

Guide de lecture critique d'un essai thérapeutique

Michel Cucherat

Service de biostatistiques, Hôpitaux de Lyon et EA 3637, Université Lyon 1, Faculté RTH Laennec, BP 8071, 69376 Lyon Cedex 08 <michel.cucherat@upcl.univ-lyon1.fr>

La lecture critique d'un compte rendu d'essai thérapeutique pour la pratique médicale a pour but de déterminer si les résultats présentés apportent la démonstration de l'intérêt clinique du traitement évalué. La réponse à cette question passe par l'évaluation de la validité interne de l'essai (réalité statistique du résultat et absence de biais) et par l'appréciation de la pertinence clinique de l'effet mis en évidence (pertinence de la question posée, représentativité des patients inclus, taille du bénéfice, balance bénéfice/risque).

Mots clés : essai clinique, lecture critique, interprétation des résultats

Pour le praticien, l'objectif de l'analyse d'un article rapportant un essai thérapeutique est d'évaluer de manière critique son résultat avant de le mettre, éventuellement, en application. Il s'agit de répondre à la question : « La balance bénéfice/risque de ce traitement est-elle suffisamment établie et cliniquement pertinente pour justifier son utilisation courante ? Et chez quels types de patients ? »

Pour cela, il convient d'analyser trois dimensions :

– la validité interne du résultat : « Le résultat est-il réel ? Non biaisé ? »

– sa cohérence externe : « Le résultat est-il concordant avec les autres connaissances disponibles sur le sujet (physiopathologiques, pharmacologiques, épidémiologiques) et avec les résultats des autres essais déjà réalisés ? »

– sa pertinence clinique et sa représentativité : « Ce résultat représente-t-il un bénéfice clinique intéressant en pratique, avec une balance bénéfice/risque favorable ? Quels types de patients sont concernés ? »

L'interprétation des résultats d'un essai thérapeutique 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, un nouvel essai devra lever ces réserves.

Ce guide de lecture reprend les grandes questions que l'on doit se poser lors de la lecture critique d'un essai clinique dont les résultats sont présentés comme pouvant conduire à une modification des pratiques¹.

¹ Les essais ne débouchant pas sur des résultats conclusifs ne peuvent pas être utilisés pour suggérer un changement de la pratique. Leur lecture critique n'est pas du champ de la pratique médicale mais plutôt de la recherche thérapeutique. Pour ces essais, la question principale qui se pose est : « Est-ce que l'absence d'effet statistiquement significatif est le témoin d'une absence de bénéfice ou d'un manque de puissance ? » L'enjeu est de déterminer s'il est nécessaire de refaire un essai éventuellement plus puissant ou conçu de manière différente...

mt

Tirés à part : M. Cucherat

Quel est le résultat mis en avant par les auteurs ?

L'analyse critique porte plus sur le résultat mis en avant par l'auteur de l'essai comme apport original de son travail que sur l'essai lui-même. En effet, un résultat peut être fragile car, par exemple, issu d'une analyse en sous-groupe alors que l'essai était irréprochable au plan méthodologique. C'est donc bien *l'analyse du résultat mis en avant* qui permet d'identifier ce problème et non pas *stricto sensu* l'analyse de l'essai.

Quelle est la validité interne de ce résultat ?

Face à un essai considéré comme conclusif par les auteurs, il est important de s'assurer que le résultat est exact, *c'est-à-dire statistiquement réel et sans biais*. La méthode utilisée pour produire ce résultat est-elle fiable et en mesure de refléter avec exactitude la réalité de l'efficacité du traitement ?

Quelle est la réalité statistique du résultat ?

– La différence est-elle statistiquement significative (ou la non-infériorité est-elle statistiquement significative, s'il s'agit d'un essai de non-infériorité) [1] ? Le p est-il inférieur à 5 % (ce qui autorise à conclure à un effet non nul avec un risque d'erreur dû au hasard faible) ?

– Est-il possible d'exclure une inflation du risque alpha liée à une répétition des tests [2], *c'est-à-dire* :

- Le résultat avancé a-t-il été obtenu sur le critère de jugement principal défini *a priori* à l'issue d'une seule comparaison statistique, ou est-il obtenu après réalisation de nombreux tests sur des critères secondaires ?

- Est-il possible d'exclure des mesures multiples du critère de jugement (mesures répétées au cours du temps) ?

- Le résultat avancé est-il issu de l'analyse effectuée avec la totalité de la population de l'essai et non pas d'une analyse en sous-groupe ?

- Si des analyses intermédiaires ont été réalisées, une méthode de contrôle de l'inflation du risque alpha a-t-elle été utilisée ?

Le résultat est-il suffisamment à l'abri des biais ?

La réalité statistique du résultat ne met pas à l'abri de biais potentiels pouvant survenir au cours de la réalisation pratique de l'essai [3]. Il convient donc de vérifier si la méthodologie employée met bien l'essai à l'abri des biais. Pour l'essai thérapeutique concluant, un biais est une situation conduisant à des résultats en faveur de l'efficacité du traitement étudié alors que celui-ci est en réalité sans

effet. La présence de biais expose au risque de conclure à tort à l'efficacité du traitement étudié. Si on n'est pas témoin de la réalité du déroulement de l'essai, l'existence de biais est impossible à déterminer. En revanche, on peut utiliser des garde-fous afin de contrôler au mieux ces biais, et notamment en vérifiant que les principes méthodologiques garantissant l'absence de biais ont été correctement mis en œuvre. Dans ce cas, l'essai est à l'abri de biais potentiels et il ne fait pas courir le risque de conclure à tort à l'efficacité du traitement. La recherche de biais repose sur la réponse à une question fondamentale : existe-t-il une différence entre les deux groupes (autre que les traitements comparés) qui pourrait faire apparaître une différence au niveau du critère de jugement en dehors de tout effet du traitement (en d'autres termes, le principe de causalité est-il respecté) ? En effet, l'essai comparatif est exposé à quatre grands types de biais : biais de sélection, biais de suivi, biais de mesure, biais d'attrition. Les questions clés posées pour la recherche de ces biais potentiels sont les suivantes.

Existe-t-il un groupe contrôle : prise en compte des facteurs de confusion ?

– Existe-t-il un groupe contrôle ?

– L'effet du traitement est-il déterminé par rapport à ce groupe contrôle ?

– Ou, en d'autres termes, l'essai répond-il au principe de comparaison ?

Est-il possible d'éliminer un biais de sélection : l'essai est-il randomisé ?

Un biais de sélection survient quand il a été possible de sélectionner des patients particuliers dans le groupe du traitement étudié par rapport à ceux du groupe contrôle. De ce fait, il est possible qu'il existe une différence pronostique entre les deux groupes des patients qui fera apparaître une différence au niveau du critère de jugement même en l'absence d'effet du traitement étudié.

– Les patients ont-ils été répartis de façon aléatoire entre les groupes de traitement ?

– Est-ce que la méthode de randomisation garantit l'imprévisibilité du traitement ? L'investigateur est-il bien dans l'ignorance du traitement qui sera alloué ? [4]

– Existe-t-il des différences dans les caractéristiques de base entre les groupes ?

– Ou, en d'autres termes, l'essai répond-il au principe de comparabilité initiale des groupes ?

Est-il possible d'éliminer un biais de réalisation ou de mesure : l'essai est-il réalisé en double aveugle ou avec des procédures à l'aveugle concernant le suivi et l'évaluation ?

Un biais de réalisation est susceptible de survenir lorsqu'il est possible de prendre en charge les patients de manière différente entre les deux groupes (en particulier au niveau des traitements concomitants). Un biais de mesure survient lorsque le critère de jugement peut être

mesuré ou recherché de façon différente entre les deux groupes.

- L'essai a-t-il été réalisé en double aveugle ?
- Si l'essai est en double aveugle, la nature du traitement attribué à chaque patient est-elle masquée de façon adéquate ?
- Si l'essai a été réalisé en ouvert, le critère de jugement est-il un critère « dur », c'est-à-dire objectif (la mortalité étant le critère le plus dur) ? (Dans ce cas, un biais de mesure est évité mais pas un biais de réalisation.)
- Les critères de jugement ont-ils été validés par une procédure centralisée et en aveugle ?
- Est-ce que les arrêts de traitement, les déviations aux protocoles et les traitements concomitants ont été recueillis et sont convenablement documentés ?
- Le recours aux traitements concomitants a-t-il été identique dans tous les groupes ?
- 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 ?
- L'évaluation du critère de jugement est-elle faite de la même façon quel que soit le traitement reçu ?
- Ou, en d'autres termes, l'essai répond-il au principe de comparabilité secondaire ?

L'essai est-il réalisé en intention de traitement ?

Est-il possible d'éliminer un biais d'attrition ?

Ce biais est moins connu mais tout aussi important, ce d'autant pour les essais en ouvert. Le biais d'attrition peut apparaître dès lors que les patients analysés ne correspondent pas aux patients randomisés, du fait de l'exclusion de patients, de l'existence de perdus de vue ou de données manquantes. Il est donc important de s'assurer que toutes les précautions ont été prises pour éviter la survenue de ce biais. Il convient de noter que les règles de contrôle du biais d'attrition dans l'essai de non-infériorité sont différentes [1]. Ce qui suit ne s'applique qu'à l'essai de supériorité.

- L'analyse a-t-elle été réalisée en intention de traiter, c'est-à-dire sur tous les patients randomisés [5] ?
- Notamment s'il s'agit d'un essai de non-infériorité, l'analyse est-elle confirmée par l'analyse *per* protocole, c'est-à-dire excluant les patients pour lesquels le protocole n'a pas été suffisamment bien suivi pour permettre une estimation fiable de l'efficacité des traitements ?
- Quelle est la robustesse du résultat vis-à-vis des patients non évaluables ?

Si le critère de jugement principal est binaire, les résultats basés sur l'hypothèse du biais maximum par exemple sont-ils donnés ? Rappelons que le biais maximum consiste à considérer les patients non évaluables comme ayant présenté un événement s'ils sont dans le groupe « nouveau traitement » et comme n'ayant pas

présenté l'événement s'ils sont dans le groupe « traitement standard ».

Si le critère de jugement principal est quantitatif, les données manquantes ont-elles été remplacées, par exemple par la dernière valeur mesurée disponible (méthode LOCF, de l'anglais *last observation carried forward*) ?

Quelle est la validité méthodologique ?

Ce résultat est-il issu d'une démarche hypothético-déductive ? C'est-à-dire que l'hypothèse testée a bien été formulée avant la réalisation de l'essai. Ou, au contraire, l'hypothèse a-t-elle été formulée au vu des résultats, hypothèse qui n'avait pas été générée lors de la conception du protocole ?

- L'objectif de l'essai est-il clairement précisé dans l'introduction ?
- L'hypothèse est-elle justifiée par des publications antérieures lors de la conception de l'essai ?

Quelle est la cohérence externe ?

- Le résultat est-il confirmé par d'autres essais ?
- La discussion présente-t-elle les résultats des autres essais adressant la même question ?
- La méta-analyse de ces essais montre-t-elle une absence d'hétérogénéité ?
- Le résultat est-il cohérent avec les connaissances fondamentales (physiopathologiques, études d'observation, etc.) ?

Quelle est la pertinence clinique ?

Les principes méthodologiques et statistiques permettent de valider la fiabilité d'un résultat, mais la signification clinique et son applicabilité aux patients vont rendre ce résultat pertinent ou non pour le soin.

Le résultat apporte-t-il un progrès thérapeutique, c'est-à-dire représente-t-il un bénéfice clinique intéressant en pratique ? À quels types de patients est-il généralisable en pratique ?

Quelle est la pertinence de l'objectif de l'essai ?

L'objectif de l'essai correspond-il à une question thérapeutique réelle, sans réponse satisfaisante actuellement ?

La question clinique est-elle réelle ? Correspond-elle à une entité nosologique posant un réel problème pour les patients ou la santé publique ?

Quelle est la pertinence des traitements comparés ?

L'effet a-t-il été déterminé par rapport à un comparateur adapté, placebo ou traitement de référence validé ? Si

des traitements ayant démontré leur efficacité existent, la question thérapeutique qui se pose n'est plus de connaître l'efficacité des nouveaux traitements par rapport au placebo mais de démontrer qu'ils apportent un plus par rapport aux meilleurs traitements disponibles.

– Si l'essai est contre placebo, n'existe-t-il vraiment aucun traitement ayant fait la preuve de son efficacité contre placebo ?

– Si l'essai est contre traitement de référence, ce traitement a-t-il montré son efficacité dans des essais thérapeutiques ? Est-il bien utilisé de façon optimale et validée (dose, schéma posologique, adaptation, surveillance) ?

Quelle est la pertinence du critère de jugement ?

Le critère de jugement du résultat mis en avant est-il cliniquement pertinent et correspond-il à l'objectif thérapeutique du traitement ? C'est-à-dire est-ce un critère clinique et non pas un critère intermédiaire [6] ?

Si le critère est composite, a-t-il une cohérence clinique ? Toutes les composantes sont-elles influencées de la même manière [ou le résultat global sur le critère composite n'est-il obtenu que par un effet sur les composantes les moins cliniquement pertinentes (les plus « molles »)] ?

Les modifications des critères de jugement secondaires sont-elles cohérentes et conformes aux hypothèses physiopathologiques ?

Quelle est la pertinence de la taille de l'effet ?

Le résultat de l'essai correspond-il à un effet suffisamment important au point de vue clinique ?

L'effet apporte-t-il un service médical suffisant au patient (c'est-à-dire résout-il de manière substantielle le problème du patient ou s'agit-il seulement d'un effet non nul mais ne modifiant que très superficiellement l'état du patient) ?

La précision de l'estimation (donnée par l'intervalle de confiance à 95 %) est-elle suffisante pour éliminer un effet trop petit pour être cliniquement pertinent [7] ?

Quelle est la pertinence des patients inclus ?

Le résultat obtenu est-il généralisable aux patients vus en pratique ?

– La définition de la pathologie utilisée dans l'essai est-elle conforme aux critères diagnostiques habituels, basés sur des méthodes diagnostiques habituelles et couramment disponibles ?

– Compte tenu des critères d'inclusion et d'exclusion, et de la description de la population incluse, n'y a-t-il pas eu une sélection excessive des patients ? Les patients sont-ils représentatifs de ceux qui sont vus en pratique médicale courante ?

– D'après les analyses en sous-groupe, l'effet du traitement est-il constant quelles que soient les caractéristiques

des patients ? Attention à ne pas surinterpréter les résultats des analyses en sous-groupe : ces analyses permettent de s'assurer de façon descriptive que les résultats sont toujours dans le même sens et qu'il n'y pas de variation notable de l'effet du traitement (interaction) en fonction des caractéristiques des patients, mais ne permettent pas de conclure sur un sous-groupe en particulier.

À l'issue de cette analyse, il convient de définir clairement la population cible du traitement en se basant à la fois sur les critères de sélection de l'étude et sur les patients effectivement inclus. Si un essai n'a pas inclus certains types de patients qui étaient pourtant éligibles d'après les critères de sélection, ces patients ne peuvent pas être considérés dans la définition de la population cible.

Quelle est la pertinence du contexte de soins ?

Le traitement a-t-il été utilisé dans un contexte de soins similaire à celui de la pratique quotidienne ?

À côté des traitements évalués dans l'essai, les patients étaient-ils traités conformément à l'état de l'art actuel ? L'essai s'est-il déroulé par-dessus (« on top ») la meilleure stratégie thérapeutique actuelle ?

Effets indésirables et balance bénéfice/risque

Ce point est *primordial* pour évaluer la pertinence clinique d'un résultat. À ce stade, on est *a priori* à même de répondre à la question du bénéfice apporté par le nouveau traitement. Cela pose alors l'épineux problème de savoir si le traitement est sûr.

– La balance bénéfice/risque est-elle acceptable ? Est-ce que le risque lié aux effets indésirables est acceptable compte tenu de la maladie traitée ?

• Le critère de jugement intègre-t-il directement les effets délétères ? C'est le cas par exemple avec la mortalité. Si une réduction de mortalité est démontrée, la balance bénéfice/risque est d'emblée favorable même si le traitement induit des effets indésirables sérieux (ceux-ci ne contrebalancent pas le bénéfice). Cependant, même dans ce cas, le profil de sécurité du traitement est important, par exemple pour choisir entre plusieurs traitements efficaces.

• Concernant les effets délétères mesurables, les principaux effets délétères du traitement sont-ils rapportés ? La fréquence de ces effets indésirables ne contrebalance-t-elle pas la totalité du bénéfice ?

• En dehors des événements rapportés, que sait-on des données externes, notamment de pharmacovigilance sur les événements indésirables rares (rarement observés au cours d'un essai thérapeutique) mais graves, voire fatals ?

– La balance bénéfice/risque est-elle plus favorable que celle d'autres traitements déjà existants ? C'est-à-dire le bénéfice est-il supérieur à celui des alternatives thérapeutiques sans surcroît de risque ? Ou le bénéfice est-il du même ordre que celui des autres traitements validés mais avec une meilleure tolérance ? Dans ce dernier cas, le

gain en tolérance soit être parfaitement bien démontré et être cliniquement pertinent compte tenu du bénéfice apporté.

Conclusion

Lorsque, à l'issue de cette analyse, aucune réserve méthodologique n'est raisonnablement possible et que le résultat obtenu est véritablement cliniquement pertinent, le traitement étudié a fait la preuve qu'il apporte un bénéfice clinique substantiel aux patients de la population cible et il peut donc être envisagé de manière systématique en pratique de soins (chez ces mêmes patients).

Références

1. Mismetti P, Laporte S, Cucherat M. Les essais de non-infériorité. *MT* 2005 ; 11(4) : 278-81.
2. Laporte S, Cucherat M. Multiplicité des tests statistiques au cours d'un essai clinique : plus on cherche, plus on trouve. *MT* 2006 ; 12(3).
3. Mismetti P, Laporte S. Méthodologie des essais cliniques : pourquoi et pour qui ? Les notions essentielles à une lecture critique. *MT* 2003 ; 9(3) : 68-72.
4. Laporte S, Quenet S, Mismetti P. La randomisation. *MT* 2003 ; 9(4) : 187-91.
5. Chabaud S, Cucherat M. L'analyse en intention de traiter et analyse *per* protocole. *MT* 2004 ; 10(5) : 350-5.
6. Kassai B, Gueyffier F, Boissel JP. Critères intermédiaires et critères de substitution. *MT* 2006 ; 12(2) : 96-103.
7. Quenet S, Laporte S, Cucherat M. Intervalle de confiance. *MT* 2004 ; 10(2) : 113-6.