



# Lecture Critique d'Article

## Glossaire

Edition 2 – Juin 2024

Cédric Dananché, Julie Haesebaert, Laure Huot, Behrouz Kassai  
Mikaïl Nourredine, Héroïse Rouzé, Fabien Subtil, Marie Viprey

Avec les contributions de Sabine Mainbourg,  
Anne-Marie Schott-Pethelaz

Ce glossaire de Lecture Critique d'Article présente des définitions unifiées de termes utilisés en LCA. Le vocabulaire peut varier selon les domaines d'application (de nombreux termes sont polysémiques). Ce glossaire ne se veut pas universel ; au-delà de la terminologie, l'essentiel est de comprendre les problèmes qui peuvent être rencontrés dans les études.


Un certain nombre de termes sont regroupés car reliés à la même notion ; ainsi, la présentation des termes ne suit pas toujours l'ordre alphabétique. De nombreux liens sont insérés pour naviguer d'un terme à l'autre.

Ce document ne se substitue pas aux cours de LCA.


La dernière version du glossaire est accessible en ligne sur le site : <http://kitmethodostat.univ-lyon1.fr>, rubrique LCA.


Légendes des symboles utilisés


 : lien vers un terme

 : terme connexe

 : méthode pour limiter un biais

 : terme polysémique








 : exemple

 : pour aller plus loin

 : éléments de calcul

## Table des matières

|   |  |    |
|---|--|----|
|    | Ajustement.....  | 8  |
|    | Aléatoire.....   | 8  |
|    | Analyse.....   | 8  |
|    | Analyse univariée.....   | 9  |
|    | Analyse multivariée.....   | 9  |
|    | Analyse en intention de traiter.....   | 10 |
|    | Analyse en intention de traiter modifiée.....  | 10 |
|    | Analyse per-protocole.....   | 10 |
|    | Analyse intermédiaire.....   | 11 |
|    | Analyse en sous-groupe (ou strate).....  | 11 |
|    | Appariement.....   | 12 |
|    | Biais.....   | 12 |
|    | Biais différentiel ou non différentiel.....  | 13 |
|    | Biais de classement (ou biais de mesure).....  | 13 |
|    | Biais de confusion.....  | 15 |
|    | Biais de sélection  | 16 |
|    | Biais d'attrition  | 17 |
|    | Biais de temps immortel.....   | 18 |
|    | Biais de contamination.....  | 20 |
|    | Biais de spectre.....  | 20 |
|   | Biais d'incorporation.....   | 21 |
|  | Biais de vérification.....   | 21 |
|  | Biais lié au délai entre les tests dans les études diagnostiques.....                                | 22 |
|  | Cas incident/Cas prévalent.....  | 23 |
|  | Causalité.....   | 23 |
|  | Comité d'éthique - CPP (Comité de protection des personnes).....                                     | 24 |
|  | Comité de surveillance et de suivi.....  | 24 |
|  | Clause d'ambivalence (ou principe d'incertitude).....  | 25 |
|  | Cohorte de sujets.....   | 25 |
|  | Comparabilité des groupes.....   | 25 |
|  | Consentement éclairé.....  | 25 |
|  | Corrélation.....   | 25 |
|  | Critère de jugement.....   | 26 |
|  | Critère de jugement clinique.....  | 26 |
|  | Critère de jugement intermédiaire.....   | 27 |
|  | Critère de substitution.....   | 27 |
|  | Critère de jugement principal et secondaire.....   | 27 |
|  | Critère de jugement principal multiple.....  | 28 |
|  | Co-primary endpoint.....   | 28 |
|  | Critère d'inclusion / non-inclusion / éligibilité (ou critère d'exclusion).....                      | 29 |
|  | Cross-Over.....  | 31 |
|  | Carry-over (effet).....  | 31 |
|  | d de Cohen.....  | 31 |
|  | Degré de significativité (valeur de p, significativité statistique).....                             | 31 |
|  | Données agrégées.....  | 31 |
|  | Dose-effet (relation dose-effet).....  | 31 |
|  | Différence moyenne standardisée.....   | 32 |
|  | Evidence Based Medicine.....   | 32 |
|  | Écart au protocole (déviations au protocole).....  | 33 |
|  | Effet nocebo.....  | 33 |

|   |    |  |    |
|---|----|--|----|
| ● Effet placebo.....  | 33 | ● Valeur prédictive positive.....  | 77 |
| ● Le placebo.....   | 34 | ● Probabilité de maladie pré-test.....   | 77 |
| ● Double placebo.....   | 34 | ● Ratio de vraisemblance positif.....  | 77 |
| ● Émulation d'essai cible (ou d'essai thérapeutique).....   | 34 | ● Ratio de vraisemblance négatif.....  | 78 |
| ● Étude : design.....   | 35 | ● Phases d'évaluation de méthodes diagnostiques.....   | 78 |
| ● Étude Cas-Témoin  .....    | 35 | ● PROMs/PREMs.....   | 79 |
| ● Étude de cohorte.....   | 37 | ● Populations.....   | 80 |
| ● Étude longitudinale  ..... | 39 | ● Population cible.....  | 80 |
| ● Étude transversale.....   | 40 | ● Population source/éligible.....  | 81 |
| ● Étude quasi-expérimentale.....  | 40 | ● Échantillon d'étude/population étudiée.....  | 81 |
| ● Designs quasi expérimentaux.....  | 41 | ● Échantillonnage stratifié.....   | 82 |
| ● Biais des études quasi-expérimentales.....  | 43 | ● Prévalence.....  | 83 |
| ● Étude : temporalité et recueil.....   | 43 | ● Prospectif  ..... | 83 |
| ● Prospectif  .....          | 43 | ● Protocole.....   | 83 |
| ● Rétrospectif  .....        | 43 | ● Puissance statistique.....   | 84 |
| ● Prolectif.....  | 44 | ● Qualités métrologiques.....  | 84 |
| ● Rétrolectif.....  | 44 | ● Exactitude / Justesse / Fidélité   |    |
| ● Épidémiologie.....  | 44 | Validité / Précision / Reproductibilité  |    |
| ● Essai clinique.....   | 44 |                     | 84 |
| ● Essai clinique contrôlé.....  | 45 | ● Questionnaire.....   | 85 |
| ● Essai clinique pragmatique.....   | 45 | ● Randomisation.....   | 86 |
| ● Essai clinique monocentrique.....   | 45 | ● Ratio de vraisemblance positif.....  | 89 |
| ● Essai clinique multicentrique.....  | 45 | ● Ratio de vraisemblance négatif.....  | 89 |
| ● Essai clinique : type d'intervention.....   | 46 | ● Régression linéaire.....   | 89 |
| ● Essai thérapeutique.....  | 46 | ● Régression logistique.....   | 89 |
| ● Essai de prévention.....  | 46 | ● Risque.....  | 89 |
| ● Essai clinique : hypothèse.....   | 46 | ● Risque alpha (ou risque de première espèce).....   | 89 |
| ● Essai de supériorité.....   | 46 | ● Risque bêta (ou risque de deuxième espèce).....  | 89 |
| ● Essai de non infériorité.....   | 46 | ● Rétrospectif    | 89 |
| ● Essai d'équivalence.....  | 47 | ● Rétrolectif.....   | 89 |
| ● Essai clinique : phases.....  | 47 |  |    |
| ● Essai de phase I.....   | 47 |  |    |
| ● Kappa de concordance.....   | 57 |  |    |
| ● Loi Jorjé.....  | 58 |  |    |
| ● Mesures d'événements / d'associations.....  | 59 |  |    |
| ● Incidence.....  | 61 |  |    |
| ● Ratio d'incidence.....  | 61 |  |    |
| ● Odds ratio (rapport de cotes).....  | 63 |  |    |
| ● Risque.....   | 64 |  |    |
| ● Différence de risque.....   | 64 |  |    |
| ● Risque relatif.....   | 65 |  |    |
| ● Méta-analyse.....   | 65 |  |    |
| ● Modèle.....   | 66 |  |    |
| ● Régression linéaire.....  | 66 |  |    |
| ● Régression logistique.....  | 68 |  |    |
| ● Modèle de Cox.....  | 70 |  |    |
| ● Interaction.....  | 70 |  |    |
| ● Monocentrique.....  | 71 |  |    |
| ● Multicentrique.....   | 71 |  |    |
| ● Multivarée (analyse).....   | 71 |  |    |
| ● Niveau de preuve.....   | 71 |  |    |
| ● Nombre de sujets nécessaires.....   | 72 |  |    |
| ● Etude comparative.....  | 72 |  |    |
| ● Etude non comparative.....  | 73 |  |    |
| ● Nombre de sujets à traiter (pour éviter un événement).....  | 73 |  |    |
| ● Objectif / Question de recherche.....   | 74 |  |    |
| ● Observance / adhésion (thérapeutique).....  | 74 |  |    |
| ● Perdu de vue.....   | 75 |  |    |
| ● Performances diagnostiques.....   | 75 |  |    |
| ● Sensibilité.....  | 75 |  |    |
| ● Spécificité.....  | 75 |  |    |
| ● Courbe ROC.....   | 76 |  |    |
| ● Valeur prédictive négative.....   | 77 |  |    |

|   |    |
|---|----|
| ● Risque relatif .....  | 89 |
| ● Run-in (période) .....  | 89 |
| ● Score de propension .....   | 90 |
| ● Sensibilité .....   | 90 |
| ● Spécificité .....   | 90 |
| ● Stratification .....  | 90 |
| ● Stratification de la randomisation .....  | 90 |
| ● Stratification en étude épidémiologique .....                                     | 91 |
| ● Sujets éligibles .....  | 91 |
| ● Survie .....  | 91 |
| ● Survie sans événement .....   | 92 |
| ● Analyse de survie .....   | 92 |
| ● Délai ou temps de participation, date de dernières nouvelles, date de point ..... | 92 |
| ● Courbe de survie .....  | 93 |
| ● Probabilité de survie .....   | 94 |
| ● Survie médiane .....  | 94 |
| ● Méthode de Kaplan-Meier .....   | 94 |
| ● Test du log-rank .....  | 96 |
| ● Taux de survenue d'événement .....  | 96 |
| ● Hazard ratio .....  | 96 |
| ● Modèle de Cox .....   | 96 |

|   |     |
|---|-----|
| ● Taux d'incidence standardisé .....  | 97  |
| ● Technique d'échantillonnage .....   | 97  |
| ● Test statistique .....  | 97  |
| ● Hypothèses d'un test statistique (fréquentiste) .....                     | 97  |
| ● Test bilatéral .....  | 97  |
| ● Test unilatéral .....   | 98  |
| ● Degré de significativité (valeur de p, significativité statistique) ..... | 98  |
| ● Risque alpha (ou risque de première espèce) .....                         | 98  |
| ● Risque bêta (ou risque de deuxième espèce) .....                          | 98  |
| ● Puissance statistique .....   | 98  |
| ● Tests paramétriques / non paramétriques .....                             | 99  |
| ● Test de tendance .....  | 99  |
| ● Tests multiples .....   | 100 |
| ● Univariée (analyse) .....   | 101 |
| ● Valeur de p .....   | 101 |
| ● Valeur prédictive négative .....  | 101 |
| ● Valeur prédictive positive .....  | 102 |
| ● Validité .....  | 102 |

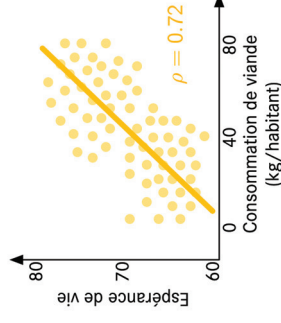
## A

### Ajustement

*adjustment*

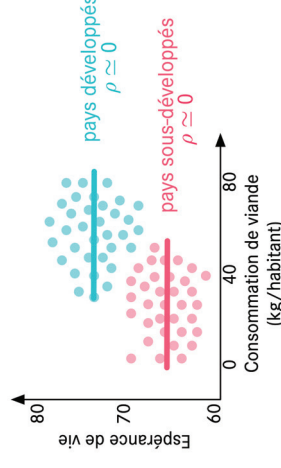
Moyen, dans une étude, de prendre en compte un facteur de confusion au moment de l'analyse. L'ajustement est réalisé dans les analyses statistiques multivariées permettant de prendre en compte plusieurs facteurs de manière simultanée.

### Analyse non ajustée



La consommation de viande semble augmenter l'espérance de vie

### Analyse ajustée



A niveau de développement comparable, la consommation de viande n'est pas associée à l'espérance de vie

L'association entre consommation de viande et espérance de vie en analyse non ajustée vient du fait que les pays développés ont, en moyenne, une espérance de vie plus élevée que celle des pays en développement, et qu'ils consomment plus de viande

### Aléatoire

*random*

Dont la survenue dépend du hasard.

### Analyse

*analysis*

Extraction et exploitation de résultats pertinents à partir d'une série de données. Le type d'analyse doit être prévu a priori dans le protocole de l'étude. Elle prendra en compte la question posée, les critères de jugement utilisés et d'autres variables, appelées covariables, qui peuvent interférer avec les critères de jugement. C'est en fonction du type de variables étudiées (qualitatives, nominales, ordinales ou quantitatives) que pourra se faire le choix des tests utilisés pour cette analyse. L'analyse peut être à visée descriptive, exploratoire (pour générer de nouvelles hypothèses, comme par exemple les analyses en sous-groupes) ou confirmatoire (en lien avec les objectifs définis).

## II Analyse univariée *univariate analysis*

Analyse dans laquelle on étudie l'association d'un seul facteur à la fois avec un critère de jugement.

## II Analyse multivariée *multivariate analysis*

Analyse dans laquelle on étudie l'action simultanée de plusieurs facteurs sur un critère de jugement. Elle permet de mesurer l'effet d'un facteur sur le critère à niveau comparable des autres facteurs : on parle d'effet ajusté du facteur. Ce type d'analyse est particulièrement approprié pour tenir compte de facteurs de confusion.

D'un point de vue statistique les analyses multivariées sont couramment effectuées par régression linéaire (critère de jugement quantitatif), régression logistique (critère de jugement binaire) et modèle de Cox (critère de jugement correspondant à un délai).

## II Analyse avec score de propension

*propensity score analysis; inverse probability weighting analysis*

L'analyse par score de propension est une analyse permettant de prendre en compte de multiples facteurs de confusion lors de l'évaluation de l'effet d'un facteur (par exemple lors de la comparaison de l'efficacité de deux traitements). On retrouve ce type d'analyse dans les études observationnelles (prise en compte du biais d'indication).

A la différence des analyses multivariées pour lesquelles l'effet du facteur étudié est ajusté sur l'ensemble des facteurs de confusion potentiels, pouvant rendre le modèle instable en cas de nombreux facteurs de confusion, l'analyse par score de propension réduit tous ces facteurs de confusion en un score. Dans le cas de la comparaison de deux traitements A et B dans une étude observationnelle (donc sans randomisation), le score correspond à la probabilité de recevoir le traitement A d'après les valeurs des différents facteurs de confusion. Les individus ayant des scores proches sont des individus ayant des valeurs de facteurs de confusion proches.

Une fois le score de propension estimé il peut être utilisé de différentes façons : par appariement, par pondération (inverse probability treatment weighting IPTW).

Au même titre que les analyses multivariées, l'analyse par score de propension est une façon de se rapprocher des résultats qui auraient pu être obtenus dans une étude randomisée (pour les facteurs pris en compte dans le score). Le score de propension n'a pas le même effet que la randomisation car il permet de contrôler les facteurs de confusion uniquement connus, bien mesurés et correctement modélisés.

## II Analyse en intention de traiter

*Intent To Treat analysis (ITT); full data set analysis*

Méthode qui consiste à tenir compte de tous les patients inclus dans les analyses (voir inclusion).

- qu'ils aient ou non eu des déviations au protocole;
- que leur critère de jugement ait été évalué ou non (par exemple non évalué car le patient a été perdu de vue); pour tenir compte des patients sans critère de jugement, des méthodes d'analyses particulières sont employées, comme l'imputation (remplacement) des données manquantes (voir hypothèse du biais maximum) et ce dans le « bras » (groupe de tirage au sort) dans lequel ils ont été randomisés au début de l'étude (qu'ils aient ou non bénéficié du traitement alloué lors de la randomisation, qu'ils aient changé de bras de traitement, ou qu'ils aient arrêté leur traitement).

L'analyse en intention de traiter permet d'éviter le biais d'attrition dans l'essai clinique. Elle a tendance à sous-estimer l'efficacité réelle du traitement, et est souvent privilégiée par les autorités de santé lors d'étude de supériorité, ces dernières considérant que l'erreur la plus grave est de déclarer efficace à tort un traitement qui ne l'est pas.

## II Analyse en intention de traiter modifiée *modified-Intention to treat analysis (mITT)*

De façon stricte, il n'existe qu'une seule définition de l'ITT et tout ce qui n'est pas ITT est une analyse en per protocole. Néanmoins il existe des analyses, proches de l'ITT, que l'on appelle ITT modifiée. Il faut être vigilant à ce qu'une analyse en per protocole ne soit pas abusivement appelée m-ITT. L'analyse m-ITT la plus acceptable (sans analyse ITT associée) est celle excluant des patients randomisés dans un groupe et qui n'ont pas reçu au moins une fois le traitement; et cela dans le cadre d'un double aveugle. Ainsi la sortie d'étude ne peut pas être en lien avec la connaissance du bras de traitement ni par le patient, ni par le médecin.

Le terme mITT englobe une grande variété de cas de figures qui, en dehors du cas précédent, sont à considérer comme des analyses discutables.

## II Analyse per-protocole *per protocol analysis*

Méthode qui consiste à analyser uniquement les données des patients inclus ne présentant pas de déviation au protocole, et ce dans le « bras » (groupe de tirage au sort) thérapeutique correspondant au traitement que le patient a réellement reçu au cours de l'étude. Les modalités de l'analyse per-protocole peuvent être variables d'une étude à l'autre. Toute analyse qui n'est pas en intention de traiter est per-protocole. L'analyse per protocole mesure l'efficacité maximale théorique du traitement expérimental (i.e. dans les meilleures

conditions possibles). L'efficacité réelle du traitement (i.e. que l'on peut espérer en condition clinique normale) se situe entre celle estimée en intention de traiter et celle en per-protocole.

A l'exception des essais de non-infériorité  $\otimes$ , l'analyse statistique principale est toujours effectuée en intention de traiter. L'analyse per-protocole est une analyse secondaire.

## II Analyse intermédiaire

*interim analysis*

Analyse effectuée avant la fin prévue de l'étude. Elle est réalisée le plus souvent lorsque l'étude est longue. Son intérêt est de pouvoir arrêter précocement l'étude, par exemple pour efficacité précoce, pour futilité, ou pour toxicité excessive. Elle doit être prévue dans le protocole  $\otimes$ , et le nombre de sujets nécessaires prend en compte le nombre d'analyses intermédiaires envisagées. Pour chaque analyse intermédiaire, un seuil de signification statistique doit être choisi.

$\otimes$  voir aussi partie Tests multiples

## II Analyse en sous-groupe (ou strate)

*subgroup analysis*

Les analyses en sous-groupe sont utilisées pour évaluer le bénéfice du traitement expérimental par rapport au traitement de référence dans différents sous-groupes (exemple : âge, sexe, mutation génétique en médecine personnalisée), et émettre des hypothèses concernant les sous-groupes pour lesquels le traitement expérimental serait plus efficace que le traitement de référence.

Elles sont utiles sur le plan clinique, le médecin ne traitant pas des patients moyens, mais des patients avec des caractéristiques individuelles ; les sous-groupes permettent de se rapprocher des caractéristiques des patients.

Sur le plan statistique, on recherche une interaction entre le bras de traitement et le sous-groupe (voir partie Modèles : interaction  $\otimes$ ). Ces analyses exposent à une inflation du risque  $\alpha$   $\otimes$  (voir partie Test statistiques : tests multiples  $\otimes$ ). Par ailleurs, l'effectif de l'étude est souvent insuffisant pour détecter les effets dans les sous-groupes (manque de puissance). Ces analyses sont le plus souvent exploratoires.

Si les analyses en sous-groupe ne sont pas prévues dans le protocole  $\otimes$  initial (hypothèses non définies a priori), les conclusions post-hoc qui en découlent nécessitent d'être validées dans des études dédiées, afin de satisfaire le raisonnement hypothético-déductif.

Dans les essais cliniques, les résultats des analyses en sous-groupes sont souvent présentés sous la forme de graphique de type forest-plot  $\otimes$ .

## II Appariement

*matching*

Technique permettant de rendre comparables deux ou plusieurs groupes sur certains facteurs de confusion  $\otimes$  déjà connus dont on veut neutraliser les effets.

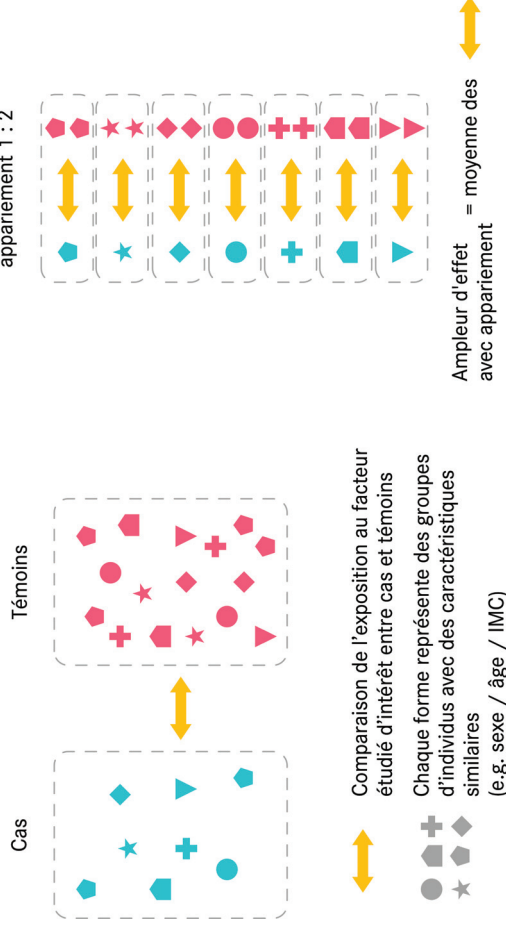
Pour chaque cas (exemple : un malade dans les études cas-témoins, ou un exposé dans les études de cohorte), on associe un ou plusieurs contrôles (témoins dans les approches cas-témoins, ou non exposés dans les approches de cohorte) qui lui sont similaires pour un ou plusieurs facteurs (exemple : âge, sexe, niveau socio-économique, ...).

Lors des analyses, l'effet du facteur étudié n'est pas évalué globalement (en comparant les groupes dans leur globalité), mais au sein des paires (en comparant les cas aux témoins au sein des paires, ou les exposés et non exposés), puis moyenné sur l'ensemble des paires.

Analyse non appariée

Analyse appariée

appariement 1 : 2



Comparaison de l'exposition au facteur étudié d'intérêt entre cas et témoins

Chaque forme représente des groupes d'individus avec des caractéristiques similaires (e.g. sexe / âge / IMC)

Exemple d'analyse appariée dans le cas d'une étude cas témoins

## B

## II Biais

*bias*

Erreur qui fausse systématiquement les résultats dans un sens donné (à la différence d'une erreur aléatoire). On distingue trois grandes familles de biais :  
- les biais de sélection


- les biais de classement / mesure
- les biais de confusion (appelé aussi « phénomène » ou « effet » de confusion).  
La définition des biais varie selon le type d'étude.  
En pratique, il n'est le plus souvent pas possible d'attester avec certitude de la présence d'un biais dans les résultats d'une étude ; en revanche, la LCA vise à identifier les biais potentiels des études et la façon dont les auteurs ont essayé (ou non) de les réduire.

### **Biais différentiel ou non différentiel** *differential bias*

Lorsqu'une étude compare deux groupes, on distingue les biais différentiels des biais non différentiels. Un biais peut s'appliquer de façon similaire (biais non différentiel) ou bien de façon différente aux deux groupes comparés (biais différentiel). Les conséquences d'un biais différentiel ou non différentiel sont en général différentes :

- un biais non différentiel conduit en général à une sous-estimation de l'association entre le facteur étudié et le critère de jugement 
- un biais différentiel fausse l'estimation de l'association entre le facteur étudié et le critère de jugement (sur-estimation ou sous-estimation).

La notion de biais différentiel / non différentiel s'évalue essentiellement dans le cas des biais de classement / mesure .

 Dans une étude cas-témoins comparant des cas de cancer du poumon à des témoins, si tous les participants (cas ou témoins) surestiment leur consommation de fruits (biais du « faking good »), il s'agit d'un biais non différentiel dont les conséquences sont de sous-estimer l'association potentielle entre fruits et cancer. En revanche si seuls les témoins ont tendance à systématiquement surestimer leur consommation de fruits, alors on risque de conclure, à tort, à un effet protecteur des fruits (association significative entre consommation de fruits et cancer).

### **Biais de classement (ou biais de mesure)** *measurement bias; misclassification bias; performance bias*

Mesure ou évaluation incorrecte du critère de jugement  ou du facteur étudié, conduisant à un biais dans les résultats de l'étude. L'évaluation incorrecte peut porter sur :

- Une variable continue : technique de dosage imparfaite, outils de mesure trop imprécis au regard du phénomène mesuré. L'évaluation de la pression artérielle par une seule mesure conduit à un résultat imprécis au vu des fluctuations importantes de PAS ; cette imprécision entraîne une sous-estimation de la relation entre la PAS et le risque d'accident cardiovasculaire (biais dans l'évaluation de l'association). Il s'agit ici d'un biais non différentiel car l'imprécision est la même dans les deux groupes.
- Une variable catégorielle, par exemple sur le statut malade / non malade lorsque l'examen pour établir le diagnostic est imparfait : l'estimation de la prévalence de la maladie sera


faussée.



Le statut répondeur ou non répondeur à une chimiothérapie peut être mal évalué si ce dernier n'est pas standardisé entre les centres ; ceci conduit à une estimation biaisée du bénéfice de la chimiothérapie.

Le terme biais de classement est préféré dans le cas des variables catégorielles, et biais de mesure dans le cas des variables continues.

Les phénomènes de biais de mesure / classement sont fréquents lors d'évaluations subjectives (par exemple évaluation de la qualité de vie par une simple échelle visuelle analogique).



**Le biais de mémorisation** est un exemple de biais de classement.


Il survient lors de l'évaluation d'une exposition ou d'un événement reposant sur l'interrogatoire d'un individu faisant appel à sa mémoire (fréquent dans les études cas-témoins). Il est différentiel  si les cas (malades) ont tendance à minimiser un comportement à risque par rapport aux témoins, ou sens inverse, ont plus souvent tendance à se rappeler de l'exposition à un facteur de risque du fait qu'ils sont malades. Le biais de mémorisation est non différentiel si les erreurs ne sont pas influencées par le statut cas ou témoin (juste un problème de mémorisation).


**Le biais de suivi**  peut être considéré comme un biais de classement / mesure : dans une étude longitudinale, le fait de connaître le traitement / l'exposition peut influencer la fréquence du suivi et donc l'évaluation du critère de jugement (biais différentiel .



Un clinicien sachant que son patient est dans le bras placebo d'un essai pourrait redouter un risque de récurrence plus élevé que s'il avait été dans le bras expérimental ; il sera tenté de réaliser des suivis plus rapprochés de son patient. En absence de différence d'efficacité entre les deux bras, ces suivis plus rapprochés dans le groupe placebo entraîneront en moyenne des temps plus courts jusqu'à la détection de récurrence que dans le bras expérimental.

 Dans un essai clinique, l'aveugle  par rapport au bras de traitement supprime ce biais.

**Biais de subjectivité de l'observateur** : dans une étude cas-témoins, savoir qu'un individu est un cas peut inciter à plus rechercher l'exposition (biais différentiel .

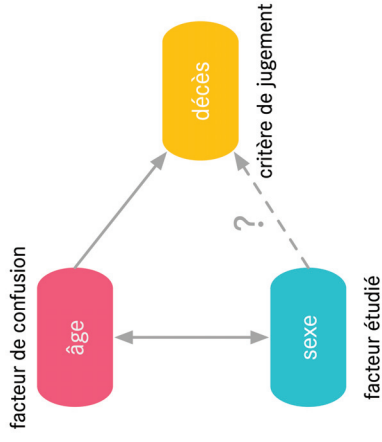
 La standardisation de l'interrogatoire et la mise en aveugle de l'évaluateur (vis-à-vis du statut cas ou témoin) permettent de limiter ce biais.

## ⓘ Biais de confusion *confounding bias*

Biais provoqué par un facteur de confusion qui conduit à une mauvaise évaluation du lien entre le facteur de risque étudié et la maladie (ou le critère de jugement ⚡). Un facteur de confusion est un facteur de risque de la maladie (du critère de jugement) qui est également associé au facteur étudié.



Le risque de décès des suites d'un infarctus du myocarde est lié à l'âge ; par ailleurs, les infarctus du myocarde sont plus tardifs chez les femmes que chez les hommes. Dans cet exemple, l'âge est un facteur de confusion dans l'analyse de l'association entre le sexe et le risque de décès des suites d'un infarctus. La simple comparaison des risques de décès entre hommes et femmes pourrait conduire à conclure à tort à un sur-risque de décès des femmes par rapport aux hommes, qui n'est en fait que le reflet d'un âge plus élevé des femmes que des hommes lors de l'infarctus.



- ➡ Dans les essais cliniques, ce biais est supprimé par la randomisation ⚡ lorsque l'essai clinique est bien conduit.
  - ➡ En dehors des essais cliniques, il peut être réduit :
    - par l'ajustement ⚡ de l'analyse sur le facteur de confusion potentiel
    - par l'appariement ⚡ des cas aux témoins sur le facteur de confusion potentiel (fréquent dans les études cas-témoins)
- Dans les études observationnelles, il n'est pas possible de garantir l'absence de biais de confusion, car tous les facteurs de confusion ne sont pas connus/mesurés.

- Attention à ne pas confondre :
- le phénomène de confusion
  - le phénomène d'interaction ⚡.

## ⓘ Biais de sélection *selection bias*

Survient lorsqu'un résultat obtenu dans l'échantillon/population d'étude est systématiquement différent de celui qui aurait été obtenu dans la population source.



| Étude observationnelle  | Étude expérimentale   |
|---|---|
| <p>Biais dans la constitution de l'échantillon, qui va se retrouver non représentatif de la population source ⚡ pour des facteurs liés au problème étudié (d'où le biais).</p> <p>Différentes raisons peuvent exister :</p> <ul style="list-style-type: none"> <li>- biais dans la constitution de l'échantillon</li> <li>- biais lié aux perdus de vue</li> </ul> <p>Exemples :</p> <ul style="list-style-type: none"> <li>- étude dont les participants sont sélectionnés parmi des personnes en activité ; cette population est en meilleur état de santé, car on omet les personnes en arrêt de travail, hospitalisées, à la retraite, donc moins susceptibles de présenter un critère de jugement ⚡ relié à l'état de santé, ... On parle du biais du travailleur sain</li> <li>- étude de dépistage portant sur des volontaires (qui sont par principe plus enclin au dépistage que la population générale)</li> <li>- étude cas-témoins analysant la relation entre le diabète de type II et le risque de cancer du poumon, recrutant les cas en service de pneumologie, et les témoins en service d'endocrinologie</li> </ul> | <p>Biais qui peut survenir dans le cas où les investigateurs sélectionneraient des patients particuliers dans le groupe traité (se rapproche du biais de confusion des études épidémiologiques).</p> <ul style="list-style-type: none"> <li>➡ Ce biais est supprimé par la randomisation.</li> </ul> <p>Un autre biais de sélection dans les essais cliniques est le biais d'attrition ⚡.</p> |

Le biais de sélection ne désigne pas le manque de représentativité de l'échantillon étudié par rapport à la population cible, qui correspond à un problème de validité externe, même s'il y contribue.





## **Biais d'attrition** *attrition bias*


Dans l'essai clinique, le biais d'attrition survient lorsque le groupe expérimental est favorisé par la « sortie de l'analyse » de certains patients. Les différentes causes possibles sont :

- les changements de bras de traitement, arrêts de traitements, ou les exclusions de certains patients post-randomisation
  - les perdus de vue lorsque ceux-ci surviennent de façon différente selon le bras de traitement ou selon le niveau du critère de jugement 
- Le biais apparaît lorsque ces situations surviennent de façon différentielle entre les bras (biais différentiel ).

 L'analyse en intention de traiter , qui oblige :

- 1°) à garder tous les patients
  - 2°) dans leur groupe de traitement alloué par la randomisation,
  - 3°) en imputant les éventuelles données manquantes
- permet de limiter le biais d'attrition.

 Dans un essai randomisé comparant l'efficacité d'une chimiothérapie combinée à une thérapie ciblée à la chimiothérapie seule, le fait d'arrêter la thérapie ciblée chez les patients ne présentant pas une réponse à 3 mois et d'exclure ces patients de l'analyse risque de surestimer l'effet de la combinaison chimiothérapie/thérapie ciblée sur la survie globale (on exclut petit à petit de ce groupe les patients pour lesquels la thérapie est la moins efficace, favorisant ainsi le groupe expérimental). L'analyse en intention de traiter  conduit à maintenir ces patients dans le groupe expérimental lors de l'analyse (même s'ils n'en n'ont pas totalement bénéficié) ; elle aboutit à des conclusions conservatrices.

 Dans un essai randomisé sur l'insuffisance cardiaque, on mesure l'efficacité d'un nouveau traitement pour augmenter la fraction d'éjection (critère de jugement principal). Le nouveau traitement a comme effet secondaire d'augmenter le décès chez les patients qui avaient la plus faible fraction d'éjection. Le critère de jugement « fraction d'éjection » ne sera alors pas disponible pour les patients décédés (donnée manquante). Sans utiliser des méthodes statistiques appropriées pour gérer ces données manquantes, on risque donc d'avoir une mesure de la fraction d'éjection surestimée dans le groupe traité car les patients avec faible fraction d'éjection sont décédés avant la mesure.


Le biais d'attrition est un type de biais de sélection .

Dans les études observationnelles, le biais d'attrition correspond au biais lié aux perdus de vue. Les analyses en intention de traiter n'ont pas de sens en étude observationnelle, mais on peut utiliser des méthodes de remplacement de données pour pallier aux données manquantes liées aux perdus de vue.


Survient lorsqu'un résultat obtenu dans l'échantillon/population d'étude est systématiquement différent de celui qui aurait été obtenu dans la population source.

## **Biais d'indication** *indication bias*

Le biais d'indication est un exemple de biais de confusion, portant sur le cas spécifique de la comparaison de deux traitements à partir de données observationnelles issues des pratiques cliniques (étude de pharmaco-épidémiologie à partir des données issues de la base de l'assurance maladie par exemple). Du fait de l'absence d'allocation par randomisation du traitement, un phénomène de confusion par indication peut survenir si le médicament d'intérêt est préférentiellement prescrit à des sujets ayant a priori un risque plus faible ou plus élevé de présenter l'événement considéré, l'estimation de l'association entre l'exposition au médicament et la survenue de l'événement d'intérêt étant alors faussée.

 On suppose qu'un médicament A plus récent ou plus efficace, aura plus tendance à être prescrit à des patients plus sévères, de moins bon pronostic. Ainsi, si l'on réalise une étude pour comparer l'efficacité en vie réelle de ce médicament A en le comparant à un médicament B, le bénéfice du médicament A sera en partie occulté (parfois totalement) par le moins bon pronostic des patients qui le reçoivent. De la même façon, on suppose que si un médicament A est susceptible d'avoir un effet indésirable, il sera moins prescrit à des patients à risque d'avoir cet effet indésirable. Dans ce cas, si l'on étudie l'association entre la survenue de cet effet indésirable après exposition au traitement A comparativement à un traitement B, le traitement A semblera donner moins d'effets indésirables que le traitement B, mais simplement car il aurait été prescrit à des patients moins à risque d'effets indésirables.

Ce biais peut être prévenu ou atténué :

- par design : en définissant des critères d'inclusion/non inclusion qui restreignent à une population homogène avec une indication comparable entre les deux traitements, ou en utilisant un médicament contrôlé susceptible d'être administré à tous les patients
- lors de l'analyse statistique, en utilisant des méthodes d'analyses ajustées, ou en ayant recours à des approches de type score de propension ou appariement .

## **Biais de temps immortel** *immortal-time bias*

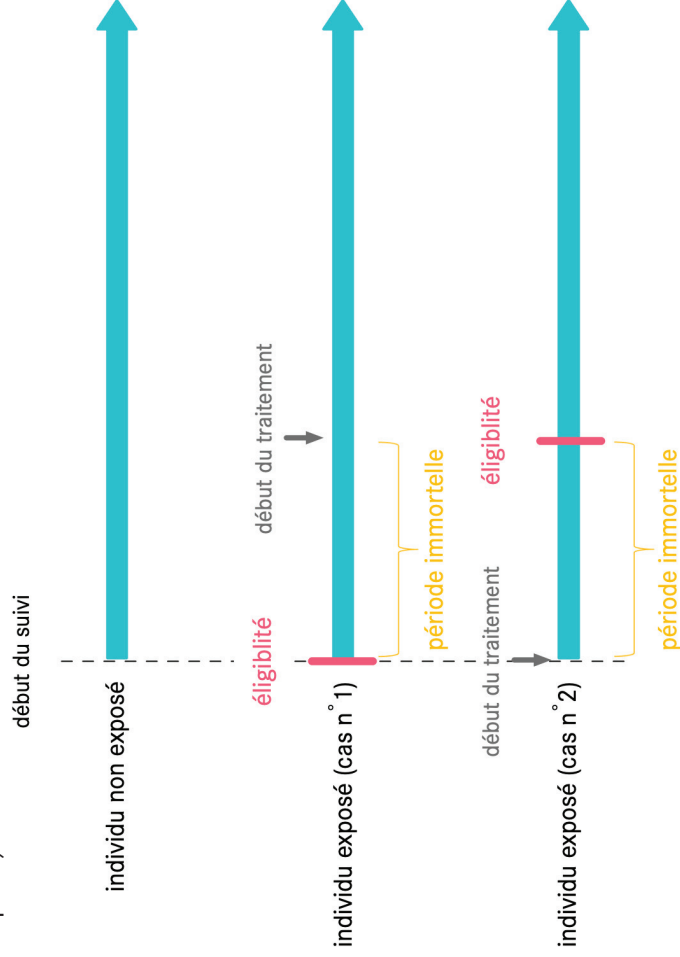
Le biais de temps immortel est un biais particulier (de type biais de sélection ) dans les études de cohorte historique  avec analyse de survie  , lorsque :

- le début de l'exposition au traitement
- ou le début du suivi (mesure des critères de jugement)
- ou le temps où tous les critères d'éligibilité sont vérifiés (par exemple, prise en compte des patients ayant au moins trois dispensations du traitement)

ne coïncident pas.

On peut illustrer ce biais par deux exemples fréquents :

- Cas n°1 : lorsque dans le groupe exposé au traitement à l'étude, un délai sépare le début du suivi (diagnostic) et le début de l'exposition au traitement, les individus exposés doivent survivre (sans événement) à cette période pour être considérés comme exposés. On ne retrouve donc aucun événement pour ces individus pendant cette période (personnes-temps « immortelles »), à la différence des individus non exposés, qui eux peuvent présenter l'évènement dès le début de leur suivi.
- Cas n°2 : lorsque les individus sont considérés comme exposés uniquement s'ils ont reçu plusieurs prescriptions ou dispensations du médicament, la période entre le début du suivi et la date où les individus sont réellement considérés comme exposés (critères d'éligibilité vérifiés) est une période de temps immortel car les individus exposés ne peuvent pas présenter l'évènement pendant cette période (sinon ils seraient considérés comme non exposés).



Dans les deux cas, cette période de temps immortel provoque un avantage de « survie » aux exposés (non survenue de l'évènement par construction) qui fausse l'association entre l'exposition au traitement et la survenue de l'évènement. Il existe différentes stratégies, méthodologiques et statistiques, pour limiter ce biais.

## Biais de contamination

*contamination bias*

Ce biais survient quand les membres d'un des groupes de l'essai reçoivent le traitement ou sont exposés à l'intervention de l'autre groupe. Ce biais entraîne une réduction de la différence observée entre les groupes. C'est un biais de classement. Il est principalement redouté lors des essais en ouvert ou lorsque l'intervention étudiée implique des modifications de pratiques des professionnels ou de l'organisation des soins.

Dans ce cas, la randomisation en clusters peut permettre de réduire le risque de contamination.

Essai évaluant l'efficacité de la réalisation d'entretiens motivationnels (technique spécifique d'entretien) par le médecin généraliste auprès des patients fumeurs, sur l'arrêt de la consommation de tabac, par rapport au suivi habituel (entretiens informationnels). Le médecin va être formé à la technique de l'entretien motivationnel pour l'étude ; de ce fait il va modifier sa façon de réaliser les entretiens avec ses patients. Il y a donc un risque qu'il modifie de manière non intentionnelle ses pratiques d'entretien avec les patients du groupe contrôle, et également un risque qu'il les modifie de manière intentionnelle, si en cours d'étude, il perçoit comme efficace l'entretien motivationnel. La randomisation en cluster (chaque médecin généraliste est randomisé et réalisera soit uniquement des entretiens motivationnels soit les entretiens classiques avec leur patients fumeurs) permet de réduire le risque de contamination entre les 2 groupes.


Essai comparant l'utilisation de nouvelles brochures éducatives pour les patients atteints d'insuffisance cardiaque sur l'adhésion aux règles hygiéno-diététiques et la survenue d'épisodes de décompensation d'insuffisance cardiaque : si les patients ont des interactions entre eux dans des groupes d'éducation thérapeutique du service, ils peuvent s'échanger des documents. La randomisation en cluster (tous les patients d'un même service sont dans le même groupe intervention ou contrôle) permet d'éviter ce biais en réduisant les échanges entre les patients des 2 groupes.

## Biais de spectre

*spectrum bias*


Dans une étude de performances diagnostiques, il s'agit de la situation où la diversité de la présentation de la maladie n'est pas représentée dans la population source, par exemple :


- lorsque tous les stades de la maladie ne sont pas représentés au sein de la population d'étude. Le plus souvent, les stades précoces (par exemple lorsqu'ils sont asymptomatiques) qui sont sous représentés. Cela conduit en général à surestimer la sensibilité, ce qui fausse l'estimation de la performance.
- lorsque certains diagnostics différentiels sont mal représentés : dans l'insuffisance cardiaque, le fait d'exclure les coronaropathies de la population non malade peut fausser l'estimation des performances du pro-BNP (car les coronaropathies ont une symptomatologie proche de l'insuffisance cardiaque)


De plus, le profil clinique des sujets de l'étude étant différent de celui de la population cible , l'extrapolation des résultats est limitée. Le biais de spectre n'est pas à proprement parler un biais, car il fait référence à un problème de validité externe.

### **Biais d'incorporation**

*Incorporation bias*

On parle de biais d'incorporation lorsque le nouveau test diagnostic évalué est une partie du test de référence. Le résultat du nouveau test est donc pris en compte pour déterminer le statut réel des sujets ce qui entraîne une surestimation des performances du nouveau test. Le biais d'incorporation est un biais de classement/mesure .


 Évaluation de la performance de différents biomarqueurs pour la détection d'une bactériémie chez le nourrisson fébrile avec utilisation de l'hémoculture comme test de référence. Si les cliniciens sont libres dans leurs investigations cliniques initiales aboutissant à la réalisation des prélèvements bactériologiques, la stratégie de référence peut donc déjà inclure en amont le dosage de différents biomarqueurs dans le sang (par exemple, la CRP, qui est réalisée en pratique). Ainsi si la CRP fait partie des biomarqueurs évalués (test évalué) alors qu'elle pouvait être incluse dans la stratégie de référence (possibilité de réaliser des hémocultures selon le taux de CRP observé), sa performance « propre » sera surestimée (biais d'incorporation) car la discrimination des « vrais » malades et des vrais « non malades » repose en partie sur ce biomarqueur.


NB : cet exemple, montre également un risque de biais de vérification , puisque les cliniciens étaient libres de réaliser les investigations qu'ils souhaitaient pour bien poser leur diagnostic avant de recourir aux hémocultures (test de référence pour le diagnostic final), l'hémoculture étant un test très invasif pour le nourrisson, les nourrissons avec un test évalué positif (avec une CRP+) ont plus de chance d'avoir une hémoculture que les nourrissons négatifs (CRP-).

On distingue donc le biais d'incorporation (utilisation du test évalué dans la stratégie de référence pour discriminer les vrais malades des vrais non-malades) et le biais de vérification (utilisation du test évalué pour décider ou non de réaliser le test de référence invasif).

### **Biais de vérification**

*Verification bias, workup bias*

Biais pouvant survenir dans des études de performances diagnostiques, lorsque le test de référence n'est pas réalisé pour l'ensemble des patients de l'étude. Ceci peut être le cas lorsque le test de référence est invasif ou coûteux. Ces patients sont soit exclus de l'étude, soit considérés comme non malades (alors que l'on ne dispose pas du test de référence), ce qui peut fausser substantiellement l'estimation de la validité du test étudié (surestimation de la sensibilité et sous-estimation de la spécificité). Le biais de vérification est un biais de classement .

 Dans le cas de l'évaluation de la performance de la procalcitoninémie (PCT) pour la détection d'une bactériémie chez le nourrisson fébrile avec l'utilisation de l'hémoculture comme test de référence : l'hémoculture est un test très invasif pour un nourrisson (prélèvement sanguin conséquent). Si les cliniciens sont libres de réaliser les investigations qu'ils souhaitaient pour bien poser leur diagnostic avant de recourir à l'hémoculture, et donc que les nourrissons avec une PCT+ ont plus de chance d'avoir une hémoculture que les nourrissons négatifs (PCT-), il existe alors un risque de biais de vérification. L'exclusion des nourrissons PCT- n'ayant pas eu d'hémoculture entraîne artificiellement une diminution de la proportion de faux négatifs (PCT-, hémoc+) et une augmentation de la proportion de faux positifs (PCT+, hémoc-), ce qui entraîne une élévation de la sensibilité et une diminution de la spécificité du test.

 Ce biais ne survient pas si tous les sujets admissibles ont bénéficié du test de référence, et ce, de façon consécutive, sans égard au résultat du test étudié.

### **Biais lié au délai entre les tests dans les études diagnostiques**

*disease progression bias*

On suspecte un biais lorsque le test évalué et le test de référence ne sont pas réalisés à des temps suffisamment proches. S'il existe un délai trop long entre la réalisation des deux tests, la maladie peut avoir progressé ou les traitements ont pu agir, ce qui empêche la comparaison des résultats.



Dans le cas de l'évaluation de la performance de la procalcitoninémie (PCT) pour la détection d'une bactériémie chez le nourrisson fébrile avec l'utilisation de l'hémoculture comme test de référence : après un stimulus septique, le taux de procalcitonine dans le sang augmente continuellement pour atteindre un pic dans les 24 h. Un prélèvement de la PCT retardé ne serait-ce que d'une heure après la prise en charge du nourrisson augmente sa spécificité du fait du moins grand nombre de faux positifs (selon le seuil de détection choisi) dans ce cas-là.

Ce biais est un biais de classement . Il ne porte pas forcément de nom spécifique, on retrouve parfois les termes de biais d'ordre séquentiel, ou biais d'évolutivité.

## C

### ▮ Cas incident / Cas prévalent

*incident case / prevalent case*

Ces termes sont particulièrement utilisés dans les études cas-témoins <sup>Ⓢ</sup>.

Le terme de cas incident est utilisé lorsqu'un patient est inclus dans une étude au moment où il développe l'événement de santé étudié. Par exemple, une étude incluant des cas incidents évaluant les facteurs de risque de diabète n'inclurait que les patients nouvellement diagnostiqués comme diabétiques à partir d'une date t correspondant au début de l'étude.

Le terme de cas prévalent est utilisé lorsqu'un patient inclus dans une étude est déjà malade au moment de son inclusion dans l'étude. Par exemple, une étude incluant des cas prévalents évaluant les facteurs de risque de diabète pourraient inclure tous les patients diabétiques, quelle que soit la durée qui s'est écoulée depuis le diagnostic de la pathologie. Les études utilisant des cas prévalents présentent généralement plus de biais que les études utilisant des cas prévalents.

Le tableau suivant présente quelques avantages et les inconvénients de l'inclusion de cas incidents ou de cas prévalents :

|                                       | Cas prévalents  | Cas incidents   |
|---------------------------------------|---|---|
| Durée nécessaire à l'inclusion        | Rapide : Cas disponibles à un instant t (patients déjà malades)   | Longue : nécessite une certaine durée de recrutement car les cas doivent être nouvellement diagnostiqués (durée d'apparition des cas) |
| Biais de sélection (survie sélective) | Possible : les cas prévalents ont survécu depuis leur diagnostic et peuvent donc présenter des caractéristiques et un pronostic spécifiques | Peu vraisemblable car tous les cas sont inclus au moment du diagnostic  |

### ▮ Causalité

*causality*

Rapport établi entre une cause et un effet, un facteur de risque <sup>Ⓢ</sup> et une maladie. Le facteur étudié est responsable (au moins en partie) de la maladie étudiée.

Association ne veut pas dire causalité. Dans les études observationnelles, certains éléments peuvent appuyer la notion de causalité.

Hill a proposé des critères pour évaluer la causalité :

- Critères internes à l'étude
  - o Force de l'association (par exemple, risque relatif élevé)
  - o Relation dose-effet entre exposition et maladie

- o Spécificité de la relation entre exposition et maladie (exposition spécifique de la maladie étudiée)
- o Chronologie : le facteur de risque précède la survenue de la maladie
- Critères externes à l'étude (bibliographie)
  - o Concordance entre les résultats des études
  - o Plausibilité biologique (mécanismes d'actions connus)
  - o Concordance avec les expérimentations in vitro ou chez l'animal
  - o Baisse de l'incidence de maladie si l'exposition est supprimée ou réduite

Il a ainsi mis en évidence, par ces critères, que les tests statistiques ne sont pas suffisants, et a reconnu l'importance de multiples autres facteurs dans la prise de décision (faisceau d'arguments).

### ▮ Comité d'éthique - CPP (Comité de protection des personnes)

*Institutional Review Board, IRB*

Groupe d'experts composé de médecins, de juristes, de philosophes, etc. Il veille au respect des questions d'éthique dans les protocoles de recherche clinique <sup>Ⓢ</sup>. En France, il existe deux types de comité : les comités de protection des personnes (CPP) relatifs à la recherche sur la personne humaine, et les comités d'éthique relatifs à la recherche sur données.

### ▮ Comité de surveillance et de suivi:

*data safety monitoring board (DSMB) ou data monitoring comitee (DMC)*

Dans le cadre des essais interventionnels, un comité de surveillance et de suivi peut être mis en place. Ce dernier examine à un rythme défini la sécurité, l'intégrité scientifique de l'essai en cours et, dans certains cas, l'efficacité (missions définies avant le démarrage de l'étude). Le comité a accès à l'ensemble des données nécessaires pour ces évaluations, par bras de traitement. Il est indépendant de l'étude (c'est-à-dire non impliqué dans l'inclusion ou le suivi des patients), souvent même indépendant de la structure menant l'étude. Il est composé de cliniciens, méthodologistes, et selon les études, d'experts particuliers comme des pharmacologues.



A la suite des réunions, le comité va produire des recommandations au promoteur pour la poursuite ou non de l'essai. C'est entre autre ce comité qui recommande l'arrêt pour efficacité ou futilité lors des analyses intermédiaires <sup>Ⓢ</sup>.

Les investigateurs n'ont accès qu'aux conclusions du comité, mais pas aux résultats qui lui sont communiqués ; ceci leur évite d'avoir accès à des informations qui pourraient modifier le déroulement de l'étude et introduire des biais.




## ▮▮ **Clause d'ambivalence (ou principe d'incertitude)**

La clause d'ambivalence existe lorsque l'état des connaissances ne permet pas de faire le choix entre plusieurs modalités thérapeutiques. Elle est un pré-requis à la réalisation d'un essai clinique comparant ces modalités thérapeutiques.

## ▮▮ **Cohorte de sujets**

Groupe de personnes suivies dans le temps de manière prospective , ayant un point commun (exemples : lieu de naissance pour l'étude Framingham, pathologie ou traitement reçu, ...). A ne pas confondre avec les études de cohorte , qui correspondent à une méthodologie d'étude particulière.

## ▮▮ **Comparabilité des groupes**

La comparabilité des groupes est obtenue lorsque la distribution des facteurs risques/pronostiques  est similaire entre les deux groupes. Contrairement aux idées reçues, la randomisation  ne conduit pas à des groupes totalement identiques, il peut y avoir des fluctuations entre les deux groupes obtenus. En revanche, elle empêche la survenue de différences systématiques entre les groupes, qui conduiraient à un phénomène de confusion .

## ▮▮ **Consentement éclairé**

*inform consent*

En cas de proposition à un patient de participer à une étude clinique soumise à la loi, il est nécessaire, pour procéder à son inclusion, d'obtenir son consentement éclairé. Celui-ci se présente sous forme d'un document écrit spécifiant les risques encourus par le patient, et dans lequel il est impérativement mentionné qu'il a le droit d'arrêter à tout moment de participer à l'étude, sans conséquence pour la poursuite des soins. Le médecin doit expliquer clairement et simplement les avantages et inconvénients de la participation à l'étude.

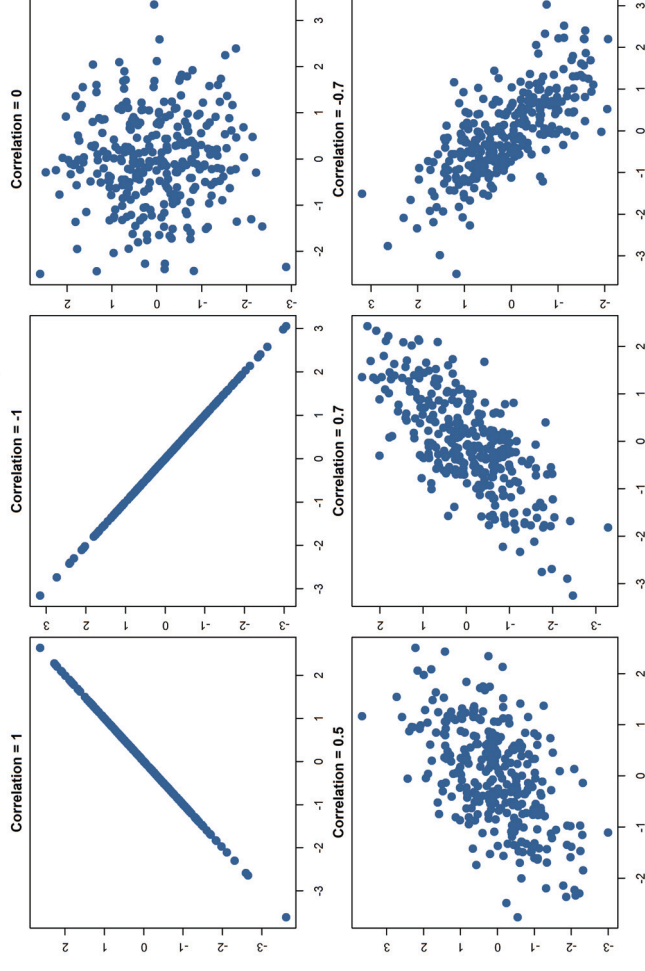
NB : les modalités de recueil du consentement varient selon le type d'étude et la réglementation relative à la recherche en santé en vigueur dans les pays concernés.


## ▮▮ **Corrélation**

*correlation*

Le coefficient de corrélation est une mesure qui quantifie la force de la relation linéaire entre deux variables quantitatives. Il est compris entre -1 et 1. Pour deux variables quantitatives sans aucun lien, le coefficient de corrélation est égal à 0. Le coefficient de corrélation augmente lorsque les deux variables évoluent conjointement de façon positive (quand X augmente, Y augmente) ou diminue lorsqu'elles évoluent conjointement de façon négative (quand X augmente, Y diminue). Il est souvent testé par rapport à la valeur 0 (absence

d'association) ou en regardant si l'intervalle de confiance contient la valeur 0 ; il est à interpréter en fonction de l'intervalle de confiance associé : un coefficient de corrélation élevé est peu crédible si l'intervalle de confiance est large.



 **Mot polysémique** : le terme corrélation est souvent employé pour signifier une association entre deux variables (quelle que soit la nature de ces variables).

## ▮▮ **Critère de jugement**

*endpoint*

Critère qui permet de mesurer l'effet du traitement dans un essai thérapeutique ou la survenue d'un événement dans une étude observationnelle.

## ▮▮ **Critère de jugement clinique**


*clinical endpoint*

Un critère de jugement clinique est un critère correspondant aux objectifs thérapeutiques du traitement (et qui est percevable par le patient). Il s'agit de la fréquence ou du délai de survenue de l'événement, de la fréquence ou du délai de survenue de récurrence, ou de complications en lien avec la maladie (par exemple pour les pathologies liées à l'athérosclérose : infarctus du myocarde, AVC, artériopathie des membres inférieurs).

Dans certains cas, l'objectif du traitement est d'améliorer la qualité de vie des patients (activité physique, fonction psychologique, vie sociale, signes fonctionnels somatiques) ; les questionnaires de qualité de vie validés peuvent constituer des critères cliniques.

### Critère de jugement intermédiaire

*intermediate endpoint*

Les critères de jugement intermédiaires ne correspondent pas directement à un objectif thérapeutique, et documentent plutôt les mécanismes d'action du traitement : il s'agit le plus souvent de paramètres biologiques ou physiologiques. Une étude s'appuyant sur un critère intermédiaire n'est pas une preuve directe d'efficacité (niveau de preuve  plus faible qu'un critère clinique).

Exemples de correspondances entre critères intermédiaires et critères cliniques pour différentes pathologies :

| Pathologie                     | Critère intermédiaire                | Critère clinique  |
|--------------------------------|--------------------------------------|---|
| Hypertension                   | Pression artérielle                  | Événements coronariens, AVC   |
| Ostéoporose                    | Densité osseuse                      | Fracture du col du fémur  |
| Infarctus du myocarde          | Taux de reperméabilisation coronaire | Décès, événements coronariens                                       |
| Diabète non insulino dépendant | Glycémie, hémoglobine glyquée        | Événements cardiovasculaires, rétinopathie, insuffisance rénale ... |

### Critère de substitution


*surrogate endpoint*

Un critère de substitution est un critère intermédiaire  dont la mesure permet une prédiction précoce et précise

- du critère clinique

- ET de l'effet du traitement sur ce critère.

Le critère de substitution doit donc être corrélé au critère clinique ; de plus, tout effet du traitement détecté par le critère de substitution doit correspondre à un effet sur le critère clinique, et vice-versa.

 En cancérologie, la survie sans progression dans les cancers colorectaux avancés traités avec une chimiothérapie à base de fluorouracil est un critère de substitution de la survie globale. Attention : la survie sans progression n'est pas un critère de substitution pour tous les cancers.

### Critères de jugement principal et secondaire


*primary endpoints, secondary endpoints*

Dans un essai clinique les critères de jugements sont regroupés en deux classes, idéalement selon l'importance clinique : principaux et secondaires. Les critères de jugements principaux sont ceux destinés à soutenir l'autorisation pour une mise sur le marché. C'est également sur les principaux que l'on souhaite démontrer un effet et dont le risque alpha doit être contrôlé.


Les critères secondaires peuvent être utilisés pour étayer le(s) critère(s) de jugement principal(principaux), ils sont également inclus pour explorer des hypothèses physiopathologiques. La démonstration d'une efficacité n'est jamais faite formellement sur l'ensemble de critères de jugement secondaires. La définition des ensembles des critères de jugement principaux et secondaires doit être fait a priori.

### Critère de jugement principal multiple

*multiple endpoints*


Quand il existe plus d'un critère de jugement principal qui permettrait de déclarer un essai clinique positif, alors les critères de jugement sont dits multiples. Dans le cas de critères de jugement multiples, l'inflation du risque alpha  doit être prise en compte, soit par nature (co-primary endpoint) soit par des méthodes statistiques (Bonferroni, hiérarchie, répartition ).

### Co-primary endpoint

Les études avec co-primary endpoints sont des études avec critères de jugement principaux multiples  pour lesquelles le bénéfice du traitement est prouvé lorsque le bénéfice est attesté simultanément sur tous les critères de jugement principaux. C'est-à-dire qu'il suffit d'un critère de jugement non significatif pour déclarer l'étude négative sur l'ensemble des critères de jugement principaux. En procédant ainsi, le risque alpha n'a pas besoin d'être réparti entre les différents critères de jugement : la p-value doit être inférieure à 0.05 pour chaque critère afin de conclure positivement pour l'étude. Pour le nombre de sujets nécessaire, à minima, le calcul d'effectif est réalisé pour chaque critère de jugement, et l'effectif maximum est retenu.



Pour un essai concernant la maladie d'Alzheimer, un co-critère de jugement primaire peut associer un critère relatif sur une échelle cognitive et un critère relatif à l'impact clinique sur le patient (par exemple l'échelle ADL : activités de la vie quotidienne) qui sont tous deux importants dans la pathologie.

 Terme polysémique : on peut retrouver le terme co-primary pour des études évaluant deux critères de jugement principaux en utilisant une répartition du risque alpha ou encore une hiérarchie pour contrôler l'inflation du risque alpha.

### Critère de jugement composite

*composite endpoint*

Un critère de jugement composite regroupe plusieurs événements cliniques (appelés composants) pour ne former qu'un seul critère de jugement. Un individu est compté une seule fois comme ayant présenté le critère de jugement composite même s'il présente plusieurs composants. Un seul test statistique est réalisé sur le composite en entier, et la

conclusion se fait sur l'ensemble du critère composite et non ses composants ; de ce fait, un seul test est réalisé, il n'y a donc pas nécessité de correction du risque alpha pour les tests multiples.

Les événements cliniques combinés doivent être du même niveau de gravité afin que le critère composite soit pertinent. Les critères composites sont utiles pour les études pour lesquelles le nombre de sujets nécessaires en prenant chacun des composants séparés serait trop élevé. Par ailleurs, le fait que l'étude soit positive sur le critère composite ne signifie pas nécessairement qu'il y a un bénéfice sur chacun des composants. A ce titre, il est recommandé que l'étude inclue comme critères de jugement secondaires chaque événement du critère composite de manière isolée.



Un critère composite est défini par l'observation d'un décès cardiovasculaire ou d'une hospitalisation pour aggravation de l'insuffisance cardiaque. Un individu qui va être hospitalisé pour une aggravation de l'insuffisance cardiaque puis va décéder ne va compter comme ayant réalisé le composite qu'une seule fois.

## II Critère d'inclusion / non-inclusion / éligibilité (ou critère d'exclusion)

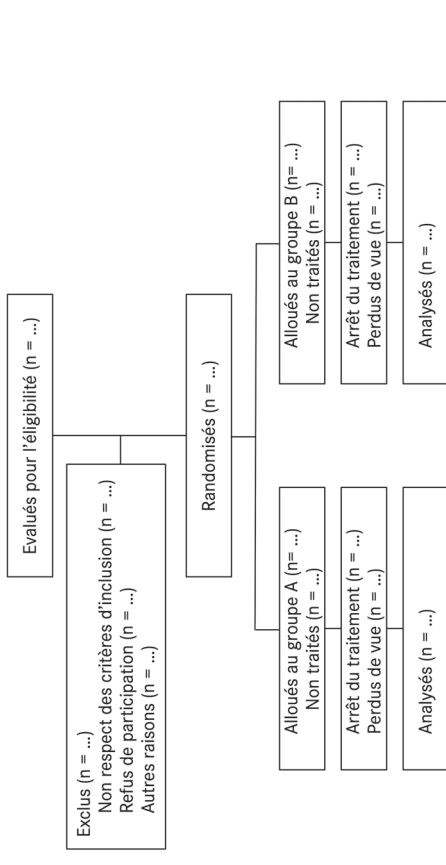
*inclusion criterion / exclusion criterion*

Ensemble de critères qui définissent de façon précise les caractéristiques des patients qui peuvent participer à une étude. Ces critères définissent la population source ☹.

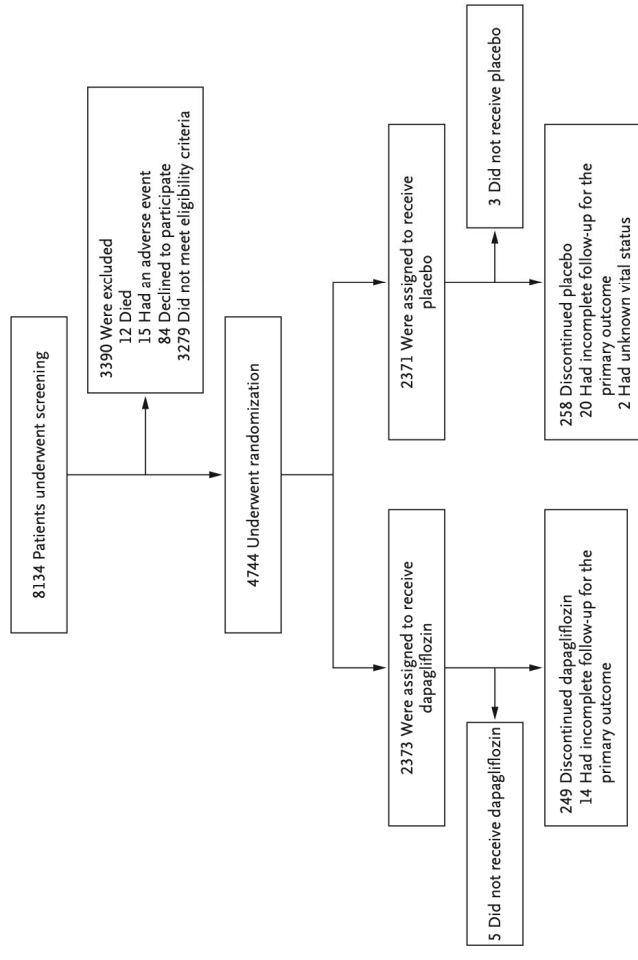
- critère d'inclusion : définit la population souhaitée pour l'étude
- critère de non-inclusion (appelé parfois critère d'exclusion) : exclut des patients pour lesquels l'étude est déconseillée

Dans les essais cliniques, on ne doit pas exclure de patients une fois inclus dans l'étude pour éviter le biais d'attrition ☹ : le traitement peut être arrêté, mais le suivi doit être assuré afin de permettre l'analyse en intention de traiter ☹ ; en revanche, un patient peut être exclu de l'analyse per protocole ☹.

La notion d'éligibilité apparaît normalement en amont des notions d'inclusion/non inclusion, au niveau du screening des patients potentiellement inclusibles dans l'étude. Elle définit la population source.



Exemple d'essai clinique sur l'évaluation de la Dapagliflozin chez les patients avec un arrêt cardiaque et faible fraction d'éjection (McMurray et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. N Engl J Med. 2019 Nov 21;381(21):1995-2008).



Les 4744 patients randomisés constituent un échantillon de la population source, et correspondent à la population analysée pour l'analyse en intention de traiter. Les patients qui n'ont pas reçu le traitement alloué par la randomisation, ayant arrêté le traitement, ou n'ayant pas un suivi permettant l'évaluation du critère de jugement principal seront exclus de la population per protocole.

épidémiologie. Pour les essais thérapeutiques, cette analyse est effectuée à partir de bras à différentes doses (au moins trois doses et un groupe placebo).

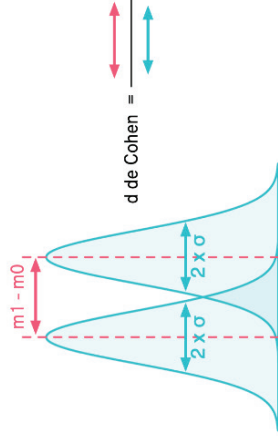
### **D** Différence moyenne standardisée

*standardized mean difference*

Les différences moyennes standardisées sont des mesures de l'écart de moyennes entre deux groupes (par exemple entre deux traitements) ramené à la variabilité au sein de ces groupes. Un exemple est le *d* de Cohen. Il correspond à la différence de moyennes divisée par l'écart-type des valeurs au sein des deux groupes :

$$\frac{m_1 - m_0}{\sigma}$$

Ce dernier mesure l'importance de l'écart entre les deux groupes au vu de l'hétérogénéité au sein des groupes.



Des abaques sont données pour aider à l'interprétation :

| d de Cohen | Ampleur d'effet |
|------------|-----------------|
| 0          | Null            |
| 0,2        | Faible          |
| 0,5        | Modérée         |
| 0,8        | Élevée          |

Le *d* de Cohen peut se généraliser à des critères de jugement binaires.

### **E**

### **E** Evidence Based Medicine

*médecine basée sur les preuves*

Selon Sackett, la médecine basée sur les preuves consiste à utiliser de manière rigoureuse, explicite et judicieuse les preuves actuelles les plus pertinentes lors de la prise de décision concernant les soins à prodiguer à chaque patient (Sackett et al; BMJ 1996).

### **C** Cross-Over

voir partie Essai clinique : designs

### **C** Carry-over (effet)

Dans les essais en *cross-over* : poursuite de l'intervention A (médicament, changement de pratique, ...) durant la période associée à l'intervention B et vice-versa. Pour un médicament, l'effet *carry-over* est dépendant de la demi-vie du médicament : lorsqu'un patient prend un médicament, et qu'il n'interrompt pas assez tôt la prise du médicament, l'effet de celui-ci peut se poursuivre. La réalisation d'une période de *wash-out* suffisante (sans prise de médicament) entre les deux périodes de traitement permet de supprimer cet effet.

### **D**

### **d** de Cohen

voir partie différence moyenne standardisée

### **D** Degré de significativité (valeur de *p*, significativité statistique)

voir partie Tests statistiques

### **D** Données agrégées

*aggregated data*

Les données agrégées sont des données résumant les données des individus d'un groupe considéré à un niveau collectif (sous forme de moyenne, de médiane, de proportion, de données cumulées, etc.).



La proportion de fumeurs dans une population donnée, le revenu médian des foyers dans une population donnée, le volume total des ventes de cigarettes dans une population donnée, etc.

### **D** Dose-effet (relation dose-effet)

*dose-response relation*

Il existe une relation dose-effet, en épidémiologie, lorsque le risque lié à un facteur varie en fonction de l'intensité de l'exposition (par exemple : plus l'exposition est grande, plus l'incidence est élevée). Pour ce qui concerne les médicaments, il existe une relation dose-effet lorsque l'effet du médicament varie en fonction de la dose reçue. On peut étudier cette relation dose-effet par des analyses spécifiques telles que les tests de tendance en



Contrairement à ce que l'on peut penser ou entendre, l'EBM n'est pas uniquement constituée par les preuves issues de la recherche scientifique ; elle est fondée sur 3 piliers :

1. Les préférences du patient et de son entourage entre différents traitements dépendent des croyances, attentes et objectifs de santé et de vie des personnes concernées. Le praticien va recueillir les préférences du patient et/ou de son entourage concernant les différentes options de traitement possibles après leur avoir expliqué les potentiels bénéfices, inconvénients, effets secondaires, et coûts de chaque option.

2. L'expérience clinique du médecin qui doit se fonder sur une analyse systématique des observations cliniques, de façon reproductible, en limitant les biais, et en évitant toute interprétation intuitive de l'information.

3. Et enfin, les données de la recherche. En effet, le clinicien doit consulter la littérature scientifique pour résoudre les problèmes cliniques et proposer une prise en charge optimale au patient.

Ces 3 piliers permettant la décision dans le soin.

## Écart au protocole (déviation au protocole)

*protocol deviations*

Ensemble de situations où le protocole ☹️ d'une étude n'a pas été suivi scrupuleusement. Il faut le mentionner lors de la restitution des résultats et dire comment cela a été pris en compte dans l'analyse, en expliquant quelle influence peuvent avoir eu ces écarts sur les résultats observés.

L'analyse en intention de traiter ☹️ inclut tous les patients même en cas de déviation au protocole, alors que l'analyse per-protocole ☹️ exclut de l'analyse les patients présentant des déviations.

## Effet nocebo

Effet nocif de la prise d'un médicament qui peut être dû à un effet physio-psychologique, par exemple lié au vécu du patient.

Dans une étude comparant une chimiothérapie à un placebo pour un cancer rare, une perte de cheveux a été observée chez certains patients du groupe placebo.

## Effet placebo

Effet bénéfique de la prise d'un médicament qui peut être dû à un effet physio-psychologique, par exemple lié au vécu du patient. L'effet placebo d'un traitement est contrôlé grâce à la comparaison à un placebo ou à un autre traitement actif.

## Le placebo

Le placebo est un traitement sans produit pharmacologiquement actif. Le placebo peut produire des effets thérapeutiques ou indésirables. Dans les essais cliniques, il contient les mêmes excipients, a la même forme galénique, le même goût (si pris per os) et la même couleur que le produit actif pour permettre de maintenir l'insu et d'estimer l'effet lié au produit pharmacologiquement actif.

## Double placebo

Dans un essai thérapeutique au cours duquel deux médicaments sont comparés, il est idéal qu'ils soient identiques sur le plan galénique et sur celui du mode d'administration (posologie) pour maintenir l'insu ☹️. Quand ce n'est pas possible (exemple : un traitement injectable et un traitement per-os), il faut prévoir que les patients recevant le traitement A prennent aussi un placebo de B ; et que les patients recevant le traitement B prennent aussi un placebo de A.

## Émulation d'essai cible (ou d'essai thérapeutique)

*target trial emulation*

L'émulation d'essai cible est une méthode générale qui permet de répondre à une question causale (ex : effet d'un traitement sur un critère de jugement) à partir de données observationnelles en limitant les biais.

L'émulation d'essai cible est un processus en deux grandes étapes : (1) définir le protocole d'un essai randomisé idéal, appelé essai cible ; (2) émuler explicitement le protocole de cet essai cible à partir des données observationnelles

Les catégories du protocole à définir sont les suivantes :

- Critères d'éligibilité
- Stratégies thérapeutiques
- Assignment des stratégies thérapeutiques
- Temps zéro et suivi
- Critères de jugement
- Contraste causal

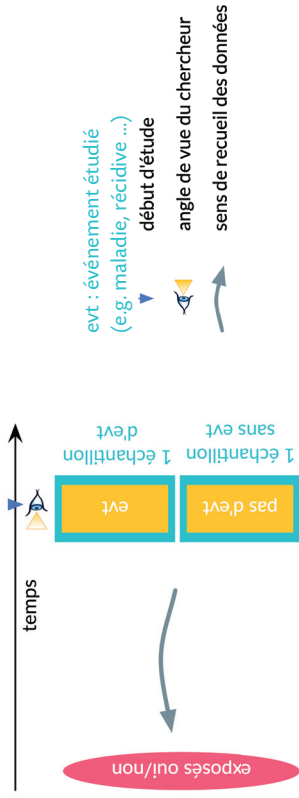
Pour chaque partie, des approches sont définies pour se rapprocher le plus possible d'un essai randomisé. Certains aspects ne peuvent jamais être reproduits à l'identique dans un essai cible, par exemple l'aveugle ou la notion de placebo. La notion de date d'inclusion (temps zéro) peut aussi être difficile à définir dans certains cas, risquant d'entraîner un biais de temps immortel ☹️.

La partie contraste causale revient à limiter les biais de sélection ☹️ dans la comparaison entre les deux groupes en l'absence de randomisation ☹️. Des approches existent, notamment en ayant recours à un score de propension ☹️.

## Étude : design

### Étude Cas-Témoin case-control study

Enquête dans laquelle on interroge comparativement des malades (cas) et des non malades (témoins) sur leurs expositions dans le passé à des facteurs de risque. L'inclusion **des sujets est effectuée selon leur statut** (malade ou non malade). L'échantillon n'est pas représentatif de la population source en termes de prévalence de la maladie, puisque l'on fixe arbitrairement le nombre de malades et de non-malades à inclure dans l'étude.

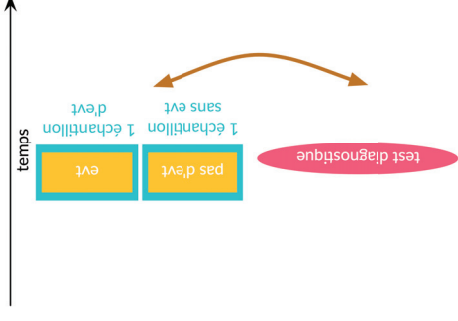


Les biais importants dans les études cas-témoins sont les biais de classement (en particulier de mémorisation) et de sélection.

Plusieurs méthodes existent pour sélectionner les cas dans l'étude : méthode des cas incidents, et méthode des cas prévalents, avec leurs avantages et limites respectifs.

Par ailleurs, ce type d'étude ne conduit pas à un échantillon de patients dans lequel la prévalence est représentative de celle dans la population source, c'est pourquoi les liens entre expositions et maladie (la mesure du risque de survenue de la maladie lié à l'exposition) sont résumés par des odds ratios (et leurs intervalles de confiance) et non par des risques relatifs.

Cas particulier des études diagnostiques : certaines évaluations de méthodes diagnostiques (phase I) sont effectuées selon un design cas-témoins : l'étude inclut un certain nombre de malades et de non-malades (avec une surreprésentation des malades par rapport à la population source) pour lesquels on étudie un biomarqueur ou un test. De ce fait, elles sont qualifiées d'études diagnostiques de type cas-témoins, même si l'étude est transversale en termes de temporalité.



### Étude Cas-Témoins nichée nested case-control study

Réalisation d'une étude cas-témoins pour laquelle les cas et les témoins sont identifiés à partir d'une cohorte précédemment constituée.

La cohorte permet d'identifier des cas incidents qui vont constituer les cas de l'étude, les témoins vont être identifiés à partir d'un sous-ensemble de non cas sélectionnés également dans la cohorte. Les témoins peuvent être appariés aux cas selon des critères d'appariement (par exemple, âge, sexe ...); dans certains cas, plusieurs témoins peuvent être appariés à un même cas pour augmenter la puissance.

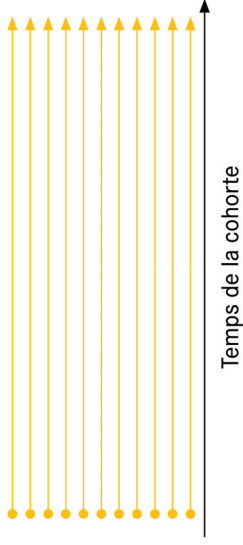
Il existe avec les études cas-témoins nichées moins de biais dans la sélection des témoins car ils appartiennent à la même cohorte initiale que les cas. De plus le biais de mémorisation est limité par rapport aux études cas-témoins classiques, car les données sont recueillies dans la cohorte de manière prospective, planifiée dans le cadre du protocole. Enfin, ce type d'étude est également plus rapide à réaliser qu'une étude cas-témoins classique, car les cas et les témoins sont déjà identifiés dans la cohorte préexistante.

Par contre les données étant déjà recueillies dans le cadre de l'étude de cohorte (ou à partir d'une base de données existante) dont les objectifs n'étaient pas ceux de l'étude cas-témoins, il est possible que celles-ci ne correspondent pas totalement aux données nécessaires.

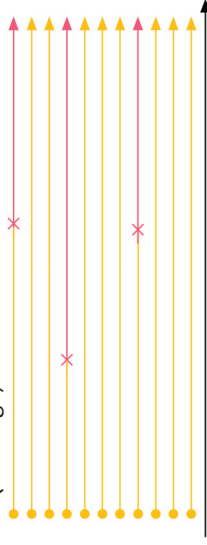
Dans une cohorte de sujets affiliés à une assurance maladie, une étude cas-témoins nichée peut être réalisée concernant l'association entre l'exposition aux benzodiazépines et le risque de survenue de maladie d'Alzheimer.

Le processus de constitution de l'étude cas-témoins nichée se fait en plusieurs étapes :

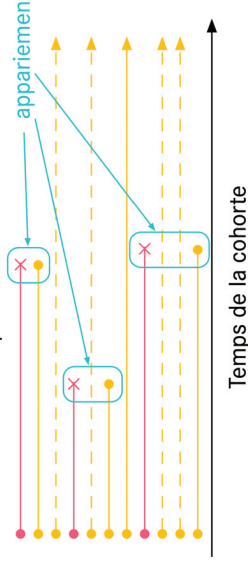
1. Identification de la cohorte de sujets âgés indemnes de maladie d'Alzheimer en début de suivi



2. Identification des cas (en rouge) = survenue de la maladie d'Alzheimer dans la cohorte



3. Identification des témoins sans maladie d'Alzheimer. Une fois les cas et les témoins sélectionnés (et éventuellement appariés), on étudie l'association entre la pathologie et l'exposition antérieure aux benzodiazépines.



### Étude de cohorte

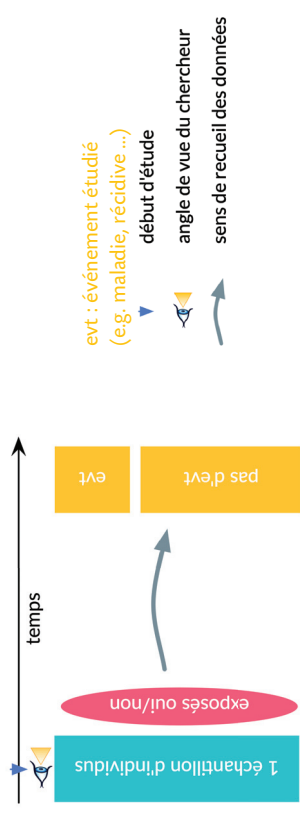
*Cohort study*

Enquête au cours de laquelle des sujets sont suivis au cours du temps afin d'analyser l'association entre des facteurs dont on a mesuré l'exposition et la survenue d'un événement de santé. L'inclusion des patients est indépendante de leur statut vis-à-vis de l'événement de santé étudié.

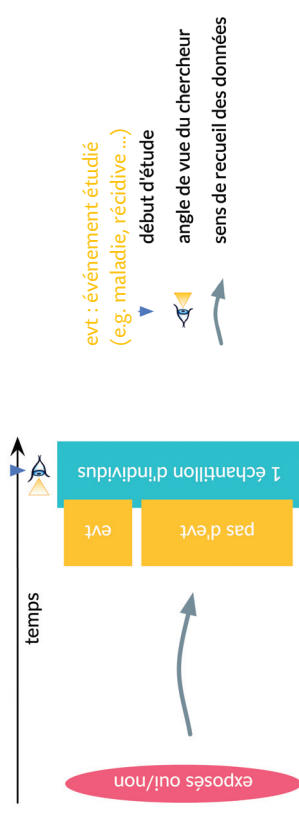
Par conséquent, l'incidence de l'événement de santé (par exemple la survenue de la maladie) dans une étude de cohorte est le reflet de celle de la population source (différent des études cas-témoins).

Les cohortes peuvent être conduites de trois façon différentes, la dernière étant une combinaison des deux premières :

- **cohorte prospective** (prospective cohort) : des patients remplissant les critères d'inclusion sont identifiés, leurs expositions sont recueillies, puis les patients sont suivis au cours du temps pour détecter la survenue future de l'événement de santé



- **cohorte historique** (historical cohort), appelée parfois cohorte rétrospective : des patients remplissant les critères d'inclusion sont identifiés ; on recueille l'exposition au facteur étudié puis la survenue de l'événement de santé ; ces informations étant déjà renseignées au sein de dossiers cliniques ou de bases de données (registre, assurance maladie, ...) : le suivi des patients est historique. Par opposition aux cohortes prospectives, les cohortes historiques sont plus sujettes à des biais de sélection, et de disponibilité des données (exposition non recueillie dans le dossier du patient). Elles sont tout de même de plus en plus fréquentes, en raison de la disponibilité de grandes bases de données (assurance maladie, registres, ...).

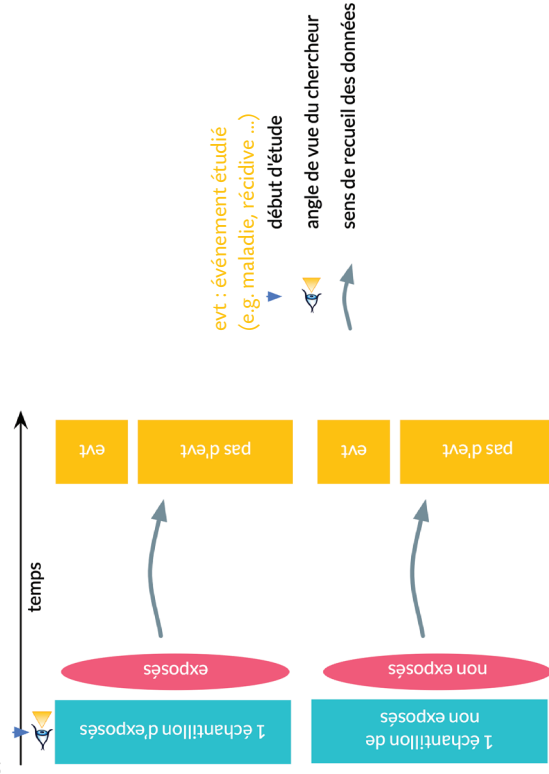


- cohorte ambidirectionnelle / historico-prospective / mixte : étude de cohorte avec des inclusions à la fois historiques et prospectives

NB : différences avec étude cas-témoins :

- o sélection des sujets indépendante du statut malade / non-malade
- o les données relatives à l'exposition sont déjà collectées (dossier clinique ou base de données selon l'étude).

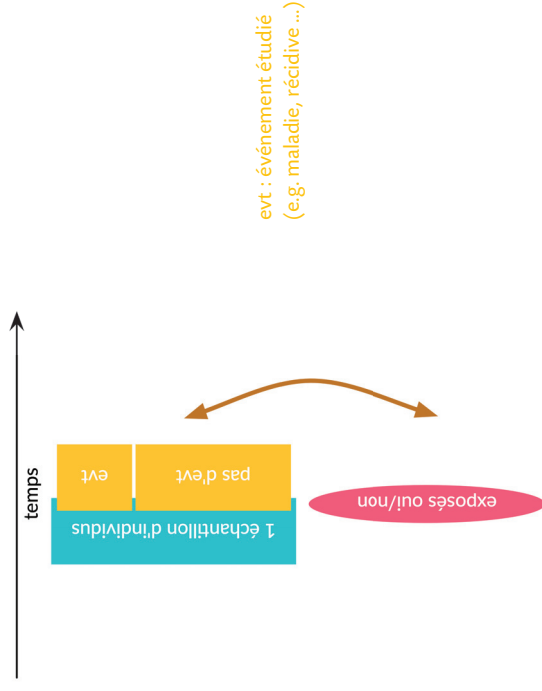
Un cas particulier des études de cohortes correspond aux études de cohorte avec un recrutement de type exposés / non exposés, pour lesquelles on sélectionne des individus exposés au facteur étudié, puis on constitue un autre groupe d'individus non exposés à ce facteur. Ce type d'étude est mené lorsque le facteur étudié est rare, ou concerne une population très spécifique (exemple : exposition professionnelle). Il existe un risque de biais de sélection  $\otimes$  plus élevé dans ce type d'étude par rapport aux cohortes classiques, en raison de la méthode de constitution des deux groupes, qui peuvent venir de populations sources différentes.



## Étude transversale

*cross-sectional*

Enquête qui consiste à mesurer simultanément ou quasi simultanément les données relatives à la maladie et aux facteurs de risques étudiés (pas de suivi dans le temps d'un même patient).



Ce design est approprié à des fins descriptives, par exemple afin d'estimer une prévalence, ou pour estimer les performances d'une méthode diagnostique (le gold standard et le test doivent être évalués au même moment).

Sous certaines conditions, ce design peut être utilisé dans le cadre analytique  $\otimes$ , afin d'évaluer l'association entre une exposition et une maladie, lorsque le facteur étudié est stable au cours du temps (par exemple sexe, ...). Dans le cas contraire, il est difficile de s'assurer que l'exposition a précédé la survenue de la maladie.

## Étude longitudinale

*longitudinal study*

Étude au cours de laquelle un suivi des patients est assuré dans le cadre de l'étude pour recueillir la survenue d'événements, par exemple pour mesurer l'incidence d'une maladie.

## Étude quasi-expérimentale

*quasi-experimental study*

Les études dites « quasi-expérimentales » regroupent différents designs d'études où l'investigateur souhaite évaluer l'effet d'une intervention, mais ne peut pas réaliser de randomisation pour différentes raisons (organisationnelles, éthiques, etc).

## Designs quasi expérimentaux

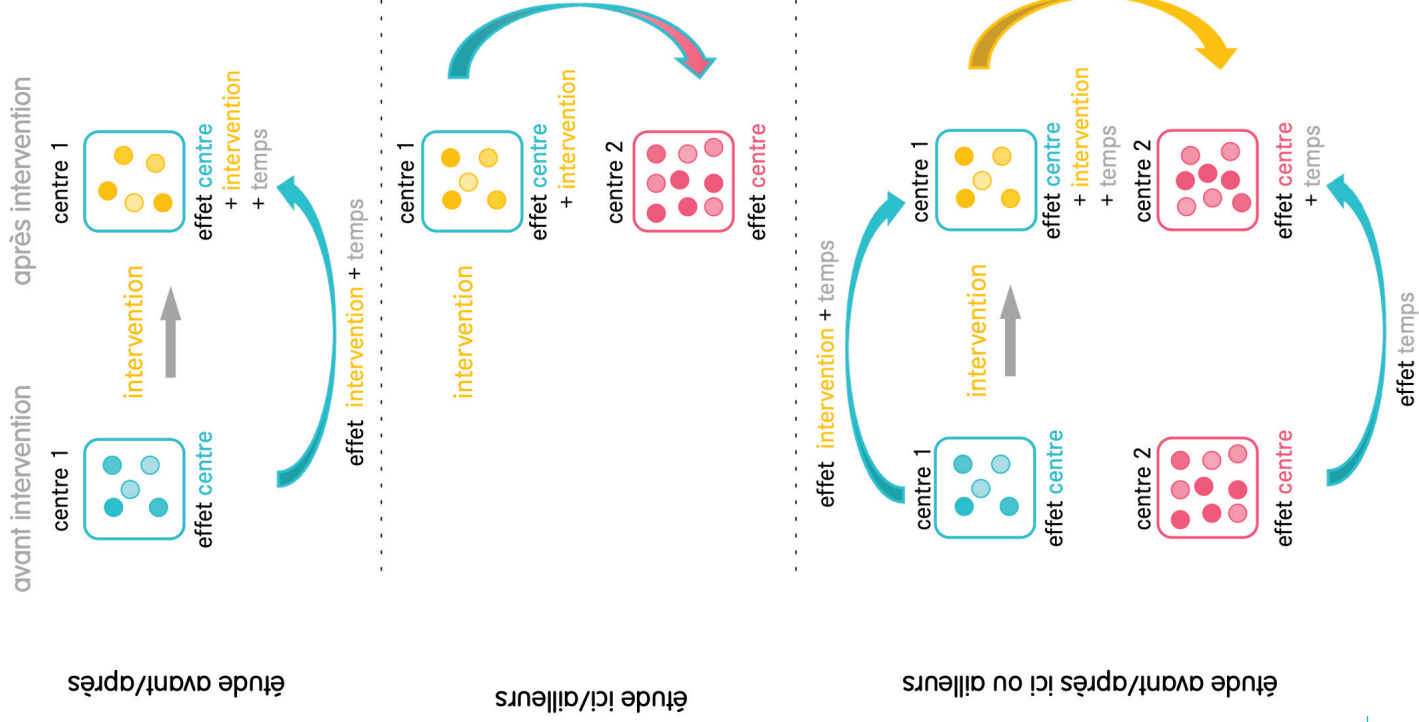
Évaluation d'une formation par simulation à l'annonce de LAT ou de décès aux urgences sur le niveau de stress des proches.

Une des manières de procéder est de mettre en place la formation dans un service d'urgence, et de comparer le niveau de stress des proches entre la période après mise en place de la formation, et la période avant mise en place. Cette étude, quasi-expérimentale, est également nommée étude « **avant-après** » (*before-after*). Ce design ne permet pas de distinguer l'évolution liée à l'effet de la formation d'évolutions temporelles liées à d'autres phénomènes (pénurie temporaire de personnel, vaste épidémie, évolutions de recommandations nationales ...).

Une autre manière de procéder est de mettre en place la formation dans un service et de comparer le stress des proches à celui de proches d'un service d'urgence d'un autre hôpital n'ayant pas bénéficié de la formation. Ce design, quasi-expérimental, est également nommé étude « **ici-ailleurs** ». Ce design ne permet pas de distinguer l'effet de la formation d'une part des différences qui peuvent exister de base entre les deux services d'urgence d'autre part (politiques de service différentes, populations différentes, ...).

La combinaison des deux stratégies, « **avant-après** » et « **ici-ailleurs** » (*difference in difference, controlled before and after study*) permet de s'affranchir de chacun des deux designs isolément (tendances temporelles naturelles et confusion liée aux différences de populations des centres).

(schéma page suivante)



### 📌 Biais des études quasi-expérimentales

Les études quasi-expérimentales souffrent d'un risque de biais plus élevé que les études expérimentales qui sont le « gold-standard » de l'évaluation de l'efficacité d'une intervention, c'est-à-dire les essais cliniques contrôlés randomisés.

Les biais les plus fréquemment retrouvés dans les études quasi-expérimentales sont :

- Le biais de sélection 📌 : il est retrouvé si la population bénéficiant de l'intervention et celle n'en bénéficiant pas ne sont pas comparables. Il est plus fréquemment rencontré que dans les essais cliniques car les populations ne sont pas randomisées.
- Le biais de classement 📌 : il est retrouvé si le recueil des expositions/de l'événement n'est pas réalisé de la même façon dans le groupe bénéficiant de l'intervention et dans celui n'en bénéficiant pas.
- Le biais de confusion 📌 : il résulte de différences au niveau de caractéristiques des populations d'étude non prises en compte dans le protocole et ayant un lien à la fois avec les expositions et l'événement étudié. Il peut notamment apparaître dans les études « avant-après » si les caractéristiques de la population ou les pratiques changent au cours du temps. Il est minimisé par la mise en œuvre d'un design avec un groupe contrôle.

### 📌 Étude : temporalité et recueil

#### 📌 Prospectif *prospective*

##### Étude prospective

Étude dont le début des inclusions est postérieur à la date de finalisation du protocole de l'étude. De cette façon, le recueil de l'exposition n'est pas influencé par la maladie (le statut malade/non malade n'est pas connu lors du recueil de l'exposition).

##### Recueil prospectif

Les données sont collectées après l'inclusion dans l'étude (cf prolectif 📌).

#### 📌 Rétrospectif *retrospective*

##### Étude rétrospective

Étude dont les données collectées sont antérieures à la date de finalisation du protocole de l'étude. De cette façon, le recueil de l'exposition est effectué en connaissant le statut vis-à-vis de la maladie (dans les études cas témoins par exemple), ce qui peut induire un biais de classement/mesure 📌.

##### Recueil rétrospectif

Les données sont déjà disponibles (dossier patient, base médico-administrative) avant l'inclusion dans l'étude (cf rétrolectif 📌).

### 📌 Prolectif

Mode de recueil de données dans une étude : les données sont collectées dans le cadre d'une étude de cohorte 📌 (prospective) suivant une description, une planification, et un objectif bien défini.

### 📌 Rétrolectif

Mode de recueil de données dans une étude : les données sont déjà recueillies avant le début de l'étude (dossier patient, registre, données de l'assurance maladie, mutuelles, ...).

### 📌 Épidémiologie

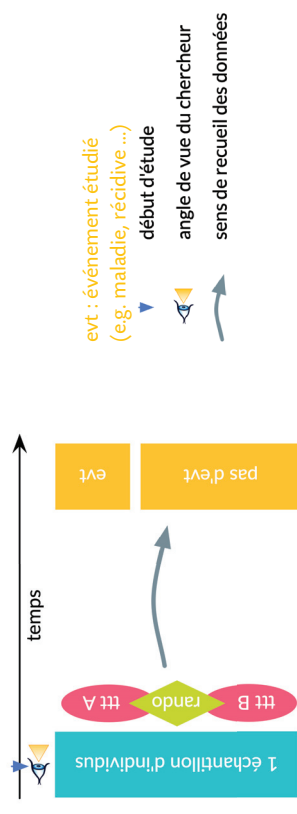
Étude de la distribution des problèmes de santé et des facteurs qui les influencent. On distingue classiquement l'épidémiologie

- descriptive : description d'une population selon différentes caractéristiques, évaluation d'une prévalence, ...
- analytique : évaluation d'une association, le facteur étudié étant observé,
- évaluative : évaluation d'une association, le facteur étudié étant expérimental (i.e. traitement, prise en charge, ...).

Les études descriptives et analytiques sont des études observationnelles.

### 📌 Essai clinique *clinical trial*

Étude expérimentale mise en place pour évaluer une intervention (traitement, stratégie thérapeutique, action préventive, ...).



### Essai clinique contrôlé *controlled clinical trial*

Essai dans lequel un groupe de sujets reçoivent le traitement/intervention de référence (ou le placebo à défaut) et un groupe de sujets reçoivent le nouveau traitement/intervention évalué.

### Essai clinique explicatif *Explanatory clinical trial*

Essai dans lequel on tente de valider une hypothèse physiopathologique ou clinique. Le cadre de la réalisation est strict avec

- (1) une partie restreinte de patients éligibles participant à l'essai,
- (2) des médecins sélectionnés pour les recruter,
- (3) l'intervention délivrée est souvent très encadrée,
- (4) le suivi des patients strict et les critères de jugement sont recueillis de manière standardisée.

### Essai clinique pragmatique *Pragmatic clinical trial*

Essai qui a pour objectif d'informer l'apport en vie réelle d'une intervention. Le cadre de la réalisation est le plus proche de la pratique habituelle,

- (1) avec une partie la plus large de patients éligibles participant à l'essai,
  - (2) des médecins non sélectionnés pour les recruter,
  - (3) intervention délivrée proche de la pratique clinique, le suivi des patients est inchangé,
  - (4) et les critères de jugement sont pertinents pour le patient et sont recueillis selon les procédures de la pratique courante.
- Les 4 caractéristiques doivent être respectées pour qu'un essai puisse être réellement considéré comme pragmatique.

### Essai clinique monocentrique *single center trial*

Qualifie un essai se déroulant dans un seul centre.

### Essai clinique multicentrique *multicenter trial*

Qualifie un essai se déroulant dans plusieurs centres à la fois. Ceci peut être utile soit en termes de faisabilité de l'étude liée à l'effectif de patients, soit pour évaluer l'hétérogénéité d'un centre à l'autre.

La réalisation d'une étude dans plusieurs centres est également un élément améliorant la validité externe de l'étude.

### Essai clinique : type d'intervention

#### Essai thérapeutique *therapeutic trial*

Essai permettant l'évaluation d'une intervention (médicament, acte, prise en charge) à visée thérapeutique chez l'homme.

#### Essai de prévention *prevention trial*

Étude mise en place pour évaluer l'efficacité d'une action de prévention. Ces essais (randomisés ou non) se font chez des personnes indemnes de maladie (prévention primaire) ou malades (prévention secondaire).

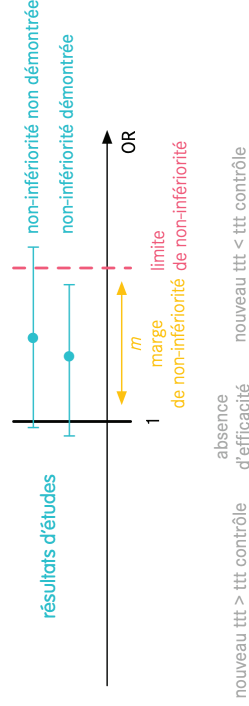
### Essai clinique : hypothèse

#### Essai de supériorité *superiority trial*

Essai clinique visant à montrer une différence entre plusieurs traitements/interventions.

#### Essai de non infériorité *non-inferiority trial*

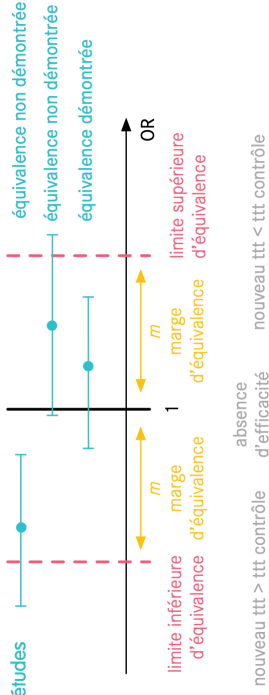
Essai clinique visant à démontrer qu'un nouveau traitement n'est pas moins efficace (inférieur) à qu'un traitement de référence, selon une marge de non-infériorité à définir ( $m$ ), mais présente d'autres avantages par ailleurs (par exemple traitement non inférieur en termes d'efficacité, mais mieux toléré). La non infériorité est démontrée lorsque l'efficacité du nouveau traitement n'est pas inférieure de plus de  $m$  points par rapport celle du traitement de référence. La difficulté consiste à définir la limite acceptable  $m$  : les recommandations de l'EMA et de la FDA proposent de ne pas perdre plus de 50 % de l'effet observé dans une étude antérieure entre le traitement de référence et le placebo (notion approfondie dans les cours).



## Essai d'équivalence

equivalence trial

Essai clinique visant à démontrer qu'un nouveau traitement n'est ni plus ni moins efficace qu'un traitement de référence, selon une marge d'équivalence à définir ( $m$ ). Ce type d'essai est réalisé, par exemple, pour la mise sur le marché de médicaments biosimilaires. L'équivalence est démontrée lorsque l'efficacité du nouveau traitement n'est pas statistiquement inférieure de plus de  $m$  points, ni statistiquement supérieure de plus de  $m$  points, par rapport à celle du traitement de référence.



## Essai clinique : phases

### Essai de phase I

Ces essais portent sur des volontaires sains (ou dans certaines situations, sur des malades, comme en cancérologie, en pédiatrie). L'étude sert à déterminer les paramètres pharmacocinétiques et la dose maximale tolérée.

### Essai de phase II

Ces essais portent sur des malades. Cette phase a pour but, entre autres, d'étudier l'efficacité pharmacologique (critères intermédiaires), biomarqueurs) du produit et de déterminer la dose optimale pour la phase III.

### Essai de phase III

Cette phase correspond aux essais cliniques comparatifs randomisés. Au cours de cette phase, on étudie l'efficacité d'un traitement sur des critères cliniques, ainsi que le rapport efficacité/tolérance.

### Essai de phase IV

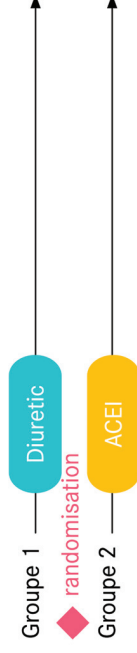
Toute étude thérapeutique réalisée après la commercialisation d'un médicament. Elle concerne principalement l'évaluation pragmatique (dans la vraie vie) de l'efficacité et de la sécurité (pharmacovigilance).

## Essai clinique : designs

### Groupes parallèles

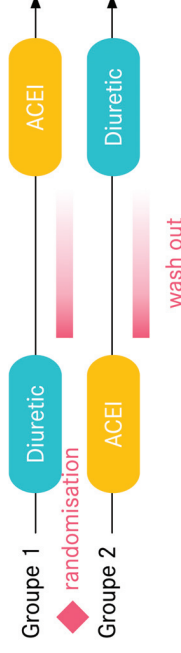
parallel randomized trial

Groupes de patients suivis en parallèle au cours d'un essai randomisé contrôlé : chaque groupe bénéficie d'une seule des stratégies étudiées pendant toute la durée de l'étude, par opposition aux essais en cross-over.



### Cross-over

Dans les essais cliniques randomisés, design dans lequel le sujet est pris comme son propre témoin. 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. Il faut s'assurer dans ce cas l'absence d'effet carry-over des traitements.



Désigne aussi dans les essais d'oncologie la possibilité pour les participants du groupe contrôle, de passer dans le bras expérimental après progression du cancer, afin d'éviter la perte de chance.

### Factoriel (plan)

Le plan factoriel est l'utilisation des mêmes patients pour effectuer simultanément deux comparaisons. La première comparaison est celle du traitement A à son placebo, et la seconde celle du traitement B à son placebo. Les patients de l'essai seront randomisés une première fois entre A et son placebo, puis une seconde fois, sans tenir compte de la nature du premier traitement reçu, entre B et son placebo. Ces deux randomisations simultanées créent quatre groupes de patients :

- 1/4 des patients recevront le traitement A et le traitement B
- 1/4 recevront le traitement A et le placebo du traitement B
- 1/4 recevront le placebo du traitement A et le traitement B
- 1/4 recevront le placebo de A et le placebo de B



L'intérêt du plan factoriel est de permettre de faire deux comparaisons sur les mêmes patients, alors qu'une étude en bras parallèles nécessiterait de constituer trois groupes (groupe placebo, groupe traitement A, groupe traitement B), et impliquerait plus de patients. La limite du plan factoriel est qu'il suppose que l'effet du traitement A sur le patient est indépendant de l'effet du traitement B et vice-versa (absence d'interaction  $\otimes$ ).

|                        | Traitement A<br>Placebo | Traitement A<br>Actif |
|------------------------|-------------------------|-----------------------|
| Traitement B - Placebo | 1/4                     | 1/4                   |
| Traitement B - Actif   | 1/4                     | 1/4                   |


### **Essai en grappes ou clusters** *cluster randomized trial*

Étude expérimentale avec randomisation de groupes d'individus bénéficiant d'interventions différentes. L'unité d'allocation de l'intervention est le groupe d'individus (cluster) et non l'individu en lui-même.

Par exemple, le cluster peut être une institution (établissement de santé, service, praticien) ou une entité administrative ou géographique (région, département, commune).

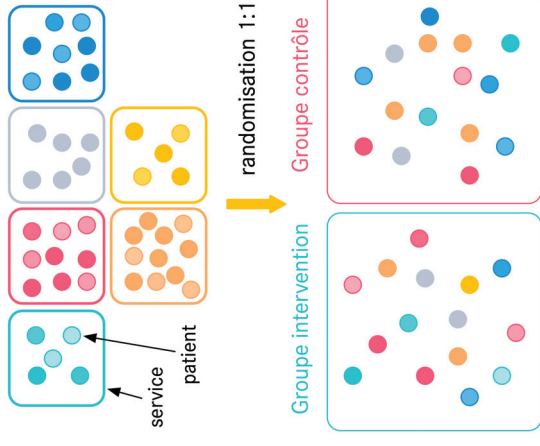
Les raisons principales pour réaliser une étude avec randomisation en cluster plutôt qu'avec randomisation individuelle sont liées :

- à la nature de l'intervention : intervention s'appliquant naturellement au niveau cluster (politique de santé, réorganisation des soins, formation de professionnels) ou intervention dispensée en groupe (éducation thérapeutique groupale)
  - au risque de contamination  $\otimes$  entre les groupes pouvant survenir en cas de randomisation individuelle
  - à des raisons de faisabilité et acceptabilité (randomisation individuelle peu acceptable éthiquement, par exemple la vaccination dans une école).
- Les essais randomisés en clusters sont très souvent utilisés pour évaluer l'efficacité d'interventions non médicamenteuses impliquant des changements de pratiques des professionnels de santé et de l'organisation des soins.

 Essai randomisé évaluant l'efficacité d'un programme de prévention de l'obésité infantile. Groupe expérimental : programme de prévention délivré par les internes de pédiatrie à chaque visite de l'enfant dès la naissance (fiches, gobelets et bol avec doses appropriées, ...). Groupe contrôle : pas de programme. La mise en place du programme dans le service implique de former tous les internes sur la communication et les techniques d'entretien. Avec une randomisation individuelle, un interne formé risque malgré tout de délivrer inconsciemment des messages aux parents d'enfants du groupe contrôle (biais de

contamination). L'unité de randomisation nécessaire pour éviter le biais de contamination est donc le service pédiatrique (le groupe contrôle correspond aux cliniques n'ayant pas bénéficié de la formation).

### Randomisation individuelle



unité de randomisation = patient

### **Essai clinique : aveugle**

 **Essai en double aveugle**  
*double blind trial*

Essai au cours duquel ni le patient, ni le médecin, ne connaissent le traitement pris. Cela permet d'éliminer l'effet placebo chez le patient et les biais de mesure  $\otimes$  liés à la subjectivité du médecin (l'évaluateur est en aveugle  $\otimes$ ). On peut parler d'essai en triple aveugle lorsque le chercheur qui analyse les résultats ne sait pas quel traitement a reçu chaque groupe de patients.

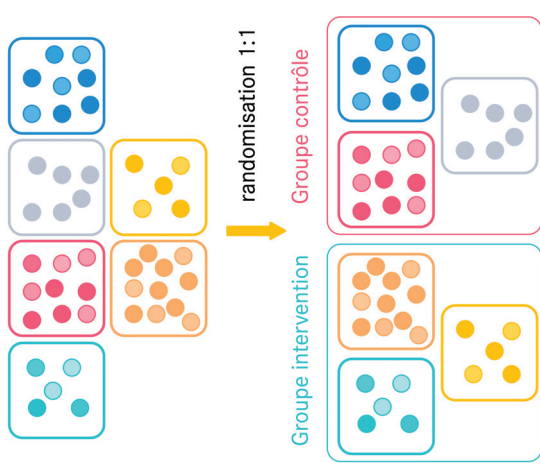
Syn. : essai en double insu.

 **Essai en simple aveugle**  
*single blind trial*

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.

Syn. : essai en simple insu.

### Randomisation en grappes



unité de randomisation = service

## Essai en ouvert

*open trial*

Essai thérapeutique dans lequel le patient et le médecin connaissent le traitement alloué. Ceci peut induire un biais de classement/mesure. Les essais en ouvert sont réalisés lorsque l'aveugle n'est pas possible pour des raisons pratiques, éthiques ou scientifiques.

Par exemple, il est peu pratique/délicat d'envisager un placebo d'une chirurgie, ou encore peu pratique de réaliser un placebo d'un traitement composé de différentes formes galéniques (réalisation auquel cas d'un double placebo), ou encore difficile de réaliser un placebo d'un traitement avec un effet secondaire pathognomonique de sa prise (coloration des urines, photosensibilité etc...).

Les essais de stratégie de prise en charge (comparaison de séquences d'exams et actes) sont également réalisés en ouvert. Ceci peut induire un biais de contamination.

## Essai séquentiel

*sequential trial*

Essai thérapeutique dont une ou plusieurs analyses seront effectuées avant la fin programmée de l'étude au cours d'analyses intermédiaires (au fur et à mesure de la survenue des événements) ; ceci peut permettre de diminuer le nombre moyen de patients à inclure dans l'étude (si l'étude est arrêtée précocement). La réalisation de ces analyses doit être prévue dans le protocole. Les analyses statistiques doivent prendre en compte la multiplicité des tests pour le contrôle du risque  $\alpha$  (un test est réalisé à chaque analyse intermédiaire). Par exemple, on dépensera 1 % du risque  $\alpha$  lors de l'analyse intermédiaire, et 4 % lors de l'analyse finale (sur les deux analyses, le risque  $\alpha$  global est bien de 5 %).

## Evaluateur aveugle

*blinded assessor*

Dans une étude, investigateur tiers qui, dans un but d'objectivité, ne connaît pas le traitement reçu par le patient (pour un essai clinique), le statut du patient vis-à-vis de la maladie (pour une étude cas-témoins, de cohorte, ou diagnostique) et qui mesure le critère de jugement (essai clinique), l'exposition (étude cas-témoins) ou le test diagnostique (étude diagnostique). Permet de limiter le biais de classement.

## Estimation ponctuelle

*estimate*

voir partie Inférence

## Exposition

*exposure*

L'exposition est le(les) facteur(s) d'intérêt faisant l'objet de l'étude.

Dans le cas des études expérimentales, l'exposition correspond à l'intervention évaluée (le patient est exposé de façon intentionnelle à une exposition : un traitement, un examen, une stratégie de prise en charge...).

Dans le cas des études analytiques, l'exposition à un ou plusieurs phénomènes extrinsèques (amiante, colorants, consommation de tabac) ou intrinsèques (sexe, l'âge, taille ...) au patient est mesurée, et on cherche à déterminer si cette exposition est un facteur de risque / un facteur protecteur.

## F

## Facteur de risque / pronostique

*risk factor / prognostic factor*

Facteur associé au risque de survenue d'un événement.

On utilise plus souvent le terme de facteur de risque pour la prévention primaire (apparition de la maladie), et le terme de facteur pronostique pour l'évolution de la maladie.

On parle de facteur protecteur pour désigner un facteur associé à une diminution du risque de l'événement.

## Factoriel (Plan)

voir partie Essai clinique : designs

## Forest-plot

Représentation graphique fréquemment utilisée pour représenter un résultat dans différentes situations.

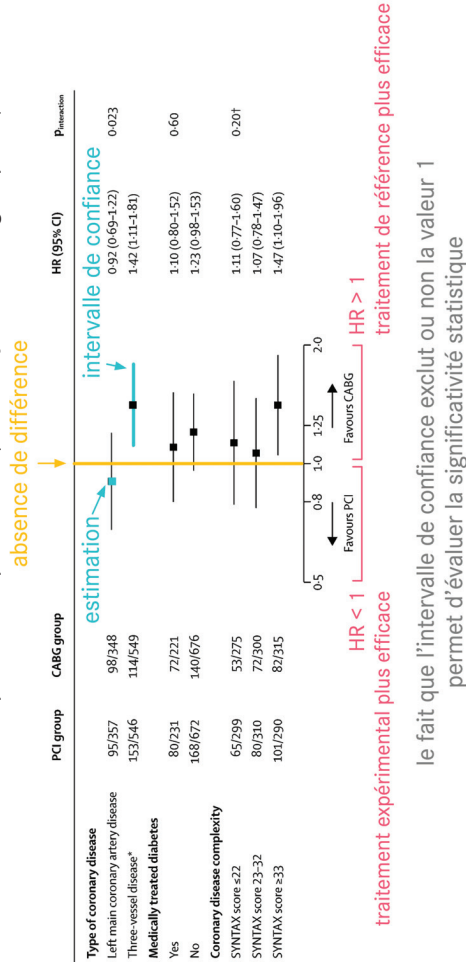
## Essais cliniques

Dans les essais cliniques, un forest-plot peut être utilisé pour décrire le bénéfice du traitement expérimental par rapport au traitement de référence dans différents sous-groupes, afin d'émettre des hypothèses concernant des sous-groupes pour lesquels le traitement expérimental serait plus efficace que le traitement de référence.

La mesure d'association entre les traitements et le critère de jugement (risque relatif ou odds ratio, hazard ratio) est représentée dans les différents sous-groupes, avec l'intervalle de confiance associé.

Les p-values d'interaction sont parfois rajoutées sur le graphique ; elles indiquent, pour chaque ensemble de sous-groupes, si le bénéfice du traitement expérimental varie selon les sous-groupes.

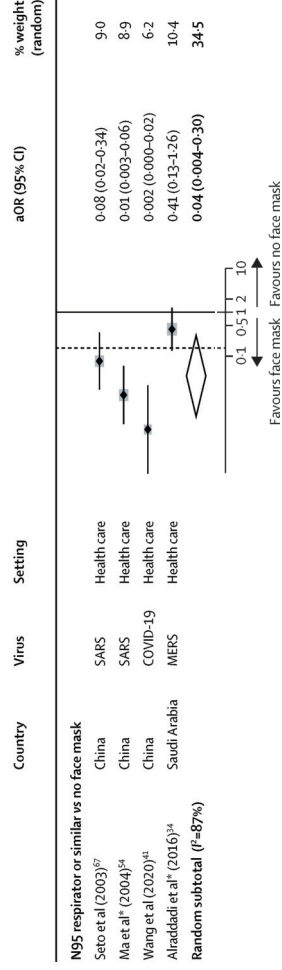
Ces résultats sont à interpréter avec précaution (voir Analyse en sous-groupe ).



Dans cet exemple, le pontage coronarien (CABG) réduit statistiquement la mortalité toute cause chez les patients tritronculaires et chez les patients dont le score SYNTAX (qui mesure la complexité de la maladie coronarienne) est supérieur ou égal à 33. Globalement, l'effet du CABG est assez homogène, à l'exception des patients présentant une atteinte prédominante sur le réseau coronaire gauche (left main coronary artery disease), pour lesquels un effet délétère (mais non significatif) semble apparaître. Les p-values d'interaction suggèrent un effet statistiquement hétérogène du traitement selon le type de coronaropathie ( $p = 0,023$ ).

### Méta-analyse

Dans les méta-analyses, le forest-plot représente les résultats du traitement expérimental par rapport au traitement contrôle retrouvés dans les différentes études. On retrouve aussi ce type de représentation pour montrer la variabilité d'une estimation entre différentes études (odds ratio, prévalence, performances de méthodes diagnostiques, ...). Souvent, la dernière ligne du forest-plot représente l'effet global du traitement expérimental, moyenné/poolé sur l'ensemble des études.




## G

### Gold Standard

Test diagnostique utilisé comme référence. Dans une démarche d'évaluation de méthode diagnostique, c'est l'examen que l'on considère comme donnant la meilleure certitude diagnostique (test donnant le résultat juste, ou à défaut le plus juste possible), et contre lequel sont évalués les nouveaux tests.

### Groupe contrôlé (témoin) / expérimental (interventionnel)

control group / experimental (interventional) group


Dans un essai contrôlé  , le groupe contrôlé (ou témoin) est le groupe qui bénéficie de la prise en charge de référence (par exemple le médicament de référence ou le placebo pour les essais thérapeutiques), par opposition au groupe expérimental/interventionnel qui bénéficie de la stratégie évaluée (par exemple le nouveau médicament).

### Groupes parallèles

voir partie Essai clinique : designs 




## H

### Hypothèses d'un test statistique

voir partie Tests statistiques 

### Hypothèse du biais maximum

worst case scenario

L'hypothèse du biais maximum est utilisée dans les essais cliniques au moment de l'analyse statistique en intention de traiter  , afin de prendre en compte toutes les déviations au protocole  (écarts au protocole) qui peuvent impacter le résultat de l'étude (exemples : changement de bras de traitement en cours d'étude, donnée manquante). Ces déviations sont prises en compte de la façon la plus défavorable possible pour le traitement expérimental. De la même façon, le critère de jugement  pour les patients perdus de vue peut être considéré de la façon la plus défavorable possible pour le traitement expérimental (par exemple, patients perdus de vue considérés en échec dans le groupe expérimental, et en succès dans le groupe témoin pour un critère de jugement binaire). En cas de supériorité statistique démontrée du traitement expérimental, la conclusion est d'autant plus convaincante lorsque l'analyse est faite selon l'hypothèse du biais maximum. L'hypothèse du biais maximum peut être adaptée à chaque étude ; elle doit être définie dans l'article.

## I

### I<sup>2</sup>

Le I<sup>2</sup> est une mesure utilisée pour évaluer l'hétérogénéité des résultats au sein d'une méta-analyse ☞. Un effet traitement peut varier en raison de fluctuations d'échantillonnage (le hasard) ou à cause de « vraies » variations (biologie, environnement etc...). Le I<sup>2</sup> répond à la question « quelle est la proportion de la variabilité de l'effet traitement entre les études reflétant une vraie différence d'effet entre les études ? ». Conceptuellement

$$I^2 = \frac{\text{vraie variabilité}}{\text{variabilité totale}}$$

Un I<sup>2</sup> à 25 % est faible, à 50 % est modéré et 75 % élevé.

En cas de I<sup>2</sup> élevé, le résultat agrégé n'est plus interprétable ; des analyses en sous-groupes sont nécessaires afin d'identifier les causes de l'hétérogénéité.

### Incidence

voir partie Mesure d'événements ☞

### Indépendance

*independency*

Événements indépendants : deux événements sont indépendants si la survenue de l'un n'influe pas sur celle de l'autre.

Échantillons indépendants : deux échantillons de mesure sont indépendants si les mesures sont effectuées sur des individus différents, et non appariés ☞ (par exemple, dans un essai en groupes parallèles ☞). Le fait que les mesures soient indépendantes ou non influence la méthode statistique utilisée (e.g. test de comparaison de moyennes pour données indépendantes ou pour données appariées).

### Inférence

*inference*

L'inférence est le processus permettant de tirer des conclusions sur une population uniquement à partir d'un échantillon issu de celle-ci. Ainsi, on associe un intervalle de confiance ☞ à une estimation d'un paramètre (par exemple à l'estimation d'un odds ratio) pour savoir où peut se situer la vraie valeur pour la population.

Il existe deux types d'inférence : l'inférence fréquentiste et l'inférence bayésienne.

### Inférence fréquentiste

Elle considère que le paramètre (par exemple un odds ratio) à estimer possède une vraie valeur, mais qu'il existe des variations dans son estimation d'un échantillon à l'autre en raison de fluctuations aléatoires.

L'intervalle de confiance ☞ indique où pourrait se situer la valeur estimée d'un paramètre si l'expérience était répétée un grand nombre de fois, au vu des résultats issus de l'échantillon. Sa largeur reflète donc les fluctuations aléatoires. Si l'intervalle de confiance est large, les fluctuations peuvent être grandes d'une expérience à l'autre ; la valeur obtenue dans l'échantillon d'origine est donc peu fiable.

La p-value ☞ indique si la valeur estimée du paramètre dans l'échantillon est compatible avec la valeur définie sous l'hypothèse nulle (par exemple, OR = 1) en raison des fluctuations aléatoires : on évalue la « fréquence » de cette valeur dans la distribution des valeurs attendues sous l'hypothèse nulle ☞ si l'expérience était répétée un grand nombre de fois. Si la fréquence est trop faible, la valeur du paramètre dans la population source est considérée réellement différente de celle de l'hypothèse nulle (l'écart ne peut pas s'expliquer par les simples fluctuations aléatoires).

☞ voir aussi partie Test statistique

### Inférence Bayésienne

☞ L'inférence fréquentiste raisonne relativement à une hypothèse nulle (souvent l'absence d'effet du traitement ou d'association) que l'on souhaite rejeter à l'issue de l'expérimentation, et considérée comme vraie (par exemple RR = 1, pente nulle, différence de moyennes nulle). Il est ensuite évalué la probabilité (vraisemblance) d'observer les données de l'étude sous cette hypothèse nulle (en lien avec les fluctuations aléatoires) ; si cette probabilité est faible (seuil classique <0,05), l'hypothèse nulle est rejetée (par exemple on considère que RR ≠ 1), et on considère qu'il y a un effet du traitement ou une association.

L'inférence Bayésienne ne calcule pas la probabilité des données observées sous une hypothèse considérée d'emblée comme vraie, mais directement la probabilité que cette hypothèse soit vraie d'après les données obtenues. Cette probabilité de l'hypothèse est évaluée au travers de la formule de Bayes, et est proportionnelle à la probabilité des données d'après l'hypothèse, ou vraisemblance, multipliée par la probabilité de l'hypothèse avant de disposer des données de l'étude, appelée aussi probabilité a priori (obtenue à partir des résultats d'autres études ou d'avis d'experts).

La différence essentielle est donc que l'inférence Bayésienne donne une crédence (degré de confiance) ou intervalle de crédibilité d'une hypothèse donnée alors que l'inférence fréquentiste donne une probabilité d'observer des données sous une hypothèse nulle qu'on souhaite rejeter. Par ailleurs, l'inférence Bayésienne reflète bien la démarche de recherche

qui consiste à actualiser les connaissances disponibles sur un sujet (information a priori) à partir des données d'une nouvelle étude. Elle est particulièrement utile pour les études de faible effectif (par exemple sur les maladies rares) pour lesquelles de l'information a priori est disponible dans la littérature.

### Estimation ponctuelle *estimate*

Résultat correspondant à une valeur approchée (estimée) d'un paramètre, la vraie valeur étant inaccessible en raison des fluctuations d'un échantillon à l'autre.

Exemple : le pourcentage d'événements estimé à partir d'un échantillon varie d'un échantillon à l'autre. L'âge moyen de survenue d'un cancer du sein dans un échantillon de femmes représentatif peut être différent de l'âge moyen de ces femmes sur l'ensemble de la France en raison du simple hasard d'échantillonnage (par hasard, l'échantillon a sur ou sous représenté les femmes âgées).

Ainsi, on associe à une estimation ponctuelle un intervalle de confiance

### Intervalle de confiance *confidence interval*

Définition utilisée couramment : fourchette de valeurs qui encadre une estimation. Par exemple, un intervalle de confiance à 95 % est tel que la vraie valeur du paramètre à estimer a 95 % de chances d'être comprise entre les deux bornes de l'intervalle. L'intervalle de confiance est un des fondements de la notion d'inférence statistique fréquentiste , qui reconnaît des fluctuations aléatoires d'un échantillon à l'autre.

Attention, cette définition est en réalité une simplification ; elle est fréquemment utilisée car elle reste valable pour les effectifs élevés.

### 

#### Kappa de concordance

Le coefficient kappa (Kappa de Cohen) est utilisé pour évaluer la concordance entre deux observateurs/méthodes dans l'évaluation d'un critère catégoriel. Une méthode simple consisterait à évaluer le pourcentage d'observations classées dans la même catégorie par les deux observateurs/méthodes, mais le résultat serait trompeur. Les observations peuvent être classées dans la même catégorie par simple hasard : pour deux catégories, ceci est d'autant plus fréquent que la vraie proportion est proche de 0 ou de 1.

Le coefficient kappa mesure la concordance entre les observateurs/méthodes au-delà de ce qui peut être attendu par le simple hasard :

### Calcul

$$\kappa = \frac{\% \text{résultats concordants} - \% \text{concordants hasard}}{1 - \% \text{concordants hasard}}$$

Il peut prendre des valeurs entre -1 et 1, les valeurs négatives indiquent un désaccord entre les évaluateurs/méthodes.

| $\kappa$    | Interprétation         |
|-------------|------------------------|
| $< 0$       | Désaccord              |
| 0.0 - 0.2   | Accord très faible     |
| 0.21 - 0.40 | Accord faible          |
| 0.41 - 0.60 | Accord modéré          |
| 0.61 - 0.80 | Accord fort            |
| 0.81 - 1.00 | Accord presque parfait |

Le coefficient peut être généralisé à plus de deux observateurs/méthodes. Il faut rester vigilant, car l'interprétation de la valeur dépend des proportions des différentes catégories.

### 

Une nouvelle technique d'IRM est évaluée pour détecter la présence de cancer de la prostate (oui/non). Lors de l'évaluation de cette technique, un préliminaire est que la classification des images soit reproductible d'un radiologue à l'autre (variabilité inter-évaluateur). Un certain nombre d'images sont évaluées par deux radiologues. La reproductibilité inter-évaluateur peut être évaluée par le coefficient kappa.

### 

#### Loi Jardé

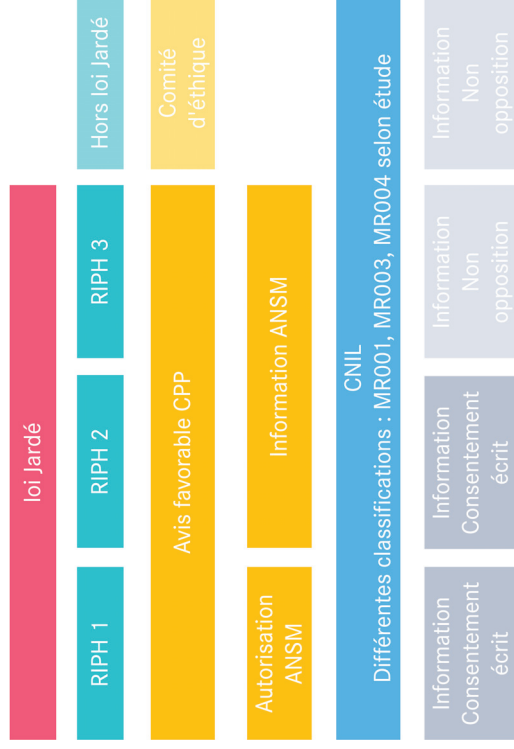
La loi Jardé est la loi française encadrant les recherches sur l'être humain en vue du développement des connaissances biologiques ou médicales. Elle est en vigueur depuis 2016 (mise en application) et répartit les recherches en santé en 3 catégories de recherches impliquant la personne humaine (RIPH) :

- RIPH 1 = Recherches interventionnelles qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle : exemple : essai thérapeutique
- RIPH 2 = Recherches interventionnelles qui ne comportent que des risques et des

contraintes minimales : exemple : étude randomisée comparant 2 stratégies diagnostiques déjà utilisées en pratique, étude nécessitant l'ajout d'un examen paramédical non réalisé habituellement

- **RIPH 3** = Recherches non interventionnelles qui ne comportent aucun risque ni contrainte dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle, sans procédure supplémentaire ou inhabituelle de diagnostic, de traitement ou de surveillance ; exemple : suivi par questionnaires (échelle d'anxiété, de dépression, de qualité de vie, de mesure de la douleur)
- Les recherches portant sur des données déjà recueillies dans la pratique ou ne recueillant pas des données de santé sont classées en dehors de la loi Jardé (exemple étude de satisfaction, analyse de données des dossiers médicaux, analyse des données de l'assurance maladie).

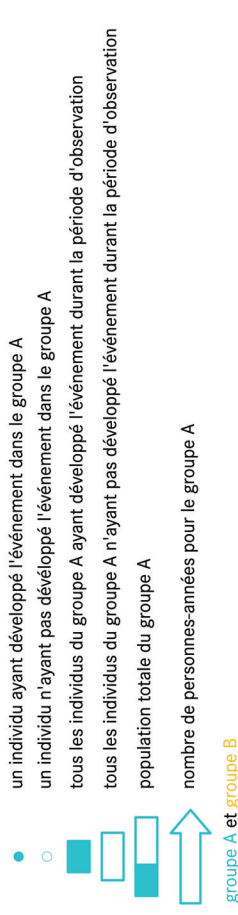
Ces différentes catégories impliquent des démarches réglementaires différentes et des modalités d'information et recueil du consentement différentes.



## Mesures d'événements/d'associations

Différents indicateurs permettent de mesurer la fréquence/vitesse d'événements/maladies, et de quantifier l'association entre un facteur étudié et l'événement. Ils sont illustrés dans le schéma ci-après, puis définis.

Chaque indicateur est à interpréter en fonction de son intervalle de confiance ; les résultats sont peu fiables si l'intervalle de confiance est large.



groupe A et groupe B

Au niveau d'une seule population : mesure de risque, d'odds, d'odds (ou cote) ou de vitesse.

Comparaison de deux populations : mesures d'association (additif ou multiplicatif) quantifiant un changement absolu (additif) ou relatif (multiplicatif) de risque, d'odds ou de vitesse du groupe A au groupe B.

| Indicateur  | Interprétation                                 | Exemple  |
|---|--|--|
| Risque <i>risk</i>  | probabilité [0, 1]                             | $R_a = 2/5 = 40\%$<br>le risque de survenue d'événement durant la période était de 40 %  |
| Risque relatif <i>relative risk</i>                           | rapport de probabilités [0, +∞[                | $RR = R_a/R_b = 2$<br>le risque du groupe A était 2 fois plus élevé que celui du groupe B  |
| Différence de risque <i>risk difference</i>                   | différence de probabilités [-1, 1]             | $DR = R_a - R_b = 20\%$<br>la différence de risque entre le groupe A et le groupe B est de 20 %  |
| Différence de risque relative <i>relative risk difference</i> | différence de probabilités relatives [-∞, +∞]  | $DRR = (R_a - R_b)/R_b = 100\%$<br>l'augmentation relative du risque du groupe A par rapport au groupe B est de 100 %                                |
| Odds  | nombre d'événement par non événement [0, +∞[   | $Odds_a = 2/3 = 0,7$<br>pour 1 patient sans événement, il y avait 0,7 patient avec événement   |
| Odds ratio  | rapport d'odds [0, +∞[                         | $OR = Odds_a/Odds_b = 2,7$<br>l'odds du groupe A était 2,7 fois plus élevé que celui du groupe B   |
| Incidence   | vitesse (nouveaux cas/ personne-temps) [0, +∞[ | $I_a = 10/1000 = 0,01$<br>pour des individus suivis durant une durée équivalente à un suivi de 1000 personnes durant 1 an, on aurait 10 nouveaux cas |
| Incidence ratio <i>incidence rate ratio</i>                   | rapport de vitesses [0, +∞[                    | $IR = I_a/I_b = 5$<br>pour un suivi équivalent, on observerait 5 fois plus de nouveaux cas dans le groupe A que dans le groupe B                     |

## 📌 Incidence

Nombre de nouveaux cas d'une maladie, survenus au cours d'une période donnée, rapporté au nombre d'unités personne-temps exposées au risque dans la population. On parle souvent de taux d'incidence, parfois de densité d'incidence.

⚠️ Attention, dans certaines études, l'incidence est parfois calculée en intégrant au dénominateur uniquement le nombre de personnes à risque, et non pas leur durée d'exposition (le nombre de personnes-années) ; ce résultat correspond au taux d'attaque. L'incidence s'estime à partir d'études de cohorte 🔄.

## 📌 Ratio d'incidence

*incidence rate ratio*

Métrique utilisée pour quantifier l'association entre un facteur d'exposition et la vitesse de survenue d'un événement. L'abréviation classique est *IRR*.

Il correspond au ratio des incidences d'événement entre le groupe de personnes présentant l'exposition étudiée et le groupe de personnes ne présentant pas l'exposition.

Le ratio d'incidence peut s'interpréter de la façon suivante : les personnes présentant l'exposition étudiée ont une incidence de survenue d'événement (vitesse) « *IRR* » fois plus élevée que les personnes non exposées (si l'*IRR* est supérieur à 1).

*IRR* > 1 : l'exposition est associée à une augmentation de la vitesse de survenue de l'événement (facteur de risque) ; *IRR* < 1 : l'exposition est associée à une diminution de la vitesse de survenue de l'événement (facteur protecteur).

La valeur de l'*IRR* doit s'interpréter avec son intervalle de confiance 🔄.

📌 L'*IRR* peut être obtenu à partir d'un modèle de régression de Poisson intégrant en offset les nombres de personne-temps exposées.

## 📌 Taux d'incidence standardisé

*standardized incidence*

Le taux d'incidence brut d'un événement de santé dépend de différentes caractéristiques liées à la population (âge, sexe...). Ainsi, si l'on veut comparer des taux d'incidence dans deux populations différentes, la comparaison de taux d'incidence bruts sera biaisée par ces facteurs. Il n'est donc pas possible de les comparer directement.

Les taux standardisés permettent de neutraliser les différences dues à des variables influençant le phénomène étudié (facteur de confusion), telles l'âge, le sexe, le niveau socio-économique, afin d'autoriser la comparaison de ces taux entre eux.

Par exemple, si l'on souhaite comparer la mortalité entre la France et la Chine (avec des niveaux de vie différents) la comparaison des taux bruts de mortalité révélera que la mortalité est plus élevée en France. Ce résultat s'explique par un âge plus élevé en France, alors que la population est plus jeune en Chine.

Pour comparer la mortalité entre les deux pays au-delà de la différence d'âge, il convient donc de réaliser une comparaison de la mortalité à âge égal (afin de gommer l'effet de l'âge sur la mortalité). Pour cela, une standardisation des taux est nécessaire.

Il existe deux méthodes de standardisation : la standardisation directe et la standardisation indirecte.

**Standardisation directe** (*direct standardization*) : aussi appelée méthode de la population type, c'est la méthode la plus fréquemment utilisée.

Dans cette méthode, on calcule les taux d'incidence que l'on s'attendrait à trouver dans les populations étudiées si elles avaient les mêmes distributions des variables de confusion (par exemple l'âge, le sexe, le niveau socio-économique...). Pour cela, on utilise la distribution des variables de confusion d'une même population de référence. La standardisation directe consiste à appliquer à la population de référence les taux spécifiques d'incidence des deux populations pour les différents strates des facteurs de confusion. Il est ainsi obtenu le nombre de cas « attendus » dans chaque strate si les populations avaient la même distribution des variables de confusion. Le taux d'incidence standardisé est obtenu en divisant le total de cas attendus par l'effectif de la population de référence.

**Standardisation indirecte** (*indirect standardization*) : aussi appelée méthode des taux-types, elle est parfois utilisée pour la comparaison de la mortalité dans deux populations lorsque les taux de mortalité spécifiques à chaque strate ne sont pas disponibles dans les populations étudiées.

Dans cette méthode, au lieu d'utiliser la structure de la population de référence, on utilise ses taux d'incidence spécifiques par strate. Ces taux d'incidence spécifiques par strates sont appliqués aux populations à comparer, que l'on a tout d'abord stratifiées selon les variables de confusion à contrôler (par exemple l'âge, le sexe, le niveau socio-économique...). De cette façon, on obtient le nombre total de cas attendus dans les deux populations à comparer si la distribution des facteurs de confusion était celle de la population de référence.

Le **Ratio de Mortalité Standardisé** (Standardized mortality ratio, **SMR**) est alors calculé en divisant le total de cas observés par le total de cas attendus dans les deux populations à comparer. Ce rapport permet de comparer la mortalité de chaque population étudiée à celle de la population de référence :

📌 Calcul

$$\text{SMR (en \%)} = \frac{\text{Nombre de décès bruts observés}}{\text{Nombre de décès attendus}} \times 100$$

Si le taux observé de décès est supérieur au taux attendu, le SMR sera supérieur à 100 %. À l'inverse, si le taux observé de décès est inférieur au taux attendu, le SMR sera inférieur à 100 %. Cette méthode permet ainsi de comparer les taux standardisés de mortalité entre les deux populations d'étude.

## Ⓜ Odds (cote)

Métrique utilisée pour quantifier la fréquence d'un événement dans une population (alternative à la métrique du risque de survenue d'événement). Il s'agit du ratio entre la probabilité de l'événement et la probabilité d'absence de cet événement (ratio entre le nombre de patients avec et sans événement).

Il est compris entre 0 et l'infini. Plus le risque d'événement est élevé, plus l'odds est élevé. Un risque de 50 % correspond à un odds de 1. Un odds de 4 peut vouloir dire que, sur un échantillon de 5 individus, 4 présentent l'événement et 1 ne le présente pas (4 contre 1).

## Ⓜ Odds ratio (rapport de cotes)

Métrique utilisée pour quantifier l'association entre un facteur d'exposition Ⓜ et un événement binaire (survenue d'une maladie, décès, etc.).

Il s'agit du rapport entre deux odds : celui estimé chez les exposés (ou dans le groupe expérimental pour un essai clinique) sur celui estimé chez les non exposés (groupe contrôle). Les individus exposés ont un odds de maladie « OR » fois plus élevé que les individus non exposés.

OR > 1 : exposition aggravante (facteur de risque) ; OR < 1 : exposition protectrice (facteur protecteur).

Si la prévalence de la maladie (ou de l'événement) est faible, l'OR est une bonne approximation du risque relatif Ⓜ.

La valeur de l'OR doit s'interpréter avec son intervalle de confiance Ⓜ. Dans les analyses univariées Ⓜ, le test du Chi 2 permet de tester si l'odds ratio est statistiquement différent de 1.

L'odds ratio de maladie peut s'estimer à partir d'une étude de cohorte Ⓜ ou cas-témoins Ⓜ ; dans les études cas-témoins, seul l'odds ratio peut être estimé pour quantifier la relation entre le facteur d'exposition et l'événement.

## Ⓜ Calculs

|                      |                      |                      |       |
|----------------------|----------------------|----------------------|-------|
|                      | <b>M<sup>+</sup></b> | <b>M<sup>-</sup></b> |       |
| <b>E<sup>+</sup></b> | a                    | b                    | a + b |
| <b>E<sup>-</sup></b> | c                    | d                    | c + d |
|                      | a + c                | b + d                | N     |

E<sup>+</sup> : exposés      E<sup>-</sup> : non exposés      M<sup>+</sup> : malades      M<sup>-</sup> : non malades      N total

$$OR = \frac{\text{Odds}(M^+ | E^+)}{\text{Odds}(M^+ | E^-)} = \frac{a/b}{c/d} = \frac{a \times d}{b \times c}$$

## Ⓜ Risque risk

En épidémiologie, un risque se définit comme la probabilité de survenue d'un événement pendant une période de temps dans une population donnée.

Pour des populations importantes, le risque se définit alors comme la fréquence de survenue d'un événement donné dans une population donnée pendant une période de temps donnée.

⚠ Attention, le terme de risque est parfois utilisé par le grand public pour parler d'incidence ou d'odds, alors qu'il s'agit de concepts différents (voir incidence Ⓜ, ratio d'incidence Ⓜ, odds Ⓜ, odds ratio Ⓜ).

Le risque de maladie s'estime à partir d'une étude de cohorte Ⓜ.

## Ⓜ Différence de risque absolute risk difference

La différence de risque (absolue) est un indicateur qui mesure l'association entre un facteur d'exposition binaire Ⓜ et un événement binaire (survenue d'une maladie, décès, etc.). Il correspond à la différence entre le risque de développer un événement étudié durant une période donnée dans un groupe de personnes présentant l'exposition étudiée (R<sub>exposés</sub>) et ce même risque dans un groupe de personnes ne présentant pas l'exposition (R<sub>non-exposés</sub>). En français, il est noté DR ou DR<sub>A</sub>, en anglais ARR (absolute risk reduction).

$$DR = R_{\text{exposés}} - R_{\text{non exposés}}$$

DR > 0 : l'exposition est associée à une augmentation du risque de survenue de l'événement (facteur de risque) ; DR < 0 : l'exposition est associée à une diminution du risque de survenue de l'événement (facteur protecteur).

Il faut être prudent dans l'interprétation : une différence de risque de 5 % lorsque le risque dans le groupe non exposé est de 5 % n'a pas la même conséquence qu'une même différence de risque, mais pour un risque dans le groupe non exposé de 50 %.

Une mesure relative a été proposée : il s'agit de la différence de risque relative (DRR), souvent appelée en anglais risk relative reduction (RRR) :

$$RRR = \frac{R_{\text{exposés}} - R_{\text{non exposés}}}{R_{\text{non exposés}}}$$

La réduction relative de risque s'interprète en pourcentage du risque de base ; ainsi, une même différence de risque n'aura pas la même signification selon le risque de base.



Les mesures d'ampleur d'effet relatives et absolues se complètent, et sont nécessaires conjointement pour une interprétation globale des effets des facteurs.

### **Risque relatif** *relative risk, ou risk ratio*

Le risque relatif est un indicateur qui mesure l'association entre un facteur d'exposition et un événement binaire (survenue d'une maladie, décès, etc.).

Il correspond au ratio du risque de développer un événement étudié durant une période donnée (par exemple une maladie) dans un groupe de personnes présentant l'exposition étudiée ( $R_{\text{exposés}}$ ), divisé par le risque de développer l'événement dans un groupe de personnes ne présentant pas l'exposition ( $R_{\text{non-exposés}}$ ).

Le risque relatif peut s'interpréter de la façon suivante : les personnes présentant l'exposition étudiée ont « RR » fois plus de risque de présenter l'événement que les personnes non exposées.

$RR > 1$  : l'exposition est associée à une augmentation du risque de survenue de l'événement (facteur de risque) ;  $RR < 1$  : l'exposition est associée à une diminution du risque de survenue de l'événement (facteur protecteur).

La valeur du RR doit s'interpréter avec son intervalle de confiance. Dans les analyses univariées, le test du Chi 2 permet de tester si le risque relatif est statistiquement différent de 1.

### Calculs

|                      |                      |                      |       |
|----------------------|----------------------|----------------------|-------|
|                      | <b>M<sup>+</sup></b> | <b>M<sup>-</sup></b> |       |
| <b>E<sup>+</sup></b> | a                    | b                    | a + b |
| <b>E<sup>-</sup></b> | c                    | d                    | c + d |
|                      | a + c                | b + d                | N     |

**E<sup>+</sup>** : exposés    **E<sup>-</sup>** : non exposés    **M<sup>+</sup>** : malades    **M<sup>-</sup>** : non malades    N total

$$RR = \frac{R_{\text{exposés}}}{R_{\text{non-exposés}}} = \frac{a/(a+b)}{c/(c+d)}$$

### **Méta-analyse** *meta-analysis*

Pour répondre à une question donnée, par exemple le bénéfice d'un traitement expérimental par rapport à un traitement de référence, on peut combiner les résultats d'une série d'études indépendantes portant sur ces traitements, selon une approche reproductible.

La méta-analyse permet d'améliorer la précision de l'estimation de la taille d'effet d'une intervention, pouvant ainsi lever le doute en cas de résultats d'études discordants ou encore augmenter la puissance statistique en augmentant le nombre d'observations. Elle est considérée comme le plus haut niveau de preuve actuellement dans le paradigme de l'EBM.

On ne peut réunir que des études homogènes en design (critère de jugement, population d'étude ...). L'hétérogénéité est donc un phénomène important à évaluer. Une forte hétérogénéité indique un effet traitement inconstant à travers les études réunies dans la méta-analyse. Le  $I^2$  est une des mesures de cette hétérogénéité. Cela doit questionner la pertinence d'un tel regroupement et nécessite des analyses supplémentaires (en sous-groupe, méta-régression).

Enfin comme la méta-analyse se base uniquement sur les études retrouvées, elle est sujette à un biais de publication (les essais positifs sont plus publiés que les essais négatifs). Les résultats des méta-analyses sont souvent représentés sous forme de forest-plot.

### **Modèle**

Représentation simplifiée d'un phénomène ou d'un processus, souvent par un outil mathématique ou informatique, dans un but explicatif ou prédictif.

### **Régression linéaire** *linear regression*

Établissement d'une relation linéaire entre une variable à expliquer Y quantitative (e.g. poids, pression artérielle, valeur de biomarqueur) et une ou plusieurs variables explicatives (X, Z ...) de la forme :

$$Y = \text{constante} + \alpha \times X + \beta \times Z + \dots + \text{résidu}$$

Les coefficients  $\alpha$  et  $\beta$  quantifient l'ampleur de l'association entre les variables X et Z et la variable Y.

La constante correspond à la valeur prédite par le modèle lorsque X et Z sont nulles. Le terme résidu correspond à l'écart entre la valeur mesurée (Y) et ce qui est prédit par le modèle ( $\text{constante} + \alpha \times X + \beta \times Z$ ).

### **Régression linéaire simple (univariée)**

On parle de régression linéaire simple (ou univariée) en présence d'une seule variable explicative :

$$Y = \text{constante} + \alpha \times X + \text{résidu}$$

$\alpha$  mesure la variation moyenne de Y (positive ou négative) lorsque X augmente d'une unité (cela correspond à la pente de la droite de régression linéaire). La constante (appelée aussi ordonnée à l'origine ou intercept) correspond à la valeur prédite de Y lorsque X vaut 0.



Relation entre % en acides gras et pression artérielle systolique (PAS) chez les femmes

Tableau dans l'article

| Factor         | Coefficient [95 % CI]   | p-value |
|----------------|-------------------------|---------|
| intercept      | 110.01 [107.09, 113.12] | < 0.001 |
| fat percentage | 0.354 [ 0.212, 0.432]   | < 0.001 |

Equation à déduire

$$PAS = 110,01 + 0,354 \times \% \text{ acides gras} + \text{résidu}$$

Interprétation

Pour une patiente avec un pourcentage en acides gras nul, la pression systolique moyenne prédite serait de **110,01 mmHg** (cette information a peu d'intérêt sur le plan clinique, car il n'est pas réaliste d'imaginer un individu avec un pourcentage en acides gras nul). Chaque augmentation d'une unité du pourcentage en acide gras est associée à une augmentation moyenne de la pression systolique de **0,354 mmHg**. Le coefficient de pente est statistiquement différent de 0 ( $p < 0,001$ ) : l'association est statistiquement significative.

### Régression linéaire multiple (ajustée)

On parle de régression linéaire multiple (ou ajustée) en présence de plusieurs variables explicatives :

$$Y = \text{constante} + \alpha \times X + \beta \times Z + \text{résidu}$$

$\alpha$  mesure la variation moyenne de  $Y$  (positive ou négative) lorsque  $X$  augmente d'une unité, à *niveau comparable* de la variable  $Z$ . Les analyses multivariées permettent de réduire le risque de confusion dans l'évaluation de l'association entre deux variables, car elle évalue l'association à niveau comparable de  $Z$  (analyse ajustée).



Relation entre % en acides gras et pression artérielle systolique (PAS) ajustée sur l'âge chez les femmes

Tableau dans l'article

| Factor         | Coefficient [95 % CI]   | p-value |
|----------------|-------------------------|---------|
| intercept      | 110.13 [107.09, 113.12] | < 0.001 |
| fat percentage | 0.332 [ 0.187, 0.498]   | < 0.001 |
| age (year)     | 0.365 [ 0.201, 0.452]   | < 0.001 |

Equation à déduire

$$PAS = 110,13 + 0,332 \times \% \text{ acides gras} + 0,365 \times \text{âge} + \text{résidu}$$

Interprétation

Pour une patiente avec un pourcentage en acides gras nul et d'âge nul, la pression systolique moyenne prédite serait de **110,13 mmHg**. A âge identique (comparable), chaque augmentation d'une unité du pourcentage en acide gras est associée à une augmentation moyenne de pression systolique de **0,332 mmHg** : entre deux patientes du même âge mais avec un écart d'une unité de pourcentage en acide gras, la différence de PAS attendue est de 0,332 mmHg. Le coefficient de pente est statistiquement différent de 0 ( $p < 0,001$ ) : l'association ajustée sur l'âge est statistiquement significative. Lorsque l'association reste

significative après ajustement sur d'autres variables, on dit que le facteur étudié est **indépendamment** associé au critère de jugement (non lié à un facteur de confusion).

N.B. : de façon symétrique, à pourcentage en acides gras identique (comparable), une augmentation d'un an est associée à une augmentation moyenne de pression systolique de 0,365 mmHg.

### Régression logistique

logistic regression

Modèle permettant d'évaluer la relation entre l'odds de survenue d'un événement  $Y$  (par exemple le décès) et une ou plusieurs variables explicatives ( $X, Z, \dots$ ), de la forme :

$$\text{Odds} = \text{Odds}_{\text{référence}} \times (\text{OR}_X)^x \times (\text{OR}_Z)^z$$

### Régression logistique simple (univariée)

On parle de régression logistique simple (ou univariée) en présence d'une seule variable explicative :

$$\text{Odds} = \text{Odds}_{\text{référence}} \times (\text{OR}_X)^x$$

- Si  $X$  est une variable catégorielle, par exemple à deux modalités (sexe homme/femme), une de ces modalités est définie en référence (par exemple homme) ; la variable  $X$  vaut alors 0 chez les hommes et 1 chez les femmes. Dans ce cas,  $\text{Odds}_{\text{référence}}$  correspond à l'odds de l'événement chez les hommes, et  $\text{OR}_X$  correspond à l'odds ratio des femmes par rapport aux hommes.

- Si  $X$  est une variable quantitative (par exemple l'âge),  $\text{OR}_{\text{age}}$  mesure par combien l'odds est multiplié lorsque l'âge augmente d'une unité.



Relation entre cirrhose et récidence après résection d'un polype gastrique par endoscopie

Tableau dans l'article

| Factor    | Coefficient [95 % CI] | p-value |
|-----------|-----------------------|---------|
| cirrhosis | 4.98 [ 1.89, 13.16]   | < 0.001 |

Forme 1

| Factor        | Coefficient [95 % CI] | p-value |
|---------------|-----------------------|---------|
| cirrhosis no  | -                     | < 0.001 |
| cirrhosis yes | 4.98 [ 1.89, 13.16]   | < 0.001 |

Forme 2

Les deux formes de tableaux précédentes peuvent se retrouver dans un article ; l'avantage de la seconde est de pouvoir identifier facilement la catégorie de référence pour le calcul de l'odds ratio.

Equation à déduire

$$PAS = \text{Odds}_{\text{référence}} \times (4,98)^{\text{cirrhose}}$$

Interprétation

Un patient avec cirrhose a un odds de récidence **4,98** fois plus élevé qu'un patient sans cirrhose (facteur aggravant). L'association entre la cirrhose et la récidence du polype est statistiquement significative ( $p < 0,001$ ) ; ceci se déduit également du fait que l'intervalle de confiance exclut la valeur 1. N.B. : l'odds de référence est rarement présenté dans les tableaux.

## Régression logistique multiple (ajustée)

On parle de régression logistique multiple (ou ajustée) en présence de plusieurs variables explicatives :

$$Odds = Odds_{référence} \times (OR_X)^X \times (OR_Z)^Z$$

$OR_X$  mesure l'effet de la variable X sur l'odds d'événement Y pour des individus comparables en termes de valeurs de la variable Z.

Si X correspond à l'âge et Z à l'IMC,  $OR_{age}$  est l'odds ratio lié à l'âge à niveau comparable d'IMC. On parle également d'odds ratio lié à l'âge, ajusté sur l'IMC.

Les analyses multivariées permettent de réduire le risque de confusion dans l'évaluation de l'association entre deux variables, car elle évalue l'association à niveau comparable de Z (analyse ajustée).



Relation entre cirrhose et récurrence après résection d'un polype gastrique par endoscopie, ajustée sur l'âge, le sexe, et la présence d'une autre maladie du foie.

Tableau dans l'article

| Factor        | Coefficient [95 % CI] | p-value |
|---------------|-----------------------|---------|
| cirrhosis     | 1.54 [ 0.33, 6.87]    | < 0.572 |
| age           |                       |         |
| ≤55/55-65     | 0.45 [ 0.13, 1.49]    | 0.124   |
| >65/55-65     | 0.38 [ 0.14, 0.97]    |         |
| male          | 1.51 [ 0.65, 3.51]    | 0.334   |
| liver disease | 2.94 [ 0.85, 11.84]   | 0.088   |

Forme 1

| Factor         | Coefficient [95 % CI] | p-value |
|----------------|-----------------------|---------|
| cirrhosis no   | -                     |         |
| cirrhosis yes  | 1.54 [ 0.33, 6.87]    | < 0.572 |
| age            |                       |         |
| 55-65          | -                     |         |
| ≤55            | 0.45 [ 0.13, 1.49]    | 0.124   |
| >65            | 0.38 [ 0.14, 0.97]    |         |
| female         | -                     |         |
| male           | 1.51 [ 0.65, 3.51]    | 0.334   |
| liver dis. no  | -                     |         |
| liver dis. yes | 2.94 [ 0.85, 11.84]   | 0.088   |

Forme 2

Les deux formes de tableaux précédentes peuvent se retrouver dans un article ; l'avantage de la seconde est de pouvoir identifier facilement la catégorie de référence pour le calcul de l'odds ratio.

### Équation à déduire

$$PAS = Odds_{référence} \times (1,54)^{cirrhose} \times (0,45)^{≤55} \times (0,38)^{>65} \times 1,51^{male} \times 2,94^{liver\ disease}$$

### Interprétation

A classe d'âge, sexe et présence d'autre maladie du foie comparables, un patient avec cirrhose a un odds de récurrence 1,54 fois plus élevé qu'un patient sans cirrhose (facteur aggravant). L'analyse ajustée montre une association non statistiquement significative entre cirrhose et récurrence du polype ( $p < 0,572$ ). Lorsque l'association n'est plus significative après ajustement sur d'autres variables, on ne peut pas dire que le facteur étudié est **indépendamment** associé au critère de jugement (l'association retrouvée en analyse univariée est liée à des facteurs de confusion).

## Modèle de Cox

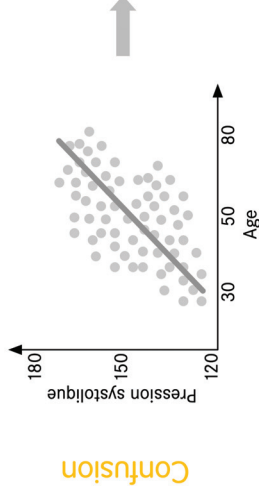
voir partie Survie

## Interaction

Il existe une interaction entre deux facteurs (par exemple X et Z) sur un critère de jugement lorsque l'effet du facteur X sur le critère de jugement n'est pas le même selon les différents niveaux du facteur Z, et vice-versa (on peut dire que Z vient modifier l'effet de X sur le critère de jugement).

D'un point de vue statistique, en cas d'interaction, le simple ajustement de l'effet de X sur Z conduit à une estimation erronée des effets de X et Z (ne pas confondre confusion et interaction). Il est nécessaire d'introduire dans le modèle l'effet du facteur X, celui du facteur Z, et leur interaction.

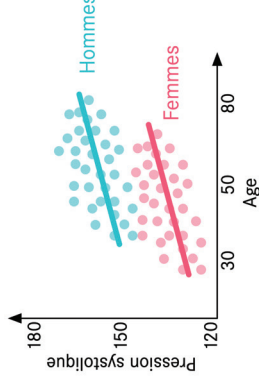
### Analyse globale



Confusion

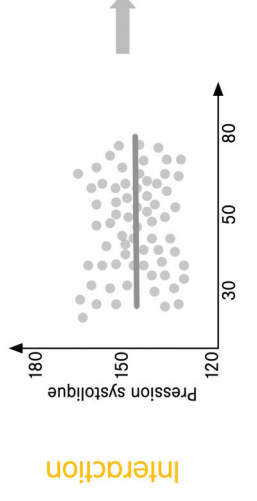
Augmentation forte de la PAS avec l'âge

### Analyse tenant compte du sexe



Effet de l'âge sur la PAS identique chez les hommes et les femmes, mais plus faible qu'en analyse globale

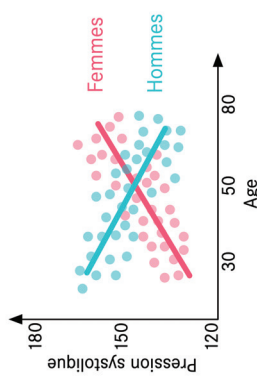
### Analyse globale



Interaction

Pas d'association entre âge et PAS

### Analyse tenant compte du sexe



♂ Diminution de la PAS avec l'âge  
♀ Augmentation de la PAS avec l'âge

Exemple artificiel concernant l'association entre l'âge et la pression systolique chez des patients hypertendus



Les effets de l'alcool et du tabac sur le risque de cancer ORL ne s'additionnent pas : l'effet de l'alcool sur le risque de cancer est beaucoup plus élevé chez les fumeurs que chez les non-fumeurs ; il existe donc une interaction entre les variables alcool et tabac.

Dans les essais cliniques <sup>Ⓢ</sup>, il est fréquent de rechercher des interactions entre le bras de traitement et d'autres variables, afin de savoir si le bénéfice du traitement expérimental par rapport au traitement de référence varie selon des sous-groupes <sup>Ⓢ</sup>.

**Monocentrique** <sup>Ⓢ</sup>  
voir partie Essai clinique <sup>Ⓢ</sup>

**Multicentrique** <sup>Ⓢ</sup>  
Voir partie Essai clinique <sup>Ⓢ</sup>

**Multivariée (analyse)** <sup>Ⓢ</sup>  
voir partie Analyse <sup>Ⓢ</sup>

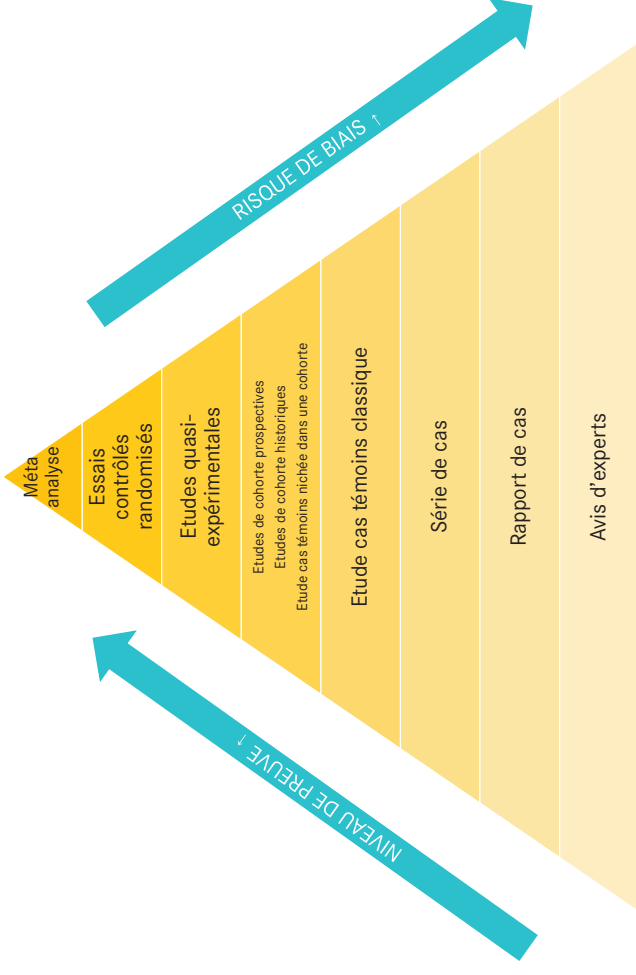


**Niveau de preuve**  
*level of evidence / hierarchy of evidence*

Le niveau de preuve d'une étude caractérise la capacité de l'étude à répondre à la question posée.

La capacité d'une étude à répondre à la question posée est jugée sur :

- l'adéquation du design d'étude à la question posée ;
  - l'existence ou non de biais <sup>Ⓢ</sup> importants dans la réalisation ;
  - l'adaptation de l'analyse statistique aux objectifs de l'étude ;
  - la puissance <sup>Ⓢ</sup> de l'étude en lien avec la taille de l'échantillon <sup>Ⓢ</sup>.
- Selon le domaine exploré (diagnostic, pronostic, dépistage, traitement, etc.), un fort niveau de preuve peut être donné par des études dont le design sera différent.



Le classement proposé ci-dessus dépend de la qualité de réalisation des études. En cas de biais dans une étude, le niveau de preuve est moindre que ce qui est dit attendu par le design. Par exemple, une méta-analyse mal réalisée (mauvaise sélection des articles, populations très différentes, ...) a un niveau de preuve plus faible que l'essai clinique seul bien mené.

**Nombre de sujets nécessaires**  
*sample size*




**Etude comparative**

Nombre de sujets défini pour une étude qui permettra de mettre en évidence une différence minimale escomptée entre deux groupes (par exemple une différence de proportions) avec une puissance définie. Ce nombre est calculé a priori ; il est lié à la différence espérée entre les deux groupes, à la puissance <sup>Ⓢ</sup> souhaitée, et au risque alpha <sup>Ⓢ</sup> fixé (ainsi qu'à des paramètres de nuisance supplémentaires, tels que l'écart-type pour une comparaison de moyennes, la proportion dans le groupe de référence pour une comparaison de proportions). L'effectif est d'autant plus élevé que la différence espérée est faible.

La différence espérée est une valeur hypothétique, qui doit être en adéquation avec les données actuelles de la science sur la maladie, et des traitements disponibles dans le cadre des essais thérapeutiques.

Attention, en cas de différence statistiquement significative, la différence à retenir est celle estimée dans le cadre de l'étude, et non la différence espérée pour le calcul d'effectif, qui n'était qu'une valeur hypothétique.

Dans une étude, le nombre de patients inclus tient compte non seulement du nombre de sujets nécessaires, mais en plus des éventuels perdus de vue. Il est donc généralement supérieur au nombre de sujets nécessaires.

Dans un essai clinique, la réalisation d'analyses intermédiaires  nécessite d'augmenter le nombre de sujets nécessaires afin de maintenir la puissance  de l'étude tout en tenant compte de la multiplicité des tests .

### Etude non comparative

Dans une étude non comparative, il est également important de définir a priori le nombre de sujets nécessaires. Celui-ci est basé sur la précision espérée de l'estimation du paramètre d'intérêt, la précision faisant référence ici à la demi largeur de l'intervalle de confiance à 95 %.

Par exemple, dans une étude estimant la prévalence d'une pathologie, si la prévalence est de 30 %, l'inclusion de 320 patients permettra d'obtenir une précision dans l'estimation de l'intervalle de confiance à 95 % de  $\pm 0,05$  (ceci signifie que, si la prévalence est bien estimée dans l'étude à 30 %, l'intervalle de confiance à 95 % sera [25 %, 35 %]). Une étude dans laquelle le paramètre d'intérêt est estimé avec un intervalle de confiance très large (i.e. avec une faible précision) est peu informative.

Dans le cadre de l'estimation d'une proportion, la précision est fonction de la valeur de la proportion ; dans le cadre de l'estimation d'une moyenne, la précision est fonction de l'écart-type des valeurs mesurées.

### Nombre de sujets à traiter (pour éviter un événement)

*number needed to treat*

Le nombre de sujets à traiter (NNT en anglais) est une mesure utilisée pour quantifier le bénéfice d'un traitement expérimental par rapport à un traitement de référence, lorsque le critère de jugement est binaire. Il correspond au nombre moyen de patients à traiter avec le traitement expérimental à partir duquel un événement (défavorable) de plus est évité par rapport au nombre d'événements observés avec le traitement de référence (pour ce même nombre de patients). Plus le NNT est faible, plus l'écart d'efficacité entre les deux traitements est élevé.

Le NNT est calculé à partir de l'inverse de la valeur absolue de la différence de risque d'événement entre les deux traitements.



Il existe un équivalent pour la survenue d'un événement indésirable (au lieu d'un événement) : NNH (number needed to harm). Il s'agit du nombre de sujets traités avec le traitement expérimental qui entraînera un événement indésirable de plus. Il est calculé par l'inverse de la différence absolue de risque d'événement indésirable entre les deux groupes.



### Objectif / Question de recherche

*objective/aim*

Le rationnel de l'étude permet de décrire l'état des connaissances actuelles, synthétisé dans l'introduction de l'article. Il permet de comprendre ce qui est connu, ce qui n'est pas connu, et la pertinence de la question de recherche principale à laquelle va permettre de répondre l'étude. Il permet de justifier l'intérêt de mener l'étude.


La question de recherche doit être posée avant la mise en œuvre de l'étude. Elle permet de formuler les hypothèses de la recherche  (sur lesquelles sont basés l'estimation de la puissance et le calcul du nombre de sujets nécessaires ). On parle de démarche hypothético-déductive.

L'objectif d'une recherche correspond à la formulation de la question à laquelle l'étude va répondre. Il doit être cliniquement pertinent et susceptible d'améliorer les connaissances et/ou de modifier la prise en charge. Il doit comporter les éléments suivants (PICOT) :

- Patient : population/pathologie concernée
- Intervention/Facteur : que cherche-t-on à évaluer/étudier ?
- Comparaison : à quoi/qui doit être éventuellement comparée l'intervention ou le facteur d'exposition ? Quelle est la méthode/groupe de référence / alternative ?
- Outcome : quel est le résultat/événement de santé mesuré ?
- Temps : à quel moment l'outcome est mesuré ?

### Observance / adhésion (thérapeutique)

*adherence*

Capacité des patients à prendre leur traitement conformément à la prescription (ou au protocole  dans le cadre d'un essai clinique). Elle comprend : l'initiation (le fait de commencer à prendre le traitement dans les temps), l'exécution (le fait de respecter le nombre de prises, le moment de prise et la dose tout au long du traitement), et la persistance (absence d'arrêt précoce du traitement).

Dans les essais cliniques, l'observance peut être évaluée en utilisant des piluliers électroniques, en demandant au patient de ramener les boîtes vides, ou par des mesures de marqueurs biologiques influencés par le médicament, dans le sang ou les urines.

**P**

**Perdu de vue**

*lost to follow-up*

Patient qui n'est pas suivi sur la totalité de la période prévue par le protocole ☹️ d'un essai ou d'une étude épidémiologique, ce qui empêche l'évaluation du critère de jugement ☹️. On ne sait pas si le patient a guéri, s'il a eu une complication ou des effets secondaires, et pourquoi il n'est pas revenu. Si la fréquence des perdus de vue n'est pas similaire entre les groupes, cela peut induire un biais d'attrition ☹️ dans les essais cliniques, ou un biais de sélection ☹️ dans les études épidémiologiques.

**Performances diagnostiques**

|                | M <sup>+</sup>    | M <sup>-</sup>    |                   |
|----------------|-------------------|-------------------|-------------------|
| T <sup>+</sup> | VP                | FP                | Nb T <sup>+</sup> |
| T <sup>-</sup> | FN                | VN                | Nb T <sup>-</sup> |
|                | Nb M <sup>+</sup> | Nb M <sup>-</sup> | N                 |

VP : vrais positifs    VN : vrais négatifs    FP : faux positifs    FN : faux négatifs

**Sensibilité**

*sensitivity*

Probabilité que le test diagnostic soit positif (T<sup>+</sup>) chez un malade (capacité du test à identifier les patients malades).

La sensibilité peut varier d'une population à l'autre en fonction du niveau de sévérité de la maladie dans la population (biais de spectre ☹️ par exemple).

**Calculs**

$$\text{Sensibilité} = \frac{VP}{VP + FN} = \frac{VP}{Nb M^+}$$

**Spécificité**

*specificity*

Probabilité que le test diagnostic soit négatif (T<sup>-</sup>) chez un non malade (capacité du test à classer correctement les patients non malades).

La spécificité peut varier d'une population à l'autre en fonction des caractéristiques de la population, par exemple l'âge pour le test des PSA (biais de spectre ☹️).

**Calculs**

$$\text{Spécificité} = \frac{VN}{VN + FP} = \frac{VN}{Nb M^-}$$

**Courbe ROC**

*ROC curve*

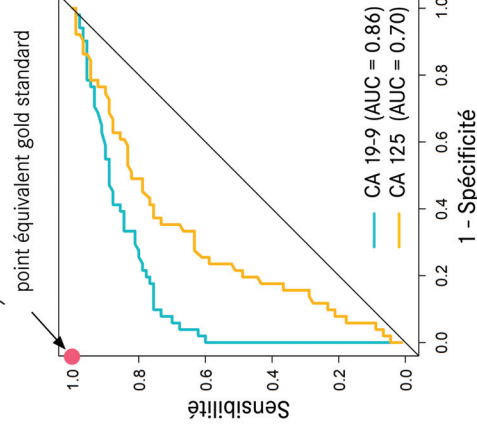
Une courbe ROC permet d'évaluer les performances diagnostiques globales d'un test à valeurs quantitatives ou ordinales, sur l'ensemble des seuils envisageables pour dichotomiser le test et l'utiliser pour la prise de décision médicale.

Il s'agit du tracé de la sensibilité en fonction du complément de la spécificité (1 - spécificité, i.e. le pourcentage de faux positifs) sur l'ensemble des seuils possibles pour le test.

Un test sans valeur diagnostique est un test dont la courbe ROC se situe sur la première bissectrice du graphique. Plus la courbe ROC se rapproche du point en haut à gauche du graphique (sensibilité = 100 %, spécificité = 100 %), meilleur est le test.

Les performances du test sont résumées globalement par l'aire sous la courbe ROC (AUC), comprise entre 0 et 1 ; AUC = 1 : test parfait ; AUC = 0.5 : test sans valeur diagnostique (qui ne fait pas mieux que le simple hasard). Attention, l'AUC s'interprète avec son intervalle de confiance ☹️. Lorsque l'intervalle de confiance exclut la valeur 0.5, le test a des performances significativement meilleures que le simple hasard.

Deux tests quantitatifs ou à valeurs ordinales peuvent être comparés au travers de leur aire sous la courbe ROC, sous réserve que les courbes ROC soient emboîtées (pas de croisement fort entre les courbes des deux tests).



### Ratio de vraisemblance négatif

*negative likelihood ratio*

Utilisé dans les études d'évaluation de performances de tests diagnostiques. Mesure par combien l'odds de maladie pré-test diagnostique est multiplié lorsque le résultat du test est négatif. Ce ratio est compris entre 0 et 1 ; le test est d'autant plus informatif que le ratio de vraisemblance négatif est faible.

Calcul

$$RV^- = \frac{VN}{VN+FN} = \frac{VN}{Nb T^-}$$

### Ratio de vraisemblance positif

*positive likelihood ratio*

Utilisé dans les études d'évaluation de performances de tests diagnostiques. Mesure par combien l'odds de maladie pré-test diagnostique est multiplié lorsque le résultat du test est positif. Ce ratio est compris entre 0 et 1 ; le test est d'autant plus informatif que le ratio de vraisemblance positif est élevé.

Calcul

$$RV^+ = \frac{VP}{VP+FP} = \frac{VP}{Nb T^+}$$

### Probabilité de maladie pré-test

*pre-test probability*

Évaluation de l'intérêt du test : le test diagnostique a-t-il des résultats différents entre des patients et des sujets sains ?

### Phases d'évaluation de méthodes diagnostiques

*positive likelihood ratio*

Utilisé dans les études d'évaluation de performances de tests diagnostiques.

Mesure par combien l'odds de maladie pré-test diagnostique est multiplié lorsque le résultat du test est positif. Ce ratio est supérieur à 1 ; le test est d'autant plus informatif que le ratio de vraisemblance positif est élevé.

Calcul

$$RV^+ = \frac{\text{Sensibilité}}{1 - \text{Spécificité}}$$

### Ratio de vraisemblance négatif

*negative likelihood ratio*

Utilisé dans les études d'évaluation de performances de tests diagnostiques. Mesure par combien l'odds de maladie pré-test diagnostique est multiplié lorsque le résultat du test est négatif. Ce ratio est compris entre 0 et 1 ; le test est d'autant plus informatif que le ratio de vraisemblance négatif est faible.

Calcul

$$RV^- = \frac{1 - \text{Sensibilité}}{\text{Spécificité}}$$

### Rapport de vraisemblance

Apport du test (gain diagnostique)

| Positif    | Négatif      |
|------------|--------------|
| 10 et plus | 0.1 et moins |
| 5 - 10     | 0.1 - 0.2    |
| 1 - 5      | 0.2 - 1      |
| 1          | 1            |
|            | Aucun        |

### Phases d'évaluation de méthodes diagnostiques

Phase 1

Évaluation de l'intérêt du test : le test diagnostique a-t-il des résultats différents entre des patients et des sujets sains ?

Phase 2


Évaluation du test dans des sous-groupes : le test détecte-t-il les différentes formes ou la sévérité de la maladie (par exemple selon le stade de la maladie) ? Le test arrive-t-il à faire le diagnostic différentiel ? Par exemple, pour le VIH, le test arrive-t-il à distinguer des patients infectés au VIH de patients ayant d'autres maladies induisant une immunosuppression.

Les études de phase 1 et 2 sont réalisées sur des petits groupes d'individus, dans le cadre d'études cas-témoins de type diagnostique.

Phase 3

Évaluation des performances diagnostiques du test dans une étude intégrant tous les patients qui seraient amenés à bénéficier du test, afin d'observer toute la diversité de la maladie et des diagnostics différentiels, et d'éviter le biais de spectre. Il s'agit d'une étude transversale observationnelle, au sens où le nouveau test n'est pas utilisé pour modifier la prise en charge du patient.

Pour s'assurer de la représentativité de l'échantillon, dans chaque centre, les patients doivent être inclus de façon consécutive (évite le biais de sélection). Ils doivent pouvoir subir


l'examen à valider et l'examen de référence, « l'étalon-or » (ou gold standard). Le test doit en outre être placé dans la stratégie diagnostique habituelle (sans être inclus dans la prise en charge, pour éviter le biais d'incorporation ) , et chaque patient faire les deux examens. Les résultats doivent être interprétés en insu sans que le médecin ait connaissance des résultats de l'autre examen.

#### Phase 4

Essai clinique randomisé (un groupe avec utilisation du test, l'autre sans utilisation du test) mesurant l'amélioration de l'état de santé grâce au recours au test.

### PROMs/PREMs

*patient-reported outcomes measures / patient-reported experience measures*

Les PROMs et PREMs contribuent à une approche des soins centrée sur le patient en intégrant son point de vue dans l'évaluation des résultats et de la qualité des soins. Ces mesures, au travers de questionnaires , sont souvent utilisées dans la recherche, la pratique clinique et les démarches d'amélioration de la qualité.

### PROMs

*patient-reported outcomes measures*

Évaluation de l'état de santé d'un patient qui provient directement du patient lui-même. Les PROMs mesurent la perception qu'a le patient de sa propre santé concernant les symptômes perçus, la qualité de vie et le bien-être général. Les PROMs sont utilisés pour recueillir des informations sur l'impact d'une maladie et de son traitement sur la vie du patient. Les PROMs sont des indicateurs pour évaluer l'efficacité d'interventions médicales du point de vue du patient : médicaments, stratégies thérapeutiques, suivi et accompagnement. Les questionnaires utilisés doivent être validés pour que les résultats puissent être interprétables et comparables.

### PREMs

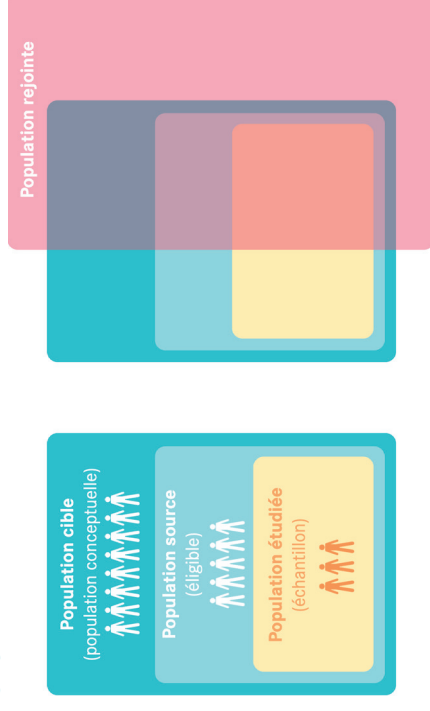
*patient-reported experience measures*

Évaluation de l'expérience subjective du patient avec les services de santé. Les PREMs mesurent les perceptions et expériences du patient concernant des dimensions des soins telles que la relation et la communication avec les professionnels, les délais, le soutien émotionnel et social, la prise en compte de l'entourage et l'expérience globale des soins. Les PREMs sont des indicateurs de la qualité des soins du point de vue du patient, pour identifier les domaines d'amélioration. Les questionnaires utilisés doivent être validés pour que les résultats puissent être interprétables et comparables.

### A

Enquêtes interrogeant les patients sur leur expérience concernant la prise de rendez-vous, la clarté des informations fournies etc. En France par exemple l'enquête nationale e-satis est envoyée à tous les patients suite à leur hospitalisation.

## Populations



### Population cible

*target population*

Population théoriquement concernée par la question posée dans l'étude, à qui on voudrait pouvoir extrapoler les résultats (population conceptuelle).

Exemple en épidémiologie : dans une étude sur les facteurs de risque de fracture ostéoporotique, la population cible pourrait correspondre aux femmes ménopausées.

Exemple en essai clinique : population pour laquelle le traitement sera recommandé à l'issue de l'étude, incluant potentiellement des patients non étudiés dans l'essai (femmes enceintes, présence de certaines comorbidités, ...).

### Population rejointe

La population rejointe correspond à la population traitée par le médicament à l'issue d'une étude. Par exemple un médicament validé pour le traitement d'une maladie est utilisé pour d'autres indications, sans étude préalable. A l'inverse, certains patients de la population cible de la thérapeutique peuvent ne pas être traités.

Par exemple, le sémaglutide validé dans le traitement du diabète de type II est actuellement prescrit pour des personnes souhaitant perdre du poids mais ne présentant pas de diabète.



### Population source/éligible *source population/eligible population*

Population à laquelle l'étude sera proposée : correspond à un sous-ensemble de la population cible, en raison de critères d'inclusion et de non inclusion.  
Exemple : pour l'étude sur les facteurs de risque de fracture ostéoporotique, exclusion des femmes avec un cancer, une SEP ...

Si les critères d'inclusion/non inclusion sont trop restrictifs, les résultats issus d'une population source ne pourront pas être généralisés à la population cible  
Exemple : dans une étude sur l'estimation de la prévalence du VIH chez les moins de 35 ans (population cible), une étude réalisée en prison (population source) conduira à des résultats non généralisables à la population cible. Cette étude en prison est uniquement généralisable à la population carcérale de moins de 35 ans.

### Échantillon d'étude/population étudiée *sample*

Partie de la population source sur laquelle porte l'étude. L'échantillon est représentatif pour une caractéristique mesurée, si cette caractéristique se distribue identiquement dans l'échantillon et dans la population source.

Les techniques utilisables pour obtenir un échantillon représentatif de la population source sont :

- le tirage au sort dans la population source
  - l'utilisation de la file active de patients (patients consécutifs consultant pour un problème de santé durant une période définie dans les centres participant à l'étude)
- sous réserve d'un échantillon de taille conséquente, car les caractéristiques des patients de l'échantillon convergent vers celles de la population source.

Pour la technique du tirage au sort, la difficulté est de disposer de registres ou de bases de données adaptées à la population d'intérêt.

L'inférence est le processus statistique permettant de généraliser les résultats d'un échantillon représentatif à la population source dont il est issu.

### Échantillonnage stratifié *stratified sampling*

Lors d'échantillonnage stratifié, un échantillon est créé à partir d'une population source, selon deux étapes :

#### 1. Création des strates

Découper la population en sous-ensembles homogènes pour une caractéristique donnée, ces sous-ensembles sont dénommés strates.

Exemples de strates : création de strates sur l'âge (strates définies sur des tranches d'âge), strates selon le sexe, la zone d'habitation ou encore sur des critères de santé (sévérité de la maladie, consommation de tabac).

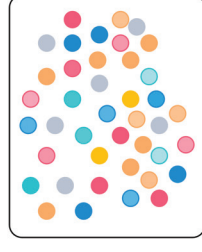
Il est possible de croiser plusieurs strates (âge et sexe par exemple).

Les facteurs de stratification sont choisis selon leur influence possible sur les critères étudiés et représentent les variables sur lesquelles on souhaite avoir une bonne représentativité de l'échantillon constitué.

Sur le schéma ci-dessous, les strates sont définies par des combinaisons de classes d'âge et de sexe :

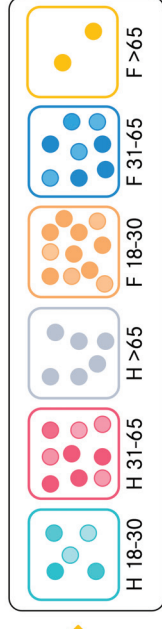
#### Echantillonnage non stratifié

Population totale

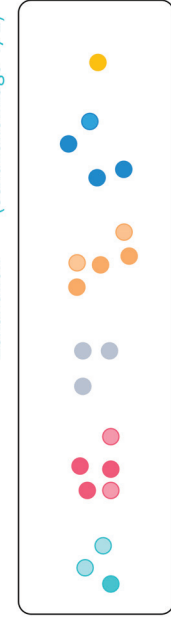


#### Echantillonnage stratifié

Population totale répartie en strates



Echantillon (échantillonnage 1/2)




Assure des caractéristiques identiques entre échantillon et population source sur les variables de stratification

Ne garantit pas des caractéristiques identiques entre échantillon et population source, risque de ne pas ou de sous-représenter les catégories rares

## 2. Constitution de l'échantillon


Les sujets sont échantillonnés par tirage au sort au sein de chaque strate. Les effectifs tirés au sort dans chaque strate sont calculés selon la proportion souhaitée de chaque strate dans l'échantillon constitué au final.

 (correspondant au schéma) : Sur une population de 3000 sujets composée de 1000 hommes et 2000 femmes adultes d'âge varié, on souhaite constituer un échantillon au 1/10<sup>ème</sup> de cette population (soit 300 sujets), représentatif pour le sexe et l'âge de la population source. Cela est possible en stratifiant sur le sexe et l'âge. Pour ce faire, on procède en plusieurs étapes :


- Création des strates sexe x âge : constitution des strates liées au sexe (une strate de 1000 hommes et une strate de 2000 femmes) puis des strates liées à l'âge (catégories 18-30, 31-65, >65) au sein de chaque strate liée au sexe. On obtient ainsi 6 strates représentant un des 2 sexes et une des catégories d'âge.
- Constitution de l'échantillon : tirage au sort 1/10<sup>ème</sup> dans chaque strate pour obtenir un échantillon constitué de 100 hommes et 200 femmes répartis selon les mêmes catégories d'âge que la population source (échantillon stratifié proportionnel).

## Prévalence





*prevalence*

Proportion d'individus présentant une caractéristique (ex : une maladie) à un moment donné, dans une population. La période durant laquelle la prévalence est estimée doit être précisée (ex : évaluation de la prévalence d'infections nosocomiales sur une journée ; évaluation de la prévalence du diabète durant un mois donné, ...). La prévalence s'estime à partir d'une étude transversale .

## Prospectif

voir partie Étude : temporalité et recueil .

## Protocole

Document qui définit précisément les objectifs, les moyens et les méthodes mis en œuvre dans le cadre d'une étude. Le protocole doit être rédigé avant le démarrage de l'étude, pour la planifier correctement, incluant : les critères de jugement  associés aux objectifs, le calcul de la taille d'échantillon , le plan d'analyse statistique, ainsi que les aspects réglementaires  et éthiques .

## Puissance statistique

voir partie Tests statistiques .

## Q

## Qualités métrologiques

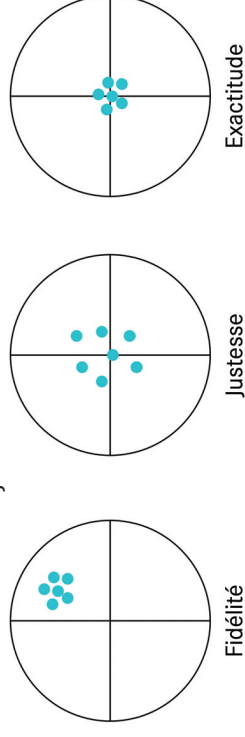
 Exactitude / Justesse / Fidélité

**Validité / Précision / Reproductibilité** 

Exactitude : combine la notion de justesse et de fidélité.


Une mesure est dite **fidèle** si les résultats obtenus lorsque la mesure est répétée sont proches les uns des autres (changent peu), sans forcément correspondre à la vraie valeur. Une mesure est **juste** si la moyenne des résultats obtenus en répétant la mesure correspond à la vraie valeur.

Une mesure est **exacte** si elle est juste et fidèle.



En recherche clinique ou épidémiologique, la notion de justesse se rapproche de la notion d'**absence de biais**, appelée aussi **validité** : en moyenne, on doit obtenir le bon résultat (même si d'une expérience à l'autre, on n'obtient pas exactement le bon résultat en raison des fluctuations aléatoires).

La notion de fidélité se rapproche de la notion de **précision**. La précision d'une estimation issue d'un ensemble de mesures est fonction de deux facteurs :

- la précision de l'outil utilisé pour réaliser les mesures individuelles, qui définit également la notion de reproductibilité  de la mesure, ou de concordance
  - o balance de Roberval vs balance au microgramme (ceci influence entre autres le nombre de décimales pertinent dans l'expression du résultat, fonction de l'erreur de mesure)
  - o fréquence des scanners suite à l'initiation d'une chimiothérapie pour évaluer le délai jusqu'à progression ; des scanners tous les mois conduiront à un délai jusqu'à progression plus précis qu'un scanner tous les 6 mois
- le nombre d'unités statistiques (par exemple le nombre de patients) contribuant à l'estimation (par exemple moyenne ou proportion) ; plus le nombre d'unités statistiques

(patients) est élevé, plus les fluctuations aléatoires se compensent au sein de l'échantillon, et donc moins les estimations seront variables d'un échantillon à l'autre.

Le terme d'**exactitude** se rapproche de la notion d'accuracy (performance pour les tests diagnostiques ).

| Étude cas-témoins, cohorte, essai clinique  |  | Étude diagnostique |
|---|--|--------------------|
| A l'échelle d'une étude, la reproductibilité définit la reproductibilité des résultats/conclusion entre les différentes études se posant la même question de recherche. | La reproductibilité correspond au fait que la valeur du test est identique/proche si la mesure est faite à nouveau. On parle aussi de fiabilité du test.<br>Exemples :<br>- plusieurs dosages d'un marqueur<br>- plusieurs évaluations d'une image<br>o reproductibilité intra-évaluateur<br>lorsque la même image est évaluée plusieurs fois par le même observateur<br>o reproductibilité inter-évaluateur<br>lorsque la même image est évaluée par plusieurs observateurs |                    |

## Questionnaire

Un questionnaire est un outils permettant l'interrogation du patient afin de recueillir différents items du ce dernier (exemple : PROMS et PREMS )


Ces questionnaires doivent être standardisés et validés pour que les résultats puissent être interprétables et comparables.



SF36 (qualité de vie générale du patient), EORTC-QLQ (qualité de vie spécifique en cancérologie), HAD (échelle d'anxiété et dépression).

Parmi les questionnaires, on distingue les auto-questionnaires et hétéro-questionnaires selon la personne qui pose les questions et recueille les réponses.

## Auto-questionnaire

Auto-évaluation = questionnaire autoadministré : les individus fournissent des informations les concernant, les personnes interrogées répondent elles-mêmes aux questions, reflétant ainsi leurs propres perceptions, expériences et opinions. Les PROMs  sont des auto-questionnaires.

Risques :

- pour les essais cliniques en ouvert, biais de mesure lié à la connaissance du groupe de traitement par la personne (privilégier auquel cas une évaluation par un hétéro-questionnaire avec un évaluateur en aveugle du groupe)
- perte de précision ou risque de biais si les questionnaires ne sont pas standardisés et validés

## Hétéroquestionnaires

Hétéro-évaluation = questionnaire hétéro-administré : un enquêteur ou une autre personne (famille, proche, soignant) administre le questionnaire et enregistre les réponses au nom du répondant. Les informations sont recueillies à partir d'une observation ou d'une interprétation externe.

Risque :

- Biais de mesure lié à l'enquêteur : influence de l'enquêteur sur les réponses obtenues, manière de formuler les questions, manière d'interpréter les réponses
- Biais de désirabilité sociale : tendance à se présenter de façon favorable devant d'autres individus en fonction de certaines normes. Ce biais peut survenir en cas de question sensible (comportements à risque, répréhensibles) : la présence de l'enquêteur peut influencer la réponse donnée par le patient.

Prévention de ce risque :





- Formation des enquêteurs, un seul enquêteur, questionnaire standardisé
- Enquêteurs en insu du groupe des patients
- Autoquestionnaire pour les sujets sensibles à risque de biais de désirabilité sociale

## R

## Randomisation

*randomization*

Répartition aléatoire des patients dans les différents bras d'intervention de l'étude.

La randomisation peut être stratifiée  selon un nombre restreint de facteurs pronostiques , afin de s'assurer que les différents niveaux des facteurs pronostiques soient répartis de façon équilibrée dans les différents bras de l'étude (ceci revient à faire une liste de randomisation par strate), et d'empêcher le biais de sélection/biais de confusion  dans les essais cliniques .

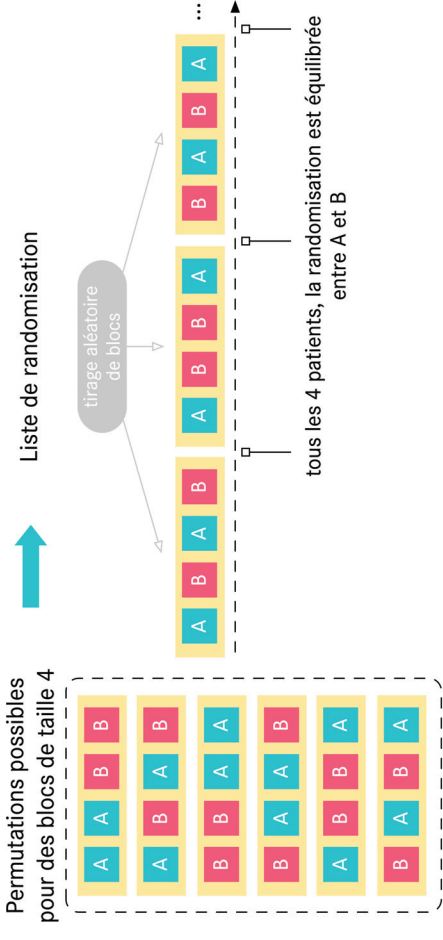
La randomisation peut être réalisée par des listes d'allocation aléatoire préétablies, ou de façon dynamique par minimisation (chaque nouveau patient sera inclus dans le bras tel que le déséquilibre des facteurs pronostiques dans les différents bras soit ainsi minimal).

La randomisation par bloc a pour but d'éviter un déséquilibre d'effectif entre les groupes en fin d'étude ; pour une randomisation par bloc de 4, tous les 4 patients inclus, le nombre de traitements expérimentaux et de référence est équilibré (elle est utilisée pour les listes d'allocation aléatoires pré-établies).

La randomisation doit être centralisée (centralized), sur un site internet (IWRS) ou via une centrale d'appel téléphonique (IVRS), pour permettre la traçabilité, et l'imprévisibilité (concealment). L'investigateur (le médecin qui randomise) ne doit pas pouvoir deviner a priori le résultat de la randomisation (clause d'ignorance).

La randomisation doit être réalisée au plus près de l'intervention, afin d'éviter des écarts au protocole (patient randomisé qui ne souhaite plus participer, dégradation de l'état général ne permettant plus l'intervention, ...). L'analyse en intention de traiter (ITT) permettra de limiter l'impact des écarts au protocole (biais d'attrition) dans les conclusions.

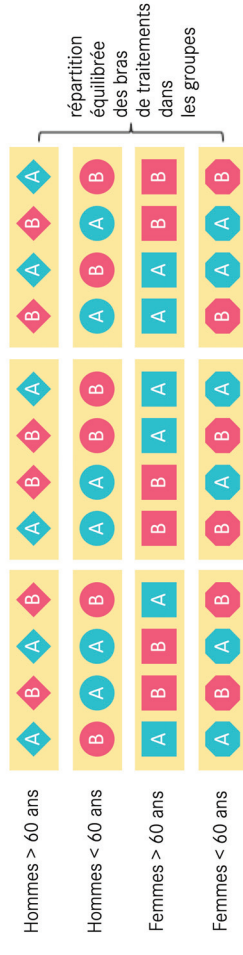
## Randomisation par blocs



## Randomisation stratifiée par blocs

exemple : stratification sur l'âge et le sexe

1 liste de randomisation différente par groupe de facteurs pronostiques



Dans certaines études, la randomisation n'est pas effectuée à l'échelle du patient car l'intervention étudiée n'est pas individuelle mais collective, mais à l'échelle d'ensembles de patients, appelés clusters (par exemple modification d'une prise en charge à l'échelle d'un hôpital ; mise en place d'un programme de prévention dans une école ; modification de pratiques de médecin généraliste, ...).

Les articles d'essais cliniques présentent les caractéristiques initiales/facteurs pronostiques des patients dans les deux groupes d'intervention ; (âge, sexe, poids, etc.). Un déséquilibre observé entre les deux groupes n'indique pas forcément un défaut de la randomisation (le hasard peut induire des déséquilibres). Il faut dans la lecture d'un article s'assurer de la

qualité de la méthode de randomisation (centralisée, imprévisible). En cas de déséquilibre important sur un facteur pronostique, l'article peut présenter dans une analyse de sensibilité des résultats ajustés sur ce facteur et le comparer aux résultats non ajustés. Si la différence est importante on peut douter de la robustesse des conclusions.

**Ratio de vraisemblance positif**  
voir partie Performances diagnostiques

**Ratio de vraisemblance négatif**  
voir partie Performances diagnostiques

**Régression linéaire**  
voir partie Modèles

**Régression logistique**  
voir partie Modèles

**Risque**  
voir partie Mesures d'événements

**Risque alpha (ou risque de première espèce)**  
voir partie Test statistique

**Risque bêta (ou risque de deuxième espèce)**  
voir partie Test statistique

**Rétrospectif**   
voir partie Étude : temporalité et recueil

**Rétrolectif**  
voir partie Étude : temporalité et recueil

**Risque relatif**  
voir partie Mesures d'événements

**Run-in (période)**  
La période run-in est la période qui se situe habituellement avant la randomisation pendant laquelle les participants reçoivent le placebo ou un traitement afin de mesurer l'adhésion aux

traitements (observance), d'exclure les placebo répondeurs, ou les patients intolérants aux traitements. Cette période run-in et la sélection des patients entraînent une modification de la population étudiée, et peut réduire la généralisabilité des résultats à la population cible, impactant ainsi la pertinence clinique de l'étude.

**S**

**Score de propension**  
*propensity score*

Le score de propension est principalement utilisé dans les études observationnelles dont la question de recherche est causale (exemple : efficacité et sécurité d'une intervention).

Le score de propension estime la probabilité qu'un individu ait reçu un traitement spécifique en fonction de ses caractéristiques observées. Il est calculé à partir de multiples variables (par exemple, l'âge, le sexe, l'état de santé initial, etc.) et vise à équilibrer les groupes traitement et contrôle en rendant les caractéristiques des sujets plus similaires. A ce titre, le score de propension vise à réduire le biais de confusion entre les groupes de comparaisons et le critère de jugement.

Une fois le score de propension estimé il peut être utilisé de différentes façons : par appariement, par pondération (*inverse probability treatment weighting IPTW*).

A l'issue de son utilisation il faut vérifier le bon équilibre entre les deux populations des facteurs de confusion pris en compte dans le score de propension, notamment à partir des standardized mean difference (SMD), qui doivent être inférieures à 0.1 (10%).

Le score de propension n'a pas le même effet que la randomisation car il permet de contrôler les facteurs de confusion uniquement connus, bien mesurés et correctement modélisés.

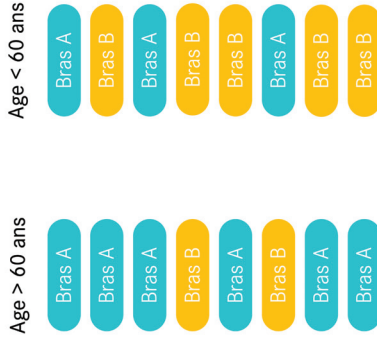
**Sensibilité**  
voir partie Performances diagnostiques

**Spécificité**  
voir partie Performances diagnostiques

**Stratification**  
**Stratification de la randomisation**

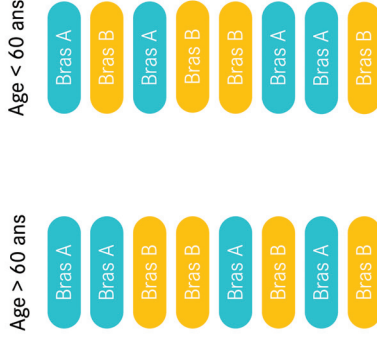
Dans un essai clinique, les patients peuvent être randomisés en tenant compte de strates (groupes de patients) définissant des facteurs pronostiques importants (par exemple la catégorie d'âge ou l'hôpital), afin d'avoir autant de placebo que de traitement expérimental dans les différentes strates. Ceci permet d'éviter un effet de confusion des variables prises en compte pour définir les strates dans l'analyse de l'efficacité du traitement.

### Randomisation non stratifiée sur la classe d'âge



Déséquilibre possible des bras de traitement selon les classes d'âge

### Randomisation stratifiée sur la classe d'âge



Répartition équilibrée des bras de traitement selon les classes d'âge

## Stratification en étude épidémiologique

La stratification en étude épidémiologique revient à faire des comparaisons selon un facteur d'exposition non pas au niveau global, mais par strate. Par exemple, pour évaluer l'association entre le sexe et la survenue d'un cancer du poumon, la survenue de cancer ne sera pas comparée globalement entre les hommes et les femmes, mais par strates selon le statut tabagique (comparer les hommes et les femmes à statut tabagique comparable). Ceci peut permettre de s'affranchir d'un effet de confusion lié aux facteurs pris en compte (tabac) dans les strates.

## Sujets éligibles

*eligible participants*

Un sujet est dit éligible dans le cadre d'une étude de recherche clinique, à partir du moment où l'ensemble de ses caractéristiques remplissent a priori, d'une part l'ensemble des critères d'inclusion, et aucun des critères de non-inclusion définis dans le protocole de l'étude.

## Survie

*survival*

La survie correspond au délai entre une date d'origine (exemple : date de diagnostic, date d'inclusion dans un essai clinique) et le décès du patient.

## Survie sans événement

La survie sans événement correspond au délai entre la date d'origine et la survenue d'une liste d'événements définis dans le critère de jugement, incluant potentiellement le décès. Par exemple, en cancérologie, la survie sans progression correspond au délai entre la date d'origine et le premier des événements entre la progression ou le décès (critère de jugement composite). Les patients vivants et sans progression sont censurés à la date de dernières nouvelles.

Attention, il faut être attentif à la définition exacte du critère de jugement : le décès peut être pris en compte ou non dans la survie sans événement.

## Analyse de survie

L'analyse de survie, au sens large, correspond à l'analyse du délai de survenue d'un événement (récidive, échec de traitement, progression, décès, ...) par rapport à une date d'origine (exemple : date de diagnostic, date d'inclusion dans un essai clinique).

L'intérêt des analyses de survie n'est pas uniquement d'étudier l'occurrence de l'événement, mais d'étudier le délai de survenue. Dans un essai clinique, deux traitements peuvent conduire au même nombre d'événements ; pour autant un des traitements peut être avantageux si le délai de survenue de l'événement est augmenté.

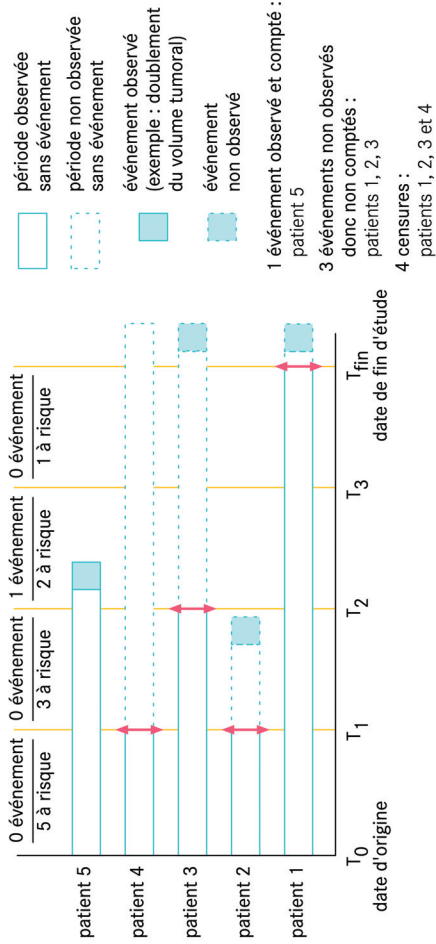
## Délai ou temps de participation, date de dernières nouvelles, date de point

Pour chaque patient, le délai retenu pour la survie (délai ou temps de participation) correspond :

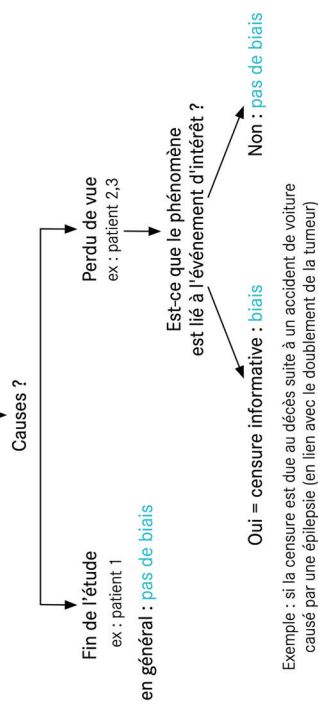
- au délai de survenue de l'événement par rapport à la date d'origine si le patient a développé l'événement au cours de l'étude
- au délai entre la date d'origine et la censure du suivi si le patient n'a pas développé l'événement au cours de l'étude

Le suivi d'un patient peut être censuré pour trois raisons :

- patient perdu de vue sans avoir développé auparavant l'événement
  - patient décédé sans avoir développé auparavant l'événement (lorsque l'événement étudié n'est pas le décès)
  - patient n'ayant pas développé l'événement à la date de fin d'étude (date de point)
- En cas de censure, la seule information disponible est que le patient n'avait pas encore développé l'événement à cette date.

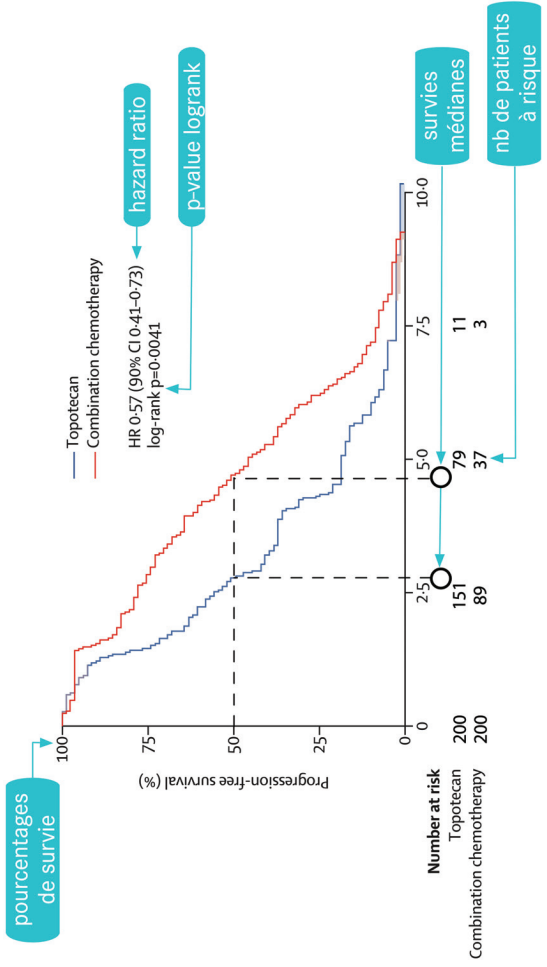


Censure : arrêt de l'observation et absence d'événement observé



### Courbe de survie survival curve

La survie dans un groupe est décrite par une courbe de survie, indiquant la probabilité de survie des patients à l'événement étudié en fonction du temps écoulé depuis l'origine. On indique en général en dessous des courbes de survie les nombres de patients encore à risque (n'ayant pas encore développé l'événement et toujours suivis dans l'étude) aux différents temps d'évaluation. La survie est de 1 au temps initial, puis diminue plus ou moins rapidement. La courbe de survie est souvent résumée par les probabilités de survie à différents délais, ou par la médiane de survie.



### Probabilité de survie

La probabilité de survie à un délai donné correspond à la probabilité qu'un individu n'ait pas encore développé l'événement d'intérêt à ce délai.

### Survie médiane median survival

Un indicateur global de la survie est la survie médiane : c'est le délai à partir duquel la probabilité de survie (ou la survie sans événement) passe en-dessous de 50 %. Plus elle est élevée, plus la survie (ou la survie sans événement) est élevée.

### Méthode de Kaplan-Meier

La méthode de Kaplan-Meier est très souvent utilisée pour obtenir des estimations non biaisées des probabilités de survie en présence de censure (au-delà du temps de censure, on ne sait pas si le patient a ou non présenté l'événement). La survie à un délai donné est calculée comme le produit des survies par intervalles de temps durant lesquels le nombre de patients suivi est constant (cf schéma ci-dessous). La méthode de Kaplan-Meier suppose que le fait qu'un patient soit censuré n'influence pas sa vraie probabilité de développer l'événement, et donc que la survie des patients encore suivis est un bon reflet de la survie des patients censurés. Ce n'est pas le cas, par exemple, lorsque des patients retournent dans leur pays d'origine en fin de vie et que l'on ne dispose plus de leurs nouvelles à partir de ce moment-là ; la censure est alors signe d'un événement imminent.

## Test du log-rank

log-rank test

Les courbes de survie de plusieurs groupes, estimées par la méthode de Kaplan-Meier, peuvent être comparées par le test du log-rank, dont l'hypothèse nulle est que les probabilités de survie à tous les délais sont égales dans les différents groupes. Le test du log-rank ne s'applique pas lorsque les courbes de survie à comparer se croisent.

## Taux de survenue d'événement

hazard

Le taux de survenue de l'événement mesure la vitesse (instantanée) d'apparition de l'événement. On peut parler d'incidence *instantanée*.

## Hazard ratio

Le hazard ratio (noté HR) entre deux groupes correspond au rapport du taux d'événement (ou d'incidence instantanée) entre ces groupes. La conclusion du hazard ratio s'interprète comme celle d'un risque relatif (HR > 1 : facteur aggravant, HR < 1 : facteur protecteur). Avant de calculer le hazard ratio, il est nécessaire de s'assurer que celui-ci est constant au cours du temps (i.e. l'effet du facteur étudié sur le taux est constant au cours du temps).

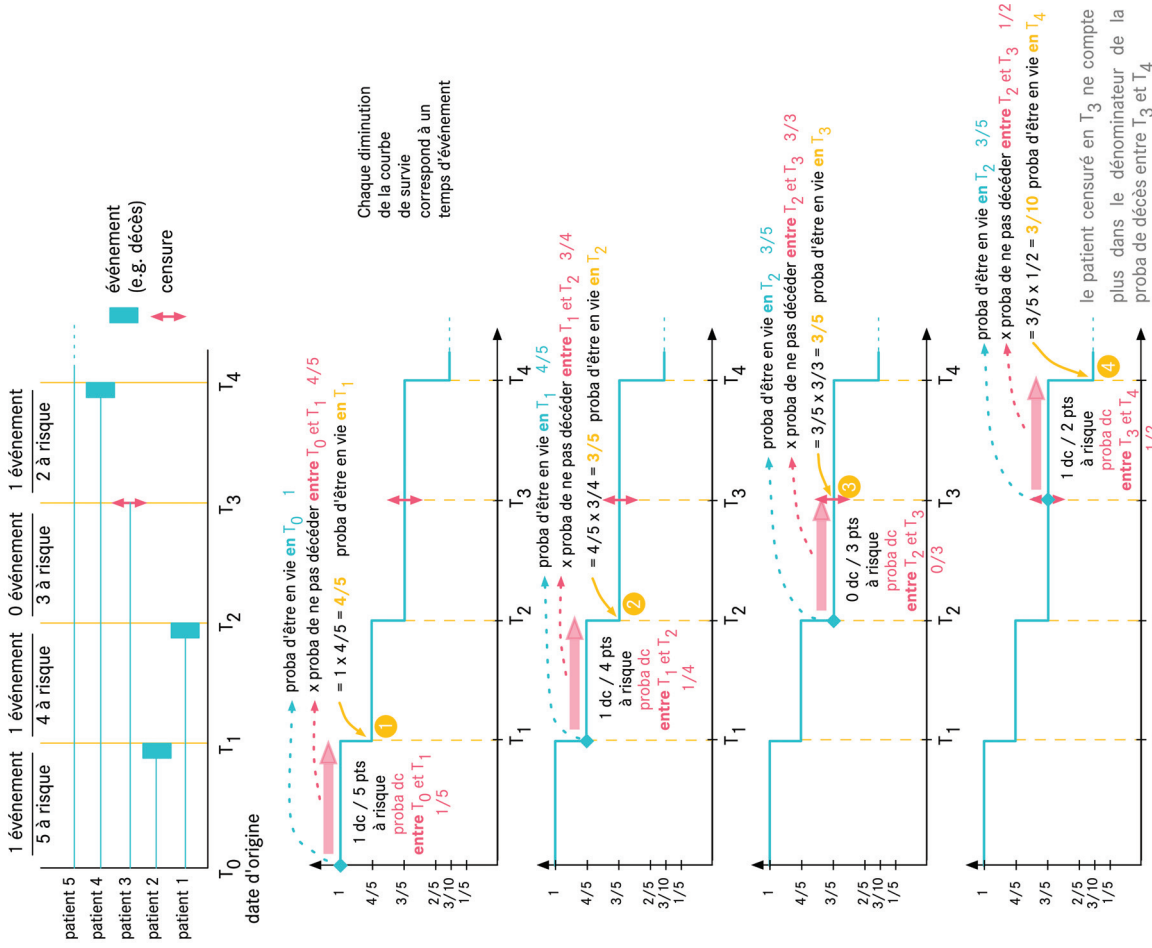
## Modèle de Cox

Cox model

Le modèle de Cox est souvent utilisé pour modéliser l'influence de facteurs sur le taux de survenue de l'événement. Ce modèle permet de calculer, pour un facteur binaire (par exemple groupe A et B), le hazard ratio (HR) associé. Lorsque le facteur étudié est quantitatif, le hazard ratio mesure par combien le taux est multiplié lorsque la variable augmente d'une unité. Par exemple, un hazard ratio de 1,1 pour l'âge signifie que le taux est multiplié par 1,1 pour chaque année supplémentaire.

Des tests issus du modèle de Cox permettent de savoir si le hazard ratio est statistiquement différent de 1.

Lorsque plusieurs facteurs sont considérés simultanément (analyse multivariée), le modèle de Cox permet d'estimer des hazard ratio ajustés ; auquel cas, on mesure le ratio de taux de survenue de l'événement entre les groupes A et B à niveau comparable des autres facteurs étudiés (ceci permet de limiter les effets de confusion).



Par souci de simplicité, il a été considéré sur cet exemple que les événements survenaient à intervalle de temps régulier, ce qui est rarement le cas



## T

### ▣ Taux d'incidence standardisé

*standardized incidence ratio (SIR), standardized mortality ratio (SMIR)*

Technique permettant de prendre en compte les différences de caractéristiques influençant le décès (par exemple l'âge) entre une population étudiée et une population de référence pour évaluer le taux de décès dans la population étudiée. Ceci permet de prendre en compte des facteurs de confusion : si la distribution de l'âge n'est pas la même entre deux populations à comparer, par exemple un pays avec une population très jeune (Chine) et une population plus âgée (France), la standardisation de l'incidence par rapport à une même population de référence rend les taux d'incidence comparables (i.e. l'effet de confusion de l'âge est retiré).

### ▣ Technique d'échantillonnage

voir partie Population d'étude

### ▣ Test statistique

voir aussi partie Inférence fréquentiste

### ▣ Hypothèses d'un test statistique (fréquentiste)

*statistical hypothesis*

Pour les tests statistiques, on définit une hypothèse nulle,  $H_0$  (souvent : pas de différence entre les groupes étudiés) et une hypothèse alternative,  $H_1$  (souvent : présence d'une différence entre les groupes). Le test statistique est construit dans le but de rejeter l'hypothèse nulle, avec une certaine probabilité d'erreur. Pour cela, il regarde la vraisemblance des données observées par rapport aux valeurs attendues sous l'hypothèse nulle considérée comme vraie ; si cette vraisemblance est trop faible, l'hypothèse nulle est rejetée en faveur de l'hypothèse alternative.

### ▣ Test bilatéral

*two-sided test*

En inférence fréquentiste, test statistique pour lequel l'hypothèse alternative correspond à l'existence d'une différence, qu'elle soit positive ou négative (par exemple, lorsque l'on compare deux traitements, le test bilatéral recherche une différence d'efficacité, que le traitement A soit supérieur ou inférieur au traitement B).

### ▣ Test unilatéral

*one-sided test*

En inférence fréquentiste, test statistique pour lequel l'hypothèse alternative correspond à l'existence d'une différence dans un seul sens (par exemple A supérieur à B) ; ce sens est fixé avant la réalisation du test.

Une limite des tests unilatéraux est de ne pas pouvoir conclure lorsque la différence observée est dans le sens contraire de celle supposée pour l'hypothèse du test.

### ▣ Degré de significativité (valeur de p, significativité statistique)

*p-value*

Souvent noté « p », le degré de significativité mesure le degré de compatibilité entre les résultats de l'étude et l'hypothèse nulle formulée lors d'un test statistique. Classiquement, l'hypothèse nulle (par exemple d'égalité des moyennes) est rejetée lorsque le degré de signification est inférieur à 0,05 (correspondant au risque alpha consenti). Dans ce cas, il y a moins de 5 % de chances de conclure à tort à une différence statistiquement significative.

### ▣ Risque alpha (ou risque de première espèce)

*alpha risk*

Probabilité de conclure à une différence alors qu'elle n'existe pas (rejet à tort de l'hypothèse nulle). Le risque alpha est fixé en amont de l'analyse (souvent à 5 %, ce seuil étant arbitraire).

### ▣ Risque bêta (ou risque de deuxième espèce)

*beta risk*

Probabilité de ne pas conclure à une différence alors que cette différence existe (non rejet à tort de l'hypothèse nulle). Il s'agit du complément à 1 de la puissance (1 – puissance). Ce risque bêta n'est pas maîtrisé.

### ▣ Puissance statistique

*power*

Dans un test statistique, probabilité de conclure à une différence qui existe bien dans la réalité (on peut faire le parallèle avec un vrai positif). C'est le complément à un du risque bêta (1-beta).

Elle dépend :

- du nombre de patients inclus (augmentation de la puissance avec le nombre de patients)
- de la différence espérée entre les groupes (augmentation de la puissance avec l'augmentation de la différence)
- de paramètres annexes (exemple : écart-type des valeurs mesurées dans les deux groupes pour la comparaison de moyennes, ...)
- du risque alpha consenti (à effectif constant, choisir un risque alpha faible, c'est-à-dire

plus strict, entraîne une baisse de puissance)  
 - du caractère uni  ou bilatéral  du test.


Dans une étude, l'effectif est calculé de sorte à atteindre une puissance suffisante (au moins 80 %) sous l'hypothèse d'une différence espérée. La puissance effective n'est jamais connue, car la vraie valeur de la différence est inconnue.

La notion de puissance s'interprète comme le pouvoir grossissant d'un microscope





Avec un plus fort grossissement (puissance), l'écart entre les deux cellules devient visible

### **Tests paramétriques/non paramétriques** *parametric/non-parametric test*

Parmi les tests statistiques, certains tests sont dits paramétriques (faisant une hypothèse sur la distribution des critères de jugement, par exemple distribution normale, qu'il faut vérifier lors de l'analyse) et d'autres non paramétriques (ne faisant pas d'hypothèse sur la distribution des critères) ; les tests non paramétriques sont moins puissants  que les tests paramétriques.

Exemples de tests paramétriques : test de comparaison de moyennes de Student, test d'ANOVA, test du Chi2 de comparaison de proportions. Tests non paramétriques : test de comparaison de médiane de Wilcoxon ou Mann et Whitney, test de Fisher de comparaison de proportions.

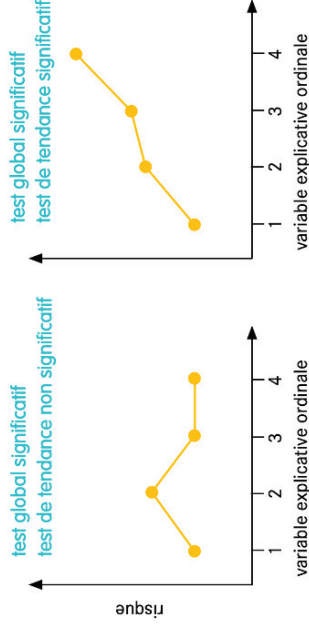
### **Test de tendance** *trend test*

Certains tests ne font qu'une seule comparaison ; ils sont alors multipliés lorsque plusieurs comparaisons sont prévues, par exemple lorsque l'on compare le risque de décès entre quatre classes d'âge (par rapport à une classe de référence par exemple). Ceci augmente le risque  $\alpha$   (tests multiples .

D'autres tests sont globaux, c'est le cas de l'ANOVA. Appliqués à plus de deux comparaisons (par exemple quatre classes d'âge), ils sont significatifs si au moins une catégorie d'âge se démarque des autres (hypothèse nulle  : toutes les classes d'âges ont le même risque de décès ; hypothèse alternative : au moins une classe d'âge a un risque de décès différent des


autres). Ils ne permettent pas d'identifier spécifiquement les classes qui se démarquent des autres (ce qui nécessiterait de faire des comparaisons deux à deux). Ainsi, une seule p-value est présentée pour quatre classes d'âge.



Enfin, certains tests permettent de tester des tendances pour une variable explicative continue ou ordinale (trend test, p-trend). Le test détecte par exemple si le risque de décès varie avec la classe d'âge de façon linéaire, que l'évolution soit positive ou négative. A nouveau les tests de tendances se résument par une seule p-value (présence ou absence de tendance). Un test global peut être significatif alors que le test de tendance ne l'est pas, comme dans l'exemple ci-dessous.



### **Tests multiples** *multiple tests*

Dans les études, il est fréquent de devoir réaliser plusieurs tests statistiques :

- comparaison de traitements selon différents critères de jugement
- comparaison de traitements selon différents sous-groupes de patients
- comparaison de plusieurs traitements expérimentaux à un traitement de référence
- réalisation d'analyses intermédiaires  en cours d'étude
- test de la significativité de plusieurs paramètres dans un modèle
- ...

Dans ce cas, le risque qu'« au moins » un de ces tests aboutisse à une fausse découverte est plus élevé que le risque  $\alpha$   classique (5 %) qui correspond au risque de fausse découverte pour « un seul test » (voir inférence fréquentiste ). Par exemple, en réalisant 20 tests, chacun au risque  $\alpha$  de 5 %, la probabilité d'aboutir à au moins une fausse découverte peut être proche de 100 %.

Il est donc fortement conseillé de limiter les tests statistiques aux questions essentielles de l'étude.


Il existe néanmoins des méthodes permettant de tenir compte de la multiplicité des tests. Parmi les plus classiques, on peut citer :

- la méthode de Bonferroni : réalisation de chacun des  $n$  tests au risque  $\alpha / n$
  - la méthode des tests hiérarchiques
- La méthode de prise en compte de la multiplicité des tests doit être spécifiée dans la partie méthodes.

### Tests hiérarchiques

*hierarchical testing*

La méthode des tests hiérarchiques est une des approches pour limiter le risque de fausse découverte sur l'ensemble des tests (voir multiplicité des tests). Elle consiste à ordonner, a priori, les tests à réaliser en fonction de leur importance (mettre en premier le test le plus important, ou qui a le plus de chances d'être significatif). Chaque test est réalisé au risque  $\alpha$  classique ; en revanche, dès qu'un test est non significatif, on ne réalise plus les tests suivants. Avec cette méthode, le risque d'avoir au moins une fausse découverte, sur l'ensemble des tests réalisés, ne dépasse pas  $\alpha$ .

 Dans la comparaison de l'efficacité de deux traitements sur la durée d'intubation et la durée globale d'hospitalisation pour des cas COVID sévères intubés, on comparera dans un premier temps les deux approches en termes de durée d'intubation au risque  $\alpha$  de 5 % ; si ce test est significatif, on comparera les deux approches en termes de durée globale d'hospitalisation au risque  $\alpha$  de 5 %. Si le premier test n'est pas significatif, on n'effectue pas de comparaison selon la durée globale d'hospitalisation.

Selon cette approche, le risque de commettre une erreur sur au moins un de ces deux tests ne dépassera pas 5 %.

La hiérarchie des tests doit être définie en amont des analyses (lors du protocole).

## U

### Univariée (analyse)

voir partie Analyse

## V

### Valeur de p

voir partie Test statistique : degré de significativité

### Valeur prédictive négative

voir partie Performances diagnostiques

### Valeur prédictive positive

voir partie Performances diagnostiques

### Validité

Il existe deux mécanismes d'erreur de l'estimation : aléatoire et systématique. Les erreurs systématiques sont des biais ; l'opposé d'un biais est la validité interne. L'opposé d'une erreur aléatoire est la précision.

La validité d'une étude a deux composantes :

- la validité de l'inférence par rapport à la population source de l'étude (c'est-à-dire l'absence de biais) : validité interne ;
- la validité de l'inférence si on généralise les conclusions à des personnes hors de cette étude (population cible) : validité externe. La validité externe est donc l'applicabilité d'une conclusion produite par une étude à l'extérieur du contexte de cette même étude ; on parle aussi de généralisabilité.

Les atteintes à la validité interne sont les biais (confusion, sélection, mesure/classement). Pour juger la validité interne, il faut analyser chaque processus de l'estimation pouvant produire un biais ainsi qu'estimer la force de ces biais sur l'estimation.

La validité externe est remise en cause lorsque certains facteurs (ex. caractéristiques des patients) sont distribués différemment entre la population de l'étude et la population cible.

Pour juger de la validité externe il faut analyser les critères de sélection de la population (caractéristiques des individus, niveau de risque). Pour obtenir une bonne validité externe, il faut que la population d'étude soit « représentative » de la population cible.

Par ailleurs, la validité externe requiert également la cohérence entre les résultats de l'étude avec ceux d'autres études sur la même hypothèse de recherche (i.e. cohérence externe).



Les résultats d'une étude cas-témoins évaluant les facteurs de risque de cancer du pancréas en Chine seront difficilement généralisables/extrapolables à une population cible française, car les caractéristiques sociodémographiques et les habitudes de vie (expositions, facteurs de confusion) sont très différentes.