

Essais thérapeutiques bases méthodologiques

Prof. Emmanuel OGER

Centre Régional de Pharmacovigilance,
de Pharmaco-épidémiologie
et d'Information sur le médicament

Présentation (1)

- **Évaluation thérapeutique**
- **Jugement de causalité : comparabilité**
 - **Randomisation : biais de sélection**
 - **Placebo, insu : biais de suivi, d'évaluation**
 - **Analyse en intention de traiter ; biais d'attrition**
- **Critère de jugement**
 - **Qualités**
 - **Critère de substitution, critère composite**

Présentation (2)

- **Nombre de sujets nécessaire**
- **Hypothèse statistique en non infériorité**
- **Jugement de signification**
 - **Quantification de l'effet**
 - **Précision de l'estimation**
- **Analyse statistique**
 - **Tests multiples, analyse de sous-groupe**
- **Extrapolation des résultats**
 - **Population cible, source, échantillon**

Évaluation thérapeutique (1)

- **Développement d'un médicament**
 - **Période d'étude chez l'animal**
 - **Période d'étude chez l'homme**
 - ✓ **Phase I : déterminer les conditions de tolérance**
 - ✓ **Phase II : déterminer les conditions d'efficacité et les modalités thérapeutiques**
 - **Étude de pharmacodynamie**
 - **Étude de pharmacocinétique**
 - ✓ **Phase III : étude de l'efficacité clinique**
 - ✓ **Phase IV : efficacité et tolérance dans les conditions usuelles de prescription**

Évaluation thérapeutique (2)

- **Essai thérapeutique de phase III**
 - **Démontrer l'efficacité d'une thérapeutique**
 - **Exemple : avec NEW on observe une fréquence d'évènements de 10%...**
 - ✓ **Est-elle réellement différente de celle observée sous traitement de référence (17%) ?**
 - Fluctuation d'échantillonnage, hasard...?
 - Test statistique
 - ✓ **La différence peut-elle être imputée au traitement évalué ?**
 - Notion de biais. Méthodologie des essais comparatifs

Évaluation thérapeutique (3)

- **Méthodologie des essais comparatifs**
 - **Minimiser l'existence de biais**
 - ✓ **Biais : erreur systématique**
 - Biais de sélection
 - Biais de confusion
 - Biais de suivi/ Biais d'exécution
 - Biais d'évaluation
 - ✓ **Notion d'erreur aléatoire**
 - **Affirmer un lien de causalité**
 - **Problèmes de l'absence de comparaison**
 - ✓ **Biais de l'histoire**
 - ✓ **Régression vers la moyenne**

Évaluation thérapeutique (4)

- **Comparaison et différence observée**
 - **Comparaison : quel comparateur ?**
 - ✓ Efficacité thérapeutique (placebo)
 - ✓ Place dans l'arsenal thérapeutique (référence)
 - **Quelle différence ?**
 - ✓ Essai de supériorité
 - ✓ Essai de non infériorité ou d'équivalence
 - **Sur quel critère ?**
 - **Dans quelle population ?**

Évaluation thérapeutique (5)

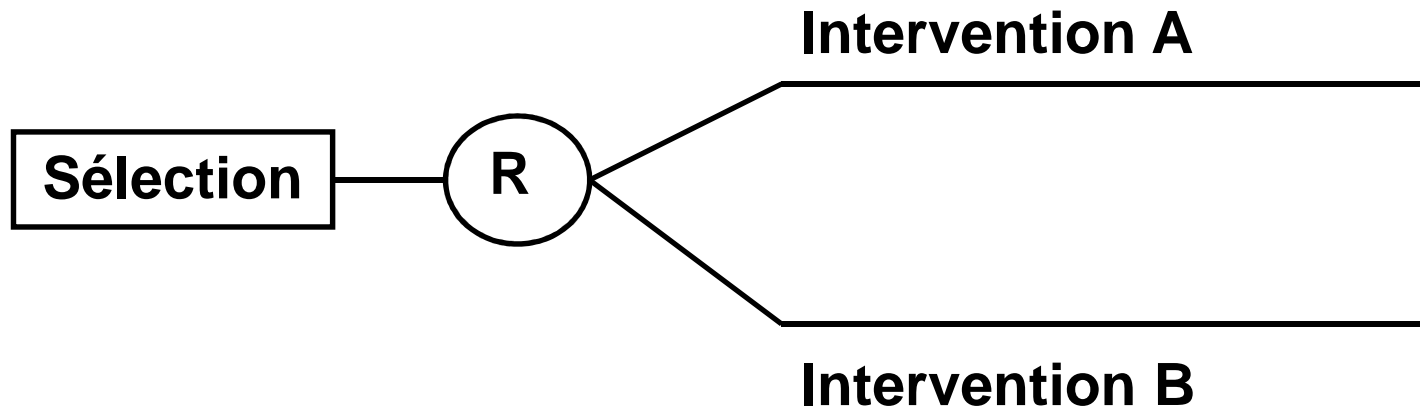
➤ Notion d'équivalence

Le progrès thérapeutique, à efficacité identique, peut correspondre, par exemple

- ✓ **A une meilleure tolérance,**
- ✓ **A un schéma d'administration et/ou une surveillance plus simple,**
- ✓ **A une durée d'hospitalisation nécessaire au traitement plus courte,**
- ✓ **A une réduction des coûts de santé...**

Jugement de causalité (1)

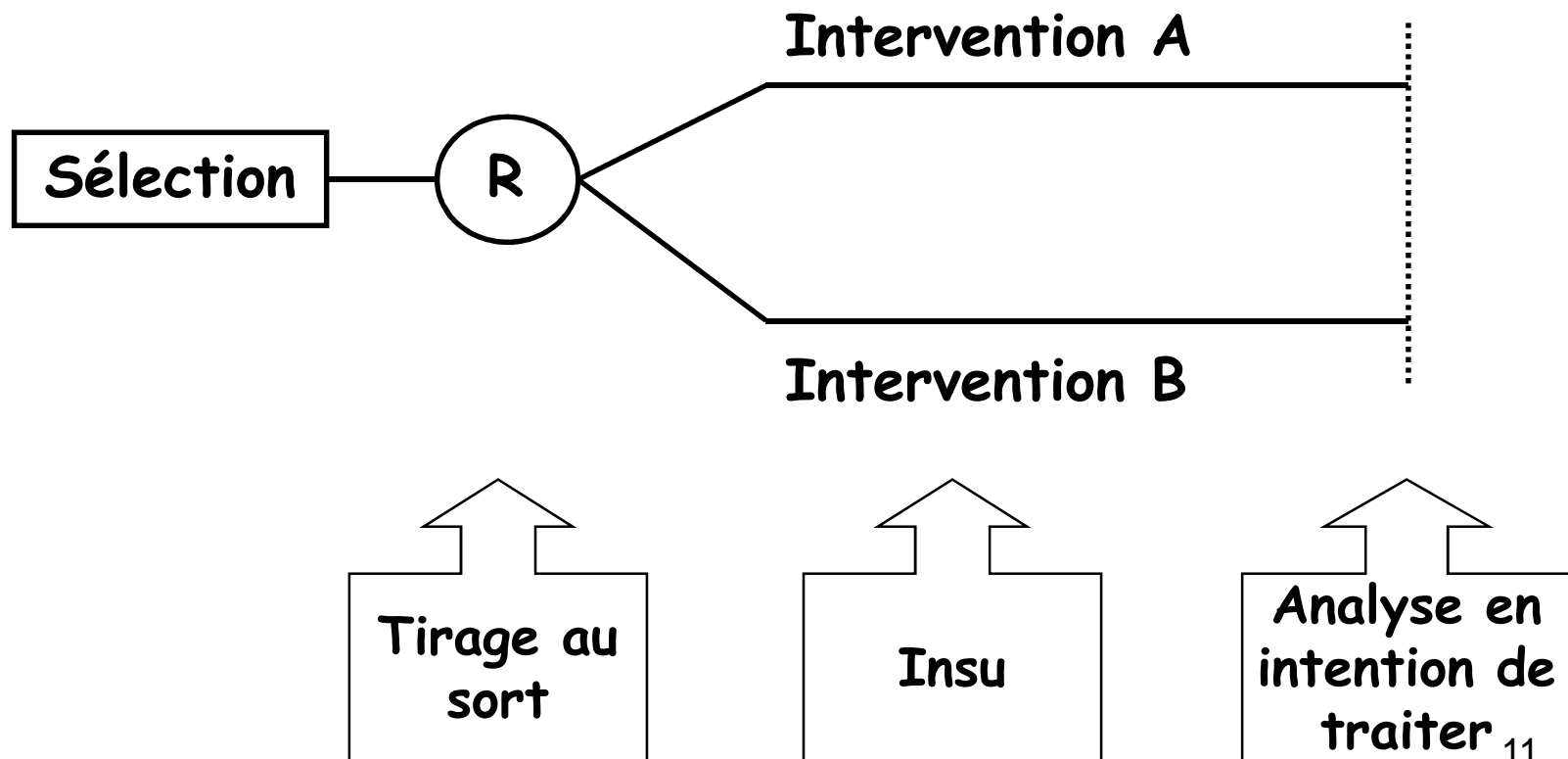
- **Peut-on imputer la différence observée à l'effet du traitement ? Lien causal ?**
 - **Les groupes sont-ils comparables ?**



Jugement de causalité (2)

- **Comparabilité initiale**
 - Randomisation
 - Biais de sélection, biais de confusion
- **Comparabilité en cours d'essai**
 - Placebo, insu
 - Biais de suivi, biais d'évaluation
- **Comparabilité finale**
 - Analyse en intention de traiter
 - Biais d'attrition

Jugement de causalité (3)



Jugement de causalité (4)

➤ Comparabilité initiale (1)

- **Randomisation : allocation aléatoire**
- **Quand randomiser ?**
 - ✓ **Après l'inclusion (« sujet bon pour l'essai »)**
 - ✓ **Le plus tard possible (avant le moment où les traitements diffèrent)**
- **Comment randomiser ?**
 - ✓ **Clause d'ignorance : imprévisibilité de l'allocation**
 - Enveloppe scellée, randomisation centralisée
 - ✓ **Tables de nombres au hasard / blocs, strates**
 - Générée par un statisticien indépendant

Jugement de causalité (5)

➤ Comparabilité initiale (2)

▪ Biais de sélection

✓ Répartition selon le choix du médecin...

✓ Allocation systématique... prévisible !

- Groupes dissemblables

- Clause d'ignorance transgressée !

- Insu impossible !

▪ Analyse des caractéristiques initiales

✓ Pas de tests statistiques

✓ Une randomisation peut produire des groupes non comparables.... par hasard!

Jugement de causalité (6)

➤ Comparabilité en cours d'essai (1)

▪ Placebo, insu

- ✓ La connaissance du traitement reçu affecte
 - Le comportement du médecin, du patient
 - La mesure du critère de jugement
- ✓ Simple insu : seul le malade ne connaît pas la nature du traitement donné
- ✓ Double insu : le malade et le médecin ne connaissent pas la nature du traitement
- ✓ Triple insu : l'évaluation (tierce personne) ne connaît pas non plus la nature du traitement

Jugement de causalité (7)

➤ Comparabilité en cours d'essai (2)

▪ Placebo

- ✓ Substance inerte dénuée de toute activité pharmacologique.
- ✓ Effet global = effet pharmacologique + effet placebo (composante non spécifique de l'effet thérapeutique, effets psychologiques de la prescription médicamenteuse).

Jugement de causalité (8)

- **Comparabilité en cours d'essai (3)**
 - **Utilisation du placebo comme comparateur**
 - ✓ Pas d'autre moyen de démontrer l'efficacité du nouveau traitement,
 - ✓ Affection bénigne,
 - ✓ Absence de traitement de référence validé.
 - **Essai contre la thérapeutique de référence**
 - ✓ Essai de supériorité,
 - ✓ Essai de non-infériorité,
 - ✓ Double placebo!

Jugement de causalité (9)

- **Comparabilité en cours d'essai (4)**
 - **Justifier l'absence de placebo : essai en ouvert**
 - ✓ Chirurgie, intervention mécanique,...
 - ✓ Insu illusoire... effet physiologique...
 - **Différence de galénique... ?**
 - ✓ Double placebo!
 - **Évaluation en insu du critère de jugement**
 - **Critère objectif**

Jugement de causalité (10)

➤ Comparabilité en cours d'essai (5)

- **Biais de suivi, biais de confusion**

- ✓ Écarts au protocole (modification de posologie)

- ✓ Traitements concomitants (source de confusion)

- ✓ Prise en charge hétérogène

- **Grâce à l'insu,**

- ✓ Il n'y a pas de modification systématique du suivi

- **Biais d'évaluation**

Jugement de causalité (11)

➤ Comparabilité finale (1)

▪ Analyse en intention de traiter (1)

- ✓ Tous les patients sont analysés dans le groupe de traitement auquel ils ont été affectés par le tirage au sort,
 - Même s'ils n'ont pas reçu le traitement (non compliance),
 - Même s'ils ont changé de traitement en cours d'essai,
 - Même s'il y a eu un écart au protocole
- ✓ Assure la comparabilité finale des groupes.

Jugement de causalité (12)

➤ Comparabilité finale (2)

▪ Analyse en intention de traiter (2)

- ✓ Permet de répondre à la question sans biais.
- ✓ Ce type d'analyse est conservatrice car elle ne favorise pas la mise en évidence d'une supériorité d'un traitement par rapport à l'autre.
- ✓ Vérifier le graphique de flux !

Jugement de causalité (13)

➤ Comparabilité finale (3)

▪ Biais d'attrition

- ✓ Les patients perdus de vue (non évalués pour le critère de jugement principal) sont exclus de l'analyse,
- ✓ « tolérable » si $< 10\%$
- ✓ Plus de 10%, risque de biais important dont l'influence sur la conclusion de l'essai peut être évaluée par l'hypothèse du pire !

Jugement de causalité (14)

➤ Comparabilité finale (4)

▪ Analyse « per protocole »

- ✓ Seuls les patients ayant suivi le protocole sont inclus dans cette analyse,
- ✓ Fourni des résultats complémentaires,
- ✓ Elle est prépondérante dans l'essai de non infériorité.

Jugement de causalité (15)

➤ Comparabilité finale (5)

▪ Analyse « per protocole » ou en ITT

- ✓ L'analyse en ITT favorise la non-infériorité.
- ✓ Dans un essai de non infériorité, l'analyse en ITT reste primordiale mais l'analyse *per* protocole, plus conservatrice sera également fondamentale.
- ✓ Pour conclure à la non infériorité, il est souhaitable que les analyses *per* protocole et en ITT donnent les mêmes résultats.

Objectif n°6 : discuter la comparabilité des groupes...

- **Comparabilité initiale**
 - Randomisation bien faite...
 - Description des caractéristiques
- **Comparabilité en cours d'essai**
 - Insu (placebo)
 - Traitement concomitants, déviation
 - Évaluation en insu du critère de jugement
- **Comparabilité finale**
 - Analyse en intention de traiter
 - Ajustement (facteurs de confusion)
 - Biais d'attrition

Objectif n°14 : relever les biais qui ont été discutés, rechercher d'autres biais...

- **Biais de sélection**
 - **Pas de randomisation ou mal faite...**
- **Biais de confusion**
 - **Tiers facteur... ajustement**
- **Biais de suivi ou d'exécution (insu)**
- **Biais d'évaluation (insu)**
- **Biais d'attrition**
 - **Sujets non évalués... non analysés**

Critère de jugement (1)

➤ **Qualités**

- **Pertinence clinique**
- **Objectivité, égalité de mesure (en insu)**
- **Exactitude et précision**
- **Reproductibilité, sensibilité**

Critère de jugement (2)

➤ Critère de substitution

▪ Pourquoi ?

- ✓ Événement clinique pertinent peu fréquent
- ✓ Critère d'efficacité plus rapidement accessible

▪ Propriétés

- ✓ Incidence > critère clinique, mesure facile
- ✓ Étape du chemin causal
- ✓ Validation par rapport au critère clinique
 - Corrélation de l'évolution (valeur prédictive)
 - Corrélation avec l'évolution sous traitement pour quantifier le bénéfice

Critère de jugement (3)

➤ Critère composite

- Les différentes composantes sont considérées comme des critères secondaires
- Valeur descriptive, concordances des estimations d'effet sur les diverses composantes
- Pas de conclusion formelle possible
- Manque de puissance (calcul global)

Objectif n°13 : discuter la nature et la précision des critères de jugement

➤ Un seul critère de jugement

- **Objectif principal, unique! Défini a priori!**
- **Calcul du nombre de sujets**
- **Pertinence clinique, précis, objectif**
- **Évalué en aveugle**
- **Critère de substitution (justification)**
- **Critère composite**

➤ Critères secondaires

- **Définis a priori, visée exploratoire!**

Nombre de sujet nécessaire (1)

- **Les éléments nécessaires au calcul**
 - **L'amplitude de l'effet que l'on souhaite être capable de détecter**
 - **Les risques consentis**
 - ✓ de ne pas détecter un effet qui existe pourtant
 - ✓ de conclure à tort à un effet
 - **Le risque de base (groupe contrôle)**
 - **La variance du critère choisi**

Nombre de sujet nécessaire (2)

➤ Les risques consentis

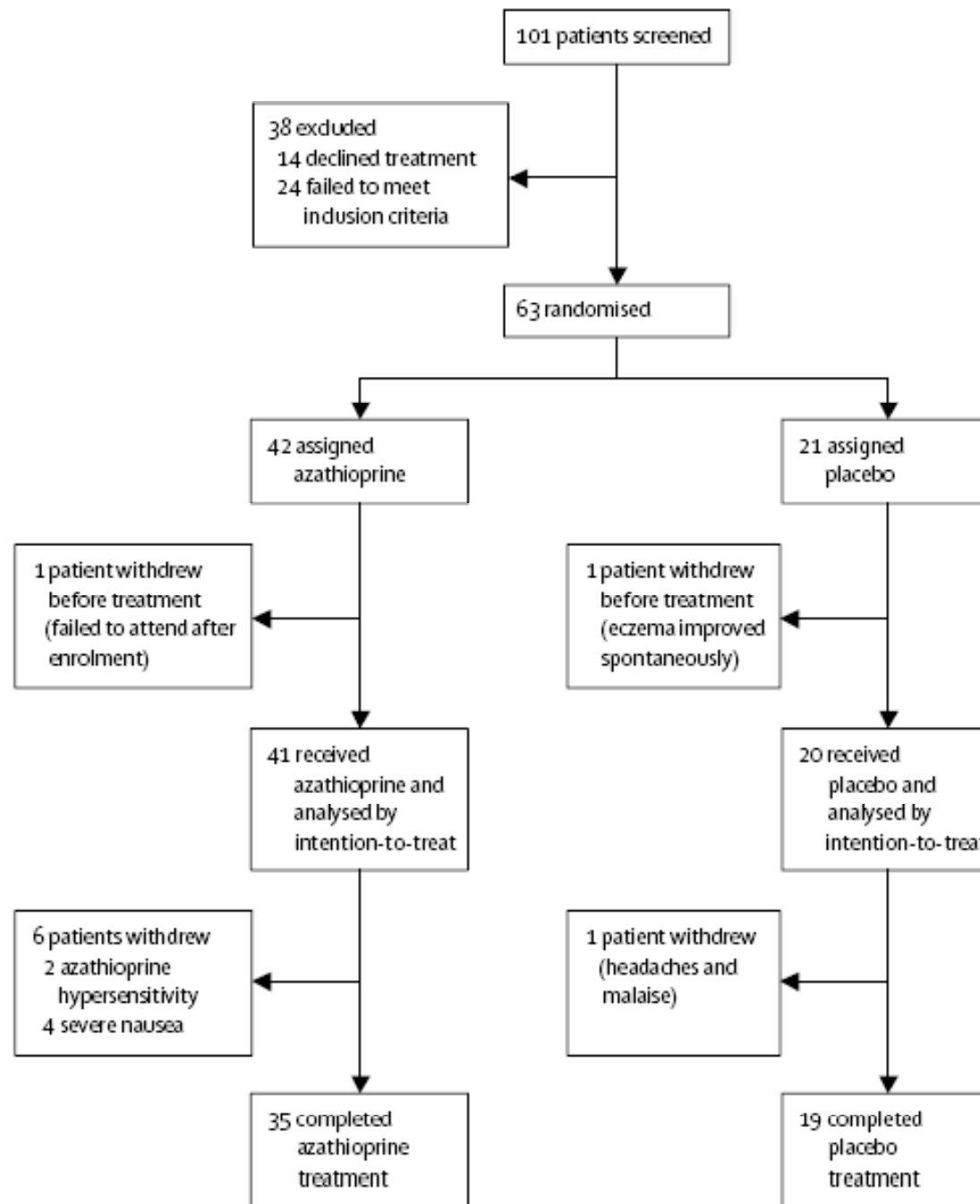
Réalité \ Conclusion	Conclusion		
	$B < A$	$B = A$	$B > A$
$B < A$		β	γ
$B = A$	α		α
$B > A$	γ	β	

Nombre de sujet nécessaire (3)

- **Conclure à une différence qui n'existe pas est grave**
risque α petit $\approx 5\%$
- **Minimiser le risque de ne pas déceler une différence réelle**
risque β petit $\approx 5\%$ à 20%
- **Le risque γ devient négligeable**

Objectif n°7 : discuter le choix des effectifs et leur cohérence...

- **Taille de l'échantillon**
 - **Calcul a priori sur une hypothèse...**
 - **Le nombre d'inclus correspond à ce calcul**
 - **Groupes équilibrés!**
- **Graphique de flux**
 - **Perdus de vue, sorties d'études**
 - **Patients évalués sur le critère principal**
- **Analyse en intention de traiter**
 - **Nombre de patients analysés / randomisés**
 - **Biais d'attrition**



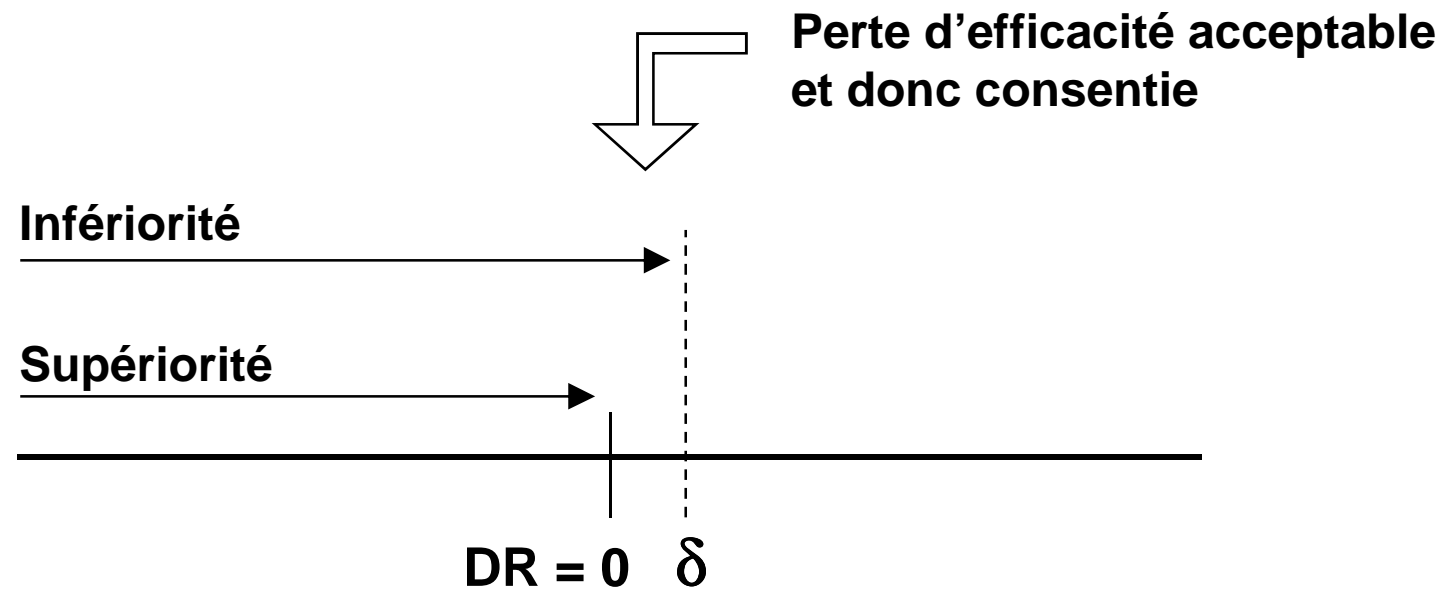
Hypothèse statistique (1)

➤ **Essai de non infériorité (1)**

- **L'efficacité de NEW est identique à celle de REF à δ près.**
- **Il s'agit en réalité d'une équivalence relative à plus ou moins δ .**
- **L'existence de cette marge est nécessaire car il est impossible de montrer la stricte équivalence.**
- **Comment définir δ ?**

Hypothèse statistique (2)

➤ Essai de non infériorité (2)



$$H_1: R_{\text{NEW}} \leq R_{\text{REF}} + \delta$$
$$DR \leq \delta$$

$$H_0: R_{\text{NEW}} > R_{\text{REF}} + \delta$$
$$DR > \delta$$

Hypothèse statistique (3)

➤ Essai de non infériorité (3)

▪ Détermination de la borne δ

- ✓ Principal problème !
- ✓ Combien d'échecs êtes-vous prêts à accepter en contrepartie de tel « progrès » ?
- ✓ Aucune règle générale... tout dépend
 - des « progrès » attendus...
 - de la pathologie... de la nature des « échecs » (décès,...)
- ✓ Effet de REF par rapport au placebo,
- ✓ Être sûr que NEW reste plus efficace que le placebo.

Hypothèse statistique (4)

➤ Essai de non infériorité (4)

▪ Détermination de la borne δ

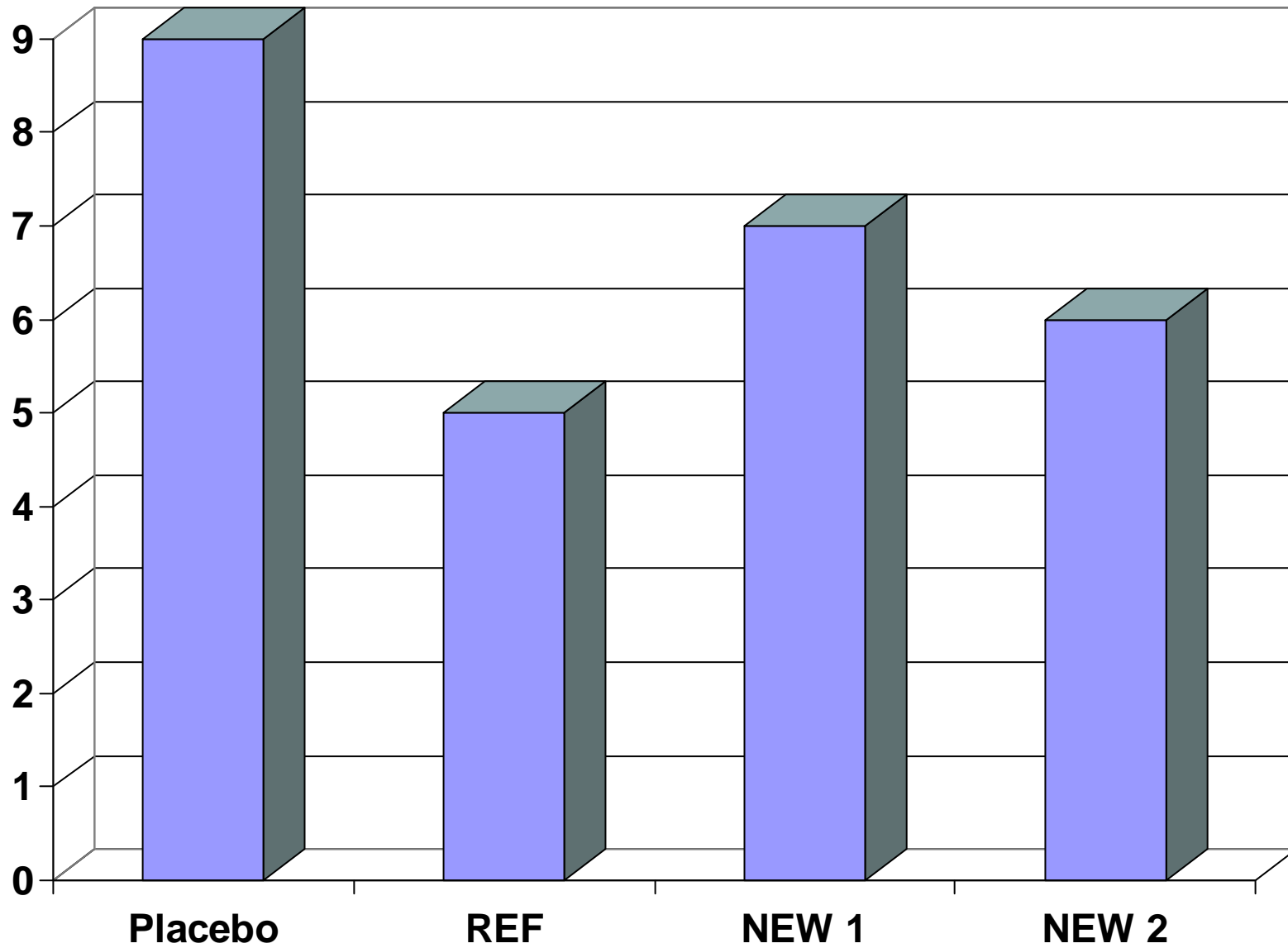
- ✓ Il faut que δ permette de ne pas perdre plus de 25 à 50 % de l'effet de REF par rapport au placebo.
- ✓ Si le taux d'échec est de 5% sous REF et de 9% sous placebo : effet de REF = 5% – 9% = – 4%
 - Pour ne pas perdre plus de 50% de cette différence, alors $\delta = 2\%$,
 - Pour ne pas perdre plus de 25% de cette différence, alors $\delta = 1\%$.

Hypothèse statistique (5)

➤ Essai de non infériorité (5)

▪ Détermination de la borne δ

- ✓ Il est préférable d'exprimer δ non pas en différence absolue mais en risque relatif
- ✓ Ainsi, les conclusions tiennent compte des incidences réellement observées :
 - hypothèse de 5 % d'échecs, borne δ identique mais exprimée différemment en :
 - différence absolue : $\delta = + 1\%$
 - en risque relatif $\delta = 6\%/5\% = 20\%$
d'augmentation du risque, soit $RR = 1,2$.



Jugement de signification (1)

➤ Quantification de l'effet (1)

▪ Indices d'efficacité pour critère binaire

✓ Fréquences d'évènements observées

✓ Quatre indices

- Risque relatif RR, réduction relative du risque = $1 - RR$

- Odds ratio OR (rapport des cotes)

- Différence des risques DR

- Nombre de sujet à traiter pour éviter un évènement NNT

▪ Indices d'efficacité pour un critère continu

✓ Moyennes observées

✓ Différences absolues ou relatives

Jugement de signification (2)

➤ Quantification de l'effet (2)

Groupe	Effectif	Évènement	Risque
NEW	400	42	0,105
REF	398	65	0,163

- $RR = 0,105/0,163 = 0,64$
- $OR = [0,10/(1-0,10)]/[0,16/(1-0,16)] = 0,60$
- $DR = 0,10 - 0,16 = -0,06$
- $NNT = 1 / DR = -17$

Jugement de signification (3)

- **Quantification de l'effet (3) : risque relatif**
 - **Mesure indépendante du risque de base**
 - **Valeur 1 importante pour l'interprétation**
 - ✓ **RR < 1 = effet bénéfique : le traitement réduit la fréquence de l'événement,**
 - ✓ **RR = 1 traduit un traitement sans efficacité,**
 - ✓ **RR > 1 témoigne d'un traitement délétère qui augmente la fréquence de l'événement.**

Jugement de signification (4)

➤ Quantification de l'effet (4)

▪ Différence des risques DR

- ✓ Taille d'effet non ajustée sur le risque de base
- ✓ En l'absence d'effet, la différence est nulle. Un effet bénéfique se traduit par une différence des risques négative et un effet délétère par une valeur positive.
- ✓ La fréquence des événements est dépendante de la durée de suivi.

Jugement de signification (5)

➤ Quantification de l'effet (5)

▪ NNT

- ✓ Nombre moyen de sujets qu'il est nécessaire de traiter pour éviter un événement.
- ✓ Le signe « moins » témoigne d'un effet bénéfique et l'interprétation en fonction du signe est identique à celle de la différence des risques.
- ✓ Le NNT, comme la différence des risques, dépend de la durée de suivi.

Jugement de signification (6)

➤ Quantification de l'effet (6)

▪ Bénéfice relatif

- ✓ Information explicative,
- ✓ Reflet direct de l'efficacité du traitement,
- ✓ Constant d'une population à l'autre, il permet de comparer différents traitements de la même maladie.

▪ Bénéfice absolu

- ✓ Reflète les conséquences au niveau d'une population,
- ✓ Spécifique d'une situation particulière.

Jugement de signification (7)

➤ Précision de l'estimation (1)

- La valeur d'un paramètre varie en fonction de l'échantillon,
 - ✓ Fluctuations aléatoires d'échantillonnage
 - ✓ Pour prendre en compte l'imprécision dans l'estimation, on fournit un intervalle dit de confiance qui permet de se faire une idée intégrant l'incertitude statistique liée au hasard.

Jugement de signification (8)

➤ Précision de l'estimation (2)

▪ Intervalle de confiance (IC) à 95 %.

- ✓ Signifie qu'il y a 95% de chance que l'IC comprenne la vraie valeur du paramètre (raccourci un peu rapide...)
- ✓ On accepte de se tromper puisqu'il y a 5% de chances que l'IC ne contienne pas la vraie valeur.

Jugement de signification (9)

➤ Précision de l'estimation (3)

- **3 facteurs font varier l'amplitude de l'IC**
 - ✓ **Le risque d'erreur accepté : plus il diminue, plus l'IC sera large.**
 - ✓ **La variabilité des mesures : si les valeurs sont très variables, l'IC sera plus large.**
 - ✓ **La taille de l'échantillon : plus il est grand, plus l'IC à 95 % est étroit, donc plus précis.**

Jugement de signification (10)

➤ Précision de l'estimation (4)

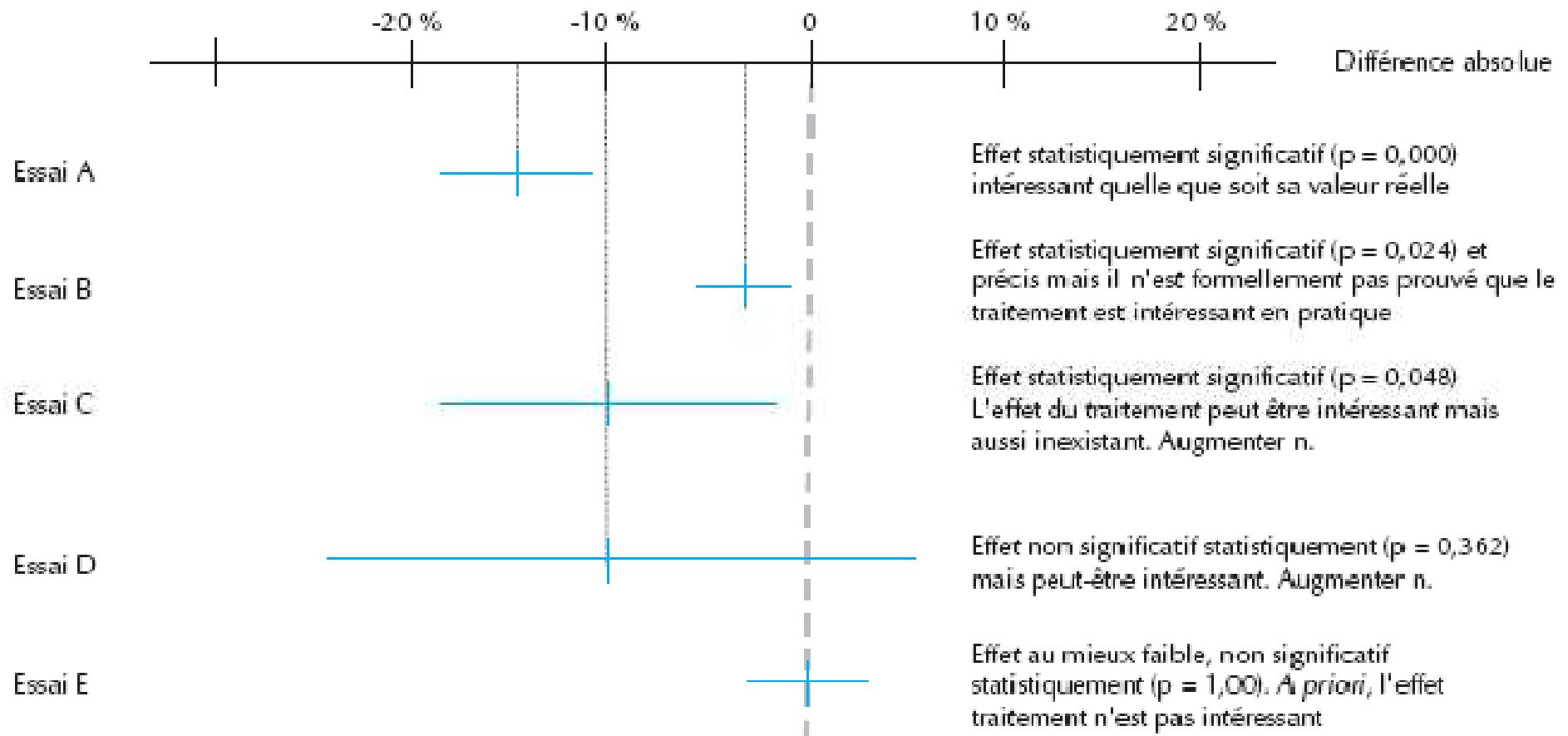
▪ IC d'une différence des risques

- ✓ Si l'IC à 95% ne contient pas 0, il est possible de conclure que les risques sont différents, que la différence observée est significativement différente de 0.
- ✓ Dans ce cas, la probabilité de conclure à tort, p , est inférieure à 0,05.

Jugement de signification (11)

- **Précision de l'estimation (5)**
 - **IC, pertinence clinique**
 - **Valeur de p , interprétation statistique**

Jugement de signification (12)



Objectif n°17 : discuter la signification clinique des résultats

Un résultat statistiquement significatif n'est pas toujours cliniquement pertinent!

La valeur de p ne reflète pas la taille de l'effet!

- **Quantification de la taille de l'effet**
- **Précision de l'effet (IC95%)**

Objectif n°16 : discuter la signification statistique des résultats

- **Intervalle de confiance à 95%**
- **Degré de signification (valeur de p)**
- **Résultat non significatif**
 - **Différence réelle**
 - ✓ **Manque de puissance**
 - ✓ **Biais**
 - **Pas de différence en réalité**
- **Tests multiples**

Objectif n°9 : vérifier que les analyses statistiques sont cohérentes

Critère	Binaire	Continu	Délai
Mesure d'effet	DR, RR	Δ moyennes	RR
Test	Chi 2 Fisher	Student	Log-rank
Modélisation	Logistique	Linéaire	Cox

Analyse statistique (1)

➤ Tests multiples

- **Multiplier les tests, c'est majorer la chance de trouver une différence significative même si le traitement est sans aucun effet.**
- **Sur 100 tests, il y aura en moyenne 5 tests significatifs (au seuil de 5%).**
- **Pour contrôler le risque alpha, il ne faut faire qu'un seul test ou utiliser une méthode d'ajustement du seuil de signification statistique.**

Analyse statistique (2)

➤ Tests multiples

- **Plusieurs situations de multiplicité des tests**
 - ✓ **Nombreux critères de jugement,**
 - Critères secondaires,
 - Composantes d'un critère principal composite,
 - ✓ **Plusieurs temps du suivi (analyse intermédiaire),**
 - ✓ **Plusieurs traitements comparés,**
 - ✓ **Différents sous-groupes...**

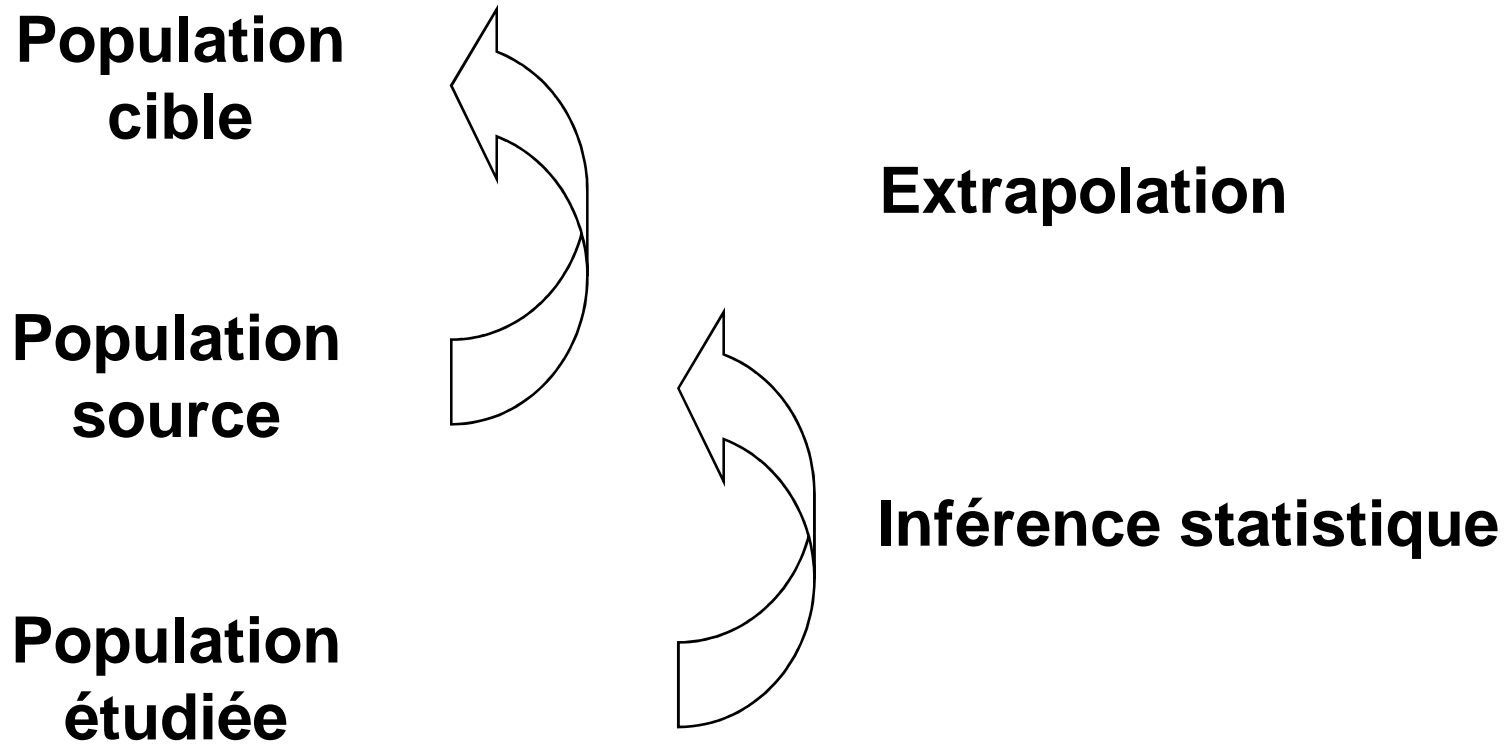
Analyse statistique (3)

➤ Analyse de sous-groupes

- En complément de l'analyse principale pour rechercher des facteurs modifiant l'effet du traitement,
- Elles ne permettent pas de conclure : analyse à titre exploratoire.
- Deux situations
 - ✓ Effet significatif sur toute de la population
 - Recherche d'interaction (hétérogénéité de l'effet)
 - ✓ Essai globalement non concluant...

Extrapolation des résultats (1)

➤ Population cible, source, échantillon



Extrapolation des résultats (2)

- **Population cible, source, échantillon**
 - **Population étudiée proche des patients vus en pratique « représentativité »**
 - **Critères d'inclusion et d'exclusion, les caractéristiques des patients inclus.**
 - **Manque de représentativité gênant si l'effet du traitement varie substantiellement en fonction des caractéristiques des patients.**

Informations complémentaires

- **Inter-mémo VG**
- **Revue Médecine & Thérapeutique**
 - **2007, vol. 13, n°4**
- **<http://www.spc.univ-lyon1.fr/lecture-critique>**