le nouveau traitement par ailleurs, perte définie par la limite de non-infériorité.

Les essais de non-infériorité sont à distinguer des essais de bioéquivalence où l'équivalence concerne les paramètres pharmacocinétiques.

2 Justification de la recherche de la non-infériorité

Un nouveau traitement ne peut être introduit dans la stratégie thérapeutique que lorsqu'il existe une preuve issue d'essais cliniques qu'il représente un progrès par rapport au traitement de référence (ou par rapport à l'absence de traitement).

En général, le progrès thérapeutique est représenté par une efficacité supérieure à celle du traitement de référence. La preuve est apportée par un essai visant à montrer la supériorité du nouveau traitement (essai de supériorité).

Cependant, dans certaines situations, une avancée thérapeutique peut, non pas être une efficacité supérieure, mais simplement une plus grande facilité d'utilisation, un traitement moins lourd pour les patients, une meilleure tolérance ou un plus faible coût. Ces avantages pourront être suffisamment intéressants pour justifier l'adoption du nouveau traitement même si son efficacité n'est pas supérieure à celle du traitement de référence, à condition qu'elle ne soit pas nettement inférieure. Même avec une efficacité légèrement moindre, ce traitement représente un progrès thérapeutique compte tenu de ces autres avantages.

La démonstration de l'intérêt du nouveau traitement sera apportée par un essai cherchant à mettre en évidence la non-infériorité de celui-ci par rapport au traitement de référence.

Par exemple, l'intérêt des héparines de bas poids moléculaires dans le traitement des thromboses veineuses profondes est une plus grande facilité d'utilisation. Il en est de même des changements de modalités d'administration d'un même produit à la recherche d'une plus grande faisabilité. Comme le remplacement d'une antibiothérapie intraveineuse par une antibiothérapie orale pour le traitement des endocardites infectieuses (étude POET [6]). Cette approche a aussi été envisagée pour la fibrinolyse avec l'alteplase à la phase aiguë de l'infarctus (essai COBALT[7]).

Dans les tumeurs primaires osseuses, une stratégie conduisant à une chirurgie conservatrice sera un progrès thérapeutique par rapport à une chirurgie d'amputation, même s'il y a une petite perte d'efficacité en termes de survie.

Exemple des HBPM dans le traitement des TVP

Jusqu'à récemment, le traitement standard des thromboses veineuses profondes consistait en l'hospitalisation et l'administration intraveineuse continue d'héparine non fractionnée (HNF) durant 5 à 10 jours, suivi par un traitement anticoagulant oral d'au moins 3 mois. L'utilisation de l'héparine non fractionnée nécessite un monitoring biologique pour l'ajustement des doses.

Les héparines de bas poids moléculaires (HBPM) présentent de nombreux avantages par rapport à l'héparine non fractionnée : leur demi-vie plus longue rend possible leur administration en 2, voire 1, prise par jour, l'adaptation de la dose par monitoring biologique n'est pas nécessaire (une adaptation au poids du patient est suffisante). Ces deux points rendent envisageable le traitement de ces patients à domicile. L'ensemble de ces avantages fait que les HBPM représentent une alternative à l'HNF intéressante en pratique même sans surcroît d'efficacité [8]. En effet, il serait tout à fait justifié d'adopter les HBPM même si elles ne s'avèrent qu'équivalentes en efficacité par rapport au traitement standard. Dans cet exemple ; le progrès thérapeutique réside dans l'amélioration de la praticabilité du traitement, et du confort du patient.

L'évaluation des HBPM dans le traitement de cette maladie s'est donc basée sur des essais de non-infériorité [8].

La désescalade thérapeutique, par exemple en cancérologie, avec un allègement des protocoles de chimiothérapie ou le recours à des traitements chirurgicaux moins délabrants représente aussi une situation où il est facilement justifiable de changer pour de nouveaux protocoles thérapeutiques d'efficacité seulement équivalente aux précédents, mais qui préservent mieux la qualité de vie des patients.

L'approche de non-infériorité ne donne des arguments permettant d'utiliser le nouveau traitement que si celui-ci présente une supériorité manifeste par rapport au traitement habituel sur d'autres points (praticabilité, tolérances, etc.). En leur absence, la perte d'efficacité consentie empêche de conclure que le nouveau traitement représente un progrès thérapeutique par rapport au précédent. Cette remarque est importante, car la tentation est grande d'évaluer en équivalence un nouveau traitement non innovant et de vanter ensuite son « efficacité équivalence » pour le faire utiliser en remplacement du précédent. Cette attitude est dangereuse, car il n'est pas possible d'exclure qu'elle conduise à remplacer un traitement par un autre, en réalité moins efficace, sans que cette substitution n'apporte un quelconque avantage.

Ces situations peuvent se généraliser de la façon suivante : le bénéfice d'un traitement est une notion multifactorielle dans laquelle intervient à la fois l'efficacité vis-à-vis des critères de jugement clinique, mais aussi la tolérance, la faisabilité, la lourdeur du traitement et le cout. La démonstration de « l'équivalence clinique » d'un nouveau traitement par rapport au traitement de référence est suffisante pour l'adoption de celui-ci chaque fois où le gain obtenu sur les autres dimensions du bénéfice représente un intérêt suffisant pour admettre une efficacité équivalente (c'est-à-dire potentiellement légèrement inférieure). Le Tableau 1 présente quelques situations de ce type.

Cependant cette prise de décision va s'appuyer sur des choix arbitraires qui consistent à décider si les avantages sont « suffisants ». En d'autres termes, quelle diminution de cout, quelle réduction de fréquence des effets indésirables représente un avantage suffisamment important pour justifier un changement. Toute la difficulté consiste à déterminer la quantité de perte d'efficacité que l'on peut consentir en regard des avantages apportés. Ce choix est le plus souvent arbitraire, dépendant du point de vue et du référentiel choisi. Il constitue la principale difficulté de la prise de décision en équivalence.

Dans certaines situations, l'essai peut servir à montrer l'avantage du nouveau traitement en même temps que sa non-infériorité. Par exemple, une fréquence d'effets indésirables moindre et une efficacité suffisante (non-inférieure). L'essai ne sera concluant que lorsque ces deux hypothèses seront vérifiées simultanément. S'ajoutera alors une problématique supplémentaire, celle des comparaisons multiples, le critère de jugement principal n'étant plus unique. Une analyse adéquate de ce type d'essai nécessitera la prise en compte de la multiplicité des critères de jugement avec par exemple une analyse hiérarchique des critères de jugement ou l'utilisation de co-critères de jugement.

Tableau 1 – Exemple d'avantages pouvant justifier une recherche de l'équivalence.

Avantage en termes de tolérance	<ul style="list-style-type: none"> • Fréquence des effets secondaires moindre • Effets secondaires moins graves
Facilité d'utilisation plus grande	<ul style="list-style-type: none"> • Voie d'administration plus simple (par exemple orale par rapport à intraveineuse, bolus à la place d'une perfusion)

	<ul style="list-style-type: none"> • Une administration par jour à la place de plusieurs ou dose unique à la place d'un traitement de plusieurs jours • Absence d'ajustement de dose
Moins d'inconvénients	<ul style="list-style-type: none"> • Traitement médical à la place d'un traitement chirurgical • Chirurgie moins délabrante • Radiothérapie moins prolongée
Coût plus faible	

3 L'impossible démonstration de l'équivalence

Il est impossible de conclure à la stricte équivalence de deux traitements en statistique.

Avec deux traitements strictement identiques évalués dans un essai thérapeutique avec un critère de jugement binaire, un risque ratio de 1 est attendu (signant bien qu'il n'y a pas eu de différence sur le critère de jugement entre les 2 bras). Cependant, même si l'estimation ponctuelle est en faveur de l'équivalence, il y aura forcément un intervalle de confiance autour de cette valeur qui déborde aussi bien de la supériorité et de l'infériorité d'un traitement par rapport à l'autre. Il est impossible en effet d'avoir un intervalle de confiance de largeur nulle.

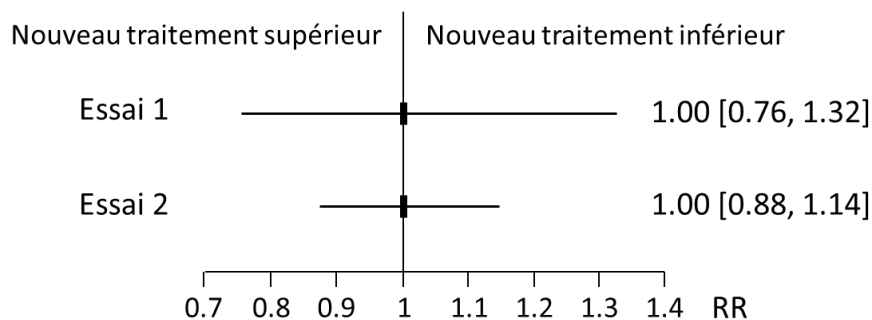
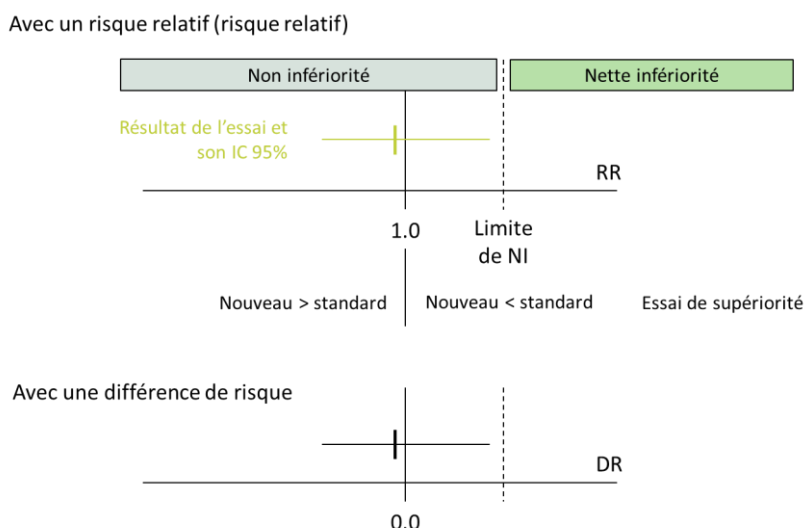


Figure 1 – Illustration de l'impossibilité de démontrer la stricte équivalence de 2 traitements.

Avec un nouveau traitement ayant la même efficacité que le traitement comparateur, un risque ratio (RR) de 1 est attendu. Mais son intervalle de confiance, plus ou moins large, débordera des 2 cotés, entre autres du côté d'une moindre efficacité. Prouver la stricte équivalence nécessiterait une largeur de confiance de largeur nulle, ce qui est impossible.

Cet obstacle est contourné par la recherche d'une « équivalence » relative, qui consiste à montrer que deux traitements ne sont pas trop différents, et que cette différence reste inférieure à un seuil préalablement fixé. Ce seuil correspond à la quantité d'efficacité que l'on peut consentir de perdre, étant donnés les autres avantages du nouveau traitement. L'introduction de cette tolérance rend la démonstration possible et conduit au concept de non-infériorité.

Ainsi la limite de non-infériorité délimite deux zones :



3.1 Seuil de non-infériorité

La décision de conclure à la non-infériorité (du nouveau traitement par rapport au traitement de référence) s'effectuera en comparant la borne supérieure de l'intervalle de confiance avec le seuil de non-infériorité choisi (Figure 2). Si cette borne est inférieure à ce seuil, il est possible de conclure à la non-infériorité avec un risque d'erreur contrôlé. En effet, toutes les vraies valeurs probables de la différence d'efficacité du nouveau traitement par rapport au traitement de référence sont inférieures au seuil préalablement fixé. Cependant si la borne supérieure est supérieure au seuil, il n'est pas possible d'exclure que le nouveau traitement soit moins efficace que le traitement de référence.

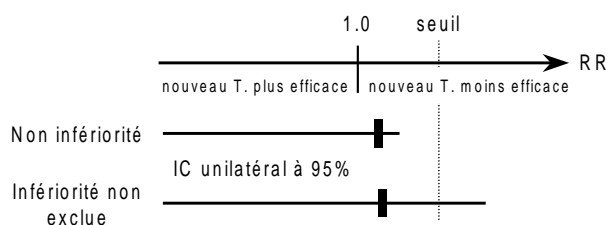


Figure 2 – Illustration du processus de décision de non-infériorité. La borne supérieure du premier intervalle de confiance est inférieure au seuil de non-infériorité choisi. Le nouveau traitement peut être considéré comme non inférior avec un risque statistique d'erreur de 2.5%. Le second intervalle de confiance englobe le seuil de non-infériorité. Dans ce cas, il n'est pas possible d'exclure que le nouveau traitement soit en fait inférieur (moins efficace) que le traitement de référence.

La conclusion de l'essai (démonstration ou non de la non-infériorité) dépend donc de la valeur du seuil. Cette valeur doit donc avoir été choisie a priori, indépendamment des résultats obtenus. Il n'est donc pas possible de transformer un essai de supériorité non-concluant en essai de non-infériorité.

3.2 Intervalle de confiance unilatéral ou bilatéral

Un intervalle de confiance unilatéral à 97.5% est utilisé. Cet intervalle de confiance correspond à un risque alpha de 2.5%. Cette valeur a été choisie pour être cohérente avec ce qui se passe dans l'essai

de supériorité. En effet avec un test bilatéral et un risque alpha de 5% (test classiquement utilisé pour les essais de supériorité), le risque alpha rattaché à la conclusion de supériorité est de 2.5% (cf. test unilatéraux/bilatéraux). Ainsi dans un essai de supériorité, le risque alpha de conclure à tort à la supériorité est de 2.5% (et le risque de conclure à tort à l'infériorité est aussi de 2.5%, ce qui au total fait 5% pour le risque global de conclure à tort). Avec un intervalle unilatéral à 97.5% dans l'essai de non-infériorité, le risque de conclure à tort à la non-infériorité est donc aussi de 2.5%, ce qui assure une cohérence entre les 2 approches.

En pratique, cependant, un intervalle de confiance bilatérale à 95% est utilisé, car la borne supérieure de cet intervalle est numériquement la même que celle de l'intervalle de confiance unilatérale à 97.5%.

3.3 Signification clinique du seuil de non-infériorité

La signification clinique du seuil de non-infériorité correspond à la plus grande perte d'efficacité par rapport au traitement de référence que l'on consent.

Par exemple, un seuil relatif de 10% (soit un risque relatif seuil de 1.1) signifie que l'on considérera le nouveau traitement comme « équivalent » (non inférieur) tant que son efficacité ne sera pas inférieure, en relatif, de 10% à celle du traitement de référence. Au maximum, le nouveau traitement, déclaré comme « équivalent », pourra entraîner une augmentation relative de la fréquence du critère de jugement de 10%.

Il est aussi possible de raisonner en différence de risque absolue. Une différence de risque de +3% signifie que l'on accepte 3 événements de plus tous les 100 patients traités par rapport à ce qui aurait été possible d'éviter avec le traitement de référence. Attention, avec les différences de risques, la signification clinique dépend aussi de la fréquence de l'événement avec le traitement standard. Si cette fréquence est de 3%, une limite de +3%, revient à accepter un doublement de la fréquence du critère de jugement comme sans conséquence clinique. Une différence de risque doit toujours être interprétée par rapport à la fréquence observée dans le groupe contrôle (qui peut être nettement inférieure à celle prévue dans le calcul de l'effectif). En effet, une différence absolue peut être acceptable par rapport à une fréquence élevée. Mais si dans l'essai cette fréquence s'avère être bien plus faible, la même différence absolue représente une augmentation relative bien plus importante que prévu.

Apparaît ici l'ambiguïté du terme non inférieur (ou « équivalent »). Si le seuil choisi correspond à une perte importante d'efficacité, dire que le nouveau traitement est non inférieur est clairement un abus de langage.

En fait, l'essai de non-infériorité ne démontre pas stricto sensu l'équivalence. Il permet simplement d'exclure que, par rapport au traitement de référence, l'efficacité du traitement étudié soit inférieure à une certaine limite. On ne peut pas dire que l'équivalence est démontrée, car il est possible que le nouveau traitement soit moins efficace que le traitement de référence. Cependant, malgré cela, il peut être acceptable d'utiliser ce nouveau traitement en pratique même s'il n'est pas exclu qu'il soit moins efficace que le traitement de référence, car il présente d'autres avantages par ailleurs. En pratique pour les essais positifs, les nuances existant dans l'interprétation de la conclusion ont tendances à disparaître quand il est souvent conclu, un peu rapidement, que le traitement étudié est *équivalent*. Cette formulation, surtout pour le béotien dans le domaine, évoque irrémédiablement l'identité, et amène à penser que les traitements sont interchangeable et conduisent au même résultat.

Les conclusions des articles utilisent fréquemment des tournures trompeuses comme « as effective as ».

En fait, un nouveau traitement montré comme équivalent au traitement de référence doit être considéré comme inférieur, jusqu'à preuve du contraire apporté par un essai de supériorité. Dans un classement hiérarchique des traitements par efficacité décroissante, le nouveau traitement arrive en seconde position derrière le traitement de référence.

La valeur du seuil de non-infériorité conditionne le nombre de sujets à inclure dans un essai de non-infériorité. En effet, plus le seuil est petit, plus la précision nécessaire à la démonstration de la non-infériorité devra être importante. Donc plus l'effectif de l'essai devra être important. Ainsi, il n'est pas réaliste de fixer arbitrairement le seuil à une valeur très petite. La valeur du seuil devra être choisie en fonction de la pathologie, du critère de jugement et de la nature et de l'importance des avantages apportés par le nouveau traitement.

4 Équivalence clinique ou non-infériorité

Non-infériorité et équivalence sont deux notions très proches. La non-infériorité correspond à une équivalence unilatérale, tandis que l'équivalence vraie est bilatérale (le nouveau traitement peut ne pas être équivalent au traitement de référence en étant soit moins efficace soit plus efficace).

Pour l'efficacité clinique, la recherche de l'équivalence est, sauf cas exceptionnel, une situation unilatérale. En effet, le nouveau traitement sera intéressant en pratique aussi bien s'il s'avère « équivalent » ou supérieur au traitement de référence, mais ne sera pas utilisé s'il s'avère inférieur. Cette description correspond à une situation unilatérale où l'on cherche à montrer avec un risque d'erreur alpha contrôlé que le nouveau traitement ne soit pas inférieur au traitement contrôle, c'est-à-dire en d'autres termes qu'il est au moins aussi efficace. L'utilisation d'un test bilatéral ferait courir le risque de ne pas conclure dans une situation où l'on ne pourrait pas considérer le nouveau traitement comme équivalent, car il serait en fait supérieur. Cette conclusion serait paradoxale et gênante en pratique, car elle entraînerait l'impossibilité de recommander un traitement en fait, supérieur à l'existant.

Très souvent, par abus de langage, le terme équivalence est employé pour désigner ce qui en réalité est une non-infériorité. Cela vient du fait que dans le domaine de l'évaluation du bénéfice clinique, le nouveau traitement peut être substitué à l'ancien à partir du moment où la non-infériorité est démontrée. La bioéquivalence est une situation d'équivalence vraie bilatérale. Il y a bioéquivalence lorsque les paramètres pharmacocinétiques du nouveau médicament ne sont ni plus ni moins élevés que ceux du traitement de référence.

5 Le choix du seuil de non-infériorité

5.1 Introduction

Contrairement à la conclusion à la supériorité, la décision de conclure à la non-infériorité repose sur une valeur arbitraire : le seuil de non-infériorité. À partir des mêmes données, la conclusion à la non-infériorité pourra être faite ou non en fonction du seuil choisi (Figure 3). Un seuil « tolérant » facilite l'obtention de la non-infériorité.

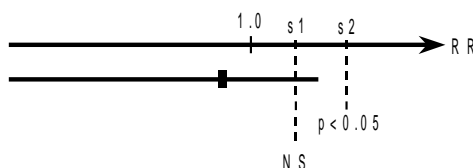


Figure 3 – En fonction de la valeur du seuil choisi, le même résultat peut conduire à un test de non-infériorité statistiquement significatif (seuil s2) ou non statistiquement significatif (seuil s1). L'obtention d'un résultat significatif demande d'accepter une plus grande perte d'efficacité.

Le choix de la limite est la principale difficulté de l'essai d'équivalence clinique. Quelle perte d'efficacité sur la mortalité justifie une chirurgie moins délabrante en cancérologie ou un traitement fibrinolytique en double bolus dans l'infarctus du myocarde à la place d'une perfusion de 90 minutes ? Le plus souvent le choix est arbitraire, reflète une échelle de valeurs, et conduit à des seuils parfois discutables. Ainsi la littérature contient plusieurs exemples dans lesquels la valeur de la limite de non-infériorité s'avère très tolérante et pour le moins discutable. Même si un essai conclut de façon statistiquement significative à la non-infériorité d'un nouveau traitement, il est tout à fait possible de rejeter cette conclusion si l'on considère que le seuil utilisé était trop tolérant. Contrairement au résultat d'un test statistique de différence qui ne peut être remis en cause (zéro est zéro, sans discussion possible).

Dans un essai comparant une héparine de bas poids moléculaire (HBPM) à une héparine non fractionnée (HNF), le traitement de référence, dans le traitement des thromboses veineuses profondes symptomatiques. La fréquence du critère de jugement (récidives thromboemboliques) attendu sous HNF était de 7 à 8%. La limite a été fixée à 5% en termes de différence absolue, ce qui signifiait que l'HBPM allait être déclarée équivalente tant qu'elle n'entraînerait pas une fréquence de récurrence de 12 à 13%. Cette limite absolue de 5% correspond en fait, à une augmentation relative de 66% de la fréquence attendue dans le groupe HNF. Ainsi les investigateurs étaient prêts à accepter que le nouveau traitement puisse multiplier par 1,66 la fréquence des récurrences, parmi lesquelles figure l'embolie pulmonaire. Un autre argument suggère que cette limite absolue de 5% est exagérée. L'HNF entraîne une réduction absolue de l'ordre de 13% par rapport au placebo. Avec la limite de 5%, on accepte une perte de plus d'un tiers (38%) du bénéfice apporté par l'HNF. Heureusement, le résultat de l'essai conduit à un intervalle de confiance dont la borne supérieure est inférieure à cette limite (1,07).

5.2 Positionnement par rapport à l'efficacité du traitement de référence

Le choix du seuil de non-infériorité peut être facilité par l'étude du bénéfice qu'apporte le traitement de référence. Le seuil sera fixé de telle façon à ne pas permettre que l'utilisation du nouveau traitement conduise à perdre l'avancée thérapeutique représentée par le traitement de référence. Entre autres, si le traitement de référence a validé son efficacité contre placebo, le seuil choisi garantira que le nouveau traitement ne puisse pas être moins bon que le placebo.

Cette démarche a été utilisée dans l'essai COBALT [5]. L'objectif de cet essai de fibrinolyse à la phase aiguë de l'infarctus du myocarde était de comparer un nouveau traitement représenté par un double bolus d'alteplase au traitement de référence, la perfusion accélérée d'alteplase. La perfusion accélérée est justifiée par les résultats de l'essai GUSTO 1 où ce traitement s'est avéré meilleur que le traitement de référence précédent, la streptokinase. Dans cet essai, la mortalité à 30 jours sous streptokinase était de 7,3% contre 6,3% avec l'alteplase. Le gain apporté par la perfusion accélérée d'alteplase est donc, en différence absolue, de 1% avec un intervalle de confiance bilatéral à 95% de 0,4% à 1,6%. Le vrai effet de l'alteplase se situe entre ces deux bornes et au pire, dans la situation la moins favorable, la différence absolue par rapport à la streptokinase n'est que de 0,4%.

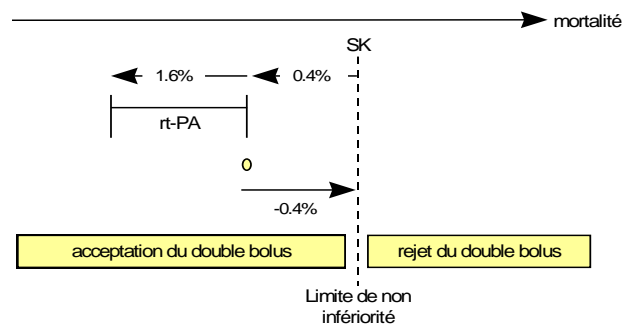


Figure 4 – Illustration du choix de la limite d'équivalence en se basant sur le bénéfice apporté par le traitement de référence. La limite est choisie de telle façon qu'il n'y ait pas de risque de régresser, c'est-à-dire de faire moins bien que le précédent traitement de référence (ici la streptokinase).

En choisissant comme limite cette valeur on est assuré que le nouveau traitement ne pourra pas être considéré comme équivalent alors qu'il fait moins bien que la streptokinase. C'est-à-dire que l'efficacité moindre que l'on pourrait tolérer en regard des avantages pratiques apportés par le double bolus d'alteplase ne conduise pas à perdre tout le bénéfice de la dernière avancée thérapeutique. La démarche de cet essai est exemplaire et montre qu'une définition rigoureuse et argumentée de la non-infériorité est possible. Elle conduit cependant à des valeurs très contraignantes demandant un nombre de sujets de même ordre de grandeur qu'un essai de supériorité.

Dans cet exemple, la perte de 100% de l'efficacité de la perfusion accélérée d'alteplase est acceptable, car, même si l'efficacité du double bolus n'est que celle de la streptokinase, le double bolus représente encore un progrès thérapeutique en termes de praticabilité et de meilleure tolérance, car il ne présente pas les effets allergiques de la streptokinase.

Un autre exemple du même domaine illustre que le choix de la limite est fréquemment subjectif. L'essai GUSTO 3 [6] comparait un nouveau fibrinolytique, la reteplase qui s'administre en double bolus au

traitement de référence qu'est la perfusion accélérée d'alteplase. À l'origine, cet essai était un essai de supériorité dont les résultats ne permettaient pas de conclure à la supériorité de la reteplase. Devant ce résultat, l'objectif de l'étude a été transformé en recherche de l'équivalence clinique. Ce changement post-hoc d'hypothèse pose un problème sur lequel nous reviendrons. Pour l'instant intéressons-nous à la limite de non-infériorité qui a été choisie. Cette limite a été choisie comme dans COBALT par référence à GUSTO 1, mais là où COBALT intégrait l'incertitude statistique en prenant la borne inférieure (0,4%) de l'intervalle de confiance de GUSTO 1, GUSTO 3 considère l'estimation moyenne de 1%. La borne supérieure de l'intervalle obtenue dans GUSTO 3 est de 0,66%, valeur qui autorise de conclure à l'équivalence avec la limite de 1%, mais pas avec celle de 0,4%. Les publications de COBALT et GUSTO 3 sont parues dans le même numéro du New England Journal of Medicine et ce contraste entre des choix différents illustre le manque de règles adoptées par tous et les dérives qui peuvent exister.

GUSTO 3 pose aussi la question de savoir s'il est licite de passer d'un objectif de supériorité à un objectif d'équivalence en fonction des résultats. À cause de ce changement post-hoc, le principe de la méthode expérimentale suivant lequel l'hypothèse doit être générée avant le recueil des données (« l'hypothesis testing » de Fisher) n'est pas respecté. L'essai ne peut pas être considéré comme démontrant l'hypothèse d'équivalence étant donné qu'il n'avait pas été conçu pour cela, mais pour démontrer l'hypothèse inverse. Dans cette situation, conclure à l'équivalence est une démarche tautologique. Les données servant à la démonstration sont celles qui ont fait générer l'hypothèse. De plus rien n'assure qu'un essai de supériorité répond au critère de qualité de l'essai d'équivalence et que le comparateur soit adéquat.

Ces restrictions méthodologiques sont prises en compte dans la discussion, mais néanmoins la conclusion de GUSTO 3, bien que prudente, suggère fortement l'équivalence.

Cette approche peut être utilisée avec le bénéfice absolu (la différence des risques) ou le bénéfice relatif (risque relatif). Bien que fréquemment utilisée jusqu'à présent, la différence absolue ne tient pas compte du risque de base. Une différence absolue limite fixée à 1% a priori sur l'hypothèse que le risque de base est de 10% correspond à un risque relatif limite de 10%. Si l'essai inclut des patients à faible risque conduisant à un risque de base de 5%, la même limite absolue de 1% correspond alors à un risque relatif limite de 20%. La tendance actuelle est d'utiliser préférentiellement le risque relatif.

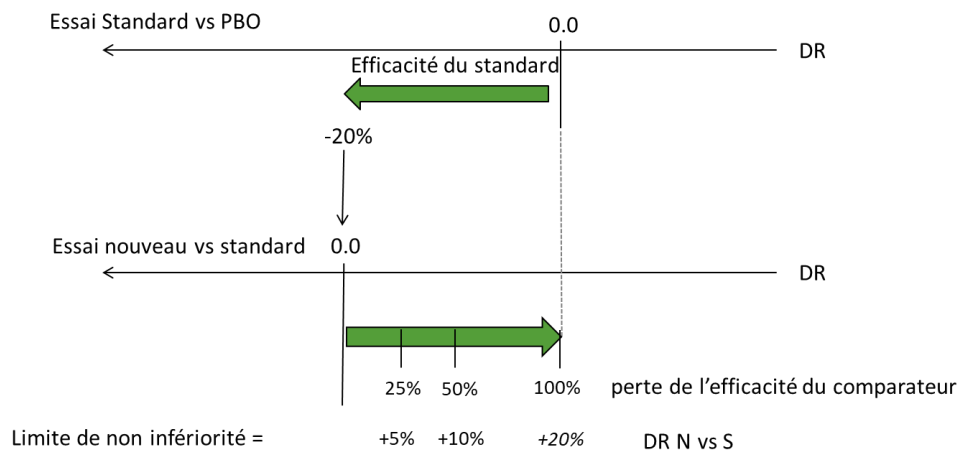
5.3 Calcul du seuil

En suivant les principes énoncés précédemment, le calcul du seuil se déroule de la façon suivante.

Considérons un essai de non-infériorité sur la mortalité du traitement N par rapport au traitement A, dans lequel on souhaite conserver 75% de l'efficacité de A. Ainsi l'éventuelle perte d'efficacité acceptable avec N ne devra pas être supérieure à 25% de l'efficacité de A.

Pour le calcul du seuil, il convient en premier de connaître l'efficacité de A par rapport à son propre contrôle. Dans le cas où A a été évalué par rapport au placebo, ce résultat peut être extrait d'une étude antérieure ou encore d'une méta-analyse des essais de A versus placebo.

Par exemple si la méta-analyse des essais de A versus placebo donne comme estimation une différence de risque de -25% avec un intervalle de confiance à 95% entre -30% et -20%.



La plus petite efficacité « garantie » avec A, donnée par la borne péjorative de l'intervalle de confiance, correspond donc à une réduction absolue de la fréquence du critère de jugement de 20%.

Vouloir la préservation de 75% de cette efficacité revient donc à autoriser au maximum une perte maximum de 25% (1-75%) de cette efficacité. Le calcul du seuil s'obtient donc par le calcul suivant :

$$20\% * (1-75\%) = +5\%.$$

Pour être considéré comme non inférieur, le nouveau traitement (N) ne devra pas entraîner une augmentation absolue du critère de jugement de plus de 5% par rapport à A.

L'essai ARISTOTLE[9] est un essai comparant l'apixaban à la warfarine dans la FA. La limite de non-infériorité a été déterminée pour préserver 50% de l'efficacité de la warfarine par rapport au placebo sur le critère AVC ou embolies systémiques :

The primary noninferiority hypothesis required that apixaban preserve at least 50% of the relative reduction in the risk of stroke or systemic embolism associated with warfarin (62%) in six previous, major randomized, controlled trials. This hypothesis provided a lower 95% confidence interval of 1.88 for the relative risk with placebo as compared with warfarin, and one half of this value was 1.44 (or 1.38 on a log scale).

A priori, l'apixaban présentait comme avantage par rapport à la warfarine de ne pas nécessiter d'ajustement de la dose par INR. Pour les concepteurs de cet essai, cet avantage justifiait de perdre jusqu'à la moitié de l'efficacité de la warfarine. Finalement l'apixaban a été montré supérieur.

5.4 Limite en différence de risque ou en risque ratio

Bien que simple, le raisonnement en différence absolue doit être évité. En effet, l'acceptabilité d'un seuil en différence absolue dépend du risque de départ. Dans l'exemple précédent, dans l'essai A versus placebo les risques étaient respectivement de 25% et 50%. Par rapport au risque obtenu sous A (25%) le seuil calculé précédemment correspond à une augmentation relative acceptée de $5\%/25\% = +20\%$. Si dans l'essai de non-infériorité on obtient un risque sous A plus faible le seuil ne représentera plus la même perte d'efficacité en relatif. Cette situation est fréquente car l'essai de non-infériorité se déroule quelques années plus tard que l'essai A versus placebo, chez des patients qui bénéficient d'autres avancées thérapeutiques concourant à diminuer encore davantage leur risque. Si dans l'essai de non-infériorité le risque avec A est de 10%, le seuil de +5% représente une augmentation relative de $5\%/10\% = +50\%$, soit plus du double ! Pour éviter cette situation, il convient

de raisonner directement en risque relatif, ce qui garantira la même augmentation relative consentie, quel que soit le risque obtenu avec A dans l'essai de non-infériorité.

6 Situation actuelle

Une partie importante des essais de non-infériorité publiés utilisent des limites de non-infériorité excessivement larges.

6.1 Étude de cas 1

L'essai SURPRISE[4] a évalué en non-infériorité le rivaroxaban par le fondaparinux dans le traitement de la thrombose veineuse superficielle. Le fondaparinux a été validé en son temps avec l'essai CALISTO[10]. Le critère de jugement est un critère composite regroupant thrombose veineuse profonde, embolie pulmonaire, extension de la thrombose veineuse superficielle, et les décès.

Une limite de non-infériorité de +4.5% en différence absolue (différence de risque) a été utilisée :

We used a non-inferiority margin of 4.5% (absolute difference between rivaroxaban and fondaparinux). [abstract]

L'article conclut à la non-infériorité, mais sans rapporter les résultats supportant cette conclusion. Les résultats sont donnés sous forme de hazard ratio et non pas de différence de risque. Il est donc impossible de savoir si la borne supérieure de l'intervalle de confiance de la différence des risques est bien en deçà de la limite de +4.5%.

La limite de non-infériorité n'est pas justifiée dans l'article. Les arguments proposés n'ont rien à voir avec ce qui est attendu comme justification, c'est-à-dire le pourcentage de l'efficacité du fondaparinux que préserve la limite.

This margin was chosen based on the considerations that both the STENOX trial and the POST registry noted that patients who received anticoagulants had events of 5–8% for an endpoint similar to the proposed primary endpoint of the SURPRISE trial.

Cette justification ne se réfère pas à l'efficacité du fondaparinux et présente des valeurs de fréquence d'événements sous anticoagulants, ce qui ne permet en rien de déterminer la perte maximale d'efficacité par rapport au fondaparinux qui peut être accepté.

Dans cet essai, la détermination de la limite de non infériorité aurait dû se faire par rapport à l'efficacité chez ces patients du fondaparinux qui est documenté dans l'essai CALISTO [10]. Les résultats de CALISTO sont présentés dans le tableau suivant :

Efficacy Outcome	Fondaparinux (N=1502) no. with event (%)	Placebo (N=1500)	Absolute Risk Reduction with Fondaparinux percentage points (95% CI)	Relative Risk with Fondaparinux % (95% CI)	P Value*
By Day 47					
Primary composite outcome†	13 (0.9)	88 (5.9)	-5.0 (-6.3 to -3.7)	0.15 (0.08 to 0.26)	<0.001

Ainsi le bénéfice minimal garanti du fondaparinux versus placebo est une différence des risques de -3.7% (borne supérieure de l'intervalle de confiance). La limite de non-infériorité maximale possible est donc de +3.7%. La limite utilisée dans SURPRISE correspond donc à une perte de plus de 100% de l'efficacité du comparateur.

Comme les résultats de SUPRISE ne sont donnés qu'en termes de hazard ratio, il est aussi possible de calculer la limite en relatif avec les résultats de CALISTO. En relatif, le bénéfice minimal garanti avec un degré de certitude de 95% est un RR de 0.26. La limite maximale est donc $1/0.26 = 3.84$. Le résultat de surprise est un HR 1.9; 95% CI 0.6–6.4 ; qui ne permet, en aucun cas de conclure à la non-infériorité, car la borne supérieure de l'intervalle de confiance est supérieure à la limite de 3.84 correspondants à la perte de la totalité du bénéfice du fondaparinux.

Cet exemple illustre bien les problématiques couramment rencontrées avec les essais de non-infériorité : absence de justification des limites de non-infériorité utilisée, utilisation de limite autorisant une perte de la totalité de l'efficacité du comparateur, conclusion induite à la non-infériorité, évaluation en non-infériorité de nouveau traitement ne présentant pas davantage par ailleurs par rapport au standard ce qui enlève toute justification à la perte d'efficacité consentie.

6.2 Étude de cas 2

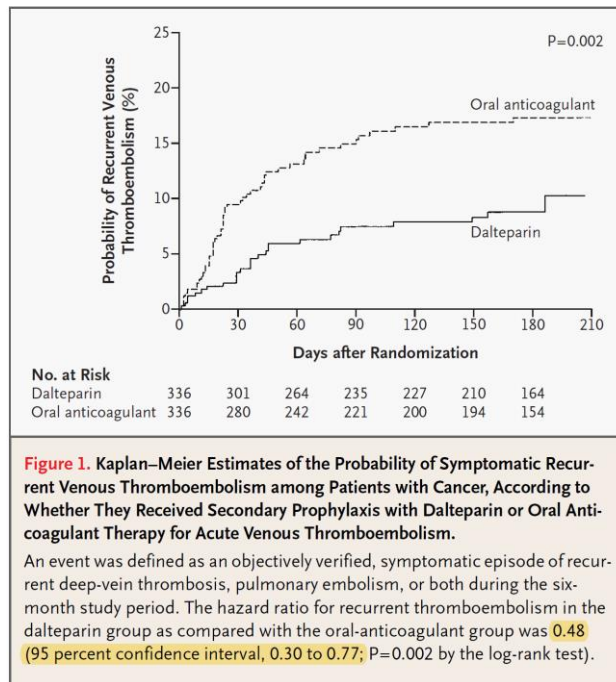
L'essai Hokusai VTE Cancer[5] évalue en non-infériorité l'edoxaban dans le traitement des épisodes thromboemboliques aigus chez des patients atteints d'un cancer. Le comparateur est la dalteparin qui a été validée en son temps (essai CLOT[11]). Le critère de jugement est un critère composite regroupant les récurrences thrombo-emboliques et les hémorragies. La limite de non-infériorité utilisée est un hazard ratio de 1.5 :

The trial hypothesis was that edoxaban would be noninferior to dalteparin with respect to the rate of primary-outcome events (recurrent venous thromboembolism or major bleeding), with an upper limit of the 95% confidence interval for the hazard ratio of less than 1.5 and a two-sided alpha level of 0.05.

Cette limite qui correspond à l'acceptation d'une augmentation relative de 50% des événements, est « justifiée » de manière complètement incorrecte :

The margin of 1.5 was chosen as the maximum difference that may be potentially clinically acceptable because of the unmet need for an alternative to parenteral low-molecular-weight heparin and the advantages of oral therapy.

À partir de CLOT il apparaît que l'effet minimal garanti de la dalteparine par rapport à la coumarine est un hazard ratio de 0.77 (HR 0.48 IC 95% 0.30, 0.77) :



La limite de non-infériorité qui correspond à la perte de la totalité de la supériorité de la dalteparine par rapport à la coumarine (qui est un anticoagulant oral) est $1/0.77 = 1.30$, valeur inférieure à la limite utilisée de 1.5. Ainsi cet essai autorise à conclure que l'édoxaban est non inférieur à la dalteparine même si celui-ci entraîne la perte de la totalité de l'efficacité apportée par la dalteparine par rapport à la coumarine, qui présente d'ailleurs exactement le même avantage d'être une prise orale que l'édoxaban.

De plus le critère de jugement composite utilisée dans cet essai combine récurrences des événements thromboemboliques et hémorragies, ce qui conduit à perdre à la fois sur le plan de l'efficacité et de la sécurité. L'utilisation de critère composite combinant des éléments qui ne sont pas des équivalents cliniques les uns des autres est problématique en non-infériorité[2].

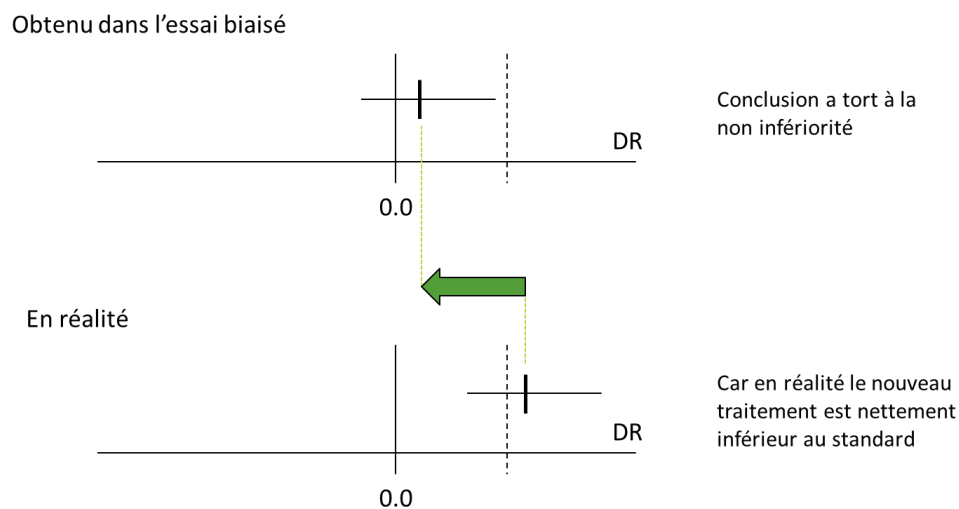
7 Enchaînement d'évaluations en non-infériorité

On assiste aussi à des évaluations successives en non-infériorité dans certains domaines comme la thrombo-prophylaxie en chirurgie. Un nouveau traitement (de 3^{ème} génération, G3) s'évalue en non-infériorité par rapport à un traitement de 2^{ème} génération (G2) déjà lui-même évalué en non-infériorité par rapport à la 1^{er} génération (qui elle avait montré sa supériorité au placebo). À chaque évaluation en non-infériorité une perte d'efficacité par rapport au comparateur est consentie, et peut se cumuler entre elles. Cette succession de perte d'efficacité potentielle expose au risque d'accepter comme non-inférieur lors de la dernière évaluation un produit dont l'efficacité pourrait être proche de celle du placebo. Par exemple, les anticoagulants directs oraux ont été évalués en non-infériorité par rapport aux HBPM, qui elles-mêmes avaient été évaluées en non-infériorité par rapport à l'HNF. L'obtention de supériorité à la place de la non-infériorité recherchée et les méta-analyse en réseau permettent d'exclure que cette 3^{ème} génération soit sans intérêt, mais le plan de développement ne permettait pas de l'assurer par construction.

8 Méthodologie, le contrôle des biais dans l'essai de non-infériorité

Le but de la méthodologie est d'éviter les biais. Les biais encourus par la recherche de la non-infériorité sont différents de ceux qui peuvent survenir dans les essais de supériorité. La méthodologie doit donc être adaptée et elle diffère des principes classiques de l'essai de supériorité.

En termes de biais, la situation redoutée avec l'essai de non-infériorité est de conclure à tort à la non-infériorité alors qu'en réalité le nouveau traitement est nettement inférieur. L'essai est alors faussement « positif ». La réelle différence en défaveur du nouveau traitement, qui aurait dû être observée entre les 2 groupes, a été réduite par les biais présents dans l'essai. Ces biais ont entraîné un décalage du résultat et de son intervalle de confiance vers 0 ou 1 (suivant qu'une différence ou un effet relatif est utilisé¹).



Dans l'essai de supériorité, un résultat faussement positif en raison de biais sera obtenu lorsque les biais augmentent la différence entre les 2 groupes (en favorisant le traitement évalué) et transforme ainsi une réelle absence de différence en apparente différence.

Il apparaît ainsi que les biais gênant dans l'essai de non-infériorité sont différents que ceux redoutés dans l'essai de supériorité. Dans l'essai de supériorité, les biais pouvant conduire à un résultat faussement positif sont des biais augmentant la différence entre les 2 groupes. Dans l'essai de non-infériorité, ce sont les biais réduisant la différence qui posent problème.

8.1 Retour sur le principe de protection contre les biais dans l'essai de supériorité

Plusieurs moyens mis en œuvre pour protéger l'essai de supériorité des biais ont pour principe de réduire la différence entre les 2 groupes afin d'éviter la survenue d'une fausse différence par biais. Ces

¹ L'effet nul (l'absence d'effet) se traduit par une différence de 0 ou un risque relatif (odds ratio, risque ratio ou hazard ratio) de 1.

principes méthodologiques sont ainsi conservateurs et prennent le parti pris de réduire les différences entre les 2 groupes plutôt que de laisser apparaître une différence à tort en raison d'un biais.

C'est le cas avec le double aveugle qui supprime le biais de réalisation que pourrait introduire l'utilisation de traitement concomitant ou de secours en faisant de sorte que la fréquence d'utilisation de ces traitements soit identique entre les 2 groupes. Si cette fréquence est importante cela fait que les 2 groupes, y compris le groupe placebo, sont traités, ce qui conduit à réduire la différence entre les 2 groupes si le traitement étudié est supérieur au control. Ainsi, le double aveugle protège contre le biais en assurant que la conséquence sur la valeur du critère de jugement de l'utilisation de traitements concomitants soit conservatrice. En double aveugle, en aucun cas les traitements concomitants sont susceptibles de favoriser le groupe expérimental. Au pire, le double aveugle handicape la mise en évidence d'une différence. En supériorité, le double aveugle peut donc conduire à un résultat faussement négatif, mais en aucune façon à un résultat faussement positif.

Ainsi les principes méthodologiques qui mettent l'essai de supériorité à l'abri des biais ne marchent pas pour les essais de non-infériorité. D'autres éléments sont donc nécessaires pour éviter les biais dans l'essai de supériorité.

8.2 Les biais de l'essai de non -infériorité et leur control

La conclusion d'un essai de non-infériorité peut être biaisée si l'efficacité développée par le traitement de référence a été moindre que ce qu'elle aurait dû être. Dans ce cas, un nouveau traitement pourtant nettement moins efficace que le traitement de référence apparaîtra à tort équivalent. Le traitement intrinsèquement le plus efficace a été ramené au niveau du moins efficace.

Tableau 2 – Liste des biais possibles

1) L'efficacité du traitement de référence est altérée
a) Le traitement de référence n'est pas administré correctement : dose trop faible ou trop forte, durée du traitement trop courte, etc.
b) Le traitement de référence est administré à des patients chez lesquels il est moins ou pas efficace
c) Le traitement de référence n'est pas le meilleur traitement possible (<i>problème d'interprétation plus que biais</i>)
d) Le traitement de référence est facilement arrêté pour effets indésirables
2) L'efficacité du nouveau traitement est renforcée par des traitements concomitants

La méthodologie de l'essai de non-infériorité doit donc veiller à ce que le traitement de référence développe correctement toute son efficacité (administration correcte) et que l'estimation de son effet reflète bien sa véritable efficacité (sensibilité et spécificité correctes de la mesure du critère). La confirmation de l'absence d'un biais à ce niveau pourrait être obtenue en incluant dans l'essai un bras placebo pour s'assurer que l'efficacité du traitement de référence est bien celle attendue. Cependant, l'emploi d'un placebo dans une situation où il existe un traitement de référence est rarement possible. Une validation externe est nécessaire.

L'utilisation de traitements concomitants peut aussi entraîner un biais. Si une forte proportion des patients des deux groupes reçoit des traitements concomitants efficaces, l'efficacité observée sera identique dans les deux groupes. Mais il ne s'agira pas de l'efficacité propre des traitements testés, mais celle des traitements concomitants. Une équivalence sera observée même si le nouveau traitement est moins efficace.

Dans un essai de non-infériorité, l'analyse en intention de traité est moins à l'abri de biais que l'analyse per-protocole. L'analyse en intention de traiter favorise l'hypothèse testée. Par contre, l'analyse per-protocole est conservatrice. En effet, plusieurs composantes de l'analyse en intention de traiter sont susceptibles de réduire la mesure de l'efficacité des traitements dans les deux groupes et en particulier celle du traitement de référence. La différence entre les groupes tend donc à diminuer, ce qui, dans le cas de l'essai d'équivalence, favorise l'hypothèse testée. Il en résulte donc un biais.

Contrairement à l'essai de supériorité, l'analyse potentiellement la moins biaisée dans l'essai de non-infériorité est l'analyse en per-protocole où seuls les patients traités en stricte conformité avec le protocole sont maintenus dans l'analyse. En pratique, une conclusion sûre n'est possible que lorsque les analyses en intention de traiter et per-protocole donnent des résultats similaires.

L'analyse en intention de traiter évalue l'équivalence des stratégies thérapeutiques.

L'analyse per-protocole évalue l'équivalence des traitements à l'intérieur des stratégies thérapeutiques.

Tableau 3 – Situations introduisant un biais vers l'absence d'effet à travers l'analyse en intention de traiter. En fait, toutes ces situations conduisent aux mêmes conséquences : la dilution et la convergence des effets.

Situations	Effet sur la différence entre les deux groupes
Patients qui n'ont pas reçu le traitement alloué	Égalisation des traitements reçus dans les 2 groupes
Arrêt prématuré du traitement de l'étude	
Déviations au protocole, administration de traitement concomitant interdit	Le groupe du traitement testé reçoit des traitements aussi efficaces que le traitement de référence
Patients inclus à tort	Patients ne pouvant pas répondre au traitement, car insensibles aux différences de traitement reçu entre les 2 groupes

Prise en compte des perdus de vue comme des valeurs manquantes	Dilution et convergence des effets des traitements
--	--

Par exemple, si le nouveau traitement est inférieur, le recours aux traitements concomitants pour échec du traitement testé sera plus fréquent dans le groupe du nouveau traitement que dans celui du traitement de référence. L'efficacité du nouveau traitement sera renforcée par celle des autres traitements et il pourra apparaître équivalent dans une analyse en intention de traiter. Par contre, dans une analyse per-protocole où les patients qui ont arrêté précocement le traitement testé ont été exclus, l'insuffisance d'efficacité du nouveau traitement apparaîtra.

Au total, l'essai de non-infériorité est extrêmement sensible à sa qualité méthodologique. Un fort taux de perdus de vue, d'écarts au protocole ou d'arrêts prématurés des traitements risque d'égaliser l'efficacité dans les deux groupes et de biaiser le résultat.

9 Méta-épidémiologie des essais de non-infériorité

Les essais de non-infériorité, tels que réalisés et publiés, ont fait l'objet de plusieurs travaux de méta-épidémiologie.

Les limites de non-infériorité utilisées correspondent souvent à une perte de plus de 100% de l'efficacité du comparateur [12, 13].

Les essais de non-infériorité sont de plus en plus utilisés[1], et dans certains domaines comme la thrombose, sont quasi exclusifs, pouvant faire craindre une faible évaluation des traitements récents [1, 14].

Globalement ces essais sont de faible qualité méthodologique et assez mal rapportée [14, 15, 16, 17].

Ainsi, ces études de méta-épidémiologie montrent que les essais de non-infériorité couramment réalisés présentent fréquemment des limites méthodologiques importantes : absence de justification des limites de non-infériorité utilisée, utilisation de limite autorisant une perte de la totalité de l'efficacité du comparateur, conclusion induite à la non-infériorité, évaluation en non-infériorité de nouveau traitement ne présentant pas davantage par ailleurs par rapport au standard ce qui enlève toute justification à la perte d'efficacité consentie.

10 Absence de différence dans un essai de différence (supériorité)

Lorsque dans un essai de supériorité celle-ci n'est pas mise en évidence de façon significative, il peut être tentant de conclure à l'équivalence. Cette conclusion n'est pas possible, car elle pose plusieurs problèmes.

La puissance est peut-être insuffisante. L'absence de différence significative ne signifie pas qu'il y a absence de différence, mais peut-être, simplement, que l'essai était insuffisamment puissant pour mettre en évidence la réelle différence qui existe entre deux traitements : « l'absence de preuve n'est pas la preuve de l'absence ».

Conclure à l'équivalence après avoir bâti l'essai pour tester une hypothèse de différence revient à changer d'hypothèse. Le principe de la méthode expérimentale, l'approche « *hypothesis testing* » de Fisher n'est pas respecté. L'essai ne peut pas être considéré comme démontrant l'hypothèse d'équivalence étant donné qu'il n'avait pas été conçu pour cela (mais pour démontrer l'hypothèse inverse). Dans cette situation, conclure à la démonstration de l'équivalence est une démarche tautologique. Les données servant à la démonstration sont celles qui ont fait générer l'hypothèse. De plus, se pose des problèmes de qualité méthodologique de l'essai (les contraintes de l'essai d'équivalence sont différentes de celles de l'essai de supériorité) et de fixation post-hoc de la limite d'équivalence.

Dans un essai incluant 1000 patients par groupe, on observe 30 événements critères de jugement dans chaque groupe. Malgré la stricte identité de ces nombres d'événements, ce résultat est loin de permettre de conclure à l'équivalence d'efficacité. Le risque relatif est bien de 1, mais avec un intervalle de confiance à 95% entre 0,61 et 1,65. Ce qui signifie que ce résultat est compatible avec, en réalité, une augmentation de la fréquence du critère de jugement par le nouveau traitement de 65%. Du fait de cette incertitude statistique, il est donc impossible de conclure à l'équivalence des 2 traitements. En fait, il est toujours impossible de conclure à la stricte équivalence entre 2 traitements, car cela nécessiterait un intervalle de confiance de largeur nulle, ce qui est impossible, car nécessitant une infinité de patients.

Références

- 1 Murthy VL, Desai NR, Vora A, et al. Increasing proportion of clinical trials using noninferiority end points. *Clinical cardiology* 2012;35:522–23 doi:10.1002/clc.22040; PMID:22806427;
- 2 Mauri L, D'Agostino RB. Challenges in the Design and Interpretation of Noninferiority Trials. *N Engl J Med* 2017;377:1357–67 doi:10.1056/NEJMra1510063; PMID:28976859;
- 3 Blackwelder WC. "Proving the null hypothesis" in clinical trials. *Controlled Clinical Trials* 1982;3:345–53 doi:10.1016/0197-2456(82)90024-1; PMID:7160191;
- 4 Beyer-Westendorf J, Schellong SM, Gerlach H, et al. Prevention of thromboembolic complications in patients with superficial-vein thrombosis given rivaroxaban or fondaparinux: the open-label, randomised, non-inferiority SURPRISE phase 3b trial. *The Lancet Haematology* 2017;4:e105-e113 doi:10.1016/S2352-3026(17)30014-5;
- 5 Raskob GE, van Es N, Verhamme P, et al. Edoxaban for the Treatment of Cancer-Associated Venous Thromboembolism. *N Engl J Med* 2018;378:615–24 doi:10.1056/NEJMoa1711948; PMID:29231094;
- 6 Iversen K, Ihlemann N, Gill SU, et al. Partial Oral versus Intravenous Antibiotic Treatment of Endocarditis. *N Engl J Med* 2019;380:415–24 doi:10.1056/NEJMoa1808312; PMID:30152252;
- 7 A comparison of reteplase with alteplase for acute myocardial infarction. *N Engl J Med* 1997;337:1118–23 doi:10.1056/NEJM199710163371603; PMID:9340503;
- 8 Koopman MM, Prandoni P, Piovella F, et al. Treatment of venous thrombosis with intravenous unfractionated heparin administered in the hospital as compared with subcutaneous low-molecular-weight heparin administered at home. The Tasman Study Group. *N Engl J Med* 1996;334:682–87 doi:10.1056/NEJM199603143341102; PMID:8594426;
- 9 Granger CB, Alexander JH, McMurray JJV, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–92 doi:10.1056/NEJMoa1107039; PMID:21870978;
- 10 Decousus H, Prandoni P, Mismetti P, et al. Fondaparinux for the treatment of superficial-vein thrombosis in the legs. *N Engl J Med* 2010;363:1222–32 doi:10.1056/NEJMoa0912072; PMID:20860504;
- 11 Lee AYY, Levine MN, Baker RI, et al. Low-molecular-weight heparin versus a coumarin for the prevention of recurrent venous thromboembolism in patients with cancer. *N Engl J Med* 2003;349:146–53 doi:10.1056/NEJMoa025313; PMID:12853587;
- 12 Pérard L, Hot A, Cucherat M, et al. Essais de non-infériorité dans la maladie thromboembolique veineuse. Une lecture critique est nécessaire! *La Revue de medecine interne* 2007;28:731–36 doi:10.1016/j.revmed.2007.05.006; PMID:17597259;
- 13 Flacco ME, Manzoli L, Ioannidis JPA. Noninferiority is almost certain with lenient noninferiority margins. *J Clin Epidemiol* 2016;71:118 doi:10.1016/j.jclinepi.2015.11.010; PMID:26607237;
- 14 Rehal S, Morris TP, Fielding K, et al. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ open* 2016;6:e012594 doi:10.1136/bmjopen-2016-012594; PMID:27855102;
- 15 Le Henanff A, Giraudeau B, Baron G, et al. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006;295:1147–51 doi:10.1001/jama.295.10.1147; PMID:16522835;
- 16 Wangge G, Klungel OH, Roes KCB, et al. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. *PLoS ONE* 2010;5:e13550 doi:10.1371/journal.pone.0013550; PMID:21048948;

- 17 Hays M, Andrews M, Wilson R, et al. Reporting quality of randomised controlled trial abstracts among high-impact general medical journals: a review and analysis. *BMJ open* 2016;6:e011082
doi:10.1136/bmjopen-2016-011082; PMID:27470506;