

Item 169 : Évaluation thérapeutique et niveau de preuve

COFER, Collège Français des Enseignants en Rhumatologie

Date de création du document 2010-2011

Table des matières

ENC :	3
SPECIFIQUE :	3
I Différentes approches d'évaluation de l'efficacité d'une modalité thérapeutique.....	3
II Principales sources d'information et niveaux de preuve.....	5
II.1 Principaux types d'articles.....	5
II.2 Types de revues médicales.....	5
II.3 Niveaux de preuve scientifique.....	6
II.3.1 Études à fort niveau de preuve.....	6
II.3.2 Études à niveau de preuve intermédiaire.....	6
II.3.3 Études à faible niveau de preuve.....	6
III Lecture critique d'une information thérapeutique.....	6
IV Limites des essais thérapeutiques.....	8

OBJECTIFS

ENC :

- Argumenter l'évaluation d'une thérapeutique et les niveaux de preuve des principales sources d'information.
- Argumenter une publication d'essai clinique ou une méta-analyse et critiquer une information thérapeutique.

SPECIFIQUE :

- Connaître les diverses approches permettant d'évaluer l'efficacité d'une modalité thérapeutique.
- Argumenter les niveaux de preuve des principales sources d'information et savoir en établir la hiérarchie.
- Connaître les limites des essais cliniques en pratique quotidienne.
- Savoir critiquer une information thérapeutique.
- Commenter les limites du niveau de preuve des thérapeutiques non médicamenteuses.

I DIFFÉRENTES APPROCHES D'ÉVALUATION DE L'EFFICACITÉ D'UNE MODALITÉ THÉRAPEUTIQUE

Les essais thérapeutiques ont pour but d'évaluer l'efficacité d'un médicament, d'une technique chirurgicale ou non médicamenteuse. Après les études toxicologiques faites sur l'animal, le médicament doit passer par quatre phases d'essais conduits chez des sujets volontaires, sains ou malades :

- l'essai de phase I a pour but de réaliser des études pharmacocinétiques chez des volontaires sains dans des laboratoires ayant l'autorisation pour ce type de recherche ;
- les essais de phase II ont pour mission de déterminer les modalités optimales d'administration (voie d'administration, posologie) et sont conduits chez des malades souvent en groupes parallèles se distinguant par la dose administrée ou les modalités d'administration ;
- les essais de phase III permettent de confirmer l'efficacité du médicament sur un grand nombre de sujets malades et sa tolérance. C'est à l'issue de ces essais que les autorités sanitaires délivreront l'autorisation de mise sur le marché (AMM) ;
- les essais de phase IV (succédant à la mise sur le marché) permettent d'étudier les effets et la tolérance à long terme.

Dans l'idéal, les essais thérapeutiques de phase III doivent être prospectifs, comparatifs, randomisés et en double insu, produisant ainsi le niveau de preuve le plus élevé.

Un essai est comparatif lorsqu'il inclut deux groupes de malades, l'un recevant le traitement et l'autre non, afin de juger objectivement de ses effets. Le traitement à l'étude peut être comparé au meilleur traitement connu (de référence) de la pathologie ou à un placebo. La comparaison d'un traitement à un placebo permet d'examiner la composante de son activité liée au principe pharmacologique du médicament, et d'évaluer objectivement ses effets indésirables. La randomisation, procédé qui consiste à s'en remettre au hasard pour attribuer à chaque patient un des traitements étudiés, permet de constituer des groupes comparables et d'appliquer aux essais thérapeutiques les théories et modèles statistiques. Tout le bénéfice apporté par la randomisation initiale sur la comparabilité des groupes peut être perdu au fil de l'étude par des différences de prise en charge des patients ou d'appréciation de l'évolution de leur état selon le traitement qu'ils reçoivent. Cela peut être limité en masquant au patient et à l'investigateur la nature réelle du traitement administré, ce qui correspond à la méthode en double insu. Les études en simple aveugle (seul l'investigateur connaît la nature du traitement) et les études ouvertes (malade et investigateur connaissent le traitement pris) ont une mauvaise validité interne du fait d'un biais de suivi et d'évaluation. Il en est de même des études cas-témoins et des études rétrospectives qui offrent des niveaux de preuve inférieurs (tableau 24.1).

Les plans expérimentaux en groupes parallèles sont les plus adaptés. Chaque malade inclus reçoit un seul des traitements comparés. Les essais croisés (cross over) ou les plans en carré latin consistent à administrer successivement à chaque malade plusieurs des traitements comparés. Cette méthode sous-entend que le malade revient à son état initial au début de chaque période thérapeutique. D'autre part, elle ne peut s'appliquer qu'à des maladies stables ou cycliques, d'intensité constante

Tableau 24.1. Classification des études selon le niveau de preuve, établie par la HAS.

Niveau 1	Essais contrôlés randomisés avec résultats indiscutables (méthodologiquement)
Niveau 2	Essais contrôlés non randomisés bien conduits
Niveau 3	Essais prospectifs non contrôlés bien menés (suivis de cohorte par exemple)
Niveau 4	Études cas-témoin, essais contrôlés présentant des biais
Niveau 5	Études rétrospectives et cas cliniques, toute étude fortement biaisée

II PRINCIPALES SOURCES D'INFORMATION ET NIVEAUX DE PREUVE

II.1 PRINCIPAUX TYPES D'ARTICLES

L'**article original** correspond à un travail personnel des auteurs confronté aux connaissances du moment. Dans le domaine de la thérapeutique, il s'agit dans le meilleur des cas d'un essai prospectif, comparatif, randomisé et en double insu. Lorsque l'étude porte sur un traitement innovant ou lorsqu'elle améliore les connaissances concernant un traitement ancien, sa publication est souvent accompagnée d'un **éditorial**. L'éditorial est habituellement demandé par le comité de rédaction à un spécialiste de la pathologie ou du traitement en question qui peut y présenter des hypothèses, replacer le travail d'un article original dans une perspective plus vaste ou en limiter la portée.

La **mise au point** est centrée sur les acquisitions récentes concernant un sujet limité. La **revue générale** a pour but de rassembler des informations plus complètes sur un thème souvent plus large.

La **revue systématique** est une revue générale répondant à une méthodologie rigoureuse de recherche des données bibliographiques, de sélection des études et d'extraction des données. La présentation des données peut se faire en intégrant les données numériques de plusieurs études selon une *méta-analyse*.

Le **cas clinique** est une observation intéressante des résultats ou effets indésirables d'une modalité thérapeutique. Il a peu de valeur scientifique.

L'**article préliminaire** présente des résultats préliminaires d'une étude qui peuvent donc porter à caution et doivent être intégrés avec prudence.

II.2 TYPES DE REVUES MÉDICALES

Les revues à comité de lecture sont à privilégier car le comité de lecture est chargé de sélectionner les manuscrits soumis. Le caractère prestigieux ou non d'une revue est reflété par le coefficient d'impact (*impact factor*) qui est la moyenne du nombre de citations (dans toutes les revues), au cours d'une certaine période, des articles publiés dans la revue en question.

Les revues sans comité de lecture n'ont pas la même valeur scientifique et n'ont souvent pas de coefficient d'impact.

II.3 NIVEAUX DE PREUVE SCIENTIFIQUE

Pour chaque étude, le niveau de preuve scientifique doit être analysé. Cette notion caractérise la capacité de l'étude à répondre à la question posée. Le niveau de preuve est apporté par la « force » du protocole de l'étude. Cette capacité se juge sur la correspondance de l'étude au cadre de travail (sujet, population, critère de jugement), l'adéquation du protocole à la question posée, l'existence ou non de biais importants dans la réalisation de l'étude, la puissance de l'étude, notamment la taille de l'échantillon. Une classification peut être déterminée, adaptée du score de Sackett.

II.3.1 Études à fort niveau de preuve

Ce sont des études contrôlées, randomisées de forte puissance, c'est-à-dire avec un risque alpha et bêta faible.

Le risque alpha, ou risque de première espèce, correspond au risque de conclure à une différence statistiquement significative entre les deux groupes soumis à comparaison alors qu'elle n'existe pas.

Le risque bêta, ou risque de deuxième espèce, correspond au risque de conclure à l'absence de différence statistiquement significative entre les deux groupes soumis à comparaison alors qu'elle existe.

La puissance de l'étude se définit comme $1 - \text{bêta}$.

II.3.2 Études à niveau de preuve intermédiaire

Elles correspondent à une présomption scientifique de l'effet du traitement étudié. Il s'agit des études contrôlées, randomisées de faible puissance, c'est-à-dire avec un risque alpha et bêta élevé.

II.3.3 Études à faible niveau de preuve

Il s'agit des autres types d'études non randomisées.

III LECTURE CRITIQUE D'UNE INFORMATION THÉRAPEUTIQUE

La lecture critique d'un essai thérapeutique a pour objectif d'évaluer la fiabilité et la pertinence de son résultat avant sa mise en application éventuelle. Pour cela, il est nécessaire d'analyser les trois points suivants :

- la *validité interne* : le résultat est-il fiable, c'est-à-dire réel et non biaisé ?
- la *cohérence externe* : le résultat est-il confirmé par d'autres ou concordant avec les connaissances sur le sujet ?
- la *pertinence clinique* : ce résultat représente-t-il un bénéfice cliniquement intéressant ?

Lors de la lecture d'un essai thérapeutique, les principales questions à se poser sont les suivantes :

- **étude comparative** ? Seule l'étude comparative permet d'établir un lien de causalité entre l'amélioration d'une maladie et la prise du traitement ;
- **étude randomisée** ? C'est-à-dire avec une répartition aléatoire des malades dans les groupes, utilisant une méthode informatique, une liste de nombres au hasard ou une table de permutation ;
- **étude contrôlée** versus traitement de référence ou placebo ?
- **étude en double insu** ? C'est-à-dire où ni l'investigateur, ni le malade ne connaissent le traitement attribué, ce qui permet d'éliminer la subjectivité des deux ;
- **définition de la population sélectionnée** ou critères d'inclusion ? Les conclusions de l'essai ne s'appliquent strictement qu'à la population ainsi définie ;
- **définition des traitements à l'étude et plan expérimental** ?
- **description des tests statistiques** ? Ils doivent être précisés et l'analyse du critère principal doit conduire à une différence statistiquement significative par rapport au groupe contrôle si l'on veut démontrer une supériorité ($p < 0,05$ %) ;
- **calcul du nombre de sujets nécessaires** ? Ce calcul correspond aux effectifs théoriques indispensables pour mettre en évidence la différence escomptée entre le traitement et le placebo. Le nombre de sujets inclus doit être en accord avec ce calcul ;
- **analyse en intention de traiter** ? La comparaison finale de l'efficacité doit porter sur la totalité des patients inclus, quelles qu'aient été les modalités réelles de traitement. L'analyse devra donc inclure les patients qui n'ont pas reçu le traitement pour des raisons diverses (effets indésirables, évolution clinique, raisons personnelles, etc.). L'analyse « sous traitement » ou per protocol juge des effets sur des groupes effectivement traités dans les termes du protocole, mais dont la comparabilité originelle n'est pas garantie ;
- **critère principal d'évaluation** ? Ce critère doit être unique, cliniquement pertinent, consensuel. Il faut préférer les critères ayant une réelle signification clinique (morbidity, mortalité) aux critères biologiques ou intermédiaires. Le critère principal doit avoir des qualités métrologiques : aisément mesurable, en corrélation avec l'évolution clinique. Une échelle, un score, un indice doivent avoir été validés ;
- **nombre de perdus de vue, de sorties d'essai** ?
- **les résultats vont-ils m'aider à soigner mes patients** ? Les critères de jugement sont-ils pertinents et les résultats sont-ils extrapolables à la population générale ? On peut apprécier l'intérêt du traitement en calculant le nombre de sujet à traiter (NST) pour éviter un événement : $1/(A - B)$ (A : fréquence de l'événement, spontanée ou sous traitement comparateur, B : fréquence sous le nouveau traitement).

Pour évaluer le rapport bénéfice/risque, on peut mettre en parallèle le NST pour éviter un événement et le NST pour voir apparaître un effet indésirable.

Une grille de lecture peut être mise au point et permet ainsi de passer en revue tous ces points.

IV LIMITES DES ESSAIS THÉRAPEUTIQUES

L'objectif principal d'un essai thérapeutique est d'identifier des relations causales entre guérison ou amélioration et prise d'un traitement. En pratique, on mesure des associations statistiques entre des critères qui représentent, de manière plus ou moins exacte, la maladie. Une association statistique ne signifie pas relation causale. Les premières limites des essais thérapeutiques sont la taille de l'échantillon et les caractéristiques des patients étudiés. En effet, l'échantillonnage étant en général le plus petit possible, les patients sont hautement sélectionnés. Ainsi, les résultats des essais sont parfois difficiles à généraliser à la population traitée « dans la vie normale ». Il est classique, par exemple, d'éliminer dans les essais les personnes âgées, les femmes enceintes, les enfants, les formes sévères de la maladie, les patients ayant des antécédents lourds ou plusieurs traitements, etc. Une autre limite vient de la grande spécificité de la question posée, ne permettant pas d'étudier d'autres aspects de l'intervention. Par exemple, le traitement d'une pathologie rhumatologique pourra être étudié par rapport à la réduction de la douleur sans tenir compte de la qualité de vie, du handicap, de la satisfaction des patients, etc., ce qui laisse un certain nombre de questions en suspens. La plupart des essais sont conçus pour apporter une réponse dans les plus brefs délais, ce qui peut laisser dans l'incertitude quant à l'efficacité à long terme. Il en est de même pour les effets indésirables. Très souvent, les nouveaux traitements sont comparés à un placebo et non aux traitements déjà disponibles. Ainsi, pour le praticien, la question persiste : « Le nouveau traitement est-il meilleur que ceux dont je dispose déjà ? »

Dans les essais concernant les thérapeutiques non médicamenteuses, les règles habituelles sont souvent impossibles à appliquer. Il en est ainsi de l'aveugle : comment comparer en double aveugle deux techniques chirurgicales, deux prises en charge physiques différentes ? De la même manière, le placebo est la plupart du temps inapplicable.