

# **GUIDE DE LECTURE CRITIQUE D'UN ARTICLE MEDICAL ORIGINAL (LCA)**

Coordinateurs : N. MEYER

Faculté de Médecine de Strasbourg  
Janvier 2017



## TABLE DES MATIERES

L'épreuve de lecture critique d'article original	3
Contexte, ECN, modalités	3
Objectifs pédagogiques	4
L'article original	7
Type, structure	8
Résumé	10
Grille de lecture générale	12
Technique de lecture	13
Niveau de preuve	14
Les grands types de recherche	15
Grilles de lecture spécifiques	16
Etudes épidémiologiques	16
L'essai clinique	19
La grille CONSORT	23
Le test diagnostique	26
Glossaire des termes utilisés	29
Exercices : chercher l'erreur	51

# L'ÉPREUVE DE LECTURE CRITIQUE D'ARTICLE ORIGINAL (LCA)

## 1. LE CONTEXTE

L'examen classant national (ou épreuves classantes nationales) informatisé qui valide le 2<sup>ème</sup> cycle des études médicales comporte une épreuve de lecture critique d'un article original. L'objectif de cette épreuve est l'incitation à lire de façon critique et à analyser le contenu d'un article. Il s'agit d'un exercice d'autoformation actuelle et future. Ce polycopié a pour but de faciliter l'apprentissage à cette lecture et son enseignement en fournissant :

- les modalités pédagogiques et docimologiques de l'épreuve
- les critères de qualité d'un article original et de son résumé
- un glossaire de méthodologie

Ces éléments devraient servir au-delà de l'apprentissage tout au long de la vie professionnelle d'un médecin quelles que soient les modalités d'exercice.

## 2. L'EXAMEN CLASSANT NATIONAL

L'examen Classant National (E.C.N.) comporte trois types d'épreuves destinées à explorer les capacités devant être acquises pour l'entrée du 3<sup>ème</sup> cycle :

- des dossiers cliniques ;
- des questions isolées ;
- une lecture critique de deux articles médicaux scientifiques.

Les modalités actuellement retenues pour l'Examen Classant National sont les suivantes :

- 18 dossiers cliniques progressifs répartis en 3 épreuves d'une durée de 3 heures chacune ;
- 1 épreuve de questions isolées d'une durée de 3 heures ;
- 1 épreuve de lecture critique d'article médical scientifique d'une durée de trois heures et portant sur 2 articles.

### 3. MODALITES DE L'EPREUVE DE LECTURE CRITIQUE D'UN ARTICLE MEDICAL ORIGINAL

L'épreuve portera sur des articles scientifiques originaux, à l'exclusion des revues générales et des articles didactiques :

- traitant d'un sujet mentionné dans le programme du 2<sup>ème</sup> cycle (B.O. n°31 – 30.08.2001),
- issus d'une revue médicale avec comité de lecture,
- en français ou en anglais.

Le résumé et éventuellement le titre et les noms des auteurs seront supprimés.

#### **Modalités de l'épreuve**

L'épreuve comporte 15 QRM ayant chacune 5 items. Les articles seront fournis, a priori, sous format papier, sauf évolution technique permettant une présentation entièrement dématérialisée.

#### **Modalités de correction**

Compte tenu de la nature de l'épreuve, la correction sera effectuée automatiquement à partir d'une grille élaborée par le jury avec l'aide d'un expert du CNCI.

### 4. OBJECTIFS PEDAGOGIQUES

#### **UE 1 : Apprentissage de l'exercice médical et de la coopération interprofessionnelle (1-20)**

##### **1. Objectifs généraux**

Entre autres objectifs, [...]

Il a un devoir de formation tout au long de la vie grâce à l'analyse critique des informations médicales et scientifiques et à l'apprentissage résultant du retour d'expériences.

L'exercice professionnel se doit de respecter un certain nombre d'éléments d'orientation qui peuvent être cliniques (recommandations et les discussions scientifiques évolutives qu'elles suscitent), éthiques (en rapport avec l'évolution des mœurs, de la société, des cultures, des valeurs qui fondent l'exercice médical et des contraintes qui viennent complexifier la décision médicale) et déontologiques ou légales (qui fondent la représentation des professionnels dans l'exercice de leur pratique et leur signifient les limites au-delà desquelles leurs responsabilités, de toutes natures, sont engagées).

##### **2. Objectifs terminaux**

#### **Item N° 3. Le raisonnement et la décision en médecine. La médecine fondée sur les preuves (Evidence Based Medicine, EBM). La décision médicale partagée.**

- Analyser les principes du raisonnement hypothético déductif et de la décision contextualisée en médecine.
- Décrire la démarche EBM ; en préciser les limites.

- Apprécier dans chaque situation clinique, le poids respectif des trois types de données constituant une approche EBM.
- Préciser la notion de niveau de preuve dans son raisonnement et dans sa décision.
- Identifier les circonstances d'une décision médicale partagée avec le patient et son entourage (voir item 319).
- Préciser les notions d'efficacité, d'efficience et d'utilité dans le raisonnement et la décision médicale.
- Comprendre et apprendre la notion de discussion collégiale pour les prises de décision en situation de complexité et de limite des savoirs.

**Item N° 14. Formation tout au long de la vie. Analyse critique d'une information scientifique et médicale. Gestion des liens d'intérêt.**

- Déduire de l'analyse d'une pratique, l'intérêt d'une actualisation des savoirs.
- Valoriser une information scientifique et médicale en fonction de sa qualité méthodologique et de sa pertinence bioclinique (voir item 323).
- Identifier les impacts potentiels d'un lien d'intérêt sur une information scientifique et médicale.
- Justifier l'utilité de la déclaration des liens d'intérêt d'un expert (auteur) et l'utilité d'une politique de gestion de ces liens potentiellement conflictuels.

**Item N° 18 - La méthodologie de la recherche expérimentale et clinique**

- Élaborer une question scientifique et effectuer une recherche documentaire sur un sujet, et une analyse critique des données existantes.
- Élaborer une hypothèse de travail et expliquer les processus expérimentaux mis en œuvre.
- Analyser, discuter et présenter un travail expérimental.
- Analyser et argumenter les grands types d'études cliniques.
- Décrire l'information du patient et le consentement éclairé.
- Connaître les grands principes de la réglementation de la recherche clinique (voir item 8).

**Item N° 20. Interprétation d'une enquête épidémiologique**

- Discuter et interpréter les résultats des principaux types d'enquête épidémiologique.

**UE 10 : Le bon usage du médicament et des thérapeutiques non médicamenteuses (318-326)**

Item 318 à 326

**1. Objectifs généraux**

L'objectif de cet enseignement transversal est de fournir, aux futurs médecins, les bases de pharmacologie médicale et de thérapeutique permettant l'usage rationnel du médicament dans un contexte de médecine praticienne, et les outils nécessaires pour assurer leur développement personnel continu dans le domaine du médicament.

Le bon usage du médicament et des thérapeutiques non médicamenteuses doit être fondé sur des données scientifiques validées et l'étudiant doit connaître les bases de la pharmacologie médicale, les principes de l'évaluation, les niveaux de preuve, les principales sources d'information et doit pouvoir critiquer un essai d'intervention ou une méta-analyse. Futur interne, il doit connaître les modalités d'utilisation des classes médicamenteuses d'utilisation courante chez l'adulte et chez l'enfant ainsi que leurs principaux effets indésirables. L'iatrogénie médicamenteuse doit être systématiquement envisagée et l'étudiant doit connaître la gestion des risques iatrogènes ainsi que l'organisation des dispositifs de signalement et de vigilance.

Seule une connaissance de la pharmacologie médicale et de la thérapeutique permettant une évaluation correcte du rapport bénéfice/risque peut conduire à une prescription de qualité.

À la fin de l'enseignement, l'étudiant doit être capable d'argumenter, en fonction de l'évolution des connaissances, la décision de prescription, le choix du médicament, les modalités de suivi et de réévaluation du traitement en tenant compte des caractéristiques du médicament, du patient, des coprescriptions et comorbidités.

#### **Item 320- Analyser et utiliser les résultats des études cliniques dans la perspective du bon usage**

- Argumenter l'évaluation d'un médicament ou d'une thérapeutique non médicamenteuse et les niveaux de preuve des principales sources d'information.
- Effet placebo et médicaments placebo, expliquer l'importance de l'effet placebo en pratique médicale et argumenter l'utilisation des médicaments placebo en recherche clinique et en pratique médicale.
- Argumenter une publication d'essai clinique ou une méta-analyse et critiquer les informations sur le médicament.
- Définir la taille d'effet et la pertinence clinique.
- Expliquer la transposabilité clinique et l'évaluation des médicaments au-delà des échantillons de population constituant les groupes d'étude.
- Interpréter une étude en pharmaco-épidémiologie.
- Sensibiliser aux liens d'intérêt et à leur impact potentiel sur l'information médicale (voir item 14).

#### **UE 12 : Formation générale à la recherche**

Cette formation, dispensée sous forme d'enseignements dirigés, porte sur la lecture critique d'articles scientifiques et a pour objectifs de :

- comprendre comment s'élabore une question scientifique
- apprendre à mener une recherche documentaire sur un sujet
- développer une analyse critique des données existantes
- comprendre comment s'élabore une hypothèse de travail
- comprendre les processus expérimentaux mis en œuvre
- savoir analyser les résultats présentés ainsi que leur discussion
- être capable de restituer et de synthétiser le contenu scientifique.

Les articles scientifiques sont choisis dans tous les domaines de la recherche biomédicale. La préparation à la lecture critique d'articles implique un travail en petits groupes tutorés, avec présentation orale et discussion interactive.

- Un stage de courte durée dans un laboratoire de recherche est souhaitable.
- La validation de cette formation permet l'obtention d'ECTS.

# L'ARTICLE ORIGINAL

## 1. LE TYPE D'ARTICLE

La bonne connaissance des principes de rédaction d'un article original fournit de précieux repères pour sa lecture critique.

Il existe **différents types d'articles** :

- original
- mise au point ou revue de la littérature
- méta-analyse
- article préliminaire
- article didactique
- lettre à la rédaction

L'article original rapporte les résultats d'une étude visant à confirmer ou infirmer une hypothèse de travail nouvelle. On distingue différents types d'articles originaux selon leurs objectifs.

TYPE D'ARTICLE ORIGINAL	OBJECTIF	SCHEMA EXPERIMENTAL
Essai thérapeutique	Evaluer l'efficacité d'un traitement	Essai comparatif randomisé, si possible en double insu
Test diagnostique	Démontrer la validité et la fiabilité d'un test	Comparaison d'un nouveau test à un test de référence
Recherche d'association et facteurs de risque	Déterminer si un facteur est responsable d'un événement	Etude de cohorte ou étude cas-témoin
Etude pronostique	Etudier le devenir des malades	Etude de cohorte
Etude descriptive	Caractériser une population	Transversale ou série de cas

## 2. STRUCTURE D'UN ARTICLE ORIGINAL

Un article original est structuré autour des **règles IMRAD** (**I**ntroduction, **M**éthodes, **R**ésultats and **D**iscussion)

La connaissance des recommandations facilite la lecture critique d'un texte.

**A. LE TITRE** : bref, clair, sans abréviation avec le maximum d'informations sur l'objet de l'article avec un minimum de mots. Reflet exact de l'article, attractif, sélectif mais non provocateur.

**B. L'INTRODUCTION** : doit permettre de **comprendre pourquoi le travail a été entrepris** et quel était son objectif

2 parties :

**1. Objectif informatif** : apporter au lecteur les éléments nécessaires et suffisants à la compréhension du travail (passerelle entre les connaissances de l'auteur et celles du lecteur).

Description du contexte qui a présidé à la mise en œuvre de l'étude.

**2. Préciser le but du travail** la question posée et l'objectif. L'hypothèse doit être clairement exposée. Chaque affirmation doit être référencée.

### **C. PATIENTS ET METHODES**

Suffisamment précis pour qu'un lecteur puisse reproduire ou vérifier le travail. Le nœud de la rigueur scientifique et de la crédibilité.

Ne comporte ni résultats ni commentaires.

Cette section décrit comment le travail a été réalisé en fournissant des informations précises, rigoureuses, devant permettre de reproduire le travail en totalité.

#### **1<sup>er</sup> OBJECTIF : les patients de l'étude (le " matériel " )**

- contient tous les détails nécessaires à l'interprétation des résultats
- doit permettre d'identifier les biais éventuels
- précise le critère principal de jugement et les critères accessoires.

#### **La description comporte :**

- les critères d'inclusion ou de non-inclusion, ils doivent être adaptés au but de l'essai (âge, sexe, CSP<sup>1</sup>, origine ethnique, symptomatologie, données morphologiques, groupes ou sous-groupes)
- le schéma expérimental : série consécutive ou non ; travail prospectif ou rétrospectif, randomisé ou non, ouvert ou en insu, primo traitement ou non
- les éléments ayant permis la construction d'un échantillon ou d'une population, calcul de la taille de l'échantillon
- traitements étudiés ; le traitement de référence est-il utilisé suivant les modalités habituelles ?
- Accord du CPP, consentement des patients.

#### **2<sup>ème</sup> OBJECTIF : ce que l'on cherche à évaluer**

- Action d'un médicament, résultats d'une intervention, valeur diagnostique d'un examen.
- Le critère de jugement principal et éventuellement un ou des critères secondaires est (sont) sont précisé(s) ainsi que les moyens permettant de les évaluer.

---

<sup>1</sup> CSP : Catégorie socio-professionnelle



### 3ème OBJECTIF : appréciation des résultats

- Valeurs normales retenues
- Tests et méthodes statistiques
- Précision : comment chiffrer un amaigrissement, comment quantifier une diarrhée
- Indiquer le nombre de malades exclus de l'analyse (en donnant la raison) ou perdus de vue

## D. RESULTATS

Ce chapitre " résultats " ne doit contenir que les résultats se référant directement aux critères de jugement.

**1. Tous les résultats** et rien que les résultats, y compris les résultats négatifs dans la mesure où ils apportent une information utile.

En cas d'essai thérapeutique la répartition des groupes et leur comparabilité à l'inclusion sont mentionnées, de même que le nombre et la cause des arrêts prématurés et des sorties d'essai, la répartition des perdus de vue et les effets secondaires.

**2. Les erreurs possibles de l'auteur**

- donner les résultats qui ne sont pas cohérents avec le but du travail (confusion)
- faire des commentaires

**3. Tableaux et figures**

Ils fournissent le *maximum d'information* sous une forme synthétique et claire. Ils évitent les répétitions entre l'exposé des résultats et leur rappel comme base de discussion dans le chapitre " discussion ".

Ils doivent être *informatifs* par eux-mêmes (légende, titre...)

Ils doivent être appelés dans le texte.

## E. DISCUSSION

**1. Précise si le but du travail exposé à la fin de l'introduction a été atteint ou non**

**2. Juge de la qualité et de la validité des résultats** : la discussion critique et objective porte sur chacun des chapitres de l'article. Commentaire sur les biais.

**3. Compare les résultats observés à ceux d'autres auteurs** en cherchant à expliquer les différences et en soulignant l'apport original.

### Les erreurs

- Répétition de ce qui a été dit dans l'introduction
- l'inexactitude des citations
  
- l'usage des temps : au passé tout ce qui relate un fait (résultat) ; au présent ce qui est opinion d'auteur et ce qui correspond à des notions bien établies

## F. REFERENCES

**1. Elles justifient tout fait énoncé** : principe fondamental de la démarche scientifique.

2. **Elles soutiennent les faits ayant amené les auteurs à formuler l'objectif du travail** (introduction), renvoient aux méthodes déjà décrites (matériel et méthodes) et argumentent les résultats critiqués dans la discussion
3. **Sélectionnées de façon pertinente**, elles doivent être appelées dans le texte

### 3. RESUME

Un bon résumé informatif doit pouvoir être lu indépendamment de l'article par un lecteur non spécialiste. Strictement fidèle au texte, il doit inciter à le lire.

#### A. PRINCIPES GENERAUX

Informatif il ne doit pas dépasser **250 mots**, ne comporter ni tableau, ni abréviation, ni référence bibliographique. Il est rédigé au passé (phrases construites avec un verbe) sauf pour l'introduction et la conclusion. Il répond aux quatre questions fondamentales :

- pourquoi ce travail a été fait ?
- comment ?
- qu'est-ce qui a été trouvé ?
- quelles conclusions ou quelle généralisation possible et autorisée
- sa construction reprend la structure IMRAD : introduction, matériel et méthode, résultats, discussion (limitée habituellement à une phrase de conclusion)
- la première phrase exprime l'idée directrice (sans répéter le titre)
- la seconde énonce la méthode de travail
- les erreurs à ne pas commettre :
  - o le résumé indicatif (au lieu d'informatif), réservé à la revue générale
  - o éviter les abréviations
  - o informations trop peu détaillées pour comprendre l'article

#### B. POINTS DE REPERES POUR LA REDACTION DU RESUME

Strictement limité à 250 mots, écrit au passé il ne comporte ni tableau, ni figure. Il est recommandé de mettre des inter-titres (introduction, méthodes, etc...) et d'indiquer un titre qui ne sont pas décomptés. Ecrit de façon très lisible en respectant la grille, il fait l'objet d'un " comptage " par l'étudiant en évitant les abréviations et selon les modalités indiquées en annexe. Le nombre de mots doit être indiqué. ( Voir la grille de rédaction et les modalités de comptage des mots en annexe )

**INTRODUCTION**            contexte : 1 à 2 phrases pour le rationnel de l'étude  
Objectif : quel est l'objectif principal ?

#### **MATERIEL ET METHODE**

##### **Schéma expérimental :**

- essai thérapeutique contrôlé, randomisé, en double insu ?
- test diagnostique : quel est le critère de référence ?

- étude pronostique : étude de cohorte ?
- étude de causalité : essai contrôlé, randomisé ? cohorte ? cas témoins ?
- étude descriptive : transversale ? série de cas ?

### **Lieu de l'étude**

- communauté, soins primaires, centre spécialisé, ambulatoire ou en hospitalisation

### **Patients**

- critères d'éligibilité (inclusion, non inclusion)
- nombre de patients inclus

### **Intervention**

- description et durée, nature de l'intervention

### **Critères de jugement**

- critère principal et critères accessoires retenus avant le début de l'étude.

### **RESULTATS**

- principaux résultats – moyennes, intervalles de confiance, signification statistique. Indiquer tous les résultats concernant le critères principal. Eviter les résultats " distractifs.

### **CONCLUSIONS**

Elles portent strictement sur les résultats, éviter les spéculations et les généralisations hâtives . Il n'y a pas de discussion des résultats. Autant que possible mettre en perspective et d'extrapoler la portée des résultats).

### **MOTS CLES (pour l'ECN)**

3 à 10 mots répertoriés dans l'Index Medicus.

Les " trucs " de rédaction du résumé (ECN)

- Ni délayage, ni style télégraphique : concision et précision.
- Ecrire les têtes de paragraphe pour débiter.
- Relever les phrases ou les fragments les plus significatifs dans l'article.
- Identifier l'objectif principal et la conclusion.
- Les mots-clés du paragraphie " méthodes "
- Les résultats principaux.
- Formuler l'hypothèse dans la première phrase du résumé.
- Respecter la cohérence.
- Ecrire au " passé " (le présent peut être employé dans la conclusion).
- Un " résumé " des résultats principaux est souvent fourni par les auteurs au début du paragraphe discussion (veillez à valider sa conformité avec les résultats).
- Ne rien écrire qui ne figure pas explicitement dans l'article.
- Exposer les résultats strictement comme dans le texte.
- Eviter les travers " les auteurs font... les auteurs analyses, proposent, estiment, etc... "
- Relire afin de vérifier que le résumé ne contient que l'information essentielle rendue sous une forme " lisible ".

# GRILLE DE LECTURE GENERALE

(Audet N. Ledère H., Pédagogie Médicale, 2001, 2, 206-212)

Une grille de lecture permet

- La compréhension globale d'une étude.
- L'évaluation de sa qualité méthodologique.
- L'évaluation de l'importance du résultat pour la pratique clinique.

## 1. LIRE DE FACON CRITIQUE : L'INVENTAIRE

### A) DETERMINER LA VALEUR SCIENTIFIQUE D'UN ARTICLE

- **pertinence de l'article** par rapport à son questionnement
- évaluer la **justesse des données**
  - . vérifier l'authenticité des faits
  - . s'assurer de l'exhaustivité des données
  - . distinguer un fait d'une opinion
  - . détecter les biais (erreurs systématiques) : évaluer si la constitution de l'échantillon permet la généralisation des résultats et si la taille est suffisante pour être pertinente
  - . dépister les relations de cause à effet erronées
  - . reconnaître les affirmations non valides
- évaluer l'**interprétation des données**
  - . critiquer le choix des tests statistiques
  - . déterminer le niveau de preuve proposé par le plan de recherche
  - . critiquer le choix du plan de recherche
- identifier les **généralisations fautives**

### B) APRES LA LECTURE, REALISER UNE SYNTHESE portant sur l'utilité clinique des résultats

- en général
- pour sa propre pratique
- capacité à induire un changement de comportement professionnel
- utilité en tant qu'outil d'autoformation en cours de lecture (métacognition)

## 2. LA TECHNIQUE DE LECTURE D'UN ARTICLE : cadre de référence

### CHOISIR CE QU'ON LIT

- motivation** :
- préoccupation fondamentale ou clinique
  - posséder une certaine maîtrise du sujet
  - disposer d'une stratégie efficace de recherche d'articles
    - banques de données, Internet
    - bibliothèque
      - expert
      - archivage d'articles

### DECIDER DE LA LECTURE

- être concerné par le sujet
- être familier avec le contenu médical et les concepts de biostatistique et d'épidémiologie clinique utilisés.

### IDENTIFIER LA PERTINENCE

- a. **Comprendre les concepts méthodologiques utilisés**
- b. **Evaluer la justesse des données**
  - distinguer un fait d'un commentaire  
(IMRAD – introduction, méthodes, résultats and discussion)
  - détecter les biais
  - vérifier l'exhaustivité des données
  - s'assurer de la qualité et de la pertinence des tableaux et figures
- c. **Evaluer l'interprétation des données**
  - niveau de preuve
  - plan de recherche
  - tests statistiques utilisés
  - généralisation des résultats en accord avec les limites permises par la méthodologie pour la population de référence

***Cette étude est ou n'est pas valide sur le plan scientifique.***

***Cette étude est ou n'est pas utile pour la pratique médicale.***

***Elle doit être confrontée à la grille de graduation des recommandations.***

**GUIDE D'ANALYSE DE LA LITTÉRATURE : NIVEAU DE PREUVE ET GRADATION DES  
RECOMMANDATIONS**

Agence Nationale d'accréditation et d'évaluation en santé (ANAES/HAS)

<b>Niveau de preuve des études</b>	<b>Grade des recommandations</b>
<b>Niveau 1</b> <ul style="list-style-type: none"><li>- essais comparatifs randomisés de forte puissance</li><li>- méta-analyse d'essais comparatifs randomisés</li><li>- analyse de décision basée sur des études bien menées</li></ul>	<b>A</b> <b>Preuve scientifique établie</b>
<b>Niveau 2</b> <ul style="list-style-type: none"><li>- essais comparatifs randomisés de faible puissance</li><li>- études comparatives non randomisées bien menées</li><li>- études de cohortes</li></ul>	<b>B</b> <b>Présomption scientifique</b>
<b>Niveau 3</b> <ul style="list-style-type: none"><li>- études cas-témoin</li></ul> <b>Niveau 4</b> <ul style="list-style-type: none"><li>- études comparatives comportant des biais importants</li><li>- études rétrospectives</li><li>- séries de cas</li><li>- études épidémiologiques descriptives (transversale, longitudinale)</li></ul>	<b>C</b> <b>Faible niveau de preuve scientifique</b>

Indépendamment de la méthode utilisée (parmi les méthodes présentées à la page précédente), un article se caractérise par l'objet principal de son étude, c'est-à-dire par le type de recherche qui y est présenté.

### Les grands types de recherche

La plupart des études s'inscrivent dans l'une ou l'autre des catégories suivantes :

- **Thérapeutique** : mesurer l'efficacité d'un médicament, d'une intervention chirurgicale, d'une nouvelle méthode d'éducation des patients ou de toute autre procédure. Le protocole de choix est l'essai comparatif randomisé.
- **Diagnostic** : montrer si un nouveau test diagnostique est valide (peut-on lui faire confiance ?) et fiable (fournit-il les mêmes résultats lorsqu'on le répète ?). Le protocole de choix est l'enquête transversale (*cross sectional survey*) où sont mis en œuvre simultanément le nouveau test et l'examen de référence (*gold standard*).
- **Dépistage** : estimer la valeur d'un test qui peut être appliqué à une vaste population et qui détecte une maladie à un stade présymptomatique. La méthode de choix est l'enquête transversale (*cross sectional survey*).
- **Pronostic** : prévoir comment évoluera une maladie diagnostiquée chez un individu à un stage précoce. La méthodologie de choix est l'étude de cohortes longitudinale.
- **Etiologie** : démontrer si un agent présumé pathogène, une pollution environnemental par exemple, est effectivement impliqué dans la survenue d'une maladie. Le protocole de choix est soit l'étude de cohortes, soit l'étude cas-témoins, selon la fréquence de la maladie mais les séries de cas peuvent également apporter des informations cruciales.

A chacun de ces grands types de recherche correspond un type d'article pouvant être décortiqué avec une grille de lecture générale mais aussi avec une grille qui lui est spécifique. On ne lit pas un article relatif à un essai thérapeutique comme un article consacré au développement d'un test diagnostique. Des grilles de lectures spécifiques vous sont donc proposées ici.

## GRILLES DE LECTURES SPECIFIQUES

### 1. Etudes épidémiologiques\*

Par niveau de preuves décroissant on distingue les études de cohorte, les études descriptives et transversales de population et les études cas témoins.

- L'étude de cohorte consiste à comparer l'incidence d'événements bien définis dans deux populations suivies pendant un temps donné, l'une exposée et l'autre non exposée. Elle permet le calcul du risque relatif. Le nombre de sujets nécessaires dépend de l'incidence attendue de l'événement dans la population et l'augmentation estimée du risque (ex tabac et cancer, dyslipidémie et accidents coronaires). Les études de cohorte peuvent contribuer à préciser les facteurs prédictifs ou pronostiques. L'étude de cohorte n'est pas réalisable pour des événements rares ou retardés nécessitant une population très importante ou une durée de suivi très prolongée.
- L'étude pronostique : c'est un cas particulier d'étude de cohorte qui s'adresse à une série de sujets déjà porteurs de la maladie à l'étude et pour lesquels on étudie le rôle pronostique (risque de rechute, de décès, d'aggravation) de différents facteurs.
- L'étude cas-témoin : elle compare les cas à des témoins en recherchant dans les deux groupes l'exposition à un facteur causal. Elle permet de calculer " l'Odds Ratio " ou rapport des cotes. Le groupe témoin doit idéalement avoir les mêmes caractéristiques que les cas. Elle est utile lorsque l'étude de cohorte n'est pas possible ou lorsqu'une étude expérimentale n'est pas possible.
- L'étude transversale descriptive d'une population permet d'établir des relations entre divers paramètres sans pouvoir affirmer qu'il existe une relation causale.
- La série de cas.

<b>L'information existe-t-elle pour chacune de ces 8 questions ?</b>	<b>La façon d'aborder la question est-elle correcte ?</b>	<b>Impact sur les conclusions</b>
<b>1.Objectif</b> - Pronostic – évolution - Impact d'une intervention - Etiologie – causalité	- Y a-t-il une hypothèse ? - L'objectif principal est-il clairement défini ?	
<b>2. Type d'étude</b> - Essai contrôlé randomisé - Etude de cohorte - Etude cas-témoins - Etude transversale - Rapport de cas – série de cas	- Le type de l'étude est-il approprié à la question posée ? - S'agit-il du type d'étude apportant le niveau de preuve le plus élevé	- Si non, les résultats de l'étude sont-ils totalement inutiles ?

\* Grille proposée par M. Velten et Mme C. Simon



<b>3. Facteur(s) étudié(s)</b> - Exposition - Intervention  - Test diagnostique	Sont-ils bien décrits ? Comment sont-ils mesurés ? - Même méthode de mesure chez tous les groupes ? - Méthode à l'aveugle ? Y a-t-il une comparaison indépendante avec l'étalon ?	Sinon ce biais de mesure menace-t-il la validité de l'étude ? <i>Idem</i> Sinon ce biais menace-t-il la validité de l'étude ?
<b>4. Critère(s) de jugement</b>	Le critère de jugement est-il unique ? Si non, a-t-on défini un critère principal Comment sont-ils mesurés ? - <i>Idem</i> question 3 Tous les critères de jugement pertinents sont-ils évalués ? En cas de critères multiples, les résultats sont-ils convergents ?	- <i>Idem</i> question 3 Si non, ceux qui ont été oubliés sont-ils importants ?
<b>5. Population source et sujets étudiés</b>	- Les critères d'inclusion et d'exclusion sont-ils précisés et pertinents ? - Y a-t-il des perdus de vue ? Si oui, leur nombre et leur répartition sont-ils précisés ? - Y a-t-il des non-réponses ? Comment ont-elles été prises en compte ? - A-t-on vérifié la robustesse des conclusions selon les hypothèses faites pour les non-réponses ?  - Y a-t-il randomisation ? A-t-elle été effectuée correctement ? Les effectifs permettant de vérifier les caractéristiques des sujets inclus et leur bonne répartition sont-ils présentés ?  - Les groupes diffèrent-ils par des caractéristiques autres que les facteurs étudiés ? - Quelle est la proportion de sujets atteignant la fin du suivi ? - Les perdus de vue sont-ils similaires aux sujets encore suivis ?	- Si des biais sont à craindre, pour chacune des questions précédentes, cela menace-t-il :  - la validité interne ?  - la validité externe ?
<b>6. Facteurs de confusion potentiels et biais</b>	- Sont-ils tous envisagés ? - Sont-ils bien contrôlés ? - Biais d'information (de mesure, cf 3.) ? Erreurs différentielles ?  Erreurs non différentielles ?  - Biais de sélection ?	- Si non, cela invalide-t-il l'étude ?  - Situation rendant l'étude inexploitable, si importantes - Conclusion possible uniquement si résultat significatif  - L'interprétation est-elle encore possible ?

<p><b>7. Résultats</b></p> <p>Type d'analyse utilisée ? Intervalles de confiance ?</p> <p>Tests statistiques</p> <p>- Si résultats positifs - Si résultats négatifs</p> <p>Force de l'association</p> <p>Cohérence externe</p>	<p>Est-elle adaptée ? A-t-on privilégié la présentation des estimations accompagnées d'un intervalle de confiance ?</p> <p>- Réalisés uniquement pour les hypothèses indiquées au départ ? - Respect des conditions d'application ?</p> <p>- Cliniquement intéressant ? - Taille de l'échantillon suffisante (puissance) ?</p> <p>- Est-elle indiquée ? - L'indicateur choisi est-il pertinent ? - Peut-on vérifier les résultats à partir des données présentées ?</p> <p>Les résultats sont-ils cohérents - avec ceux obtenus par d'autres études épidémiologiques ? - avec ceux obtenus dans d'autres disciplines, par d'autres méthodes ?</p>	<p>- Si non, peut-on les calculer à partir des données présentées ?</p> <p>- Le risque global d'erreur de première espèce est-il conservé ?</p> <p>- Si non, l'étude est-elle utile ? - Si insuffisant, l'étude est-elle utile ou non concluante ?</p>
<p><b>8. Conclusions des auteurs ?</b></p> <p>Réponses aux questions ? Vérification de l'hypothèse ? Objectif atteint ?</p>	<p>- Les conclusions répondent-elles à l'objectif ?</p>	<p>Les résultats sont-ils acceptables appliqués à la population source (validité) ?</p> <p>Les résultats peuvent-ils être extrapolés aux situations qui vous intéressent pour votre propre pratique (applicabilité) ?</p>

## 2. L'essai clinique (J.L. IMBS)

La lecture critique d'une publication d'essai clinique procède en trois temps :

- identification de l'objectif de l'étude,
- analyse méthodologique,
- pertinence des résultats pour la pratique clinique.

### 1. OBJECTIF

L'essai ne peut répondre qu'à une seule question : cet **objectif (principal)** doit être clairement décrit. Sa définition fera choisir un **critère de jugement principal** qui sera, entre autre, utilisé pour calculer le **nombre de sujets nécessaire**. Des **objectifs secondaires** peuvent y être associés : ainsi une analyse de la sécurité d'emploi du médicament (la tolérance) peut s'ajouter à un objectif principal ciblé sur l'évaluation de l'efficacité.

Il est souhaitable que l'utilité thérapeutique de la question posée soit explicitée, ou du moins située dans le contexte actuel (ces données étant référencées dans la bibliographie).

Schématiquement, il faut vérifier que :

- le but de l'essai est défini,
- de façon compréhensible ;
- la réponse à la question posée pourrait avoir une application clinique.

### 2. METHODOLOGIE

Un essai **prospectif, contrôlé** (vs placebo ou vs un **traitement de référence** reconnu), **randomisé** selon un **plan expérimental** adéquat mené en **double-insu** sur un **échantillon** parfaitement défini dont l'effectif aura été **calculé** selon la **puissance** choisie, utilisant un **critère de jugement** adapté à l'objectif, offrira un **niveau de preuve** optimal.

Ces paramètres seront considérés selon 3 questions (a, b, c)

a) *Comment les réponses aux traitements testées vont-elles être mesurées ?*

- Selon un plan expérimental correctement choisi (groupes parallèles, chassé-croisé, factoriel, séquentiel)
- Le critère d'évaluation (principal) répond bien à l'objectif de l'étude. Le choix des horaires de sa mesure, la sensibilité et la spécificité de l'instrument de mesure, éventuellement la variabilité inter-individuelle (paramètre utile pour le calcul du nombre de sujets nécessaire) sont connus. Autrement dit, il faut vérifier que : un critère principal est défini pour mesurer l'efficacité ; ce critère principal a une signification médicale claire ; il est adapté au but de l'essai.  
Chaque fois que possible c'est un critère quantitatif qui sera choisi (plutôt que qualitatif).
- L'échantillon de la population étudiée est parfaitement défini par des **critères d'inclusion et de non-inclusion**. Il faut se souvenir que des critères d'inclusion étroits permettront d'obtenir une homogénéité des patients inclus favorable à la détection d'une significativité statistique mais que les résultats obtenus sur cet échantillon très particulier pourront ne pas être extrapolables à la généralité des patients souffrant de l'affection considérée.
- Le choix du **traitement de référence** (ou traitement " contrôle ") répond aux données actuelles du traitement de l'affection considérée, qu'il s'agisse de l'emploi d'un **placebo** ou d'une spécialité pharmaceutique qui devra alors être utilisée à la posologie et pour la durée reconnues dans l'indication faisant l'objet de l'étude.

b) *Comment la comparabilité des groupes est-elle satisfaite ?*

- La répartition des traitements est faite par **tirage au sort**. La méthode utilisée est correcte (le plus souvent : table de nombre au hasard) S'il existe une stratification, elle est justifiée sur le plan clinique et ne multiplie pas les sous-groupes de façon excessive et préjudiciable à la puissance de l'étude.
- Le **double-insu** (éventuellement **simple insu**) est bien assuré, au besoin avec la participation d'observateurs externes (lorsque la mesure d'un critère de jugement risque de lever l'aveugle). La méthode du " double-placebo " est éventuellement utilisée, à bon escient.
- Le nombre de patients sortis de l'étude ou perdus de vue n'est pas différent selon les groupes (ou les périodes).

c) *Comment la puissance de l'essai est-elle assurée ?*

- Le **nombre de sujet nécessaire** a été calculé. Les risques  $\alpha$  (1<sup>ère</sup> espèce) et  $\beta$  (2<sup>ème</sup> espèce) acceptés sont indiqués. Cet effectif est atteint.

- Le nombre de sorties d'étude (< 10 à 15 % des inclus), de **perdus de vue** (< quelques %), ou de valeurs manquantes, n'est pas invalidant.
- L'**observance** des traitements comparés à été mesurée. Les moyens de favoriser cette observance ont été mis en œuvre.
- En cas d'**essai multicentrique**, le choix des centres participants est raisonnable et des efforts ont été faits pour standardiser au niveau de chacun d'eux les procédures de l'étude. Un tirage au sort séparé, équilibré par centre, a été fait.
- Le recueil des données et leur saisie ont été organisés de façon à réduire les erreurs de transcription.

### 3. ANALYSE DES RESULTATS

- a) La comparabilité des groupes expérimentaux à l'inclusion (au début de l'étude) est vérifiée. Le choix des critères de comparabilité est raisonnable.
- b) Des tableaux rendant compte des effectifs de sujets inclus, sortis d'étude, perdus de vue, avec les effectifs utilisés pour une analyse en **intention de traiter** ou en "**per protocole**" sont fournis. Les chiffres qu'ils mentionnent sont concordants, en particulier avec le texte de la publication.
- c) Le test statistique utilisé pour analyser la différence entre les valeurs du critère de jugement principal recueillies pour chaque groupe expérimental est bien choisi. L'analyse statistique porte sur la totalité des patients inclus dans l'essai (analyse en intention de traiter). Il existe une différence avec un  $p < 0.05$ . Ce résultat est retrouvé par l'analyse per protocole. Cette différence statistiquement significative est cliniquement intéressante.
- d) Les résultats d'éventuelles analyses en sous-groupe ne sont donnés qu'à titre informatif.
- e) Les résultats de l'analyse d'éventuels critères de jugement secondaires sont cohérents.

### 4. CONCLUSION

Elle reprend les faits démontrés sans donner d'opinion ou formuler de nouvelle(s) hypothèse(s). Elle indique si le résultat obtenu peut être extrapolé à l'ensemble des formes cliniques de l'affection traitée ou doit rester limité au type d'échantillon recruté pour l'étude. Elle souligne l'éventuel bénéfice thérapeutique pratique qui résulte de l'essai.

\* Les notions explicitées dans le glossaire figurent en **caractères gras** lors de leur première citation dans cette grille.

## LES DIX COMMANDEMENTS DE L'ESSAI THERAPEUTIQUE

(Grille de lecture d'un essai thérapeutique ou de prévention)

1. Essai contrôlé ?
2. Etude randomisée ?
3. Essai en double aveugle ?
4. Nombre de sujets nécessaires calculé.
5. Définition claire de la population.
6. Un critère principal de jugement unique et clinique pertinent.
7. Analyse en intention de traiter.
8. Une analyse en fin d'essai sur la totalité des patients.
9. Différence statistiquement significative.
10. Une différence cliniquement significative.

## **ANALYSE D'UN ARTICLE D'ESSAI THERAPEUTIQUE A L'AIDE DE LA GRILLE CONSORT.**

---

**La grille CONSORT (Consolidated Standards of Reporting Trials) est un outil-guide international destiné à évaluer la qualité de la rédaction d'un essai contrôlé randomisé.**

1. Titre et résumé.
2. Introduction : rationnel.
3. Critères d'inclusion, d'exclusion.
4. Description des traitements alloués.
5. Objectifs spécifiques de l'étude et hypothèses.
6. Critères de jugement principal et secondaire.
7. Détermination des effectifs nécessaires.
8. Méthode de randomisation.
9. Méthode illustrant les unités thérapeutiques après randomisation.
10. Définition des investigateurs responsables de l'inclusion, de la randomisation et de la distribution des unités thérapeutiques.
11. Simple aveugle ou double aveugle.
12. Méthodes statistiques utilisées pour le critère de jugement principal, analyse par sous-groupes ?  
Ajustements ?
13. Diagramme décrivant le nombre de patients à chaque stade de l'étude (flow chart) : screening, inclusion, randomisation, allocation des différents traitements, suivi de l'étude (perdu de vue, arrêt de traitement), nombre sur lequel a porté l'analyse.
14. Date de la période de recrutement et follow-up.
15. Caractéristiques cliniques et démographiques au début de l'étude des deux groupes comparés.
16. Nombre de participants dans chaque groupe et nombre de patients sur lequel porte l'analyse (analyse en intention de traiter).
17. Résultats précis avec intervalle de confiance à 95 % sur le critère de jugement principal et critères de jugement secondaires.
18. Analyses par sous-groupes, analyses après ajustement, statistiques.
19. Description des effets secondaires dans chaque groupe.
20. Interprétation des résultats, discussion des biais ou imprécisions et des dangers éventuels.
21. Applicabilité clinique des résultats de l'étude, les résultats de l'étude sont-ils extrapolables à la population générale (validité externe).
22. Interprétation générale des résultats et recommandations.

## QUESTIONS COMPLEMENTAIRES A LA GRILLE CONSORT

- Pertinence de la question : importance, originalité.
- L'hypothèse testée est-elle clairement exprimée ?
- Quelle est la comparabilité exprimée ?
- Quelle est la comparabilité des groupes ?
- Les facteurs pronostiques susceptibles d'intervenir ont-ils été pris en compte (stratification dans la randomisation, ajustements lors de l'analyse statistique).
- Cause des arrêts prématurés de traitement.
- Pertinence clinique d'une différence observée.
- Quel est le nombre de sujets à traiter pour éviter un événement ?
- L'étude est-elle éthique ?
- Les conclusions de l'étude sont-elles en rapport avec l'hypothèse de travail et concernent-elles le critère de jugement principal ?

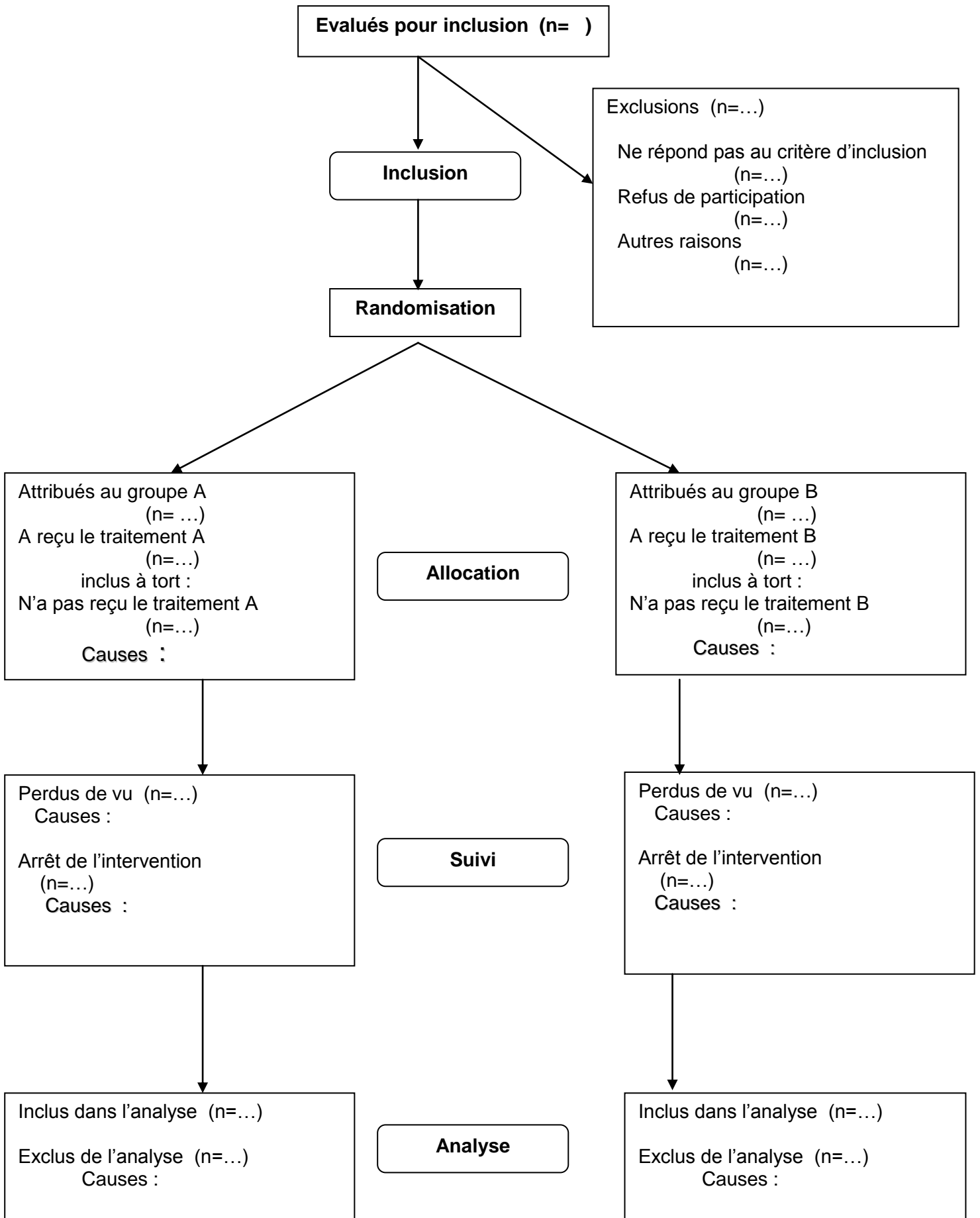
## LECTURES CONSEILLEES

- Dossier " Essais cliniques " paru dans la Revue du Praticien, 2000, 50, p. 827 à 869, et en particulier : BERGMANN J.F. et CHASSANY O. Lire un compte-rendu d'essai clinique, p. 838 à 845.
- BOUVENOT G., VILLANI P., AMBROSI P. Lecture critique de la publication d'un essai clinique. Presse Med., 2002, 31 : 1061-8.
- Dossier " La vie du médicament " paru dans la Revue du Praticien, 2002, 52 : 480-515

## DIAGRAMME DE FLUX (EN ANGLAIS : FLOWCHART)

Le diagramme de flux décrit sous forme d'organigramme le recrutement, la randomisation et le suivi de l'ensemble des patients lors de l'étude. Il permet de connaître rapidement le nombre d'inclusions, l'équilibre des groupes et l'éventuelle perte de cet équilibre au long de l'essai. Il permet donc de connaître à chaque étape de l'étude le devenir des sujets ayant participé à l'étude. Le diagramme de flux donné ci-dessous est le diagramme de la grille CONSORT.





### 3. Le test diagnostique\*

NB : par " test ", on entend toute information dont le résultat peut être utilisé dans une démarche décisionnelle, le plus souvent dans un but diagnostique. Ce peut être un examen biologique ou d'imagerie, un signe clinique, une procédure ou une stratégie (c'est à dire la combinaison de plusieurs tests)

1. Identifier ce qui est évalué : examen biologique ou autre, signe clinique, stratégie....
  - Le test évalué est-il suffisamment bien décrit pour être reproduit ?
  
2. Identifier l'examen de référence (" gold standard ")
  - Le test étudié doit être comparé à l'examen le plus performant disponible pour affirmer le diagnostic. Le choix de cet examen de référence est capital, il doit être justifié dans le texte avec des références adéquates. Lorsque cet examen n'existe pas, cela peut être une vérification autopsique ou chirurgicale, plus rarement une évolution clinique qui confirme a posteriori le diagnostic.
  - Tous les patients étudiés doivent avoir les deux tests, celui qui est évalué et celui de référence, et l'interprétation de chacun de ces tests doit être faite de façon indépendante, en " aveugle ".
  - La maladie et l'absence de maladie sont elles bien définies ?
  
3. La population étudiée est elle correctement décrite ?
  - Quelle est la fréquence de la maladie dans cette population ? Cette fréquence sera elle identique dans les populations auxquelles les auteurs destinent ce test ?
  - S'il y a des stades de gravité différente pour cette maladie, sont-ils mesurés ?
  
4. La reproductibilité du test est-elle analysée (variabilité intra/inter observateur) ? L'interprétation du test dépend elle de l'expérience d'un examinateur (par ex. examen radiologique) ?
  
5. Comment sont exprimées les qualités diagnostiques du test : sensibilité, spécificité, valeurs prédictives, rapports de vraisemblances (voir glossaire) ?
  
6. Les intervalles de confiance sont-ils fournis ?
  
7. L' " applicabilité " du test est elle discutée en tenant compte : de la fréquence de la maladie dans d'autres populations ? de la faisabilité technique ? d'une évaluation économique ? Ce test a t-il été placé dans le contexte des autres examens utilisés et de leur séquence ?

---

\* Grille proposée par B. Goichot et S. Rohr

## Rappels utiles pour le "test diagnostique"

	Malade	Non malade	Total
Test positif	VP	FP	VP+FP
Test négatif	FN	VN	FN+VN
Total	VP+FN	FP+VN	n

Lecture verticale :  $Se = VP/(VP+FN)$   $Sp = VN/(VN+FP)$

Lecture horizontale :  $VPP = VP/(VP+FP)$   $VPN = VN/(VN+FN)$

Sensibilité (Se) = probabilité d'avoir un test positif quand on est malade

VPP = probabilité d'être malade quand on a un test positif

Spécificité = probabilité d'avoir un test négatif quand on n'est pas malade

VPN = probabilité de ne pas être malade quand on a un test négatif

Sensibilité et spécificité décrivent les qualités *intrinsèques* du test.

Les valeurs prédictives dépendent des qualités intrinsèques du test et de la *probabilité pré-test*, c'est à dire de la prévalence de la maladie dans la population étudiée. Elles décrivent l'utilité diagnostique du test dans *une population donnée*.

Les *rappports de vraisemblance* (LR likelihood ratio) décrivent l'apport d'un test au diagnostic :

Le rapport de vraisemblance positif (L) d'un test est la vraisemblance de la présence de la maladie lorsque le test est positif. Il varie de 0 à l'infini. Si  $LR = 1$ , cela signifie que la probabilité après avoir fait le test est identique à la probabilité pré-test, donc que le test n'apporte rien au diagnostic. Plus il est élevé, plus le " gain diagnostique " est important.

Le rapport de vraisemblance négatif ( $\lambda$ ) est la vraisemblance d'être malade lorsque le test est négatif. Il varie entre 0 et 1. Plus il est proche de 0, plus il permet d'exclure le diagnostic.

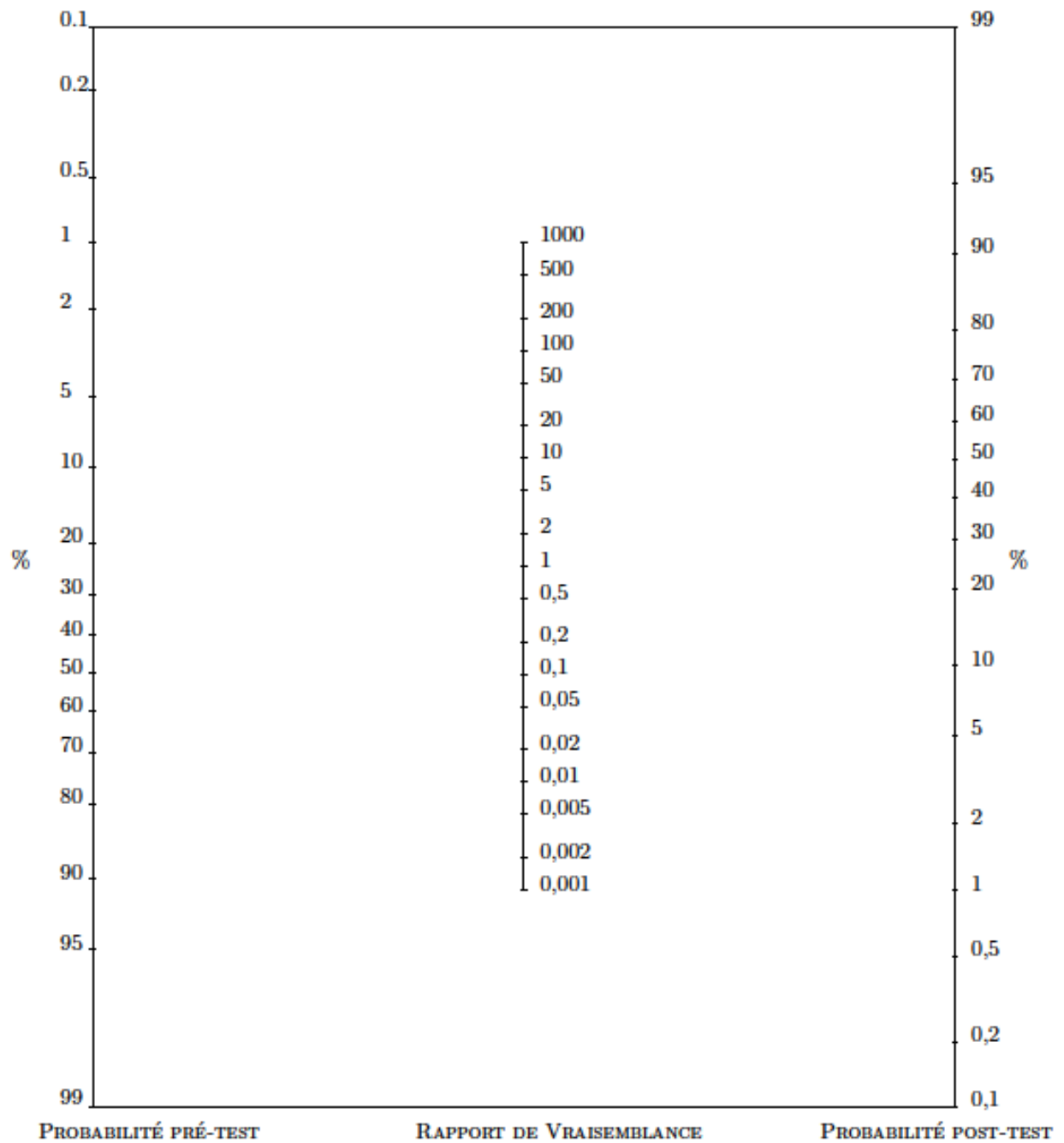
$$L = Se/(1-Sp) = (VP/malades)/(FP/non malades)$$

$$\lambda = (1-Se)/Sp = (FN/malades)/(VN/non malades)$$

Les rapports de vraisemblance sont indépendants de la prévalence de la maladie et sont un bon indice de la " valeur diagnostique " d'un test.

A partir de la probabilité pré-test d'être malade et du rapport du vraisemblance positive, le nomogramme de Bayes donne la probabilité post-test d'être malade. Pour cela, il faut prolonger sur l'axe de la probabilité post-test une droite passant par la probabilité pré-test et le rapport de vraisemblance positif.

# Nomogramme de *Bayes*



Réalisé avec  $\text{\LaTeX}$  par N. Meyer.  
D'après T.J. Fagan, NEJM;293:257.

# GLOSSAIRE DE TERMES SOUVENT UTILISES DANS LA REDACTION DES ARTICLES MEDICAUX

Nicolas MEYER<sup>①</sup>, Erik-André SAULEAU<sup>①</sup>, Jean-Louis IMBS<sup>②</sup>

<sup>①</sup> Laboratoire de Biostatistiques et Informatique médicale – Faculté de Médecine – 67085 Strasbourg Cedex, Pôle de Santé Publique, GMRC, CHU de Strasbourg  
<sup>②</sup> Institut de Pharmacologie – Faculté de Médecine – 11, rue Humann – 67085 Strasbourg Cedex

*Ce glossaire a pour objectif de rappeler le sens des principaux termes utilisés dans la rédaction d'article médicaux. Il va à l'essentiel des notions nécessaires dans la méthodologie biostatistique tout en limitant l'utilisation des formules et cite les principales composantes de l'essai clinique contrôlé. Il a été complété par certaines définitions figurant dans le glossaire proposé sur le site de l'ECN.*

## Ajustement

Méthode statistique permettant d'estimer l'effet propre d'un facteur sur un critère de jugement en tenant compte des effets d'autres facteurs.

## Aléatoire

Dont la survenue dépend du hasard. L'attribution aléatoire d'un traitement est obtenu par un tirage au sort du traitement à attribuer à chaque sujet. Le tirage au sort des groupes de sujets permet de les rendre initialement comparables pour tous les facteurs connus ou inconnus qui pourraient influencer sur le critère que l'on mesure. Souvent, cette hypothèse de comparabilité initiale des groupes est vérifiée numériquement et statistiquement lors de l'analyse des résultats sur les caractéristiques relevées au moment de l'inclusion.

## Analyse en intention de traiter

Dans un essai thérapeutique, faire une analyse en intention de traiter consiste à inclure dans l'analyse statistique tous les sujets ayant été inclus dans l'étude, que ces sujets aient ou n'aient pas *in fine* reçu le traitement attribué par le tirage au sort.

En toute rigueur, ce que l'essai contrôlé compare, c'est l'intention de traiter par un traitement A à l'intention de traiter avec un traitement B. Il faut donc conserver dans l'analyse tous les patients inclus dans l'étude selon leur répartition au hasard (randomisation). Cette allocation aléatoire est l'une des garanties que les deux groupes de patients (traité par A ou par B) sont comparables. Toute soustraction de l'analyse de patients (sortis d'étude, ou ne satisfaisant pas aux critères d'inclusion mais inclus par erreur ou pour lesquels des valeurs sont manquantes, ou encore perdus de vue) risque d'induire un biais.

Cette approche prend la réalité clinique en compte : il est par exemple impossible de continuer à traiter un patient par un traitement qu'il ne supporte pas uniquement pour des exigences d'analyse statistique. En corollaire, il faut signaler que cette analyse ne permet que de mesurer l'efficacité clinique, c'est à dire une efficacité définie dans le cadre de la pratique courante. L'analyse en intention de traiter peut être associée avec une analyse selon le protocole (ou "per protocole") dans laquelle des patients peuvent être exclus de l'analyse pour des raisons variées comme celles citées ci-dessus.

### **Analyse per protocole**

Cette analyse ne porte que sur les patients qui ont terminé et réalisé entièrement l'étude conformément au protocole et pour lesquels toutes les mesures prévues sont disponibles.

### **Analyse en traitement reçu**

Les patients sont analysés en fonction du traitement effectivement reçu. Un patient randomisé dans un groupe A recevant le traitement B sera analysé comme appartenant au groupe B. L'analyse est donc différente d'une analyse en per-protocole dans laquelle ce sujet aurait été exclu.

### **Analyse intermédiaire**

Analyse effectuée avant l'inclusion de tous les sujets prévus. Elle est réalisée le plus souvent lorsque l'étude est longue car elle permet éventuellement de conclure précocément à l'efficacité ou l'inefficacité du traitement étudié. Elle doit être prévue dans le protocole, et le nombre de sujets nécessaires prend en compte le nombre d'analyses intermédiaires qui sont prévues. Pour chaque analyse intermédiaire, un seuil de signification spécifique doit être choisi.

### **Analyse séquentielle**

L'analyse statistique d'une étude est classiquement réalisée une seule fois, à la fin de l'étude, lorsque toutes les données ont été recueillies. Dans la méthode séquentielle, l'analyse est réalisée après l'inclusion de chaque sujet afin de détecter le plus rapidement possible un produit efficace ou un produit dangereux ou inutile. Les études avec analyses intermédiaires rentrent dans le cadre des analyses séquentielles.

Ce type d'analyse nécessite des techniques statistiques particulières et permet en général de réduire le nombre de sujets nécessaire.

### **Appariement**

Technique permettant de rendre comparables deux ou plusieurs groupes, en particulier par rapport à certains facteurs de confusion déjà connus dont on veut neutraliser les effets (voir appariement).

### **Appariement (former des paires)**

Rendre comparables deux groupes en terme de facteurs de confusion potentiels. Pour chaque cas (exemple : un malade), on associe un ou plusieurs sujets qui lui sont similaires pour un ou plusieurs facteurs (exemple : âge, sexe, niveau socio-économique).

### **Biais**

Erreur *systématique* introduite par une erreur méthodologique. En présence d'un biais, l'échantillon sur lequel porte l'analyse n'est pas représentatif de la population à laquelle on s'intéresse et les résultats sont faussés.

### **Biais d'attrition**

Le biais d'attrition intervient lorsque des patients randomisés sont exclus de l'étude ou écartés de l'analyse au fur et à mesure que l'étude progresse ou encore lorsque les sujets quittent d'eux même l'étude. Ce biais est fréquent et assez facile à détecter lorsque les effectifs sont suffisamment bien définis dans l'article. Dans la pratique ce biais est limité par une bonne organisation de l'étude. Ce biais est induit par le

fait que les sujets qui quittent prématurément l'étude ne sont en général pas comparables aux sujets qui restent dans l'étude.

### **Biais d'évaluation**

Ce biais survient lorsque l'évaluation de l'effet d'un traitement ou d'une exposition est influencée par des facteurs subjectifs, survenant notamment lorsque l'étude n'est pas menée en double aveugle. Même si l'étude est en double-aveugle, il faut parfois faire intervenir un évaluateur externe pour mesurer objectivement le critère de jugement (encore appelé triple aveugle).

### **Biais de classement**

Biais dans la mesure du facteur de risque ou dans la détermination de la maladie. Ce biais consiste à classer le facteur de risque ou la maladie dans une catégorie erronée (présente alors qu'elle est absente ou vice versa). Cette erreur est quasiment inévitable puisqu'aucun outil de mesure (interrogatoire, examen, test) n'est parfait.

*Exemple : un comportement à risque minimisé par le malade, ou simplement non recherché dans le questionnaire.*

### **Biais de confusion**

Un biais de confusion est un biais entraîné par l'absence de la prise en compte d'un tiers-facteur pouvant influencer simultanément le critère de jugement et le facteur d'exposition. Ce facteur peut modifier artificiellement l'intensité de la relation entre le facteur d'exposition et la maladie si il n'est pas pris en compte lors de l'analyse statistique. Des méthodes d'ajustement peuvent prendre en compte un facteur de confusion. Un exemple classique en est le biais lié à l'exposition à l'alcool dans une étude sur l'effet du tabagisme dans les cancers ORL. Un autre exemple est celui du risque de suicide (maladie) en prison (exposition) pour lequel la présence d'une pathologie psychiatrique est un facteur de confusion.

### **Biais de sélection**

Ce biais existe lorsque deux (ou plusieurs) groupes construits ne sont pas comparables. C'est le cas lorsque l'on effectue une comparaison avec un groupe de référence (témoin) non construit à partir de la même population que celle des autres groupes. Ce biais existe donc essentiellement dans les études observationnelles, sans randomisation. Deux points sont à souligner :

- Dans un essai thérapeutique, la randomisation doit être effectuée selon les bonnes règles (p.ex. en centralisée si l'étude est multicentrique, équilibrée ou stratifiée si nécessaire).
- La bonne comparabilité des groupes doit être vérifiée sur les facteurs les plus importants car la randomisation ne donne pas une garantie absolue de comparabilité.

### **Biais de suivi**

Ce biais est introduit lorsque la comparabilité n'est pas maintenue tout au long de l'étude. Les causes de ce biais sont nombreuses comme par exemple les déviations au protocole, les traitements concomitants qui peuvent être différents, les arrêts du traitement pour diverses raisons.

### **Caractérisation (description statistique) d'une variable qualitative**

Une variable qualitative peut avoir deux modalités, comme la présence ou l'absence d'un signe (variable dichotomique), ou plus que deux niveaux, comme lors de l'évaluation de la gravité d'une maladie

(variable polychotomique, souvent ordinale si il y a un ordre naturelle des modalités). La distribution d'une variable dichotomique suit une loi binomiale, celle d'une variable polychotomique suit une loi multinomiale. On décrit de telles variables en calculant dans l'échantillon l'effectif et la fréquence de chaque modalité, c'est à dire le nombre et la proportion de sujets ayant telle ou telle modalité de la variable. On peut également ajouter l'intervalle de confiance de ces fréquences.

### **Caractérisation (description statistique) d'une variable quantitative**

Une variable quantitative est une variable numérique représentant souvent une dimension physique (longueur, valeur d'un dosage). La description statistique d'une variable quantitative se fait à l'aide de la moyenne et de son écart type ou de sa variance. D'autres paramètres sont souvent utilisés comme les valeurs extrêmes et les quantiles (quartiles, déciles). Dans le domaine de la biologie, la répartition des valeurs est souvent gaussienne, ce qui permet d'utiliser des tests dits paramétriques (Analyse de la variance ou test de Student). Lorsque ce n'est pas le cas, il convient d'utiliser des tests dits non paramétriques (Test de Mann-Whitney par ex.).

### **Causalité**

Relation de cause à effet entre un facteur et un critère de jugement. Cette notion est employée lorsque l'on souhaite prouver que la variation d'un facteur explicatif est la cause des variations d'un facteur à expliquer. En recherche clinique, la seule méthode permettant de prouver une relation de cause à effet est la méthode expérimentale. Rechercher une relation de cause à effet entre deux variables nécessite la construction d'une étude expérimentale dans laquelle l'une des variables est contrôlée (l'expérimentateur en fixe la valeur), la variable à expliquer étant simplement observée.

Remarque : L'existence d'une liaison entre deux variables obtenues autrement que par la méthode expérimentale ne prouve pas la causalité. Lorsqu'un coefficient de corrélation entre deux grandeurs quantitatives est statistiquement différent de zéro, on a uniquement montré que les deux variables n'étaient pas indépendantes (lorsque les distributions sont gaussiennes), la corrélation de deux mesures n'impliquant pas nécessairement un lien de causalité. En épidémiologie, la causalité repose sur un faisceau d'arguments et ne pas conclure à la causalité à partir d'une seule étude.

### **Cohorte de sujets**

Groupe de personnes suivies dans le temps de manière prospective.

### **CPP**

Comité de Protection des Personnes. Il est chargé de valider la conformité d'un protocole avec la législation, de vérifier que la sécurité des personnes se prêtant à la recherche est assurée notamment par l'évaluation de la balance bénéfice-risque. Si les aspects éthiques sont également pris en considération par le CPP, ce Comité doit être distingué du Comité d'Éthique qui s'occupe plus particulièrement de ces aspects.

Tous les protocoles de recherche clinique doivent être soumis à un CPP pour avis. Les études épidémiologiques sans investigations invasives sortent de ce champ. Il remplace le CCPPRB (Comité Consultatif de Protection des Personnes se prêtant à la Recherche Biomédicale) depuis 2006.



## Comparabilité des groupes

Un des critères de qualité d'un essai contrôlé est la comparabilité des groupes. Les groupes doivent être obtenus par randomisation à partir d'une population initiale clairement définie. Il faut veiller à maintenir cette comparabilité afin d'éviter d'introduire des variations non contrôlées dans l'étude ce qui conduirait à des biais de suivi, d'attrition ou d'évaluation. Il est important de vérifier, une fois l'étude terminée et ses résultats saisis, qu'il n'y avait effectivement pas de différence entre les groupes au début de l'étude. Exceptionnellement, le hasard (la randomisation) peut faire qu'une telle différence existe, rendant alors très difficile l'interprétation de l'étude.

## Comparaison de plusieurs groupes

Dans un essai contrôlé on peut être amené à comparer la moyenne d'une variable quantitative dans plusieurs groupes. La méthode statistique à utiliser est l'analyse de la variance à un facteur à plusieurs modalités (plusieurs groupes). L'analyse de la variance a pour but de comparer les moyennes tout en gardant le risque  $\alpha$  global égal à 5%. L'analyse conclut dans une première étape à l'absence d'une différence entre les groupes ou à la présence d'au moins une différence. Dans le second cas, l'analyse est complétée par la localisation des différences entre couples de groupes (par le test de Newman – Keuls ou le test de Scheffe).

Lorsque la comparaison porte sur une variable qualitative, le test statistique à utiliser est le test du  $\chi^2$ .

## Comparaison multiple

En augmentant le nombre de tests statistiques de comparaison sur un jeu de données, on augmente le risque de trouver au moins une différence statistiquement différente à tort. On dit que l'on modifie le risque  $\alpha$  initial fixé. Par exemple, si le risque  $\alpha$  est fixé à 0,05 pour un test, pour deux tests, le risque de conclure à tort à une différence (qui n'existe pas dans la réalité) sur au moins un des deux tests sera de 0,08 et pour trois tests il sera de 0,11. Une façon de corriger cette modification du risque est de protéger le niveau de risque du test selon la méthode de Bonferroni en baissant le risque  $\alpha$  initial en fonction du nombre de tests prévus (prendre un risque  $\alpha' = \alpha / n$  pour  $n$  tests).

## Consentement éclairé

Libre acceptation (exprimée par écrit) d'une personne envisageant de participer à un essai donné. Cette acceptation formelle ne doit être demandée qu'après information (dite "éclairée") de la personne, précisant les objectifs, les bénéfices, les risques et les inconvénients potentiels, liés à l'essai ; la personne doit également être informée de ses droits et responsabilités. Son acceptation doit être recueillie avant son inclusion dans l'étude.

## Critère de jugement (= critères d'efficacité, en anglais : *end points*)

L'essai contrôlé a pour objectif de comparer l'efficacité de deux traitements sur une maladie. Cette efficacité est jugée à l'aide d'un critère de jugement. Le ou les critères permettant de mesurer les effets d'un médicament dépendent des buts de l'essai et, évidemment, de la classe du produit étudié. Le critère de jugement principal de l'efficacité doit être choisi de façon à répondre à l'objectif principal de l'étude. En somme, il s'agit du mode de mesure du bénéfice que le patient peut retirer du médicament. Le critère de jugement peut-être qualitatif ou quantitatif. On utilise par exemple un taux de succès, une concentration sérique ou un taux de survie au temps  $t$  pour comparer les groupes. Dans un essai contrôlé, il n'y a qu'un

seul critère de jugement principal : c'est par rapport à lui que les effectifs nécessaires sont définis. D'autres paramètres que cette variable principale peuvent être intéressants à étudier : ils constituent des critères de jugement secondaires.

Il existe trois types de critères de jugement : clinique, intermédiaire, de substitution.

### **Critère clinique**

Mesure quantitative ou qualitative traduisant cliniquement l'état de santé du patient. On jugera par exemple l'efficacité d'un traitement sur une maladie par :

- amélioration de la probabilité de survie (le critère est ici le décès) ;
- prévention d'un état morbide ;
- disparition d'un symptôme douloureux, gênant ou invalidant ;
- guérison plus rapide ;
- amélioration de la qualité de la vie.

### **Critère d'inclusion / d'éligibilité / d'exclusion**

Ensemble des critères spécifiant les caractéristiques des personnes à inclure ou à exclure de l'étude.

### **Critère intermédiaire**

Critère de jugement, généralement biologique ou paraclinique, qui est modifié par le traitement mais qui n'est pas directement un résultat clinique. La modification par un traitement d'un critère intermédiaire n'implique pas un effet clinique mais est un argument en faveur de l'effet du traitement sur la pathologie.

### **Critère de substitution** (en anglais : surrogate end point)

Certains critères de jugement sont utilisés en lieu et place d'un autre critère. Cette substitution se fait quand :

- la preuve est faite que l'amélioration du critère de substitution coïncide certainement avec l'amélioration du critère substitué ;
- le critère de substitution est généralement d'acquisition plus facile que celle du critère substitué, d'où son utilisation.

Un exemple en est la baisse de la pression artérielle dans le traitement de la maladie hypertensive, laquelle devrait être théoriquement jugée sur ses conséquences (survenue d'un AVC par exemple). Un critère intermédiaire devient un critère de substitution lorsque l'on peut montrer que le critère intermédiaire est très corrélé avec le critère clinique qu'il est censé remplacer.

### **Critère composite**

Le critère de jugement utilisé dans une étude est le plus souvent simple, c'est-à-dire qu'il n'est constitué que d'une seule mesure (un dosage biologique unique, une mesure échographique unique) ou un seul évènement (décès). Un critère composite est, au contraire, un critère qui est constitué de plusieurs éléments de jugement. Cela peut-être par exemple une survie sans évènement, où on compte comme évènement la rechute d'un cancer, sa progression, le décès ou l'apparition d'une métastase. Le critère composite peut aussi être obtenu par combinaison de plusieurs paramètres biologiques comme par exemple les différents éléments d'un bilan lipidique.

Les critères composites ont l'avantage de donner une vision globale de la maladie et permettent parfois d'augmenter la puissance de l'étude mais ils sont souvent difficiles à interpréter, notamment lorsque les différents éléments qui le composent ont dans l'étude des évolutions discordantes.

### **Cross-Over (essai en chassé-croisé)**

Essai thérapeutique où le sujet est pris comme son propre témoin. Chaque patient reçoit l'ensemble des traitements à comparer. Dans le cas de deux traitements A et B, un groupe de patients reçoit le traitement A puis le traitement B, l'autre groupe de patients reçoit le traitement B puis le traitement A.

### **Description d'une population**

La description de la population étudiée permet au lecteur d'apprécier le contexte dans lequel l'étude est effectuée. Elle est nécessaire afin de connaître avec précision les critères d'inclusion et d'exclusion de l'étude, permettant au lecteur d'évaluer la portée de l'étude. Une description complète facilitera la comparaison des résultats avec d'autres études et pourra faciliter une méta-analyse.

La population étant définie par le problème étudié, l'étude sera en général effectuée sur un échantillon représentatif de cette population.

### **Différence de risque : voir Mesures de risque**

### **Degré de signification (ou de significativité)**

Risque (au sens de pari) de se tromper lorsque l'on conclut à une différence dans un test statistique alors qu'elle n'existe pas dans la réalité. Lors du test, on calcule la probabilité d'observer une différence au moins aussi importante que celle que l'on a obtenue : c'est le degré de significativité « p ». Si cette probabilité est inférieure à un seuil classiquement fixé à  $\alpha=5\%$  (0,05), on conclut que les deux groupes diffèrent. En procédant de cette manière, une différence apparaîtra comme significative dans 5% des cas où elle n'existe pas en réalité. Parmi 100 essais thérapeutiques comparant chacun deux traitements ayant en réalité des effets identiques, pour 5 d'entre eux on conclura à tort à une différence significative.

### **Double – insu (ou double aveugle, traduit littéralement de l'anglais "double-blind")**

Cette notion s'applique à un essai contrôlé au cours duquel ni le patient, ni le médecin ne connaît celui des traitements testés que reçoit le patient. Les patients ont cependant reçu une information appropriée leur expliquant quels sont les traitements utilisés dans l'étude. Les patients savent seulement qu'ils reçoivent l'un des traitements comparés mais ils ignorent lequel.

L'essai contrôlé en double insu a pour objet de limiter les biais qui pourraient s'introduire dans l'étude par la connaissance du traitement car patients et médecins pourraient alors manquer d'objectivité. Cela permet ainsi d'éliminer l'effet placebo chez le patient et les biais de mesure liés à la subjectivité du médecin. Dans l'essai en triple aveugle, le chercheur qui recueille les critères de jugement diffère de celui qui applique les traitements et il ne sait pas quel groupe de patients a reçu quel type de traitement.

### **Ecart-type**

L'écart-type d'une série de valeurs quantitatives quantifie la variabilité de cette série, c'est-à-dire la dispersion des mesures autour de la moyenne des mesures. Cette grandeur est utilisée dans de nombreux calculs en statistique, par exemple pour calculer un intervalle de confiance, calculer des effectifs nécessaires, etc...

## **Echantillon**

L'objectif de la recherche clinique est d'obtenir des résultats (des connaissances) présentant un caractère général (universel) valide sur l'ensemble d'une population. L'étude qui amènera cette connaissance ne peut que très rarement être effectuée sur toute la population. C'est pour cette raison qu'elle sera en général effectuée sur un échantillon représentatif de la population.

C'est à partir des connaissances obtenues sur l'échantillon que l'on induit les connaissances sur la population. On détermine certaines grandeurs à partir de l'échantillon, celles-ci constituant des estimations des grandeurs correspondantes de la population. Lorsque l'échantillon est bien constitué, la statistique inductive donne des indications sur la valeur des paramètres d'intérêts (risque relatif ou écart de moyennes par ex.) dans la population avec un certain risque d'erreur. On peut ainsi, à partir de l'échantillon, caractériser la population en effectuant une estimation ponctuelle de certaines grandeurs ou une estimation par intervalle de confiance de ces même grandeurs. Cette démarche nécessite au préalable la vérification de certaines hypothèses.

Constituer un échantillon, c'est tirer au hasard une sous-population d'effectif  $n$  dans une population (population parente) de taille  $N$ . La précision des paramètres estimés augmente avec la taille de l'échantillon.

*Remarque* : Tirer au hasard ou au sort, ou randomiser (de " random ", hasard en anglais), permet d'éliminer toute subjectivité (biais) dans la constitution de l'échantillon. Dans la pratique, pour obtenir un échantillon correct d'une population définie par des critères d'inclusion et d'exclusion, on peut se servir de tables numériques fournissant des séries de nombres aléatoires ou utiliser des programmes informatiques.

**Effectifs nécessaires : voir « nombre de sujets nécessaire »**

## **Effet *nocebo***

Effet négatif de la prise d'un médicament qui n'est pas lié aux propriétés physico-chimiques de la molécule, mais au fait même que l'on prend un traitement.

## **Effet *placebo***

Effet positif de la prise d'un médicament qui n'est pas lié aux propriétés physico-chimiques de la molécule, mais au fait même que l'on prend un traitement.

## **Enquête longitudinale**

Enquête au cours de laquelle des informations sont recueillies de façon longitudinale, c'est-à-dire à différents temps tout au long de l'étude. La durée de l'étude est en général définie à l'avance et peut être assez longue, durant souvent plusieurs années. Elle peut être prospective (idéalement) ou rétrospective.

## **Enquête transversale**

Enquête qui consiste à recueillir simultanément ou quasi simultanément les données relatives à la maladie et aux facteurs de risques étudiés.

## **Essai contrôlé (Essai comparatif randomisé, Essai thérapeutique)**

Essai dans lequel on dispose d'un groupe dit « contrôle » servant de référence. Idéalement, les groupes comparés doivent être constitués par un tirage au sort dans la population d'intérêt (définie par des critères d'inclusion et d'exclusion). Dans un tel essai, on peut mettre en évidence des relations de cause à effet. Cette étude est donc démonstrative, c'est-à-dire que la différence éventuellement mise en évidence peut être attribuée au facteur étudié. L'essai thérapeutique est une forme particulière d'essai contrôlé dans lequel on compare un ou plusieurs traitements à une référence.

Selon l'organisation, l'essai peut être en simple aveugle (simple insu) ou en double aveugle (double insu) selon que le patient seul ou le patient et le médecin ignorent quel est le médicament attribué.

### **Essai de non infériorité<sup>1</sup>**

Il existe de plus en plus de traitements reconnus, efficaces, validés. Un nouveau médicament dans une classe thérapeutique où existe un tel traitement de référence ne pourra pas, éthiquement, être comparé à un placebo dans le cadre d'un essai de supériorité. Ce nouveau médicament sera alors en général comparé au traitement de référence. Dans ce cas, le non rejet de l'hypothèse  $H_0$  (postulant que les traitements sont identiques) dans un essai de supériorité ne permet cependant pas de conclure à une égalité. Il est peut alors être intéressant de montrer que le nouveau médicament, à défaut d'être plus efficace que le traitement de référence, n'est pas moins efficace. C'est pour répondre à cette situation de plus en plus fréquente que la méthode des essais de non infériorité s'est développée.

Actuellement, un nouveau traitement présente souvent sur une plus grande facilité d'utilisation, ou une meilleure sécurité d'emploi ou encore un moindre coût, alors que son efficacité est similaire au(x) médicament(s) de référence. Ces avantages peuvent être jugés suffisants pour permettre d'accepter une certaine perte d'efficacité. Il s'agira alors de montrer que la différence d'efficacité entre les traitements comparés (nouveau médicament vs médicament de référence) est inférieure à un seuil préalablement fixé. Ce seuil correspond à la quantité d'efficacité qu'il est jugé acceptable de perdre face aux autres avantages du nouveau médicament. C'est le plus souvent une limite unilatérale qui sera fixée par ce seuil définissant la non infériorité. Cette méthodologie est très particulière à cette situation, clairement distincte des essais de supériorité.

Le choix du seuil de non infériorité relève d'une discussion d'experts pesant les avantages potentiels du nouveau médicament face à la perte d'efficacité acceptée. Il dépend de la classe thérapeutique, du critère de jugement principal, de l'efficacité du médicament de référence telle qu'elle est reconnue au moment de l'étude, et doit être défini à l'avance. En général, on conserve au moins 75 % de l'effet du traitement de référence par rapport à celui du placebo.

Plusieurs exigences méthodologiques découlent logiquement de ces concepts :

- Les résultats sont exprimés par l'intervalle de confiance à 95 % de la moyenne de la différence d'efficacité sur le critère de jugement principal entre les deux traitements. La non infériorité est démontrée si cet  $IC_{95\%}$  ne chevauche pas le seuil de non infériorité.
- Il est nécessaire d'éviter tout biais qui amoindrirait une différence sinon significative (choix des posologies, limitative du nombre de perdus de vue, homogénéité des sorties d'étude entre les bras, calcul du nombre de sujets).

---

<sup>1</sup> Note : Les essais de non infériorité font intervenir un domaine d'équivalence unilatéral défini par ce seuil. C'est un domaine d'équivalence bilatéral qui est fixé dans les " essais d'équivalence ", non traités ici car rarement utilisés dans les essais thérapeutiques

- L'efficacité du médicament de référence doit correspondre aux données de la littérature (validation externe).

### Essai en simple aveugle

Essai au cours duquel le patient ne connaît pas le traitement qu'il reçoit. Le médecin connaît le traitement que chaque patient reçoit. Cela permet normalement de neutraliser les effets placebo et *nocebo*.

### Essai ouvert

Essai thérapeutique n'utilisant pas de procédure de simple ou de double aveugle. Ce type d'essai, parfois comparatif, est souvent mené sur petit groupe de sujets, permettant d'étudier la faisabilité d'un essai comparatif à la plus grande échelle.

### Etude cas – témoins ; Enquête cas - témoins

Une enquête cas-témoins (ECT) est une étude dans laquelle l'échantillon est constitué en se basant sur le fait qu'un sujet présente la maladie (cas, M+) ou non (témoin, M-). On connaît les cas (les malades sont recrutés dans les hôpitaux ou les cabinets en ville) et l'on cherche des témoins (non malades). On dispose alors de deux groupes, plus ou moins comparables, sur lesquels on recueille le facteur à étudier, par exemple l'exposition (E+) ou non (E-) à un facteur de risque potentiel. Ce type d'enquête permet de construire un tableau de contingence à quatre cases contenant les effectifs pour les situations M+E+, M+E-, M-E+, M-E-. On peut calculer une liaison entre le facteur maladie (M) et le facteur exposition (E). Etant donnée la structure rétrospective de l'ECT, on ne peut pas calculer le risque relatif. Par contre, on peut calculer le rapport des cotes ou Odds Ratio (voir ce terme dans les mesures de risque) qui converge vers le risque relatif lorsque la maladie est rare. Ce type d'enquête est plus facile à réaliser qu'une étude prospective. Elle est cependant l'objet d'un certain nombre de biais.

Ce type d'étude est fréquemment effectué chez les patients en milieu hospitalier, certains patients malades étant pris comme « cas », d'autres patients non malades étant pris comme « témoins ». Cette situation pose souvent de nombreux problèmes d'interprétation.

	M +	$\bar{M}$ -
E +	a	b
$\bar{E}$ -	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).

Il faut remarquer que les taux de malades et de non malades ne sont pas représentatifs de la population. Par conséquent, les rapports [ a / ( a + b ) ] et [ c / ( c + d ) ] n'ont pas de sens (double trait vertical rappelant cette notion) et le calcul du risque relatif n'a pas de sens. L'odds-ratio peut, lui, être calculé.

### Etude d'une série de cas

Ce type d'étude correspond à l'analyse d'une série de cas sans pour autant avoir un objectif de comparaison avec un groupe témoin. Elle permet de faire le point sur une méthode dans un but descriptif ou ayant comme objectif préliminaire de faire une étude plus structurée dans un deuxième temps.

### Etude de cohorte ; Enquête de cohorte

Une enquête de cohorte est une étude dans laquelle un échantillon est constitué à partir de la population d'intérêt et qui est suivi dans le temps.

### Etude Exposé – non exposé

Une étude exposés - non exposés est une étude de cohorte dans laquelle les individus sont classés au début de l'étude dans différents groupes selon leur niveau d'exposition au facteur E et sont suivi dans le temps jusqu'à l'apparition éventuelle de la maladie d'intérêt (M). Ce type d'enquête permet de construire un tableau de contingence contenant les effectifs pour les situations M+E+, M+E-, M-E+, M-E-. Il est possible de déterminer s'il existe une liaison statistique entre les deux facteurs (par exemple par le calcul du test du  $\chi^2$ ) et l'on peut calculer le risque relatif. La liaison obtenue n'est cependant pas une liaison de cause à effet. Ce type d'enquête demande cependant une longue période d'observation pendant laquelle la comparabilité initiale s'estompe fréquemment.

	M +	$\bar{M}$ -
E +	a	b
$\bar{E}$ -	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d). La proportion de malades dans chaque groupe d'exposition est *a priori* représentatif du risque de maladie dans les mêmes groupes au niveau de la population correspondante. Le calcul du risque relatif est alors pertinent.

### Etude pronostique

Formuler un pronostic, c'est estimer les probabilités respectives des divers modes d'évolution d'une maladie. L'étude pronostique doit s'attacher à analyser l'évolution d'un groupe de sujet à partir d'une date ou d'un événement. Les études pronostiques s'appuient sur les études de cohorte ou de survie.

### Etude transversale ; Enquête transversale

C'est une étude dans laquelle l'exposition au facteur de risque et la présence de la maladie sont recueillies simultanément. Elle se distingue donc de l'étude cas-témoin et de l'étude exposée – non exposées où le recueil de ces deux informations sont dissociées dans le temps. Ce type d'étude permet de donner une vision de la relation entre maladie et exposition à un instant donné. Elle permet d'obtenir la prévalence de diverses maladies.

Elle a un certains nombre d'inconvénients, en particulier elle ne permet pas d'apprécier la séquence " Exposition au risque → maladie ".

Il est possible de déterminer s'il existe une liaison statistique entre les deux facteurs (calcul du  $\chi^2$ ) et calculer le risque relatif ou un odds-ratio. La liaison obtenue n'est cependant pas une liaison de cause à effet.

	M +	$\bar{M}$ -
E +	a	b
$\bar{E}$ -	c	d

Les termes a, b, c, d représentent des effectifs. L'effectif total est (a + b + c + d).

Ce type d'enquête est souvent difficile à mettre en place et demande de grands effectifs lorsque la maladie est rare dans la population.

## **Facteur de risque**

Un facteur de risque pour une maladie est un facteur dont la présence est associée statistiquement à une augmentation de la fréquence de la maladie. Lorsque le lien entre le facteur et la maladie n'est pas causal, on dit que le facteur est un *marqueur* de risque pour préciser que le facteur, si sa présence indique un plus grand risque d'avoir la pathologie en question, n'est pas forcément la cause de cette pathologie.

**Hazard ratio** : voir Mesures de risque

## **Incidence (taux d')**

Nombre de nouveaux cas dans une période de temps donnée dans une population donnée.

## **Intervalle de confiance**

L'intervalle de confiance d'un paramètre (proportion de malades, risque relatif, odds-ratio, écart de moyennes) est un intervalle de valeur dans lequel la vraie valeur du paramètre a une probabilité donnée de se trouver.

La vraie valeur du paramètre (dans la population) est inconnue et elle est estimée à partir d'un échantillon. L'estimation de la valeur d'un paramètre dans une population à partir d'un échantillon est entachée d'une incertitude puisque d'autres échantillons aléatoires issus de la même population donneraient d'autres estimations. Il est possible de calculer pour un paramètre un intervalle tel qu'il y ait 95% de chances pour que la vraie valeur du paramètre de la population soit dans cet intervalle. Cet intervalle est appelé « intervalle de confiance ».

On peut et il faut calculer l'intervalle de confiance pour tout paramètre statistique et pour toute mesure de risque. Ces intervalles de confiance doivent normalement être présents dans un article correctement rédigé. *Exemple* : OR = 1.53 [1.08 – 2.2]. Les valeurs entre les crochets indiquent respectivement la borne basse et la borne haute de l'intervalle de confiance de l'odds-ratio obtenu dans l'étude. Ces valeurs traduisent que la vraie valeur de l'OR a 95% de chance de se trouver dans un intervalle allant de 1.08 à 2.2.

## **Indépendance**

Neutralité de l'influence d'un événement A sur la survenue d'un événement B : deux événements sont indépendants si la probabilité d'observer l'un des événements ne dépend pas de la probabilité d'observer l'autre événement.

## **Intervention**

Une intervention est l'application d'une mesure (au sens large) destinée à améliorer la santé d'une population. La mise en place d'une campagne de vaccination, l'utilisation d'une nouvelle thérapeutique, la modification de l'accès aux soins sont des exemples d'intervention.

## **Latéralité d'un test statistique**

Un test statistique est soit bilatéral, lorsque l'on ne sait pas *a priori* quel est le sens de la différence entre les groupes que l'on observera, soit unilatéral lorsque la différence ne peut se faire, pour certaines raisons, que dans un sens ou n'a d'intérêt que dans un sens.



## Médicament

Selon l'OMS (Organisation Mondiale de la Santé), est un médicament toute substance ou composition administrée pour guérir ou prévenir une maladie ou un symptôme, établir un diagnostic, modifier ou restaurer les fonctions de l'organisme.

## Méthode expérimentale

L'expérience scientifique est une comparaison où tous les facteurs confondants (par ex. pour l'essai clinique : régression à la moyenne, effet placebo, évolution spontanée, terrain pathologique particulier, précision des instruments de mesure des critères de jugement), connus ou inconnus, jouent de la même manière sur les termes de la comparaison. Seul change un facteur, dit facteur expérimental, dont on fait varier l'intensité ou les modalités d'application pour juger de son impact sur un phénomène. L'hypothèse porte sur l'effet de ce facteur. L'hypothèse testée, sur laquelle s'appliqueront les tests statistiques, doit avoir été parfaitement formulée avant le début de l'expérience.

L'essai contrôlé est l'application à la recherche clinique des principes de la méthode décrite par Claude Bernard.

## Mesures de risque

Considérons deux groupes de sujets : l'un est exposé à un facteur et l'autre n'est pas exposé. On souhaite savoir si l'exposition au facteur augmente la fréquence de la maladie par rapport à l'absence d'exposition. Pour comparer les deux groupes, on calcule différents indices, appelés de façon générique " mesures de risque ". Ces mesures de risques sont :

**Le Risque Absolu (RA)**

**Le Risque Relatif (RR)**

**La Différence de Risque = Risque attribuable = Réduction absolue du risque (DR)**

**La Réduction Relative du Risque (RRR)**

**Le nombre de sujets à traiter (NST) (en anglais " *Number Needed to Treat, NNT* ")**

**L'Odds-Ratio (OR)**

**Le rapport de risque (Hazard Ratio, HR)**

Les cinq premières mesures sont utilisées dans les études prospectives tandis que la sixième (OR) peut-être utilisée à la fois dans les études prospectives ou rétrospectives (études cas-témoins). Le septième est obtenu dans une étude de survie.

Ces différentes mesures peuvent être résumées à partir du tableau ci-dessous. Les deux lignes correspondent aux deux niveaux d'exposition (exposition présente / absente) que l'on peut prendre au sens large et donc remplacer par l'attribution du traitement à l'étude ou d'un traitement de référence. Les deux colonnes correspondent au résultat de l'étude : sujets sains ou malades, que l'on peut remplacer de façon plus générale par la présence ou l'absence d'un évènement d'intérêt.

	M+	M-
E+	a	b
$\bar{E}$ -	c	d

Les termes " a, b, c, d " représentent des effectifs. L'effectif total est (a + b + c + d). Les termes  $a / (a + b)$  et  $c / (c + d)$  représentent les risques absolus (taux d'incidence) d'être atteint par la maladie dans l'un et l'autre groupe.

Les mesures de risque se calculent de la façon suivante :

Le **risque absolu** : c'est le risque de développer ou de présenter la maladie dans un groupe d'exposition donné. Le risque absolu chez les exposés est donc :  $RA^{E+} = a/(a+b)$  et chez les non-exposés, le risque absolu vaut :  $RA^{E-} = c/(c+d)$ . Le risque absolu n'est donc pas comparatif, par définition, contrairement aux autres mesures de risques. Ces deux risques absolus sont la base du calcul de plusieurs mesures de risques : RR, DR, RRR et NST.

Le **risque relatif**, noté RR :

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

C'est le rapport du risque absolu chez les exposés sur le risque absolu chez les non-exposés. Le risque relatif peut être calculé sur une étude de cohorte.

La **différence de risque**, DR :

$$DR = [a/(a+b)] - [c/(c+d)]$$

C'est la fréquence chez les exposés moins la fréquence chez les non exposés. Selon les besoins, on peut bien sûr calculer  $DR = [c/(c+d)] - [a/(a+b)]$  pour avoir une différence de risque de sens opposée (positive au lieu d'être négative ou vice-versa).

La DR est aussi appelée **risque attribuable**, RA lorsque l'on peut donner un rôle causal au facteur de risque.

La **réduction relative du risque**, RRR :

$$RRR = \frac{a/(a+b) - c/(c+d)}{c/(c+d)} = RR - 1$$

C'est la part de la différence de risque rapportée au risque dans le groupe non exposé. Il est en général utilisé dans les essais thérapeutiques où il mesure la diminution relative du risque en présence du traitement étudié.

Le **nombre de sujets à traiter**, NST (en anglais : " *Number needed to treat* "):

C'est le nombre de sujets à traiter pour éviter l'apparition d'un cas supplémentaire.

$$NST = \frac{1}{DR}$$

Remarque : dans cette formule, la différence de risque DR doit être exprimée par une valeur comprise entre 0 et 1 et pas par un pourcentage, qui varie de 0 à 100%. Donc, pour une différence de risque de 13%, le  $NST = 1/0.13 = 7.8$ , soit huit sujets. Il faut traiter en moyenne 8 sujets pour éviter un évènement.

L'Odds-Ratio, OR :

$$OR = \frac{ad}{bc}$$

On peut le calculer dans une étude prospective (étude de cohorte ou essai thérapeutique, même si ce n'est alors pas le meilleur indice) ou rétrospective (cas-témoins). Il constitue une bonne approximation du risque relatif quand la fréquence de la maladie est faible.

Tous ces termes sont des estimations obtenues à partir d'un échantillon d'individus. Il convient dès lors de leur associer un **intervalle de confiance**.

Le **rapport de risque** (ou hazard ratio, HR) est un rapport de taux d'incidence. Il est obtenu lors de la comparaison de la durée de survie entre deux groupes. Il ne faut pas le confondre avec le risque relatif même s'il peut s'interpréter de manière similaire.

### Mesures répétées

Une mesure est dite répétée lorsqu'elle est réalisée de manière itérative chez un même individu. L'analyse de ce type de données implique des méthodes statistiques adaptées à cette situation.

### Méta-analyse

Une méta-analyse est une « analyse d'analyses ». Elle consiste à colliger les données de plusieurs études réalisées sur une même problématique avec des protocoles les plus proches possibles puis à faire une analyse statistique globale afin de profiter de la puissance apportée par l'augmentation du nombre de cas. Elle tente de trancher des controverses apparues à la suite d'essais contradictoires. Elle tient compte de la variabilité des résultats des différentes études.

### Méthodes multivariées

Les modèles utilisés en recherche clinique sont complexes, une variable comme par exemple la valeur d'un dosage sanguin dépendant en général de plusieurs facteurs. On cherche souvent à effectuer une analyse « globale » en étudiant les relations entre une variable (dite variable à expliquer) et plusieurs autres variables (dite variables explicatives). De telles méthodes sont donc appelées multivariées. Citons par exemple la régression multiple, la régression logistique ou encore le modèle de Cox pour l'étude des facteurs pronostics dans les études de survie, etc... Un modèle multivarié va par exemple mettre en relation la valeur d'un dosage sanguin avec le sexe, l'âge, les antécédants médicaux et la valeur d'autres dosages, tous mesurés chez tous les sujets de l'étude.

Ces méthodes ou modèles multivariés doivent cependant faire suite aux méthodes univariées, évaluant le lien entre une variable à expliquer et une unique variable explicative, qui permettent de comprendre plus facilement les relations pouvant exister entre variables.

Il faut remarquer que dans toute étude on cherche à approcher la réalité médicale et biologique à travers des modèles définis. Il faut s'assurer que le modèle utilisé est capable de décrire le phénomène et par conséquent en valider son utilisation. Ceci est une démarche qui reste difficile car de nombreux facteurs tels que le modèle lui-même (additif, multiplicatif) peuvent intervenir.

### Moyenne

Il s'agit de la moyenne arithmétique définie comme la somme des valeurs divisée par l'effectif. Cette valeur a son importance puisque dans les modèles concernés par la recherche clinique il est fréquent de rechercher une différence de valeur moyenne entre deux ou plus de deux groupes lorsque la variable est quantitative. Le test statistique est construit pour tester une différence entre les valeurs moyennes des groupes.

**Nombre de sujets à traiter : voir Mesures de risque**

### **Nombre de sujets nécessaire**

L'objectif d'un essai thérapeutique est de montrer si une différence cliniquement pertinente peut être attribuée au traitement étudié, au risque  $\alpha$  de conclure à tort que la différence existe. Ne pas conclure lorsque le test est non significatif peut signifier que la différence attendue n'existe pas ou que le test manque de puissance. Pour éviter cette situation, la puissance augmentant avec le nombre de sujets de l'étude, il faut contrôler la puissance statistique du test par un calcul des effectifs nécessaires ce qui permet de se donner une probabilité élevée de trouver l'effet attendu.

Le calcul du nombre de sujets nécessaire est basé sur les paramètres suivants :

- le risque  $\alpha$  (risque de conclure à une différence qui n'existe pas) accepté (le plus souvent égal à 0,05, mais d'autres valeurs sont possibles) ;
- le risque  $\beta$  (risque de ne pas observer une différence alors qu'elle existe) accepté, le plus souvent égal à 0,10 ou 0,20 (ce qui correspond à une puissance statistique de 90 ou 80%) ;
- la différence jugée intéressante sur le plan clinique entre l'effet thérapeutique observé dans chacun des deux groupes ;
- la variabilité du critère de jugement (d'efficacité) principal, autrement dit la dispersion des valeurs mesurées sur l'échantillon, quantifié par l'écart-type ou la variance.

Remarques :

- L'importance de définir un objectif principal, impliquant l'utilisation d'un critère de jugement principal pour sa mesure, apparaît parfaitement : le calcul du nombre de sujets nécessaire sera fait selon la variabilité de cette mesure (et non de celle d'autres critères de jugement ou d'efficacité, dits alors secondaires) et la puissance de l'étude en découle directement.
- Le calcul des effectifs nécessaires demande la connaissance de certaines grandeurs assez faciles à obtenir dans les cas de comparaisons entre deux groupes, mais beaucoup plus difficile à obtenir dans des plans d'étude complexes, comme par exemple dans l'analyse de la variance à plusieurs facteurs ou de nombreuses grandeurs sont à connaître.
- Une fois le calcul du nombre de sujets fait, il est habituel d'augmenter le effectifs de 10 ou 15% de façon à compenser les éventuelles sorties d'étude.

### **Objectif principal, secondaire**

Toute étude doit être assortie d'un objectif principal, ce pour quoi on réalise l'étude. Cet objectif est par exemple de connaître le taux de survie d'un certain groupe de patient atteint d'un cancer donné ou bien de comparer un taux de guérison pour deux traitements. La définition de l'objectif principal est importante car

elle permet de spécifier la mesure qui servira de critère de jugement principal. Ainsi, pour connaître le taux de survie de patients atteints d'un cancer, il faudra définir et calculer un taux de survie à 1 an ou 5 ans par exemple et recueillir les éléments permettant de calculer ce taux.

L'objectif principal est unique (le plus souvent) et c'est sur lui que repose le calcul du nombre de sujets nécessaire via le critère de jugement principal lui correspondant.

Des objectifs secondaires sont très fréquemment utilisés. Ils ne servent pas à calibrer l'étude mais permettent d'explorer certains aspects du problème.

### **Observance** (en anglais : compliance)

Degré de coïncidence entre le comportement d'une personne et les conseils et prescriptions donnés par son médecin.

### **Odds ratio : voir Mesures de risque**

### **Perdus de vue**

Malades ayant quitté l'étude avant sa fin et pour lesquels aucune information n'est plus disponible depuis lors. Cette situation est différente de celle où le malade reste accessible mais pour lequel manquent certaines informations (valeurs manquantes). Un pourcentage de  $\geq 5\%$  de perdus de vue dans un essai est inacceptable.

### **Plan d'expérience** (plan expérimental)

Un plan d'expérience ou plan expérimental est l'agencement d'une expérimentation mise en œuvre dans le cadre de la méthode expérimentale.

Les plans les plus courants sont :

- **plan en groupes parallèles :**

Les participants sont randomisés en 2 groupes (ou plus) dont chacun va recevoir un des traitements testés.

- **plan en chassé-croisé (ou cross-over) :**

Chaque sujet est son propre témoin : il reçoit un traitement A pendant une première période et le traitement B pendant une deuxième période, l'allocation aléatoire s'appliquant à l'ordre d'application des traitements (A-B ou B-A) et non pas aux traitements, de sorte qu'autant de patients recevront d'abord B ou A.

- **plan factoriel :**

Utilisé pour évaluer la contribution de chacun des traitements prescrits en association. Par exemple, un plan factoriel conçu pour évaluer l'effet antihypertenseur de l'association d'un bêta-bloquants et d'un diurétiques comportera 4 groupes de traitements (parallèles) en double insu : association bêta-bloquant + diurétique ; bêta-bloquant + placebo du diurétique ; diurétique + placebo du bêta-bloquant ; placebo du bêta-bloquant + placebo du diurétique.

### **Prévalence**

Dans une population, la prévalence d'une maladie est la proportion de sujets présentant la maladie à un instant  $t$ .

### **Principe du test statistique**

La comparaison de deux traitements est réalisée sur le plan statistique par un « test statistique ».

Le principe est le suivant :

- On veut comparer l'effet de deux médicaments A et B.
- On fait d'abord l'hypothèse que ces effets sont identiques : c'est l'hypothèse « nulle »  $H_0$ . On pose ensuite une hypothèse dite « alternative »  $H_A$  spécifiant que la différence d'effet entre les deux groupes est supérieure ou égale à une valeur cliniquement pertinente.
- On applique ensuite chaque traitement à un groupe de sujets obtenu par tirage au sort.
- On quantifie la différence d'effet entre les deux traitements, par exemple par un écart de taux de succès ou par un écart de moyennes.
  - Si cette différence est « grande », on admet que les traitements diffèrent et on accepte l'hypothèse alternative. Cette conclusion est effectuée au risque  $\alpha = 0,05$  d'avoir rejeté  $H_0$  à tort.
  - Dans le cas contraire (différence « petite »), on admet que les traitements ne diffèrent pas et on conserve l'hypothèse nulle : on dit qu'aucune différence significative entre A et B n'a été mise en évidence. On ne peut toutefois pas affirmer que leurs effets sont identiques. La théorie statistique indique que cette conclusion est effectuée sous le risque  $\beta$  de ne pas avoir mis la différence en évidence si elle existe. Ce risque peut être diminué en augmentant les effectifs.

En vertu des lois de la statistique, les conclusions obtenues sur l'échantillon de patients sont extrapolées à la population des patients.

### **Puissance d'un essai clinique**

C'est l'aptitude d'un essai à mettre en évidence l'effet du médicament testé. Cette puissance relève de quatre composantes :

#### **- Puissance statistique :**

Le nombre de malades inclus est un élément essentiel de la puissance d'un essai clinique. La puissance statistique  $(1-\beta)$  est la probabilité de trouver une différence qui existe. Elle est complémentaire du risque  $\beta$  ou risque de 2<sup>e</sup> espèce qui est le risque de ne pas trouver une différence qui existe. Dans une étude, le calcul du nombre de sujets est réalisé de manière à assurer la puissance statistique souhaitée. Le nombre de sujets augmente avec la puissance et vice-versa.

#### **- Puissance pharmacologique :**

Elle est inhérente à la logistique de l'essai. Ainsi, il faut que le médicament à tester et médicament de référence soient administrés à posologies adéquates, que l'observance des traitements soit satisfaisante, que la fréquence des prises soit adaptée à la pharmacocinétique des médicaments testés.

#### **- Puissance (ou qualité) méthodologique :**

Le plan d'expérience doit être adapté à la situation explorée : par ex. un essai en chassé-croisé ne peut être utilisé lorsque la pathologie traitée varie spontanément et rapidement. La qualité de la réalisation de l'étude intervient également : sorties d'études, valeurs manquantes, participants perdus de vue, sont autant de paramètres susceptibles de réduire la puissance d'une étude.

#### **- Puissance (ou qualité) analytique :**

La méthode statistique doit être adaptée au protocole choisi et aux données recueillies. Ainsi, l'utilisation répétée des mêmes variables pour des comparaisons multiples demande une adaptation du seuil

du risque  $\alpha$  et entraîne une diminution de la puissance éventuellement compensée par une augmentation des effectifs de l'échantillon.

### **Randomisation**

C'est une méthode qui consiste à constituer des groupes de manière aléatoire, par tirage au sort des sujets ou unités qui les composent. La propriété importante de la randomisation est que chaque groupe est constitué de sujets similaires et donc que les groupes ainsi construits sont comparables. La randomisation permet d'éviter le biais de sélection et permet par conséquent de conclure à l'action du traitement lorsqu'une différence significative apparaît. Lors d'une randomisation, les allocations d'individus doivent être définies à l'avance. On pourra se servir de tables de randomisation pour affecter les individus aux divers groupes (dans le cas de deux groupes l'un des groupes recevra le traitement A, l'autre le traitement B).

Dans certains cas on peut utiliser des méthodes permettant d'équilibrer les effectifs tous les  $n$  patients (randomisation par blocs) ou randomiser de manière à représenter de façon équilibrée un paramètre dans les groupes (par exemple l'âge) en randomisant par stratification.

Comparer des patients recevant un nouveau traitement à des patients ayant reçu un ancien traitement n'est pas une étude randomisée et il est presque certain que des biais sont présents dans l'étude.

**Réduction absolue du risque : voir Mesures de risque**

**Réduction relative du risque : voir Mesures de risque**

**Référence** (médicament ou traitement de)

C'est le traitement du groupe servant de contrôle. Son choix dépend de la situation thérapeutique étudiée :

- il n'existe aucune thérapeutique pour la maladie considérée : placebo ;
- il existe un traitement établi (soit par des essais contrôlés antérieurs, soit par une longue et indiscutable expérience clinique) : médicament validé à la posologie active, ou, si l'éthique le permet, un placebo.

### **Régression vers la moyenne**

Soit la situation suivante : on mesure dans un premier temps un paramètre quantitatif (par exemple la cholestérolémie) dans un échantillon représentatif d'une population. Si l'on retient pour le second temps de l'étude par exemple les 25% de sujets ayant les valeurs les plus élevées de la cholestérolémie, une seconde mesure de la cholestérolémie chez ces sujets montrera une cholestérolémie moyenne plus basse que celle obtenue initialement sur ces mêmes sujets. La cholestérolémie moyenne de ce groupe a donc *régressé vers la moyenne* générale de la population. Ce phénomène s'explique par le fait que dans le groupe des sujets ayant des valeurs initialement hautes, certains de ces sujets avaient une valeur haute en raison d'une variation aléatoire à la hausse de leur valeur habituelle alors que les autres ont habituellement une cholestérolémie élevée. Le groupe sélectionné au premier temps est donc composé de deux types de sujets : ceux qui ont en permanence une valeur élevée de cholestérolémie et ceux qui ont une cholestérolémie haute de façon ponctuelle. Plus le second groupe est quantitativement important (en raison

d'une forte variabilité intra-individuelle de la mesure) plus le phénomène de régression vers la moyenne sera fort.

Ce phénomène est observé chaque fois qu'est suivie l'évolution d'un paramètre au sein d'un échantillon tronqué (i.e. limité par un seuil) d'une population.

Ainsi, l'identification de personnes hypertendues dans une population consiste à employer un seuil de pression artérielle. Lors des répétitions des mesures, la moyenne des valeurs de pression artérielle des sujets ainsi sélectionnés lors d'une première mesure va se rapprocher de la moyenne des valeurs de la PA de l'ensemble de la population. La régression à la moyenne est d'autant plus forte que le seuil choisi est plus extrême. Ceci explique des erreurs d'analyse, telles que par exemple : l'effet du traitement est d'autant plus marqué que les taux de départ sont plus élevés.

### **Représentatif**

Un échantillon est dit représentatif de sa population si la répartition des différentes caractéristiques des sujets qui composent l'échantillon ne diffère pas de la répartition dans la population. Un échantillon tiré au sort est représentatif.

### **Rétrospectif**

Une étude est dite rétrospective lorsque l'on y recueille des informations et des mesures concernant des événements qui se sont déroulés avant le début en pratique de l'étude.

### **Risque alpha ou risque de 1<sup>er</sup> espèce**

C'est le risque, accepté dans une étude, de conclure à une différence significative alors qu'en réalité elle n'existe pas. Ce risque est fixé arbitrairement et le plus souvent à 5% en médecine et en biologie mais d'autres valeurs peuvent se justifier.

### **Risque beta ou risque de 2<sup>e</sup> espèce**

C'est le risque, accepté dans une étude, de ne pas conclure à une différence significative alors qu'elle existe. Ce risque peut être diminué en augmentant les effectifs de l'étude.

**Risque absolu : voir Mesures de risque**

**Risque attribuable : voir Mesures de risque**

**Risque relatif : voir Mesures de risque**

### **Sensibilité**

La sensibilité est une notion qui caractérise les qualités diagnostiques d'un signe clinique (S), d'un test clinique ou paraclinique (T). La sensibilité est la probabilité conditionnelle de la présence du signe  $S^+$  ou du test positif  $T^+$  lorsque la maladie est présente ( $M^+$ ). C'est la probabilité d'avoir un test positif lorsque l'on présente la maladie.

	$M^+$	$M^-$
$S^+$ ou $T^+$	a	b



S <sup>-</sup> ou T <sup>-</sup>	c	d
----------------------------------	---	---

La sensibilité =  $[ a / ( a + c ) ]$  = Probabilité ( S<sup>+</sup> | M<sup>+</sup> ) ou encore Probabilité ( T<sup>+</sup> | M<sup>+</sup> )

L'effectif a correspond aux vrais positifs, l'effectif c correspond aux faux négatifs

**Signification/significativité statistique** : voir principe du test statistique.

### Simple – insu

Etude contrôlée dans laquelle le patient ne sait pas dans quel groupe de traitement il est affecté (il ignore quel traitement lui est attribué). Cette méthode a pour objectif d'éviter les biais liés aux réponses du patient.

### Spécificité

La spécificité est une notion qui caractérise les qualités diagnostiques d'un signe clinique (S), d'un test clinique ou paraclinique (T). La spécificité est la probabilité conditionnelle de l'absence du signe S ou de la négativité du test T lorsque la maladie est absente (M<sup>-</sup>). C'est la probabilité d'avoir un test négatif lorsque l'on ne présente pas la maladie.

	M <sup>+</sup>	M <sup>-</sup>
S <sup>+</sup> ou T <sup>+</sup>	a	b
S <sup>-</sup> ou T <sup>-</sup>	c	d

La spécificité =  $[ d / ( b + d ) ]$  = Probabilité ( S<sup>-</sup> | M<sup>-</sup> ) ou encore Probabilité ( T<sup>-</sup> | M<sup>-</sup> )

L'effectif d correspond aux vrais négatifs, l'effectif b correspond aux faux positifs.

### Survie

Une étude de survie est une étude dans laquelle on quantifie la distribution de la survie des sujets d'un groupe. La survie s'entend au sens large comme le temps s'écoulant entre l'entrée dans l'étude et la survenue d'un événement d'intérêt (rechute, aggravation de la maladie, décès). Au sens strict, elle ne concerne que la survie jusqu'au décès. La méthode de Kaplan-Meier et le modèle de Cox sont les deux méthodes statistiques les plus couramment utilisées pour décrire et analyser la survie d'un groupe de patients. Le modèle de Cox permet en outre l'étude des facteurs pronostiques indépendants. Ces méthodes tiennent compte de la durée de suivi du sujet jusqu'à l'apparition ou non de l'événement, puisque cet événement ne sera pas forcément observé chez tous les sujets. En effet, l'étude devant pouvoir être menée en un temps raisonnable, certains patients ne présenteront pas l'événement avant la fin de l'étude.

### Test diagnostique

Un test diagnostique est un examen clinique (signe, symptôme) ou paraclinique (radiographie, dosage biologique, questionnaire etc) dont le résultat permet de classer un sujet soit comme sujet sain soit comme malade. Pour évaluer l'intérêt de ce test diagnostique, on fait appel aux notions de Sensibilité, Spécificité, valeurs prédictives positives et négatives.

## Tests non paramétriques

Méthodes statistiques permettant la comparaison de deux ou plusieurs groupes sans avoir à faire d'hypothèses sur la distribution des variables étudiées. Les tests non paramétriques doivent être utilisés lorsque les conditions nécessaires à l'utilisation des tests paramétriques (en général caractère gaussien d'une distribution) ne sont pas satisfaites. Ces tests sont en général moins puissants que les tests paramétriques. Un exemple en est le test de Mann & Whitney.

## Valeurs manquantes

Les valeurs manquantes sont un problème très fréquent en recherche médicale. Lors de la lecture d'un article médical, il faut essayer de détecter d'éventuelles valeurs manquantes (par des effectifs variables d'une analyse à l'autre) et essayer d'évaluer l'existence de biais potentiels liés à ces données manquantes. Cela est néanmoins difficile à faire.

## Valeurs prédictives positives et négatives

Ces grandeurs mesurent la performance d'une stratégie diagnostique ou d'un test diagnostique.

La Valeur Prédictive Positive (VPP) est la proportion de malades parmi les sujets ayant un test positif. C'est le rapport du nombre de vrais positifs sur le nombre total de test positifs.

La VPP correspond à la probabilité pour un sujet d'être malade lorsque le test est positif.

	M <sup>+</sup>	M <sup>-</sup>
S <sup>+</sup> ou T <sup>+</sup>	a	b
S <sup>-</sup> ou T <sup>-</sup>	c	d

$$VPP = a / (a+b) = \text{Probabilité} ( M^+ | S^+ )$$

La Valeur Prédictive Négative (VPN) est la proportion de non-malades parmi les sujets ayant un test négatif. C'est le rapport du nombre de vrais négatifs sur le nombre total de tests négatifs.

La VPN correspond à la probabilité pour le sujet de ne pas être malade lorsque le test est négatif.

$$VPN = d / ( c+d ) = \text{Probabilité} ( M^- | S^- )$$

## Validité interne / externe

Une étude présente une validité interne si son organisation permet de répondre à ses objectifs. Une étude présente une validité externe si elle est cohérente avec les connaissances actuelles sur le sujet.

## CHERCHER L'ERREUR...<sup>1</sup>

(Lecture critique de résumé d'essai contrôlé)

### Questions

#### 1 **Diabète :**

Une étude a été menée sur 72 sujets tirés au sort dans une population de diabétiques de la maturité et ne recevant auparavant aucun traitement médicamenteux, pour savoir si un mois de traitement par un nouveau biguanide permettrait de faire diminuer la glycémie. A l'entrée dans l'essai, les moyennes de glycémie pour les sujets étaient de  $8,3 \pm 2,1$  mmol. Après un mois de traitement, la glycémie moyenne était de  $6,2 \pm 1,3$  mmol. La différence entre ces deux chiffres est statistiquement significative,  $p < 0,01$ . Cette étude montre donc l'efficacité de ce biguanide pour faire baisser la glycémie.

#### 2 **Antihypertenseur :**

Afin de tester l'efficacité d'un nouvel antihypertenseur inhibiteur de l'enzyme de conversion, un essai croisé a été effectué en médecine de ville chez 124 patients ayant tous une hypertension artérielle essentielle d'intensité modérée. Après deux semaines de wash out, tous les patients ont eu le nouveau traitement antihypertenseur pendant six semaines. Après le cross over, tous les patients ont eu six semaines de placebo. Une analyse de variance a montré que les chiffres tensionnels diminuaient significativement plus après les six semaines de traitement par IEC qu'après les six semaines de traitement par placebo. Ces différences significatives se retrouvent aussi si on fait une comparaison de moyennes et permettent de conclure qu'il existe bien un effet antihypertenseur supérieur à celui du placebo pour ce nouvel IEC

#### 3 **Chirurgie oesophagienne :**

Depuis Janvier 1995, le centre de chirurgie oesophagienne de l'hôpital X possède un système d'agrafage intra-oesophagien permettant de réaliser des sutures d'excellente qualité. La qualité de ce système d'agrafage a été prouvée, car durant l'année 1995, 73 patients ont pu bénéficier de ce nouvel appareillage. Il n'a été noté des complications locales post opératoires que chez 4 de ces patients. En revanche, dans le même centre chirurgical, avec les mêmes chirurgiens, durant l'année 1994 et en l'absence d'appareil d'agrafage, 83 sutures manuelles de l'oesophage avaient été réalisées et entraîné 13 complications locales post-opératoires. La différence dans la fréquence des complications est statistiquement significative en faveur de l'appareil d'agrafage (  $\chi^2$  corrigé  $p < 0,02$ ).

---

<sup>1</sup> \* Publié avec l'aimable autorisation du Professeur J.F. BERGMANN (Paris-Lariboisière)

#### 4 **Dyspepsie :**

Afin d'étudier l'efficacité d'un nouveau médicament antidyspeptique, une étude en parallèle double aveugle randomisée contrôlée versus placebo a été menée. Tous les patients inclus avaient une dyspepsie cotée par une échelle validée, avec un score moyen de  $3,2 \pm 1$  dans le groupe traité par placebo et de  $3,4 \pm 0,8$  dans le groupe traité par le nouvel antispasmodique (différence non significative à l'entrée dans l'étude). Après six semaines de traitement, le score moyen de dyspepsie est passé à  $1,4 \pm 0,3$  dans le groupe traité par l'antispasmodique et à  $2,8 \pm 0,5$  dans le groupe traité par placebo. La différence entre ces deux scores est statistiquement significative ( $p < 0,05$ ). Cette étude prouve l'efficacité de ce nouvel antidyspeptique comparé au placebo en terme d'amélioration d'un score de dyspepsie.

#### 5 **Laxatif :**

Afin d'étudier l'efficacité d'un nouveau laxatif, il a été mené un essai randomisé contrôlé versus placebo. Deux groupes de 30 patients ont reçu soit le placebo, soit le laxatif. Les deux groupes étaient comparables à l'entrée dans l'essai, avec une moyenne de 2,2 selles par semaine. Après 15 jours de traitement, le nombre de selles dans le groupe placebo est passé de  $2,2 \pm 0,8$  à  $2,8 \pm 0,9$  (différence non significative). En revanche, dans le groupe traité par le nouveau médicament, le nombre moyen de selles est passé de  $2,2 \pm 0,6$  à  $3,8 \pm 0,3$  (différence significative  $p < 0,01$ ). Ces résultats prouvent l'efficacité supérieure du nouveau laxatif par rapport au placebo en terme de nombre de selles.

#### 6 **Radiothérapie dans le cancer du côlon :**

Afin d'évaluer l'efficacité de la radiothérapie adjuvante en association avec la chirurgie du cancer du côlon, une étude a été menée de façon randomisée et contrôlée. Tous les patients chez qui l'indication d'une colectomie pour cancer avait été posée ont été tirés au sort. La moitié des patients a eu une colectomie dans la clinique chirurgicale A; l'autre dans le centre radiochirurgical B. La moyenne de survie à 5 ans a été de 38 % chez les malades avant eu une colectomie seule et de 58 % chez les malades ayant eu colectomie plus radiothérapie. La différence entre les deux groupes est statistiquement significative. Cette étude randomisée, contrôlée et menée en parallèle montre la supériorité de l'association radiothérapie plus chirurgie par rapport à la chirurgie seule.

#### 7 **Surfactant artificiel :**

Afin de tester l'efficacité d'un nouveau surfactant artificiel à utiliser chez les enfants prématurés en prévention de la maladie des membranes hyalines, une étude contrôlée a été menée dans une maternité parisienne. Le surfactant artificiel était instillé par un appareil spécial dont le Docteur X avait la responsabilité. Tous les enfants prématurés nés dans cette maternité lorsque le Docteur X était présent ont eu une instillation de surfactant artificiel. Les enfants nés dans cette maternité en l'absence du Docteur X n'ont pas eu de surfactant et ont constitué le groupe contrôle. Il y a eu finalement le même nombre de patients dans les deux groupes. La survie à deux mois a été finalement de 82 % dans le groupe traité par surfactant artificiel et de 64 % dans le groupe sans surfactant. Cette différence est

statistiquement significative et confirme l'efficacité de ce produit pour prévenir la mortalité des enfants prématurés.

8 **Yohimbine :**

Afin d'évaluer l'efficacité de la Yohimbine dans l'impuissance, une étude en demi cross over a été menée. Tous les patients incluables étaient dans un premier temps randomisés pour recevoir soit le placebo, soit la Yohimbine. Après trois semaines de traitement, les patients ayant reçu du placebo recevaient la Yohimbine, mais pour des raisons éthiques, les patients sous Yohimbine dans la première phase de l'étude ne recevaient pas le placebo. 22 patients ont été d'emblée traités par Yohimbine et 13 par placebo. 89 % des patients sous Yohimbine ont noté une amélioration de leur potentiel sexuel contre 49 % des patients sous placebo. Lorsque ces derniers ont été traités par la Yohimbine, 60 % d'entre eux ont noté une amélioration de leur potentiel sexuel. Dans tous les cas de figure, les résultats obtenus par la Yohimbine sont supérieurs à ceux obtenus par le placebo. Ceci confirme la supériorité de ce traitement par rapport au placebo pour le traitement des impuissances.

9 **Oxygénateur cérébral :**

On a comparé sur 2 groupes de 120 malades un oxygénateur cérébral à un placebo. Le critère d'évaluation principal était évalué par l'échelle d'appréciation clinique en gériatrie (EACG). Globalement les résultats ne montraient pas de différence significative. Cependant, si l'on éliminait les gens trop anxieux (score d'Hamilton > 18), en ne prenant en compte que les femmes (les hommes semblent moins répondeurs), la différence devient hautement significative sur la tranche d'âge 70 à 80 ans, ce qui montre bien l'activité du produit chez les gens âgés.

10 **Troubles liés au vieillissement :**

On a évalué l'activité du V. par une étude randomisée en double aveugle contre placebo, chez des patients atteints de troubles liés au vieillissement. Les critères d'évaluation déterminés à J0 J30 J90 comportaient 7 tests psychocliniques, 14 tests psychologiques, les résultats montrent la supériorité significative du V. en ce qui concerne les performances au test de Pacaud dès J 30, au trait marking test et au temps de réaction complexe à J90. Cette étude montre un effet significatif du V. en faveur d'une amélioration chiffrée des fonctions cognitives chez le sujet âgé.

11 **Extraits ribosomiaux :**

On a étudié l'effet d'injections d'extraits ribosomiaux dans la protection des poussées infectieuses chez des malades prédisposés. L'étude a consisté en l'observation comparative du "confort infectieux" de 26 patients traités depuis 2 ans par le produit, en regard des 2 années précédentes. Le nombre d'épisodes

infectieux a diminué de 72 %. Le nombre de jours d'antibiotiques de 69 %, le nombre de jours d'arrêt de travail de 64 %.. L'efficacité d'un tel traitement apparaît évident.

#### 12 **Artérite :**

Afin d'évaluer un nouveau traitement de l'artérite stade 2, on compare un extrait d'une plante africaine au tartrate d'I. Le nombre de sujets nécessaire pour mettre en évidence une amélioration de 40 % du périmètre de marche, au risque alpha de 5 %, avec une puissance de 90 %, est de 80 sujets par groupe. 236 patients répondants aux critères d'inclusion sont donc tirés au sort, pour recevoir l'un ou l'autre des traitements. Après un mois de run-in placebo (période de latence prévue au protocole avant l'introduction des principes actifs), 51 patients ayant eu une amélioration de 30% de leur périmètre de marche sont exclus de l'essai (placebo répondeurs ). 101 patients ont alors été traités par la plante, et 84 par le tartrate d'I. Au terme de l'essai, 63 % des patients du groupe "plante" ont vu leur périmètre de marche amélioré de 40 %, alors qu'ils n'étaient que 47 % dans le groupe "tartrate". Cette différence est hautement significative ( $p < 0,01$ ) en faveur de la plante.

#### 13 **Cancer du poumon :**

Le VP 16 a été comparé au VP 17 dans le traitement du cancer épidermoïde du poumon non opérable. Le critère principal était la survie à 6 mois. Au terme de l'essai, 37 % des patients traités par VP 16 et 43 % des patients traités par VP 17 étaient encore en vie (différence NS). Mais si l'on étudie le groupe des patients qui avait une réduction de la masse tumorale de plus de 50 %, après la troisième cure, on constate que la survie est de 46 % dans le groupe VP 16, et de 59 % dans le groupe VP 17 (différence significative  $p < 0,05$ ). On conclut donc que le VP 17 est plus efficace que le VP 16, en terme de survie à 6 mois, chez les patients répondeurs