

Evaluation critique des résultats des essais cliniques

Principes mis en oeuvre dans Intérêt-Thérapeutique.org

Michel Cucherat

24 novembre 2015

Table des matières

1	Introduction	5
2	Principe général de l'évaluation des résultats des essais cliniques	7
2.1	Les éléments de la force de conviction d'un résultat	7
2.2	le spin des conclusions	8
2.3	Les essais concernés	9
2.4	Logique générale de l'évaluation des nouvelles propositions thérapeutiques	9
2.4.1	Nécessité d'avoir des preuves positives	9
2.4.2	Respect de la démarche hypothético-déductive	10
2.5	TODO	10
3	Qu'est ce qu'un résultat	11
3.1	Éléments constitutifs d'un résultat	11
3.2	Résultat positif ou négatif	12
3.3	Sens du résultat	12
3.4	Au total	13
4	Classification des résultats en faveur de l'efficacité	15
4.1	Notes	17
4.1.1	Pourquoi une prise de risque minimale dans l'acceptation (et la recommandation) de nouveau traitement	17
5	Les éléments de la force de conviction d'un résultat d'essai thérapeutique	19
5.1	Pourquoi des points positifs n'estompent pas d'autres réserves méthodologique	19
8	Résultat faussement positif dû au hasard	31
9		33
9.1	Généralités	33
9.1.1	Le test statistique et la "p value"	34
9.1.2	L'inflation du risque alpha	35
9.1.3	démarche hypothetico déductive et cohérence externe du résultat	36
9.2	Situation où il est possible d'exclure que le résultat soit un faux positif dû au hasard	37
9.3	Situation où il n'est pas possible d'exclure que le résultat soit un faux positif dû au hasard	38
9.4	Résultat montrant un surcroît d'effets indésirables, risque de faux positif	38
8	Résultat faussement positif dû au hasard	31

9		33
9.1	Généralités	33
9.1.1	Le test statistique et la "p value"	34
9.1.2	L'inflation du risque alpha	35
9.1.3	démarche hypothetico déductive et cohérence externe du résultat	36
9.2	Situation où il est possible d'exclure que le résultat soit un faux positif dû au hasard	37
9.3	Situation où il n'est pas possible d'exclure que le résultat soit un faux positif dû au hasard	38
9.4	Résultat montrant un surcroît d'effets indésirables, risque de faux positif	38
10	Résultat faussement positif dû à un biais	41
10.1	Situations où il est possible d'exclure que le résultat soit un faux positif dû à un biais	43
10.2	Situations où il n'est pas possible d'exclure que le résultat soit un faux positif dû à un biais	43
11	Pertinence clinique	45
11.1	Situation assurant la pertinence clinique du résultat	46
11.2	Situation de manque de pertinence clinique	47
12	Bénéfice net	49
12.1	Notion de bénéfice clinique net	49
12.1.1	Bénéfice clinique net quantitatif	50
12.1.2	Bénéfice net qualitatif	51
12.2	Le bénéfice n'est pas susceptible d'être contre balancé quantitativement par des effets indésirables de même gravité	51
12.3	Il n'est pas possible d'exclure que ce bénéfice puisse être contre-balancé quantitativement par des effets indésirables de même gravité rendant le bénéfice net nul ou défavorable	52
12.4	Bénéfice net défavorable, suggéré ou démontré ($DR > 0$, NS ou $p < 0.05$)	53
12.5	Bénéfice net défavorable qualitativement	54
12.6	Ce résultat a été obtenu sur un critère qui est déjà un critère de bénéfice net	55
13	Démarche hypothetico déductive	57
13.1	Situations où la nature hypothetico déductive du résultat est acceptable	57
13.2	Situations où la nature hypothetico déductive du résultat n'est pas acceptable	57
14	Cohérence externe	59
14.1	Situation assurant la cohérence externe du résultat	60
14.2	Situation d'absence de cohérence externe du résultat	60
14.2.1	L'appréciation de la cohérence externe impossible	60
14.2.2	Remise en cause de la cohérence externe du résultat	60
14.3	références	61
15	Les différents types de résultats	63
15.1	Résultat démontré, cliniquement pertinent, bénéfice net assuré	64
15.2	Bénéfice cliniquement pertinent démontré mais la balance bénéfice risque n'est pas assurée	65
15.3	Bénéfice clinique net défavorable (alerte de safety, $DR > 0$ NS ou $p < 0.05$)	65

15.4	Résultat démontré, non cliniquement pertinent	66
15.5	Résultat démontré mais absence de cohérence externe	67
15.6	Résultat seulement suggéré (cliniquement ou non cliniquement pertinent)	67
15.7	Résultat seulement suggéré mais possibilité d'upgrading	68
15.8	Résultat non significatif	68
15.9	Résultat non significatif pour un critère de sécurité	69
15.10	Infériorité ($p < 0.05$), démontré ou suggéré (sur un critère d'efficacité cliniquement pertinent)	69
15.11	Résultat montrant un défaut de sécurité, démontré ou simplement suggéré, sans bénéfice en regard	70
15.12	Surcroît d'effet indésirable acceptable compte tenu du bénéfice apporté (bénéfice net favorable)	70
15.13	Effet délétère remettant en cause un bénéfice obtenu par ailleurs	71
15.14	Effet délétère sous estimé par biais	71
15.15	Résultat ne montrant pas de défaut de tolérance pouvant contrebalancer le bénéfice	72
15.16	Non infériorité cliniquement pertinente démontré	72
15.17	Non infériorité démontré mais non cliniquement pertinente	72
15.18	Non infériorité suggéré mais non formellement démontré	73
15.19	Données non matures	74
16	Intérêt thérapeutique	75
16.1	L'essai apporte un ou plusieurs résultats en faveur de l'intérêt thérapeutique du traitement évalué	75
16.2	L'essai n'apporte pas de résultat en faveur de l'intérêt thérapeutique du traitement évalué	76
17	Cohérence externe d'un résultat	77
17.1	Les différents cas de figure	77
17.1.1	Le résultat considéré est en cohérence avec les résultats des autres études similaires	77
17.1.2	L'appréciation de la cohérence externe impossible	77
17.1.3	Remise en cause de la cohérence externe du résultat	77
17.2	Exemple : infarctus et dabigatran	78
18	But évaluation	79
18.1	Logique générale de l'évaluation	79
18.2	Prise en considération du risque de faux négatifs	81
19	Conséquences des décisions basées sur des preuves imparfaites	83
19.1	introduction	83
19.2	Place de ces preuves imparfaites dans le décision	84
19.3	calcul de la probabilité qu'un traitement soit réellement efficace après un essais "concluant"	84
19.4	Conséquences de l'utilisation de preuves imparfaites pour recommander l'utilisation d'un traitement	86
20	Chapitre 1	89
20.1	Bibliographe	89
20.2	Introduction	89

20.2.1	section 1	90
20.2.2	section 2	92
20.3	Précision [ref-section-precision]	92
20.4	Renvois et références	93
20.5	fichier d'aide (section help)	93

Chapitre 1

Introduction

Le site Intérêt-Thérapeutique.org propose une évaluation standardisée des résultats des essais thérapeutiques des nouveaux traitements. Les résultats sont classés en fonction de leur validité méthodologique et de leur pertinence clinique.

La finalité est d'identifier les résultats qui apportent des éléments de preuve de l'intérêt thérapeutique des nouveaux traitements.

Ce document décrit les principes mis en œuvre pour réaliser cette analyse et les critères utilisés pour déterminer si un résultat constitue ou non une preuve formelle de l'intérêt thérapeutique du nouveau traitement évalué dans cet essai.

L'approche utilisée consiste à classer les résultats produits par l'essai en fonction de leur validité méthodologique et de leur pertinence clinique.

Chapitre 2

Principe général de l'évaluation des résultats des essais cliniques

L'évaluation de l'intérêt thérapeutique d'un traitement (médicamenteux ou autre) s'effectue à l'aide des essais thérapeutiques contrôlés, randomisés, en double aveugle, correctement conçus et réalisés.

En fonction des résultats obtenus et de la robustesse de sa méthodologie, un essai thérapeutique peut ainsi apporter ou non des éléments de preuve de l'intérêt thérapeutique d'un traitement.

Pour apporter une preuve de l'intérêt thérapeutique d'un traitement, un essai doit fournir au moins un résultat qui démontre parfaitement un bénéfice cliniquement pertinent débouchant sur un bénéfice net favorable après prise en considération des effets indésirables.

Cependant, tous les résultats produits par un essai thérapeutiques, indépendamment de leur teneur, ne se valent pas et certains ne sont pas suffisamment fiables ou pertinents.

Note :

Le fait qu'un résultat puisse être, apparemment, en faveur de l'efficacité du traitement n'est pas suffisant pour un faire une preuve de l'intérêt thérapeutique du traitement. Encore faut-il, qu'il soit solide méthodologiquement, qu'il puisse permettre, de conclure statistiquement sans risque, qu'il représente une plus-value médicale et qu'il ne soit pas susceptible d'être complètement contre-balancer par des effets indésirables

L'objectif du site web Intérêt-Thérapeutique.org est de réaliser une analyse systématique de tous les résultats produits par une étude clinique afin de déterminer quels sont ceux qui sont suffisamment fiables et pertinents pour prouver l'intérêt thérapeutique du traitement évalué.

2.1 Les éléments de la force de conviction d'un résultat

La force de conviction d'un résultat d'essai thérapeutique dépend de plusieurs éléments :

- La valeur de ce résultat : le résultat est-il en faveur d'un effet bénéfique ou non (en fonction de la valeur du risque ratio, du hazard ratio, etc.),
- La fiabilité de l'étude qui a produit ce résultat, son aptitude à mesurer correctement l'effet réel du traitement, sans biais (validité interne),

- La solidité statistique du résultat, c'est à dire la possibilité d'utiliser le résultat pour conclure à l'intérêt du traitement sans prendre un risque inconsidéré de faire cette conclusion à tort,
- La pertinence clinique du résultat, c'est à dire la plus-value médicale apportée par le traitement (en quelque sorte l'importance du service médical rendu par le traitement),
- La balance-bénéfice risque, c'est-à-dire, compte tenu de l'importance et de la nature des effets indésirables du traitement, existe-t-il un **bénéfice net** substantiel pour le patient (le bénéfice recherché, apporté par le traitement n'est pas contrebalancé en totalité par un « contre-bénéfice » produit par les effets indésirables).

Ainsi en fonction de la méthode de l'essai, de la solidité statistique du résultat, de la nature du critère de jugement, de l'importance du bénéfice apportée, de la nature et de la fréquence des événements indésirables, des patients inclus dans l'essai, un résultat donné peut constituer ou non une preuve de l' **intérêt thérapeutique du traitement** .

En pratique, la détermination de l'intérêt thérapeutique d'un traitement à partir du compte rendu d'un essai thérapeutique (article ou rapport d'essai) nécessite d'évaluer les résultats obtenus suivant toutes ces dimensions. Seuls les résultats qui ne soulèvent pas de réserves sur une quelconque de ces dimensions, seront en mesure de prouver l'intérêt thérapeutique du traitement. Les autres ne font que suggérer, au mieux, le bénéfice du traitement, mais n'en apportent pas la démonstration nécessaire à l'utilisation du traitement en pratique courante. Ces résultats doivent être alors confirmés dans un nouvel essai qui les démontrera de façon formelle ou qui les invalidera. En effet, l'expérience montre que beaucoup de résultats positifs, mais insuffisamment forts, n'ont pas été confirmés dans les essais ultérieurs mis en place pour les confirmer (cf. fiche concept : **Preuves empiriques de la fragilité des résultats non formellement démontrés**).

Ainsi, un essai ne donne pas toujours des résultats démontrant l'intérêt du traitement. Il peut avoir un design ou une méthodologie impropre à fournir des résultats suffisamment fiables ou suffisamment cliniquement pertinents. Ainsi, même s'il est, apparemment, en faveur d'un bénéfice, un résultat peut être invalidé par son manque de fiabilité ou de pertinence.

2.2 le spin des conclusions

La nécessité de faire une évaluation critique systématique des article rapportant un essai clinique est rendu incontournable par l'existence de pratique d'enjolivement des conclusion, désigné par « **spin des conclusions** ». Ainsi, les résultats mis en avant dans les conclusions de ces articles et utilisés pour conclure positivement ne sont pas forcément des résultats ayant toutes les qualités requises pour permettre de conclure à l'intérêt thérapeutique du traitement.

Ainsi, certains papiers, même publiés récemment dans des revues prestigieuses, proposent des conclusions non supportées par les résultats et/ou abusivement valorisées pour soutenir un effet favorable du traitement étudié.

De ce fait, il est indispensable d'effectuer une évaluation critique des essais et de leurs résultats avant de suivre éventuellement les conclusions des auteurs et d'adopter les nouveaux traitements.

2.3 Les essais concernés

L'approche utilisée ici est centrée sur la recherche des nouveaux traitements représentant un réel progrès thérapeutique, c'est-à-dire apportant une amélioration de la prise en charge des patients par rapport au meilleur des traitements précédemment disponibles.

Il existe aussi des essais, moins ambitieux, qui cherchent seulement à montrer qu'un nouveau traitement peut représenter une ressource thérapeutique efficace sans vraiment chercher à détrôner le traitement de référence (par exemple à l'aide d'essai de non-infériorité, mais c'est aussi le cas d'essais de supériorité).

Ces essais ont pour objectif d'introduire sur le marché un nouveau traitement, mais ne se préoccupe pas de faire progresser la prise en charge de la maladie.

Dans cette logique, les nouveaux traitements évalués par ces essais ne peuvent pas être utilisés en première intention (car on ne sait pas s'ils représentent la meilleure option à proposer aux patients) et ils constituent seulement des options alternatives en cas de réelles impossibilités de donner à certains patients le traitement de référence.

Les résultats de ces essais seront donc systématiquement considérés comme ne montrant pas l'intérêt thérapeutique de ces nouveaux traitements, le terme intérêt thérapeutique étant compris dans ce site comme amélioration de l'existant (un nouveau traitement n'a d'intérêt que s'il permet de faire mieux que les traitements précédents).

Cependant, ces traitements peuvent avoir, à un niveau individuel, un certain intérêt pour le médecin, par exemple quand le traitement de référence a échoué ou n'est pas utilisable ou supporté. Ce sera par exemple le cas avec :

- Un essai de non-infériorité versus le traitement de référence avec un nouveau traitement qui n'a pas d'avantage par ailleurs et dont l'objectif est seulement de montrer que le nouveau traitement a une efficacité suffisante (cf. fiche concept [non infériorité](#))
- Un essai versus placebo alors qu'il existe déjà un traitement supérieur au placebo. Dans ce cas, le résultat n'a pas de pertinence clinique car la question médicale qui se pose n'est plus celle de l'existence d'un traitement efficace, mais bien celle de l'existence d'un traitement plus efficace de l'actuel.

2.4 Logique générale de l'évaluation des nouvelles propositions thérapeutiques

2.4.1 Nécessité d'avoir des preuves positives

Un autre élément de la logique utilisée est qu'en absence de démonstration formelle de l'efficacité, la situation est alors considérée comme identique à celle où l'efficacité serait inexistante.

En effet, avant d'engager la communauté à proposer un nouveau traitement à de nombreux patients, il est indispensable d'avoir la preuve de l'efficacité du traitement. De plus, les traitements ont un cout qui engage la collectivité et même, dans certains systèmes de santé, directement les patients eux-mêmes.

Bien sûr l'absence de preuve n'est pas la preuve de l'absence, mais ce questionnement ne concerne que le chercheur qui s'interroge sur l'opportunité de mettre en place un nouvel essai s'il pense que les résultats négatifs obtenus ne remettent pas en cause l'hypothèse théorique de l'intérêt de ce traitement. Il y a donc une lecture critique du chercheur qui est différente de celle du clinicien ou du décideur.

2.4.2 Respect de la démarche hypothético-déductive

Un résultat ne peut être considéré comme scientifiquement établi que s'il correspond à l'hypothèse de l'essai et respecte ainsi la démarche hypothético-déductive. Autrement, il s'agit seulement d'un résultat post-hoc qui n'a qu'une valeur exploratoire.

En effet, l'approche utilisée doit respecter strictement la démarche scientifique hypothético-déductive et doit écarter toute tentation d'avoir une lecture inductive des résultats (cf. fiche concept [démarche inductive et déductive](#)).

Souvent, la tentation est forte de généraliser les résultats observés dans une étude (démarche inductive), mais cette généralisation, au rang de loi, d'une simple observation, correspond à une [faute de logique \(logical fallacy\)](#) désignée par « Post hoc, ergo propter hoc » (« après cela donc à cause de cela »).

Ainsi toute exploitation de résultats ne découlant pas directement de l'hypothèse spécifique testée par l'essai (résultats [post hoc](#)) n'a aucune valeur de preuve et ne représente pas un fait scientifique prouvé. Tout au plus il permet de générer une nouvelle hypothèse à vérifier de manière déductive dans un nouvel essai thérapeutique.

Cette exploitation inductive des résultats, bien que néanmoins très largement pratiquée, n'est pas en mesure de fournir des éléments de preuves de l'intérêt thérapeutique des traitements. Cela conduit à récuser par exemple, pour la question de l'efficacité, les résultats provenant des critères qui n'étaient pas des critères principaux ou provenant de sous-groupes pour lesquels aucune hypothèse spécifique n'avait été formulée.

2.5 TODO

Il convient de distinguer plusieurs situations :

- l'évaluation d'un résultat "positif", montrant le cas échéant ¹ un bénéfice du traitement
- l'évaluation d'un résultat "positif", en faveur d'un effet indésirable du traitement
- l'évaluation d'un résultat "négatif", montrant le cas échéant l'absence de bénéfice du traitement
- l'évaluation d'un résultat montrant le cas échéant l'absence d'effet indésirable
- l'évaluation d'un résultat de non infériorité

Les principes mis en jeu ne sont pas les mêmes en fonction de ces différentes situations :

- Pour les résultats pouvant déboucher le cas échéant sur une conclusion de bénéfice ou de non infériorité on cherchera à exclure formellement que le résultat puisse être faussement positif. Seuls les résultats parfaitement solide (méthodologiquement et statistiquement) seront acceptés pour s'autoriser à faire ce type de conclusion (si le résultat s'avère aussi cliniquement pertinent et robuste vis à vis des effets indésirables du traitement).
- pour les résultats pouvant déboucher le cas échéant sur une conclusion effet indésirable l'analyse s'effectuera suivant le principe de précaution. En raison de l'enjeux, un simple signal sera pris en considération même si l'effet indésirable n'est pas formellement démontré. Le traitement sera écarté si aucun bénéfice pouvant rendre acceptable ce risque n'a été parfaitement démontré.
- pour les résultats pouvant déboucher le cas échéant sur une conclusion d'absence d'effet ...
- pour les résultats pouvant déboucher le cas échéant sur une conclusion d'absence d'effet indésirables : raisonnement sur le bénéfice clinique net

1. s'il s'avère solide et cliniquement pertinent

Chapitre 3

Qu'est ce qu'un résultat

Dans ce site et dans ce document, le terme « résultat » est utilisé pour désigner la mesure de l'effet du traitement évalué par rapport au traitement contrôle sur un critère de jugement donné.

L'effet traitement est mesuré, suivant le cas, par un risque ratio ou un odds ratio avec les critères de jugement binaire; par un hazard ratio avec les critères de temps jusqu'à événement (comme la survie par exemple) et par une différence de moyenne (standardisée ou non) avec les critères continus.

Cette mesure de la taille de l'effet est donnée encadré par son intervalle de confiance à 95% qui représente l'incertitude avec laquelle on connaît cet effet. Un résultat peut être statistiquement significatif ou non.

Exemple : Exemples de résultats tels qu'ils sont rapportés dans les abstracts

The rate of the primary outcome was 1.27% per year in the N group, as compared with 1.60% per year in the C group (hazard ratio, 0.79; 95% confidence interval [CI], 0.66 to 0.95; $P < 0.001$ for noninferiority; $P = 0.01$ for superiority).

At 12 months, the primary end point—a composite of death from vascular causes, myocardial infarction, or stroke—had occurred in 9.8% of patients receiving N as compared with 11.7% of those receiving C (hazard ratio, 0.84; 95% confidence interval [CI], 0.77 to 0.92; $P < 0.001$).

There was no significant difference in the duration of overall survival between the N group and the placebo group (median, 15.7 and 16.1 months, respectively; hazard ratio for death in the N group, 1.13).

3.1 Eléments constitutifs d'un résultat

Un résultat est constitué par 5 éléments :

- Une comparaison de 2 traitements : le traitement évalué comparé à un traitement contrôle (traitement actif de référence ou placebo)
- Un critère de jugement (mesuré à un temps donné)
- Une **taille d'effet** (risque relatif, hazard ratio, odds ratio, différence de moyennes, etc.)

- Un intervalle de confiance, à 95% le plus souvent, donnant la précision avec laquelle on connaît la taille de l'effet
- Une « p value » documentant la réalité statistique du résultat

Taille d'effet [Les indices d'efficacité pour les critères binaires](#) [Les indices d'efficacité pour les critères continus](#) [Les indices d'efficacité pour les critères de temps jusqu'à événement](#)

3.2 Résultat positif ou négatif

Un résultat est dit positif lorsqu'il permet de conclure à l'existence d'un effet du traitement. Il est dit négatif dans le cas contraire.

Pour la question de l'efficacité, on recherche un résultat positif, permettant de conclure à l'efficacité (versus placebo) ou à la supériorité (versus comparateur actif) du traitement étudié. Un résultat négatif empêchera de conclure à l'efficacité du traitement.

Pour la sécurité, un résultat positif montrant un surcroît d'effet délétère avec le traitement étudié permettra de conclure à un défaut de sécurité du traitement étudié. Par contre, un résultat négatif, ne "montrant" pas d'effet délétère du traitement va plutôt dans le sens espéré (absence d'effet indésirable) mais ne permettra cependant pas de faire cette conclusion (cf. chapitre ???).

3.3 Sens du résultat

Il existe deux grands types de résultats en terme de sens du résultat : les résultats allant dans le sens d'un intérêt du traitement évalué et ceux allant dans le sens d'un effet délétère ou d'une infériorité du traitement évalué.

La logique et les règles d'évaluation changent complètement entre ces 2 types de résultats car les décisions correspondantes sont aussi complètement opposées.

Un résultat *allant dans le sens d'un intérêt du traitement* permet d'envisager une décision de recommander le traitement ; contrairement à un résultat montrant un effet délétère ou une infériorité, qui lui, fait plutôt envisager l'arrêt du développement du traitement ou une recommandation de non utilisation (si le traitement est disponible).

Par exemple, dans un essai d'un stent carotidien un résultat montrant une réduction du risque de récurrence d'AVC sera un résultat allant dans le sens d'un intérêt du traitement évalué. Il pourra être à l'origine d'une décision de conclure à l'intérêt du stent carotidien. Dans le même essai, une augmentation des infarctus du myocarde sera alors un résultat allant dans le sens d'un effet délétère. Ce résultat nous fera nous interroger sur l'innocuité du stent et nous conduira éventuellement à une décision de ne pas utiliser ce stent carotidien.

Avec les résultats allant dans le sens d'un intérêt du traitement évalué, l'évaluation s'effectue avec la crainte d'accepter à tort cet intérêt. L'accent sera donc mis sur le risque que le résultat soit faussement positif et la preuve formelle de l'intérêt du traitement est absolument recherchée. Toutes limites ou réserves empêchant de conclure à la démonstration du résultat conduiront donc à le récuser et à le faire considérer comme seulement suggéré et non formellement démontré.

Avec un résultat *montrant une infériorité ou un effet délétère*, on ne cherche plus l'obtention d'une démonstration formelle. En vertu du principe de précaution, le résultat sera accepté même s'il ne présente pas toutes les garanties méthodologiques et statistiques. En effet, ce type de résultat ne risque pas de conduire à l'utilisation à tort du traitement puisqu'il est, au contraire, de nature à conduire à l'arrêt du développement (et donc à la non utilisation du traitement). De ce fait, les exigences de démonstration que le résultat ne soit pas un faux positif (dû au hasard ou à un biais) tombent. La seule difficulté que

présente cette logique d'interprétation est celle liée à l'excès de précaution qui conduirait la communauté à se priver d'un traitement pourtant efficace devant un résultat suggérant à tort un effet délétère ou une infériorité. La prise en considération du bénéfice clinique net permet de limiter ce risque (cf. [chapitre sur le bénéfice net](#)).

3.4 Au total

Au total, la logique mise en oeuvre dans le domaine de l'évaluation des nouvelles proposition thérapeutique, conduit aux 4 règles globales suivantes.

Un résultat "positif", fiable, cliniquement pertinent, correspondant à l'objectif (hypothèse) de l'essai, débouchant sur un bénéfice clinique net, fournit une preuve de l'intérêt thérapeutique.

Un résultat "négatif" (sur un critère d'efficacité) ne donne aucun argument pour utiliser le traitement, ce qui revient à considérer que le traitement n'a pas d'intérêt jusqu'à preuve du contraire.

Un résultat en faveur de l'infériorité du traitement ou d'un effet indésirable (non contre balancé par un bénéfice et conduisant donc in fine à un bénéfice net défavorable quantitativement ou qualitativement), même s'il n'est pas formellement démontré (absence d'hypothèse, réserves méthodologiques possibles), conduit à ne pas utiliser le traitement par précaution.

Un résultat "négatif" sur un critère de sécurité ne dédouane pas le nouveaux traitement tant qu'il n'y a pas de démonstration d'un bénéfice clinique net favorable.

Chapitre 4

Classification des résultats en faveur de l'efficacité

Afin de déterminer leur aptitude à prouver l'effet du traitement évalué, les résultats en faveur de l'efficacité du traitement ("résultats positifs") sont classés en fonction de leur validité en termes de fausse positivité (validité interne), de pertinence clinique et du bénéfice net qu'ils représentent. Ces dimensions sont détaillées dans le chapitre sur [les éléments de la validité d'un résultat](#) .

TABLE 4.1 – Classification des résultats (en faveur de l'efficacité)

Validité interne	Pertinence clinique	Bénéfice net	En faveur de l'intérêt thérapeutique
résultat formellement démontré	résultat cliniquement pertinent	bénéfice net favorable démontré	oui
idem	idem	absence de démonstration d'un bénéfice net favorable ou bénéfice net non favorable (suggéré ou démontré)	non
idem	résultat insuffisamment cliniquement pertinent	sans objet	non
résultat seulement suggéré	sans objet	sans objet	non
résultat non significatif			non

En termes de **fausse positivité dû au hasard ou aux biais** (validité interne), les résultats se répartissent en deux catégories :

- Les résultats **démontrés** qui sont issus d'une méthodologie et d'une démarche statistique garantissant au mieux que le résultat n'est pas un faux positif dû à un biais ou au hasard. En quelques mots, cela sous entend que le résultat est à l'abri d'un biais de sélection en raison d'une randomisation imprévisible, d'un biais de suivi et de mesure en raison d'une réalisation en double insu et d'un biais d'attrition en raison d'une analyse en ITT avec remplacement des données manquantes (pour l'essai de supériorité). Au niveau statistique cela signifie que le résultat est statistiquement significatif (au moins au seuil de 5% bilatéral) et qu'il n'est pas issu d'une démarche

de répétition des comparaisons et des tests engendrant multiplicité et inflation du risque alpha (multiplicité des critères de jugement, analyse de sous groupe, analyse intermédiaire non protégées, etc.).

- Les résultats seulement **suggérés** car il n'est pas possible d'exclure qu'ils puissent être dus à un biais ou au hasard en raison d'un défaut méthodologique ou de démarche statistique. Les résultats post hoc sont assimilés à cette catégorie.

Note :

les résultats non statistiquement significatifs ne permettent pas de conclure. Ils ne donnent aucun argument en faveur du bénéfice du traitement. Du fait de l'absence de signification statistique, il est complètement déraisonnable d'exclure que le résultat ne soit pas entièrement dû au hasard.

En termes de **pertinence clinique**, les résultats se classent en résultats cliniquement pertinents et non cliniquement pertinents (manque de pertinence clinique du comparateur, du critère de jugement, de la taille de l'effet, de la population étudiée, etc. . .).

En termes de **bénéfice clinique net** (balance bénéfice risque), il existe aussi deux situations :

- Les événements indésirables ne sont pas en mesure de contrebalancer le bénéfice (de manière qualitative ou quantitative, cf. chapitre [bénéfice net](#)). Il y a démonstration que le bénéfice net est favorable.
- Les événements indésirables peuvent contrebalancer complètement le bénéfice (soit qualitativement, soit quantitativement) : il y a un risque d'événements indésirables plus grave que le bénéfice apporté ou la fréquence des événements indésirables est du même ordre de grandeur que le bénéfice. Il y a absence de démonstration que le bénéfice net est favorable ou mise en évidence que ce bénéfice net n'est pas favorable (de manière démontré ou simplement suggéré)

Tous les croisements de ces 4 dimensions ne sont pas intéressants (cf. tableau introductif). Par exemple, avec un résultat non significatif il est inutile de se poser la question de la pertinence clinique, ou de la balance bénéfice risque.

Seul les **résultats démontrés, cliniquement pertinents et correspondant à un bénéfice clinique net favorable** fournissent la **preuve de l'intérêt thérapeutique** du traitement et peuvent être utilisés pour justifier une utilisation systématique du traitement en pratique courante.

Les résultats seulement suggérés ne représentent pas des preuves suffisantes de l'intérêt thérapeutique du traitement car le risque que ce résultat soit un résultat faussement positif dû à un biais ou au hasard est trop élevée pour qu'il soit possible de baser une pratique.

Les résultats insuffisamment cliniquement pertinents ne représentent pas non plus des preuves de l'intérêt thérapeutique ainsi que les résultats ne correspondant pas à un bénéfice clinique net favorable.

TABLE 4.2 –

Type de résultat en faveur de l'efficacité	Bénéfice net	en faveur de l'intérêt thérapeutique
Démonstration d'un bénéfice cliniquement pertinent	assuré a (bénéfice net favorable, $DR < 0$, significatif)	oui
idem	non garanti (bénéfice net non significatif)	non
idem	défavorable ($DR > 0$, significatif ou NS)	non
Démonstration d'un bénéfice non cliniquement pertinent	défavorable (le traitement n'ayant pas montré de bénéfice, quelque soit les risques, le bénéfice net est défavorable)	non
Suggestion du bénéfice (cliniquement pertinent ou non)	sans objet	non

a : Soit quantitativement comme le montre une différence des risques du bénéfice net < 0 et statistiquement significative soit parce qu'aucun effet indésirable capable de remettre en cause

4.1 Notes

Pour conclure à l'intérêt thérapeutique avec une prise de risque de se tromper minimale il faut une preuve incontestable (au delà de tout doute raisonnable) que le traitement apporte un réel bénéfice cliniquement pertinent débouchant sur un bénéfice clinique net.

La logique est d'avoir la certitude au delà de tout doute raisonnable que le bénéfice sera au rendez-vous. Un moindre doute a un quelconque niveau (faux positif, pertinence, bénéfice clinique net) rend la valeur de preuve d'un résultat caduque.

c'est au traitement d'apporter la preuve irréfutable (au delà de tout doute raisonnable) de son intérêt clinique et non pas au détracteur (utilisateurs) d'apporter la preuve du non intérêt. Ainsi toute insuffisance dans la solidité de la preuve invalide la preuve et laisse le traitement démuné.

C'est en ligne avec les principes de la méthode scientifique où une théorie est acceptée quand elle a résisté à toutes les tentatives de réfutation qu'elle a subie Popper, ouvrage d'épistémologie.

Pourquoi la charge de la preuve au traitement Pourquoi une prise de risque minimale dans l'acceptation (et la recommandation) de nouveau traitement

4.1.1 Pourquoi une prise de risque minimale dans l'acceptation (et la recommandation) de nouveau traitement

une prise de risque minimale dans l'acceptation (la recommandation ou l'utilisation) d'un nouveau traitement est nécessaire étant donné les enjeux de la décision.

- pour ne pas abuser de la confiance des patients, répondre à une exigence éthique
- pour ne pas bloquer à tort la poursuite de la recherche
- engagement de la collectivité, problème de cout
- à partir du moment où un traitement est adopté l'évaluation cesse et il devient impossible de récupérer une mauvaise décision
- des effets indésirables rare mais graves peuvent être détecter après la hase d'évaluation car ils sont rares. Si pas de bénéfice certain, dans ces cas la décision apparait a posteriori comme ayant été vraiment incongrue

Chapitre 5

Les éléments de la force de conviction d'un résultat d'essai thérapeutique

La force de conviction d'un résultat d'essai thérapeutique en faveur de l'efficacité du traitement évalué dépend de plusieurs éléments :

- Le risque de conclure à tort à l'intérêt du traitement à partir de ce résultat uniquement dû fait du hasard
- Le risque que ce résultat soit en fait un résultat faussement positif (faux positif) dû à un biais
- Sa pertinence clinique (c'est-à-dire la valeur ajoutée médicale qu'apporte le traitement)
- La balance bénéfice risque
- Sa cohérence externe, c'est à dire sa cohérence avec les résultats produits par les autres études similaires

Auxquels on peut adjoindre un 6ème, la nature hypothetico déductive (qui peut aussi être considérée comme une forme particulière de faux positif dû au hasard).

Les deux premiers éléments concernent la possibilité qu'un résultat positif puisse l'être à tort, alors qu'en réalité le traitement n'est pas efficace et que, donc, l'essai aurait dû conduire à un résultat négatif. Il existe deux causes qui peuvent produire facilement des résultats faussement positifs : les erreurs statistique et les biais. §

Chacun de ces éléments sont discuté dans les 4 chapitre suivants.

5.1 Pourquoi des points positifs n'estompent pas d'autres réserves méthodologique

Une seule réserve portée sur un des éléments de la validité d'un résultat conduit à son invalidation. L'existence de points satisfaisant sur les autres éléments de la validité ne permet pas de compenser la réserve effectuée. En effet, les éléments de la validité d'un résultat (contrôle du risque de faux positif dû au hasard et protection contre les biais) s'enchaînent en série pour former une chaîne et une seule défaillance sur l'un de ses maillons affaiblie la totalité de la chaîne.

Par exemple, un défaut de contrôle d'un biais particulier ne peut pas être rattrapé par un contrôle parfait des autres biais. Les autres moyens mis en oeuvres ne peuvent pas suppléer la possibilité de faux positif consécutive à ce défaut méthodologique. Par exemple, la randomisation ne compense pas l'absence de double aveugle et ne peut pas empêcher la survenu d'un résultat faux positif dû à un biais de mesure ou de suivi. Il ne

sert donc à rien de faire un scoring de « qualité » des essais étant donné que l'existence d'une seule et unique réserve remet en cause la fiabilité du résultat. Un risque de résultat faux positif existe même si la protection contre les biais est optimale. Ainsi pour un critère de jugement secondaire, un contrôle des biais optimal n'empêche pas l'inflation du risque alpha engendrée par l'exploration des critères secondaires.

Une méthodologie parfaite permettant d'exclure que le résultat soit un faux positif (dû au hasard ou à un biais) ne récupère pas un défaut de pertinence clinique du résultat et ne peut pas transformer un résultat sans intérêt pratique en un résultat cliniquement pertinent.

La forte pertinence clinique potentielle d'un résultat (comme par exemple une réduction de mortalité totale) ne compense pas un défaut de fiabilité du résultat (protection contre le risque de faux positif dû à un biais ou au hasard). Ainsi, même la mortalité totale est sujette aux fluctuations aléatoires d'échantillonnage ou au biais. La force du critère de jugement ne change en rien la problématique et n'autorise en rien un moindre contrôle de la fiabilité du résultat. Ce point est fondamental, car on observe que souvent se développe un sur enthousiasme en pratique en fonction de l'intérêt clinique du résultat et que par exemple un résultat de mortalité est souvent mis en avant sans tenir compte des réserves méthodologiques sous-jacentes.

Une taille d'effet importante ne permet pas d'exclure la possibilité d'un biais ou d'une erreur aléatoire.

Chapitre 6

Résultat faussement positif dû au hasard

Chapitre 7

est il possible d'accepter une décision basée sur ce résultat est il possible d'accepter que ce résultat ne soit pas dû au hasard \$

7.1 Généralités

Un résultat d'essai thérapeutique peut montrer le bénéfice du traitement par erreur, uniquement du fait du hasard. Avec un traitement en réalité sans effet, les fluctuations aléatoires d'échantillonnage peuvent faire apparaître une différence entre les 2 groupes de l'essai thérapeutique qui pourrait être prise, à tort, pour l'effet du traitement.

Note : résultat faux positif

Un résultat d'essai faussement positif (faux positif) est un résultat qui semble montrer le bénéfice du traitement (une différence de mortalité entre les 2 groupes par exemple) mais cela faussement car le traitement est en réalité sans effet. Les résultats faux positifs proviennent de deux grandes causes : les biais et le hasard.

Il existe aussi la possibilité d'avoir un résultat faussement négatif, c'est-à-dire ne concluant pas à l'efficacité d'un traitement qui pourtant en réalité est efficace.

En lecture critique, seul le risque de résultat faux positif est considéré car ces 2 risques ne sont symétriques. Le plus redouté pour la décision médicale est celui de faux positif. Car une fois qu'un traitement est accepté comme étant efficace, il ne sera plus possible de revenir en arrière. En effet, à partir de ce moment, il ne devient plus possible de faire un nouvel essai avec ce traitement qui est considéré comme efficace. Il ne sera pas possible de récupérer l'erreur initiale.

Par contre un faux négatif ne conduit pas à l'adoption du nouveau traitement. Cela peut être aussi gênant car cela prive le patient d'une ressource thérapeutique qui pourrait leur apporter un bénéfice. Mais cette erreur est récupérable car dans cette situation il est toujours possible de faire un nouvel essai. Et cette problématique n'est plus une problématique de décision médicale ou de santé publique, mais concerne les développeurs du traitement.

En lecture critique, la première question à se poser est celle de la possibilité que le résultat positif que l'on est en train de considérer soit un pur produit du hasard et qu'en réalité, le traitement n'est pas efficace.

7.1.1 Le test statistique et la "p value"

En pratique, il convient donc, avant d'accepter un résultat d'essai en faveur de l'efficacité du traitement, de s'assurer que ce résultat « positif » ne soit pas, en réalité, un résultat faux positif dû au hasard. L'outil à notre disposition pour faire ce diagnostic est simplement le test statistique et la valeur de p. En fait le risque de résultat faux positif est le risque d'erreur statistique alpha qui est le concept central du test statistique et de la signification statistique. Ainsi une différence statistiquement significative nous autorise à conclure à l'existence de l'efficacité du traitement car le risque alpha de faux positif est faible (<5% en général) et que l'on consent à prendre ce risque car on l'estime peu probable.

Ainsi quand le test statistique rattaché au résultat en considération est statistiquement significatif, on considère que le risque alpha est faible, acceptable dans notre prise de décision et qu'il est possible d'accepter la positivité du résultat. Cependant l'obtention d'un p statistiquement significatif n'est pas suffisant par lui-même pour raisonnablement écarter la possibilité d'un résultat faux positif. Encore faut-il pouvoir écarter une situation d'inflation du risque alpha induite par une multiplicité des comparaisons statistiques non gérée.

Note : le risque alpha

Un résultat d'essai thérapeutique peut montrer le bénéfice du traitement par erreur, uniquement du fait du hasard. Ainsi, avec un traitement en réalité sans effet, les fluctuations aléatoires d'échantillonnage peuvent faire apparaître une différence entre les 2 groupes de l'essai thérapeutique qui sera prise, à tort, pour l'effet du traitement.

En pratique, il convient donc, avant d'accepter un résultat d'essai en faveur de l'efficacité du traitement, de s'assurer que ce résultat « positif » ne soit pas, en réalité, un résultat faux positif dû au hasard. L'outil à notre disposition pour faire ce diagnostic est le test statistique et la valeur de p. En fait le risque de résultat faux positif est le risque d'erreur statistique alpha qui est le concept central du test statistique et de la signification statistique.

Ainsi quand le test statistique rattaché au résultat considéré est statistiquement significatif, on considère que le risque alpha est faible et qu'il est possible d'accepter la positivité du résultat. Il est possible d'écarter, raisonnablement avec un risque (probabilité) d'erreur résiduelle de moins de 5%, la possibilité que le résultat soit un faux positif dû au hasard.

Cependant l'obtention d'un p statistiquement significatif n'est pas suffisant pour formellement écarter la possibilité d'un résultat faux positif. Encore faut-il pouvoir écarter une situation d'inflation du risque alpha induite par une multiplicité statistique non gérée.

L'inflation du risque alpha survient quand on a la possibilité de faire la même conclusion sur l'efficacité du traitement à partir de plusieurs tests statistiques. A partir du moment, où à partir de plusieurs tests, on conclut à l'efficacité du traitement quand au moins l'un d'entre eux est significatif, le risque de conclure à tort à l'efficacité du traitement est bien plus important que 5%. En effet, dans cette démarche on consent un risque de 5% de conclure à tort à partir du 1er test effectué, puis à nouveau 5% sur le deuxième, etc. jusqu'à l'obtention d'un test significatif pris alors comme démonstration de l'efficacité. En effet, avec un traitement sans effet, si l'on réalise 100 tests statistiques (en multipliant les critères de jugement par exemple), 5 d'entre eux

seront statistiquement significatif par définition puisque chaque test à un risque de 5% de permettre de conclure à tort à l'effet du traitement. 5% pour chaque test cela fait en moyenne 5 tests positifs à tort. Ainsi, avec tout traitement sans effet, il sera possible d'avoir au moins un test significatif si l'on a multiplié les comparaisons statistique (à partir de 20 tests il est quasi certain d'avoir au moins un p significatif sur l'ensemble).

risque alpha et test statistique

7.1.2 L'inflation du risque alpha

<p> TODO Reformulation du problème de la signification stat dans le n contexte de la multiplicité</p>

Est-ce que le résultat est suffisamment probant pour faire reposer la décision de recommander le traitement sur lui (sans courir trop de risque de prendre cette décision à tort)

L'inflation du risque alpha survient quand on a la possibilité de faire la même conclusion sur l'intérêt du traitement à partir de plusieurs tests statistiques. Par exemple, pouvoir conclure à un quelconque intérêt du traitement à partir de plusieurs critères de jugements. À partir du moment, où, à partir de plusieurs tests, on conclut à l'efficacité du traitement quand au moins l'un d'entre eux est significatif, le risque de conclure à tort à l'efficacité du traitement est bien plus important que 5%. En effet, dans cette démarche on consent un risque de 5% de conclure à tort à partir du 1er test effectué, puis à nouveau 5% sur le deuxième, etc. jusqu'à l'obtention d'un test significatif pris alors comme démonstration de l'efficacité. En effet, avec un traitement sans effet, si l'on réalise 100 tests statistiques (en multipliant les critères de jugement par exemple), 5 d'entre eux seront statistiquement significatifs par définition puisque chaque test à un risque de 5% de permettre de conclure à tort à l'effet du traitement. Ainsi, avec tout traitement sans effet, il sera possible d'avoir au moins un test significatif si l'on a multiplié les comparaisons statistique (à partir de 20 tests il est quasi certain d'avoir au moins un p significatif sur l'ensemble).

Note : : inflation du risque alpha

En l'absence d'inflation du risque alpha, un résultat statistiquement significative permet de considérer, avec un risque minime et parfaitement contrôlé, que le résultat n'est probablement pas un faux positif du au hasard. En effet un $p < 0.05$ signifie que, si en réalité le traitement n'a pas d'effet, l'essai avait moins de 5% de chance d'obtenir une différence entre les 2 groupes au moins aussi importante que celle qui est observée. De ce fait il est considéré que le résultat n'est probablement pas un faux positif. Cette déduction n'est pas certaine à 100%. Il reste toujours un risque d'erreur résiduel quand on accepte l'efficacité du traitement à partir d'un résultat statistiquement significatif. En effet, un test significatif ne nous assure pas que le risque alpha est nul mais simplement qu'il est contrôlé et suffisamment petit (<5% si un tel seuil de signification a été choisi au protocole) pour que l'on puisse accepter de prendre ce risque de conclure à tort.

Encore faut-il que ce risque soit réellement inférieur à 5% et non pas supérieur. En effet, un $p < 0.05$ ne correspond pas forcément à une prise de risque contrôlée si ce petit p a été obtenu dans un contexte de multiplicité des comparaisons. Dans ce cas il y a inflation du risque alpha.

En effet, si pour prendre la décision de dire que le traitement présente un intérêt, nous avons à notre disposition le résultat de multiples comparaisons statistiques (par exemple en raison de multiples critères de jugement dans l'essai), le risque global de conclure à tort à l'intérêt de ce traitement à partir de n'importe laquelle de ces comparaisons va être bien plus important que le risque habituellement consentie de 5%. Car on va prendre un risque de 5% de conclure à tort à l'intérêt du traitement sur chacune de ces comparaisons et ces prises de risque vont donc se cumuler. Par exemple, s'il y a 100 critères de jugement, nous en aurons 5 de significatif même si le traitement n'a aucun effet (les 5% de faux positif acceptée par définition d'un risque alpha à 5%). Ainsi, si l'on a la possibilité de choisir la comparaison qui nous arrange pour conclure à l'efficacité du traitement, il y en aura presque toujours une de significative (si les critères ne sont pas trop corrélés), même si le traitement n'a aucun effet, à partir du moment où plusieurs comparaisons s'offrent à nous. Le nombre de résultats significatifs augmente avec le nombre de comparaisons disponibles. On dit que l'on va à la « pêche à la ligne ».

Pour que le contrôle du risque alpha de faux positif soit efficace il faut que la décision de conclure à l'intérêt du traitement ne puisse se faire qu'à partir d'une seule et unique comparaison, définie a priori ou qu'une méthode statistique de contrôle de l'inflation du risque alpha ai été mise en oeuvre (méthode séquentielle hiérarchique, ajustement du seuil de la signification statistique par la méthode Bonferroni, etc.. cf. section ???).

Dans l'essai thérapeutique, les situations qui peuvent engendrer une multiplicité des comparaisons si aucune mesure adaptée n'est prise, sont les suivantes :

- Multiplicité des critères de jugements sans critère de jugement principal défini a priori (ou sans recours à une méthode séquentielle hiérarchique ou à un ajustement du seuil de la signification)
 - Réalisation d'analyse en sous-groupe pour chercher un sous-groupe de patients permettant de conclure à l'intérêt du traitement
 - Réalisation d'analyse intermédiaire sans méthode statistique adaptée
 - Mesures répétées du critère de jugement au cours du suivi des patients
 - Multiplicités des bras de traitement (plusieurs doses du même traitement par exemple)
- [multiplicité des comparaisons et inflation du risque alpha](#)

7.1.3 démarche hypothetico déductive et cohérence externe du résultat

Deux autres problématiques peuvent être rattachées à celle des résultats faussement positifs : celle de la cohérence externe du résultat et celle du respect de la démarche hypothético déductive.

La cohérence externe d'un résultat est assurée lorsque celui-ci est confirmé par d'autres études. L'outil d'exploration de la cohérence externe est la méta-analyse. En méta-analyse, la cohérence externe se traduit par l'absence d'hétérogénéité entre les études signifiant que le résultat est bien retrouvé à l'identique dans toutes études disponibles. A l'inverse, la méta permet d'identifier un résultat artefactuel ou dû au hasard, en montrant que ce

résultat est différents de ceux obtenus par les autres essais. (cf. [fiche concept cohérence externe](#)).

Le respect de la démarche hypothético est indispensable pour assurer la validité scientifique du résultat. Le résultat correspond à une hypothèse formulée avant la réalisation de l'étude et qui était l'objet de l'étude. Ainsi le résultat est le reflet de la confrontation à la réalité de cette hypothèse, confrontation qui va confirmer ou infirmer cette hypothèse. Dans ce cas l'essai était un essai de confirmation (« confirmatory study ») s'inscrivant dans une démarche déductive. A l'inverse, les résultats obtenus en dehors de l'hypothèse de l'essai n'ont que peu de valeur car ils sont purement exploratoires car produit par une démarche inductive. Ce sont des résultats post-hoc qui demande à être confirmés par une étude prospective spécialement conçu pour cela. Cf. [fiche concept sur la démarche hypothetico déductive](#) .

7.2 Situation où il est possible d'exclure que le résultat soit un faux positif dû au hasard

Le risque que ce résultat soit faussement positif dû au hasard peut être raisonnablement écarté. En effet, ce résultat est statistiquement significatif (à un seuil de signification au plus égal à 5%) et il n'y a pas eu de multiplicité des comparaisons dans la recherche de ce résultat. De ce fait il n'y a pas eu d'inflation du risque alpha. On peut donc considérer que le risque alpha de conclure à tort à l'intérêt du traitement à partir de ce résultat est parfaitement contrôlé et inférieur à 2.5% (5% en bilatéral, voir [fiche concept test unilatéral/bilatéral](#)). En effet, les tests réalisés dans les essais thérapeutiques sont bilatéraux mais un seul côté de cette bilatéralité (la supériorité du nouveau traitement versus le control) conduit à la conclusion de l'intérêt du traitement. Le risque alpha consentie sur cette conclusion est donc de 2.5% (pour un test bilatéral de 5%). Il est ainsi acceptable. Une inflation du risque alpha, qui conduirait à un risque de conclure à tort à l'intérêt du traitement bien supérieur à 5% et donc inacceptable, est écartée car il n'y a eu de multiple possibilité pour trouver un résultat significatif en faveur de cet intérêt.

L'absence d'inflation du risque alpha est assurée car :

- Ce résultat a été obtenu sur le critère de jugement principal défini a priori ou par une autre méthode garantissant l'absence d'inflation du risque alpha en cas d'utilisation de critère de jugement multiple (« co-primary endpoint ») : méthode séquentielle hiérarchique, ajustement du seuil de la signification (par la méthode de Bonferroni, Hochberg) par répartition du risque alpha global entre les co-critères de jugement primaires, etc..
- Ce résultat n'est pas obtenu dans un sous groupe de patient mais bien au niveau de la population de l'essai toute entière ; ou une méthode de contrôle du risque alpha a été utilisée si ce résultat correspond à une sous population de l'essai (méthode séquentielle hiérarchique)
- Ce résultat n'est pas issu d'une analyse intermédiaire réalisée sans protection contre l'inflation du risque alpha (cf. [fiche concept analyses intermédiaires](#))
- Ce résultat n'est pas issu d'un essai multibras sans prise en compte de la multiplicité engendrée

7.3 Situation où il n'est pas possible d'exclure que le résultat soit un faux positif dû au hasard

Il n'est pas possible d'écarter la possibilité que le résultat soit un faux positif dû au hasard car il n'y a pas eu de contrôle strict du risque alpha en raison d'une multiplication des comparaisons effectuées à la recherche de l'intérêt du traitement (entraînant une inflation du risque alpha). De ce fait il est impossible de prendre le risque de conclure à l'intérêt du traitement à partir de ce résultat, même si le p est inférieur à 5%, car le risque de faire cette conclusion à tort (erreur statistique alpha) est trop important (cf. supra) et il est supérieur au 5% bilatéral / 2.5% unilatéral habituellement consenti.

L'inflation du risque alpha peut provenir de plusieurs situations :

- Le résultat n'a pas été obtenu sur le critère de jugement principal de l'étude mais sur un critère de jugement secondaire ; ou il n'y a pas eu de définition d'un critère de jugement principal
- Le résultat a été obtenu dans une démarche séquentielle hiérarchique après un test non significatif. De ce fait, bien que le p nominal obtenu sur ce résultat soit inférieur à 5% il n'est pas possible de considérer que ce résultat est statistiquement significatif (cf. fiche concept)
- Le résultat a été obtenu sur un co-critère de jugement principal (co-primary endpoint) mais avec un p , bien qu'inférieur à 5%, qui est néanmoins supérieur au seuil de la signification ajusté pour prendre en compte la multiplicité des critères primaires.
- Le résultat a été obtenu sur un sous groupe (le fait que le sous groupe soit défini a priori au protocole n'empêche pas l'inflation du risque alpha. C'est éventuellement la condition préliminaire indispensable pour mettre en œuvre un ajustement de la signification statistique, mais, en soit, cette mention n'empêche pas l'inflation du risque alpha. Idem au fait que l'essai soit significatif sur tous les patients)
- Le résultat a été obtenu lors d'une analyse intermédiaire non prévue ou non protégée contre l'inflation du risque alpha

7.4 Résultat montrant un surcroît d'effets indésirables, risque de faux positif

Le résultat montre un surcroît d'événements indésirables (défaut de sécurité). Compte tenu des conséquences potentielles graves d'un défaut de sécurité, ce type de résultat n'a pas besoin d'être démontré formellement pour être pris en considération. Un résultat seulement suggéré est suffisant pour s'interroger sur un éventuel effet délétère et faire comme si cet effet était démontré. De plus, ce type de résultat n'est pas en mesure de conduire à une adoption à tort du traitement mais à son rejet. Il n'y a donc pas de crainte particulière liée à un résultat faussement positif. Ainsi, dans le domaine de la sécurité, le principe de précaution prévaut et il n'est pas nécessaire d'obtenir une démonstration formelle pour décider.

Avec le principe de précaution, la logique d'interprétation change et l'accent n'est plus mis sur le contrôle du risque de faux positif (conclure à tort à l'existence de l'effet) mais bien sur celui du risque de faux négatif (ne pas mettre en évidence à tort un effet indésirable). On accepte de conclure à tort à un défaut de sécurité car le risque d'utiliser à tort un traitement non sûr est considéré comme non acceptable. Si le nouveau traitement était comparé à un traitement actif, ce type de résultat signifie une moindre sécurité du nouveau traitement par rapport au précédent. Cela correspond à un défaut de sécurité du nouveau traitement

par rapport à la meilleure option actuellement possible.

Les conséquences de la suspicion ou de la démonstration d'un surcroît d'un événement indésirables varient en fonction du contexte et de l'existence ou non d'un bénéfice à mettre en regard.

La mise en évidence d'un surcroît d'événements indésirables peut ne pas être rédhibitoire si, en regard, le traitement apporte un bénéfice sur des événements de même gravité que celle de l'événement indésirable induit. Tout dépend de l'existence ou non d'un bénéfice clinique net favorable (cf. section ???). Par contre si aucun bénéfice n'existe en regard de ce surcroît d'événement indésirable (sur des événements de même gravité que les événements indésirable), l'intérêt thérapeutique du traitement évalué est inexistant. L'intérêt thérapeutique sera aussi remis en cause si un bénéfice existe ne débouchant sur un bénéfice clinique favorable (cf. section a ???). Dans ce, le surcroît d'événements indésirables viens contrebalancer complètement le bénéfice apporté par le traitement.

Chapitre 8

Résultat faussement positif dû au hasard

Chapitre 9

est il possible d'accepter une décision basée sur ce résultat est il possible d'accepter que ce résultat ne soit pas dû au hasard \$

9.1 Généralités

Un résultat d'essai thérapeutique peut montrer le bénéfice du traitement par erreur, uniquement du fait du hasard. Avec un traitement en réalité sans effet, les fluctuations aléatoires d'échantillonnage peuvent faire apparaître une différence entre les 2 groupes de l'essai thérapeutique qui pourrait être prise, à tort, pour l'effet du traitement.

Note : résultat faux positif

Un résultat d'essai faussement positif (faux positif) est un résultat qui semble montrer le bénéfice du traitement (une différence de mortalité entre les 2 groupes par exemple) mais cela faussement car le traitement est en réalité sans effet. Les résultats faux positifs proviennent de deux grandes causes : les biais et le hasard.

Il existe aussi la possibilité d'avoir un résultat faussement négatif, c'est-à-dire ne concluant pas à l'efficacité d'un traitement qui pourtant en réalité est efficace.

En lecture critique, seul le risque de résultat faux positif est considéré car ces 2 risques ne sont symétriques. Le plus redouté pour la décision médicale est celui de faux positif. Car une fois qu'un traitement est accepté comme étant efficace, il ne sera plus possible de revenir en arrière. En effet, à partir de ce moment, il ne devient plus possible de faire un nouvel essai avec ce traitement qui est considéré comme efficace. Il ne sera pas possible de récupérer l'erreur initiale.

Par contre un faux négatif ne conduit pas à l'adoption du nouveau traitement. Cela peut être aussi gênant car cela prive le patient d'une ressource thérapeutique qui pourrait leur apporter un bénéfice. Mais cette erreur est récupérable car dans cette situation il est toujours possible de faire un nouvel essai. Et cette problématique n'est plus une problématique de décision médicale ou de santé publique, mais concerne les développeurs du traitement.

En lecture critique, la première question à se poser est celle de la possibilité que le résultat positif que l'on est en train de considérer soit un pur produit du hasard et qu'en réalité, le traitement n'est pas efficace.

9.1.1 Le test statistique et la "p value"

En pratique, il convient donc, avant d'accepter un résultat d'essai en faveur de l'efficacité du traitement, de s'assurer que ce résultat « positif » ne soit pas, en réalité, un résultat faux positif dû au hasard. L'outil à notre disposition pour faire ce diagnostic est simplement le test statistique et la valeur de p. En fait le risque de résultat faux positif est le risque d'erreur statistique alpha qui est le concept central du test statistique et de la signification statistique. Ainsi une différence statistiquement significative nous autorise à conclure à l'existence de l'efficacité du traitement car le risque alpha de faux positif est faible (<5% en général) et que l'on consent à prendre ce risque car on l'estime peu probable.

Ainsi quand le test statistique rattaché au résultat en considération est statistiquement significatif, on considère que le risque alpha est faible, acceptable dans notre prise de décision et qu'il est possible d'accepter la positivité du résultat. Cependant l'obtention d'un p statistiquement significatif n'est pas suffisant par lui-même pour raisonnablement écarter la possibilité d'un résultat faux positif. Encore faut-il pouvoir écarter une situation d'inflation du risque alpha induite par une multiplicité des comparaisons statistiques non gérée.

Note : le risque alpha

Un résultat d'essai thérapeutique peut montrer le bénéfice du traitement par erreur, uniquement du fait du hasard. Ainsi, avec un traitement en réalité sans effet, les fluctuations aléatoires d'échantillonnage peuvent faire apparaître une différence entre les 2 groupes de l'essai thérapeutique qui sera prise, à tort, pour l'effet du traitement.

En pratique, il convient donc, avant d'accepter un résultat d'essai en faveur de l'efficacité du traitement, de s'assurer que ce résultat « positif » ne soit pas, en réalité, un résultat faux positif dû au hasard. L'outil à notre disposition pour faire ce diagnostic est le test statistique et la valeur de p. En fait le risque de résultat faux positif est le risque d'erreur statistique alpha qui est le concept central du test statistique et de la signification statistique.

Ainsi quand le test statistique rattaché au résultat considéré est statistiquement significatif, on considère que le risque alpha est faible et qu'il est possible d'accepter la positivité du résultat. Il est possible d'écarter, raisonnablement avec un risque (probabilité) d'erreur résiduelle de moins de 5%, la possibilité que le résultat soit un faux positif du au hasard.

Cependant l'obtention d'un p statistiquement significatif n'est pas suffisant pour formellement écarter la possibilité d'un résultat faux positif. Encore faut-il pouvoir écarter une situation d'inflation du risque alpha induite par une multiplicité statistique non gérée.

L'inflation du risque alpha survient quand on a la possibilité de faire la même conclusion sur l'efficacité du traitement à partir de plusieurs tests statistiques. A partir du moment, où à partir de plusieurs tests, on conclut à l'efficacité du traitement quand au moins l'un d'entre eux est significatif, le risque de conclure à tort à l'efficacité du traitement est bien plus important que 5%. En effet, dans cette démarche on consent un risque de 5% de conclure à tort à partir du 1er test effectué, puis à nouveau 5% sur le deuxième, etc. jusqu'à l'obtention d'un test significatif pris alors comme démonstration de l'efficacité. En effet, avec un traitement sans effet, si l'on réalise 100 tests statistiques (en multipliant les critères de jugement par exemple), 5 d'entre eux

seront statistiquement significatif par définition puisque chaque test à un risque de 5% de permettre de conclure à tort à l'effet du traitement. 5% pour chaque test cela fait en moyenne 5 tests positifs à tort. Ainsi, avec tout traitement sans effet, il sera possible d'avoir au moins un test significatif si l'on a multiplié les comparaisons statistique (à partir de 20 tests il est quasi certain d'avoir au moins un p significatif sur l'ensemble).

risque alpha et test statistique

9.1.2 L'inflation du risque alpha

<p> TODO Reformulation du problème de la signification stat dans le n contexte de la multiplicité</p>

Est-ce que le résultat est suffisamment probant pour faire reposer la décision de recommander le traitement sur lui (sans courir trop de risque de prendre cette décision à tort)

L'inflation du risque alpha survient quand on a la possibilité de faire la même conclusion sur l'intérêt du traitement à partir de plusieurs tests statistiques. Par exemple, pouvoir conclure à un quelconque intérêt du traitement à partir de plusieurs critères de jugements. À partir du moment, où, à partir de plusieurs tests, on conclut à l'efficacité du traitement quand au moins l'un d'entre eux est significatif, le risque de conclure à tort à l'efficacité du traitement est bien plus important que 5%. En effet, dans cette démarche on consent un risque de 5% de conclure à tort à partir du 1er test effectué, puis à nouveau 5% sur le deuxième, etc. jusqu'à l'obtention d'un test significatif pris alors comme démonstration de l'efficacité. En effet, avec un traitement sans effet, si l'on réalise 100 tests statistiques (en multipliant les critères de jugement par exemple), 5 d'entre eux seront statistiquement significatifs par définition puisque chaque test à un risque de 5% de permettre de conclure à tort à l'effet du traitement. Ainsi, avec tout traitement sans effet, il sera possible d'avoir au moins un test significatif si l'on a multiplié les comparaisons statistique (à partir de 20 tests il est quasi certain d'avoir au moins un p significatif sur l'ensemble).

Note : : inflation du risque alpha

En l'absence d'inflation du risque alpha, un résultat statistiquement significative permet de considérer, avec un risque minime et parfaitement contrôlé, que le résultat n'est probablement pas un faux positif du au hasard. En effet un $p < 0.05$ signifie que, si en réalité le traitement n'a pas d'effet, l'essai avait moins de 5% de chance d'obtenir une différence entre les 2 groupes au moins aussi importante que celle qui est observée. De ce fait il est considéré que le résultat n'est probablement pas un faux positif. Cette déduction n'est pas certaine à 100%. Il reste toujours un risque d'erreur résiduel quand on accepte l'efficacité du traitement à partir d'un résultat statistiquement significatif. En effet, un test significatif ne nous assure pas que le risque alpha est nul mais simplement qu'il est contrôlé et suffisamment petit (<5% si un tel seuil de signification a été choisi au protocole) pour que l'on puisse accepter de prendre ce risque de conclure à tort.

Encore faut-il que ce risque soit réellement inférieur à 5% et non pas supérieur. En effet, un $p < 0.05$ ne correspond pas forcément à une prise de risque contrôlée si ce petit p a été obtenu dans un contexte de multiplicité des comparaisons. Dans ce cas il y a inflation du risque alpha.

En effet, si pour prendre la décision de dire que le traitement présente un intérêt, nous avons à notre disposition le résultat de multiples comparaisons statistiques (par exemple en raison de multiples critères de jugement dans l'essai), le risque global de conclure à tort à l'intérêt de ce traitement à partir de n'importe laquelle de ces comparaisons va être bien plus important que le risque habituellement consentie de 5%. Car on va prendre un risque de 5% de conclure à tort à l'intérêt du traitement sur chacune de ces comparaisons et ces prises de risque vont donc se cumuler. Par exemple, s'il y a 100 critères de jugement, nous en aurons 5 de significatif même si le traitement n'a aucun effet (les 5% de faux positif acceptée par définition d'un risque alpha à 5%). Ainsi, si l'on a la possibilité de choisir la comparaison qui nous arrange pour conclure à l'efficacité du traitement, il y en aura presque toujours une de significative (si les critères ne sont pas trop corrélés), même si le traitement n'a aucun effet, à partir du moment où plusieurs comparaisons s'offrent à nous. Le nombre de résultats significatifs augmente avec le nombre de comparaisons disponibles. On dit que l'on va à la « pêche à la ligne ».

Pour que le contrôle du risque alpha de faux positif soit efficace il faut que la décision de conclure à l'intérêt du traitement ne puisse se faire qu'à partir d'une seule et unique comparaison, définie a priori ou qu'une méthode statistique de contrôle de l'inflation du risque alpha ai été mise en oeuvre (méthode séquentielle hiérarchique, ajustement du seuil de la signification statistique par la méthode Bonferroni, etc.. cf. section ???).

Dans l'essai thérapeutique, les situations qui peuvent engendrer une multiplicité des comparaisons si aucune mesure adaptée n'est prise, sont les suivantes :

- Multiplicité des critères de jugements sans critère de jugement principal défini a priori (ou sans recours à une méthode séquentielle hiérarchique ou à un ajustement du seuil de la signification)
 - Réalisation d'analyse en sous-groupe pour chercher un sous-groupe de patients permettant de conclure à l'intérêt du traitement
 - Réalisation d'analyse intermédiaire sans méthode statistique adaptée
 - Mesures répétées du critère de jugement au cours du suivi des patients
 - Multiplicités des bras de traitement (plusieurs doses du même traitement par exemple)
- [multiplicité des comparaisons et inflation du risque alpha](#)

9.1.3 démarche hypothetico déductive et cohérence externe du résultat

Deux autres problématiques peuvent être rattachées à celle des résultats faussement positifs : celle de la cohérence externe du résultat et celle du respect de la démarche hypothetico déductive.

La cohérence externe d'un résultat est assurée lorsque celui-ci est confirmé par d'autres études. L'outil d'exploration de la cohérence externe est la méta-analyse. En méta-analyse, la cohérence externe se traduit par l'absence d'hétérogénéité entre les études signifiant que le résultat est bien retrouvé à l'identique dans toutes études disponibles. A l'inverse, la méta permet d'identifier un résultat artefactuel ou dû au hasard, en montrant que ce

résultat est différents de ceux obtenus par les autres essais. (cf. [fiche concept cohérence externe](#)).

Le respect de la démarche hypothético est indispensable pour assurer la validité scientifique du résultat. Le résultat correspond à une hypothèse formulée avant la réalisation de l'étude et qui était l'objet de l'étude. Ainsi le résultat est le reflet de la confrontation à la réalité de cette hypothèse, confrontation qui va confirmer ou infirmer cette hypothèse. Dans ce cas l'essai était un essai de confirmation (« confirmatory study ») s'inscrivant dans une démarche déductive. A l'inverse, les résultats obtenus en dehors de l'hypothèse de l'essai n'ont que peu de valeur car ils sont purement exploratoires car produit par une démarche inductive. Ce sont des résultats post-hoc qui demande à être confirmés par une étude prospective spécialement conçu pour cela. Cf. [fiche concept sur la démarche hypothetico déductive](#) .

9.2 Situation où il est possible d'exclure que le résultat soit un faux positif dû au hasard

Le risque que ce résultat soit faussement positif dû au hasard peut être raisonnablement écarté. En effet, ce résultat est statistiquement significatif (à un seuil de signification au plus égal à 5%) et il n'y a pas eu de multiplicité des comparaisons dans la recherche de ce résultat. De ce fait il n'y a pas eu d'inflation du risque alpha. On peut donc considérer que le risque alpha de conclure à tort à l'intérêt du traitement à partir de ce résultat est parfaitement contrôlé et inférieur à 2.5% (5% en bilatéral, voir [fiche concept test unilatéral/bilatéral](#)). En effet, les tests réalisés dans les essais thérapeutiques sont bilatéraux mais un seul côté de cette bilatéralité (la supériorité du nouveau traitement versus le control) conduit à la conclusion de l'intérêt du traitement. Le risque alpha consentie sur cette conclusion est donc de 2.5% (pour un test bilatéral de 5%). Il est ainsi acceptable. Une inflation du risque alpha, qui conduirait à un risque de conclure à tort à l'intérêt du traitement bien supérieur à 5% et donc inacceptable, est écartée car il n'y a eu de multiple possibilité pour trouver un résultat significatif en faveur de cet intérêt.

L'absence d'inflation du risque alpha est assurée car :

- Ce résultat a été obtenu sur le critère de jugement principal défini a priori ou par une autre méthode garantissant l'absence d'inflation du risque alpha en cas d'utilisation de critère de jugement multiple (« co-primary endpoint ») : méthode séquentielle hiérarchique, ajustement du seuil de la signification (par la méthode de Bonferroni, Hochberg) par répartition du risque alpha global entre les co-critères de jugement primaires, etc..
- Ce résultat n'est pas obtenu dans un sous groupe de patient mais bien au niveau de la population de l'essai toute entière ; ou une méthode de contrôle du risque alpha a été utilisée si ce résultat correspond à une sous population de l'essai (méthode séquentielle hiérarchique)
- Ce résultat n'est pas issu d'une analyse intermédiaire réalisée sans protection contre l'inflation du risque alpha (cf. [fiche concept analyses intermédiaires](#))
- Ce résultat n'est pas issu d'un essai multibras sans prise en compte de la multiplicité engendrée

9.3 Situation où il n'est pas possible d'exclure que le résultat soit un faux positif dû au hasard

Il n'est pas possible d'écarter la possibilité que le résultat soit un faux positif dû au hasard car il n'y a pas eu de contrôle strict du risque alpha en raison d'une multiplication des comparaisons effectuées à la recherche de l'intérêt du traitement (entraînant une inflation du risque alpha). De ce fait il est impossible de prendre le risque de conclure à l'intérêt du traitement à partir de ce résultat, même si le p est inférieur à 5%, car le risque de faire cette conclusion à tort (erreur statistique alpha) est trop important (cf. supra) et il est supérieur au 5% bilatéral / 2.5% unilatéral habituellement consenti.

L'inflation du risque alpha peut provenir de plusieurs situations :

- Le résultat n'a pas été obtenu sur le critère de jugement principal de l'étude mais sur un critère de jugement secondaire ; ou il n'y a pas eu de définition d'un critère de jugement principal
- Le résultat a été obtenu dans une démarche séquentielle hiérarchique après un test non significatif. De ce fait, bien que le p nominal obtenu sur ce résultat soit inférieur à 5% il n'est pas possible de considérer que ce résultat est statistiquement significatif (cf. fiche concept)
- Le résultat a été obtenu sur un co-critère de jugement principal (co-primary endpoint) mais avec un p , bien qu'inférieur à 5%, qui est néanmoins supérieur au seuil de la signification ajusté pour prendre en compte la multiplicité des critères primaires.
- Le résultat a été obtenu sur un sous groupe (le fait que le sous groupe soit défini a priori au protocole n'empêche pas l'inflation du risque alpha. C'est éventuellement la condition préliminaire indispensable pour mettre en œuvre un ajustement de la signification statistique, mais, en soit, cette mention n'empêche pas l'inflation du risque alpha. Idem au fait que l'essai soit significatif sur tous les patients)
- Le résultat a été obtenu lors d'une analyse intermédiaire non prévue ou non protégée contre l'inflation du risque alpha

9.4 Résultat montrant un surcroît d'effets indésirables, risque de faux positif

Le résultat montre un surcroît d'événements indésirables (défaut de sécurité). Compte tenu des conséquences potentielles graves d'un défaut de sécurité, ce type de résultat n'a pas besoin d'être démontré formellement pour être pris en considération. Un résultat seulement suggéré est suffisant pour s'interroger sur un éventuel effet délétère et faire comme si cet effet était démontré. De plus, ce type de résultat n'est pas en mesure de conduire à une adoption à tort du traitement mais à son rejet. Il n'y a donc pas de crainte particulière liée à un résultat faussement positif. Ainsi, dans le domaine de la sécurité, le principe de précaution prévaut et il n'est pas nécessaire d'obtenir une démonstration formelle pour décider.

Avec le principe de précaution, la logique d'interprétation change et l'accent n'est plus mis sur le contrôle du risque de faux positif (conclure à tort à l'existence de l'effet) mais bien sur celui du risque de faux négatif (ne pas mettre en évidence à tort un effet indésirable). On accepte de conclure à tort à un défaut de sécurité car le risque d'utiliser à tort un traitement non sûr est considéré comme non acceptable. Si le nouveau traitement était comparé à un traitement actif, ce type de résultat signifie une moindre sécurité du nouveau traitement par rapport au précédent. Cela correspond à un défaut de sécurité du nouveau traitement

par rapport à la meilleure option actuellement possible.

Les conséquences de la suspicion ou de la démonstration d'un surcroît d'un événement indésirables varient en fonction du contexte et de l'existence ou non d'un bénéfice à mettre en regard.

La mise en évidence d'un surcroît d'événements indésirables peut ne pas être rédhibitoire si, en regard, le traitement apporte un bénéfice sur des événements de même gravité que celle de l'événement indésirable induit. Tout dépend de l'existence ou non d'un bénéfice clinique net favorable (cf. section ???). Par contre si aucun bénéfice n'existe en regard de ce surcroît d'événement indésirable (sur des événements de même gravité que les événements indésirable), l'intérêt thérapeutique du traitement évalué est inexistant. L'intérêt thérapeutique sera aussi remis en cause si un bénéfice existe ne débouchant sur un bénéfice clinique favorable (cf. section a ???). Dans ce, le surcroît d'événements indésirables viens contrebalancer complètement le bénéfice apporté par le traitement.

Chapitre 10

Résultat faussement positif dû à un biais

Une cause de résultat d'étude faussement positif est l'existence d'un biais dans l'étude. Un biais est un défaut de conception ou de réalisation de l'essai qui fait que, quelle que soit la réelle efficacité du traitement, l'essai montrera toujours une différence entre les 2 groupes favorable au traitement évalué. Ainsi, même si le traitement est sans effet, le résultat de l'essai sera « positif » et constituera donc un « faux positif ». En lecture critique, il convient donc de se poser la question si le résultat en faveur de l'effet du traitement en cours de considération n'est pas simplement le fait d'un biais. Il est facile d'imaginer des circonstances conduisant à un biais et à un résultat faussement positif. Par exemple, si on s'est arrangé pour que tous les patients à mauvais pronostic se retrouvent dans le groupe contrôle et que ceux pronostic soient inclus dans le groupe du traitement expérimental. Ou bien encore, un essai où les patients du groupe expérimental reçoivent en forte proportion un traitement concomitant efficace tandis que les patients du groupe contrôle ne le reçoivent jamais. Ainsi, même en l'absence de tout effet du traitement, l'essai ne pourra donner qu'un résultat montrant une différence entre les 2 groupes, en faveur du traitement étudié. L'essai est donc biaisé. Ici l'erreur n'est pas aléatoire et la réplication de l'essai avec le même défaut (le même biais), conduira au même résultat.

S'il est facile d'imaginer ces situations biaisées, il est par contre très difficile de les détecter a posteriori, en analysant les compte-rendus des essais. Par exemple, comment détecter un recours asymétrique à des traitements concomitants si la fréquence d'utilisation de ces traitements dans les 2 groupes n'est pas rapporté dans la publication ou le rapport. Pour cette raison, il a été inventé des méthodes qui empêchent la survenue des biais : les principes méthodologiques. Ainsi, si ces méthodes sont correctement appliquées dans l'essai, il n'est plus nécessaire de se poser la question de l'existence de biais car l'essai est à l'abri des biais.

En lecture critique, pour écarter la possibilité que le résultat puisse être un faux positif dû aux biais on analysera simplement le degré de protection de l'essai vis-à-vis des biais, c'est-à-dire la mise en oeuvre correcte de ces principes méthodologiques. Détecter un biais effectif est très difficile, voire impossible. Par contre il est possible de s'assurer que l'essai est à l'abri des biais et que, donc, par conséquence, le résultat n'est pas dû à un biais.

Plusieurs biais peuvent venir fausser le résultat d'un essai thérapeutique et donner un résultat positif même en l'absence de toute efficacité du traitement.

Le **biais de sélection** survient quand il y a eu un processus de sélection qui a conduit à mettre préférentiellement les patients à bon pronostic dans le groupe expérimental et les patients à mauvais pronostic dans le groupe contrôle. Il y a donc possibilité d'un

biais de sélection chaque fois que le processus d'allocation des traitements aux patients pourra effectuée une telle sélection. Mettre l'essai à l'abri du biais de sélection nécessite donc d'écarter totalement la possibilité d'une sélection des patients à bon pronostic dans le groupe expérimental. Le moyen d'y parvenir est la randomisation c'est-à-dire, confier cette allocation des traitements à un processus aléatoire qui donne la même probabilité de recevoir chaque traitement à tous les patients quel que soit leur pronostic (risque de base) ou leur valeur du critère de jugement. Encore faut-il que ce processus de randomisation soit imprévisible pour éviter cette sélection opportune des patients. Dans un essai en ouvert, l'imprévisibilité de la randomisation est garantie par une inclusion centralisée des patients qui empêche que la décision définitive d'inclure ou non un patient puisse survenir après avoir pris connaissance de la nature réelle du traitement que recevrait le patient s'il est inclus.

le biais de sélection Un **biais de suivi** (appelé aussi biais de réalisation) survient quand les patients des 2 groupes ne reçoivent pas les mêmes soins durant la période de l'essai, par exemple, en termes de traitements concomitants efficaces. Mettre l'essai à l'abri du biais de suivi nécessite donc d'écarter totalement la possibilité d'un suivi ou d'une prise en charge différente des patients entre les 2 groupes. Le moyen d'y parvenir est l'administration des traitements en insu (double aveugle). Ainsi il est impossible de connaître la nature exacte du traitement que reçoit le patient. Il devient donc impossible qu'une prise en charge différente ait pu être appliquée entre les deux groupes et l'essai est à l'abri du biais de suivi. Bien sûr, s'il y a un recours important à un traitement concomitant efficace chez presque tous les patients, cela conduira à ce que tous les patients des 2 groupes bénéficient d'un traitement efficace, ce qui masquera l'effet du traitement évalué. Bien que l'essai donnera dans ce cas un résultat erroné si le traitement évalué est efficace, cela ne sera pas un faux positif mais à un faux négatif. L'essai est donc à l'abri de donner un résultat faux positif. On dit aussi que le processus est conservateur. Par abus de langage on dit aussi que l'essai est à l'abri des biais, même si le double aveugle peut fausser (biaiser) le résultat dans le sens d'un faux négatif, car seuls les faux positifs sont redoutés.

Un **biais de mesure** survient quand le critère de jugement ne sera pas mesuré de la même façon entre les 2 groupes. Si le processus de diagnostic des événements critères de jugement est plus sensible dans le groupe contrôle, cela entraînera une surestimation de la fréquence de ces événements dans le groupe contrôle pouvant, in fine, conduire à une différence entre les deux groupes en faveur du traitement évalué.

le biais de suivi et le biais de mesure Un **biais d'attrition** survient quand des patients sont exclus de l'analyse en raison de leur évolution sous traitement. L'attrition désigne la perte de patients entre l'inclusion (la randomisation) et l'analyse. Si par exemple, 200 patients ont été inclus mais si seulement 180 participent à l'analyse finale, il y a eu une attrition de 20 patients (10%). Cette attrition peut provenir du fait que certains patients sont perdus de vue (ils ne se sont pas présentés à la visite finale de l'essai) ou qu'ils ont demandé à sortir de l'étude. Mais l'attrition peut aussi être engendrée par les investigateurs qui sortiraient des patients de l'essais car ils ont été inclus à tort, il y a eu une erreur dans l'application de leur traitement, etc.

Au total, ces sorties d'études peuvent induire un biais (le biais d'attrition) en soustrayant de l'information utile à l'estimation de l'effet du traitement.

Pour éviter ce biais il convient qu'il n'y ait pas d'attrition et que tous les patients inclus dans l'essai participent à l'analyse. Cela est l'objet du principe de l'analyse en intention de traiter : tous les patients inclus sont analysés. L'investigateur ne doit procéder à aucune exclusion de l'analyse même si les patients n'ont pas reçu le traitement prévu (d'où l'origine du nom "intention de traiter").

Cependant, même avec l'intention d'exclure aucun patients, si le critère de jugement n'est pas disponible pour certains patients, il ne va pas être possible de prendre en compte tous les patients dans l'analyse. Pour éviter cela, un remplacement des données manquantes du critère de jugement est effectuée avec une méthode conservatrice comme le biais maximum pour les variables binaires ou LOCF pour les variables quantitatives.

le biais d'attrition Ainsi il ne sera possible d'exclure qu'un résultat positif¹ soit un faux positif dû à un biais que si l'essai a mis en oeuvre tous les moyens de se protéger contre ces biais :

- Une randomisation imprévisible pour se protéger contre le biais de sélection
- Une réalisation en double aveugle réel pour se protéger contre un biais de suivi et un biais de mesure
- Une analyse en intention de traiter avec remplacement des données manquantes de manière conservatrice pour se protéger contre le biais d'attrition

En cas de manquement sur l'un de ces principes, il n'est plus possible d'exclure que le résultat considéré soit un faux positif dû au hasard. Ce résultat ne pourra donc pas être considéré comme démontré.

10.1 Situations où il est possible d'exclure que le résultat soit un faux positif dû à un biais

Il n'est possible d'exclure qu'un résultat positif soit un faux positif dû à un biais que si l'essai a mis en oeuvre, de manière satisfaisante, tous les moyens disponibles pour se protéger contre ces biais :

- Une randomisation imprévisible pour se protéger contre le biais de sélection
- Une réalisation en double aveugle réel pour se protéger contre un biais de suivi et un biais de mesure
- Une analyse en intention de traiter avec remplacement des données manquantes de manière conservatrice pour se protéger contre le biais d'attrition

10.2 Situations où il n'est pas possible d'exclure que le résultat soit un faux positif dû à un biais

Plusieurs biais peuvent venir fausser le résultat d'un essai thérapeutique et donner un résultat positif même en l'absence de toute efficacité du traitement (cf. supra).

Ainsi il n'est possible d'exclure qu'un résultat positif soit un faux positif dû à un biais que si l'essai a mis en oeuvre tous les moyens de se protéger contre ces biais :

- Une randomisation imprévisible pour se protéger contre le biais de sélection
- Une réalisation en double aveugle réel pour se protéger contre un biais de suivi et un biais de mesure
- Une analyse en intention de traiter avec remplacement des données manquantes de manière conservatrice pour se protéger contre le biais d'attrition

En cas de manquement sur l'un de ces principes, il n'est plus possible d'exclure que le résultat considéré soit un faux positif dû à un biais. Ce résultat ne pourra donc pas être considéré comme démontré mais seulement suggéré et ne constituera pas un argument utilisable pour conclure à l'intérêt thérapeutique du traitement.

1. d'un essai de supériorité. Les moyens de protection contre les biais diffèrent dans l'essai de @1@.

Chapitre 11

Pertinence clinique

La pertinence clinique correspond à la valeur médicale du résultat : représente-t-il un progrès thérapeutique substantiel ?

Un résultat cliniquement pertinent assure que le bénéfice du traitement à un réel sens médical et qu'il correspond à un réel bénéfice clinique pour les patients.

La notion de pertinence clinique recouvre plusieurs éléments :

- La pertinence clinique de la question médicale envisagée ;
- La validité du traitement contrôle, placebo ou traitement actif ; la pertinence de la stratégie de traitement par-dessus laquelle le nouveau traitement a été évalué ; le contexte de soins.
- La pertinence clinique du critère de jugement
- L'importance de la taille de l'effet raisonnablement garanti (intervalle de confiance)
- La représentativité des patients inclus dans l'essai
- La balance bénéfice-risque

Pour qu'un résultat puisse prétendre démontrer un réel bénéfice clinique il faut que :

- le résultat soit obtenu sur un critère de jugement clinique et non pas intermédiaire

Un **critère clinique** est un critère qui correspond directement à un objectif thérapeutique recherché par le patient et le médecin (réduire la mortalité, réduire la douleur, réduire la fréquence de survenu d'un événement clinique, d'une complication, etc...). A l'opposé un critère intermédiaire est un simple marqueur de l'activité biologique du traitement (baisse de la pression artérielle, anomalie radiologique, etc..). Les critères intermédiaires sont insuffisants pour assurer la pertinence clinique car il existe de nombreux exemples où une action pharmacologique ne s'est pas traduite en bénéfice clinique lors de la réalisation des essais sur critère clinique.

- Le comparateur soit pertinent .

Un résultat obtenu versus placebo n'a de sens médical que si aucun traitement n'a démontré son bénéfice auparavant. Dans ce cas, il est licite pour le nouveau traitement de se positionner par rapport au placebo afin d'établir son bénéfice dans l'absolu. Par contre, si un traitement à déjà démontré son intérêt dans cette pathologie, la question médicale qui se pose alors, n'est plus celle de l'efficacité du nouveau traitement dans l'absolu mais bien celle de savoir si le nouveau traitement représente un progrès thérapeutique par rapport au traitement précédent. Dans ce cas, seule une comparaison au traitement de référence est pertinente. Dans le cas d'un traitement actif, celui-ci doit être utilisée de manière optimale afin de déployer toutes son efficacité.

- Le contexte de soin et la stratégie de base soit pertinent .

Le nouveau traitement a été évalué par-dessus une stratégie de base adaptée comportant tous les traitements validés du domaine. Les patients n'ont pas été sous traités. Le contexte de soins est adapté et contemporain (Par exemple évaluation d'un anti-agrégant plaquettaire dans l'angor instable dans un contexte de soin où le recours à la PCI est faible. Ce contexte de soins n'est pas représentatif de la pratique actuel où un nombre substantiel de patients sont traités de manière invasive).

- La situation pathologie abordée dans l'essai doit être réelle et correspondre à un réel problème médical non encore résolu .

Il apparaît parfois que l'entité nosologique à laquelle s'intéresse l'essai est de définition récente. Il convient alors d'écarter une situation de " [disease mongering](#) " où une entité nosologique a été soit créer de toutes pièces soit artificiellement mise au premier plan en exagérant son importance, sa gravité ou son retentissement.

- La taille de l'effet observé soit suffisamment importante pour avoir un sens clinique .

En effet, un essai peut montrer de manière statistiquement significative un effet non nul du traitement mais celui-ci peut être de taille insuffisamment importante pour représenter une amélioration sensible de la condition des patients. Par exemple, dans l'incontinence urinaire de la femme, un traitement faisant passer de 12 à 9 le nombre moyen de fuites hebdomadaires ne change en rien la vie de ces patientes : en moyenne, sur une semaine elles auront encore au moins une fuite par jour. Le traitement ne leur permettra pas de se dispenser de porter des protections et il y aura en moyenne chaque jour une protection souillée. Un tel effet, même s'il est réel, ne présente aucun intérêt clinique.

- Le bénéfice mis en évidence soit raisonnablement extrapolable aux patients qui posent se problème dans la réalité du quotidien des médecins .

Les patients inclus dans l'essai doivent avoir été inclus à l'aide de critère conforme à la définition actuelle de la maladie considérée. Aucune hyper sélection arbitraire n'a été effectuée/

- La balance en le bénéfice apportée et les risques encourus soit favorable .

Ce point fait l'objet d'une évaluation à part entière à l'aide du bénéfice clinique net.

11.1 Situation assurant la pertinence clinique du résultat

Le résultat est cliniquement pertinent assurant que le traitement apporte un bénéfice clinique tangible pour les patients et représentant un progrès thérapeutique (cf. fiche concept sur la [pertinence clinique](#)).

Cette pertinence clinique du résultat est assurée par :

- La pertinence du critère de jugement : le critère est un critère clinique et non pas un critère intermédiaire. S'il s'agit d'un critère de substitution (surrogate), celui-ci est parfaitement bien validé et il est possible de garantir que le bénéfice mesuré sur le surrogate implique un bénéfice clinique
- La pertinence du traitement comparateur, placebo ou traitement de référence. S'il s'agit d'un traitement actif, celui-ci est utilisé de manière optimale afin de déboucher sur une comparaison loyale ne favorisant pas artificiellement le nouveau traitement
- La pertinence du contexte de soins qui a été comparable au contexte de soins actuel de ces patients. Ainsi le nouveau traitement n'a pas été évalué dans un contexte où les patients étaient sous-traités et où il était donc facile pour lui de montrer sa supériorité.

- La pertinence de la taille du bénéfice. Ce bénéfice est connu avec suffisamment de précision pour s'assurer que même dans la situation du pire (borne péjorative de l'intervalle de confiance) le bénéfice apportée est encore suffisamment important
- L'extrapolabilité du résultat aux patients qui seront ciblés dans la pratique médicale réelle. Cette extrapolabilité est raisonnablement garanti par la représentativité de patients inclus (pas d'hyper sélection de ces patients, critères de sélection large et absence de critère d'exclusion inutile) et par l'absence raisonnable de facteur modifiant substantiellement l'effet du traitement (recherche d'interaction dans les analyses en sous-groupe).

11.2 Situation de manque de pertinence clinique

Le résultat manque de pertinence clinique et ne démontre donc pas que le traitement apporte un réel bénéfice clinique aux patients (cf. section 4.3).

Ce manque de pertinence peut provenir :

- D'un manque de pertinence du critère de jugement : critère intermédiaire et non pas clinique
- D'un manque de pertinence au niveau du traitement comparateur : utilisation du placebo alors qu'il existe (au moins) un traitement ayant démontré qu'il apportait un bénéfice ; utilisation d'un traitement actif qui n'a pas démontré son efficacité (donnant ainsi une comparaison ininterprétable) ; utilisation non optimal du traitement de référence favorisant ainsi indument le groupe expérimental et conduisant à une comparaison déloyale pour le groupe contrôle.
- D'une taille d'effet insuffisante ne garantissant pas un changement sensible de la condition des patients ou d'une précision insuffisante dans l'estimation de la taille de l'effet
- D'un doute sur l'extrapolabilité du résultat aux patients qui posent se problème dans la réalité du quotidien des médecins (sélection excessive des patients dans l'essai, critère d'exclusion non justifié par le mécanisme physiopathologique du traitement ou par la sécurité, mise en évidence de facteurs modifiant l'effet du traitement de façon cliniquement pertinente dans les analyses en sous-groupes par exemple, etc.)
- balance bénéfice risque est appréciée spécifiquement par le bénéfice clinique net.

Chapitre 12

Bénéfice net

12.1 Notion de bénéfice clinique net

Le bénéfice net est l'outil nécessaire à l'évaluation de la safety. Il permet de faire une confrontation de l'efficacité avec la sécurité et de déterminer si, in fine, le traitement apporte plus d'avantages que d'inconvénients.

Note : L'impossible évaluation de la sécurité des traitements

Sans passer par le calcul du bénéfice net, la gestion de la sécurité est impossible pour de nombreuses raisons.

Il est impossible de *démontrer l'absence* d'effets indésirables. En effet, ce n'est pas parce qu'une différence non statistiquement significative a été obtenue sur un critère de sécurité, qu'il est possible de conclure à l'absence d'un surcroît de risque sur ce critère. L'absence de signification statistique n'implique pas l'absence de différence. Ainsi, malgré ce qui couramment fait, il est impossible de conclure à l'absence d'effet indésirable à l'issue d'un essai où aucun critère de jugement de sécurité ne sort significatif.

Cette problématique peut s'illustrer à l'aide de l'intervalle de confiance. En cas de différence non significative, la borne supérieure de l'intervalle de confiance du risque relatif est supérieure à 1, compatible avec la possibilité d'une augmentation de fréquence de l'événement indésirable considéré. Ne pouvant exclure raisonnablement un surcroît de risque, il n'est pas possible de conclure à l'absence de sur-risque. Par exemple, un risque ratio de 0.99 (IC à 95% entre 0.50 et 1.89) sur des hémorragies majeures avec un anticoagulant ne permet pas vraiment de conclure à l'absence de sur-risque hémorragique car, avec la précision de cette estimation, il n'est pas possible de raisonnablement exclure une augmentation de 89% de la fréquence des hémorragies. Une telle possibilité ne correspond pas vraiment, stricto sensu, à une absence de surrisque.

Il est aussi possible qu'un essai montre un effet délétère de manière statistiquement significatif sans que cela remette en cause l'intérêt du traitement si cet effet indésirable est rare, de gravité comparable au bénéfice recherché et ne contre-balance pas quantitativement le bénéfice obtenu. Ainsi, un résultat statistiquement significatif de défaut de sécurité n'est pas non plus un argument pour récuser automatiquement un traitement. Tout dépend de la balance numérique entre ces 2 composantes de la balance bénéfice risque.

En fait tout dépend de la balance entre le bénéfice et les risques et qui n'est appréciable qu'en passant par un calcul ou une estimation directe du bénéfice clinique net.

12.1.1 Bénéfice clinique net quantitatif

Le *bénéfice net* quantifie la balance entre le nombre d'événements évités par le traitement (la cible du traitement) et le nombre d'événements indésirables (de même gravité ou de gravité plus importante) induits par le traitement. La quantification s'effectue pour 100 ou 1000 patients traités en fonction des niveaux de risque.

Numériquement le bénéfice net s'exprime sous la forme d'une **différence de risque (DR)** obtenue par la somme de la différence de risque du critère d'efficacité et de la différence de risque du critère de sécurité. Il est aussi possible de mesurer directement ce bénéfice clinique net dans un essai en utilisant un critère composite¹ regroupant les événements que l'on cherche à prévenir et les événements indésirables de même importance clinique (comme par exemple dans l'essai RE-LY).

Exemple :

un fibrinolytique a été évalué versus placebo à la phase aigüe de l'infarctus du myocarde. La mortalité à 30 jours (critère d'efficacité) dans le groupe fibrinolytique est de 5%, versus 10% dans le groupe contrôle; soit une différence de risque de $DR1 = 5\% - 10\% = -5\%$. Le traitement de 100 patients par le fibrinolytique évite 5 décès à 30 jours.

Par ailleurs, le fibrinolytique augmente la fréquence des AVC invalidants (par AVC hémorragiques) à 2% versus 1%. Les AVC invalidants sont de gravité comparable aux infarctus mortels. S'ils surviennent en même nombre que les décès évités, ils contrebalanceraient complètement le bénéfice et rendraient le fibrinolytique sans intérêt thérapeutique. La différence de risque sur les AVC invalidant est $DR2 = 2\% - 1\% = +1\%$. Le traitement de 100 patients induit 1 AVC invalidant.

Le bénéfice net est $DR1 + DR2 = -5\% + (+1\%) = -4\%$. Au total le bénéfice net est favorable au fibrinolytique. Les AVC invalidants induits ne contrebalance pas le bénéfice obtenu sur la mortalité et au total, il y a une réduction net de 4% des décès après déduction des AVC invalidants induits.

Bien sûr ce calcul est une approximation car il ne prend pas en compte la co-survenue chez un même patient des 2 types d'événements. Seul l'estimation du bénéfice absolu par un critère composite le permet. Mais cependant, l'approximation effectuée ici est conservatrice et ne peut pas surestimer le bénéfice net.

Il est indispensable de prendre en compte les incertitudes statistiques existantes sur ces estimations et de calculer l'intervalle de confiance du bénéfice net. Par exemple, pour un bénéfice conduisant à une $DR = -1,66\%$ IC 95% (-2,52%; -0,79%), et pour un effet délétère de $DR = 0,47\%$ IC 95% (0,05%; 0,89%), le bénéfice clinique net débouche sur une $DR = -1,19\%$ avec un IC95% entre -2,15% et -0,23%.

En fonction de la signification statistique du bénéfice net plusieurs cas de figure se présentent (voir les sections suivantes) :

- Bénéfice net favorable et statistiquement significatif ($DR < 0$, significatif)
- Bénéfice net favorable mais non significatif ($DR < 0$, NS)
- Bénéfice net défavorable ($DR > 0$, NS ou significatif)
- Intervalle de confiance très large (aucune conclusion possible)

1. cette approche est la seule réalisable quand l'analyse repose sur les courbes de survie

TABLE 12.1 –

Bénéfice net favorable démontré	DR<0, significatif	Ce bénéfice n'est pas susceptible d'être contre balancé quantitativement par des effets indésirables de même gravité
Bénéfice net favorable mais non significatif	DR<0, NS	Il n'est pas exclu que ce bénéfice puisse être contre balancé quantitativement par des effets indésirables de même gravité rendant le bénéfice net nul ou défavorable
Bénéfice net défavorable	DR>0, NS ou significatif	
bénéfice net défavorable qualitativement		
Défaut de safety sans aucun bénéfice en regard		

Une difficulté de cette approche est de mettre de part et d'autre de la balance bénéfice risque, des événements de gravité clinique comparable. Il faut qu'en termes d'embêtement pour le patient, d'altération de la fonction, de mise en jeu du pronostic vital, les événements pris pour mesurer le bénéfice et le risque soient comparables. En général, le choix du critère pour le bénéfice est imposé par le design de l'étude sous la forme du critère de jugement principal. Il est alors nécessaire d'identifier parmi les événements indésirables rapportés ceux qui sont, au moins, de même gravité et qui pourraient contre-balancer le bénéfice.

12.1.2 Bénéfice net qualitatif

Indépendamment de la problématique de l'équilibre numérique, des effets indésirables peuvent aussi remettre en question l'intérêt thérapeutique d'un traitement s'ils sont de gravité bien supérieure au bénéfice apporté. C'est le cas, par exemple, avec des effets indésirables potentiellement mortels pour un traitement antalgique mineur utilisé en petite traumatologie. Le risque est alors *qualitativement* disproportionné par rapport au bénéfice. Aucune quantification n'est nécessaire et c'est simplement un jugement de valeur qualitatif qui amène à la décision.

12.2 Le bénéfice n'est pas susceptible d'être contre balancé quantitativement par des effets indésirables de même gravité

Le bénéfice clinique net est favorable et statistiquement significatif (DR<0, p<0.05). Il est donc démontré que les événements indésirables pris en considération dans le calcul du bénéfice net ne sont susceptibles de remettre en cause ce bénéfice (Figure 2, situation 1). Les résultats de cet essai démontrent que, quantitativement, la balance bénéfice risque est favorable. La différence de risque (DR) calculée pour le bénéfice clinique net est favorable (DR<0) et statistiquement significatif. Il est donc statistiquement démontré que les événements indésirables considérés ne remettent pas en cause quantitativement le bénéfice.

Par exemple, avec les données du tableau suivant, le calcul du bénéfice clinique net donne une DR de -4.4%, IC à 95% entre -8.4% et -0.4%.

TABLE 12.2 –

Critère	groupe traité	groupe contrôle	différence des risques (IC à 95%)
critère de bénéfice	230/1201	287/1207	-4.6% [-7.9%;-1.4%]
critère de sécurité	113/1201	111/1207	0.2% [-2.1%;2.5%]
bénéfice clinique net			-4.4% [-8.4%;-0.4%]

Il faut cependant ne pas trop personnaliser ce bénéfice clinique et veiller à ne pas faire de sur-interprétation. Ainsi il est excessif de dire qu'en cas de bénéfice clinique net favorable, le nombre d'événements évités est supérieur au nombre d'événements indésirables induit et le traitement bénéficie à un nombre de patients supérieur à celui de ceux auxquels il nuit. En effet il est impossible d'identifier ces patients (qui ont bénéficié ou pâti du traitement).

Les nombres observés ne sont que l'expression de probabilité au niveau d'un groupe. Ainsi si on obtient un bénéfice se traduisant par une différence de risque de -5%, cela signifie que un groupe de 100 patients on aura, en moyenne, 5 évènements de moins sous traitement que sans traitement, mais cela ne veut pas dire que seulement 5 sujets bénéficient du traitement. Ces 5 sujets sont simplement la traduction sur un groupe d'une probabilité de faire l'évènement passant par exemple de 20% à 15%.

Ce type de résultats ne permet pas de dire que dans le groupe initial il y avait 5% de patients répondant au traitement et que ceux qui feront quand même l'évènement même sous traitement sont des non-répondeurs.

En fait une hypothèse de ce type (existence de répondeurs/non-répondeurs) conduit exactement au même résultat qu'une situation où tous les patients bénéficient du traitement avec un risque diminué pour tous.

Les résultats d'un essai ne permettent pas de discerner ces 2 hypothèses (qui conduisent aux mêmes résultats en fréquence). Le premier modèle (existence de répondeurs) ne peut être validé qu'en montrant qu'il est possible d'identifier a priori, et de manière reproductible, les sujets à coup sûr répondeurs des non-répondeurs. En l'absence d'une telle démonstration, les principes scientifiques, principes de parcimonie ou du rasoir d'Occam, impose de retenir que le modèle le plus général, c'est-à-dire le moins spéculatif (faisant le moins de sous hypothèses), c'est-à-dire celui du bénéfice pour tous.

Ainsi, a priori, lorsque l'on initie un traitement chez un patient, il aura une probabilité plus forte d'en bénéficier que d'avoir un EI, il a donc, au total, une probabilité positive d'en tirer un bénéfice. Pour une discussion de cette transposition des résultats moyens d'un essai à l'échelon individuel cf. la fiche concept ???.

12.3 Il n'est pas possible d'exclure que ce bénéfice puisse être contre-balancé quantitativement par des effets indésirables de même gravité rendant le bénéfice net nul ou défavorable

Il existe une tendance non significative à un bénéfice net favorable (DR<0 NS)

Ce résultat est démontré et cliniquement pertinent, mais le calcul du bénéfice net ne permet pas de raisonnablement exclure que les événements indésirables puissent complètement contre-balancer le bénéfice (Figure 2, situation 2).

En effet, le calcul du bénéfice net ne permet pas d'exclure, raisonnablement avec un degré de certitude de 95%, que le nombre d'événements indésirables induit soit supérieur au nombre d'événements cibles évités. La borne supérieure de l'intervalle de confiance de

la différence de risque du bénéfice clinique net est supérieure à zéro, compatible avec un bénéfice net défavorable (plus d'EI engendrées que d'événements évités).

Par exemple, avec les données du tableau suivant, le calcul du bénéfice clinique net donne une DR de -3.0%, IC entre -7.0% et 1.0%.

TABLE 12.3 –

Critère	groupe traité	groupe contrôle	différence des risques (IC à 95%)
critère de bénéfice	241/1201	287/1207	-3.7% [-7.0%;-.4%]
critère de sécurité	107/1201	99/1207	0.7% [-1.5%;2.9%]
bénéfice clinique net			-3.0% [-7.0%;1.0%]

Cette situation provient du fait que l'incertitude statistique sur la fréquence des événements indésirables induits est très importante, ce qui se rend à son tour, incertaine l'estimation du bénéfice net. Compte tenu de cette incertitude et malgré le bénéfice démontré, l'intervalle de confiance de l'estimation du bénéfice net est très large et la borne supérieure devient positive. Il n'est pas possible de conclure, avec suffisamment de certitude, à une balance bénéfice risque acceptable. Les données ne permettent pas de démontrer que le bénéfice est quantitativement supérieur au risque.

12.4 Bénéfice net défavorable, suggéré ou démontré (DR>0, NS ou p<0.05)

Le calcul du bénéfice clinique net débouche sur une différence de risque défavorable (DR>0), en faveur d'un surcroît possible d'événements indésirables supérieur quantitativement aux événements évités par le traitement. La différence de risque du bénéfice net est défavorable de manière statistiquement significative ou non (Figure 2, situation 3 et 4). En raison de l'application du principe de précaution en sécurité, ce résultat n'a pas besoin d'être statistiquement significatif pour être considéré comme devant être pris en compte au niveau décisionnel.

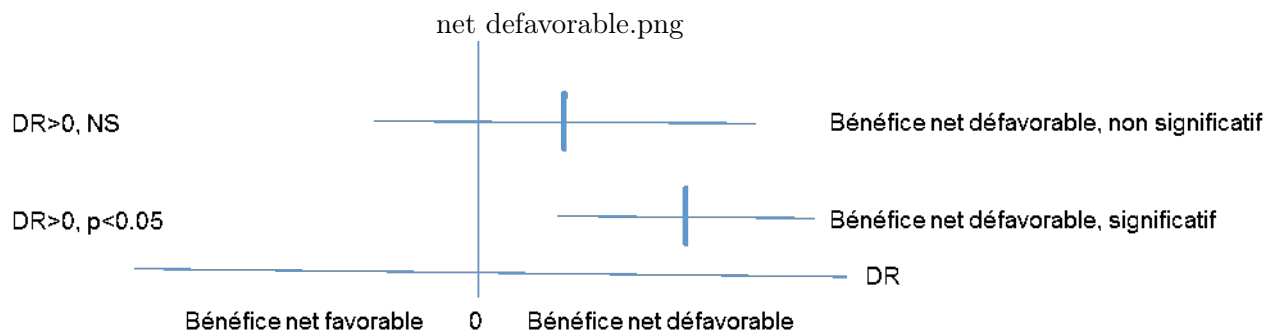


FIGURE 12.1 – Figure 1 -

Par exemple, avec les données du tableau suivant le calcul du bénéfice clinique net donne une DR = +2.9%, IC 95% entre -0.9% et 6.7%.

TABLE 12.4 –

Critère	groupe traité	groupe contrôle	différence des risques (IC à 95%)
critère de bénéfice	79/1201	106/1207	-2.2% [-4.3%;-0.1%]
critère de sécurité	260/1201	200/1207	5.1% [1.9%;8.2%]
bénéfice clinique net			2.9% [-0.9%;6.7%]

Il faut cependant remarquer qu'une différence des risques positive avec un intervalle de confiance très large est non informative et peut ne pas correspondre à un réel bénéfice net défavorable. Cependant la largeur de l'intervalle de confiance montre que l'on ne connaît pas avec suffisamment de précision la *sécurité* pour s'assurer que le bénéfice clinique net est acceptable. Il est donc inutile de différencier ce cas des autres recouvert par cette situation, car en pratique la conclusion reste la même : le bénéfice observée ne peut pas être accepté comme une preuve de l'intérêt thérapeutique du traitement.

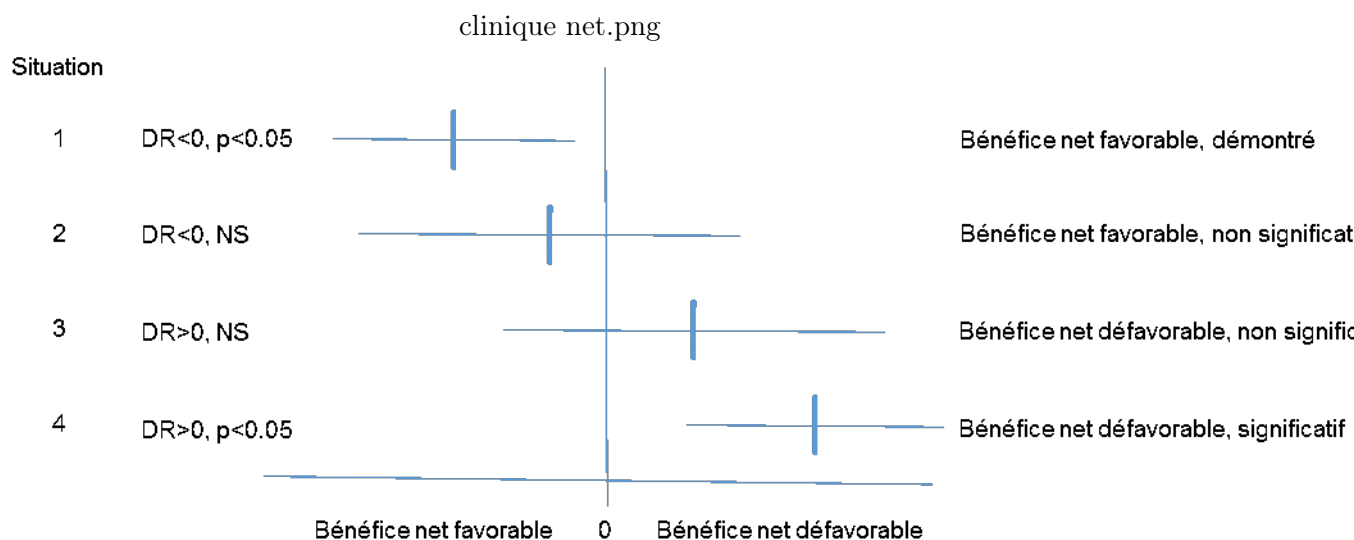


FIGURE 12.2 – Figure 2 - Illustration graphique des différentes situations concernant le bénéfice clinique net

12.5 Bénéfice net défavorable qualitativement

Ce bénéfice démontré ne se traduit pas en un bénéfice clinique net pour le patients en raison du risque encouru d'un (ou de plusieurs) évènements indésirables dont la gravité est disproportionnée par rapport au bénéfice apporté.

Le bénéfice apporté par le traitement est donc contrebalancé qualitativement par ces évènements indésirables. Ces évènements indésirables contre-balançant le bénéfice peuvent avoir été montrés par cet essai ou une autre étude (par exemple observationnelle) ou simplement observé dans le cadre de la pharmacovigilance.

12.6 Ce résultat a été obtenu sur un critère qui est déjà un critère de bénéfice net

Ce résultat a été obtenu sur un critère qui intègre par essence le bénéfice et le risque (critère composite de bénéfice clinique net ou mortalité totale). Il n'est donc pas nécessaire de faire un calcul de bénéfice clinique net, ce résultat en est une estimation directe. Un résultat statistiquement significatif sur ce critère apporte la démonstration d'un bénéfice non contrebalancé quantitativement par des effets indésirables de gravité comparable.

Les décès de toutes causes est un critère de ce type. A partir du moment où il est démontré que le nouveau traitement apporte une réduction de mortalité globale, il ne devient plus nécessaire de s'interroger sur des effets indésirables de même gravité, étant donné que ces événements indésirables sont intégrés directement dans ce critère. Il convient cependant s'il existe un possible surcroît de décès par événements indésirables de prendre en compte cette information, si par exemple il existe un autre traitement d'efficacité similaire avec un meilleur profil de sécurité.

Chapitre 13

Démarche hypothetico déductive

- 13.1 Situations où la nature hypothetico déductive du résultat est acceptable
- 13.2 Situations où la nature hypothetico déductive du résultat n'est pas acceptable

Chapitre 14

Cohérence externe

La cohérence externe d'un résultat est assurée quand celui-ci est en phase avec les autres résultats disponibles obtenus sur études similaires. Dans ce cas, la robustesse du résultat est renforcée et il est alors possible d'exclure que le résultat soit simplement un artefact non reproductible.

En effet un résultat unique, sur un concept inédit, fait toujours appréhender qu'il puisse s'agir d'un artefact. En effet, il y existe toujours la possibilité d'un biais résiduel non détecté ou d'un résultat faussement positif dû au hasard. Classiquement, dans la démarche expérimentale, il est demandé une réplication des résultats avant leur acceptation définitive.

Au niveau statistique, l'existence d'une confirmation lors d'une réplication permet aussi de réduire le risque alpha. En effet, au niveau d'un essai clinique, le risque alpha est habituellement contrôlé à moins de 2.5% en unilatéral. Ce niveau de risque, bien que habituellement accepté, reste somme toute très élevé pour décider de l'adoption définitive d'un nouveau traitement qui va concerner de très nombreux patients dans le futur. De plus compte tenu du nombre de traitement en évaluation (plus de 10000) et du fait qu'environ la moitié de ces traitements ne confirme pas leur intérêt, l'utilisation d'un risque alpha à 2.5% conduit à un nombre important de traitements adoptés à tort. La disponibilité d'au moins deux essais indépendants statistiquement significatif permet de réduire le risque alpha à un niveau beaucoup plus satisfaisant (0.0625%). Ainsi dans de nombreux domaines, les agences d'enregistrement attendent (sans pour autant formellement l'exiger) deux essais pivots concluants.

Dans des situations où un seul essai n'est envisagé, pour garantir la robustesse du résultat unique qui sera produit, un abaissement du seuil de la signification statistique prévu au protocole a été adopté (cf. HPS???).

Cette approche est intéressante car il s'avère difficile en pratique de faire la réplication d'un essai, particulièrement quand celui-ci concerne montre une réduction de mortalité ou d'événements cliniques graves. Dans ce cas, l'acceptabilité d'un nouvel essai conduisant à ne pas traiter des patients est faible. La faisabilité des réplifications dans ces domaines est en général compromise (souvent dès la soumission à un comité d'éthique). Indirectement, cette constatation montre que l'importance apportée à ce principe méthodologique par la communauté médicale est faible.

En général, le deuxième essai est réalisée dans une population de patients quelque peu différente. Même s'il ne s'agit alors pas d'une stricte réplication, l'obtention d'un résultat concluant dans cette deuxième étude contribue quand même à renforcer la cohérence externe du premier résultat en montrant une reproductibilité à travers différentes population de patients.

Par contre, dans tous les domaines l'enjeu du traitement ne concerne pas la mortalité

ou les événements clinique sérieux, la réalisation de 2 essais similaire ou quasi-similaire indépendant est faisable.

14.1 Situation assurant la cohérence externe du résultat

14.2 Situation d'absence de cohérence externe du résultat

14.2.1 L'appréciation de la cohérence externe impossible

L'appréciation de la cohérence externe du résultat considéré n'est pas possible, car il n'existe aucun n'autre essai similaire. Il n'existe pas non plus d'essais du même traitement dans d'autres situations cliniques, ni d'essais d'autres molécules de la même classe pouvant permettre une appréciation partielle indirecte de sa cohérence externe.

En l'absence de réplication du résultat il n'est pas possible d'exclure un résultat artefactuel ou un biais de publication.

14.2.2 Remise en cause de la cohérence externe du résultat

Un résultat peut être remis en cause pour manque de cohérence externe quand les autres essais similaires existants ne le confirment pas. La non-confirmation se juge en terme de taille de l'effet et non pas en terme de signification statistique. En effet, des discordances de signification statistiques entre essais peuvent simplement provenir des différences de puissance des études sans que l'effet du traitement en varie d'un essai à l'autre. Il y a non confirmation quand les résultats quantitatifs des essais varient de manière importante, avec une hétérogénéité des résultats supérieurs à ce qui est attendu uniquement du fait des fluctuations aléatoires d'échantillonnage.

Ainsi un résultat sera récusé pour manque de cohérence externe quand la méta-analyse intégrant ce résultats et ceux des autres études similaires ne montrent pas d'effet statistiquement significatif et ne confirme donc pas le résultat considéré. Les autres résultats récusent en quelque sorte le résultat considéré. Celui-ci s'avère, ainsi, être particulièrement favorable à l'effet du traitement, de manière non entièrement reproductible. Au total, il n'est donc pas possible l'existence d'un effet du traitement puisse qu'au total, en considérant la totalité de l'information apportée par tous les essais faisant la même évaluation, il n'est pas possible de conclure à cet effet.

Dans ces circonstances, le résultat analysés s'avère donc être très vraisemblablement un résultat faussement positif qui peut être dû à un biais non identifié de l'essai ou à des facteurs modificateurs de l'effet (une hétérogénéité existe alors) ou dû simplement aux hasard à travers les fluctuations aléatoires d'échantillonnages (il n'est pas mis en évidence d'hétérogénéité puisque les différences entre les résultats des essais proviennent uniquement du hasard).

Exemple :

L'étude CHAOS fourni un résultat en faveur du bénéfice de la vitamine E en prévention cardiovasculaire en montrant une réduction de la fréquence des décès cardiovasculaires par rapport au placebo de ??? réf. Cependant ce résultat s'avère unique, car il n'est pas retrouvé dans les ??? autres essais ayant le même objectif, qui sont tous non concluants.

La méta-analyse regroupant tous ces résultats ne montre pas de différence statistiquement significative et aucune hétérogénéité importante n'est détectée. Il est ainsi possible de conclure que le résultat positif de CHAOS est très vraisemblablement un faux positif (dû hasard ou à un biais non détecté¹). Cela conduit à rejeter le résultat de CHAOS pour manque de cohérence externe et à considérer que la vitamine E n'a pas apporté la preuve de son intérêt sur les décès cardiovasculaires.

14.3 références

[\]\(http://blog.scielo.org/en/2014/02/27/reproducibility-of-research-results-the-tip-of-the-iceberg/\)](http://blog.scielo.org/en/2014/02/27/reproducibility-of-research-results-the-tip-of-the-iceberg/) [

Chapitre 15

Les différents types de résultats

Un essai thérapeutique produit un grand nombre de résultats : un pour chaque critère de jugement (principaux, secondaires ou de sécurité). Ce nombre peut être augmenté par l'existence de sous groupe, d'analyses en ITT et en per protocole, etc. Tous ces résultats n'ont pas la même validité et force de conviction. A partir des 4 éléments faisant la validité d'un résultat :

- la possibilité que le résultat soit un faux positif dû au hasard,
- la possibilité que le résultat soit un faux positif dû à un biais,
- sa pertinence clinique,
- sa robustesse vis-à-vis des effets indésirables (bénéfice clinique net),

il est ainsi possible de distinguer plusieurs catégories de résultats (pour la définition précise de l'une de ces entités voir la section suivante correspondante) :

1. Résultat allant dans le sens d'un effet bénéfique du traitement (supériorité versus traitement actif ou effet bénéfique versus placebo)
 - (a) Bénéfice démontré et cliniquement pertinent, bénéfice net favorable démontré [1]
 - (b) Bénéfice démontré et cliniquement pertinent, bénéfice net favorable NON démontré ($DR < 0$, NS) [12]
 - (c) Bénéfice démontré mais bénéfice net DÉFAVORABLE objectivé ou suspecté ($DR > 0$, NS ou $p < 0.005$) [11]
 - (d) Bénéfice démontré mais NON cliniquement pertinent [2]
 - (e) Bénéfice suggéré, cliniquement pertinent [3] (qq soit le bénéfice net, si bénéfice net non favorable mettre en avant l'événement indésirable catégorie 8)
 - (f) Bénéfice suggéré, NON cliniquement pertinent [4] (si bénéfice net non favorable -> mettre en avant l'événement indésirable catégorie 8)
 - (g) Résultat non statistiquement significatif [5]
 - (h) Résultat non statistiquement significatif pour un critère de sécurité [21]
 - (i) (Pour mémoire car ce résultat est en fait un résultat allant dans le sens d'un effet délétère, sur un critère d'efficacité, infériorité vs traitement actif ou effet délétère vs placebo, suggéré ou démontré [13])
2. Résultat allant dans le sens d'un effet délétère
 - (a) Effet délétère suspecté ou démontré, sans aucun bénéfice par ailleurs [8]
 - (b) Effet délétère suspecté ou démontré, mais bénéfice par ailleurs
 - i. Effet délétère ne remettant pas en cause le bénéfice [14]

- ii. Effet délétère remettant en cause le bénéfice (quantitativement ou qualitativement) [15]
 - (c) Effet délétère sous estimé empêchant l'appréciation correcte du bénéfice clinique net [16]
 - (d) Supériorité (effet bénéfique, moins d'EI) -> devient un critère de bénéfice (démonstré ou suggéré)
3. Résultat allant dans le sens de l'équivalence de 2 traitements (non infériorité)
 - (a) Non infériorité démontrée, cliniquement pertinente [6] = progrès thérapeutique
 - (b) Non infériorité démontrée (non cliniquement pertinente) [7]
 - (c) Non infériorité seulement suggéré [9]
 4. Résultat ne montrant pas de défaut de tolérance pouvant contrebalancer le bénéfice [17]

15.1 Résultat démontré, cliniquement pertinent, bénéfice net assuré

Il est possible d'écarter que le résultat soit un résultat faussement positif dû au hasard ou dû à un biais. Le résultat est cliniquement pertinent et n'est pas susceptible d'être contre balancé, quantitativement ou qualitativement, par des effets indésirables de gravité au moins identique au bénéfice apporté. Ce résultat, démontré et cliniquement pertinent, fournit ainsi une preuve de l'intérêt thérapeutique de ce traitement, qui peut être à la base de son utilisation systématique en pratique.

TABLE 15.1 – Table 1

faux positif dû au hasard	faux positif dû à un biais	pertinence clinique	bénéfice clinique net	intérêt thérapeutique
exclu	exclu	assurée	assuré favorable	OK

Ce résultat est issu d'une méthodologie et d'une démarche statistique garantissant correctement que le résultat ne soit pas un faux positif dû à un biais ou au hasard.

Le résultat est à l'abri des biais :

- biais de sélection en raison d'une randomisation imprévisible,
- biais de suivi et de mesure en raison d'une réalisation en double insu
- biais de mesure en raison aussi de la réalisation de l'essai en double insu
- biais d'attrition en raison d'une analyse en ITT avec remplacement des données manquantes.

Au niveau statistique cela signifie que le résultat est statistiquement significatif (au moins au seuil de 5% bilatéral) et qu'il n'est pas issu d'une démarche statistique engendrant multiplicité et inflation du risque alpha (multiplicité des critères de jugement, analyse de sous groupe, analyse intermédiaire non protégées, etc.) Le résultat est cliniquement pertinent (aussi bien en termes de comparateur, de critère de jugement, de taille d'effet et de population ciblée). Il démontre donc que le traitement apporte un bénéfice cliniquement pertinent (et représentant une plus value médicale avérée). Le calcul du bénéfice net montre que les événements indésirables de gravité identique au bénéfice apportée ne contre balancent pas numériquement le bénéfice. De même, il n'a pas identifié d'effets indésirables de gravité bien supérieure au bénéfice apportée qui invalideraient de façon qualitative la balance bénéfice risque.

15.2 Bénéfice cliniquement pertinent démontré mais la balance bénéfice risque n'est pas assurée

Un bénéfice cliniquement pertinent est démontré mais il n'est pas possible d'exclure que les événements indésirables puissent le contre balancer en totalité quantitativement. Un effet délétère est connu, en général objectivé dans l'essai lui-même, mais le bénéfice clinique net n'est pas assuré car il est non statistiquement significatif. L'estimation (calculée ou directement mesurée par un critère composite) de la différence de risque correspondant au bénéfice clinique net est favorable ($DR < 0$) mais elle est non statistiquement significative ($DR < 0$, NS). Il est donc impossible de raisonnablement exclure que les événements indésirables puissent contre balancer en totalité le bénéfice obtenu. En effet, le calcul du bénéfice net ne permet pas d'exclure que, raisonnablement avec un degré de certitude de 95%, le nombre d'événements indésirables induit soit supérieur au nombre d'événements cibles évités. In fine, il n'est pas possible de conclure à un bénéfice clinique net favorable.

Un résultat de ce type est insuffisant pour conclure à l'intérêt thérapeutique du traitement en raison de l'incertitude existant sur la balance bénéfice risque.

TABLE 15.2 –

faux positif dû au hasard	exclu
faux positif dû à un biais	exclu
pertinence clinique	assurée
bénéfice clinique net	non assuré
intérêt thérapeutique	non

Ce raisonnement se base sur le calcul du bénéfice net absolu qui s'obtient en soustrayant de la différence de risque (DR, absolute risk reduction ARR) des événements cibles évités la différence des risques des événements indésirables induits. Bénéfice net, différence des risques (DR) = DR(événements cible évités) – DR(événements indésirables) Dans ce cas la borne supérieure de l'intervalle de confiance de ce bénéfice net absolu est inférieure à zéro montrant, que compte tenu de la certitude statistique apportée par les données disponibles, il n'est pas possible d'exclure que le nombre d'événements indésirable ne soit pas supérieur au nombre d'événement cible évité. Cette situation peut avoir deux causes :

- Le nombre d'événement indésirable induit est effectivement supérieur au nombre d'événements évités
- L'incertitude sur le nombre d'événements indésirables induits est très importante ce qui rend à son tour incertaine l'estimation du bénéfice net. L'essai est insuffisamment puissant pour estimer correctement la sécurité du traitement compte tenu de l'importance relative des événements indésirables par rapport au bénéfice.

En raison de cette incertitude, l'intervalle de confiance de l'estimation du bénéfice net est très large et la borne supérieure devient positive. Compte tenu de cette imprécision, il n'est pas possible de conclure, avec suffisamment de certitude, à une balance bénéfice risque acceptable. Les données ne permettent pas de démontrer que le bénéfice est quantitativement supérieur au risque.

15.3 Bénéfice clinique net défavorable (alerte de safety, $DR > 0$ NS ou $p < 0.05$)

Le résultat est en faveur d'un bénéfice (sur le critère considéré) mais celui-ci est compensé en totalité, quantitativement ou qualitativement, par des événements indésirables

de même gravité ou de gravité plus importante. Si l'excès d'événements indésirables est qualitatif, le bénéfice net est défavorable (statistiquement significatif ou non). Il n'est pas en faveur d'une réduction nette d'événements. Si cet excès est qualitatif, il existe des événements indésirables de gravité clinique supérieure au bénéfice apporté par le traitement.

TABLE 15.3 – Table 3

faux positif dû au hasard exclu	faux positif dû à un biais exclu	pertinence clinique assurée	bénéfice clinique net défavorable quantitativement ($DR > 0$, NS ou $p < 0.05$) ou qualitativement	intérêt thé- rapeutique non
---------------------------------------	--	-----------------------------------	---	-----------------------------------

Cela signifie que le traitement présente un problème de sécurité et que, même s'il apporte un bénéfice réel et pertinent, celui-ci est insuffisant, quantitativement ou qualitativement, vis-à-vis des événements indésirables.

Cette situation peut provenir :

- Soit de la survenue d'événements indésirables de gravité supérieure au bénéfice apportée. Le risque que fait courir le traitement est disproportionné par rapport au bénéfice qu'il apporte. C'est par exemple le cas, si l'événement indésirable est une hépatite cytolytique pour un antalgique utilisée en petite traumatologie (bénéfice démontré sur une échelle analogique visuelle de douleur).
- Soit de la survenue d'événements indésirables de gravité comparable au bénéfice apporté avec une telle fréquence qu'ils contre balancement quantitativement le bénéfice apporté. Au total le bénéfice clinique net est négatif. Par exemple, un fibrinolytique a la phase aiguë de l'infarctus du myocarde qui engendre deux fois plus d'AVC invalidants (par hémorragie) qu'il n'évite de décès.

TABLE 15.4 – Les différents cas de figure de la balance bénéfice risque quantitative possible avec un bénéfice démontré

Bénéfice net	Conclusion	Intérêt thérapeutique
$DR < 0$, Significatif	bénéfice non contre balancé par les événements indésirables	oui
$DR < 0$, NS	il n'est pas possible d'exclure que le bénéfice soit contrebalancé par un surcroît d'événements indésirables	non
$DR > 0$, significatif	Au total le traitement induit plus d'événements indésirables qu'il n'évite d'événements cibles de son action	non
$DR > 0$, NS	il n'est pas possible d'exclure que le bénéfice soit contrebalancé par un surcroît d'événements indésirables	non

Exemple : ???

15.4 Résultat démontré, non cliniquement pertinent

Il est possible d'écarter que le résultat soit un résultat faussement positif dû au hasard ou dû à un biais, mais le résultat n'est pas cliniquement pertinent. Il ne correspond donc pas à un bénéfice intéressant en pratique. Bien que réel, ce résultat n'a pas suffisamment de valeur médicale. Il ne peut donc pas être considéré comme représentant un réel progrès thérapeutique à même de changer les pratiques. Ce type de résultat est insuffisant pour considérer que le traitement a un intérêt thérapeutique avéré.

TABLE 15.5 – Table 5

faux positif dû au hasard	faux positif dû à un biais	pertinence clinique	bénéfice clinique net	intérêt thérapeutique
exclu	exclu	non	sans objet	non

Le manque de pertinence clinique d'un résultat peut provenir de plusieurs causes.

15.5 Résultat démontré mais absence de cohérence externe

Le résultat est formellement démontré mais manque de cohérence externe. Il n'est pas confirmé par les autres études comparables. Ce résultat n'est pas retrouvé dans la méta-analyse regroupant cet essai et les autres essais disponibles abordant la même question thérapeutique. Le résultat de la méta-analyse n'est plus statistiquement significatif et montre un effet traitement moindre. Une hétérogénéité existe entre tous ces résultats suggérant que le résultat de cet étude est réellement discordant par rapport à ceux des autres études. La réplication de ce résultat n'est donc pas assurée. En l'absence d'éléments pour expliquer l'obtention de résultat ne confirment pas l'effet du traitement sur ce critère, il convient de considérer ce dernier comme étant artefactuel.

Ce manque de cohérence de ce résultat avec ceux des autres études lui fait perdre son caractère probant et il ne peut plus représenter une preuve formelle de l'intérêt thérapeutique du traitement.

15.6 Résultat seulement suggéré (cliniquement ou non cliniquement pertinent)

Le résultat est en faveur de l'efficacité du traitement mais il n'est pas possible d'écarter qu'il puisse être un résultat faux positif dû au hasard (erreur statistique) ou dû à un biais. De ce fait, la prise de risque est trop importante dans l'acceptation de ce résultat comme preuve de l'efficacité du traitement.

TABLE 15.6 – Table 6

faux positif dû au hasard	faux positif dû à un biais	pertinence clinique	bénéfice clinique net	intérêt thérapeutique
non exclu sur une ou les 2 catégories	-	sans objet	sans objet	non

De nombreuses circonstances peuvent conduire à cette situation :

- Le résultat a été obtenu dans un cadre de réalisation de multiples tests statistiques : critère de jugement secondaire, sous groupe, analyse intermédiaire non protégée contre l'inflation du risque alpha, mesures répétées au cours du temps, essai multi-bras sans gestion de la multiplicité engendrée, etc.
- Le résultat est post hoc et n'est donc pas issu d'une démarche hypothético-déductive
- L'essai n'est pas à l'abri des biais : randomisation prévisible, absence de double aveugle, présence d'une attrition, analyse non réalisée en intention de traiter, données manquantes non remplacées, etc.

En effet un risque d'erreur alpha non contrôlé ne permet absolument pas de savoir le risque de conclure à tort est acceptable ou pas (puisque que ce risque peut être supérieur

au seuil de la signification statistique fixé). De ce fait il n'est pas possible de conclure à la démonstration du résultat.

De même, l'absence de protection complète contre les biais ne permet pas d'exclure que le résultat puisse être faussement positif en raison d'un biais (visible ou non). N'ayant pas cette garanti, il n'est donc pas possible de prendre le risque de conclure à la démonstration du résultat. Il est seulement suggéré de ce fait.

Ces résultats seulement suggérés et non démontrés sont insuffisants pour donner un argument en faveur de l'intérêt thérapeutique avéré du nouveau traitement. Ils peuvent cependant inciter à réaliser un nouvel essai de confirmation. La réalité de ce type de résultat n'étant pas acquise, il est donc inutile de s'interroger sur leur pertinence clinique et le bénéfice clinique net.

Les éléments justifiant ce verdict sont précisés et justifiés pour les 2 catégories « faux positif dû au hasard » et « faux positif dû aux biais ».

15.7 Résultat seulement suggéré mais possibilité d'upgrading

????

15.8 Résultat non significatif

Le résultat n'est pas statistiquement significatif, il ne peut donc pas servir à justifier un éventuel intérêt du traitement, même s'il va dans le bon sens.

Un résultat non significatif est synonyme d'absence de résultat. Il ne permet pas de conclure à l'effet du traitement évalué et il ne permet pas non plus de conclure à l'absence de cet effet (cf. §concept résultat NS)

Ainsi, il n'est pas possible, sans prise de risque exagérée, d'écarter la possibilité que ce résultat soit faussement positif dû au hasard. De ce fait, ce résultat ne peut pas être considéré comme apportant une quelconque preuve de l'efficacité du traitement.

TABLE 15.7 –

faux positif dû au hasard	non exclu
faux positif dû à un biais	sans objet
pertinence clinique	sans objet
bénéfice clinique net	sans objet
Intérêt thérapeutique	non

Le concept de signification statistique est directement lié à celui de résultat faussement positif dû au hasard. La valeur de p (p -value) est la quantification du risque d'erreur alpha liée au résultat considéré. C'est la probabilité d'obtenir une différence au moins aussi importante que celle observée sans que le traitement soit efficace (hypothèse nulle). Il s'agit donc de la quantification du risque d'erreur de conclure à tort à l'efficacité du traitement (alors que le traitement est en réalité sans effet). Si ce risque est supérieur à un seuil appelé seuil de la signification statistique (et qui est fixé en général à 5% en bilatéral) il est impossible de conclure à l'existence de l'efficacité du traitement car ce risque alpha est trop important. Par contre si la p value est inférieure au seuil, on accepte de conclure car le risque alpha sous-jacent est considéré comme suffisamment petit pour accepter de courir un tel risque.

Note : Note

Attention : $p < 0.05$ ne signifie pas forcément que le résultat est statistiquement significatif. Un seuil de la signification statistique plus petit a pu être choisi a priori dans le protocole de l'étude. Une méthode permettant des comparaisons multiples a pu aussi être utilisée, conduisant à diminuer le seuil à atteindre par les p individuels afin de maintenir l'inflation du risque alpha consécutive au seuil habituel de 5% globalement

[fiche concept la multiplicité des comparaisons](#) En cas de résultat non statistiquement significatif il est impossible de faire une conclusion. Si l'essai compare 2 traitements actifs, il est impossible de conclure à l'équivalence des 2 traitements, car une différence non significative peut provenir soit d'une réelle absence de différence mais aussi d'un manque de puissance, sans qu'il soit possible de faire la part entre les 2 possibilités.

[le risque alpha](#)

15.9 Résultat non significatif pour un critère de sécurité

L'absence de signification statistique ne signifie pas l'absence d'effet indésirable. En effet, un résultat NS peut provenir de 2 causes indistinguables : l'absence d'effet et le manque de puissance.

[l'évaluation de la sécurité](#)

15.10 Infériorité ($p < 0.05$), démontré ou suggéré (sur un critère d'efficacité cliniquement pertinent)

Le résultat montre que le nouveau traitement est inférieur au traitement contrôle (ou au placebo) sur un critère cliniquement pertinent. Ce résultat est statistiquement significatif. Si la comparaison était versus placebo, ce type de résultat correspond à la mise en évidence d'un effet indésirable et remet en cause l'intérêt du traitement. Même si le critère de jugement était initialement un critère d'efficacité, ce résultat révèle un problème de sécurité.

Si le nouveau traitement a été comparé à un traitement actif, ce résultat montre l'infériorité du nouveau traitement qui ne peut pas être considéré alors comme une alternative de première intention au traitement de référence utilisée dans le groupe contrôle. Éventuellement, ce nouveau traitement peut alors être envisagé en alternative à ce traitement de référence en cas d'échec de celui-ci ou d'intolérance, mais après qu'il y ait eu une démonstration de la supériorité de son efficacité par rapport au placebo putatif (cf. fiche concept placebo putatif et essai de non infériorité d'efficacité suffisante??). En effet, ce résultat peut signifier bien plus qu'une relative infériorité par rapport au traitement de référence et être en réalité le reflet d'un effet délétère (qui rend le traitement en fait « inférieur » au placebo). Il convient donc d'écarter cette possibilité avant de conclure que le traitement, bien qu'inférieur au traitement de référence, a tout de même une efficacité suffisante pour représenter une alternative de 2ème intention. Les comparaisons indirectes peuvent être une manière d'explorer cette problématique.

15.11 Résultat montrant un défaut de sécurité, démontré ou simplement suggéré, sans bénéfice en regard

Ce résultat montre un effet délétère (défaut de sécurité ou infériorité d'efficacité), avec la mise en évidence de l'infériorité du nouveau traitement par rapport à son contrôle (traitement actif ou placebo). Aucun bénéfice du traitement n'ayant été démontré par ailleurs sur un critère de même gravité clinique, cet effet délétère ne peut donc pas être justifié par un gain en événements évités par ailleurs. Le calcul d'un bénéfice clinique net est sans objet (absence de bénéfice brut, cf. la discussion du bénéfice net section 4.4.). Il peut ainsi s'agir, par exemple, d'une augmentation des hémorragies mortelles sans que le traitement est démontré qu'il apporté une réduction d'une mortalité de cause spécifique.

Cet effet délétère peut être soit démontré soit simplement suggéré. Dans les 2 cas, le résultat est suffisamment fort pour invalider l'intérêt du résultat. En effet, en cas de suggestion d'effet délétère, la situation n'est alors plus de rechercher des arguments pour l'utilisation du traitement mais plutôt pour sa non-utilisation. La logique du raisonnement s'inverse. Il n'est plus question de minimiser le risque de conclure à tort à l'efficacité (risque alpha ou biais), mais bien de minimiser celui de ne pas conclure à tort à l'existence d'une d'un effet délétère (risque bêta). De ce fait, la recherche d'une démonstration statistique stricte (résultat significatif sans inflation du risque alpha et absence formelle de biais) n'est plus de mise et une simple tendance (non significative ou non totalement à l'abri des biais) autorise d'avancer l'infériorité (le surcroît d'événements indésirables) par précaution. Il faut cependant être conscient que ce principe induit une inflation du risque de résultat faussement négatif (non démonstration ou acceptation du bénéfice du traitement) parfois très importante, comme en safety, et conduit à un risque élevé de conclusion faussement positive en raison des défauts de safety par exemple.

La recherche d'un bénéfice, qui aurait permis de compenser cet effet délétère, a portée sur les critères de gravité similaire à celle de cet effet indésirable. L'essai ne montre aucun bénéfice de ce type. Par exemple, si l'effet délétère est observé sur un critère de mortalité, comme des décès toxiques en cancérologie ou des hémorragies mortelles en anti-coagulation. Il a été recherché un bénéfice sur la mortalité totale ou sur une mortalité de cause spécifique. Si aucune réduction de mortalité totale ou de cause spécifique n'est observée, il est impossible d'accepter un effet délétère de ce type.

Exemple : Prasugrel dans ACCOAST ([http : www.ncbi.nlm.nih.gov/pubmed/23991622](http://www.ncbi.nlm.nih.gov/pubmed/23991622)) TRITON avec les hgie mortelles

15.12 Surcroît d'effet indésirable acceptable compte tenu du bénéfice apporté (bénéfice net favorable)

Ce résultat montre une augmentation statistiquement significative de cet évènement indésirable, mais cette augmentation n'est pas suffisamment importante (quantitativement ou qualitativement) pour remettre en cause le bénéfice obtenue par ailleurs. Le bénéfice clinique net reste favorable (Cf. [benefice net](#))

C'est l'image en miroir (du point de vue des évènements indésirables) d'un résultat démontré et cliniquement pertinent débouchant sur un bénéfice clinique net favorable démontré malgré un surcroît d'événements indésirables statistiquement significatif. Cette augmentation d'événements indésirables n'est pas suffisamment importante (quantitativement ou qualitativement) pour remettre en cause le bénéfice obtenue.

??? Exclure une sous estimation par manque de sensibilité et de spécificité. Intérêt

de l'adjudication recueil sensible : détection performante classification spécifique : guide, définition a priori, adjudication

15.13 Effet délétère remettant en cause un bénéfice obtenu par ailleurs

Ce résultat montre un effet délétère (surcroît d'événements indésirables) qui remettant en cause (quantitativement ou qualitativement) un bénéfice obtenu par ailleurs sur des événements de même gravité clinique. De ce fait, ce bénéfice débouche sur l'absence de démonstration d'un bénéfice clinique net favorable ($DR < 0$ NS) ou, pire, sur un bénéfice net défavorable, montré ($DR > 0$ $p < 0.005$) ou non exclus ($DR > 0$, NS) (Cf. [discussion du bénéfice net](#)).

Exemple : ATLAS ACS 2 - TIMI 51 , rivaroxaban dans ACS, ([http : www.ncbi.nlm.nih.gov/pubmed/22077192](http://www.ncbi.nlm.nih.gov/pubmed/22077192))

15.14 Effet délétère sous estimé par biais

Ce résultat suggère plutôt l'absence de l'effet indésirable considéré mais ce résultat n'est pas à l'abri des biais.

Pour les résultats de sécurité en faveur de l'absence d'un effet indésirables les s'inversent par rapport à ceux pouvant affecter la recherche de l'efficacité. En effet, pour la sécurité le biais redouté est celui qui conduit à un résultat faussement négatif (qui ne montre pas à tort l'existence d'un effet indésirable), tandis que pour l'efficacité le biais redouté est celui qui conduit à un résultat faussement positif (permettant de conclure à tort à l'efficacité). Hors les moyens méthodologiques mis en oeuvre pour protéger des biais la recherche de l'efficacité sont conservateurs, c'est à dire qu'ils tendent à réduire les différences entre les 2 groupes comparés afin de ne pas faire conclure à tort à l'existence d'une différence. C'est justement, pour la recherche de la sécurité, des situations qui pourraient faire conclure à tort à l'absence de défaut de sécurité (à l'absence d'effet indésirable).

Pour ce résultat, il n'est pas possible d'exclure qu'il soit apparemment en faveur de l'absence d'effet indésirable du fait d'un biais conservateur, favorisant l'émergence de résultat faussement rassurant négatif.

Les situations non à l'abri des biais pour l'évaluation de la sécurité d'un traitement sont :

- l'analyse en ITT
- le remplacement des données manquantes
- l'utilisation de méthodes peu sensible et/ou peu spécifique pour la recherche de l'effet indésirable considéré
- l'utilisation de critère de jugement très large (induisant une dilution d'un effet indésirable dans des événements non spécifiques)
- l'utilisation du traitement de comparaison de manière à maximiser le nombre d'événements indésirables qu'il engendre (dose, co-pathologies, etc.)
- l'absence de double aveugle

[les biais potentiel de l'évaluation de la safety](#)

15.15 Résultat ne montrant pas de défaut de tolérance pouvant contrebalancer le bénéfice

Ce résultat ne suggère pas l'existence d'un effet indésirable qui pourrait contrebalancer qualitativement ou quantitativement le bénéfice obtenu.

Le critère de jugement est un critère de sécurité d'emploi (safety) et aucune différence statistiquement significative n'a été notée entre les 2 groupes. Une différence non significative ne permet pas de conclure à l'absence de surcroît de cet événement indésirable mais ce résultat n'est pas susceptible de contre balancer le bénéfice obtenu car :

- soit la gravité de cet événement est faible comparé à la gravité des événements prévenus par le traitement
- soit l'événement est très rare et ne peut numériquement compenser le nombre d'événements évités. Malgré le manque de précision avec lequel on connaît l'effet du traitement sur ce critère (intervalle de confiance très large ou incalculable), le bénéfice clinique net après intégration de ces événements reste favorable au traitement étudié.

15.16 Non infériorité cliniquement pertinente démontré

Le résultat permet de conclure à la non-infériorité et montre une perte d'efficacité potentielle acceptable compte tenu de l'avantage par ailleurs apporté par le traitement. Cela sous entend aussi que cet avantage est parfaitement démontré dans cet essai où dans un autre (ou que cet avantage est implicite) et que son importance clinique permet d'accepter une perte d'efficacité au moins aussi importante que celle compatible avec le résultat démontré.

non infériorité

15.17 Non infériorité démontré mais non cliniquement pertinente

Le résultat montre une non-infériorité démontré statistiquement et méthodologiquement mais qui n'est cependant pas acceptable car entraînant potentiellement une perte d'efficacité inacceptable compte tenu de l'avantage par ailleurs. La perte d'efficacité potentielle sous jacente à la démonstration de la non-infériorité n'est pas acceptable car non compensé par un avantage par ailleurs, démontré ou implicite, suffisamment intéressant pour justifier cette perte d'efficacité. Soit l'avantage par ailleurs n'est pas formellement démontré, soit il est insuffisamment cliniquement pertinent pour justifier d'accepter de perdre autant en efficacité que ce que laisse possible la démonstration de non infériorité.

Au total le nouveau traitement ne démontre pas qu'il représente un progrès thérapeutique par rapport au comparateur utilisé dans l'essai. Tout au plus, ce résultat permet d'établir que le nouveau traitement à une efficacité suffisante (par un raisonnement de type placebo putatif), mais cela est insuffisant pour justifier l'utilisation du traitement en 1er intention à la place du traitement de référence et cela permet seulement d'en faire une alternative de 2ème intention (en cas d'intolérance au traitement de référence par exemple). «< faire une fiche placebo putatif et essai d'efficacité suffisante»> Exemple : RELY 110 ou ROCKET

non infériorité

15.18 Non infériorité suggéré mais non formellement démontré

Le résultat suggère une non-infériorité (la borne péjorative de l'intervalle de confiance est inférieure à la limite de non infériorité fixée au protocole) mais ce résultat ne peut pas être considéré comme formellement démontré car il n'est pas possible d'exclure qu'il puisse être faussement positif dû au hasard et/ou à un biais (cf. les réserve(s) émise(s)).

Compte tenu de ces réserves sur la validité interne du résultat, sa pertinence clinique n'a pas été évaluée.

Pour pouvoir écarter la possibilité qu'un résultat suggère faussement la non infériorité il faut :

- Pouvoir écarter une multiplicité de la recherche de non infériorité (résultat non obtenu sur un critère de jugement principal, résultat de sous groupe, d'analyse intermédiaire non protégée, switch d'hypothèse de la supériorité vers la non infériorité, formulation post-hoc d'hypothèse, etc. . .)
- Pouvoir écarter un biais de sélection, de suivi et de mesure.
- Pouvoir écarter un biais vers l'absence d'effet, c'est-à-dire une situation où, même en cas de réelle différence entre les 2 traitements comparés, le résultat ne montrera pas de différence. Il sera biaisé vers l'absence d'effet (0 pour une différence ou 1 pour un risque relatif). Ce biais peut survenir quand l'essai n'a aucun pouvoir discriminant entre les 2 groupes : patients non malades ou à très faible risque, analyse en intention de traiter, remplacement des données manquantes sur le critère de jugement par une méthode conservatrice, utilisation d'une mesure peu sensible ou peu spécifique du critère de jugement, etc.

Les mécanismes de biais dans un essai de non infériorité sont mes mêmes que ceux dans les essais de supériorité. Pour rendre un traitement en réalité inférieur simplement non inférieur il faut que les biais induisent une surestimation de l'effet du nouveau par rapport au standard comme pour biaisé un essai de supériorité, c'est à dire transformer l'absence d'effet en supériorité. Ainsi tout les mécanismes de biais discutés pour l'essai de supériorité s'appliquent sans changement à l'essai de non infériorité, sauf pour le biais d'attrition.

Un biais vers l'absence d'effet survient dans un essai lorsque l'essai a une faible aptitude à mettre en évidence une différence réelle. L'essai n'a pas de « sensibilité » pour détecter les différences et va donc conduire à un résultat biaisé vers l'absence de différence. Cela survient par exemple quand l'essai mesure le critère de jugement avec une très faible spécificité engendrant de nombreux faux positifs dans les 2 groupes et in fine à une même fréquence du critère de jugement entre les 2 groupes comparés, quel que soit la réelle différence d'efficacité entre les 2 traitements.

Pour détecter cette absence de sensibilité d'un essai de non infériorité, il conviendrait d'avoir un 3ème groupe, recevant un placebo. Ce groupe permettrait de valider l'aptitude de l'essai à mettre en évidence des différences en montrant que l'on retrouve la différence attendue entre le traitement standard et le placebo (obtenue dans l'essai antérieur validant le traitement de référence utilisé dans l'essai de non-infériorité). Cependant ce troisième groupe n'est pas toujours possible (il entraîne une perte de chance pour les patients si le traitement standard a démontré qu'il apportait un bénéfice clinique substantiel) et il ne sera possible d'écarter ce biais que si le résultat a été obtenu dans les circonstances suivantes :

- Analyse en per protocole
- Mesure du critère de jugement optimale

- Absence de remplacement des données manquantes si elles sont en faible nombre ; ou remplacement par une méthode inflationniste (biais maximum inversé par exemple)
- Patients à risque notoire
- Absence d'utilisation de traitements concomitants efficaces dans les 2 groupes (cf. fiche concept [non infériorité](#))

15.19 Données non matures

bla bla

(cf. fiche concept [analyses intermédiaires](#))

Chapitre 16

Intérêt thérapeutique

Pour apporter des éléments de preuve en faveur de l'intérêt thérapeutique du traitement évalué, un essai doit avoir produit au moins un résultat démontré, cliniquement pertinent, débouchant sur un bénéfice clinique net favorable ou un résultat démontrant une non-infériorité cliniquement pertinente.

En l'absence de ce type de résultat, un essai n'apporte pas d'éléments de preuve pour l'intérêt thérapeutique du traitement étudié. Sans ces éléments de preuve, l'acceptation de l'intérêt du traitement ne repose que sur des arguments spéculatifs, comme le mécanisme d'action du traitement, ou de nature purement inductif, basés sur l'observation. À côté de la réflexion épistémologique, l'expérience a aussi montré la fragilité potentielle de ces arguments en apportant de nombreux contre-exemples de non confirmation d'arguments de ce type par des d'essais plus rigoureux ([liste d'exemple](#)). De ce fait, il est maintenant admis que ces d'arguments ne sont pas suffisants pour assoir sur une pratique courante.

16.1 L'essai apporte un ou plusieurs résultats en faveur de l'intérêt thérapeutique du traitement évalué

L'essai a produit au moins un résultat qui démontre un bénéfice cliniquement pertinent avec bénéfice net favorable et il n'y a pas de critère débouchant sur un bénéfice net défavorable. De par ce résultat, l'essai apporte donc un élément de preuve en faveur de l'intérêt thérapeutique du traitement évalué.

Une autre situation apportant un élément de preuve pour l'intérêt du traitement est la démonstration d'une non infériorité cliniquement pertinente. Dans ce cas le nouveau traitement entraîne une possible perte d'efficacité mais qui, compte tenu des avantages du traitement par ailleurs, est acceptable et, au total, le nouveau traitement apporte un progrès thérapeutique par rapport au traitement de référence. Cette situation est, en quelque sorte, équivalente à une démonstration de la supériorité du nouveau traitement sur un critère composite regroupant le critère de bénéfice sur lequel est démontré la non infériorité et le critère mesurant l'avantage par ailleurs. Comme la non-infériorité est cliniquement pertinente, cela valide la pertinence clinique du regroupement de ces 2 critères. De ce fait, l'essai apporte un élément de preuve pour l'intérêt thérapeutique du nouveau traitement par rapport au traitement standard.

Bien que prépondérant dans la décision de conclure à un intérêt thérapeutique, un bénéfice démontré n'implique pas de facto cette conclusion. D'autres éléments entre en ligne de compte comme la gravité ou le risque résiduel de la maladie qui va conditionner la rentabilité du traitement (apprécie par le bénéfice absolu ou absolute risk reduction, ARR). Ce dernier aspect n'est pas documenté dans ce site car il dépend fortement du contexte

(contexte de santé publique, des autres traitements disponibles, du contexte économique, etc.). Cf. [rentabilité du traitement](#) .

16.2 L'essai n'apporte pas de résultat en faveur de l'intérêt thérapeutique du traitement évalué

Les résultats produits par cet essai sont insuffisant pour constituer un élément de preuve de l'intérêt thérapeutique du traitement évalué. Ces résultats peuvent être insuffisant soit en raison de limites méthodologiques ou statistiques qui ne permettent pas de constituer des démonstrations formelles, soit en raison d'une insuffisance de pertinence clinique.

Au total, aucun bénéfice cliniquement pertinent et avec un bénéfice net favorable n'a été démontré :

- Soit parce qu'aucun résultat ne peut être considéré comme formellement démontré (en raison de limites méthodologiques ou statistiques)
- Soit parce qu'aucun résultat ne montre un bénéfice cliniquement pertinent ;
- Soit parce qu'il existe un bénéfice mais que celui-ci est susceptible d'être entièrement contre balancé par les effets indésirables (qualitativement ou quantitativement). In fine aucun bénéfice net favorable n'a été montré ;
- Ne démontre qu'une non-infériorité non cliniquement pertinente.

L'essai n'apporte donc pas d'argument pour la décision de conclure à l'intérêt thérapeutique du traitement évalué.

Une seule démonstration de non infériorité est en général insuffisante pour considérer que l'intérêt thérapeutique du traitement est avéré. En effet, sauf en cas de non infériorité parfaitement compensée par des avantages par ailleurs notables, parfaitement démontrés et cliniquement pertinents (dans ce cas le progrès thérapeutique apporté par le nouveau traitement est évident), la non infériorité ne permet pas de positionner le nouveau traitement par rapport à son comparateur.

Les essais de non infériorités peuvent aussi être vu comme des essais d'efficacité suffisante, mais dans ce cas, le raisonnement se fait par rapport au placebo putatif (ou par rapport au comparateur ayant servi à valider le traitement de référence utilisé dans l'essai de non infériorité).

Deux situations plus une en résumé :

- Non infériorité pertinente avec avantages par ailleurs pertinent et parfaitement démontré = progrès thérapeutique = intérêt thérapeutique avérée
- Non infériorité non pertinente = pas d'intérêt thérapeutique (peut être efficacité suffisante)
- Utilisation de l'essai de non infériorité en efficacité suffisante versus placebo putatif, évaluation du biais d'extrapolation

[Les conséquences des décisions basées sur des preuves imparfaites](#)

Chapitre 17

Cohérence externe d'un résultat

Un dernier élément de validité d'un résultat est sa cohérence externe, c'est-à-dire sa cohérence avec d'autres résultats estimant la même chose.

Il s'agit de vérifier si le ce résultat est confirmé ou infirmé par d'éventuels autres résultats concernant la même comparaison de traitements dans la même pathologie et les mêmes patients et sur le même critère de jugement, un problème de cohérence.

17.1 Les différents cas de figure

17.1.1 Le résultat considéré est en cohérence avec les résultats des autres études similaires

17.1.2 L'appréciation de la cohérence externe impossible

L'appréciation de la cohérence externe du résultat considéré n'est pas possible, car il n'existe aucun n'autre essai similaire (portant sur la même comparaison de traitement et sur le même type de patients).

Conséquences ???

17.1.3 Remise en cause de la cohérence externe du résultat

Un résultat peut être remis en cause pour manque de cohérence externe quand les autres essais similaires existants ne le confirment pas. La non-confirmation se juge en terme de taille de l'effet et non pas en terme de signification statistique. En effet, des discordances de signification statistiques entre essais peuvent simplement provenir des différences de puissance des études sans que l'effet du traitement en varie d'un essai à l'autre. Il y a non confirmation quand les résultats quantitatifs des essais varient de manière importante, avec une hétérogénéité des résultats supérieurs à ce qui est attendu uniquement du fait des fluctuations aléatoires d'échantillonnage.

Ainsi un résultat sera récusé pour manque de cohérence externe quand la méta-analyse intégrant ce résultats et ceux des autres études similaires ne montrent pas d'effet statistiquement significatif et ne confirme donc pas le résultat considéré. Les autres résultats récusent en quelque sorte le résultat considéré. Celui-ci s'avère, ainsi, être particulièrement favorable à l'effet du traitement, de manière non entièrement reproductible. Au total, il n'est donc pas possible l'existence d'un effet du traitement puisse qu'au total, en considérant la totalité de l'information apportée par tous les essais faisant la même évaluation, il n'est pas possible de conclure à cet effet.

Dans ces circonstances, le résultat analysés s'avère donc être très vraisemblablement un résultat faussement positif qui peut être dû à un biais non identifié de l'essai ou à des facteurs modificateurs de l'effet (une hétérogénéité existe alors) ou dû simplement au hasard à travers les fluctuations aléatoires d'échantillonnages (il n'est pas mis en évidence d'hétérogénéité puisque les différences entre les résultats des essais proviennent uniquement du hasard).

Exemple :

L'étude CHAOS fourni un résultat en faveur du bénéfice de la vitamine E en prévention cardiovasculaire en montrant une réduction de la fréquence des décès cardiovasculaires par rapport au placebo de ??? réf. Cependant ce résultat s'avère unique, car il n'est pas retrouvé dans les ??? autres essais ayant le même objectif, qui sont tous non concluants.

La méta-analyse regroupant tous ces résultats ne montre pas de différence statistiquement significative et aucune hétérogénéité importante n'est détectée. Il est ainsi possible de conclure que le résultat positif de CHAOS est très vraisemblablement un faux positif (dû au hasard ou à un biais non détecté¹). Cela conduit à rejeter le résultat de CHAOS pour manque de cohérence externe et à considérer que la vitamine E n'a pas apporté la preuve de son intérêt sur les décès cardiovasculaires.

17.2 Exemple : infarctus et dabigatran

Chapitre 18

But évaluation

Le but de l'évaluation est d'évaluer si un résultat donné, proposé (avancé) pour justifier l'intérêt d'un traitement, est suffisamment fiable et pertinent pour être accepté comme preuve de cet intérêt thérapeutique.

18.1 Logique générale de l'évaluation

L'évaluation a donc pour mission d'informer un décideur sur le risque de prendre une mauvaise décision qu'il encoure en suivant ce que suggère le résultat considéré

Ici, le décideur (individu ou institution) peut être :

- un médecin décidant de l'intérêt d'adopter un nouveau traitement dans son arsenal thérapeutique,
- un enseignant de médecine décidant d'introduire le nouveau traitement dans ses cours,
- un agent de régulation décidant d'octroyer une autorisation de marché,
- un payeur décidant de prendre en charge le nouveau traitement.

Tous ces décideurs ont la même crainte, qu'ils souhaiteraient absolument éviter étant donné les enjeux pour les patients et le système de santé, celle de conclure à tort à l'intérêt du traitement à partir d'un résultat de recherche clinique suggérant abusivement (à tort, faussement) l'intérêt du traitement. Étant donné cette crainte, tous ces décideurs souhaiteraient ne suivre que des résultats de recherche clinique suffisamment sûrs pour que leur risque de prendre une mauvaise décision soit le plus faible possible (voir quasiment inexistant).

TABLE 18.1 – caption

Décideur	Décision recherchée	Crainte
Médecin (clinicien)	Adopter un nouveau traitement dans son arsenal thérapeutique	Traiter des patients avec un traitement sans intérêt
Enseignant de médecine	Introduire le nouveau traitement dans ses cours	Enseigner à tort un traitement
Agent de régulation	Octroyer une autorisation de marché	Donner l'AMM à un traitement sans intérêt
Payeur	Prendre en charge le nouveau traitement	Rembourser un traitement sans intérêt

Il existe une autre possibilité de prendre décision inappropriée. Celle de ne pas retenir l'intérêt thérapeutique à tort d'un traitement pourtant efficace net ainsi de priver de ma-

nière plus ou moins définitive les patients d'une ressource thérapeutique pourtant efficace. Cet aspect est discuté dans la section ???.

Hors tous les résultats de recherche clinique suggérant un intérêt du nouveau traitement ne se valent pas et certains sont plus sûrs que d'autres en fonction du type de l'étude qui les ont produits, de la protection contre les biais de l'étude, du contrôle du risque d'erreur statistique dû au hasard qui a été effectué dans l'étude, de la concordance du résultat avec l'objectif initial de l'étude, etc..

Cela revient donc à évaluer, devant un résultat apparemment en faveur de l'intérêt du traitement, le risque que ce résultat dit « positif » soit en réalité faussement positif (positif à tort). Un résultat faussement positif est un résultat d'apparence positive (en faveur de l'intérêt du traitement), mais d'apparence seulement, car en réalité, le traitement est dénué d'intérêt. « Les chiffres sont en faveur de l'intérêt, mais à tort ».

La conclusion que suggère le résultat est fautive et en désaccord avec la réalité de l'intérêt du traitement. C'est donc un résultat d'étude faussement positif, c'est un faux positif. (il existe aussi une possibilité d'avoir des résultats faussement négatifs : le résultat n'est en faveur de l'intérêt du traitement à tort, car le traitement a réellement un intérêt)

Plusieurs phénomènes peuvent être à l'origine de la genèse de résultats faussement positifs. De ce fait, la possibilité de résultats faux positifs est une réelle question d'intérêt.

Les phénomènes qui peuvent conduire à des résultats faux positifs sont les suivants :

- le phénomène de confusion : ce qui est pris pour un effet du traitement est en fait l'effet d'un autre facteur, donnant un résultat identique à ce que donnerait un traitement efficace. De ce fait ce résultat, en faveur de l'effet du traitement, peut survenir indépendamment de tout effet réel du traitement étudié. L'effet de ce tiers facteur a été confondu avec l'effet du médicament. Ces tiers facteurs sont appelés facteurs de confusion.
- Les biais qui font qu'une étude donnera des résultats en faveur de l'intérêt du traitement par (défaut) construction, quel que soit le réel intérêt du traitement
- Les erreurs statistiques de première espèce (risque d'erreur alpha)
- Les artefacts, les découvertes fortuites comme celles issues d'une observation sans hypothèse préalable produites d'un raisonnement inductif et non pas déductif
- Les erreurs de raisonnements ou de logique (logical fallacy ou formal fallacy) comme les résultats post hoc

TABLE 18.2 – caption

Phénomène pouvant générer des faux positifs	Solutions
phénomène de confusion	groupe contrôle ou ajustement sur la totalité des facteurs de confusion
biais	protection contre les biais (la notion de biais et de confusion est partiellement similaire)
erreurs statistiques de première espèce	test statistique permettant de contrôler le risque de première espèce à un niveau choisi
erreurs de raisonnements ou de logique	ne pas les faire (éviter les raisonnements post hoc, etc. . .)
découvertes fortuites comme celles issues d'une observation sans hypothèse préalable (raisonnement inductif)	n'accepter que les résultats hypothétiques déductifs (raisonnement déductif)
spéculation sur un mécanisme	vérification du résultat attendu

Devant la multitude de ces phénomènes pouvant conduire à des résultats faux positifs, il a été développé au cours du temps, des outils et des principes qui ont pour but d'éviter de

produire des résultats faux positifs en recherche scientifique. Ce fût le but du développement de la méthode scientifique .

Ces principes nous assurent que, s'ils sont respectés (appliqués), les résultats positifs obtenus ne sont pas des résultats faussement positifs. Cette assurance n'est cependant pas absolue, mais tout aura été mis en œuvre pour limiter ce risque et il est en général accepté que pour éliminer ce risque résiduel de faux positif il convient de vérifier tout résultat positif par une réplification (une autre étude) ???

Dans sa recherche de prendre des décisions au plus juste, le décideur souhaiterait n'utiliser que des résultats pour lesquels il est raisonnablement possible d'écarter la possibilité qu'ils puissent être faussement positifs (positifs à tort).

Or, en dehors du plein respect de ces principes évoqué précédemment, il n'est pas possible de raisonnablement exclure qu'un résultat positif le soit à tort. Ce point illustre le fait qu'un résultat ne respectant pas ces principes est forcément un résultat faussement positif, mais il n'est pas possible d'exclure qu'il le soit. Il est tout à fait possible d'un résultat ne respectant pas ces principes soit à raison positif et reflète bien la réalité, mais le risque de prendre la mauvaise décision à partir d'un résultat ne respectant pas ces principes n'est pas contrôlé et n'est pas minimisé ; ainsi il est sûr qu'il y a prise de risque à suivre un résultat ne respectant pas ces principes. Le maximum n'a pas été fait pour minimiser le risque de prendre une décision erronée.

Liste de résultat s'étant avéré être des faux positifs ultérieurement après la réalisation de nouvelle étude ou d'étude plus solide

TABLE 18.3 – caption

Résultat	Type d'étude à l'origine de ce résultat	Réfuté par	Explication du faux positif
statine et BPCO	Etude observationnelle	RCT	???
statine et choc septique	Etude observationnelle	RCT	???

18.2 Prise en considération du risque de faux négatifs

Pourquoi ne pas considérer le risque de prendre une mauvaise décision en ne reconnaissant pas l'intérêt du nouveau traitement dans l'élaboration des exigences de niveau de démonstration ?

La démarche d'évaluation est entièrement centrée sur la minimisation du risque d'utiliser à tort un nouveau traitement. En symétrie, il existe aussi le risque de ne pas utiliser à tort un traitement (pourtant efficace). Adopter des exigences élevées de démonstration du bénéfice fait courir le risque que certains traitements efficaces ne soient pas retenus, car les résultats produits par leurs essais n'atteignent pas le niveau d'exigences de solidité demandé. Cette situation pourrait être perçue comme dommageable aussi pour les patients, car conduisant à une perte de chance.

Cependant la solution ne doit pas être un abaissement des exigences, mais bien une amélioration de la qualité des études de ces traitements. En effet, abaisser les exigences conduirait, certes à diminuer le taux de faux négatif, mais aussi et surtout à augmenter celui de faux positifs. Au total, la probabilité qu'un traitement adopté soit efficace diminuerait (cf. fiche concept sur ???).

En effet, de nombreuses raisons conduisent à privilégier le contrôle du risque de faux positifs par rapport à celui de faux négatifs.

Un faux négatif est toujours rattrapable tandis qu'un faux positif ne l'est pas (ou très rarement). Il est toujours possible de refaire un essai après un premier négatif et rattraper ainsi un éventuel résultat faussement négatif. Par contre une fois qu'un nouveau traitement est adopté, aucun nouvel essai (sauf cas exceptionnel¹) ne sera plus entrepris et il y aura ainsi plus aucune possibilité de se rendre compte que le résultat ayant conduit à l'adoption du nouveau traitement était un faux positif. Ainsi il convient d'être bien plus vigilant vis-à-vis des faux positifs que des faux négatifs étant donné le caractère irrémédiable de ces premiers contrairement aux derniers.

L'utilisation à tort d'un traitement sur la base d'un résultat faux positif entraîne aussi une perte de chance pour le patient car cela prive le patient d'éventuels autres traitements efficaces (cf Milwaukee sentinelle). Cela donne aussi une fausse impression que les besoins sont couverts, limitant les efforts de recherche d'un traitement supplémentaire.

L'adoption à tort d'un nouveau traitement conduit aussi à une asymétrie des bénéfices : alors qu'aucun bénéfice n'est apporté aux patients ou à la santé publique, le fabricant tire de son traitement les bénéfices attendus.

Les résultats faussement positifs conduisent aussi à un gaspillage direct des ressources collectives ou de celles des patients en fonction du système de santé.

In fine, il convient aussi de remarquer que la communauté biomédicale à l'heure actuelle favorise nettement le contrôle des faux positifs au détriment de celui des faux négatifs, puisqu'une puissance de 80% est couramment employée, correspondant à un taux accepté de faux négatifs de 20%, alors que le taux couramment admis de faux positifs est seulement de 2.5% (seuil de signification de 5% en bilatéral).

1. Les exemples d'essais entrepris après l'adoption d'un nouveau traitement dans le but de remettre en question leur intérêt sont extrêmement rares : CAST avec les antiarythmiquesref???

Chapitre 19

Conséquences des décisions basées sur des preuves imparfaites

19.1 introduction

L'évaluation recherche les preuves optimales de l'intérêt thérapeutiques d'un traitement. Parfois, pour certaines conditions cliniques, l'obtention de ces preuves optimales est un objectif très difficilement atteignable : nombre de sujets nécessaires pour mettre en évidence le bénéfice attendu disproportionné par rapport à la population de malade existant (maladies rares), réalisation du double insu difficile (voir dans quelques cas impossible), etc. Dans ces situations on est donc amené à prendre des décisions concernant la recommandation de traitement sur la base de preuves imparfaites. Se pose alors la question des conséquences de telles décisions.

Dans certains cas de figures on est tenté d'adopter des traitements à partir de preuves non complètement optimales. Quelles sont les conséquences de ce type de décision

L'optimalité d'une preuve en faveur de l'intérêt thérapeutique d'un nouveau traitement se définit en termes de risque que cette preuve soit faussement en faveur de l'intérêt du traitement. Ce risque est minimal avec une preuve optimale, répondant à toutes les exigences méthodologiques et augmente lorsque toutes ces exigences ne sont remplies (cf. section ???).

Mais, mêmes avec toutes les conditions nécessaires, le risque qu'une preuve optimale soit faussement en faveur de l'intérêt du traitement n'est pas nulle et, le souvent, s'établit à 2.5% (ce qui correspond au risque alpha de la partie en faveur de la supériorité d'un test statistique bilatéral habituellement effectué avec un seuil de la signification à 5%).

Lorsque une preuve ne remplit pas toutes les exigences méthodologiques, ce risque de faux positif augmente mais, en l'état actuel des connaissances, il n'est pas possible de quantifier cette augmentation. Ainsi, par exemple, il est impossible de quantifier le risque qu'un résultat provenant d'un essai réalisé sans insu soit un faux positif. La seule certitude est qu'il n'est pas possible de garantir que ce risque soit inférieur à 2.5%. En effet chaque fois qu'un essai en double insu ne confirme pas les résultats d'un premier essai réalisé sans insu, cela montre que le risque de faux positif des essais en ouvert est supérieur à 2.5% (celui des essais en double insu).

L'évaluation peut ainsi être vue comme une sorte de filtre qui empêche que trop de traitements sans effet soient considérés à tort comme ayant un intérêt thérapeutique et

soient in fine adopté en pratique. En exigeant des preuves optimales, seul 2.5% de ces traitements sans effet filtrent et sont finalement recommander pour la pratique.

Dans les situations où ils n'est pas possible d'obtenir des preuves optimales, l'évaluation sera moins performantes et plus de 2.5% des traitements sans effets soumis à l'évaluation seront finalement adoptés.

Pour percevoir les réelles conséquences de cette plus grande perméabilité du filtre aux traitements sans effet il est nécessaire de raisonner en termes de probabilité qu'un traitement recommandé soit, en fait, sans intérêt. En effet, du point de vue du médecin, la question est : quelle est la probabilité que ce traitement que je m'apprête de prescrire soit efficace? et du point de vue du patient, la question est : quelle est la probabilité que le traitement que je m'apprête de suivre soit efficace? Pour calculer cette probabilité il convient de recouvrir au théorème de Baye.

19.2 Place de ces preuves imparfaites dans le décision

Ce résultat n'est donc pas démontré mais seulement suggéré. Il ne peut pas être considéré comme un éléments de preuve de l'intérêt thérapeutique du traitement évalué. De ce fait il ne peut pas être utilisé pour justifier une décision d'utiliser ou de recommander le traitement évalué. Il a seulement une valeur exploratoire ou cognitive. Il n'intéresse pas directement le clinicien mais le chercheur. Compte tenu de son absence de démonstration, ce résultat n'est pas en mesure de guider le clinicien dans le choix d'un traitement ou de l'élaboration d'une recommandation.

19.3 calcul de la probabilité qu'un traitement soit réellement efficace après un essais "concluant"

Les explications de ce calculs sont donnés dans [risques statistiques vu comme des taux de filtration](#) .

Le calcul de cette probabilité peut s'effectuer par dénombrement (ce qui évite de recourir aux théorèmes du calcul des probabilités). Les 2 approches sont équivalentes étant donné peut être vue comme la fréquence de survenu de l'événement stochastique d'intérêt.

Considérons la population théorique de l'ensemble des traitements qui seront évalués pour une question thérapeutique donnée. La proportion de traitement efficace est p et celle de traitement sans efficacité $1 - p$. Les essais réalisés pour évaluer les traitements sans efficacité ont tous un risque de produire un résultat faux positif (dû au hasard et au biais) égal à α . Pour les essais réalisés avec les traitement efficace, le risque de faux négatif est β .

Au total, sur l'ensemble des essais réalisés, les taux de faux positifs (FP) et de vrai négatif (VN) seront

$$FP = (1 - p)\alpha$$

$$VN = (1 - p)(1 - \alpha)$$

Les taux de vrai positif (VP) et de faux négatif (FN) seront

$$VP = p(1 - \beta)$$

$$FN = p\beta$$

Au total le nombre de traitements étiquetés comme efficace sera $VP + FP$ et celui des traitements considérés sans effet sera $VN + FN$.

Mais parmi les traitements considérés comme efficaces, la proportion θ de traitement vraiment efficace est de $VP/(VP + FP)$ soit

$$\theta = \frac{p(1 - \beta)}{p(1 - \beta) + (1 - p)\alpha}$$

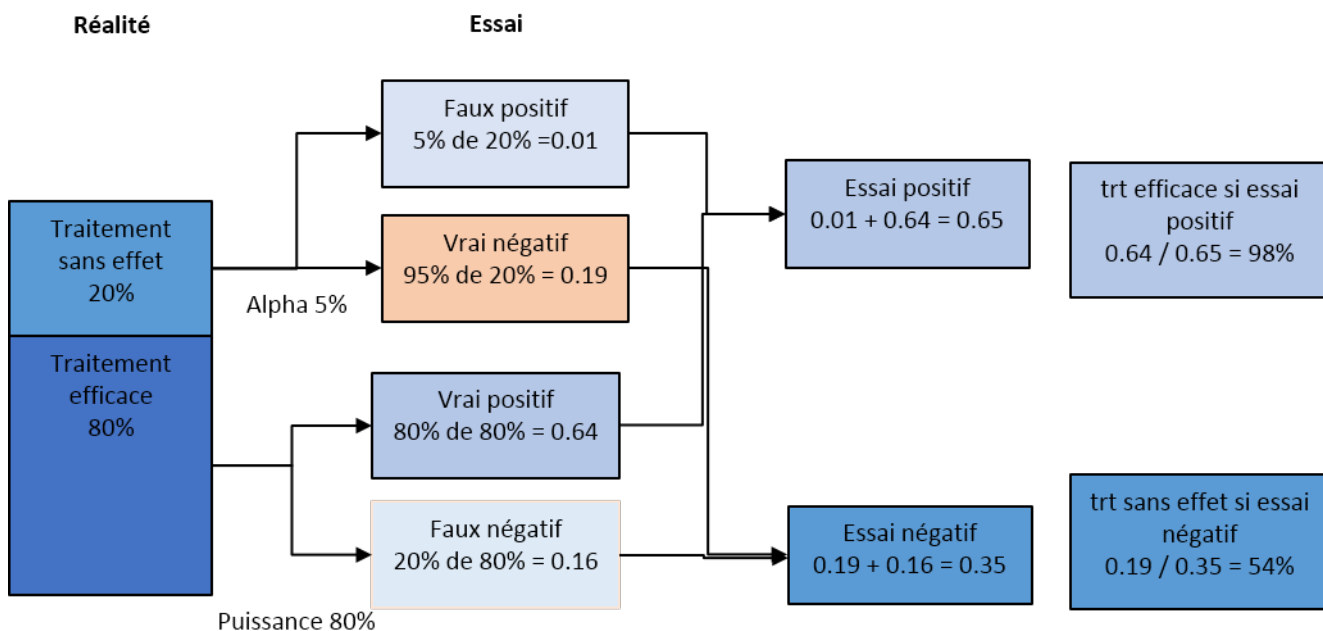


FIGURE 19.1 – Figure 1 - Illustration du calcul de la probabilité que le traitement soit efficace si l'essai est positif.

En d'autre termes, la probabilité qu'un traitement ayant donné lieu à un essai concluant soit réellement efficace dépend :

- de la probabilité a priori p que le traitement soit efficace
- du risque α qu'avait l'essai de donner un résultat faussement positif (dû au hasard ou à un biais)
- de la puissance statistique $1 - \beta$ de l'essai

Ainsi une étude concluante sera d'autant moins prédictive de l'efficacité du traitement que :

- l'hypothèse ayant conduit à l'étude était très spéculative ou inexistante (comme lors d'une découverte fortuite dans une étude de data-mining par exemple où l'on cherche sans hypothèse préalable)
- l'essai a des lacunes méthodologiques (qui augmente le risque d'avoir un faux positif dû au hasard)
- il y a eu une multiplicité des comparaisons statistiques lors de la recherche de l'efficacité du traitement
- l'essai avait une faible puissance
- le critère de jugement était peu prédictif du bénéfice réel que peut apporter un traitement (cas des critères intermédiaires).

Ainsi, un résultat concluant n'a pas du tout la même valeur prédictive de l'existence de l'efficacité du traitement (valeur de démonstration) suivant qu'il s'agit d'une preuve optimale ou d'une preuve imparfaite.

Ce développement est strictement identique à la théorie des tests diagnostiques. Le risque de vrai positif est identique à la sensibilité d'un test diagnostique, le taux de vrai négatif à sa spécificité. La probabilité que le traitement soit réellement efficace à l'issue d'une étude concluante est la valeur prédictive positive (VPP). La probabilité a priori que le traitement soit efficace est la probabilité a priori que le patient soit malade (prévalence de la maladie).

19.4 Conséquences de l'utilisation de preuves imparfaites pour recommander l'utilisation d'un traitement

Quel est l'acuité de cette problématique en pratique ?

Quelle est la valeur prédictive de la réelle efficacité des résultats couramment utilisés pour conclure à l'efficacité du traitement ?

La probabilité a priori que le nouveau traitement soit efficace est un élément important de la problématique. Cette probabilité peut être vue comme étant le reflet de l'efficacité des études des phases pré cliniques et précoces chez l'homme (1 et 2) de sélectionner des traitements candidats de valeur. Cette probabilité n'est pas très facile à cerner mais il est évident qu'elle peut avoir des valeurs très différentes en fonction des circonstances. Par exemple, si un nouveau médicament appartient à une classe déjà très étudiée et où de nombreuses autres molécules ont déjà démontré dans des bons essais qu'elles apportent un bénéfice (comme par exemple les statines en prévention secondaire), la probabilité a priori (avant la phase 3) que cette nouvelle molécule soit efficace est certainement très importante. Par contre si cette nouvelle molécule est la première de sa classe, sa réelle efficacité clinique est nettement plus incertaine. L'hypothèse de l'efficacité du traitement est bien plus spéculative que dans le cas précédent. Il est possible de cerner la valeur de cette probabilité en faisant le dénombrement des études de phase 3 qui ont été des échecs. Ce taux d'échec est d'environ 50% [3], ce qui conduit à une probabilité a priori de 50%.

De plus, la décision d'adopter un nouveau traitement sont souvent prises à partir de preuves imparfaites (essai non comparatif, non randomisé, essai randomisé en ouvert, données manquantes non remplacées, absence de critère de jugement principal, etc.). Cela est particulièrement fréquent avec les traitements ne donnant pas lieu à un enregistrement réglementaire, avec les l'évaluation des stratégies ou dans des domaines comme celui des dispositifs ou parfois en oncologie ref d'essais non contrôlé. De plus, dans ces situations, va se surajouter au risque intrinsèque de faux positifs de ces études une amplification du phénomène par le biais de publication. La non publication des résultats négatifs (qui sont les vrais négatifs pour les traitements sans efficacité), va augmenter le risque de faux positifs ??? Dans ces conditions, le risque global d'avoir un résultat faussement positif peut être très élevé (au delà de 50%).

En pratique, il est assez fréquent de rencontrer des essais de puissance insuffisante.

En particulier, il existe un domaine où il est difficile de faire des preuves optimales. Ainsi, si dans une pathologie rare, la puissance des essais évaluant les nouvelles propositions thérapeutiques est disons de l'ordre de 50% et que, compte tenu des limites méthodologiques des études réalisées, les résultats positifs obtenus ont une probabilité de 25% d'être faussement/abusivement positifs, la proportion des traitements adoptés qui sont en fait sans efficacité est de 33% si la probabilité a priori d'efficacité est de 50%. Si les hypothèses a priori sont encore plus spéculatives et moins étayées (comme c'est assez fréquemment le

cas dans ce domaine ref à trouver), cette proportion de traitement inutile monte à 67% (pour une probabilité à priori de 20%).

Les conséquences de l'utilisation de preuves imparfaites de l'intérêt thérapeutiques des nouveaux traitements sont donc importantes. L'adoption des traitements à partir de preuves non optimale (essais insuffisamment puissants, non à l'abri des biais, sans démarche hypothetico déductive forte) peut conduire à ce que plus de la moitié des traitements adoptés le soit à tort et que la probabilité qu'un patient reçoive un traitement inefficace peut être de l'ordre de 50%.

Ces résultats montrent la toute fragilité des résultats d'essais significatifs. Dans de nombreux cas ce résultat "positif", souvent à l'origine d'un grand enthousiasme, n'a que finalement très peu de valeur predictive de l'efficacité du traitement : cela chaque fois que l'essai a une puissance insuffisante, présente des limites méthodologique (essai en ouvert, ITT sans remplacement des données manquantes) et/ou statistique (inflation du risque alpha, résultat de sous groupe, résultat obtenu sur un critère de jugement secondaire), ou que le résultat considéré est purement exploratoire (post hoc).

Un argument souvent apporté pour justifier des essais rapides ne cherchant pas toutes les garanties est d'augmenter les chances des patients en accélérant la mise à disposition des traitement, cela dans une démarche compassionnelle. Hors le résultat va être inverse en augmentant la probabilité de perte de chance pour les patients (en augmentant le risque qu'il est un traitement sans effet).

[3]

[3]

[3]

[3] PNAS

References bibliographique

- [1] référence Kola :2004bq inexistante [0]
- [2] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. [Internet]. Nat Rev Drug Discov. 2010 [cited 2014 May 6]. Available from : [http : www.nature.com/doifinder/10.1038/nrd3078](http://www.nature.com/doifinder/10.1038/nrd3078) [None]
- [3] référence Ioannidis :2005bw inexistante [0]
- [4] Hamel MB, Darrow JJ, Avorn J, Kesselheim AS. New FDA Breakthrough-Drug Category — Implications for Patients. N Engl J Med. 2014 Mar 27;370(13) :1252–8. [None]
- [5] référence Johnson :2013cv inexistante [0]

Chapitre 20

Chapitre 1

20.1 Bibliographe

ref par PMID [3]

pour les fichiers de références provenant de paper ou endnote appel à référence [3] [3]

20.2 Introduction

TABLE 20.1 –

A	B	L
1	2	3
3	4	5

TABLE 20.2 – caption

A	B
zzz	er

Ce document présente la **syntaxe** du langage perso de markdown [linked phrase](#)

— item 1

— item 2

— item 3

fin de la liste

la suite est [dans le chapitre double insu](#) .

Des précisions sont apportées dans la section précision

Tableau en mode liste

TABLE 20.3 – Table 2 - caption [refTableau1]

col 11	col 2	col 3
col 21	col 2	col 3
col 31	col 2	col 3

Liste hiérarchique :

1. item 1

(a) item 1.1

(b) item 1.2

2. item 2

et voici un exemple

Exemple : exemple d'un brut

exemple

exemple

dshfg jsdghfg jsfgq jfgjs fd

Note : titre de la short note

test de la short note, not more than 1 paragraph ended by a dollar sign

20.2.1 section 1

ceci est la section 1 ¹ sdhj kqhjs qdhs kdhkdsq

un nouveau paragraphe

1. note 1

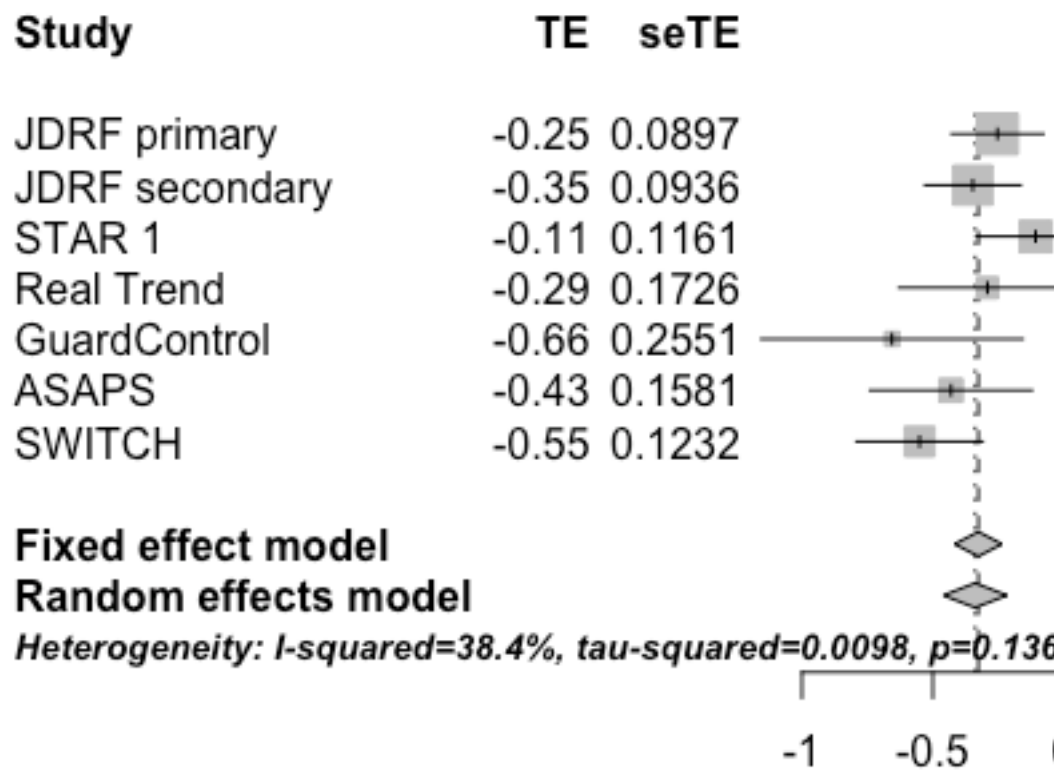


FIGURE 20.1 – Figure 1 - ceci est une figure

formulation alternative pour une figure

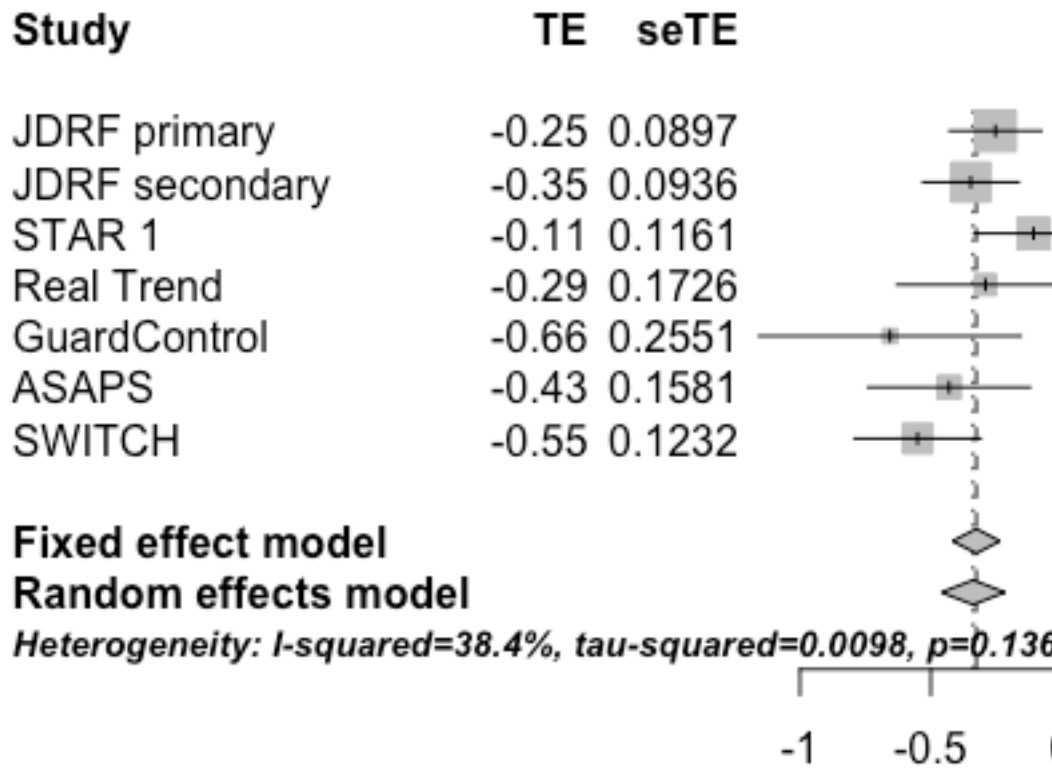


FIGURE 20.2 – Figure 2 - titre de la figure

un paragraphe saisie sur 2 lignes qui seront regroupées

20.2.2 section 2

Ceci est la section 2
avec 2 paragraphe

20.3 Précision [ref-section-precision]

Des precisions

20.4 Renvois et références

liens [texte du lien](#)

note de bas de page²

liens vers un fichier [dans le chapitre double insu](#) .

label \$label label

20.5 fichier d'aide (section help)

[les biais de suivi et de mesure, le double insu](#)

2. note de bas de page

References bibliographique

- [1] Lega JC, Bertoletti L, Gremillet C, Chapelle C, Mismetti P, Cucherat M, Vital-Durand D, Laporte S. Consistency of safety and efficacy of new oral anticoagulants across subgroups of patients with atrial fibrillation. PLoS One 2014 ;9(3) :e91398 [24622012]
- [2] Vicaut E, Cucherat M. [Noninferiority trials : several simple principles]. Presse Med. 2007 Mar ;36(3 Pt 2) :531–5. [17289337]
- [3] Verde PE. Meta-analysis of diagnostic test data : a bivariate Bayesian modeling approach. Stat Med. 2010 Dec 30 ;29(30) :3088–102. [21170904]