

La lecture critique des essais thérapeutiques *The critical reading of clinical trials*

● M. Cucherat*, M. Lièvre*

La lecture critique d'un compte-rendu d'essai thérapeutique a pour objectif d'évaluer la fiabilité et la pertinence de son résultat avant sa mise en application éventuelle. En fin de compte, pour un clinicien, la question est de savoir si les résultats justifient un changement de sa pratique et, pour un décideur de santé publique, s'ils justifient de modifier l'offre de soins dans le système de santé.

La réponse à cette interrogation s'obtient par l'analyse des trois points suivants :

① **La validité interne.** Le résultat est-il fiable, autrement dit, est-il réel et non biaisé ?

② **La cohérence externe.** Ce résultat est-il confirmé par d'autres, et est-il concordant avec les autres connaissances sur le sujet ?

③ **La pertinence clinique et la représentativité.** Ce résultat représente-t-il un bénéfice cliniquement intéressant, et est-il extrapolable aux patients à traiter en pratique ?

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 (*encadré 1*).

Encadré 1. Les trois axes de l'interprétation d'un résultat d'essai clinique pour la pratique médicale.

✓ La **validité interne** permet de s'assurer que le résultat obtenu est issu d'une démarche hypothético-déductive et reflète bien la réalité, car il n'est dû ni à un biais ni au hasard. L'absence de biais dépend de la qualité méthodologique 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 n'est pas unique en son genre, mais qu'il s'intègre dans un cadre logique : confirmation par d'autres essais, cohérence avec les connaissances fondamentales, épidémiologiques, etc. La méta-analyse est un des outils d'évaluation de la cohérence externe.

✓ La **pertinence clinique** permet de s'assurer que le résultat de l'essai correspond à un effet de taille suffisante, concerne un critère cliniquement pertinent et est extrapolable à la pratique :

- intervalle de confiance de l'effet permettant d'éliminer un effet de trop petite taille, sans intérêt en pratique ;
- comparateur adapté, placebo ou traitement de référence validé ;
- patients de l'essai proches des patients vus en pratique médicale courante et non excessivement sélectionnés.

VALIDITÉ INTERNE

Validité de la démarche de recherche

Afin de respecter le principe de la méthode expérimentale, le résultat avancé doit fournir la réponse à une question préalable découlant naturellement des connaissances et des données disponibles, une hypothèse dont le test était l'objet spécifique de l'essai. En lecture critique, 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.

S'il n'est pas issu d'une démarche hypothético-déductive, le résultat est de nature inductive. Il *suggère* un effet, mais ne peut le démontrer.

Réalité statistique du résultat

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 % . Il

* EA 643, faculté Laennec, 69000 Lyon.

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, d'analyses en sous-groupes, de recherche de l'effet répétée au cours du temps, d'analyses intermédiaires non protégées.

Absence de biais

Il y a 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. L'analyse critique doit pouvoir éliminer la possibilité 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. Le **tableau I** récapitule les biais pouvant affecter l'essai thérapeutique.

COHÉRENCE EXTERNE

La cohérence externe d'un essai se juge en confrontant son résultat à la somme des connaissances accumulées sur la question.

La cohérence entre les résultats de plusieurs essais thérapeutiques de même nature est évaluée de façon idéale par la méta-analyse, principalement grâce au test d'hétérogénéité. Toujours réalisée de façon post-hoc, l'explication d'éventuelles discordances est une démarche purement inductive à laquelle on ne peut accorder que peu de confiance : on découvre toujours a posteriori une théorie pour expliquer un résultat non prévu !

La compatibilité du résultat avec les autres connaissances du domaine (données épidémiologiques, pharmacologiques, physiopathologiques, etc.) est généralement acquise, étant donné que l'hypothèse testée dans un essai a été construite à partir de ces informations.

Quatre situations différentes remettent en cause la cohérence externe d'un résultat positif :

- ❶ **L'essai est le seul disponible.** Il est impossible d'écarter une origine artéfactuelle du résultat, sa reproductibilité est inconnue.
- ❷ **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.
- ❸ **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.
- ❹ **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. Cependant, s'il est confirmé par ceux d'autres essais, ce résultat gagnera en crédibilité : il est peut-être en avance sur les connaissances fondamentales.

Biais	Origine du biais	Contrôle du biais
Biais de confusion	Le biais de confusion est le biais entraîné par l'absence de prise en considération des facteurs de confusion.	L'essai comparatif, comportant un groupe contrôle utilisé comme référence pour la mesure de l'effet du traitement, permet de prendre en compte les facteurs de confusion.
Biais de sélection	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 du traitement.	La randomisation a pour but d'éviter le biais de sélection en créant deux groupes comparables en moyenne.
Biais de suivi	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 du traitement.	Le double insu est un élément central pour empêcher l'apparition de ce biais.
Biais d'évaluation	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.
Biais d'attrition	Le biais d'attrition survient quand des patients randomisés sont écartés de 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. L'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, conduit à une destruction non aléatoire de la comparabilité initiale des groupes.	Tous les patients randomisés doivent être inclus dans l'analyse. Le taux de perdus de vue doit être proche de zéro. Pour éviter ce biais, l'analyse doit être réalisée en intention de traiter.

Tableau I. Les principaux biais de l'essai thérapeutique.

PERTINENCE CLINIQUE

L'évaluation de la pertinence clinique permet de s'assurer que le bénéfice apporté par le traitement est suffisamment important et concerne un critère cliniquement pertinent, que le rapport bénéfice/effets indésirables est acceptable, et que ce résultat est extrapolable aux situations de la pratique médicale courante.

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, il n'est pas exclu que le nouveau traitement soit, en fait, moins efficace que ce dernier. 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 versus un traitement actif, sans que ce dernier ait été évalué, le résultat en sera peu informatif. 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, on ne peut en déduire qu'il est efficace et qu'il se serait révélé supérieur au placebo. Cependant, dans certains cas où des traitements sont standard 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 standard de soins.

Dans les essais de supériorité versus 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).

Pertinence clinique du 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).
- ✓ Le critère est composite (il regroupe plusieurs événements cliniques ; par exemple, infarctus mortel et non mortel), la com-

posante qui a le plus de poids (la plus fréquente et/ou le plus modifiée par le traitement) a une faible pertinence clinique (critère non clinique, ou critère clinique secondaire).

✓ Les critères continus (par exemple : échelles) posent assez souvent des problèmes au niveau de l'interprétation de leur pertinence clinique ou de la pertinence des effets observés.

Traitements concomitants

Il faut aussi analyser les traitements interdits par le protocole et leurs taux d'utilisation effective. On peut alors découvrir que le nouveau traitement a montré une efficacité chez des patients ne recevant pas tous les traitements validés ou les recevant à dose insuffisante. Dans ce cas, la réelle efficacité du nouveau traitement utilisé conjointement avec les autres est inconnue, de même que le risque d'interaction délétère.

Représentativité des patients inclus dans l'essai

La question de l'extrapolabilité du résultat est de savoir si l'efficacité obtenue sur les patients inclus dans l'essai est informative vis-à-vis de l'efficacité du traitement pour tous les patients ciblés.

L'analyse de la population étudiée vérifie que celle-ci est constituée de patients proches des patients vus en pratique, et qu'elle n'est pas composée de patients particuliers. Cette "représentativité" des patients est acquise quand aucun type de patients vus en pratique n'a été systématiquement évincé de l'essai. Ce point est jugé d'après les critères d'inclusion et d'exclusion et les caractéristiques des patients effectivement inclus.

La présence de certains types de patients en faible proportion ne signifie pas obligatoirement que ces patients aient 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.

Un manque de représentativité de la population étudiée n'est gênant 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. Le test d'interaction des analyses en sous-groupes oriente la réponse à cette question.

Un scénario fâcheux serait un traitement inefficace chez la majorité des patients vus en pratique, mais seulement efficace chez des sujets très particuliers et chez lesquels l'essai aurait été conduit. 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 représentatifs 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 vus en pratique. L'autre est qu'il existe peu d'exemples dans lesquels 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 cependant être 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 cela est une affaire de nuance. Un résultat obtenu sur des patients notablement hypersélectionnés n'est pas aussi convaincant 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 (*encadré 2*).

Encadré 2. Essais pragmatiques ou explicatifs.

Paramètre	Essai	
	pragmatique	explicatif
Objectif	Documenter le bénéfice qu'apportera le traitement aux patients qui seront traités en pratique	Évaluer et expliquer l'effet du traitement dans des conditions très précises
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'autres facteurs
Critères de sélection	Simple, correspondant aux critères utilisés en pratique pour décider de traiter, mesurés par des examens de routine	Précis, définissant parfaitement les patients ciblés, mesurés par des techniques sophistiquées
Pertinence clinique	Forte	Faible

Taille et précision de l'effet

L'estimation de la taille de l'effet doit être suffisamment précise pour pouvoir éliminer la possibilité que l'effet puisse être petit, et donc sans intérêt pratique. Cette estimation 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 de la limite de non-infériorité dans les essais de ce nom. 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 tout entière), les choix politiques de santé publique.

● **Intervalle de confiance.** L'intervalle de confiance (IC) traduit la précision de l'estimation de la taille de l'effet réalisée

par l'essai. L'IC à 95 % est un intervalle de valeurs qui a 95 % de chances de contenir la véritable valeur du paramètre estimé. L'IC 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 du fait des fluctuations d'échantillonnages.

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

Évaluation du rapport 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 rapport les effets bénéfiques et les effets négatifs :

- ✓ "Le risque lié aux effets indésirables est-il acceptable compte tenu de la maladie traitée ?"
- ✓ "Les effets indésirables ne contrebalancent-ils pas la totalité du bénéfice ?"

L'évaluation du rapport bénéfice/risque est la partie la plus difficile de la lecture critique des essais, car elle ne fait pas seulement appel à des notions objectives. Elle implique, en effet, l'estimation du bénéfice minimal désirable et celle du risque maximal tolérable, valeurs subjectives qui peuvent varier d'une personne à l'autre. Idéalement, le patient devrait être impliqué dans la décision lorsque le rapport bénéfice/risque n'est pas nettement favorable ou défavorable. Cette solution n'est cependant pas sans soulever des problèmes éthiques : certains patients sont prêts à troquer le soulagement immédiat d'une maladie bénigne, mais pénible, contre un hypothétique événement grave dont ils ont du mal à appréhender le risque. Si l'accident survient, la responsabilité, ne serait-ce que morale, du médecin reste engagée. L'évaluation doit à ce moment non seulement faire intervenir les données propres à l'essai, mais aussi toutes les connaissances concernant le médicament et les autres options thérapeutiques. La prise en compte des phénomènes de société comme le niveau de tolérance au risque et l'exigence de soins est aussi importante pour aborder ce délicat problème, particulièrement si l'on est décideur de santé publique.

Les éléments à prendre en compte sont les suivants :

- ✓ la gravité des effets indésirables par rapport à la gravité de la maladie,
- ✓ la gravité des effets indésirables par rapport à celle des événements évolutifs évités par le traitement.

● **La fréquence des effets indésirables.** À cet égard, il faut remarquer que les essais thérapeutiques ne sont pas capables de mettre en évidence des événements graves de fréquence très faible. L'information provenant d'autres essais du médicament doit être prise en compte pour évaluer le rapport bénéfice/risque. La fréquence des effets indésirables est un paramètre capable de déterminer leur gravité : des événements bénins, mais pénibles et fréquents, peuvent peser autant que de rares événements graves.

● **Le rapport bénéfice/risque des autres options thérapeutiques.** Des traitements qui auraient un rapport bénéfice/risque dans l'absolu de l'essai analysé peuvent être inacceptables en comparaison aux autres traitements disponibles.

CONCLUSION

La lecture critique des essais thérapeutiques est une des bases principales de la médecine factuelle (*evidence-based medicine*). Bien qu'elle s'appuie avant tout sur une analyse méthodologique et statistique des données, elle fait appel également aux connaissances cliniques et humaines, ce qui en fait la grandeur et la difficulté. Son développement à la faveur de la réforme du deuxième cycle des études médicales ne peut qu'améliorer la qualité des soins, en réduisant le nombre de décisions prises à partir d'arguments émotionnels ou politiques. ■

POUR EN SAVOIR PLUS

– Bouvenot G, Eschwege E. *Therapeutic trial. Principles of interpretation. Rev Prat* 1991 ; 41, 19 : 1853-7.
 – Bouvenot G, Vray M. *Internal validity and impact of the results of clinical trials (éditorial). Rev Med Int* 1994 ; 15, 1 : 9-12.

– Bouvenot G, Vray M. *Essais cliniques : théorie, pratique et critique. Paris : Médecines-Sciences Flammarion ; 1996.*

– Cucherat M, Boissel JP, Leizorovicz A. *La méta-analyse des essais thérapeutiques. Paris : Masson ; 1997.*

– Moher D. *Assessing the quality of randomized controlled trials : an annotated bibliography of scales and checklists. Controlled Clinical Trials* 1995 ; 16 : 62-73.

– Peto R, Pike MC, Armitage P et al. *Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Br J Cancer* 1976 ; 34 : 585-612.

– Pocock SJ. *Clinical trials. A practical approach. Chichester : John Wiley & Sons ; 1983.*

– Poynard T. *Évaluation de la qualité méthodologique des essais thérapeutiques randomisés. Presse Med* 1988 ; 17 : 315-8.

– Schulz KF, Chalmers I, Hayes RJ, Altman DG. *Empirical evidence of bias : dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA* 1995 ; 273 : 408-12.

– Boissel JP, Bossard N, Cucherat M et al. *L'information thérapeutique. Paris : Masson ; 2000.*

– Buyse M. *Analysis of clinical trial outcomes : some comments on subgroup analyses. Controlled Clinical Trials* 1989 ; 10 : 187S-194S.

– Chalmers TC, Smith SJ, Blackburn B. *A methode for assessing the quality of a randomized clinical trial. Controlled Clinical Trial* 1981 ; 2 : 31-49.

OUI, JE M'ABONNE AU BIMESTRIEL *La Lettre du Pharmacologue*

Merci d'écrire nom et adresse en lettres majuscules

Collectivité

à l'attention de

Particulier ou étudiant

M., Mme, Mlle

Prénom

Pratique : hospitalière libérale autre.....

Adresse e-mail

Adresse postale

.....

Code postal Ville

Pays.....

Tél.....

Merci de joindre votre dernière étiquette-adresse en cas de réabonnement, changement d'adresse ou demande de renseignements.

À découper ou à photocopier

ABONNEMENT : 1 an

FRANCE/DOM-TOM/EUROPE

- 74 € collectivités
- 59 € particuliers
- 37 € étudiants*

*joindre la photocopie de la carte

ÉTRANGER (AUTRE QU'EUROPE)

- 94 € collectivités
- 79 € particuliers
- 57 € étudiants*

*joindre la photocopie de la carte

+

ET POUR 10 € DE PLUS !

- 10 €, accès illimité aux **26 revues** de notre groupe de presse disponibles sur notre site **vivactis-media.com** (adresse e-mail gratuite)

+

RELIURE

- 10 € avec un abonnement ou un réabonnement

MODE DE PAIEMENT

- carte Visa, Eurocard Mastercard N°

Signature : _____ Date d'expiration

- chèque (à établir à l'ordre de **La Lettre du Pharmacologue**)
- virement bancaire à réception de facture (réservé aux collectivités)

EDIMARK - 62-64, rue Jean-Jaurès - 92800 Puteaux

Tél. : 01 41 45 80 00 - Fax : 01 41 45 80 25 - E-mail : contacts@vivactis-media.com

Total à régler €
À remplir par le souscripteur