

BASE DE STATISTIQUES

Aide à la réalisation du travail de fin
d'étude

Dr E. Engelman

01.04.2023



POURQUOI DES ÉTUDES CONTRÔLÉES ?

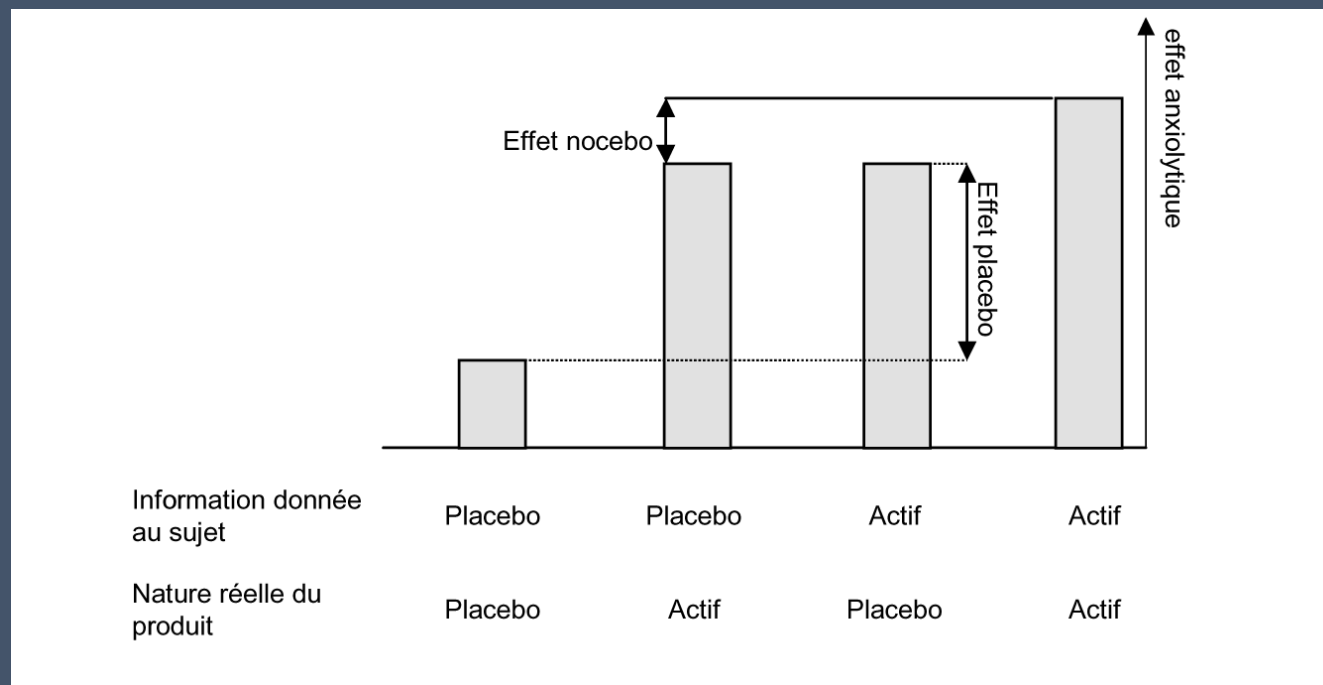
- Pourquoi faire une **étude contrôlée** plutôt que de se baser sur des résultats ponctuels anecdotiques (au mieux case report) ?
- Facteurs de **confusions** (« *confounding factors* ») = biais de confusion
 - l'évolution naturelle de la maladie
 - l'effet placebo
 - la régression à la moyenne
 - l'effet des traitements concomitants
 - plus bien d'autres

Effet placebo

Volontaires soumis à une épreuve stressante

Avant cette épreuve, les sujets prenaient des gélules qui contenaient soit un anxiolytique soit un placebo. De plus ils recevaient une information sur la nature du produit qu'ils prenaient, mais cette information ne correspondait pas forcément à la réalité. Ce plan expérimental crée quatre groupes de sujets :

- ceux qui reçoivent le placebo et à qui l'on dit qu'ils prennent le placebo,
- ceux qui reçoivent l'anxiolytique mais à qui l'on dit qu'ils prennent le placebo,
- ceux qui reçoivent le placebo mais à qui l'on dit qu'ils prennent l'anxiolytique,
- ceux qui reçoivent l'anxiolytique et à qui l'on dit qu'ils prennent l'anxiolytique.



BIAIS: DÉFINITION STATISTIQUE

- C'est une **erreur *systematique*** entre une estimation et la véritable valeur du paramètre estimé
- Elle se reproduit à l'identique, et contrairement aux erreurs aléatoires, ***ne se compense pas en moyenne***
- Dans le langage courant on appelle « biais » la cause des biais statistiques

PRISE EN COMPTE DES FACTEURS DE CONFUSION

- L'effet de ces facteurs est pris en compte grâce à l'utilisation d'un groupe de référence
- N'est pas suffisant pour tous les biais cf. le biais de sélection, biais de traitement, de suivi, de mesure ...
- On a décrit environ 120 types de biais différents

MAUVAIS GROUPES CONTRÔLES

- **Contrôles historique**
 - Rien ne garanti que les anciens patients soient comparables au nouveaux
 - Les autres traitement concomitants ont aussi évolués. Un meilleur résultat avec les nouveaux patients signifie simplement que la prise en charge globale est meilleur, pas que le nouveau traitement étudié a un effet
- **Contrôle géographique**
 - Patients traité dans le service A vs. le service B
 - Pas de garantie de la similitude des patients ni des traitements concomitants de base

COMMENT CONTRER LES BIAIS ?

- La randomisation
 - Seule l'allocation **aléatoires** des traitements permet de constituer la comparabilité des groupes
- Le double insu («*double blind*»)
 - Evite le biais de suivi et de mesure

EN PRATIQUE



STATISTICIAN

**I'M NOT ARGUING
I'M JUST EXPLAINING
WHY I'M RIGHT**

RANDOMISATION EN PRATIQUE (1)

- Le moyen le plus simple: la constitution d'une liste à l'aide d'un programme informatique ou d'une liste de nombres au hasard
 - Dans un essai multicentrique: liste disponible dans chaque centre
 - Aussi effectuée à l'aide d'enveloppe scellée qui renferme le numéro (pour étude en aveugle) ou en claire la nature du traitement (étude ouverte)
 - A noter que dans un contexte anesthésico-chirurgicale il est souvent impossible d'effectuer une étude en double aveugle en cas de réalisation d'un acte technique (essayer de séparer celui qui fait de celui qui mesure l'effet !)
- **Inconvénient:**
 - Permet de prendre connaissance anticipée de la nature du traitement du prochain patient
 - Permet d'arranger le moment où un patient sera inclus afin de recevoir le traitement souhaité
 - Autrement dit: il est matériellement possible de « tricher » ou de simplement donner le traitement qu'on pense être le meilleur à quelqu'un qui en bénéficiera le plus et d'avoir un biais de sélection

RANDOMISATION EN PRATIQUE (2)

- Seule une procédure centralisée et extérieure de randomisation garantit suffisamment son imprévisibilité
 - Donc: le patient est enregistré dans l'étude
 - La randomisation s'effectue le plus tard possible, et le plus près possible du moment où le traitement expérimental vs. standard est administré
 - Contact est pris avec le centre de randomisation (téléphone, Internet) et le numéro ou nature du traitement est obtenu
 - Du fait de la centralisation, tous les patients sont connus, et empêche de faire disparaître un patient de l'étude « ni vu, ni connu », et permet de détecter un biais de suivi (mais ne l'empêche pas)

SIGNIFICATION STATISTIQUE

- Lorsque $p < 0.05$ ($\alpha = 5\%$) la différence est dite statistiquement significative

le p n'est pas

p n'est pas la probabilité de l'hypothèse nulle

p n'est pas la probabilité d'absence de différence

p n'est pas la probabilité que le traitement n'ait pas d'effet

$p < 0.05\%$ ne signifie pas qu'il y a moins de 5% de chance que le traitement soit sans effet

p n'est pas $\Pr(H_0)$ ou $1 - \Pr(H_1)$
 p n'est pas la probabilité de l'hypothèse nulle

le p est

p est la probabilité d'obtenir le résultat observé si l'hypothèse nulle est vraie

p est la probabilité d'observer une différence au moins aussi importante si en réalité il n'y a pas de différence

p est la probabilité d'obtenir le résultat qui a été observé si le traitement est en réalité inefficace

il y a moins de 5% d'observer le résultat obtenu si le traitement est sans effet

$p = \Pr(\text{résultat}/H_0)$
 p est la probabilité conditionnelle du résultat sous l'hypothèse nulle

ALPHA = 5%: EST-CE RAISONNABLE ?

- D'autres valeurs sont proposées comme 1%
- Si 400 spécialités pharmaceutiques sont testées par un seul essai:
 - Avec un $\alpha=5\%$
 - il existe un risque que 20 de ces produits soient en réalité inefficace
 - Avec un $\alpha=1\%$
 - il existe un risque que 4 de ces produits soient en réalité inefficace
- Donc, si large diffusion du traitement (vaccin par ex.) prendre un risque de 5% de conclure à tort à son efficacité est trop important. 1% serait ok
- Par contre pour une maladie très rare, sans traitement efficace consentir un risque $\alpha=10\%$ peut s'envisager

CHOIX DE L'ENDPOINT PRIMAIRE (PRIMARY ENDPOINT)*

- Est le critère qui servira à mettre en évidence l'efficacité du traitement
- Vous oblige à exprimer en langage courant ce que vous cherchez et ce que vous voulez montrer
- Une étude peut comporter plusieurs endpoints primaires (voir problème des comparaisons multiples)
- Un endpoint primaire peut-être un composite de plusieurs critères cliniques

*le critère de jugement principal, pour les français !

ENDPOINTS PRIMAIRES MULTIPLES

- Par exemple 3 EP
- La réalisation de plusieurs (3) test statistiques augmente le risque de faire à tort une conclusion d'efficacité
 - 5% sur le premier test, 5 % sur le 2^e, 5% sur le 3^e
 - Avec les 3 critères on la probabilité de trouver au moins une différence significative à tort est 14.3%
 - Ce risque augmente avec le nombre de comparaisons
- Donc, il faut corriger ce phénomène et si l'ont veut garder 5% d'erreur globalement pour l'ensemble de l'analyse il faut considérer chaque EP significatif si $p < (0.05/3)$ càd $p < 0.017$

ENDPOINTS SECONDAIRES (SECONDARY ENDPOINTS)

- Documenter les bénéfices secondaires
 - par ex. si EP est mortalité totale, le traitement peut aussi réduire la fréquence des événements non-mortels
- Critères pour documenter le mécanisme d'action
 - Par ex. causes spécifiques de mortalité
- Critères intermédiaires
 - Par ex. taux de reperfusion coronaire dans l'essai sur la fibrinolyse dans l'infarctus
- Critères correspondant aux effets délétères du traitement
- Les composants individuels d'un EP composite

PUISSANCE DES ENDPOINTS SECONDAIRES

- La puissance statistique d'un essai est déterminée pour l'endpoint primaire
- La puissance pour les endpoints secondaires n'est pas contrôlée et peut être faible
 - Quand la fréquence de base de l'endpoint secondaire est plus petite que celle de l'EP
 - Quand l'effet attendu sur l'endpoint secondaire est plus petit que pour de l'EP

MULTIPLICITÉ DES ENDPOINTS SECONDAIRES

- Correction de α
- $P < (0.05 / k+1)$
 - K = nombre d'endpoints secondaires
 - +1 pour l'endpoint primaire

PUISSANCE ET NOMBRE DE SUJETS NÉCESSAIRES (1)

- La puissance statistique d'un essai mesure son aptitude à mettre en évidence l'effet d'un traitement si celui-ci existe
- Le risque β est la probabilité d'obtenir un faux résultat négatif (alors que le traitement est efficace)
- La puissance est la probabilité (exprimée en %) d'obtenir un vrai résultat positif (mettre en évidence l'efficacité d'un traitement réellement efficace)

PUISSANCE ET NOMBRE DE SUJETS NÉCESSAIRES (2)

- La puissance d'un essai dépend de:
 - *La taille de l'effet à mettre en évidence: différence entre les incidences (critère binaire) ou différence entre moyennes (critère continu)*
 - il faut définir *a priori* la taille de l'effet recherché
 - Le nombre de sujets
 - Au plus il y a de sujets au plus, la puissance est grande
 - Pour un petit effet, il faut une grande puissance statistique donc beaucoup de sujets
 - Pour les critères binaires: le risque de base sans traitement
 - Si la fréquence de base est petite il faut plus de puissance que si fréquence élevée
 - Pour un critère continu: la variance
 - Au plus la variabilité entre sujet grande, au plus il faut de la puissance, donc de sujets
- Le risque d'erreur α consenti (la fameuse valeur de p)
 - Ne pas oublier la correction pour multiplicité si plusieurs endpoints primaires

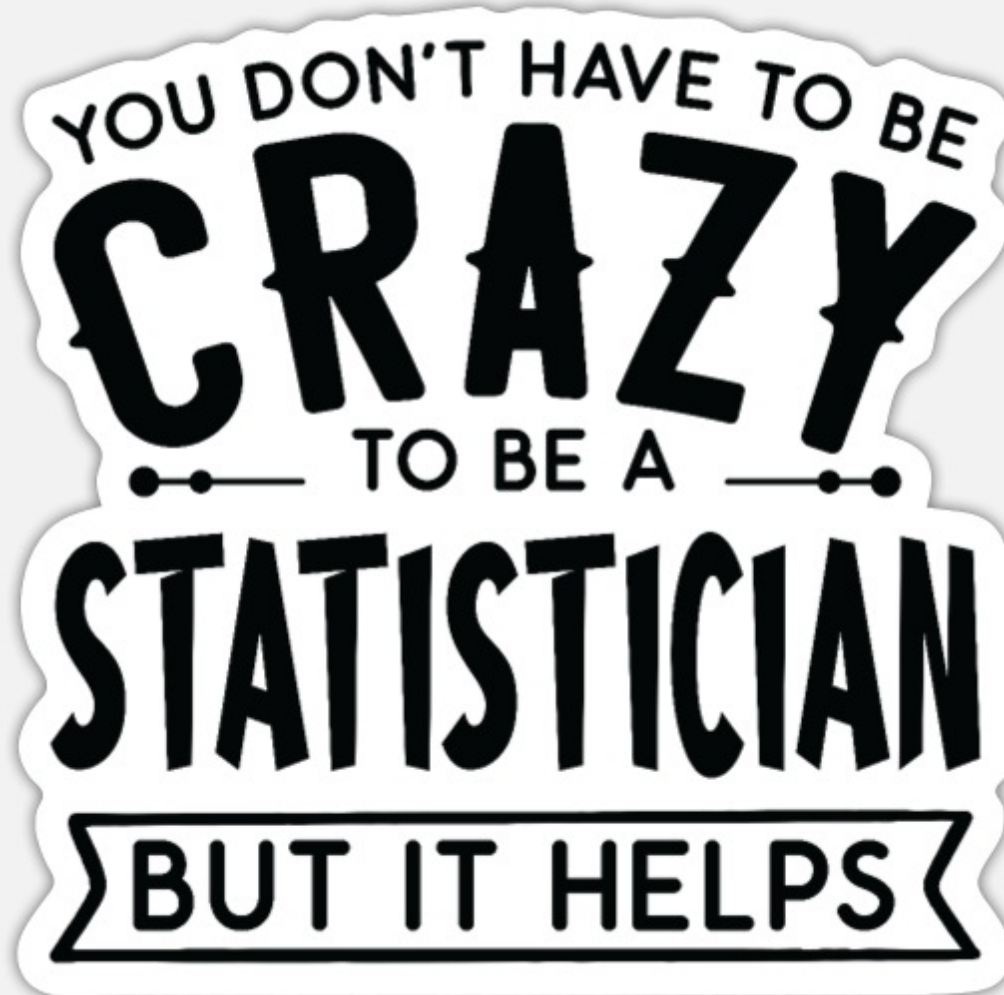
PUISSANCE ET NOMBRE DE SUJETS NÉCESSAIRES (3)

- Pour un primary endpoint continu
- Vous devez fournir:
 - La réduction de la valeur moyenne que vous désirez pour considérer le traitement comme meilleur
 - La variance des valeurs (représentant la variabilité individuelle)
 - L'erreur α que vous acceptez: c'est-à-dire quel est la valeur de p que vous visez ($p < 0.05$, 0.01 , autre)
 - Quel puissance désirez vous ? 80 % ou 90% sont les standards souvent utilisés

PUISSANCE ET NOMBRE DE SUJETS NÉCESSAIRES (3)

- Pour un primary endpoint binaire
- Vous devez fournir:
 - La réduction de l'incidence de l'évènement que vous désirez pour considérer le traitement comme meilleur
 - L'incidence de base de l'évènement sans traitement
 - L'erreur α que vous acceptez: c'est-à-dire quel est la valeur de p que vous visez ($p < 0.05$, 0.01 , autre)
 - Quel puissance désirez vous ? 80% ou 90% sont les standards souvent utilisés

EN PRATIQUE



PLAN D'EXPÉRIENCE (1)

- **Groupes parallèles: parallel arms**
 - Archétype de l'essai thérapeutique – un groupe contrôle est comparé à un groupe actif (ou plusieurs k groupes actifs)
 - Si plusieurs groupes: α/k si chaque groupe comparé au placebo, $\alpha/2k$ si en plus chaque groupe actif est comparé aux autres
- **Plan factoriel: factorial design**
 - Répond à 2 questions différentes dans le même essai
 - Permet un gain de temps et de patients
 - Il est nécessaire que les 2 traitements n'interagissent pas entre eux

PLAN D'EXPÉRIENCE (2)

PLAN FACTORIEL (2)

- On a en fait 4 « traitements »
 - traitement A
 - placebo de A
 - traitement B
 - placebo de B
- $\frac{1}{4}$ des patients recevront A et B
- $\frac{1}{4}$ des patients recevront A et placebo B
- $\frac{1}{4}$ des patients recevront placebo A et B
- $\frac{1}{4}$ des patients recevront placebo A et placebo B

PLAN D'EXPÉRIENCE (3)

PLAN FACTORIEL (3)

- Analyse finale assez complexe – pour faire simple
 - Pour un critère binaire:
 - test de Mantel Haenszel ou un test de Logrank stratifié ou des données de survie
 - Analyse multivariée de régression logistique ou un modèle de Cox incluant une interaction entre les traitements
 - Pour un critère continu
 - Analyse de variance à 2 facteurs, mais nécessite l'égalité (ou la proportionnalité) des effectifs des 4 groupes

PLAN D'EXPÉRIENCE (4)

- Essai croisé: *cross over*
- Chaque patients reçoit tous les traitements de l'essai, administrés lors de périodes successives
- Utilise le patient comme son propre contrôle
- Avantage: forte comparabilités des groupes contrôles et traités (mêmes patients)

PLAN D'EXPÉRIENCE (5)

ESSAI CROISÉ (2)

- Peut recevoir d'abord le traitement actif ou d'abord le contrôle dans un ordre aléatoire
- Chaque patient produit donc un critère de jugement pour chaque traitement
- Par exemple valeur de *peak flow* chez un asthmatique stable sous bronchodilatateur ou placebo

PLAN D'EXPÉRIENCE (6)

ESSAI CROISÉ (3)

- Nombreuses conditions d'applicabilité
 - L'endpoint doit pouvoir être mesuré à plusieurs reprises
 - L'effet du traitement ne doit pas être irréversible
 - Une période lavage (wash out) doit être prévue entre les 2 traitements
 - La maladie ne doit pas évoluer notablement entre les 2 périodes
 - Pas d'interférence entre l'ordre d'administration et l'effet
 - Le nombre de perdus de vue doit être limité en première période
 - La mesure de l'endpoint doit être dépourvue d'effet de conditionnement
- Concrètement ne peut s'utiliser que si
 - Maladie chronique stable, endpoint intermédiaire, effet du traitement disparaît rapidement à l'arrêt, apparition rapide de l'efficacité

PLAN D'EXPÉRIENCE (7)

ANALYSE DE SURVIE

- Ne concerne pas uniquement la « survie » au sens vie / mort
- Signifie en fait persistance d'une situation
- Donc, peu s'utiliser pour l'analyse de nombreux endpoints et situation
 - Par exemple l'étude des durées de périodes exemptes de nausées peut parfaitement s'analyser par une technique statistique « d'analyse de survie »

PLAN D'EXPÉRIENCE (8)

- Etude de non-infériorité
- Ne veut pas dire que les traitements sont exactement équivalents
- On doit accepter une perte d'efficacité dont la taille est à définir
 - Correspond aux marges de non-infériorité à calculer
 - Cela doit être compenser par un autre avantage potentiel



INDICES D'EFFICACITÉ CRITÈRES BINAIRES (1)

- Données nécessaires
 - Fréquences de survenue du critère de jugement dans le groupe expérimental et contrôle
 - Effectifs des groupes
- RISQUE RELATIF
 - Risque = nombre d'évènements / effectif
 - $RR = \text{risque avec traitement} / \text{risque sans traitement}$
 - Interprétation du RR
 - $RR=1$: absence d'effet
 - $RR<1$: réduction du risque avec le traitement
 - $RR>1$: augmentation du risque avec le traitement

INDICES D'EFFICACITÉ CRITÈRES BINAIRES (2)

Tableau 2 – Exemple de résultat apporté par un essai avec un critère de jugement binaire

	Effectif	Événements	Risque
traitement étudié	250	21	0.08 (8%)
traitement contrôle	246	36	0.15 (15%)

- Dans notre exemple:
 - $RR=0.08/0.15=0.53$
- Réduction du risque relatif:
 - $RRR=(1-0.53)\times 100=47\%$
- Différence de risque
 - $Diff=0.15 - 0.08=0.07$
- Number Needed to Treat NNT = nombre moyen de patients à traiter pour éviter 1 événement
 - $NNT = 1/Diff = 1/0.07=14$

INDICES D'EFFICACITÉ CRITÈRES BINAIRES (3)

- ODDS RATIO - RAPPORT DES COTES
 - Cf cote des parieurs (courses de chevaux)
 - OR =
$$\frac{0,08/(1-0,08)}{0,15/(1-0,15)}$$

INDICES D'EFFICACITÉ CRITÈRES BINAIRES (4)

Tableau 3 . Relation entre l'odds ratio (OR) et le risque de base (R0) dans une situation où le risque relatif (RR) est constant et égal à 0,8.

R0	RR	R1=R0*RR	OR
0.02	0.80	0.02	0.80
0.10	0.80	0.08	0.78
0.20	0.80	0.16	0.76
0.40	0.80	0.32	0.71

- OR est une estimation du RR lorsque la fréquence de l'évènement est faible
- Pour une fréquence élevée, OR donne une image de surestimation de l'effet thérapeutique

INDICES D'EFFICACITÉ POURQUOI LES OR ?

- OR utilisé dans les études cas témoins (étude rétrospectives)
- Mettre en évidence des facteurs qui peuvent contribuer à l'apparition d'une maladie
 - comparer des sujets malades (les cas) avec des sujets non malades (les témoins), mais qui sont similaires par ailleurs
- Dans ce type d'étude: nombre de cas est fixé par l'investigateur, sans que l'on connaisse la population dont-ils sont issus
 - On ne peut déterminer le RR car on ne connaît pas le dénominateur

INDICES D'EFFICACITÉ

RR vs OR : PRO CON (1)

- En pratique, le risque relatif est préféré à l'odds ratio car :
 - il est d'interprétation intuitive
 - c'est un indice qui vient naturellement à l'esprit pour mesurer l'effet d'un traitement
 - il donne une valeur similaire à celle de OR car dans une grande majorité des situations le risque de base est peu important
- Inconvénient du RR
 - Il ne peut pas prendre la même plage de valeur en fonction du risque de base
 - Le RR max. possible diminue avec l'augmentation du risque de base sans traitement
 - RR n'est pas symétrique pour un décrire l'opposition présence / absence de l'évènement

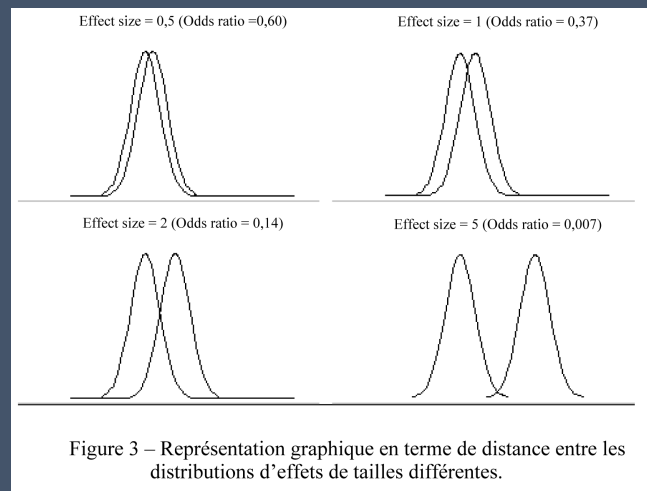
INDICES D'EFFICACITÉ

RR VS OR : PRO CON (2)

- Inconvénient du RR (suite)
- Avec exemple initial:
 - Groupe traité: l'absence de l'événement est de $1-8\%=92\%$
 - Groupe contrôle: l'absence de l'événement est de $1-15\%=85\%$
 - Nous avons vu que le traitement entraîne une réduction relative de 47% de la fréquence de l'événement
 - Par contre le traitement n'augmente que de $((1-0,92 / 0,85)*100\%)=8\%$ la fréquence relative des patients exempts de l'événement (et non 47%)
 - Le traitement pourrait paraître moins efficace avec le critère de jugement de prévention de l'événement par rapport au critère d'échec du traitement
- Conclusion:
 - le RR ne peut pas être considéré comme étant une caractéristique universelle d'un effet traitement
 - Le OR ne présente pas ces inconvénients: il peut avoir la même valeur quel que soit le risque de base, ou le sens de l'analyse

INDICES D'EFFICACITÉ CRITÈRE CONTINU (1)

- Différence des moyennes et intervalle de confiance de la différence
 - Différence absolue
 - Différence relative (en %)
- Effet standardisé
 - Différence des moyennes standardisé par l'écart type commun aux 2 groupes
 - Permet le calcul de l'effect size qui exprimera le % de valeurs non communes



PREDICTION AND ANALYSIS

- Predicting categorical variables from categorical variables
 - Log-Linear model
 - Logistic regression
- Predicting continuous variables from continuous variables
 - Simple linear relationship
 - Multiple regression
- Predicting continuous variables from categorical variables
 - One-way or multi-way ANOVA
 - Analysis of covariance
 - Multivariate analysis of variance
 - Repeated measures ANOVA
- Predicting categorical variables from continuous variables
 - Logistic regression
 - Discriminant analysis