

**GENERALITES
SUR LA METHODOLOGIE
DES ESSAIS CLINIQUES
(ESSENTIEL)**

PLAN

1) Avant le recueil des données :

- Justification de l'essai
 - Objectif, équivalence, supériorité
- Principes méthodologiques
 - Comparaison, causalité, significativité
- Mise en évidence du bénéfice clinique
 - Eviter les biais
 - Contrôle
 - Randomisation
 - Double insu
 - Critères de jugement
 - Principal, secondaires,
 - Cliniques, substitution, caractéristiques, composites
- Sélection des sujets avant recueil des données
 - Comparabilité, critères d'inclusion, de non inclusion
 - Stratification des groupes
- Plans expérimentaux
 - Groupes parallèles, cross over, carrés latins

PLAN (SUITE)

2) Après le recueil des données

Sélection des sujets après le recueil des données
Intention de traiter (ITT), per protocole

Expression des résultats

Paramètres numériques : RR, RRA, RRR, NST
Méta analyses Odd ratio (OR)

Analyse statistique

Fluctuations d'échantillonnage
Risque α première espèce ; petit p
Risque β deuxième espèce
Puissance de l'essai
Intervalle de confiance (IC)
Calcul du nombre de sujets
Relation entre IC et test statistique

Analyses d'essais cliniques

Analyses en sous groupes

essai non concluant, concluant

JUSTIFICATION DE L'ESSAI

L'essai apporte t-il une information utile ?

Les participants sont-ils convaincus ?

Le traitement expérimental à l'essai est-il susceptible d'apporter quelque chose de nouveau ?

I) METHODOLOGIE AVANT RECUEIL DES DONNÉES

ESSAI CLINIQUE

Un objectif et un seul :

- Comparer un nouveau traitement à un traitement témoin
- Impact du traitement sur un symptôme précis
- Détermination de la dose, de la durée de traitement
- Acceptabilité de formes galéniques
- Interactions médicamenteuses
- Utilisation chez des patients à risque
- Précautions d'utilisation

ESSAIS D'EQUIVALENCE

Les essais d'équivalence clinique (ou de non infériorité) ont pour objectif de montrer que deux traitements sont équivalents en terme d'efficacité clinique.

Attention, ne pas confondre avec Essais de Bioéquivalence (utilisés dans les dossiers de médicaments génériques) où l'équivalence ne concerne que les paramètres pharmacocinétiques.

Dans certaines situations, une avancée thérapeutique peut ne pas être une efficacité supérieure, mais simplement une plus grande facilité d'utilisation ou une meilleure tolérance.

Avantages des essais d'équivalence :

Fréquence moindre des effets secondaires

Effets secondaires moins graves

Facilité d'utilisation plus grande :

voie d'administration plus simple

réduction de posologie (*ex once daily*)

absence d'ajustement de dose

Lourdeur du traitement plus faible

Traitement médical au lieu de chirurgical

Chirurgie moins délabrante

Radiothérapie moins prolongée

Coût plus faible.

ESSAIS DE SUPERIORITE

Dans une pratique fondée sur les preuves (« *evidence based medicine* ») un nouveau traitement n'est adopté que lorsqu'il existe une preuve, issue d'essais cliniques, qu'il représente une avancée thérapeutique par rapport au traitement de référence.

NB La commission d'AMM demande de plus en plus d'essais de supériorité. De même que la Commission de transparence , pour l'ouverture au remboursement.

Il faut préciser dans l'hypothèse de l'essai :

- l'importance de la différence que l'on souhaite mettre en évidence.
- La puissance statistique de l'essai pour montrer cette différence, si elle existe.

TROIS GRANDS PRINCIPES DE LA METHODOLOGIE DES ESSAIS CLINIQUES

COMPARAISON

- Comparaison d'un traitement par rapport : à un autre, à un placebo, à rien
- Comparaison de populations comparables :
 - Définition des populations, Définition de critères d'inclusion et de non inclusion (exclusion) Il faut limiter les facteurs de variabilité.
- Comparaison dans des circonstances comparables.

CAUSALITE

La différence est-elle imputable au traitement ? Ou au hasard ?

Essai en aveugle

Tirage au sort ou randomisation

Critère de jugement fiable

SIGNIFICATIVITE

Significativité = raisonnement statistique

La différence observée est-elle significative ?

Une seule question pour une réponse de probabilité cohérente

Facteurs de variabilité (qui empêchent la comparaison)

Plus les sujets d'un groupe sont différents, plus on peut s'attendre à des réponses variables au traitement, plus il sera difficile de tirer des conclusions générales. D'où l'intérêt de constituer des groupes homogènes.

Liste des facteurs de variabilité

1 Caractéristique de l'état pathologie traité :

cadre nosologique : définition précise de l'état pathologique
forme clinique (symptomatique, évolutive, étiologique)

2 Pathologie associée + traitements correspondants

3 Caractéristiques démographiques et physiologiques des sujets : âge, sexe, poids, taille, race, niveau intellectuel, profil psychologique, position sociale, régime, activité physique, compréhension de la langue

4 Facteurs extérieurs : personnel soignant, climat, saison, environnement, hospitalisé ou ambulatoire, activité professionnelle.

5 Traitement :

Posologie, voie d'administration, horaire des prises, réalité des prises, forme galénique, biodisponibilité, observance, fidélité au rendez-vous, encadrement familial.

Exemple : Cas de la Posologie

Posologie fixe : recommandée dans un but de standardisation et donc diminution de la variabilité.

En fait il peut y avoir une augmentation de la variabilité car certains sujets reçoivent trop, d'autres pas assez. C'est fonction de la sensibilité individuelle.

Ex dose fixe contre placebo

dose contre produit de référence : dose optimale ajustée

Posologie adaptée initialement :

Fonction du poids, de la surface corporelle, gravité de la maladie, des taux plasmatiques de la substance (pharmacocinétique).

Posologie ajustée :

Par paliers progressifs prédéterminés jusqu'à l'apparition d'effets secondaires trop gênants, ou jusqu'à l'obtention de l'effet thérapeutique.

Ex : certains médicaments : anticoagulants, lithium.

Horaire des prises :

Horaire est un facteur important de l'efficacité .

Absolu ou pas par rapport aux repas.

Traitements antérieurs et associés
(qui peuvent empêcher la comparaison)

Arrêt des traitements antérieurs :

Sevrage : wash out

La durée est proportionnelle à la rémanence de l'effet thérapeutique du traitement arrêté.

Arrêt du traitement testé :

Continuation du traitement prescrit jusqu'à l'ouverture du code de randomisation, c'est-à-dire à la fin de la période d'observation du dernier malade entré dans l'essai.

Traitements associés :

Traitements associés permis ou non (critères d'exclusion)

exemple: la psychothérapie est permise pour un essai sur les psychotropes

Sélection des sujets avant l'étude (pour pouvoir comparer)

Problèmes liés au choix du sujet :

- Favoriser la mise en évidence de la différence recherchée
- Sélectionner un groupe homogène (limiter au minimum les facteurs de variabilité).
- Sélectionner un groupe représentatif permettant l'extrapolation. C'est un compromis entre l'homogénéité et la représentativité :
Si ↑ Homogénéité => ↓ Représentativité de la population réelle.
- Recrutement réaliste, en fonction des critères d'exclusion et d'inclusion.
- Respecter les impératifs éthiques :
Exemple : femme enceinte à cause des risques tératogènes.
- Définition précise des critères d'inclusion et d'exclusion.

Critères d'inclusion (critères positifs) :

Définition précise de la maladie (cas limites, cas frontières)

Formes cliniques acceptées (symptomatiques, évolutives, étiologiques)

Age, sexe

Patients ayant donné leur consentement

Aptes intellectuellement à participer (langue)

Contraception efficace pour les femmes en période d'activité génitale.

Critères de non inclusion ou exclusion (critères négatifs) :

Formes cliniques exclues, âges extrêmes

Affections associées (ex insuffisance rénale)

Traitements associés qu'il faut préciser (ex : incompatibles)

Patients non « compliants », intellectuellement inaptes, itinérants (gens du voyage)

Femme enceinte ou susceptible de l'être

Race (problème métabolique).

Mise en évidence du bénéfice clinique

La mise en évidence du bénéfice clinique se fait :

- **par des critères de jugement** ou d'évaluation (*endpoint / outcome*) qui mettent en évidence l'effet recherché.
- **En évitant les phénomènes qui interfèrent** dans la relation entre le traitement et le critère de jugement. Ces phénomènes sont **les facteurs de confusion et les biais**.

Facteurs de confusion (ou facteurs confondants) :

1) Evolution naturelle de la maladie

Exemple : Après une prise d'eau minérale pendant 5 jours, 78 % des sujets présentant un rhume banal n'ont plus de symptômes.

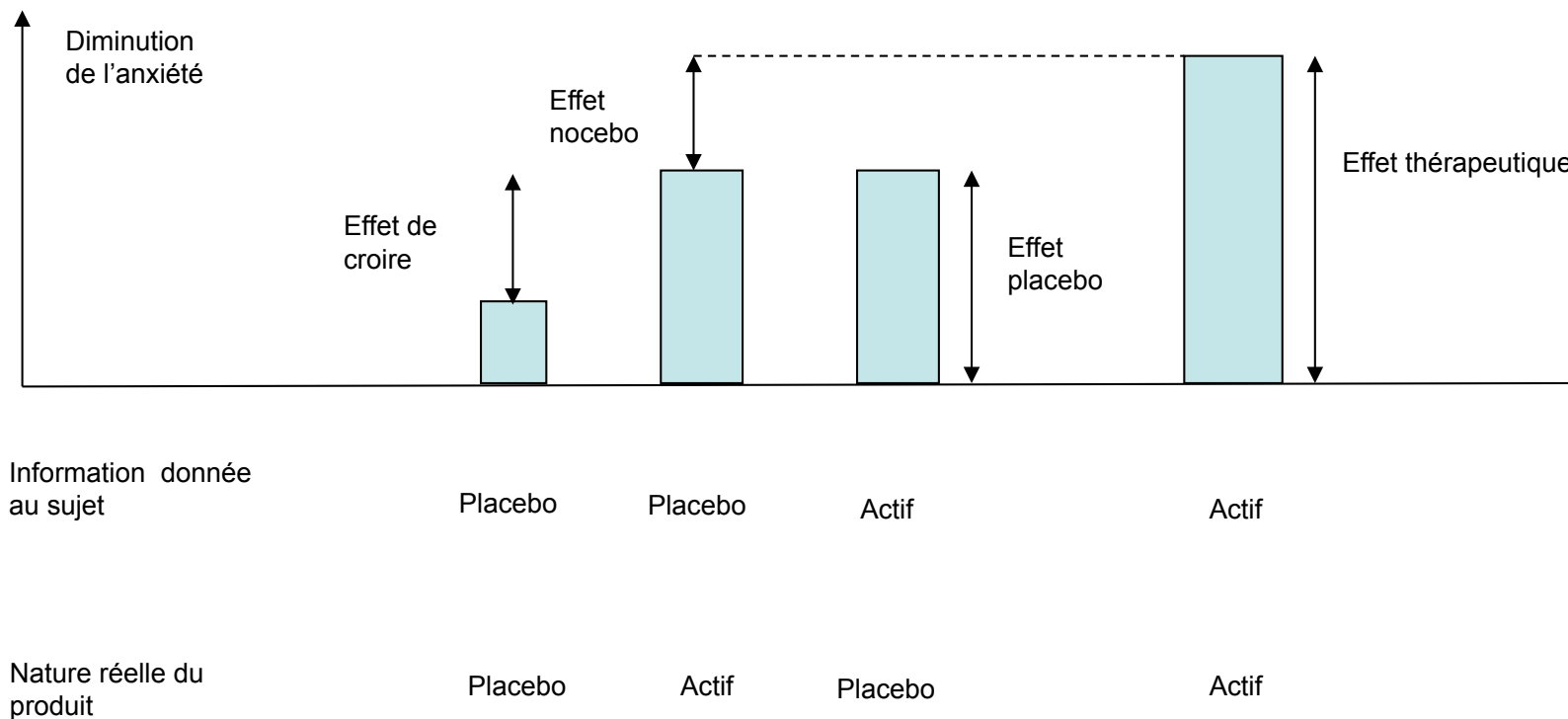
Cela n'apporte pas la preuve de l'efficacité de l'eau minérale. L'effet observé correspond à l'évolution naturelle du rhume.

L'évolution naturelle peut être confondue avec un effet du traitement. Elle est appelée facteur confondant.

Facteurs de confusion (ou facteurs confondants) :

2) Effet placebo : la prise en charge médicale peut améliorer l'état d'un patient en dehors de tout traitement actif. L'effet placebo peut être expliqué par des effets de conditionnement, d'auto ou d'hétéro suggestion.

Démonstration expérimentale de l'effet placebo (stress et anxiolytiques)



Les biais

Pour un essai thérapeutique, un biais est un facteur qui va conduire à un résultat non conforme à la réalité, c'est-à-dire un résultat biaisé ou erroné.

Exemple : risque de conclure à l'efficacité d'un traitement qui en est en réalité dépourvu.

Liste des principaux biais :

Biais de confusion : biais introduit par les facteurs de confusion.

Le risque de biais de confusion est supprimé par l'utilisation d'un **groupe contrôle**.

Biais de sélection : les deux groupes ne sont pas comparable initialement.

La **randomisation** empêche la survenue d'un biais de sélection.

Biais de suivi (ou de réalisation) : destruction de la comparabilité des groupes au cours du suivi. Ce biais est évité par l'utilisation du **double insu**.

Biais d'attrition : retrait de certains patients de l'analyse. Ce biais est contrôlé par **l'analyse en intention de traiter**.

Biais d'évaluation : le critère de jugement n'est pas recherché de la même façon entre les groupes. Le double insu supprime le risque de biais.

Comparaison à un Groupe contrôle (éviter les facteurs de confusion)

Conséquence de la prise d'un traitement (ce qui est observé sur le groupe traité) = Effet du traitement (ce qui est à déterminer) + Effet des facteurs confondants (ce qui pollue le résultat)

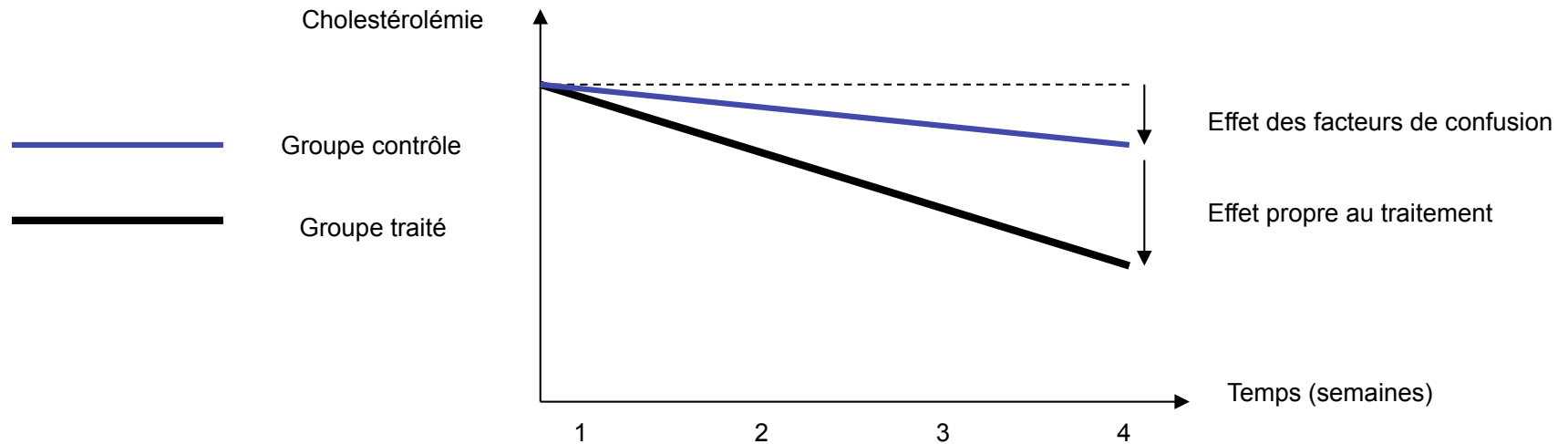
Solution : travailler avec un groupe contrôle (ou témoin), qui ne reçoit pas le traitement étudié, mais qui va subir les mêmes influences en provenance des facteurs de confusion que le groupe traité (avec le traitement étudié). Le contrôle permet la comparaison.

Conséquence de la prise d'un non traitement (ce qui est observé sur le groupe contrôle) = Pas de traitement + Effet des facteurs confondants (ce qui pollue le résultat)

Du fait de l'existence d'un groupe contrôle, l'essai est dit « contrôlé ». C'est un essai comparatif qui compare un groupe traité à un groupe contrôle non traité, afin de déduire l'effet propre d'un traitement.

Illustration de l'intérêt d'un groupe contrôle

Détermination de l'effet hypocholestérolémiant propre à un traitement.



Le groupe contrôle peut prendre **un traitement de référence** si on cherche à mesurer l'efficacité « relative » par rapport à un médicament disponible.

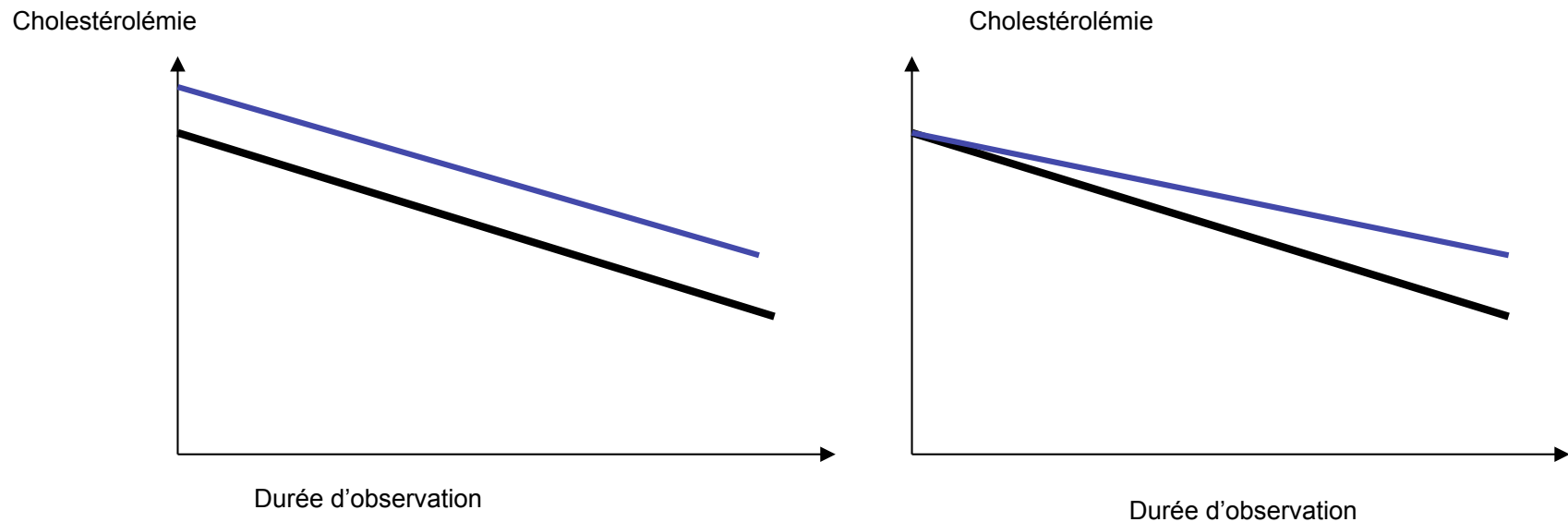
Si la question est d'évaluer l'efficacité « absolue », alors le groupe contrôle ne doit recevoir aucun traitement actif. Il va prendre un médicament « placebo » (même forme mais pas de principe actif) pour qu'il y ait : 1) la même prise en charge médicale, 2) que l'effet placebo soit intégré dans les facteurs de confusion et 3) qu'il y ait la possibilité, par la galénique, de faire un double insu.

Randomisation

(éviter le biais de sélection, pouvoir évaluer la cause liée au traitement)

Notion de biais de sélection

Une condition évidente pour que la différence entre deux groupes représente uniquement l'effet du traitement est que les deux groupes soient comparables avant l'application du traitement.



Si les deux groupes n'ont pas initialement la même cholestérolémie moyenne, la différence observée après traitement peut être simplement la répercussion de la différence initiale. Pour éviter cela il faut que les deux groupes soient identiques au début de l'essai. **Si il y a une différence initiale, il y a un biais de sélection.**

La Randomisation, qui répond au principe de causalité.

Pour éviter un biais de sélection, la nature du traitement que reçoit un patient ne doit dépendre d'aucun autre facteur susceptible d'influencer le résultat (gravité, caractéristiques du patient, contexte des soins) **Seule une attribution au hasard** peut garantir la totale indépendance de la nature du traitement donné vis-à-vis de ces facteurs. Pour obtenir cette indépendance les traitements sont attribués aléatoirement aux patients. Cette allocation aléatoire s'appelle « randomisation ».

Randomisation par tirage au sort honnête.

Réalisation : dés, pile ou face, papiers dans un chapeau

Tables de nombres au hasard (Ex Tables de Fischer & Yates)

Exemples pour deux traitements :

0,1,2,3,4 pour traitement A

5,6,7,8,9 pour traitement B

279 354 608 : 9 malades

Pour trois traitements

1,2,3, Traitement A

4,5,6, Traitement B

7,8,9 Traitement C

752 196 328 375 12 malades

Le moment du tirage au sort : cela doit être préparé à l'avance.

La liste pré établie des traitements servira à préparer les conditionnements numérotés de médicaments ou de placebo.

Cela permettra de préparer les enveloppes de randomisation (codage) à ouvrir en cas d'urgence (par exemple pharmacovigilance).

<p style="text-align: center;">DOUBLE INSU (Eviter les biais de réalisation ou d'évaluation)</p>

Biais de réalisation

La comparabilité doit être maintenue au cours de l'essai et que rien ne vienne la détruire durant le suivi (follow up) ou lors de la mesure du critère de jugement.

Ex : un groupe qui reçoit plus de traitements concomitants qu'un autre.

Ex : la mesure du critère ne s'effectue pas de la même façon dans les groupes.

Double insu (double blind ou double aveugle) **qui répond au principe de causalité**

Ce principe consiste à faire que tous les patients, quelle que soit l'appartenance à un des groupes, apparaissent identiques.

Ceci est obtenu en ne révélant pas la nature exacte du traitement reçu par les patients.

Ni le patient, ni le médecin ne connaissent le médicament ou le placebo.

NB : Simple insu : seul le patient ne connaît pas le produit qu'il prend.

Les deux médicaments, traitement et placebo, sont identiques dans leur apparence (forme galénique et conditionnement).

Le recours au double insu est particulièrement important quand le critère de jugement est de nature subjective (Ex : l'évaluation par le médecin).

PLANS EXPERIMENTAUX : GROUPE PARALLELES

Groupe paralleles : il y a autant de groupe qu'il y a de traitement à comparer.

Exemples : « 3 groupe » ou « 3 bras »

placebo

produit à tester

produit de référence (comparateur)

Avantages :

Simplicité d'organisation

Simplicité d'analyse

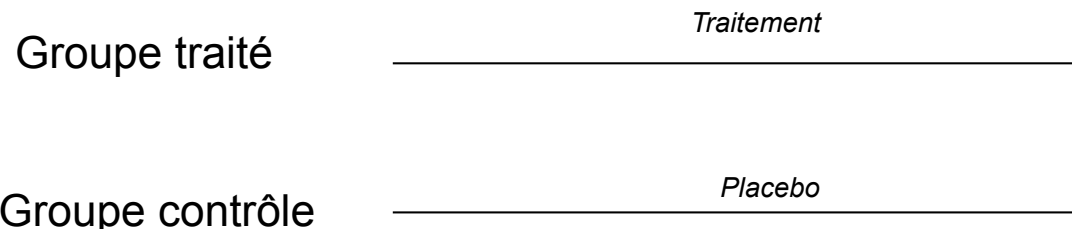
Nécessité de bons critères d'inclusion (homogénéité des groupe) et bonne randomisation pour assurer la comparabilité des groupe même s'il y a des différences inter individuelles.

Inconvénients :

Variabilité des résultats entre les sujets est maximale.

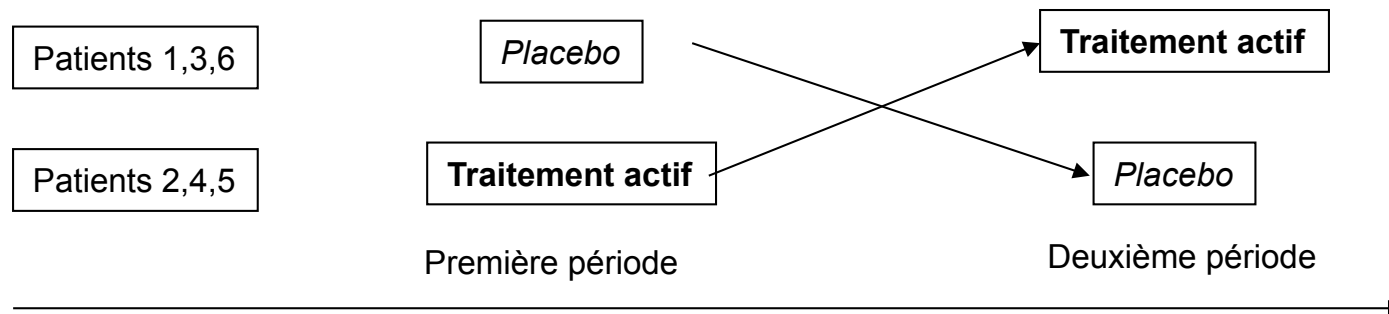
Nombre de sujets important.

Risque d'erreur due à la constitution de groupe non comparables.



PLANS EXPERIMENTAUX : ESSAI CROISÉ (CROSS OVER)

Tous les patients reçoivent le traitement étudié et le traitement contrôle dans un ordre aléatoire. L'essai croisé permet de comparer les traitements au sein du même malade (comparaison intra individuelle). Le critère de jugement est mesuré après chaque période.



Avantages :

Moins de sujets

Variation intra individuelle (même patient) < variations inter individuelles (entre patients)

Inconvénients :

Etat basal stable

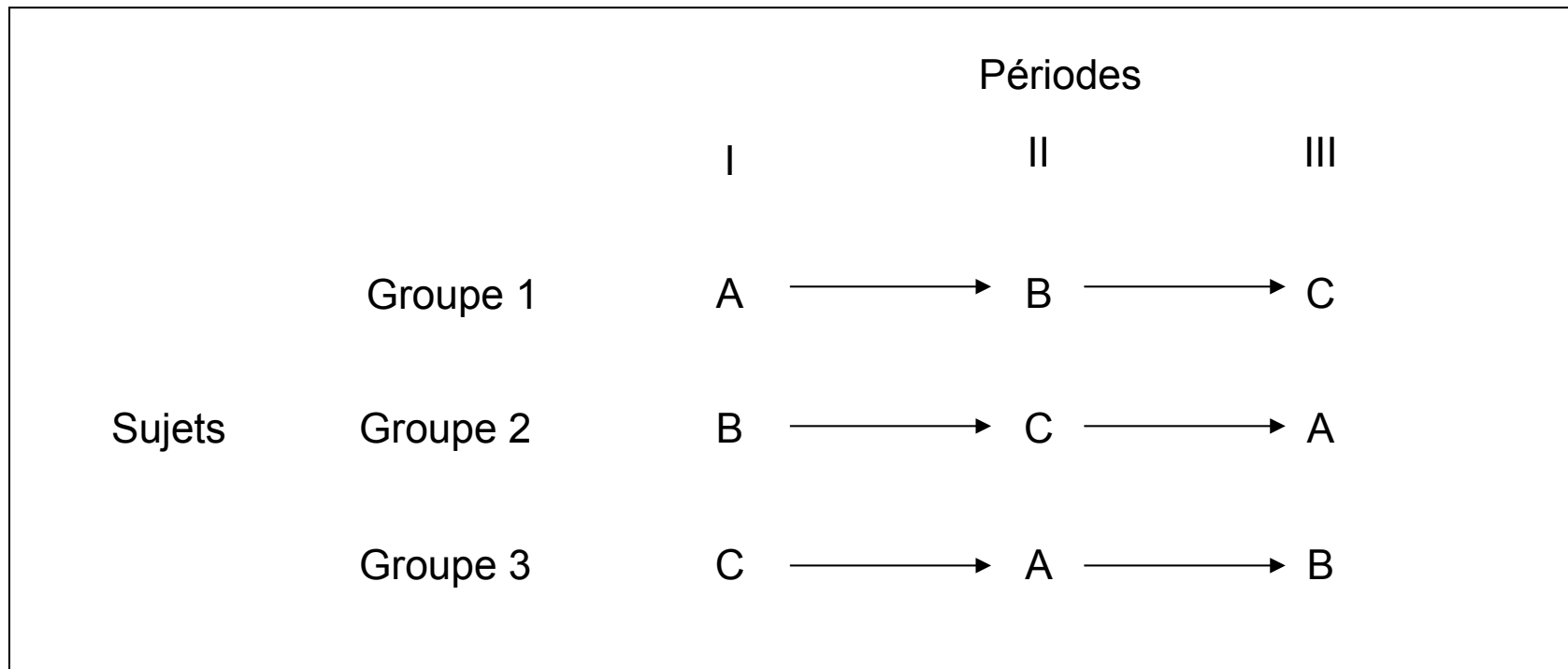
Perte d'intérêt lorsque l'essai est relativement prolongé; nécessité de période thérapeutique brève. Il faut qu'aucun des traitements administrés au cours d'une des phases de l'essai n'ait d'influence sur les résultats de la phase suivante.

Effet tardif ou rémanent (irréversibles ou retardés) : dans ces cas, effet de sevrage intermédiaire avec wash out.

PLANS EXPERIMENTAUX : CARRES LATINS

Le nombre de sujets est égale à un multiple du nombre de traitements.
Chaque sujet reçoit successivement chacun des traitements.

Chacun des traitements est administré à un des sujets à chacun des périodes de l'essai.



Influence de 3 facteurs : sujets, périodes, traitements.

STRATIFICATION (SOUS GROUPES AVANT ETUDE)

Essais stratifiés : Ces essais étudient simultanément deux ou plusieurs strates qui sont l'équivalent de sous-groupes.

Strates : sous groupes de malades au pronostic comparable

Strate 1 : pression artérielle diastolique < 100 mm Hg

Strate 2 : pression artérielle diastolique > 100 mm Hg.

Avantages de la stratification :

- Limitation de différences due à la constitution des groupes comportant des proportions différentes de malades dissemblables.
- Recrutement plus large et plus hétérogène, donc plus facile.
- Représentativité du groupe plus large => plus grande généralisation des résultats.
- Comparaison statistiques plus sensibles. La variabilité à l'intérieur des strates est diminuée.
- Possibilité d'étudier en plus de l'effet des traitements, l'influence des facteurs pronostiques et leur interaction.

Inconvénients de la stratification :

- Limitation du nombre de strates, du nombre de facteurs que l'on peut prendre en compte indépendamment des uns des autres.

CRITERES DE JUGEMENT

Le critère de jugement (ou d'évaluation) appelé « **endpoint** » ou « **outcome** » va permettre une quantification, c'est-à-dire une mise en évidence numérique de l'effet du traitement.

Les critères de jugement sont des variables dont la valeur numérique est susceptible de changer sous l'effet du traitement.

Le protocole de l'essai devra définir :

- Un critère principal (efficacité)
- Éventuellement des critères secondaires de bénéfice clinique (ex tolérance)

CRITERE PRINCIPAL

Il permet de documenter l'efficacité du médicament. C'est sur ce critère que sera portée uniquement la conclusion globale de l'essai, quels que soient les résultats obtenus avec les critères secondaires. C'est un critère choisi a priori.

L'utilisation du critère principal permet également le calcul du nombre de sujets nécessaires à inclure dans l'essai.

NB Dans certains cas exceptionnels, il est possible de choisir 2 critères principaux, mais il y a nécessité de différence statistiquement significative retrouvée sur les deux tests. L'utilisation de deux critères doit être mentionnée dans le protocole.

CRITERES SECONDAIRES

Objectif : Tentation d'utiliser ces critères secondaires lorsque aucune différence n'est mise en évidence sur le critère principal.

Conséquence :

Augmentation du nombre de tests statistiques qui vont être effectués.

Possibilité plus grande que 5% (p) de trouver à tort une différence statistiquement significative.

Exemple : les calculs montrent qu'avec la réalisation de 5 critères, on a 23 % de chances d'obtenir, sur au moins un test, une différence statistiquement significative entre les deux médicaments dont l'efficacité serait identique.

Attention :

La présentation des critères secondaires d'une étude clinique dans un document promotionnel de produit fait l'objet d'une recommandation de l'AFSSAPS (1998).

Cette recommandation stipule que les résultats des critères secondaires ne peuvent être présentés qu'accompagnés de ceux des critères principaux.

TYPES DE CRITERES DE JUGEMENT

CRITERES CLINIQUES

Ces critères correspondent à un événement clinique ou un symptôme d'importance pour le patient.

Exemples :

*décès,
accident vasculaire cérébral,
douleur,
hospitalisation.*

L'objectif du traitement est d'agir sur ces critères cliniques.

La pertinence clinique du critère doit être justifiée dans le protocole.

Caractéristiques des données cliniques ou biologiques (critères de qualité) :

Intègre : existence même de la donnée.

Documentée : originale, identifiable, interprétable et disponible

Conforme : recueillie en conformité avec le protocole

Consistante : similarité entre données sources et cahier d'observation

Exacte : ne pas présenter de biais

CRITERES DE SUBSTITUTION

Ils correspondent à une mesure de laboratoire ou à un signe physique qui se substitue aux critères cliniques.

Ils permettent également d'expliquer le mécanisme d'action du médicament.
Ils doivent être idéalement corrélés aux critères cliniques. Une modification du critère de substitution entraîne une modification du critère clinique.

Exemples :

Taux de cholestérol

Tension artérielle

Masse tumorale

Densité osseuse

Vitesse de sédimentation

Erosion gastrique.

Le choix est plus facile :

- Action plus rapide du médicament
- Différence entre deux médicaments plus grande sur les critères de substitution, ce qui donne une plus grande puissance et entraîne moins de patients.

Caractéristiques d'un critère de jugement :

Fiable, Reproductible, Facilement disponible

Évalué comme un effet dose / réponse, c'est-à-dire corrélé positivement à une augmentation de la probabilité de survenue d'un événement clinique.

Consistant : cette qualité correspond à la reproductibilité des mesures faites par le même observateur et à la concordance des mesures faites par des observateurs différents.

Stable : un critère mesuré est stable s'il ne fluctue pas trop dans le temps. Dans le cas contraire il est labile. *Idéal : même observateur entraîné.*

Pour augmenter la stabilité => augmentation de la standardisation des techniques ; augmentation de la standardisation des conditions de mesure (*horaires, température extérieure, circonstances physiologiques*) .

L'instrument de mesure du critère de jugement doit être :

Sensible : c'est-à-dire qu'un résultat positif devrait détecter les patients ayant une augmentation du risque.

Spécifique : c'est-à-dire qu'un résultat négatif devrait exclure les patients sans augmentation du risque.

Valeur prédictive positive : si le test est positif, il y ait bien l'affection.

Valeur prédictive négative : si le test est négatif, il n'y ait pas l'affection.

**Critères de jugement (instrument de mesure)
: Sensibilité et Spécificité**

Méthode de mesure	Maladie présente	Maladie absente
Signe présent (critère) +	A (VP vrais positifs)	B (FP Faux positifs)
Signe absent (Critère) -	C (FN faux négatifs)	D (VN vrais négatifs)

- **Sensibilité** = $A / A + C$ le plus possible de diagnostic positif
- **Spécificité** = $D / B + D$ diagnostiquer le moins souvent possible les cas n'ayant pas l'affection
- **Valeur prédictive positive** = $A / A + B$ quand le test est +, on veut qu'il y ait bien l'affection.
- **Valeur prédictive négative** = $D / C + D$ quand le test est -, on veut qu'il n'y ait pas l'affection.

CRITERES COMPOSITES

Un critère composite prend en compte plusieurs types d'événements cliniques.

Exemple :

décès,
infarctus,
hospitalisation

La survenue de ces événements pré définis est comptabilisée une seule fois par patient, même si un second événement survient plus tard chez ce même patient. C'est pour cette raison que la somme de chacun des événements est généralement supérieure au nombre de survenue d'événements du critère composite.

Avantages : augmenter la puissance de l'essai tout en nécessitant moins de patients, par rapport à un essai ayant un critère d'évaluation simple, pour mettre en évidence une différence significative.

Illustration : risque de 5 % avec une puissance de 90 %. Le nombre de patients nécessaire est de 1400 pour mettre en évidence une différence significative avec un taux d'événements de 20 %. C'est seulement 700 patients avec un taux d'événements de 40 %.

Inconvénients : Choix des composantes du critère composite (action homogène), mais homogénéité d'action pas toujours retrouvée.

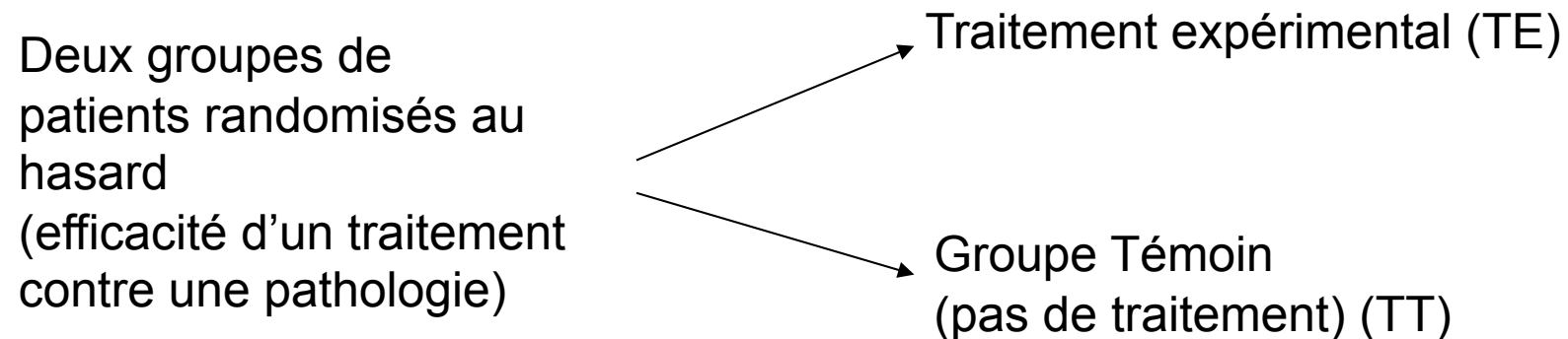
Interprétation assez délicate : nécessité réglementaire (AFSSAPS) d'analyser également de façon individuelles les différentes composantes.

II) METHODOLOGIE APRES RECUEIL DES DONNEES

EXPRESSION DES RESULTATS ET ANALYSE STATISTIQUE (SIGNIFICATIVITE)

Dans un essai thérapeutique, **la mise en évidence de l'effet du traitement étudié repose sur l'observation d'une différence** au niveau du critère de jugement entre un groupe de patients recevant ce traitement et un groupe ne le recevant pas.

Lorsque le critère de jugement est la survenue d'un événement clinique, **l'efficacité du traitement se traduit par une fréquence moindre de l'événement dans le groupe traité par rapport au groupe contrôle.**



SELECTION DES SUJETS A LA FIN DE L'ETUDE

Etude en intention de traiter (ITT) :

Tous les participants à l'étude restent étudiés dans le groupe auquel ils ont été assignés, même s'ils n'ont pas entièrement achevé le protocole.

Analyse qui se rapproche de la réalité.

Etude *per protocole* :

Les patients sont exclus de l'analyse lorsque le protocole initial n'est pas totalement respecté. Analyse qui se rapproche des conditions idéales.

*Exemple illustré : 2 groupes de 100 patients, A traités, B placebo
A : 50 sorties pour intolérance; 25 améliorations; 25 stagnations.
B : 0 sorties ; 25 améliorations ; 75 stagnations.*

Si on fait une analyse per protocole :

A : 50 % d'amélioration (25/50)

B : 25 % d'amélioration (25/100)

Si on fait une analyse en Intention de traiter (ITT)

A : 25 % (25 /100)

B : 25 % (25 / 100)

Les résultats d'une étude peuvent être donnés à la fois en ITT et en *per protocole*. Cela illustre le biais d'attrition.

PARAMETRES NUMERIQUES D'EXPRESSION DES RESULTATS

	Malades (événements)	Sains (non événements)
Groupe expérimental (n=30)	A = 1	B = 29
Groupe témoin (n=30)	C =9	D =21

Calcul du Risque Relatif (RR) : c'est le risque de tomber malade dans le groupe expérimental par rapport à celui de tomber malade dans le groupe témoin.

- TEE (taux d'événements dans le groupe expérimental)

$$TEE = a/(a+b) = 1 / (1+29) = 1/30 = 0,033$$

C'est le risque de tomber malade si l'on fait partie du groupe expérimental

- TET (taux d'événements dans le groupe témoin)

$$TEE = c/(c+d) = 9 / (9+21) = 9/30 = 0,3$$

C'est le risque de tomber malade si l'on fait partie du groupe témoin

Le Risque Relatif (RR) est le rapport de ces deux taux.

$$RR = TEE / TET = 0,033 / 0,30 = 0,11 \text{ Soit un risque de } 11 \%$$

Si le traitement est efficace, ce risque va être inférieur à 1. Plus il est faible, plus le traitement est efficace. Si $RR > 1$, le produit étudié a augmenté le risque de tomber malade.

Calcul de la Réduction du Risque Absolu (RAA) : c'est la réduction du risque entre le fait de tomber malade dans le groupe témoin et le fait de tomber malade dans le groupe expérimental. Le RAA est exprimé en pourcentage.

$$\text{RAA (\%)} = \text{TET} - \text{TEE} = 30 - 3,3 = 26,7 \%$$

Ce résultat a peu de signification si on ne sait pas que le taux d'événements dans le groupe témoin est de 30 % (si on ne connaît pas le TEE).

Calcul de la Réduction du Risque Relatif (RRR) : c'est la réduction du risque entre le fait de tomber malade dans le groupe témoin et le fait de tomber malade dans le groupe expérimental, rapporté au risque de tomber malade dans le groupe témoin. Le RRR est exprimé en pourcentage.

$$\text{RRR (\%)} = (\text{TEE} - \text{TET}) / \text{TET} = 1 - \text{RR} = 1 - 0,11 = 0,89 \text{ soit } 89 \% \text{ de réduction}$$

C'est-à-dire la mesure en quoi le traitement modifie le risque par rapport à la valeur neutre du risque en dehors de toute intervention

NST : Nombre de sujets à traiter pour prévenir un événement supplémentaire.

$$\text{NST} = 1 / (\text{TET} - \text{TEE}) = 1 / \text{RAA} = 1 / 0,267 = 4$$

Il faut 4 patients de plus pour prévenir un événement supplémentaire, c'est-à-dire pour empêcher la survenue d'un autre cas de maladie.

EXPRESSION DES RESULTATS D'ETUDES (META ANALYSES)

Méta analyses : revue de synthèse de multiples essais randomisés.

ODD RATIO (relative odds ou rapport de côtes)

L'odd (ou côte) d'un patient (dans un groupe donné) est le nombre de patients qui présentent l'événement par rapport au nombre de patients qui ne le présentent pas.

Exemple odde = $a/b = 1/29 = 0,034$; oddT = $c/d = 9/21 = 0,43$

Odd ratio (OR) = oddE/oddT = $0,034 / 0,43 = 0,08$

3 cas pour l'odd ratio (OR) :

OR = 1 La proportion de gens malades par rapport aux sains ne varie pas entre le groupe expérimental et le groupe traité => le traitement n'a pas d'effet.

OR > 1 Le traitement augmente la proportion de malades. Il est néfaste.

OR < 1 Le traitement diminue la proportion de malades. Il est efficace. Plus le OR est faible, plus le traitement est efficace.

Attention : ne pas confondre RR et OR :

RR : nombre de patients malades, par rapport au nombre total

OR : nombre de patients malades, par rapport au nombre de non malades.

ANALYSE STATISTIQUE (SIGNIFICATIVITE)

FLUCTUATIONS D'ECHANTILLONNAGE

La survenue d'un événement clinique chez un patient est en partie imprévisible. Cela s'apparente donc à un phénomène aléatoire. Pour un sujet donné, il est impossible de prévoir à coup sûr la survenue ou non de l'événement.

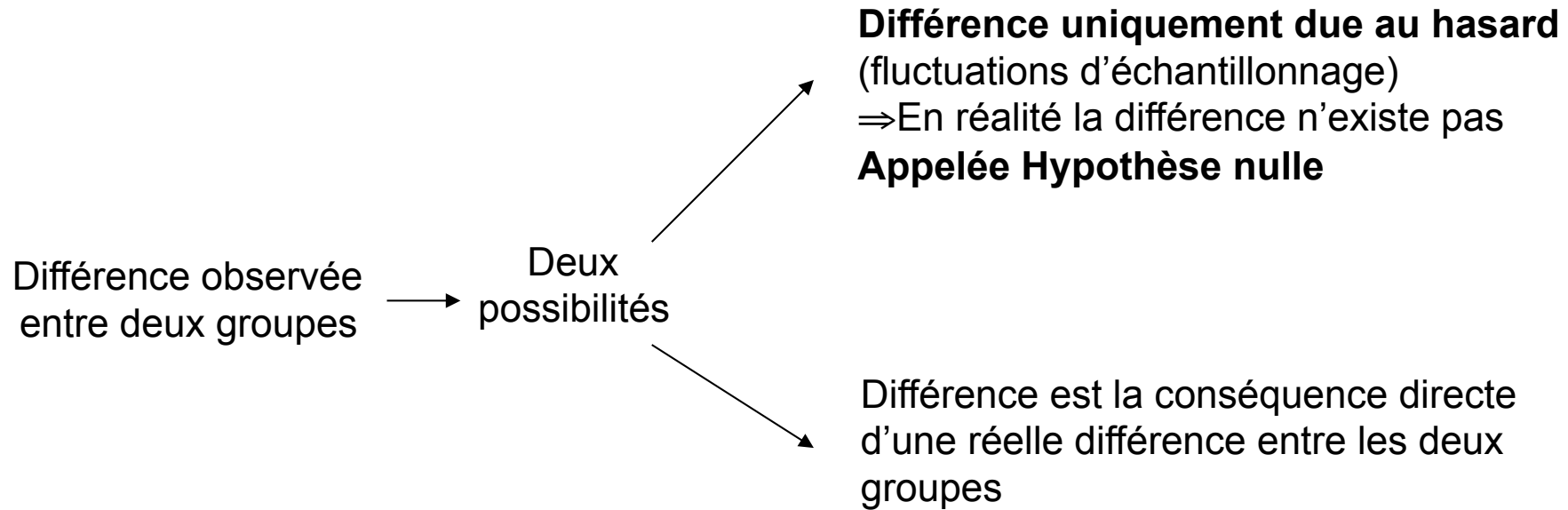
Exemple : la survenue sur une période de 5 ans d'un événement cardiovasculaire chez un sujet hypercholestérolémique est imprévisible.

	Nombre d'événements observés à 5 ans
50 patients p=10%	7 (14 %)
50 patients p=10%	5 (10 %)
50 patients p=10%	5 (10 %)
50 patients p=10%	3 (6 %)

Les pourcentages observés diffèrent entre les groupes : ce sont les **fluctuations aléatoires d'échantillonnage**.

Les valeurs vont fluctuer autour de la vraie valeur. Comme dans ces groupes, tous les patients ont le même risque ces différences observées sont dues au hasard.

RISQUE α PREMIERE ESPECE



S'il n'y a pas de moyen de faire la part des choses entre les deux possibilités, Alors on ne peut pas tirer de conclusion en pratique.

La solution à ce dilemme est apportée par le **test statistique (test d'hypothèse)** qui consiste à rejeter l'hypothèse nulle.

TEST STATISTIQUE = RISQUE α , PETIT P

Devant une différence observée, le test statistique permet de calculer la probabilité d'observer ce résultat, si en réalité il n'y a pas de différence entre les deux groupes. Cette probabilité s'appelle **p**.

C'est la probabilité que les fluctuations aléatoires d'échantillonnages donnent la différence observée, alors qu'il n'y a pas de différence en réalité.

Le petit p est une quantification du risque de faire une erreur de première espèce (appelé risque α) si l'on décidait de conclure à l'existence d'une différence entre les deux groupes.

Le risque α de première espèce, c'est le risque d'énoncer à tort une conclusion positive, c'est-à-dire de conclure à une différence significative alors qu'elle n'existe pas.

PETIT P ET SIGNIFICATIVITE STATISTIQUE

En pratique, on ne portera la conclusion d'une différence entre les deux groupes, que si le risque de se tromper est suffisamment petit.

Il a été convenu que le risque acceptable d'erreur est de 5 %.

Conséquence :

Devant une différence observée, on ne conclura à l'existence de cette différence seulement si le risque de se tromper est inférieur à 5 %, c'est-à-dire si $p < 5\%$ ou $p < 0,05$.

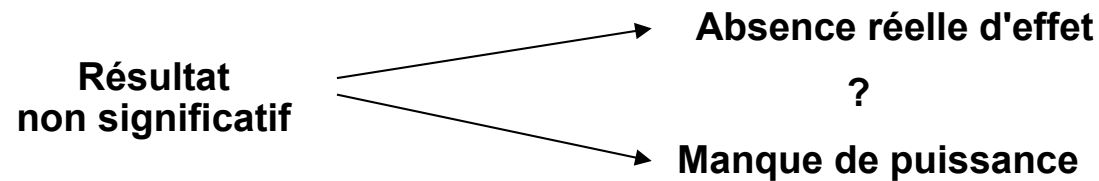
On parle aussi de **seuil** de significativité.

Signification statistique :

$p < 5\%$: c'est une différence « **statistiquement significative** ». La différence est suffisamment importante par rapport aux fluctuations d'échantillonnage.

$P > 5\%$: c'est une différence dite « **non statistiquement significative** ». La différence n'est pas suffisamment importante par rapport aux fluctuations d'échantillonnage pour pouvoir exclure qu'elle soit un artéfact dû au hasard.

DIFFERENCE NON SIGNIFICATIVE



Attention : Une différence non significative n'est pas synonyme d'absence d'effets.

La comparaison est peut être insuffisamment puissante pour mettre en évidence une différence qui existe.

L'absence de preuves (ns) n'est pas la preuve de l'absence (d'effets).

D'autres valeurs de seuil peuvent être utilisées, plus contraignantes. Ex $p=0,01$ (1%)

Le seuil de 5 % n'est pas totalement négligeable.

Exemple 4000 spécialités dans le VIDAL avec un seul essai thérapeutique.

Avec un risque de 5 %, cela fait 200 spécialités qui peuvent être enregistrées à tort.

RISQUE β DEUXIÈME ESPECE ET PUISSANCE

Si la différence observée n'est pas significative au seuil α choisi (p) :

- Soit il n'existe pas de différence,
- Soit cette différence n'est pas détectée au test.

C'est pour différencier ces deux conclusions que la connaissance du risque de deuxième espèce est indispensable. Ce risque est appelé risque β .

C'est le risque d'énoncer à tort une conclusion négative, c'est-à-dire de conclure à l'équivalence de deux échantillons qui sont en réalité différents (faux résultat négatif).

On considère le plus souvent le risque $1 - \beta$ encore appelé « Puissance du test ».

La puissance est donc la probabilité d'obtenir un vrai résultat positif, c'est-à-dire mettre en évidence l'efficacité d'un traitement réellement efficace.

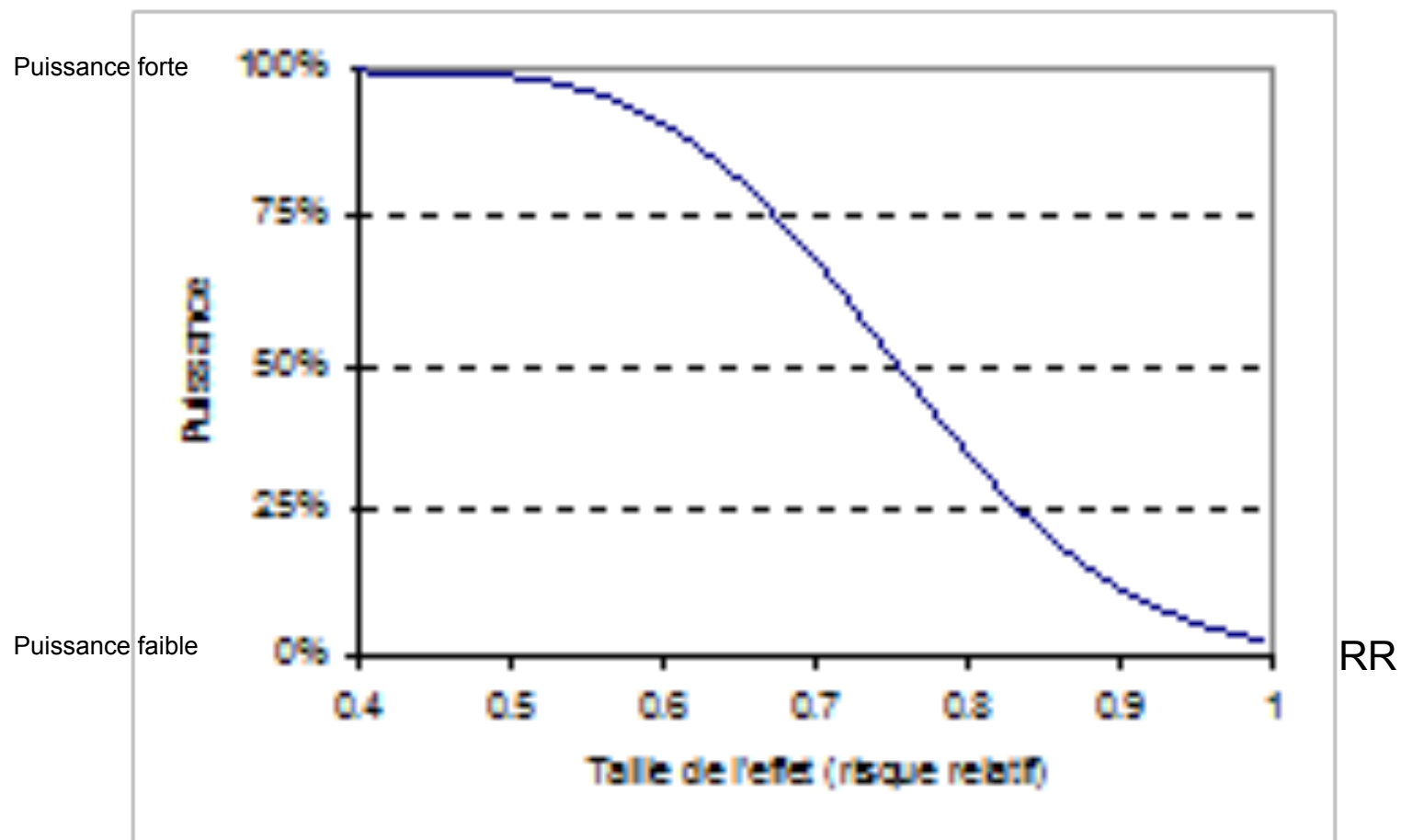
Exemple d'illustration :

La puissance est similaire au grossissement d'un microscope.

Un grossissement suffisant (puissance) est nécessaire pour montrer que deux points très proches l'un de l'autre, mais cependant séparés, sont distincts.

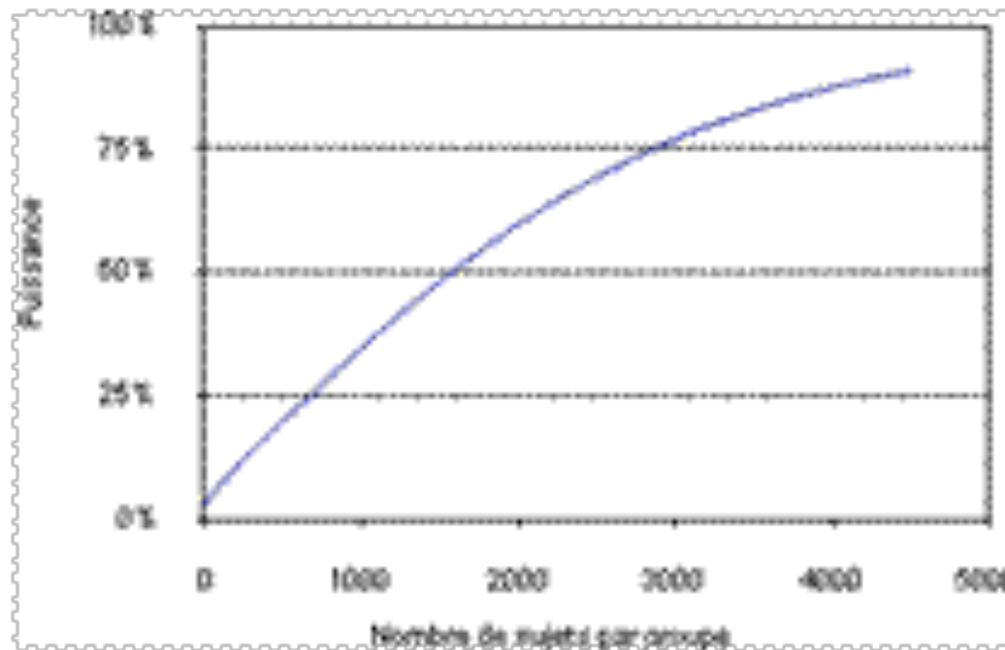
Avec un grossissement insuffisant (manque de puissance), ces deux points ne paraissent faire qu'un et on en distingue pas la différence.

RELATION ENTRE LA PUISSANCE ET LA TAILLE DE L'EFFET



Plus l'effet du traitement est faible, plus il faut de la puissance statistique pour le mettre en évidence. Ce paramètre n'est pas contrôlable par l'investigateur, c'est une caractéristique du traitement étudié, en quelque sorte sa puissance pharmacologique ou thérapeutique.

RELATION ENTRE LA PUISSANCE ET LE NOMBRE DE SUJETS



Risque de
base de 10 %
et RR de 0,8

Plus le nombre de patients est important, plus l'essai est puissant.

L'effectif est le paramètre sur lequel l'investigateur (le promoteur) peut agir pour contrôler la puissance de son essai.

En particulier lorsque l'effet recherché est petit, il est nécessaire d'inclure un grand nombre de patients.

En revanche, un effectif plus faible est suffisante pour mettre en évidence des effets conséquents.

CALCUL DU NOMBRE DE SUJETS

Éléments pour calculer le NSN (nombre de sujets nécessaires) :

- Critère principal
- Variance du critère principal
- Risques α et β
- Caractère du test :
 - A est-il différent de B
 - A est-il supérieur à B
- Différence minimale cliniquement intéressante du critère principal que l'on souhaite mettre en évidence.

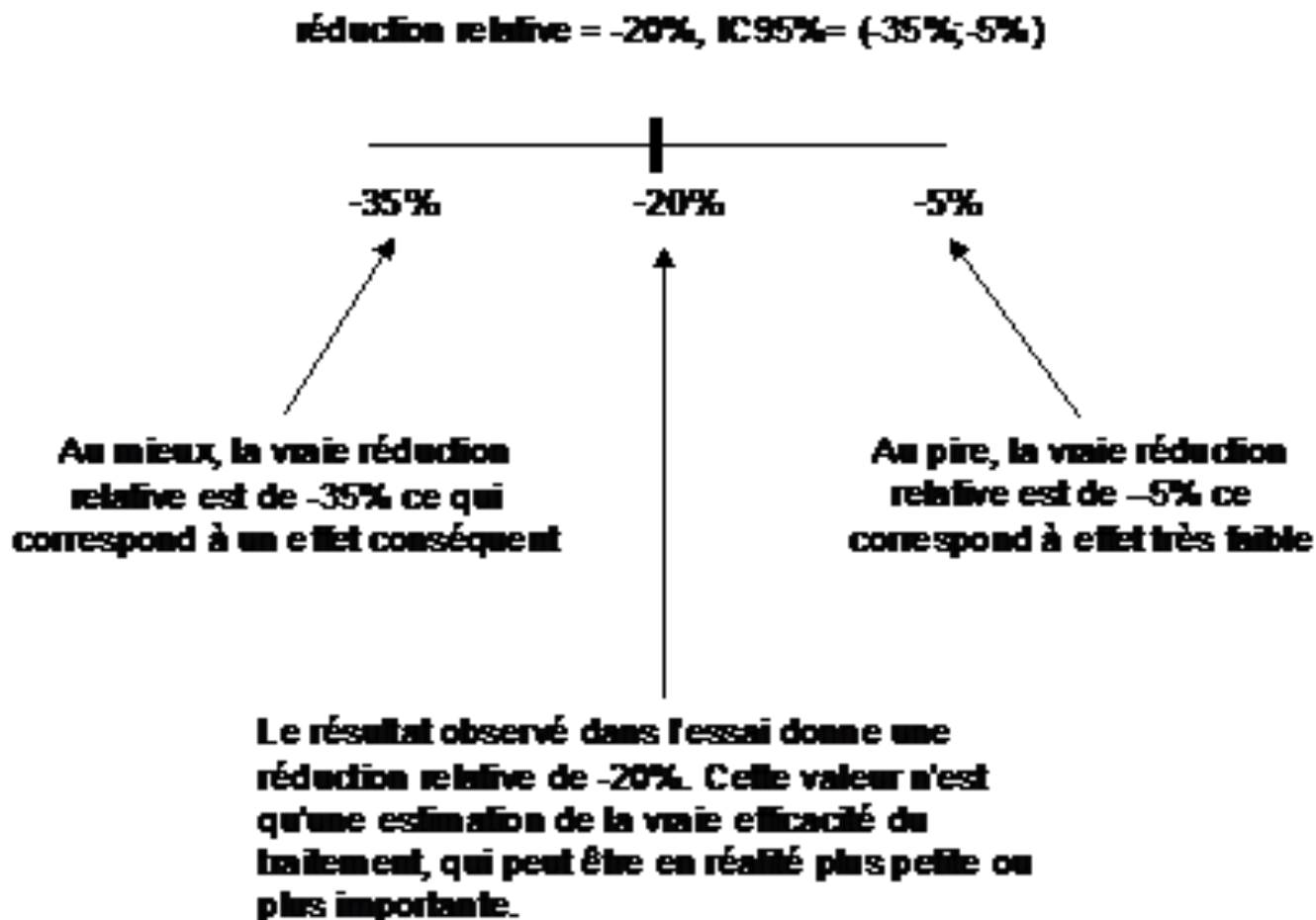
= > FORMULES MATHÉMATIQUES DE CALCUL

Ce nombre de sujets est déterminé a priori dans le protocole afin de garantir la puissance de l'effet.

INTERVALLE DE CONFIANCE

L'intervalle de confiance (IC) est un intervalle de valeurs qui a 95 % de chances de contenir la véritable valeur du paramètre estimé. Il permet de visualiser la précision avec laquelle l'effet du traitement est connu.

Exemple : Réduction de mortalité de 20 % avec IC 95 % de (-35% : -5%)



CALCUL DE L'INTERVALLE DE CONFIANCE

L'IC est calculé à partir de la valeur observée du critère étudié (*exemple : différence entre deux taux d'événements*) et l'écart type de l'estimation.

$$\text{IC 95 \%} = [(\text{RRA} - 1,96 \text{ ET}) ; (\text{RRA} + 1,96 \text{ ET})]$$

Illustration :

Essai randomisé d'un vaccin anti coquelucheux versus placebo.

1670 nourrissons vaccinés : 71 cas de coqueluche

1665 nourrissons non vaccinés : 240 cas de coqueluche

Ecart type = 0,99

Calculer TEE ? TET ? RRA ? Calculer IC 95 %

Faire la représentation graphique ?

Solution :

TEE = $71/1670=4,3 \%$; TET = $240/1665=14,4 \%$

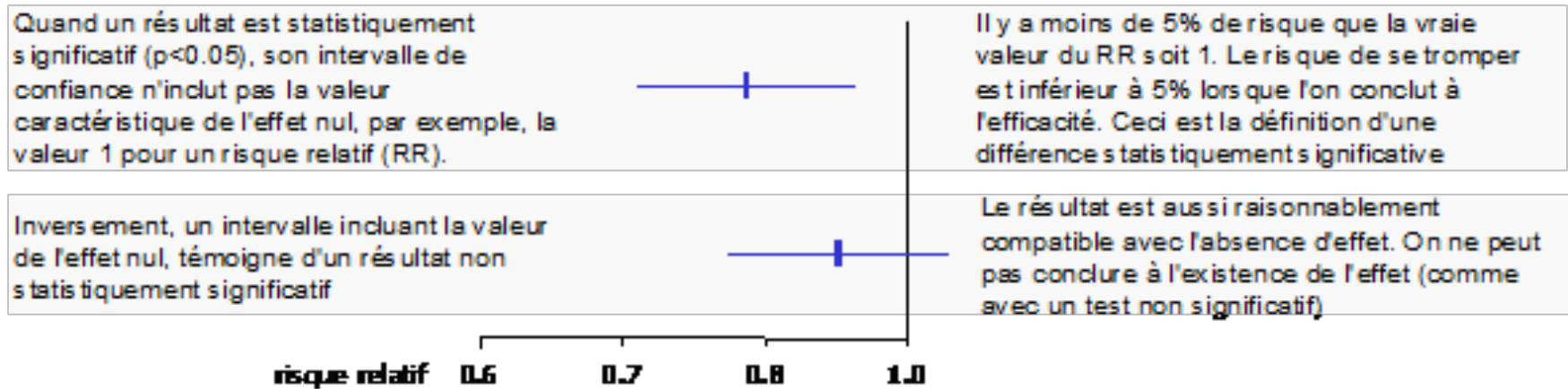
RRA = $14,4 - 4,3 = 10,1 \%$

IC = $[(10,1 - (1,96 * 0,99)) ; (10,1 + 1,96 * 0,99)]$

IC = $10,1 \%$ (8,2 % ; 12 %)



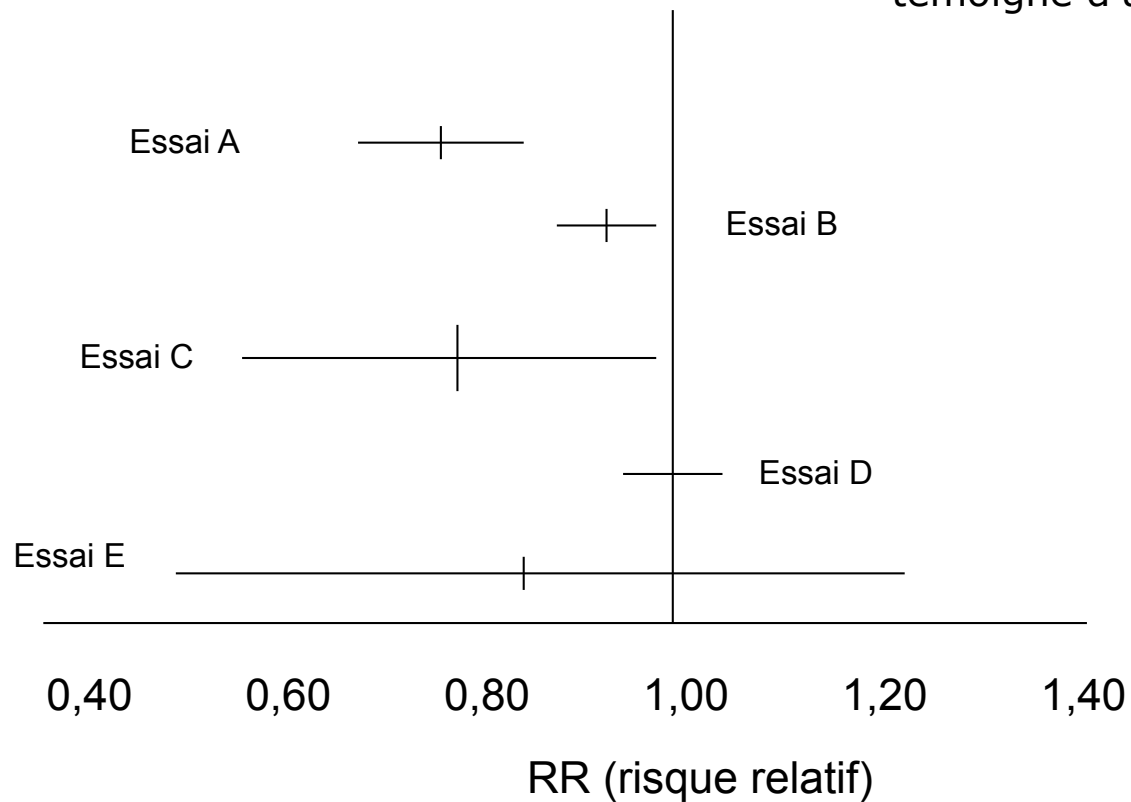
RELATION ENTRE INTERVALLE DE CONFIANCE ET TEST STATISTIQUE



INTERPRETATION DES INTERVALLES DE CONFIANCE

Essai	RRR	IC 95%	p
A	-23%	[-30%;-16%]	0,000
B	-6%	[-10% ; -1%]	0,024
C	-23%	[-41% ; -1%]	0,043
D	0%	[-4% ; 4%]	1,000
E	-19%	[-48% ; 27%]	0,362

RRR : « réduction » relative de risque. Par convention dans cet exemple, une RRR négative signe une réduction de risque.
A l'opposé, une valeur positive témoigne d'une augmentation.



INTERPRETATION DES INTERVALLES DE CONFIANCE (Suite)

Essai A : Il existe un effet statistiquement significatif, de taille importante et connu avec précision. Dans le pire des cas, cet effet est encore de - 16 %, ce qui correspond à une réduction relative du risque satisfaisante.

Essai B : L'interprétation de ce résultat est qu'il existe un effet statistiquement significatif, que l'effet du traitement est connu avec précision (l'intervalle est étroit) mais qu'il n'est pas formellement prouvé que le traitement soit intéressant en pratique. Dans la meilleure des situations (- 10%) la taille de l'effet reste faible et peu intéressant en pratique.

Essai C : L'interprétation est difficile. L'effet est statistiquement significatif, mais la taille de l'effet n'est pas connue avec précision (intervalle large) Mais la borne inférieure (- 1%) rend possible que l'effet réel soit quasiment nul. En pratique il est difficile de recommander l'utilisation de ce traitement.

Essai D : L'effet n'est pas significatif. Au mieux la réduction est très faible (- 4 %). L'intervalle de confiance est précis . Ce traitement n'est d'aucune utilité en pratique. Cet exemple montre la supériorité de l'IC sur le test statistique.

Essai E : La réduction relative n'est pas significative. Ce résultat n'autorise pas à conclure à l'absence d'effet. L'intervalle est large mais en grande partie du côté favorable, ce qui renforce la possibilité de l'existence de l'effet. D'où pour cet essai de réaliser avec une plus grande puissance.

ANALYSE EN SOUS-GROUPES (APRES RECUEIL DES DONNEES)

Des analyses en sous-groupes sont réalisées en complément du résultat principal d'un essai thérapeutique. Ces analyses présentent un intérêt dans la recherche de facteurs modifiant l'effet du traitement, mais elles ne permettent pas de conclure.

ESSAI NON CONCLUANT :

Dans un essai qui n'a pas montré de différence significative, le but des analyses en sous-groupes est de rechercher le ou les sous-groupes dans lesquels existerait un effet du traitement statistiquement significatif.

	Effet du traitement	p
Essai en entier	0,92	NS
Hommes	0,92	NS
Femmes	0,99	NS
Diabétiques	0,78	p<0,01

L'analyse en sous-groupe peut permettre de trouver des patients chez lequel le traitement marche. Mais ce type d'analyse se heurte à des difficultés méthodologiques.

Les résultats sont de nature exploratoire et ces nouvelles hypothèses devront être confirmées par de nouveaux essais.

ESSAI CONCLUANT :

Dans un essai concluant qui a montré une différence statistiquement significative, le but des analyses en sous-groupes serait de rechercher le ou les sous-groupes dans lesquels le traitement serait le plus efficace et surtout ceux dans lesquels il serait inefficace.

	Effet du traitement	p
Essai en entier	0,78	p<0,05
Age < 75	0,65	p< 0,01
Age > 75	0,90	NS
Antécédents d'infarctus	0,97	NS

L'objectif est de mieux définir le population cible.

Mais l'absence de différence significative ne signifie pas qu'il y a absence d'effet car la puissance de comparaison au niveau du sous-groupe n'est pas assurée.

DIFFERENCES ENTRE STRATES ET SOUS GROUPES

ANALYSE EN SOUS-GROUPES

Les analyses en sous-groupes sont décidées après le recueil des données. Il s'agit d'analyse « post-hoc » dont la valeur est purement exploratoire. Aucune hypothèse préalable n'a été formulée. Aucun calcul de sujet n'a été réalisé pour garantir une puissance suffisante.

ESSAIS STRATIFIES

Les réserves méthodologiques des analyses en sous-groupes sont levées grâce aux mesures suivantes :

- formulation explicite des hypothèses concernant les différentes strates,
- calcul du nombre de sujets nécessaires pour chacune des strates,
- randomisation indépendante par strate,
- prise en compte de la multiplicité des tests statistiques.