

UE10 - Santé, Société, Humanité - Épidémiologie
Le 03/05/2017 de 15h30 à 17h30
Dr Fadia Dib
Ronéotypeur : Anaïs Rimbert
Ronéoficheur : Gaspard de Moustier

Cours 5 – Evaluation thérapeutique

Le cours a peu changé par rapport à l'année dernière.

Plan

I. Objectif de l'évaluation thérapeutique

II. Etudes quasi expérimentales (avant-après)

- a) Facteurs pouvant expliquer la réponse à un traitement
- b) Inconvénients des études avant-après

III. Essais cliniques contrôlés randomisés

- a) Principe
- b) Validité interne
- c) Validité externe
- d) Nombre de sujets nécessaires

IV. Phases d'études cliniques

I.Objectif de l'évaluation thérapeutique

L'évaluation thérapeutique a pour but d'évaluer l'efficacité et la tolérance(++ une thérapeutique de même efficacité mais comprenant moins d'effets indésirables sera plus intéressante) d'une intervention thérapeutique. Elle s'oppose à l'évaluation diagnostic, qui évalue un test diagnostic.

L'intervention thérapeutique comprend tout ce qui est médicaments, dispositifs médicaux, conseils hygiéno-diététiques, chirurgie, acupuncture, rééducation...

exemple que nous allons reprendre tout au long du cours : le Programme Sophia Asthme, étude menée par l'assurance maladie prévoit un accompagnement des personnes asthmatiques via des appels téléphoniques (les appels sont ici l'intervention thérapeutique). Le but est de savoir si cela améliore significativement la prise en charge et diminue la morbidité.

II. Etudes quasi expérimentales (avant-après)

Il s'agit d'une étude comparative de l'état des patients avant l'intervention thérapeutique et de leur état après intervention. Ces études ne sont néanmoins pas optimales dans le domaine de l'évaluation thérapeutique, en raison de biais. On y préfère donc les essais contrôlés randomisés (ECR), utilisés comme gold standard ou méthode de référence.

Exemple : On réalise des injections intradiscales de dextrose (=intervention thérapeutique) chez des patients lombalgiques chroniques. Le critère de jugement utilisé sera la douleur, évaluée par l'échelle EVA. Lorsqu'on observe la douleur baseline (avant injection) et la douleur à deux mois, on va être tenté de conclure que l'amélioration est due au traitement.

Injections intradiscales de dextrose chez les patients lombalgiques chroniques

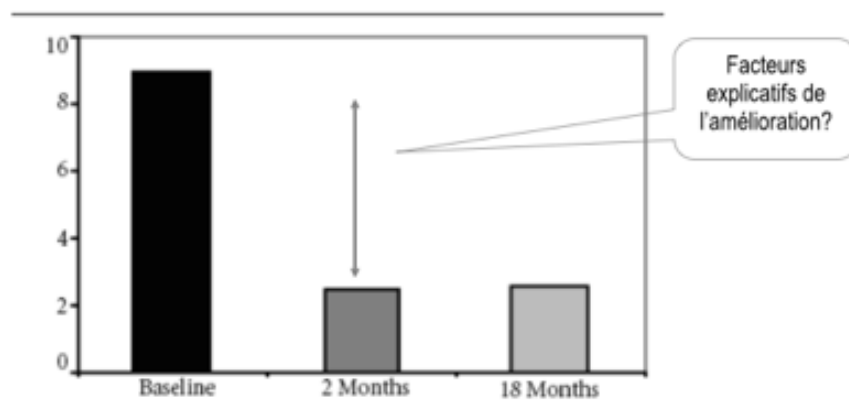


Fig 4. Pain scores in responders

a) Facteurs pouvant expliquer la réponse à un traitement

➤ L'histoire naturelle de la maladie.

L'effet du médicament est « confondu » avec l'évolution spontanée de la maladie. Elle concerne notamment les maladies aiguës et les maladies chroniques qui évoluent par poussées, c'est-à-dire des pathologies qui tendent spontanément vers une amélioration des symptômes voire la guérison.

Exemple : La grippe dure en moyenne 5-6 jours. Ainsi, si l'on mène une étude avant-après, on observera une guérison. Néanmoins, on ne peut pas attribuer cette amélioration à un traitement prescrit au cours de l'épisode viral.

➤ L'effet Hawthorne

Il s'agit d'une amélioration de la maladie liée à des modifications du comportement du patient, liées à des mécanismes psychologiques, à l'attention particulière accordée au patient lorsqu'il participe à une étude.

Exemple : changement de l'alimentation, de l'activité physique, motivation à arrêter de fumer, ...

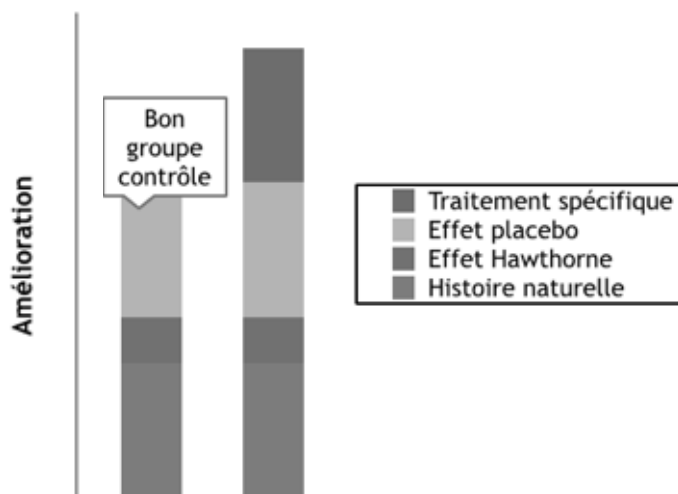
➤ L'effet placebo

Il s'agit des effets thérapeutiques non spécifiques associés à l'administration d'un placebo. Un placebo est une substance inerte dénuée d'activité pharmacologique mais perçue par le malade comme un médicament. Il a deux composantes : - l'effet malade : le simple fait de recevoir un traitement même inefficace peut modifier positivement l'évolution de la maladie ;

- l'effet médecin : le fait de savoir que le patient reçoit le nouveau traitement (ou ne le reçoit pas) amène le médecin à manquer d'impartialité.

➤ L'effet réel du traitement

Il faut éliminer une action possible des trois facteurs précédemment cités afin de conclure à un effet réel du traitement évalué. On parle de contrôler les trois facteurs (histoire naturelle de la maladie, effet Hawthorne, effet placebo) car pour conclure à l'effet réel, on utilise un groupe contrôle.



Je place ce schéma là pour montrer comment le groupe contrôle permet de contrôler les autres facteurs que celui du traitement. En effet on observe une différence d'amélioration significative entre le groupe de gauche (le groupe contrôle, qui reçoit le placebo) et le groupe de droite (groupe expérimental, qui reçoit le traitement) : cette différence est l'effet réel du traitement. Nous verrons plus tard comment choisir un bon groupe contrôle.

En cas d'absence de groupe contrôle, les conclusions de l'étude avant-après seront erronées, biaisées.

b) Inconvénients des études avant-après

L'absence de contrôle des facteurs d'amélioration autres que le traitement (l'histoire de la maladie, l'effet Hawthorne et l'effet placebo) entraîne des conclusions biaisées.

L'évaluation de l'amélioration sur un seul groupe de patients entraîne une évaluation absolue (non relative à un groupe de patients contrôle).

Les études avant-après ne permettent pas de créer un lien de causalité entre le traitement et l'amélioration, donc on ne peut pas conclure à l'efficacité du produit.

Exception : parfois l'étude randomisée n'est pas possible, on accepte alors de conclure par une étude avant-après mais si et seulement si on utilise un groupe contrôle (dont on va mimer la randomisation).

III. Essais cliniques contrôlés randomisés (+++)

a) Principe

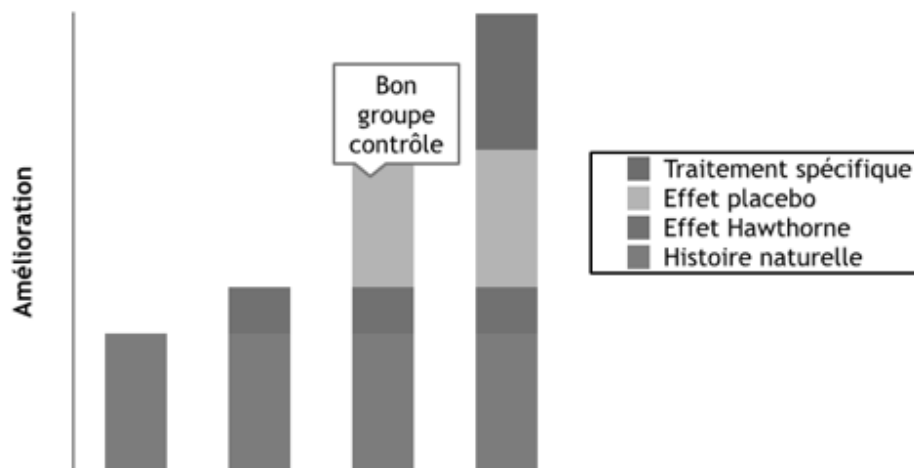
On prélève un échantillon dans la population cible, c'est-à-dire dans l'ensemble de la population pour laquelle on espère généraliser l'intervention thérapeutique (ex pour l'étude Sophia Asthme : la population asthmatique dont l'asthme est mal maîtrisé). A partir de cet échantillon, on va distinguer un groupe expérimental à qui on donne le traitement à évaluer, et un groupe contrôle à qui on donne un placebo ou le traitement de référence. Ces deux groupes seront évalués sur des critères de jugement.

Un critère de jugement est le paramètre à mesurer pour dire si l'intervention est efficace ou non. Il doit être pertinent cliniquement, et va servir à calculer le nombre de sujets nécessaires à l'essai. *Ex de l'étude Sophia Asthme : les critères de jugements peuvent être la qualité de vie, le nombre d'hospitalisation, le nombre d'arrêt de travail, la symptomatologie.*

On distingue les critères « durs » qui sont des critères mesurés objectivement, ex : *nombre d'hospitalisations, nombre de jours d'hospitalisation, nombre de décès, résultats d'examens biologiques* ; et les critères « mous » qui sont mesurés subjectivement, déclarés par le patient, ex : *douleur, qualité de vie.*

Les deux groupes (expérimental et contrôle) sont comparés sur les mêmes critères de jugement.

Le groupe contrôle doit être comparable au groupe expérimental, afin de contrôler les facteurs de biais que sont l'histoire naturelle de la maladie, l'effet Hawthorne et l'effet placebo (*la prof demande d'être capable de reconnaître les biais d'une étude++*). Ainsi, les différences observées entre les groupes seront uniquement attribuables à l'effet réel de l'intervention thérapeutique.



b) Validité interne

La validité interne est la qualité méthodologique de l'étude. Elle évalue à quel point les résultats d'une étude sont corrects pour la population étudiée. Elle dépend donc de la méthodologie utilisée lors de l'étude et de la présence de biais.

Le biais est une différence systématique entre l'effet mesuré et l'effet réel. La présence de biais entraîne une mauvaise validité interne : on peut alors remettre en cause la conclusion clinique de l'étude. Il existe différents types de biais :

➤ Le biais de sélection

Il s'agit de la constitution de deux groupes non comparables, c'est-à-dire que les caractéristiques des patients ne sont pas réparties de façon uniforme entre les 2 groupes. Cela concerne l'âge, le sexe, les morbidités associées (ex : obésité), la sévérité de la maladie ainsi que des caractéristiques inconnues qui pourraient influencer l'évolution de la pathologie.

Ex de l'étude Sophia Asthme : si les groupes proviennent de régions géographiques différentes, alors l'exposition à la pollution va différer et induire plus de crises d'asthme dans l'un des deux groupes, sans que cela soit en rapport avec le traitement testé.

Comment faire pour assurer la comparabilité initiale des groupes ? réponse : la randomisation !

Il s'agit d'allouer le traitement (expérimental ou placebo) de façon aléatoire. Ainsi, on devrait obtenir des groupes comparables avec une même répartition en âge, sexe, etc. La randomisation ne garantit pas que les deux groupes soient identiques, mais augmente fortement la probabilité qu'ils le soient.

Attention ! Même si l'on effectue une randomisation, si l'effectif est limité, il existe un risque que les groupes ne soient pas comparables (*piège déjà tombé aux partiels+++*). Pour résumer, la randomisation permet d'assurer, si l'échantillon est suffisamment grand, l'équilibre des caractéristiques connues et inconnues des patients dans les deux groupes, et donc la comparabilité initiale des groupes.

La randomisation est régie par deux principes :

- Le principe d'assignation secrète, ce qui signifie simplement que seul le hasard doit décider du traitement administré. La séquence d'allocation, ou tirage au sort, doit être cachée au médecin, qui pourrait influencer son choix d'inclure le patient dans l'étude, et créer ainsi un biais de sélection. La randomisation doit donc être imprévisible.
- La clause d'ambivalence (équipoise), qui implique que tout patient inclus dans un essai est à même de recevoir indifféremment l'un ou l'autre des traitements que le hasard lui assignera. Il ne doit alors pas avoir de contre-indication à l'un ou l'autre des traitements (traitement expérimental et traitement de référence). Exemple : comparaison de deux traitements de la douleur lombaire. Le traitement du groupe 1 consiste en des injections intra-discales, ce qui est contre-indiqué pour les patients ayant des troubles de la coagulation. Le groupe 2 reçoit quant à lui des AINS, dont la contre-indication est l'ulcère gastro-duodénal. Cela signifie que les patients inclus dans l'étude ne doivent présenter ni troubles de la coagulation, ni ulcère gastro-duodénal, et pourront être assignés dans le groupe 1 ou le groupe 2 sans biais de sélection.

➤ Le biais de performance/mesure

Le biais de performance est la différence systématique dans la prise en charge des patients, c'est-à-dire que la prise en charge de chaque patient va être un peu différente. Ce biais peut avoir lieu dans la réalisation de l'intervention, dans l'intensité du suivi, dans la prescription et l'utilisation de co-interventions.

Par exemple, si l'on pense qu'un patient prend le traitement expérimental, on aura tendance à les encourager à être parfaitement observant, on leur prescrira plus d'examens complémentaires, etc.

Le biais de mesure est la différence systématique lors de l'évaluation des critères de jugement, que ce soit la mesure des critères d'efficacité ou celle des effets indésirables (donc des critères de tolérance).

Par exemple, si l'on pense qu'un patient prend le traitement expérimental, on aura tendance à influencer le patient dans la mesure de sa douleur, en lui disant qu'il a bonne mine, que ça a l'air d'aller mieux.

Comment faire pour maintenir la comparabilité initiale des deux groupes ?

Réponse : l'insu !

Si nous avons cherché à construire des groupes comparables au début de traitement par un tirage au sort/randomisation, il faut impérativement que ces groupes restent comparables tout au long de l'essai pour contrôler tout ce qui n'est pas le traitement évalué. L'insu consiste à ne pas informer les différents participants à l'étude (médecins évaluateurs, équipe soignante, patients) du type de traitement reçu par les patients, depuis le début de l'étude jusqu'au recueil du critère d'efficacité. Cela va permettre d'éviter des modifications systématiques de leur comportement pouvant biaiser les résultats de l'étude.

- Si le malade connaît la nature de son traitement (expérimental ou placebo), il y a un risque de :
 - Modifier son comportement dans un sens différent selon le groupe auquel il appartient
 - Modifier l'observance du traitement
 - Retenir sur son mode de vie
 - Fausser son propre jugement sur l'effet qu'il ressent
- Si le médecin évaluateur connaît la nature du traitement reçu par le malade, il y a un risque de :
 - Modification de la qualité des soins
 - Modification de l'écoute et des conseils donnés
 - Modification de la surveillance des effets indésirables et l'évaluation des traitements
 - Modifier la façon d'interroger le patient
 - Influencer la mesure de l'efficacité.

Ainsi, l'insu est le terme politiquement correct pour dire double-aveugle. Il permet d'assurer la meilleure égalité possible du suivi, indépendamment des effets des traitements. Enfin, l'insu permet d'éviter les biais de performance et de mesure.

Comment faire pour obtenir l'insu ?

Tout d'abord, le placebo doit être indiscernable du traitement actif afin d'être perçu par le malade comme un médicament malgré l'absence d'activité pharmacologique. Pour cela, il faut :

- Une similarité des caractéristiques physiques (les comprimés ont le même aspect, couleur, consistance, forme, goût. Si le traitement expérimental s'administre en injections ou perfusions, le placebo sera alors du sérum physiologique, et la poche et la tubulure seront alors masquées).
- Une similarité des effets indésirables (ex : si le traitement expérimental colore les urines en rouge, alors le placebo doit colorer les urines en rouge).
- Une similarité du schéma posologique.

Néanmoins, l'insu n'est pas toujours possible à obtenir pour des questions éthiques ou logistiques, notamment en cas de chirurgie ou d'éducation thérapeutique. Lorsque l'intervention thérapeutique est un dispositif ou appareil médical, on peut ruser en laissant l'appareil éteint, ou lorsque l'intervention thérapeutique est l'acupuncture, on peut utiliser des aiguilles superficielles.

Il est d'autant plus important de choisir un critère de jugement principal « dur »/objectif lorsque l'insu n'est pas respecté ou n'est pas possible (+++). En effet, si l'insu n'est pas respecté, alors cela signifie que le patient et/ou le médecin connaît l'attribution du traitement (expérimental ou placebo/de référence). Il y a alors un fort risque de biais pour des critères « mous »/subjectifs tels que la douleur ou la qualité de vie, on cherche donc à limiter ce risque en utilisant des critères « durs »/objectifs tels que des résultats d'examens biologiques.

L'évaluation du critère de jugement comparable dans les deux groupes se fait

- En situation d'insu du traitement reçu
- En standardisant les conditions et techniques de mesures
- Si possible, avec une centralisation de la mesure du critère de jugement pour améliorer la reproductibilité. *Par exemple, si l'interprétation des dosages biologiques a lieu dans des laboratoires différents (mesure non centralisée), alors on obtiendra des résultats différents. Il en est de même pour des examens opérateur-dépendants comme les radiographies, dont l'interprétation va différer selon les examinateurs (« la qualité ne sera pas la même s'il s'agit d'un externe ou d'un senior »).*

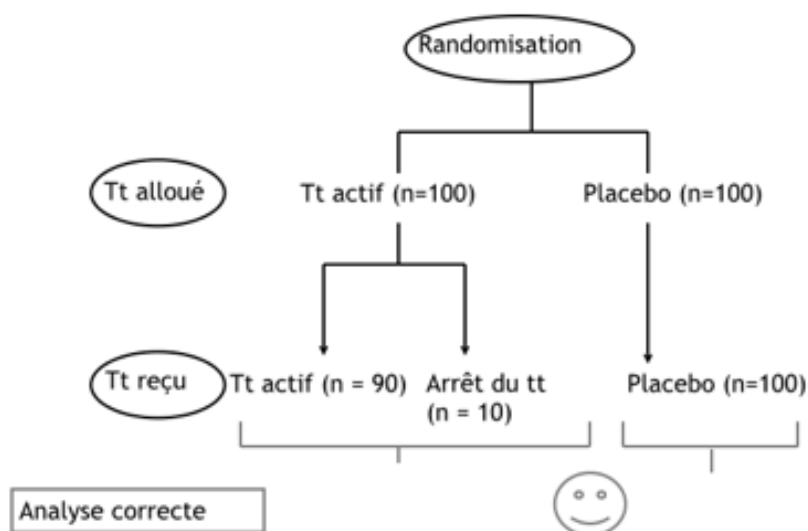
➤ Le biais d'attrition

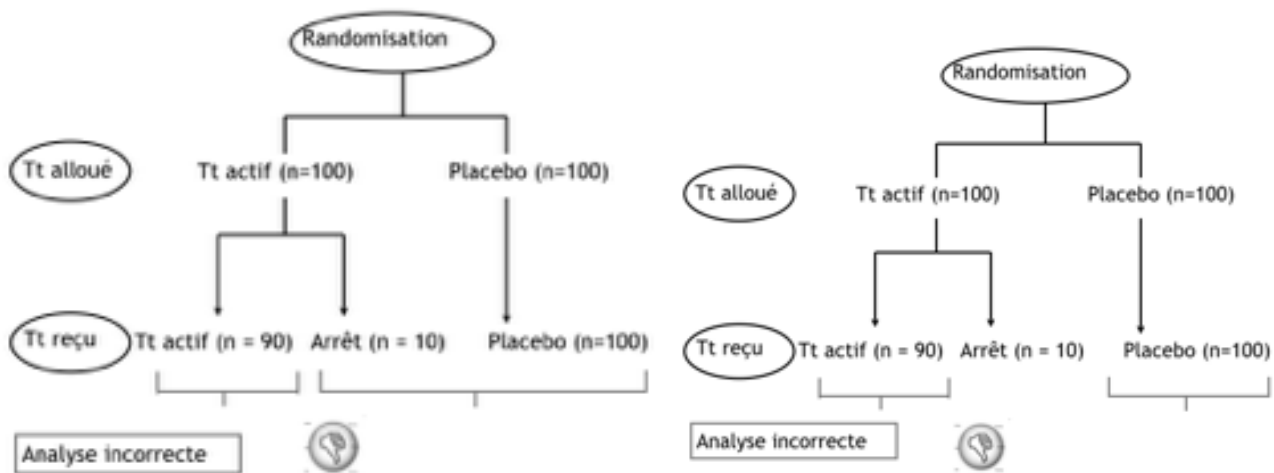
L'attrition est le contraire de la rétention, c'est la différence systématique survenant au moment de l'analyse statistique, quand des patients randomisés sont écartés de l'analyse statistique. Le biais d'attrition correspond, comme son nom l'indique, à l'attrition (diminution) de l'effectif des patients au cours de l'étude du fait du retrait de certains patients dans l'analyse, tels les patients perdus de vue, ou les patients qui sont non analysés dans leur groupe (*par exemple, on analyse uniquement des patients de sexe féminin et de moins de 40ans*).

Ne pas prendre en compte ces malades dans l'analyse peut être source de biais car il est impossible de démontrer a priori que ces données manquantes sont indépendantes du traitement évalué. *Par exemple, un patient sera perdu de vue, il aura arrêté le traitement du fait de trop nombreux effets indésirables.*

Comment faire pour maintenir la comparabilité initiale des groupes ? Réponse : l'analyse en intention de traiter (ITT)

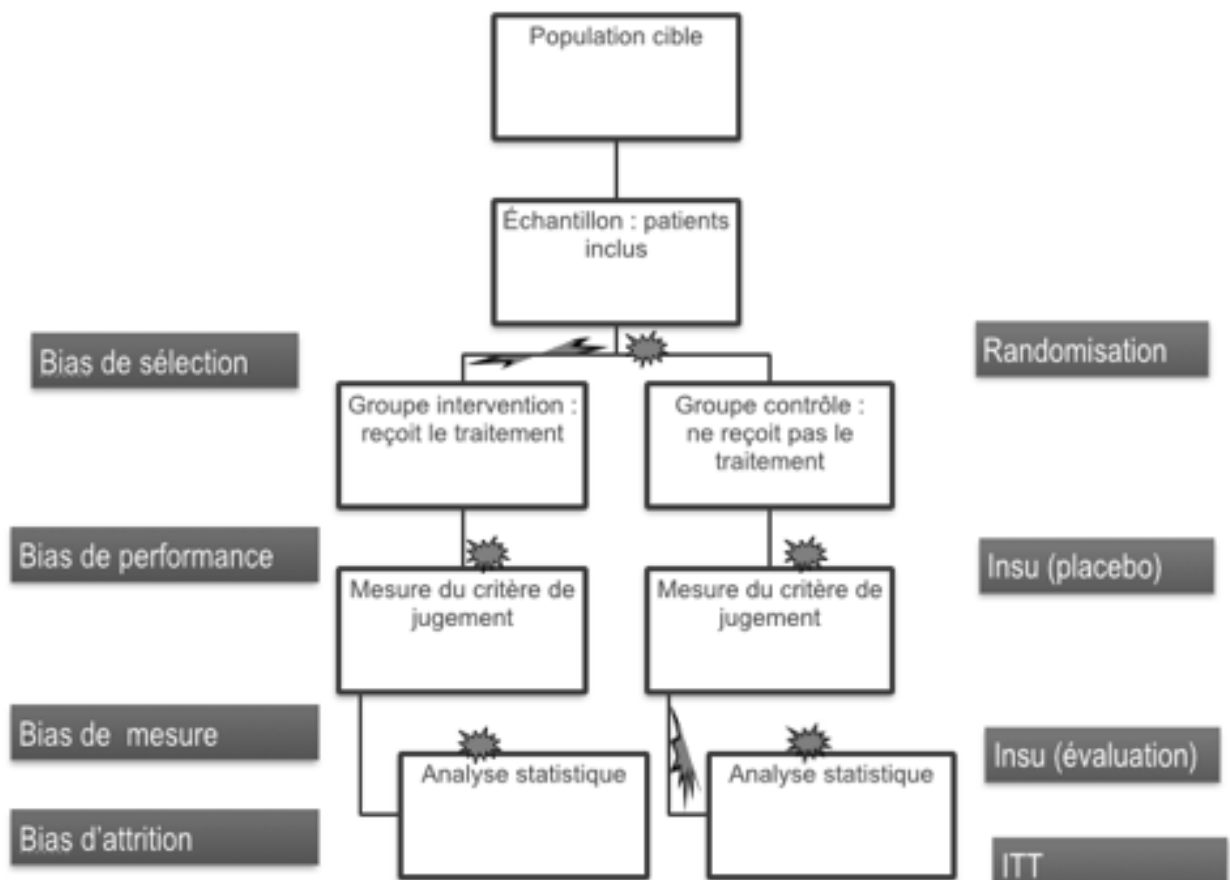
L'ITT consiste à analyser au long de l'étude tous les patients randomisés dans leur bras de randomisation/leur groupe initial, attribué au hasard. Ainsi, un patient faisant parti du groupe recevant le traitement expérimental, mais ayant arrêté de prendre son traitement ne doit PAS être considéré comme un patient du groupe placebo, mais doit être toujours analysé comme un patient recevant le traitement expérimental. De plus, les patients les plus compliants ont le meilleur pronostic ! Les patients non observants ou ne pouvant pas recevoir leur traitement ont un moins bon pronostic que les autres, même lorsque le traitement reçu est un placebo. En résumé, l'exclusion des patients qui arrêtent le traitement est un biais d'attrition.





S'il peut paraître illogique de garder dans un groupe un malade qui n'a pas reçu le traitement de ce groupe, c'est pourtant la seule façon de conserver la comparabilité initiale des groupes. De plus, si malgré tout, les résultats de l'analyse en ITT restent favorables au traitement, on sera d'autant plus convaincu de l'efficacité de ce traitement. Il ne faut jamais exclure un patient de l'analyse.

Schéma Bilan :



b) Validité externe

La validité externe permet de juger de l'applicabilité des résultats à la population générale, de la possibilité de généraliser des résultats (à un autre service hospitalier, une autre région, un autre pays). Elle dépend :

- Des critères d'éligibilité (critères d'inclusion et d'exclusion à l'étude, *exemple : si l'on prend seulement les patients les moins graves, le traitement ne sera généralisable qu'à ce type de patients et non à tous les malades*)
- Du mode de recrutement (désigne où sont recrutés les patients, *par exemple par un médecin en cabinet de ville, par des registres d'état civil*)
- Des centres (nombre et type)
- De la prise en charge, de l'adhérence au traitement
- Des co-interventions (autorisées? Identiques dans les 2 groupes?)

c) Nombre de sujets nécessaires

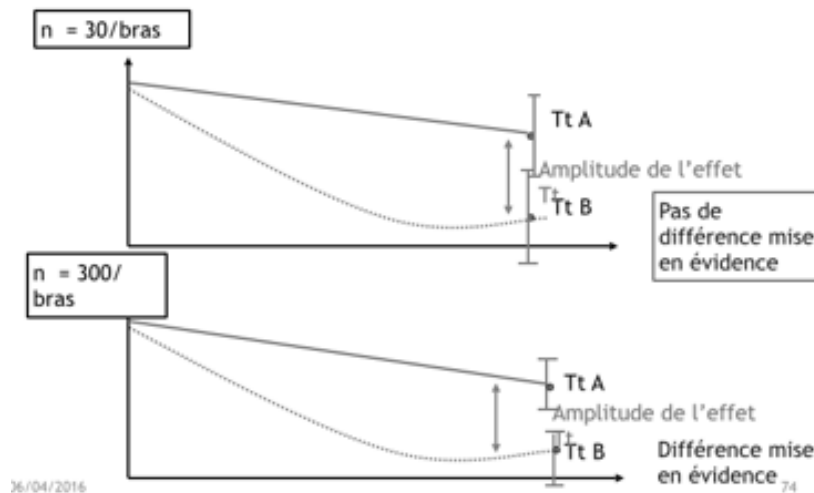
Ce n'est ni un critère de validité interne, ni un critère de validité externe. Le nombre de sujets nécessaires à l'étude permet d'obtenir une validité statistiques. Il dépend de plusieurs paramètres :

- Les risques consentis. Ils comprennent le risque de 1ère espèce ou risque d'erreur α . C'est le risque de conclure à une supériorité de A sur B alors qu'elle n'existe pas. On fixe en général ce risque à 5% ($p < 0,05$). En résumé, si l'essai est concluant avec A supérieur à B, j'ai 5% de chance de me tromper. Les risques consentis comprennent aussi le risque de 2^{ème} espèce ou risque β . C'est le risque de conclure (à tort) à l'absence de supériorité de A par rapport à B, alors que cette supériorité existe. Ce risque varie souvent entre 5 et 20%. Il permet de calculer la puissance de l'étude = $1 - \beta$. Par exemple, si $\beta = 20\%$, alors la puissance de l'étude est de $1 - \beta = 80\%$. En résumé, si l'essai n'est pas concluant, j'ai 20% de chance de me tromper en disant que A n'est pas supérieur à B.
- L'effet du produit de référence (B) ou du placebo. Cet effet est antérieurement connu via la littérature et des expériences cliniques.
- La différence escomptée ou attendue, entre l'effet de A (traitement expérimental) et l'effet de B ou du placebo. *Par exemple : Un traitement antihypertenseur de référence (B) fait baisser la pression artérielle diastolique de 10 mm de Hg. De combien faut-il qu'un traitement A fasse baisser la pression artérielle diastolique pour qu'il soit plus utile que B : de 15, 20, 25 mm de Hg ?*

Plus le nombre de sujets inclus est important, plus la différence d'efficacité entre A et B est facile à mettre en évidence. Il faut avoir suffisamment de sujets, autrement dit suffisamment de puissance. Plus la taille de l'effet est modéré, plus il faut de sujets pour détecter cette différence.

Exemple de l'étude Sophia Asthme : les appels ont pour effet d'améliorer un peu la qualité de vie des patients. Cependant, si on prend un nombre de sujets $n=30/\text{groupe}$, soit ce résultat ne sera pas visible car l'effet est trop faible, soit ce résultat sera visible mais ne nous permettra pas de conclure à une différence réelle ou une différence imputable aux seules fluctuations d'échantillonnage. A contrario, si $n=300/\text{groupe}$, la différence entre le groupe contrôle et le groupe expérimental pourra être mise en évidence.

Nombre de sujets nécessaires



QCM pour s'entraîner :

On souhaite comparer une trithérapie à une quadrithérapie dans le traitement de l'infection à VIH. On réalise cet essai avec au total 300 sujets. On trouve 50% de succès sous quadrithérapie, et 30% sous trithérapie. Lorsqu'on fait le test, on rejette l'hypothèse nulle avec un degré de signification inférieur à 10⁻⁶. Parmi les propositions suivantes, laquelle est vraie ?

- a) On ne peut pas conclure au bénéfice de la quadrithérapie, car on ne sait pas comment les groupes ont été constitués.
- b) La quadrithérapie est plus efficace que la trithérapie (au risque de 1%)
- c) On apprend que l'essai a été réalisé en double aveugle. On peut alors conclure au bénéfice de la quadrithérapie (au risque de 1%)
- d) Quelles que soient les conditions de réalisation de l'essai, on ne peut pas conclure au bénéfice de la quadrithérapie.
- e) On apprend que l'essai a été réalisé avec tirage au sort lors de l'attribution du groupe de traitement. On ne peut cependant pas conclure au bénéfice de la quadrithérapie (au risque 1%).

Réponse : a).

On rappelle que rejeter l'hypothèse nulle signifie que l'étude est concluante d'un point de vue statistique.

Le problème dans cet énoncé est que l'on ignore comment ont été constitués les groupes. Or, il est nécessaire que les groupes soient randomisés pour conclure de l'efficacité clinique du traitement (on rejette la proposition c) qui répartie les groupes en double-aveugle mais sans randomiser). L'étude a donc un biais de sélection, qui engendre une mauvaise validité interne : on ne peut pas faire confiance aux résultats de l'étude.

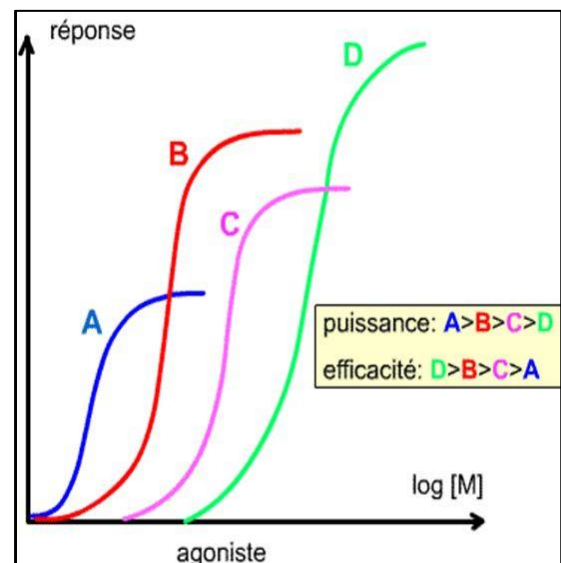
Attention aux pièges : une différence statistiquement significative ne signifie pas que le traitement est efficace (+++). Il ne faut pas oublier de prendre en compte les biais. Pour faire la preuve de l'efficacité (ou causalité), il faut nécessairement :

- Une différence d'amélioration statistiquement significative entre le groupe traité et le groupe contrôle
- Que l'essai clinique soit bien conduit (sans biais), avec une bonne validité interne.

III. Phases d'études cliniques

La phase I détermine les conditions de tolérance humaine. Elle implique donc d'étudier la posologie entraînant les premiers effets indésirables et la pharmacocinétique. Elle se réalise sur un petit nombre de personnes (10 à 20), souvent des volontaires sains, parfois d'emblée des malades si le médicament est toxique (*ex : chimiothérapie*).

La phase II détermine l'efficacité pharmacologique. Elle étudie les conditions d'efficacité, les modalités thérapeutiques (posologie efficace, mode d'administration), les courbes puissance-efficacité. Elle se réalise sur un petit nombre de personnes (50 à 100), qui sont des volontaires malades.



La phase III est la phase d'essai clinique randomisé (gold standard), qui détermine l'efficacité thérapeutique. Elle se réalise sur de grands effectifs (>1000, selon le nombre de sujets nécessaires pour démontrer un effet) qui sont des volontaires malades.

La phase IV a lieu après la mise sur le marché du médicament. Elle détermine l'efficacité et la tolérance dans les conditions usuelles de prescription, ainsi que les effets secondaires rares ou tardifs. Elle se réalise sur un très grand nombre de patient (>1million) puisqu'il s'agit de la population cible totale et non plus un échantillon.

Ce qu'il faut retenir :

- La nécessité du groupe contrôle pour expliquer la réponse au traitement
 - Les biais : sélection, mesure/performance et attrition
 - Eviction des biais par la randomisation, insu et ITT
- ➔ Il faut donc être capable de reconnaître quand l'étude est biaisée, identifier le biais, et savoir faire la différence entre une étude significative d'un point de vue statistique (risque α , risque β , nombre de sujets) et significative d'un point de vue clinique (randomisation, insu, ITT).



Voilà, c'était le dernier cours de l'année ! Bon courage dans vos révisions !