La lecture critique des essais cliniques

http://www.spc.univ-lyon1.fr/lecture-critique

Michel Cucherat
Faculté Laennec - Lyon
michel.cucherat@upcl.univ-lyon1.fr

Version 1, 30/10/2004

1 INTRODUCTION

La lecture critique d'un compte-rendu d'essai thérapeutique a pour objectif de faire une évaluation critique de son résultat avant de le mettre en application.

Pour un médecin, il s'agit de répondre à la question « le bénéfice apporté par ce traitement est-il établi avec suffisamment de fiabilité et est-il cliniquement pertinent pour justifier son utilisation en pratique ? »

Pour répondre à cette interrogation, il convient d'analyser les trois points suivants :

1. **La validité interne** : Est-ce que le résultat est fiable, c'est-à-dire est-il réel et non biaisé ?

2. **La cohérence externe** : Est-ce que ce résultat est concordant avec les autres connaissances sur le sujet ? Ce résultat n'est pas isolé et il existe tout un faisceau de preuves factuelles concordantes les unes avec les autres.

3. **La pertinence clinique et la représentativité** : Ce résultat représente-t-il un bénéfice intéressant et est-il extrapolable aux patients à traiter en pratique ? Quelle est la taille de l'effet ? L'estimation de la taille de l'effet est-elle précise ? L'effet observé correspond-il à un effet cliniquement pertinent ? et finalement les patients de l'essai sont-ils représentatifs des patients rencontrés couramment en pratique médicale ?

La lecture critique consiste donc à analyser s'il est licite de répondre par l'affirmative à ces trois groupes de questions ou si, au contraire, il existe des points conduisant à émettre des réserves sur la réalité et/ou la pertinence du résultat considéré. Dans ce dernier cas, il est alors nécessaire d'attendre les résultats d'essais supplémentaires.

La **validité interne** (« *internal validity* ») permet de s'assurer que le résultat obtenu reflète bien la réalité et qu'il n'est pas dû à un biais ou au hasard. De plus, le résultat doit être issu d'une démarche hypothético-déductive pour être valide sur un plan méthodologique (essai de confirmation).

L'absence de biais dépend de la méthode utilisée, de sa capacité à éviter les différents biais et de la qualité de réalisation de l'essai. La réalité statistique du résultat est confirmée ou infirmée par le test d'hypothèse.

La **cohérence externe** permet de s'assurer que le résultat pris en considération n'est pas unique en son genre mais qu'il s'intègre dans un cadre logique : le résultat est-il confirmé par d'autres et est-il cohérent avec les connaissances fondamentales, épidémiologiques, etc… ? En général, un résultat n'est pas interprété de manière isolée, mais mis en perspective par rapport aux autres de même nature, par exemple par une méta-analyse.

La **pertinence clinique** (« *clinical relevance* ») permet de s'assurer que le résultat de l'essai correspond à un effet suffisamment important et concernant un critère cliniquement pertinent. L'estimation de la taille de l'effet est suffisamment précise pour pouvoir raisonnablement éliminer un effet de trop petite taille, sans intérêt en pratique. L'effet a été déterminé par rapport à un comparateur adapté, placebo ou traitement de référence validé. Les patients de l'essai doivent être représentatifs des patients vus en pratique médicale courante afin d'assurer **l'extrapolabilité** (« *extrapolability* ») du résultat: même définition de la maladie, pas de sélection excessive sur le sexe, l'âge, les comorbidités, etc.. En particulier, ils ne doivent pas avoir été sur-sélectionnés.

Tableau 1. Principales composantes des validités internes et externes et de la pertinence clinique d'un résultat.

Validité interne	▪ Absence de biais ▪ Réalité statistique du résultat ▪ Résultat issu d'une démarche hypothético-déductive méthodologiquement valide
Cohérence externe	▪ Résultat confirmé par au moins un autre ▪ Méta-analyse sans hétérogénéité ▪ (cohérence avec les données fondamentales : biologiques, épidémiologiques, physiopathologiques,…)
Pertinence clinique et extrapolabilité	▪ Taille du bénéfice cliniquement pertinente ▪ Précision de l'estimation suffisante pour éliminer un effet trop petit pour être cliniquement pertinent ▪ Critère pertinent et correspondant à un objectif thérapeutique ▪ Comparateur adapté ▪ Définition de la pathologie et patients représentatifs de ceux vus en pratique médicale courante ▪ Contexte de soins similaire à celui de la pratique courante

2 LA VALIDITE INTERNE

2.1 Réalité statistique du résultat

2.1.1.1 Rappel

Une différence observée entre deux groupes peut être soit réelle, soit due au hasard. Le test statistique permet d'évaluer la réalité statistique de la différence. Le risque de conclure à tort à un effet du traitement devant une différence en réalité due au hasard est le risque alpha (risque de faux positif). Un résultat statistiquement significatif limite ce risque à moins de 5%.

Pour conclure à la réalité statistique d'un résultat, il convient que la différence observée soit statistiquement significative.

Il est aussi nécessaire d'écarter une situation d'inflation du risque alpha résultant d'une répétition des tests statistiques rencontrée en cas :

- d'absence de critère de jugement principal fixé avant l'obtention des résultats,
- d'analyses en sous groupes,
- de recherche de l'effet répétée au cours du temps,
- d'analyses intermédiaires non protégées.

2.1.1.2 Points à vérifier

Les différents points à vérifier pour s'assurer de la réalité statistique de l'effet du traitement sont les suivants :

- Le résultat est-il statistiquement significatif à un seuil inférieur ou égal à 5% ?
- Peut-on considérer que le risque alpha a été parfaitement contrôlé pour le résultat avancé ?
- Le test statistique utilisé est-il adapté ?

2.1.1.3 Situations problématiques

- Le résultat avancé est issu d'une analyse en sous-groupes. Il existe un fort risque d'inflation du risque alpha. De plus, ce résultat n'est pas issu d'une démarche hypothético-déductive, ce qui limite sa valeur épistémologique.
- Le résultat est obtenu sur un critère de jugement qui n'a pas été clairement défini a priori comme étant le critère de jugement principal.
- Des analyses intermédiaires sont réalisées sans protection contre l'inflation du risque alpha.
- Il y a une répétition de la recherche de l'effet au cours du temps.
- Il y a absence du calcul préalable de l'effectif nécessaire. L'analyse porte alors sur un nombre arbitraire de sujets. Le choix du moment de l'analyse est peut-être conditionné par les résultats obtenus.
- Un ajustement est effectué sur des variables non prévues a priori. Un ajustement post-hoc sur des variables trouvées déséquilibrées entre les groupes entraîne un biais. Les variables d'ajustement doivent être prédéfinies et non pas déterminées en fonction des résultats.
- Des mesures multiples du critère de jugement chez le même patient sont analysées comme si elles étaient mesurées chez des sujets différents (le nombre d'unités statistiques sur lesquelles se base l'analyse statistique est supérieur au nombre de patients).

2.2 Valeur méthodologique du résultat

Afin de respecter le principe de la méthode expérimentale, le résultat avancé doit correspondre directement à l'hypothèse formulée a priori, et dont le test était l'objet spécifique de l'essai.

Il convient d'éliminer la possibilité que l'hypothèse ait pu être formulée après la prise de connaissance des résultats de l'essai (formulation post-hoc). Dans ce cas, « l'expérience » ne peut que confirmer l'hypothèse puisque celle-ci a été formulée à partir de ses résultats. Cette situation tautologique enlève toute valeur aux conclusions. L'introduction doit justifier de manière prospective l'hypothèse de l'essai, ses objectifs cliniques et les analyses en sous-groupes prévues. L'introduction doit démontrer que l'hypothèse testée découle naturellement des connaissances et des données disponibles avant le début de l'essai et que celui-ci a été spécifiquement entrepris pour la tester.

Un résultat non issu d'une démarche hypothético-déductive est de nature inductif. Il suggère alors un effet, mais ne peut le démontrer.

Tout changement post-hoc (ou définition post-hoc) de l'hypothèse testée, du critère de jugement, de la population cible supprime sa valeur déductive à un résultat et le transforme en un résultat inductif, exploratoire. Ce type de résultat suggère alors un effet, mais ne peut le démontrer.

2.3 Absence de biais

L'analyse critique doit pouvoir éliminer la possibilité de l'existence d'un biais. Les situations propices à l'apparition des différents biais sont à rechercher, soit au niveau d'un défaut méthodologique, soit au niveau d'un défaut de réalisation.

> Rappel : Il ya biais quand la différence observée entre les deux groupes à la fin de l'essai est due à un autre facteur que le traitement étudié

Les différents biais pouvant affecter un essai thérapeutique vont être passés en revue. Pour chacun d'entre eux, un bref rappel de son mécanisme est effectué, puis les points à vérifier pour s'assurer que le résultat en est exempt sont listés. Pour terminer, une liste des situations à fort risque de biais est proposée.

2.3.1 Biais de confusion

2.3.1.1 Rappel

Le biais de confusion est le biais entraîné par l'absence de prise en considération des facteurs de confusion. Pour l'éviter, l'essai doit être comparatif et doit comporter un groupe contrôle contemporain utilisé comme référence.

2.3.1.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais de confusion il convient de se poser les questions suivantes.

- Existe-t-il un groupe contrôle ?
- L'effet du traitement est-il déterminé par rapport à ce groupe contrôle ?

2.3.1.3 Situations à fort risque de biais

Dans les situations suivantes, le risque de biais de confusion est fort et remet en cause la validité interne du résultat obtenu.

- Malgré la présence d'un groupe contrôle, l'effet est mesuré par une comparaison avant – après dans le groupe traité.

2.3.2 Biais de sélection

2.3.2.1 Rappel

Le biais de sélection survient lorsque les deux groupes de l'essai ne sont pas comparables. Une différence entre les deux groupes peut alors apparaître en dehors de tout effet traitement.

La randomisation a pour but d'éviter le biais de sélection en créant, en moyenne, deux groupes comparables.

2.3.2.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais de confusion, il convient de se poser les questions suivantes.

- La méthode de randomisation garantie-t-elle l'imprévisibilité du traitement alloué à un patient ?
  En effet, il est particulièrement important qu'un investigateur ne puisse pas connaître ou prédire le groupe auquel sera alloué le prochain patient. À ce titre, une « pseudo randomisation » basée sur la date de naissance du patient ou le jour de la consultation est inacceptable. L'utilisation d'enveloppe scellée n'est pas optimale, surtout pour les essais en ouvert. Seules les procédures centralisées, téléphone, fax, informatique donnent suffisamment de garantie.

- Les groupes issus de la randomisation sont-ils comparables ?
  Pour juger de cela, il convient de vérifier qu'il n'existe pas de déséquilibre entre les groupes au niveau des principaux facteurs pronostiques (ou d'autres variables conditionnant le critère de jugement).

2.3.2.3 Situations à fort risque de biais

Dans les situations suivantes, le risque de biais de confusion est fort et remet en cause la validité interne du résultat obtenu.

- Le groupe contrôle n'est pas constitué de patients contemporains, mais de témoins historiques ou de témoins géographiques (en fait, il n'y a pas eu de randomisation).

- Le processus de randomisation était prévisible. Il était possible pour les investigateurs de sélectionner les patients dans les groupes de l'essai.

2.3.3 Biais liés à l'absence ou un défaut de double insu

2.3.3.1 Rappel

L'absence, ou une mauvaise réalisation, du double insu est susceptible d'entraîner différents biais : biais de suivi, biais d'évaluation. Dans certaines situations, la réalisation d'un double insu n'est pas possible pour des raisons éthiques ou pratiques (cf. tableau 2). Dans ce cas, les essais ne peuvent être réalisés qu'en simple insu ou en ouvert. Les points spécifiques à cette situation seront abordés dans une section suivante.

2.3.3.2 Questions à se poser pour vérifier les précautions prises pour éviter les biais

Afin de vérifier si le double insu a été réalisé de façon satisfaisante il convient de se poser les questions suivantes :

- Le traitement du groupe contrôle est-il indiscernable du traitement du groupe traité ? Les deux groupes doivent recevoir un traitement qui a la même forme (gélule, perfusion IV, etc.), la même apparence (couleur, volume, conditionnement, étiquetage,), le même goût, etc…

- En cas de différence entre les traitements comparés (voie d'administration, forme galénique, etc.), une technique de double placebo a-t-elle été employée ?

- Le code du traitement figurait-il sur les boîtes de traitements (par exemple code A, B)

2.3.4 Biais de suivi

2.3.4.1 Rappel

Un biais de suivi survient lorsque les deux groupes ne sont pas suivis de la même manière au cours de l'essai. La comparabilité initiale est alors détruite et une différence peut apparaître en dehors de tout effet traitement. Le double aveugle est un élément central pour empêcher l'apparition de ce biais. À côté de l'évaluation de la qualité du double aveugle, d'autres points spécifiques du biais de suivi sont à prendre en considération.

Des points d'analyse spécifiques de l'essai en ouvert sont exposés dans la section suivante.

2.3.4.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais de suivi, il convient de se poser les questions suivantes en plus de l'analyse du double aveugle :

- Est-ce que les arrêts de traitements, les déviations aux protocoles et les traitements concomitants ont été recueillis et sont convenablement documentés ?
  Ces informations sont nécessaires pour répondre aux questions suivantes.

- Le recours aux traitements concomitants a-t-il été aussi fréquent dans tous les groupes ? Une différence dans les traitements concomitants peut faire disparaître l'effet du traitement étudié, ou, à l'inverse, faire apparaître une fausse différence.
  Une différence dans les traitements concomitants peut aussi être le reflet de l'effet du traitement étudié. Avec un traitement efficace, la fréquence de recours aux traitements de seconde ligne est réduite. Par exemple, un traitement doté d'un effet antalgique puissant entraîne une diminution de l'utilisation des antalgiques de seconde ligne prévus dans le protocole.
  Un traitement ayant une mauvaise tolérance entraîne une augmentation de consommation des traitements prescrits en raison de cette mauvaise tolérance. Par exemple, des antiémétiques avec une chimiothérapie anticancéreuse.

- Les taux de déviation au protocole sont-ils similaires dans les deux groupes ?

- Les taux d'arrêt du traitement de l'étude sont-ils similaires dans les deux groupes ?
  En sachant que les différences observées peuvent être dues à une différence de tolérance des produits et non pas à une situation potentiellement biaisée.

2.3.5 Biais d'évaluation

2.3.5.1 Rappel

Le biais d'évaluation survient quand la mesure du critère de jugement n'est pas réalisée de la même manière dans les deux groupes. Le double insu limite le risque de biais d'évaluation.

2.3.5.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais d'évaluation, il convient de se poser les questions suivantes.

- L'évaluation du critère de jugement est-elle faite de la même façon quel que soit le traitement reçu ?

- Le traitement est-il susceptible d'influencer sur la mesure du critère de jugement ?

- Dans un essai en ouvert, la mesure du critère de jugement est-elle subjective ? La connaissance du traitement reçu par le patient peut influencer la mesure du critère de jugement. Avec ce type de critère, si le double aveugle est impossible (par exemple psychothérapie), l'évaluation des patients doit se faire, en insu du traitement reçu, par un évaluateur indépendant des médecins ayant en charge les patients (triple aveugle).

2.3.6 Essai en ouvert

2.3.6.1 Rappel

Dans certaines situations, la réalisation d'un double insu n'est pas possible pour des raisons éthiques ou pratiques. Dans ce cas, les essais ne peuvent être réalisés qu'en simple insu ou en ouvert. La méthodologie employée n'empêchant pas la survenue d'un biais, il convient d'analyser soigneusement les marqueurs permettant de vérifier que le suivi et l'évaluation des critères de jugement se sont effectués de manière identique dans les deux groupes.

Seules quelques situations très particulières empêchent la réalisation d'un double insu (cf. tableau 2). En dehors de ces situations, la non-réalisation de l'essai en double insu est insatisfaisante.

Tableau 2 – Liste des situations où l'absence de double insu est « acceptable ».

Un des traitements comparés est une intervention chirurgicale ou invasive (radiologie interventionnelle comme une angioplastie).
Un des traitements comparés nécessite un appareillage lourd dont il est impossible de faire un simulacre comme la radiothérapie.
Un des traitements comparés s'accompagne d'effet indésirable ou d'une toxicité évocatrice qui laisse deviner la nature du traitement dans presque tous les cas : chute de cheveux dans des chimiothérapies anticancéreuses.
Les traitements comparés sont des stratégies de prise en charge : traitement à domicile versus traitement hospitalier.
Un des traitements comparés concerne une prise en charge améliorée : stroke unit, kinésithérapie, aide à domicile, etc.
Le traitement factice risque d'avoir un effet : faux massage, placebo de chewing-gum pour l'arrêt du tabac, etc.
Un des traitements comparés délivre son action de façon évidente et non dissimulable. Il est donc impossible d'en faire un simulacre sans effet : (chirurgie,) dans une certaine mesure kinésithérapie, cure thermale, physiothérapie (chaleur), etc.
D'une manière générique, toutes les situations où la réalisation d'un traitement « placebo » ayant la même apparence que le traitement étudié s'avère trop compliqué à réaliser ou illusoire, par exemple, quand l'action du traitement est directement visible (comme la chirurgie, le recours à une aide humaine, etc.).

2.3.6.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais dans un essai en ouvert, il convient de se poser les questions suivantes.

- Le critère de jugement est-il un critère « dur », dont l'évaluation n'est pas subjective ? Le décès est le critère le plus sûr dans un essai en ouvert car il ne demande aucune interprétation. Par contre, l'utilisation d'événements cliniques est moins robuste. Dans certains

cas, le diagnostic de survenue de l'événement clinique peut être subjectif et influencé par la connaissance du traitement du patient.

- En cas d'utilisation d'événements cliniques comme critère de jugement, l'adjudication s'est-elle effectuée de manière centralisée, indépendante et en insu de la connaissance du traitement ?

- L'essai est réalisé en ouvert alors que sa réalisation en double insu était éthiquement et pratiquement possible.
  La justification de l'absence d'aveugle pour des raisons pratiques, principalement de coûts, ne doit pas être acceptée trop facilement. L'expérience montre que, même avec des critères de jugement « durs » (mortalités), il existe une surestimation de l'effet dans les essais en ouvert par rapport aux essais en double aveugle. Les situations où il est impossible de réaliser un double insu sont rares. Par exemple, la nécessité d'une adaptation posologique en fonction d'un paramètre biologique n'est pas un obstacle insurmontable à la réalisation d'un double aveugle. Une procédure d'ajustement centralisé peut être mise en place.

2.3.7 Biais d'attrition

2.3.7.1 Rappel

Le biais d'attrition survient quand des patients randomisés sont écartés de l'analyse. Tous les patients randomisés doivent être inclus dans l'analyse. Les patients inclus mais non analysés correspondent soit à des perdus de vue, soit à des données manquantes, ce qui rend dans les deux cas le critère de jugement principal manquant.

2.3.7.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais d'attrition il convient de se poser les questions suivantes.

- Le nombre de patients analysés est-il égal au nombre de patients randomisés ?

- Qu'elle est la robustesse du résultat vis-à-vis de l'hypothèse du biais maximum ?

- Est-ce qu'une méthode de remplacement des données manquantes a été utilisée ? Dans ce cas, le nombre de patients analysés correspond au nombre de patients randomisés même si de nombreuses valeurs étaient manquantes. Ces méthodes nécessitent des hypothèses sur la nature des données manquantes. Même si elles sont pour la plupart conservatrices, leur utilisation ne doit pas faire oublier le problème initial et le risque de biais.

2.3.8 Biais liés à l'absence d'analyse en intention de traiter

2.3.8.1 Rappel

Différentes situations peuvent conduire à une destruction de la comparabilité initiale des groupes, comme, par exemple, une analyse en « per-protocole » où les patients inclus à tort, traités par erreur avec un mauvais traitement, ayant arrêté le traitement de l'étude ou ayant reçu des traitements concomitants sont exclus de l'analyse. Ces exclusions secondaires sont susceptibles de biaiser le résultat, principalement en détruisant la comparabilité initiale des groupes et du fait que les exclusions sont potentiellement liés à l'effet du traitement. Pour éviter ce biais, l'analyse doit être réalisée en intention de traiter.

Le diagramme de flux de patients des recommandations « CONSORT » permet de juger de la population soumise à l'analyse.

2.3.8.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'un éventuel biais, il convient de se poser les questions suivantes.

- L'analyse a-t-elle était faite en intention de traiter ?
  C'est-à-dire tous les patients inclus dans l'essai ont-ils été analysés dans le groupe dans lequel ils ont été randomisés, quel que soit le traitement qu'ils ont reçu ?

- Les deux situations suivantes sont-elles exclues : des patients randomisés mais non traités ne sont pas retenus pour l'analyse, des patients alloués à un groupe mais traités par erreur avec le traitement d'un autre groupe ne sont pas analysés.

2.3.9 Biais des essais de non-infériorité

2.3.9.1 Rappel

Les biais spécifiques affectent l'essai de non-infériorité, en particulier, tout ce qui concourt à faire disparaître l'effet des traitements étudiés. La situation est inversée par rapport à l'essai de supériorité où ces situations n'entraînent pas de biais mais simplement une perte de puissance.

2.3.9.2 Questions à se poser pour vérifier les précautions prises pour éviter le biais

Afin de vérifier l'absence d'éventuels biais dans un essai de non-infériorité, il convient de se poser les questions suivantes.

- Le traitement de référence a-t-il développé sa pleine efficacité ?
  Les conditions d'administration du traitement de référence (dose utilisée, schéma d'administration, observance des patients) doivent garantir l'obtention de l'efficacité optimale du traitement de référence. Si ce n'est pas le cas, un nouveau traitement, en réalité, inférieur au traitement de référence, apparaîtrait comme non-inférieur.

- Les patients inclus sont-ils similaires aux patients chez lesquels le traitement de référence a été validé ?

- Les patients inclus présentent-ils un risque suffisamment élevé pour permettre à l'effet du traitement de se manifester. La fréquence du critère de jugement doit être proche de celle qui est attendue et qui a été utilisée dans le calcul du nombre de sujets.

- L'analyse en intention de traiter donne-t-elle les mêmes résultats que l'analyse en intention de traiter ? Dans l'essai de non-infériorité, l'analyse per-protocole est la plus sensible et la moins biaisée. Cependant, elle ne reflète pas la vraie vie. L'analyse en intention de traiter est plus représentative de la pratique courante, mais elle est conservatrice et a tendance à faire disparaître les différences. Il convient donc de considérer simultanément ces deux analyses pour avoir à la fois une vue non biaisée et représentative de la réalité.

2.3.9.3 Situations à fort risque de biais

Dans les situations suivantes, le risque de biais dans l'essai de non-infériorité est fort et remet en cause la validité interne du résultat obtenu.

- La mesure du critère de jugement est peu sensible et/ou peu spécifique. La mauvaise performance diagnostique de cette mesure tend à égaliser les résultats des deux groupes, et peut gommer une différence en défaveur du traitement étudié.

- De nombreux patients sont exclus de l'analyse per-protocole.

- Il existe un fort taux d'écarts au protocole.

- Le taux de données manquantes était élevé et des techniques de remplacements ont été utilisées. Ces techniques sont conservatrices et elles sont susceptibles de faire disparaître une réelle différence entre les traitements.

3 COHERENCE EXTERNE

La cohérence externe d'un essai se juge en confrontant son résultat à ceux des autres essais abordant la même question. En fait, il s'agit d'un problème global : est-ce que tous les résultats obtenus sur cette questions sont concordants entre eux ? ou, au contraire, existe-t-il des essais qui ont donné des résultats différents des autres. Dans ce dernier cas, il est légitime de chercher le « pourquoi » de ces discordances. Mais nous verrons, que les explications que l'on peut trouver sont peu robuste, et, il est en général bien difficile de conclure sans faire un nouvel essai.

La méta-analyse est l'outil de choix pour examiner la cohérence externe d'un résultat.

Une autre composante de la cohérence externe est représentée par la compatibilité du résultat avec les autres connaissances du domaine : données épidémiologiques, pharmacologiques, physiopathologiques, etc. En général, cette concordance est acquise étant donné que l'hypothèse testée dans un essai a été construite à partir de ces informations.

3.1 Cohérence externe d'un résultat positif

Quatre situations différentes remettant en cause la cohérence externe d'un résultat positif.

1. **L'essai est le seul disponible**. Il est impossible d'écarter une origine artefactuelle du résultat, sa reproductibilité est inconnue.

2. **L'essai introduit une hétérogénéité en méta-analyse**. Le résultat n'est pas compatible avec les autres essais disponibles. Il diffère par autre chose qu'une fluctuation aléatoire. Se pose alors la question de l'explication de cette différence.

3. **La méta-analyse ne donne pas de résultat statistiquement significatif** et il n'y a pas d'hétérogénéité. Le résultat positif de l'essai est dû au hasard. Il s'agit d'un faux positif. Cette manifestation du risque d'erreur statistique alpha est corrigée par la méta-analyse. Cette situation sous-entant aussi que la quantité d'information amenée par l'essai considéré est faible vis-à-vis de celle apportée par les autres essais de la méta-analyse.

4. **Le résultat n'est pas compatible avec les connaissances physiopathologiques, pharmacologiques ou épidémiologiques**. La valeur prédictive du résultat est faible en raison d'une faible probabilité a priori de l'hypothèse, cette dernière n'étant pas ou peu étayée par les connaissances fondamentales.

3.2 Cohérence externe d'un résultat négatif

En cas de résultat négatif, contrastant avec des résultats antérieurs positifs, individuellement ou en méta-analyse, il convient de discuter de la possibilité d'un défaut de puissance du nouvel essai ou d'un biais de publication.

L'essai considéré est peut-être moins puissant que les autres, ce qui expliquerait son résultat non significatif. Ce manque de puissance doit cependant être objectivé : plus petite taille ou risque de base plus faible et absence d'hétérogénéité statistique avec les autres essais. Bien souvent, le dernier essai en date est le plus important. L'hypothèse d'un manque de puissance n'est alors pas tenable, ce qui conduit vers la deuxième possibilité qui est celle d'un biais de publication expliquant les résultats positifs antérieurs : les résultats négatifs obtenus précédemment n'ont pas été publiés.

L'existence d'un résultat négatif peut aussi être expliquée par des facteurs directement liés à l'efficacité du traitement, comme la dose utilisée, le mode d'administration, les caractéristiques des patients, la définition de la maladie, etc.

3.3 Explications des discordances

Lorsque les discordances constatées entre les différents essais sont trop importantes pour être dues au hasard, différents facteurs peuvent être analysés pour tenter de les expliquer. Il convient alors de rechercher s'il existe des différences entre les essais au niveau des caractéristiques des patients, des traitements ou des autres caractéristiques de l'étude qui pourraient expliquer les discordances de résultats.

En cas de résultats discordants, il est difficile de conclure formellement car les explications trouvées a posteriori sont fragiles.

En général, une explication est assez facilement trouvée car de très nombreux facteurs peuvent être considérés. Cependant, la confiance à accorder aux explications trouvées est faible. En effet, on découvre toujours a posteriori une théorie pour expliquer un résultat non prévu ! Il faut cependant être conscient que cette démarche est purement inductive.

Expliquer les discordances, c'est finalement choisir, entre deux types de résultats, les positifs et les négatifs, ceux que l'on considère représenter la réalité. C'est une démarche inductive et non pas hypothético-déductive. On change de paradigme : la conclusion finale est purement inductive, même si les données qui ont été prises en considération étaient issues d'une démarche hypothético-déductive. Si, de façon post-hoc, on trouve des arguments pour expliquer les résultats négatifs obtenus dans certains essais, cette explication va être utilisée pour avancer que ce sont les essais positifs qui représentent la réalité. Même si les essais positifs sont, à l'origine, de véritables essais de confirmation, ils perdent cette valeur car ils sont préférés aux autres sur la foi d'un raisonnement « a posteriori », purement inductif.

Ainsi, la sélection a posteriori de certains résultats dans un ensemble de résultats contradictoires pose donc un certain nombre de problèmes épistémologiques et conduit à des conclusions discutables. Dans l'idéal, un nouvel essai est nécessaire pour confirmer l'hypothèse avancée pour expliquer les discordances.

4 PERTINENCE CLINIQUE

L'évaluation de la pertinence clinique (« *clinical relevance* ») permet de s'assurer que le bénéfice apporté par le traitement est suffisamment important et concerne un critère cliniquement pertinent, que la balance bénéfice / effets indésirables est acceptable et que ce résultat est informatif pour les situations de la pratique médicale courante (résultat extrapolable).

4.1 Objectif de l'essai

L'objectif de l'essai (« *aims* » ou « *objective* ») doit être clairement formulé. L'hypothèse doit pouvoir être justifiée à partir des connaissances disponibles au moment de la planification de l'essai.

Pour être parfaitement défini, l'objectif doit préciser :

- le traitement testé,
- le traitement contrôle (placebo ou traitement actif),
- s'il s'agit d'une recherche de supériorité ou d'équivalence,
- le critère de jugement principal (et le moment de sa mesure),
- les patients concernés : maladie et éventuellement caractéristiques particulières.

« Montrer que l'alteplase entraîne une réduction supplémentaire de la mortalité à court terme par rapport à la streptokinase à la phase aiguë de l'infarctus » est un objectif correctement formulé.
« Évaluer le ramipril dans l'hypertension » n'est pas un objectif énoncé avec suffisamment de précision.

L'objectif de l'essai doit correspondre à une question pertinente, représentant un problème réel de thérapeutique (jugé par l'expérience du lecteur et non pas uniquement à partir de l'argumentaire de l'essai), pour lequel il n'y a pas encore de solution satisfaisante (jugée par l'étude de la méta-analyse centrée sur la question afin de connaître et de juger les résultats obtenus par les traitements concurrents).

4.2 Traitement de comparaison adapté

Un essai contre placebo, alors qu'il existe un traitement ayant déjà montré son efficacité par rapport au placebo, ne peut pas justifier l'utilisation du nouveau traitement en remplacement du traitement de référence. En effet, la supériorité du nouveau traitement par rapport au traitement de référence est inconnue. Il n'est pas exclu que le nouveau traitement soit, en fait, moins efficace que ce dernier.

À défaut de mieux, il est possible de comparer l'efficacité de ces deux traitements dans une comparaison indirecte. Tout au plus, le nouveau traitement pourra servir d'alternative en cas d'intolérance ou d'effet secondaire avec le traitement de référence, mais il ne peut pas faire l'objet d'une utilisation systématique.

Cette règle est à nuancer, par exemple, dans les cas où la démonstration de l'efficacité du traitement de référence est peu solide. La comparaison d'un nouveau traitement au placebo, et non pas à ce traitement, peut être alors justifiée.

A contrario, si l'essai est réalisé contre un traitement actif, sans que ce dernier ait été évalué, le résultat de l'essai sera peu informatif vis-à-vis du nouveau traitement. L'efficacité du traitement servant à la comparaison étant inconnue, il n'est pas possible d'exclure que celui-ci soit en fait délétère. Même si le nouveau traitement s'avère supérieur, il n'est pas possible d'en déduire qu'il est efficace et qu'il se serait révélé supérieur au placebo. Cependant dans certains cas où des traitements sont standards du fait de l'usage, ce type d'essai montre que les nouveaux traitements permettent de faire mieux (ou au pire moins mal) que les pratiques standards de soins de cette maladie.

Dans les essais de supériorité contre traitement de référence, ainsi que dans les essais de non-infériorité, il convient aussi de se méfier d'une utilisation non optimale du comparateur entraînant une diminution de son efficacité. Dans ce cas un nouveau traitement moins efficace que le traitement de référence pourrait donner l'impression du contraire. Ce genre de situation se détecte en recherchant si le traitement de référence est prescrit dans les conditions où il a révélé son efficacité optimale (dose, schéma d'administration, types de patients, etc.) et s'il a été effectivement utilisé de cette façon (dose moyenne, dose moyenne rapportée au poids, durée moyenne, fréquence des arrêts de traitement).

4.3 Pertinence clinique critère de jugement

La pertinence du critère de jugement peut être remise en cause dans de nombreuses situations :

- Le critère reflète seulement un mécanisme biologique ou pharmacologique non directement lié à l'objectif thérapeutique (critère intermédiaire).

- Dans un critère composite, la composante qui a le plus de poids (qui est la plus fréquente et/ou qui est le plus modifiée par le traitement) a une faible pertinence clinique (critère non clinique, ou critère clinique secondaire).

- Les critères continus posent assez souvent des problèmes au niveau de l'interprétation de leur pertinence clinique ou de la pertinence des effets observés.

4.4 Traitements concomitants

La liste des traitements interdits par le protocole doit aussi être analysée ainsi que les taux d'utilisation effective des différents traitements concomitants. Elle peut révéler que le nouveau traitement n'a montré une efficacité que chez des patients sous traités au niveau des co-traitements et ne bénéficiant pas de tous les traitements validés. Dans ce cas, la réelle efficacité du nouveau traitement utilisé conjointement avec les autres traitements est inconnue, de même que le risque d'interaction statistique et/ou pharmacologique.

4.5 Praticabilité du traitement évalué

Une partie de la pertinence clinique d'un résultat dépend de la praticabilité du traitement étudié. Une utilisation complexe, une mauvaise tolérance, un coût élevé ou la nécessité d'investissement important, sont autant de facteurs qui peuvent limiter la pertinence d'un résultat, même en cas d'efficacité substantielle.

La documentation des changements de posologie qui ont eu lieu dans l'essai permet d'apprécier quelle fût la praticabilité du traitement tel qu'il était prévu au protocole. De nombreuses diminutions de posologie peuvent laisser supposer une posologie initiale inadéquate. De multiples abandons du traitement révèlent un traitement soit trop lourd soit mal supporté. Dans ces situations, le traitement sera difficilement utilisable en pratique même s'il montre un bénéfice malgré ces arrêts de traitement.

L'analyse des effets secondaires documente aussi la tolérance du traitement et sa praticabilité. Un traitement mal supporté peut ne pas présenter d'avantage par rapport à un autre mieux toléré, même s'il possède une efficacité plus importante (cf. justification de l'essai d'équivalence).

La réalisation d'essai pragmatique est particulièrement nécessaire avec les traitements mal supportés. En effet, dans le cadre très protocolisé d'un essai, un traitement est moins facilement arrêté pour mauvaise tolérance que dans un contexte plus libre : les patients sont plus sollicités pour continuer et/ou des mesures d'accompagnement sont plus facilement prises. Dans ce contexte, le bénéfice mesuré surestimera celui qui sera effectivement obtenu dans la vraie vie où le traitement sera plus facilement arrêté.

4.6 Représentativité des patients inclus dans l'essai

La représentativité des patients est acquise quand aucun type de patients ciblés n'a été systématiquement évincé de l'essai

L'analyse de la population étudiée vérifie que celle-ci est constituée de patients représentatifs des patients vus en pratique et qu'elle n'est pas composée de patients sélectionnés. Il s'agit de déterminer si la population étudiée est proche de la population cible.

Ce point est jugé d'après la définition des critères d'inclusion et d'exclusion (est-ce que les critères de sélection sont trop étroits ou trop larges ?) et d'après la description des patients effectivement inclus (ces patients ont-ils en moyenne les mêmes caractéristiques que la population cible ?).

Les critères de sélection déterminent la population qui était ciblée. La description de la population incluse confirme dans quelles mesures cette cible a effectivement été atteinte. En effet, il n'est pas impossible qu'un essai n'arrive pas à inclure les patients qu'il recherche et ceci pour diverses raisons. Par exemple, les investigateurs hésitent à inclure certains types de patients, les patients sont pris en charge de façon qui exclue leur inclusion dans l'essai, etc.

La présence de certains types de patients en une faible proportion ne signifie pas obligatoirement que ces patients ont été systématiquement évincés du recrutement. Ils ne représentent peut-être qu'une faible proportion de la population cible et il est donc normal qu'ils soient peu représentés dans l'échantillon de l'essai.

4.6.1 Extrapolation des résultats

La question de l'extrapolabilité du résultat est la suivante : est-ce que l'efficacité obtenue sur les patients inclus dans l'essai est informative vis à vis de l'efficacité du traitement chez des patients présentant d'autres caractéristiques ?

Un manque de représentativité de la population étudiée n'est gênante que si l'effet du traitement varie substantiellement en fonction des caractéristiques des patients. Dans ce cas, l'efficacité observée chez certains types de patients n'est pas le reflet de l'efficacité chez d'autres types de patients ou dans la population cible de la thérapeutique.

Un scénario fâcheux serait un traitement inefficace chez la majorité des patients de la population cible mais seulement efficace chez des sujets très particuliers et chez lesquels l'essai aurait été conduit. Toutefois, il existe peu d'exemple de ce type.

Cependant l'accent est souvent mis sur une éventuelle possibilité de ce type, conduisant à déclarer tout résultat d'essai non extrapolable sous le prétexte que, presque toujours, les patients inclus dans les essais ne sont pas représentatif des patients tout venant. Cette attitude, bien que couramment enseignée, est certainement exagérée pour deux raisons. La première est que, dans les essais qui prennent garde à ce point, les patients recrutés sont très proches de ceux de la population cible. L'autre raison est qu'il existe peu d'exemples où l'efficacité du traitement change de façon notable d'un type de patient à l'autre. Ainsi, même si les patients inclus dans un essai ne sont pas strictement représentatifs de la population générale, le résultat peut être cependant retenu. L'extrapolabilité du résultat ne sera remise en cause que s'il existe des arguments positifs forts faisant craindre un changement d'efficacité important en dehors des patients de l'essai.

Bien entendu tout ceci est une affaire de nuance. Un résultat obtenu sur des patients notablement hyper-sélectionnés n'est pas aussi convainquant qu'un résultat obtenu sur une large variété de patients, même si dans ce dernier cas la population de l'essai n'est pas totalement représentative.

Très souvent la population dépend de la façon dont est posée la question de recherche comme le montre le tableau suivant :

Question	**Population de l'essai**
Effet chez les patients porteurs d'un angor y compris chez les diabétiques	Diabétiques présents dans la même proportion que leur part relative dans la population générale des patients angoreux. Cette population ne permet pas de répondre à la question « que fait le traitement chez les diabétiques ? »
Effet chez les patients angoreux diabétiques	Échantillon d'angoreux diabétiques uniquement

4.6.2 Patients représentatifs des patients vus en pratique

Pour être représentative de la pratique médicale de tous les jours, l'inclusion des patients doit être basée sur des critères larges, peu sélectifs tels qu'utilisés en pratique pour définir la maladie cible. L'essai a comme but de documenter la pratique médicale telle qu'elle sera faite avec ce traitement. C'est un essai pragmatique dont le but est de savoir si l'utilisation du traitement permet en pratique d'atteindre les objectifs thérapeutiques.

La méthode de l'essai thérapeutique peut aussi être utilisée avec une finalité plus cognitive. Ces essais explicatifs ont alors pour but d'expliquer des mécanismes et se rapprochent de la recherche

fondamentale. L'objectif de ces essais est d'étudier les mécanismes d'effet des traitements (justification théorique) mais pas d'apporter la justification factuelle de leur utilisation. Il est alors nécessaire de sélectionner un groupe de patients le plus homogène possible, afin de se placer dans les meilleures conditions expérimentales possibles. Étant donné que les patients de ces essais ne sont pas représentatifs de ceux qui sont vus en pratique, leurs résultats ont une faible pertinence clinique.

Tableau 3 – Différences au niveau de la population ciblée et des critères d'inclusion entre essai pragmatique et essai explicatif

	Essai pragmatique	**Essai explicatif**
Objectif	*Documenter le bénéfice qu'apportera le traitement aux patients qui seront traités en pratique*	*Connaître l'effet du traitement dans des conditions très précises dans un but d'explication*
Finalité	*Pratique : valider une pratique médicale*	*Cognitive : connaissance et explication*
Population visée	Représentative des futurs patients qui seront traités	La plus homogène possible afin de contrôler au maximum l'influence d'autre facteur
Critères	Simples correspondant au critère utilisé en pratique pour décider un traitement. Faisant appel à des examens de routine	Précis définissant parfaitement les patients recherchés. Faisant appel à des techniques sophistiquées

Ces principes généraux conduisent aux oppositions listées dans le tableau 4.

Au total, un essai pragmatique n'utilise des critères de sélection que si ces critères sont indispensables pour définir la population cible du traitement. De ce fait ces critères seront ensuite utilisés dans la pratique médicale.

Tableau 4 – Oppositions entre essais pragmatiques et essais explicatifs.

Essai pragmatique	Essai explicatif
Pas d'age limite (excepté >18 ans pour des raisons légales)[1]	Plage d'age parfaitement définie et limitée
Pas de sélection sur le pronostic	Sélection de patient à bon pronostic pour limiter le risque de données manquantes liées au décès du patient.
Pas d'exclusion de comorbidité[2]	Exclusion de toute comorbidité ne faisant pas partie de l'objectif de l'essai
Aucune limitation de l'utilisation des traitements validés sauf raisons particulières	Exclusion des traitements concomitants
Définition courante de la maladie (attention parfois les essais font changer la définition des maladies, exemple IDM) Définition clinique et biologique (sérologie) d'une maladie virale	Définition hautement spécialisée Diagnostic et typage virologique dans une maladie virale
Utilisation de méthode diagnostique simple et disponible en routine Infarctus du myocarde diagnostiqué par la clinique et l'ECG	Utilisation de méthode hautement spécialisée (de plus haute sensibilité et spécificité) Infarctus diagnostiqué par la scintigraphie

4.6.3 Détermination de la population cible

Dans l'approche pragmatique, tous les patients porteurs de la maladie appartiennent potentiellement à la population cible de la thérapeutique, car, si en pratique le traitement confirme son intérêt, tous ces patients seront susceptibles d'être traités. De ce fait il convient de valider le traitement sur un échantillon des patients chez lesquels il sera utilisé dans la pratique médicale future. Cette démarche globale n'est cependant envisageable que s'il n'existe a priori aucune raison de suspecter que l'effet du traitement puisse être nettement différent chez certains types de patients. Dans le cas contraire, où il existe des arguments pour penser que le traitement sera chez certains types de patients sans efficacité ou délétère, il n'est plus opportun de regrouper ces patients avec les autres dans un même essai. En effet, un effet délétère chez une faible proportion des patients peut être masqué par le bénéfice obtenu chez les autres patients.

Dans ce cas, il est nécessaire de disposer de deux essais : un répondant à la question générale avec exclusion de la sous-population de patients répondant potentiellement de façon différente au

[1] Par exemple, l'essai de prévention GISSI (évaluant en plan factoriel la vitamine E et les huiles de poisson polyinsaturée) {GISSI-prevenzione investigators, 1999 #413} a porté sur tous les patients ayant présenté un infarctus du myocarde depuis moins de 3 mois sans distinction d'âge.

[2] Dans la mesure où il n'y a pas lieu de suspecter de modification notable de l'effet du traitement en fonction de certaines covariables. Dans le cas contraire, la sélection sur ces covariables n'enlève en rien la représentativité de la population de l'essai puisque ces critère de sélection seront utilisés aussi en pratique.

traitement. L'autre essai répond à la question spécifique chez ces patients. Ces deux essais peuvent être réalisés sous la forme d'un seul essai stratifié.

La difficulté de cette approche est de déterminer, a priori, s'il y a lieux de suspecter que l'effet du traitement va être différent chez certains types de patients. La tendance spontanée est d'ailleurs de surestimer la variabilité de l'effet. Il existe relativement peu de situations où une interaction forte a été mise en évidence entre l'effet et les caractéristiques des patients. Par exemple, il est fréquemment avancé que les patients à haut risque ne doivent pas être étudier avec des patients à bas risque dans un même essai. Il n'existe pourtant pas de relation directe entre le risque et le bénéfice. Au contraire, il s'avère fréquemment que le risque relatif est identique quel que soit le risque de base. Cependant, pour les traitements s'accompagnant d'effet indésirables fréquents et/ou graves la balance bénéfice-risque peut être conditionné par le risque de base des patients.

Le Tableau 5 récapitule quelques situations pour lesquelles il est licite d'envisager des variations dans l'effet traitement et la séparation de la population cible en plusieurs sous-populations étudiées séparément.

Tableau 5 – Situations pouvant faire suspecter des variations dans l'effet du traitement (liste non exhaustive)

- Variations dans la pharmacocinétique du médicament entraînant suivant les cas une diminution de l'effet (en cas de diminution des concentrations) ou une augmentation des effets indésirables (en cas d'élévation de la concentration) : insuffisance rénale, inhibition ou induction enzymatique, etc…
- Modification, voire disparition, de la cible d'action du traitement.
- Trait génétique conduisant à une plus grande susceptibilité aux effets indésirables (par l'intermédiaire de la pharmacocinétique ou de la pharmacodynamie) ou à une modification du mécanisme d'action ou de la cible du traitement (modification des récepteurs par exemple).
- Risque de base faible laissant la possibilité aux effets indésirables de contrebalancer le bénéfice apporté.
- Stade d'irréversibilité des lésions ou des processus morbides.
- etc.

4.6.4 Conséquences du regroupement de patients trop hétérogènes

Le mélange de patients différents dans un même essai n'est pas gênant tant qu'il n'existe pas de fortes modifications de l'effet du traitement entre les différents types de patients. Ce qui est le plus gênant ce n'est pas les variations dans le risque de base mais celles dans la taille, voir le sens de l'effet du traitement (interaction). En effet, les différences de risques de bases auront simplement une répercussion au niveau de la puissance de l'essai. Par contre l'existence d'une interaction conduit à mesurer avec un seul indicateur un effet variable d'un type de patient à l'autre

Le mélange de types de patients ne bénéficiant pas du traitement de la même manière peut rester cependant intéressant en rendant compte de l'effet "moyen" sur ces différents types de patients vus en pratique et traités de la même façon.

4.7 Taille et précision de l'effet

L'estimation de la taille de l'effet (« *size of effect* ») doit être suffisamment précise pour pouvoir éliminer la possibilité que l'effet puisse être petit et donc sans intérêt en pratique. Cette infirmation est apportée par l'intervalle de confiance du résultat.

La difficulté réside dans la fixation d'une valeur pour le « plus petit bénéfice intéressant en pratique ». Cette détermination pose des problèmes similaires à celui de la fixation du « delta » dans les essais de non-infériorité. La détermination de cette valeur, qui ne peut être qu'arbitraire, doit prendre en compte de nombreux facteurs :

- la gravité de la pathologie,
- les autres possibilités de traitement,
- les autres intérêts du traitement (tolérance, coût, facilité d'administration, satisfaction des patients),
- la fréquence de la maladie : un petit bénéfice peut correspondre, si la maladie est fréquente, à un nombre substantiel d'événements sérieux évités au niveau de la population toute entière,
- les choix politiques de santé publique.

4.7.1 Intervalle de confiance

L'intervalle de confiance (« *confidence interval* ») traduit la précision de l'estimation de la taille de l'effet réalisée par l'essai.

Le but de l'estimation est de déterminer la vraie valeur d'un paramètre, par exemple, la vraie réduction relative de mortalité. Cependant, la valeur estimée dans un échantillon peut être assez loin de la vraie valeur du fait des fluctuations aléatoires d'échantillonnage, c'est-à-dire du fait du hasard. L'intervalle de confiance permet de prendre en compte cette incertitude aléatoire dans la présentation des estimations.

L'intervalle de confiance (IC) à 95% est un intervalle de valeurs qui a 95% de chance de contenir la véritable valeur du paramètre estimé. Avec un peu moins de rigueur, il est possible d'admettre que l'IC représente la fourchette de valeurs à l'intérieur de laquelle nous sommes certains à 95% de trouver la vraie valeur recherchée. L'intervalle de confiance est donc l'ensemble des valeurs raisonnablement compatibles avec le résultat observé (estimation ponctuelle). Il donne une expression formelle de l'incertitude rattachée à une estimation ponctuelle du fait des fluctuations d'échantillonnages.

Par exemple (figure 1), une réduction de mortalité de -20% avec un IC 95% de [–35% ;-5%] signifie que bien qu'une baisse de –20% ait été observée ponctuellement dans l'essai, il n'est pas possible d'exclure que l'efficacité du traitement soit en réalité plus petite (au pire elle peut être de –5%) ou plus grande (au mieux de –35%).

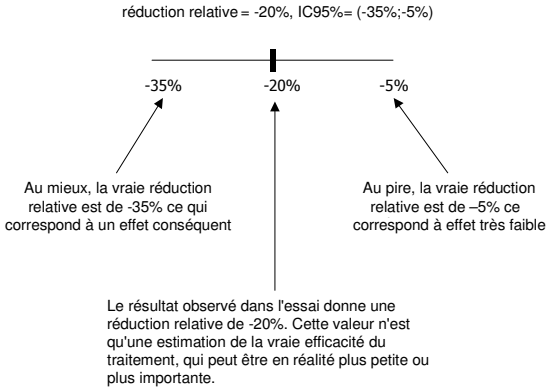


Figure 1 – Signification des bornes d'un intervalle de confiance

4.7.2 Interprétation

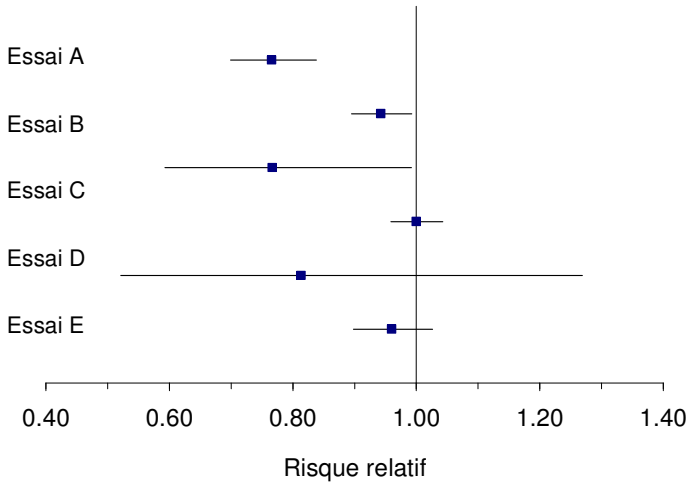
La borne péjorative de l'intervalle de confiance représente le plus petit effet du traitement que l'on ne peut pas raisonnablement exclure.

Les intervalles de confiance permettent de visualiser le plus petit effet du traitement que l'on ne peut pas raisonnablement exclure. Ce plus petit effet est la borne péjorative de l'intervalle de confiance.

Tableau 6 – Exemple de 5 situations différentes (ces données sont représentées graphiquement sur la figure).

Essai	RRR	IC 95%		p
A	-23%	-30%	-16%	0,000
B	-6%	-10%	-1%	0,024
C	-23%	-41%	-1%	0,043
D	0%	-4%	4%	1,000
E	-19%	-48%	27%	0,362

RRR : réduction relative de risque. Par convention, une RRR négative signe une réduction de risque. A l'opposé, une valeur positive témoigne d'une augmentation.



4.7.2.1 Interprétation des intervalles de confiance dans le cas d'un résultat significatif

Dans l'essai A (Cf. tableau et figure), le traitement entraîne une réduction relative du risque (RRR) de -23% (IC95% [-30%,-16%]) ; une valeur de RRR négative signe une réduction du risque, à l'inverse une valeur positive une augmentation. L'interprétation de ce résultat est qu'il existe un effet statistiquement significatif, de taille importante et connue avec précision. Ce traitement est intéressant en pratique car quel que soit la valeur réelle de l'effet, celle-ci reste intéressante. Dans le pire des cas, cet effet est encore de -16% ce qui correspond à une réduction relative du risque satisfaisante.

Le traitement dans l'essai B entraîne une réduction relative du risque de -6% (IC95% [-10% ; -1%]). L'interprétation de ce résultat est qu'il existe un effet statistiquement significatif, que l'effet du traitement est connu avec précision (l'intervalle de confiance est étroit) mais qu'il n'est pas formellement prouvé que le traitement soit intéressant en pratique. En effet, même dans la meilleure des situations, c'est à dire celle où l'effet réel serait proche de la borne inférieure (-10%), la taille de l'effet reste faible et peu intéressante en pratique.

Le traitement dans l'essai C entraîne une réduction relative du risque de -23% (IC95% [-41% ;-1%]). L'interprétation de ce résultat est qu'il existe un effet statistiquement significatif , la taille de l'effet n'est pas connue avec précision mais il se pourrait que cet effet soit de taille intéressante. En effet l'estimation ponctuel (-23%) témoigne d'un effet substantiel de même que la borne inférieure de l'intervalle (-41%). Cependant l'incertitude sur ce résultat est grande est-il est aussi possible que l'effet réel soit quasiment nul (proche de la borne supérieure, -1%). En pratique, il est difficile de recommander l'utilisation de ce traitement car il existe une possibilité que le traitement soit peu efficace. Un essai supplémentaire qui permettra d'améliorer la précision de l'estimation de l'effet en méta-analyse pourrait être souhaitable.

4.7.2.2 Interprétation des intervalles de confiance dans le cas d'un résultat non-significatif

Dans l'essai D, le traitement entraîne pas de modification relative du risque (RRR=0%, IC95% de [-4% ;+4%]). Ce résultat n'est pas significatif (p=1.00). Aux mieux, il pourrait exister une réduction très faible de 4% qui ne présente pas beaucoup d'intérêt en pratique. Bien qu'en toute rigueur, il ne soit pas possible de conclure à l'absence d'efficacité, l'interprétation de l'intervalle de confiance autorise à conclure que très probablement ce traitement serait d'aucune utilité en pratique. Cet exemple montre la supériorité de l'approche par les intervalles de confiance sur celle utilisant uniquement des tests statistiques. En utilisant l'approche test statistique il est impossible de conclure. Par contre, avec l'approche basée sur les intervalles de confiance et étant donné la précision du résultat, il est licite de conclure à l'absence d'intérêt de ce traitement : même si celui-ci a une efficacité non nulle, la taille de l'effet serait trop petite pour être intéressante en pratique.

Le traitement dans l'essai E entraîne une réduction relative non significative de -19% (IC à 95% de [-48%,+27%]). Il apparaît clairement que ce résultat non significatif n'autorise pas à conclure à l'absence d'effet. En effet, ce résultat est compatible avec une réduction relative de -48%, effet de taille conséquente. De plus l'intervalle est en très grande partie du coté favorable ce qui renforce la possibilité de l'existence de l'effet. En conclusion, il est possible que le traitement soit efficace et que cette efficacité soit suffisamment importante pour être intéressante en pratique. Ce résultat encourage à réaliser un nouvel essai de plus grande puissance.

4.8 Évaluation de la balance bénéfice-risque

Les conséquences sur l'organisme d'un traitement, médicamenteux ou d'autre nature, ne sont jamais exclusivement bénéfiques, mais s'accompagnent d'effets indésirables, plus ou moins sévères, plus ou moins intenses ou fréquents.

La prise de décision doit donc mettre en balance les effets bénéfiques et les effets négatifs :

- « Est-ce que le risque lié aux effets indésirables est acceptable compte tenu de la maladie traitée ? »,

- « est-ce que les effets indésirables ne contrebalancent-ils pas la totalité du bénéfice ? ».

À ce niveau, le but de l'interprétation des résultats est de déterminer si la balance bénéfice risque est favorable ou non au traitement. Le raisonnement et les éléments d'interprétation vont être différents en fonction de la gravité relative des événements indésirables par rapports à la maladie ou à la nature du bénéfice apporté.

Tableau 7 – Quelques situations où bénéfice et effets indésirables sont de natures différentes

Situations	Bénéfice	Effets indésirables
Oméprazole dans ulcère duodénale	Augmentation du taux de cicatrisation	Troubles neuropsychiatrique, rares troubles hématologiques
Statines dans la prévention secondaire des maladies cardiovasculaire	Réduction de mortalité et de morbi-mortalité	Myalgies parfois associées à une rhabdomyolyse, se compliquant exceptionnellement d'insuffisance rénale aiguë
IEC insuffisance cardiaque	Réduction de mortalité et de morbi-mortalité	Toux, rare œdème de Quincke, augmentation modérée de la créatinine

4.8.1 Effets indésirables de gravité supérieure à la maladie

Lorsque les événements indésirables sont de gravité bien supérieure à la pathologie traitée, la balance bénéfice/risque sera défavorable lorsque le risque encouru est disproportionné par rapport au bénéfice apporté. C'est le cas, par exemple, de la survenue de syndromes de Lyell potentiellement mortels avec la prise d'AINS pour soulager des traumatismes bénins. Par contre, avec des chimiothérapies anticancéreuses, un risque faible de syndrome de Lyell est acceptable compte tenu de la gravité de la maladie et du bénéfice attendu sur la mortalité.

En fait, la gravité des événements indésirables est à confronter, non pas à la gravité de la maladie, mais à la nature du bénéfice. Ainsi, la survenue de syndromes de Lyell avec un antiémétique est peu acceptable même dans un contexte oncologique.

Tableau 8 – Quelques situations où le risque lié aux effets indésirables a été jugé excessif par rapport au bénéfice apporté.

Situation	**Risque lié aux effets indésirables**	**Bénéfice prouvé**
Alostron dans le colon irritable	Infarctus mésentérique et occlusion intestinale (dont certains cas mortels)[3]	Diminution de l'inconfort et des douleurs abdominales
Troglitazone dans le diabète de type 2	Insuffisance hépatique conduisant dans certains cas au décès ou à la transplantation[4]	Diminution des taux d'hémoglobine glycosylée (pas d'effet prouvé sur les complications du diabète)

[3] ce traitement a été retiré de la commercialisation 9 mois après son introduction sur le marché aux USA à la suite du décès de 5 patients.

[4] Le troglitazone a été retiré du marché 3 ans après sa commercialisation à la suite de 90 cas d'insuffisance hépatique dont 70 mortels ou ayant nécessités une transplantation.

Cette évaluation repose sur un jugement de valeur et va donc dépendre de l'appréciation qui est faite de la gravité des événements indésirables et de l'importance de leur fréquence de survenue. Cette appréciation repose sur des échelles de valeur qui peuvent être différentes suivants les médecins, les patients et qui dépendent des choix sociétaux. Ces éléments expliquent les différences d'appréciation des mêmes données qui se rencontrent parfois dans l'évaluation de la balance bénéfice risque.

4.8.2 Effets indésirables de même nature que la maladie

Certains événements indésirables sont de même nature que les événements que le traitement cherche à prévenir. Ils sont donc susceptibles de contrebalancer en partie ou en totalité le bénéfice apporté par le traitement (cf. tableau 9).

C'est, par exemple, le cas des décès par hémorragie cérébrale induit par l'utilisation des fibrinolytiques à la phase aiguë de l'infarctus du myocarde. Dans ce cas le critère de jugement utilisé pour l'efficacité, le décès, prends aussi en compte les effets délétères et évalue directement la balance bénéfice risque. Par exemple, avec la streptokinase comparée au placebo, la mortalité totale est réduite de 23% ce qui permet de dire que, même si il existe des effets délétères mortels, ceux-ci sont quantitativement minoritaire à coté des effets bénéfiques du traitement. Cette situation est la plus propice à une tentative de quantification du rapport bénéfice risque (cf. section suivante).

Fréquemment, les événements indésirables graves s'accompagne d'un cortège d'événements plus bénins mais qui entraînent une gêne substantielle des patients. Ces deux dimensions doivent être prise en compte dans la décision.

Tableau 9 – Exemples de situations où les effets indésirables sont de même nature que le bénéfice apporté par le traitement

Situation	Effets indésirables	Bénéfice
Aspirine en prévention des maladies coronariennes ischémiques (American Physicians' Health study)	Augmentation de la fréquence des accidents cérébraux hémorragiques	Diminution de la fréquence des événements coronariens (infarctus)[5]
Œstrogènes et hypercholestérolémie (CDP)	Augmentation de la fréquence des thromboses veineuses	Diminution de la fréquence des événements coronariens (infarctus)

Les effets indésirables bénins posent des problèmes identiques lorsque le bénéfice apporté par le traitement est de même nature. Par exemple, dans le syndrome d'instabilité vésicale, un traitement qui améliorerait la symptomatologie en diminuant le nombre de mictions, mais qui s'accompagnerait d'effets indésirables fréquents de type cholinergique (sécheresse buccale et oculaire, nervosité, somnolence, etc.) aurait une balance bénéfice risque défavorable. Pour un grand nombre de patients, le désagrément induit par le traitement serait analogue à celui qui est soulagé par le traitement. Bien entendu cette appréciation est susceptible de varier en fonction du patient, suivant son vécu de la symptomatologie urinaire.

[5] Au total il n'est pas possible de mettre en évidence de réduction significative de la mortalité totale.

4.8.3 Synthèse

Au total, la balance bénéfice risque est défavorable dans les situations suivantes dont les principales caractéristiques sont récapitulées dans le Tableau 10 :

- la gravité des événements indésirables fait courir un risque disproportionné par rapport à la gravité de la maladie ou par rapport à la nature du bénéfice apporté,

- la fréquence des événements indésirables contrebalance le bénéfice.

Des facteurs externes rentrent aussi en ligne de compte dans l'appréciation de la balance bénéfice risque, comme :

- les risques inhérents aux autres traitements déjà existant : la balance bénéfice/risque d'un nouveau traitement sera jugé défavorable s'il est moins bien toléré que les traitements déjà disponibles, même si dans l'absolu, celle-ci pourrait être acceptable ou si son efficacité est légèrement supérieure à celle des autres,

- la différence de perception entre la personne exposée au risque, la victime potentielle, et le décideur (cf. tableau ci-dessous). La perception du risque par les patients vient donc s'interposer entre l'évaluation de la balance bénéfice risque et la décision qui doit tenir compte de ce filtre déformant.

	Analyse du décideur	Analyse de la victime
Nature du bénéfice	Social, donc général	Particulier, donc individuel
Nature du détriment	Risque, donc probabilité	Danger, donc intolérable

Tableau 10 – Les deux situations dans lesquelles les effets indésirables compromettent la balance bénéfice risque (BBR).

Effets indésirables de gravité supérieure à la celle de la maladie …	**Effets indésirables de même gravité que la maladie …**
… rendent la BBR défavorable quand ils font courir un risque disproportionné par rapport à la gravité de la maladie ou au bénéfice apporté.	… rendent la BBR défavorable quand ils contrebalance la quasi totalité du bénéfice.
Il s'agit souvent **événements spécifiques** qui ne surviennent habituellement pas dans le cadre de la maladie.	Il s'agit **d'événements non spécifique**s qui sont communément rencontrés dans la maladie. Le surcroît d'événement engendré par le traitement est donc à distinguer du bruit de fond naturel. Une comparaison est nécessaire.
Mise en évidence en **pharmacovigilance**, avec quantification à l'aide **d'études d'observations.**	Mise en évidence par **comparaison à un groupe contrôle** (analyse de sécurité d'un essai d'efficacité).

5 CAS PARTICULIER DE L'ESSAI NEGATIF

5.1 Essai de supériorité

Trois types de circonstance peuvent être à l'origine d'un essai négatif, c'est-à-dire d'un essai dont le résultat n'est pas statistiquement significatif :

1. un manque de puissance statistique,
2. un essai trop conservateur,
3. l'absence d'efficacité du traitement testé.

5.1.1 Manque de puissance statistique

Un essai dont l'effectif est insuffisant, vis-à-vis de la taille de l'effet à mettre en évidence ou du risque de base de l'événement, présente un manque de puissance statistique et a une forte probabilité d'obtenir un résultat non significatif même si l'effet du traitement est non nul (cf. puissance statistique).

Ce manque de puissance est à évoquer dans les situations suivantes.

- L'essai a arrêté de recruter avant d'attendre l'effectif nécessaire en raison, par exemple, de difficultés de recrutement.
- Les patients inclus avaient un risque de base plus faible que celui utilisé pour le calcul de l'effectif.
- La taille de l'effet est plus petite que celle attendue.

5.1.2 Essai trop conservateur

Un essai trop conservateur est un essai qui, par construction ou par défaut, limite l'importance de la différence entre les groupes, et sous-estime ainsi la taille de l'effet. Les circonstances de l'essai ne sont pas propices à la pleine et entière manifestation de l'efficacité du traitement.

Il est possible de suspecter un essai trop conservateur dans les circonstances suivantes.

- Le schéma d'utilisation du traitement étudié prévu au protocole est non optimal : dose insuffisante (par exemple par rapport au poids moyen des patients), durée insuffisante, etc. Cependant, si en pratique le traitement étudié n'a pas pu être utilisé de manière optimale en raison de nombreux arrêts de traitement liés à une mauvaise tolérance ou à la survenue de nombreux effets indésirables, le résultat négatif de l'essai n'est pas à mettre sur le compte d'un excès de conservatisme. Il reflète la réalité qui est celle d'un traitement peu efficace car mal supporté.

- Le taux d'utilisation des traitements concomitants dans l'essai est important, apportant, entre autres, au groupe contrôle un bénéfice thérapeutique identique à celui développé dans le groupe expérimental. Les deux groupes étant traités de manière similaire, aucune différence n'est susceptible d'apparaître même en cas d'efficacité du traitement étudié.

- La qualité de réalisation est médiocre avec un fort taux d'inclusion à tort, d'écart au protocole, etc. En cas de faible qualité de la réalisation de l'essai, le principe de l'intention de traiter entraîne un certain degré de conservatisme indispensable pour éviter un biais.

- La mesure du critère de jugement est trop peu spécifique et/ou sensible.

- Les patients recrutés sont insensibles au traitement : patients non malades, ne répondant pas au traitement ou au contraire trop sévèrement atteints et inaccessibles à l'efficacité du traitement

- etc.

Note : Un essai trop conservateur est une source de biais pour la recherche de la non-infériorité.

5.1.3 Manque réel d'efficacité du traitement

Le manque réel d'efficacité du traitement étudié ne sera vraiment suspecté que si l'essai permet d'exclure la possibilité que la taille de l'effet est supérieure au plus petit effet intéressant cliniquement. Cela est obtenu quand le plus grand effet compatible avec l'intervalle de confiance du résultat[6] est inférieur au plus petit effet intéressant cliniquement. Ce point est développé dans le chapitre démonstration de l'effet.

5.2 Essai de non-infériorité négatif

Avec les essais de non-infériorité, un résultat négatif (non concluant) peut provenir soit d'un manque de puissance soit d'un traitement étudié inférieur au traitement contrôle.

Un manque de puissance conduit à un résultat négatif en donnant un intervalle de confiance large, dont la borne supérieure se trouve au-delà de la limite de non-infériorité. La figure 2 représente trois situations de résultats non significatif dans un essai de non infériorité. Les cas A et B correspondent à un manque de puissance. En B, même si le nouveau traitement apparaît plus efficace que le traitement contrôle, l'intervalle est encore trop large pour conclure à la non-infériorité.

[6] matérialisée par sa borne inférieure.

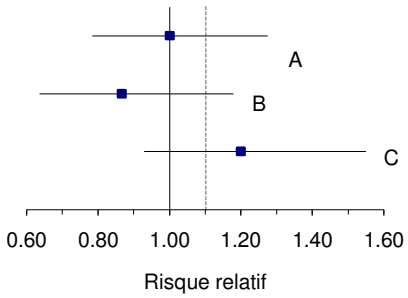


Figure 2 – Trois situations conduisant à des résultats non significatifs dans un essai de non infériorité. Le trait vertical en pointillé représente la limite de non infériorité.

Le cas C évoque un nouveau traitement moins efficace que le traitement de référence.