

# Méthodologie des études épidémiologiques

## Concepts de base

### Introduction

### Historique

- ⇒ **Avant 1950** : l'épidémiologie est centrée sur les maladies infectieuses (étude des épidémies)
- ⇒ **A partir de 1950** : le champ de l'épidémiologie s'est élargi à toutes les pathologies, notamment chroniques (évolution de la définition de l'épidémiologie)

### Définition de l'épidémiologie

Etude de la **distribution** dans **le temps** et dans **l'espace** des **états de Santé** dans les **populations humaines** et de leurs **déterminants** ayant pour but la **prévention** des problèmes de santé.

### 3 branches

- Epidémiologie descriptive
- Epidémiologie analytique (ou étiologique)
- Epidémiologie évaluative

- **Epidémiologie descriptive**

Étude de la **fréquence** et de la **répartition** des états de Santé dont le but est de:

- ⇒ **Etablir des taux ou proportions de personnes** présentant un état de santé donné dans une population définie en tenant compte du facteur temps
- ⇒ **Étudier les variations de fréquence des problèmes de santé** selon les caractéristiques des personnes (âge, sexe, profession...), la zone géographique, le temps

Utilité :

- Outil de planification sanitaire (détermination des priorités de Santé publique)
- Permet d'élaborer des hypothèses étiologiques

- **Epidémiologie analytique (ou étiologique)**

Étude des **causes** des problèmes de Santé par l'analyse de l'association entre exposition à différents facteurs et états de santé

- ⇒ Identification de **facteurs de risque**

Utilité :

- Identification de groupes de population à haut risque
- Ciblage des interventions sanitaires : dépistage, prévention, information

- **Epidémiologie évaluative**

Étude de l'évaluation des **interventions** de Santé, préventives et curatives, menées en collectivité

Utilité :

- Mesure des **effets** d'une intervention de Santé par rapport à ses objectifs
- Afin **d'aider à la prise de décision** et à la **planification** (au niveau des politiques de Santé)

## Champs de l'épidémiologie

Cadre	Epidémiologie descriptive	Epidémiologie analytique	Epidémiologies évaluative
<b>Epidémiologie de population</b> (sujets sains et sujets malades)	Répartition de la maladie	Facteurs de risque	- Action de prévention primaire (= lutte contre les facteurs de risques) - Action de dépistage
<b>Epidémiologie clinique ou Recherche clinique</b> (sujets malades)	Nosographie de la maladie (= étude de la répartition des ≠ types de maladie)	Facteurs pronostiques	- Méthodes diagnostiques - Méthodes thérapeutiques

## Indicateurs de santé

**Objectifs des indicateurs de Santé** : doit répondre à 2 questions

⇒ Quelle est la **fréquence** du phénomène de Santé à un moment donné?

Indicateurs statiques : la prévalence

Obtenus par des enquêtes transversales.

⇒ Quelle est la **vitesse de survenue** du phénomène de Santé pendant une période déterminée?

Indicateurs dynamiques : incidence de la maladie et mortalité

Obtenus par des enquêtes longitudinales.

## Indicateurs de Morbidité

**Définition de la morbidité** : passage de l'état de non-maladie à l'état de maladie, progressif et mal défini.

- **Morbidité ressentie** : individuelle donc subjective
- **Morbidité objective** : observée si toute la population était dépistée systématiquement pour toutes les maladies
- **Morbidité diagnostiquée** : dépend de la morbidité ressentie et reflète le fonctionnement du système de soins  
C'est la seule qu'on peut mesurer : Prévalence et Incidence

### • Prévalence

La prévalence est une proportion de malades présents **à un moment donné** dans une population

$$\frac{\text{nombre de sujets atteints de la maladie dans la population au temps } t}{\text{effectif de la population au temps } t \text{ (malades + non malades)}}$$

S'exprime sous forme d'un %, sans unité : nombre de cas pour 100, 1000, 10000, ...

Elle inclut tous les cas de maladie : anciennement diagnostiqués et récemment diagnostiqués

Seule : pas toujours bon indicateur, car son augmentation n'est pas toujours péjorative

Dépend de 2 dimensions :

- **Durée de la maladie** : si elle augmente, la prévalence augmente
- « **Vitesse d'apparition** » des nouveaux cas de la maladie (= taux d'incidence) : si elle augmente, la prévalence augmente

Si on s'intéresse à l'étiologie de la maladie, la seconde dimension « vitesse d'apparition » de la maladie est la plus intéressante

⇒ Introduction d'un nouvel indicateur : le taux d'incidence (TI)

Si durée et incidence sont constants dans le temps : Prévalence = TI × temps

- **Incidence cumulée**

C'est le rapport du nombre de **nouveaux cas d'une maladie** survenus pendant un période de temps donnée, divisé par la population à risque de développer la maladie pendant cette période.

$$\frac{\text{nombre de nouveaux survenant pendant une période } \Delta t}{\text{effectif de la population à risque pendant la période } \Delta t}$$

S'exprime sous forme d'un %, sans unité : nombre de cas pour 100, 1000, 10000, ...personnes pendant la période  $\Delta t$

La Période de mesure et généralement systématique (annuelle, mensuelle, hebdomadaire,...) et doit être spécifiée.

Pour que le calcul ait un sens, la population doit rester stable pendant la période.

L'augmentation de l'incidence est toujours péjorative.

Cas particulier : le **taux d'attaque** (utilisé pour les maladies infectieuses sur de courtes périodes)

Même signification que l'Incidence Cumulée en cas d'épidémie.

Période de mesure pas systématique mais dépend de la durée de l'épidémie.

= Incidence cumulée de la maladie pendant la durée de l'épidémie

## Indicateurs de Mortalité

Ce sont des indicateurs **dynamiques**.

Similaire à l'incidence mais l'évènement étudié n'est plus la survenue de la maladie mais la survenue du décès.

- **Taux brut de mortalité**

$$\frac{\text{nombre de décès enregistrés dans la population pour une période donnée}}{\text{effectif de la population pendant la même période}}$$

Remarque : pour que le calcul ait un sens, la population doit rester stable pendant la période

- **Taux brut de mortalité par cause**

$$\frac{\text{nombre de décès dus à cette cause pour une période donnée}}{\text{effectif de la population pendant la même période}}$$

- **Taux spécifiques de mortalité**

$$\frac{\text{nombre de décès du sous – groupe pour une période donnée}}{\text{effectif de la population du sous – groupe pendant la même période}}$$

Caractéristiques différents de sous groupes : âge, sexe, ...

- **Létalité**

= Part des décès dus à la maladie parmi les patients atteints de cette maladie

Exemple :

$$\frac{\text{nombre de décès par cancer en France en 2010}}{\text{nombre de sujets atteints de cancer en France en 2010}}$$

Témoigne de la gravité de la maladie et de la qualité des soins.

La létalité peut diminuer et la mortalité augmenter

⇒ Progrès de la médecine mais augmentation plus rapide de l'incidence

Mortalité : concept d'épidémiologie et Santé Publique

Létalité: concept de recherche clinique

# Mesures de risque, d'association et d'impact

## Définition du Risque

Probabilité de survenue d'un événement (décès ou maladie) à un moment donné pendant un intervalle de temps défini.

En épidémiologie :

- Risque de survenue d'un décès = mortalité
- Risque de survenue d'une maladie = incidence

Cette probabilité est différente selon les individus et variable selon certaines caractéristiques.

## Définition des facteurs de risque

Ils sont de nature probabiliste.

Un facteur F est un facteur de risque pour une Maladie M si l'exposition au facteur F modifie la probabilité d'apparition de la maladie M (incidence).

$$P_F(M) \neq P_{\bar{F}}(M)$$

Comment mesurer l'effet d'un facteur de risque ?

- **Mesures d'association** : Risque Relatif (RR) et Odds Ratio (OR)
- **Mesures d'impact** : Excès de Risque (ER) et Risque attribuable (RA)

Complémentarité des deux types d'indicateurs en matière de santé publique.

## Mesures d'association

- **Risque relatif (RR)**

Rapport des taux d'incidence de la maladie dans les deux groupes d'exposition à un facteur

$$RR = \frac{P_F(M)}{P_{\bar{F}}(M)} = \frac{I_F}{I_{\bar{F}}}$$

$I_F$  : Incidence de la maladie chez les exposés

$I_{\bar{F}}$  : Incidence de la maladie chez les non-exposés

Quantification multiplicative

- ⇒ Mesure la force de l'association
- ⇒ Mesure le rôle étiologique du facteur
  - Si  $RR = 1$ , le facteur ne constitue pas un risque
  - Si  $RR > 1$ , le facteur constitue un risque
  - Si  $RR < 1$ , le facteur est protecteur

- **Odds ratio (OR)**

### Odd ratio de maladie

$$OR = \frac{P_1 / (1 - P_1)}{P_0 / (1 - P_0)}$$

$P_1$  : probabilité d'être malade chez les exposés

$P_0$  : probabilité d'être malade chez les non exposés

### Odd ratio d'exposition

$$OR = \frac{P_{E1} / (1 - P_{E1})}{P_{E0} / (1 - P_{E0})}$$

$P_{E1}$  : probabilité d'être exposé chez les malades

$P_{E0}$  : probabilité d'être exposé chez les non malades

Odd ratio de maladie = Odd ratio d'exposition

Odd ration : son interprétation est moins facile et sa compréhension moins immédiate que le risque relatif. Cependant il est très utilisé en épidémiologie, car il est **estimable à partir de tous les types d'enquête** contrairement au risque relatif.

Lorsque la maladie est rare (prévalence < 1%) :  $OR \approx RR$

$RR$  et  $OR$  sont des **indicateurs multiplicatifs du risque**.

## Mesure d'impact

- **L'excès de risque (ER)**

$$ER = P_F(M) - P_{\bar{F}}(M) = I_F - I_{\bar{F}}$$

L'excès de risque est la différence entre les deux probabilités précédentes : c'est un indicateur additif du risque.

Parmi les sujets exposés, certains auraient pu contracter la maladie même en l'absence du facteur F. L'excès de risque exprime donc l'augmentation de l'incidence de M, due à la présence du facteur F.

- **La proportion de risque attribuable (PRA)**

Risque attribuable dans la population totale.

= Proportion de cas de la maladie que l'on peut attribuer au facteur de risque dans l'ensemble de la population (exposés + non exposés).

$$PRA = \frac{p \cdot (RR - 1)}{p \cdot RR + (1 - p)}$$

$p$  = proportion de personnes exposées dans la population

**Mesure l'impact du facteur de risque dans la population.**

**N'a de sens que si le facteur est un facteur causal** de la maladie.

Son interprétation nécessite beaucoup de **prudence** car elle est directement liée à l'estimation du risque relatif et donc sensible à tous les biais qui peuvent affecter l'estimation de ce dernier.



### Exemple de la complémentarité des deux types de mesure

Mortalité /  $10^5$  / an

	Cancer du poumon	Maladie coronarienne
Fumeurs F+	166	599
Non fumeurs F-	7	422

Prévalence de fumeurs dans la population = 29%

- **RR**

Cancer du poumon =  $166/7 = 23,7$

Maladie coronarienne =  $599/422 = 1,4$

Le risque de décéder :

- d'un cancer du poumon est environ 24 fois plus important chez les gros fumeurs que chez les non fumeurs.
- d'une maladie coronarienne est 1,4 fois plus important chez les gros fumeurs que chez les non fumeurs.

Le risque lié au tabac est plus fort pour le décès par cancer du poumon que pour le décès par maladie coronarienne.

- **ER**

Décès par cancer du poumon =  $166-7 = 159$  pour  $10^5$

Décès par maladie coronarienne =  $599-422 = 177$  pour  $10^5$

Le tabac cause chaque année :

- 159 décès supplémentaires par cancer du poumon pour  $10^5$  habitants.
- 177 décès supplémentaires par maladie coronarienne pour  $10^5$  habitants.

- **PRA**

Décès par cancer du poumon = 87%

Décès par maladie coronarienne = 11%

Sur 100 nouveaux cas de décès par cancer du poumon, 87 seraient évitables si le FR tabac était évincé.

Sur 100 nouveaux cas de décès par maladie coronarienne, 11 seraient évitables si le FR tabac était évincé.

## Interprétations

En termes de Santé Publique, le rôle du tabac est aussi important pour la maladie coronarienne que pour le cancer du poumon car il entraîne environ le même nombre de décès chaque année (ER égaux).

En revanche, le tabac expose plus au risque de décès par cancer du poumon qu'au risque de décès par maladie coronarienne (RR plus élevé).

Enfin, le tabac est le facteur de risque prépondérant pour le cancer du poumon, alors que les facteurs sont multiples pour la maladie coronarienne, ce qui explique sa plus grande fréquence.

## Conclusion sur les facteurs de risques

La connaissance du risque et des facteurs de risque est essentiel en termes de prévention.

⇒ Permet de prendre des décisions en Santé Publique

Complexe :

- Peu de vrais FR (causaux)
- Souvent des FR multiples pour une même maladie
- N'expliquent que peu la maladie (hormis le tabac qui explique environ 90% des cancers du poumon)

# Erreur en épidémiologie

Toute mesure est sujette à l'erreur.

Deux types d'erreur

- L'erreur aléatoire : manque de précision
- L'erreur systématique : **biais**

Quand on planifie une enquête :

- **Objectif de précision** : absence d'erreur aléatoire  
Calcul du NSN (Nombre de Sujets Nécessaire)
- **Objectif de Validité** : absence de biais  
Précautions méthodologiques selon, le type d'enquête

## Biais

### Définition

Distorsion dans l'estimation d'une mesure en épidémiologie, liée à une erreur systématique  
Tend à fournir une estimation du paramètre mesuré différente de la « vraie » valeur de la population

En résumé, un biais introduit une **erreur systématique** qui s'ajoute à **l'erreur aléatoire** liée aux fluctuations d'échantillonnage.

### Types de Biais

- **De Sélection**

Résulte de la façon dont l'échantillon est choisi au sein de la population

Ex: sélection des sujets sur le volontariat pour une enquête de satisfaction

Le taux de réponse sera vraisemblablement lié à la satisfaction

- **De Classement ou d'Information**

Résulte d'erreurs de mesure sur l'exposition ou sur la maladie

Ex: mesure de l'exposition aux solvants par simple interrogatoire ou par investigation plus poussée sur le poste de travail et les tâches effectuées

Le taux d'exposés ne sera vraisemblablement pas le même selon les deux modes de mesure de l'exposition

- **De Confusion**

Résulte de l'influence de tiers facteurs sur l'association entre l'exposition et la maladie

## Phénomène de confusion

Une association statistique entre une exposition E et une maladie M peut être due à un facteur de confusion X :

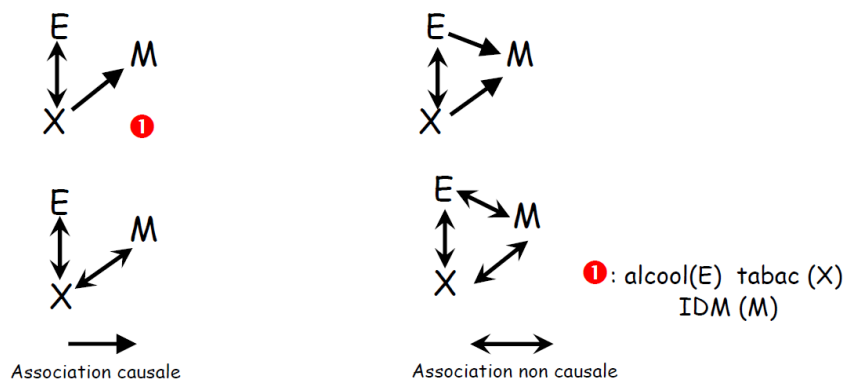
- ⇒ Le facteur de confusion est lié à la fois à la maladie et à l'exposition
- ⇒ L'association observée entre l'exposition et la maladie peut être partiellement expliquée par le facteur de confusion
- ⇒ Le RR brut lié à l'exposition est modifié lorsqu'on ajuste sur le facteur de confusion

Un facteur de confusion sur la relation entre un facteur de risque et une maladie est un tiers facteur qui, lorsqu'on le prend en compte, **modifie l'effet apparent du facteur de risque étudié.**

## Notion de facteur de confusion

Pour être facteur de confusion sur la relation entre une exposition E et une maladie M, un facteur X doit satisfaire aux **3 conditions** suivantes :

- Être un facteur de risque pour M (causal ou non)
- Être associé à E
- Ne pas être une conséquence de E



**Négliger les facteurs de confusion** dans l'analyse des liens entre facteurs de risque et maladie conduit à des interprétations erronées.

Un **problème majeur** en épidémiologie est de **tenir compte de tous les facteurs de confusion**: parfois on n'en soupçonne même pas l'existence.

La **prise en compte** des facteurs de confusion lors des études épidémiologiques se fait.

- Soit lors de la **planification des enquêtes**
- Soit lors de l'**analyse** des enquêtes par des méthodes statistiques appropriées

# Les enquêtes

## Définition

Procédure visant à rechercher, à rassembler, à **recueillir des informations sur l'état de santé** d'une partie ou de la totalité d'une **population** (sujets sains et/ou malades)

Dans le but d'**estimer un indicateur** :

- Descriptif (prévalence, incidence, mortalité,...)
- Étiologique (RR, ER, OR)

## Typologie des enquêtes

### Selon le nombre de sujets

- Exhaustive (toute la population concernée par l'enquête)
- Sur Échantillon (le plus souvent)

### Selon l'objectif

- Descriptif
- Étiologique
- Évaluatif

### Selon la chronologie

- Transversale (à un moment donné)
- Longitudinale (suivi dans le temps)
- Prospective (meilleure qualité que rétrospectif)
- Rétrospective

### Selon la procédure

- Observationnelle
- Expérimentale

On distingue :

### Les enquêtes expérimentales (= étalon-or)

- ⇒ **Imputation causale** (relations de cause à effet)

### Les enquêtes d'observation

- ⇒ **Pas d'imputation causale** au sens strict  
Nécessité d'un Faisceau d'arguments

## Procédure expérimentale

**Intervention** délibérément choisie appliquée à des **sujets choisis** selon une **procédure déterminée par l'expérimentateur**.

Enquête **toujours prospective**.

Randomisée ou non.

Procédure de choix en **Recherche Clinique** (recherche chez le sujet malade).

Exemple : évaluation de l'effet d'un nouveau traitement

Constitution de deux groupes identiques en tout sauf en ce qui concerne le traitement à évaluer.

Par tirage au sort ou randomisation.

Ainsi une différence entre les deux groupes à l'issue de l'expérience pourra être imputée à la différence de traitement.

⇒ **Imputation causale**

## Procédure d'observation

Expérimentation pas toujours possible

Exemple : en épidémiologie (recherche en population: sujets non malades et sujets malades), quand on recherche l'effet d'un facteur de risque (cholestérol) pour une maladie (coronaropathie): le facteur «cholestérol» n'est pas contrôlable

**Procédure d'observation** : on va rechercher le lien entre le niveau de cholestérol et la présence ou l'absence de coronaropathie

Mais **facteurs de confusion** possibles (les deux groupes ne sont pas constitués au hasard) :

Les sujets hypercholestérolémiques sont aussi plus souvent obèses, sédentaires...que les normocholestérolémiques.

⇒ **Biais possibles**

⇒ **Pas d'imputation causale**

## Enquêtes (ou études)

- **Observationnelles**

**Transversale ou enquête de prévalence** : enquête descriptive

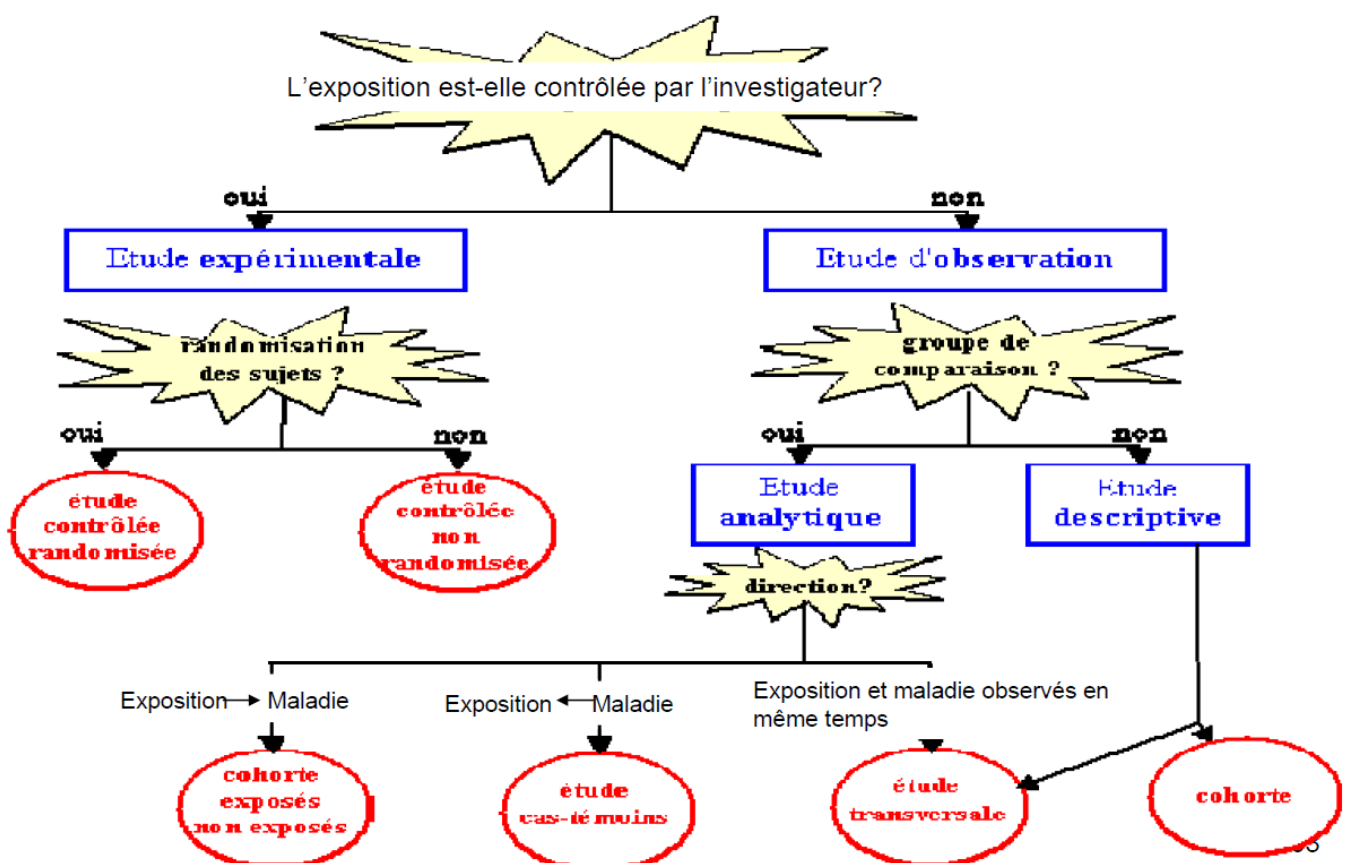
**Cohorte** (recueil prospectif de l'information) : enquête descriptive ou étiologique, se rapproche le plus de l'expérimentation

**Cas témoins** (recueil rétrospectif de l'information sur l'exposition) : enquête étiologique

- **Expérimentales**

Cas particulier de l'**essai thérapeutique comparatif** : enquête évaluation

## Schémas d'étude



## Etudes observationnelles

- **Enquête de Prévalence**

### Définition

Enquête dont l'objectif est d'**estimer la prévalence** d'une maladie dans une population à un moment donné

### Principe

Enquête **transversale**, c'est à dire:

- Réalisée à un moment donné (période brève)
- Pas de suivi des patients

Sur échantillon **représentatif** d'une population bien définie

Recueil d'informations :

- Sur la présence ou l'absence d'une maladie
- Sur des caractéristiques autres au moment de l'enquête : socio-démographiques, exposition à différents facteurs

### Champ d'application : Epidémiologie descriptive

**Image instantanée** du phénomène étudié dans la population

- ⇒ Ne permet **pas de connaître l'incidence** de la maladie : on observe seulement les cas existants au moment de l'enquête, les cas prévalents
- ⇒ Ne permet **pas d'étudier les aspects dynamiques** des phénomènes observés : la chronologie entre exposition à une nuisance et apparition d'une maladie ne peut pas être étudiée par ces enquêtes
- ⇒ Type d'enquête mal adapté à la recherche de Facteurs de Risque

**Estimation de la prévalence** avec son intervalle de confiance à 95% pour quantifier la précision de l'estimation.

**Difficulté majeure** : obtenir un **échantillon représentatif** de la population étudiée

Le tirage au sort est la seule méthode qui peut « garantir » la représentativité d'un échantillon

- Base de sondage nécessaire
- Problème des non réponders

**Risque de biais de sélection** : l'échantillon observé est différent par certaines caractéristiques (qui peuvent être liées au phénomène étudié) de la population totale

- La prévalence calculée sur l'échantillon risque de ne pas être une bonne estimation de celle de la population
- Pas de possibilité d'extrapoler les résultats à la population
- Le biais de sélection est difficile à quantifier

### Intérêt en Epidémiologie et en Santé Publique

- **Planification Sanitaire** : prévoir les moyens de prise en charge d'une maladie
- **Décrire la répartition temporo-spatiale d'une maladie**
- ⇒ **Hypothèses étiologiques**

Souvent le préalable à des enquêtes étiologiques de recherche de FR



- **Enquête de cohorte**

**Toujours longitudinale** : suivi des sujets dans le temps

**Souvent prospective** (parfois rétrospective: reconstitution de l'exposition et de la survenue de la maladie *a posteriori*)

**Schéma temporel** : suivi prolongé avec mesures régulières en vue de détecter l'apparition de symptômes, maladie, décès...

Population : cohorte constituée d'un groupe de sujets sélectionné sur certains critères (âge, lieu de résidence, exposition à une nuisance ...)

**Objectif :**

- **Soit Descriptif** : pas de groupe de référence
- **Soit étiologique** : enquête exposés-non exposés

**Indicateurs produits :**

- **Incidence** (enquête de cohorte descriptive)
- **RR – ER** si étiologique (enquête exposés-non exposés)

### Enquête exposés-non exposés

**Objectif**

- Vérifier l'hypothèse **d'une relation causale entre l'exposition à un FR et la survenue d'un événement de santé**
- Par comparaison de la fréquence de survenue de l'évènement (taux d'incidence) dans deux groupes de sujets: les exposés au FR et les non-exposés au FR

### Schéma de l'enquête



## Principe

- **Choix des sujets**
  - Tous indemnes de la maladie au début de la période d'observation
  - Répartition des sujets en deux groupes : exposés et non exposés
  - Souvent sur échantillon
  - Recueil de données initiales : sur l'exposition au FR, sur l'absence de la maladie, sur les facteurs de confusion
- **Période d'observation** : durée variable liée au délai susceptible de s'écouler entre l'exposition au FR et la survenue de la maladie (plusieurs années voire dizaines d'années)
- **Suivi répété des sujets** pendant toute la période d'observation
  - Mesure d'éventuelles modifications d'exposition
  - Observation de l'incidence de la maladie selon les différents niveaux d'exposition
- A la fin de la période d'observation, **on dispose de toutes les données nécessaires au calcul des incidences**
  - Indicateurs obtenus : incidence chez les exposés et non exposés, risque relatif et excès de risque

	<b>M+</b>	<b>M-</b>	<b>Total</b>
<b>F+</b>	<b>a</b>	<b>b</b>	<b>a+b</b>
<b>F-</b>	<b>c</b>	<b>d</b>	<b>c+d</b>

Incidence chez les exposés

$$P_F(M) = \frac{a}{a+b}$$

Incidence chez les non exposés

$$P_{\bar{F}}(M) = \frac{c}{c+d}$$

Risque relatif

$$RR = \frac{P_F(M)}{P_{\bar{F}}(M)}$$

Excès de risque

$$P_F(M) - P_{\bar{F}}(M)$$

## Biais

### De sélection

- Peu au moment de la sélection des groupes.
- Mais surtout en raison des perdus de vue liés à la longueur de l'enquête (appelé aussi biais de perdu de vue).

En principe, **peu de biais de classement** car recueil prospectif de l'information (pas d'information manquante sur l'exposition)

### De confusion

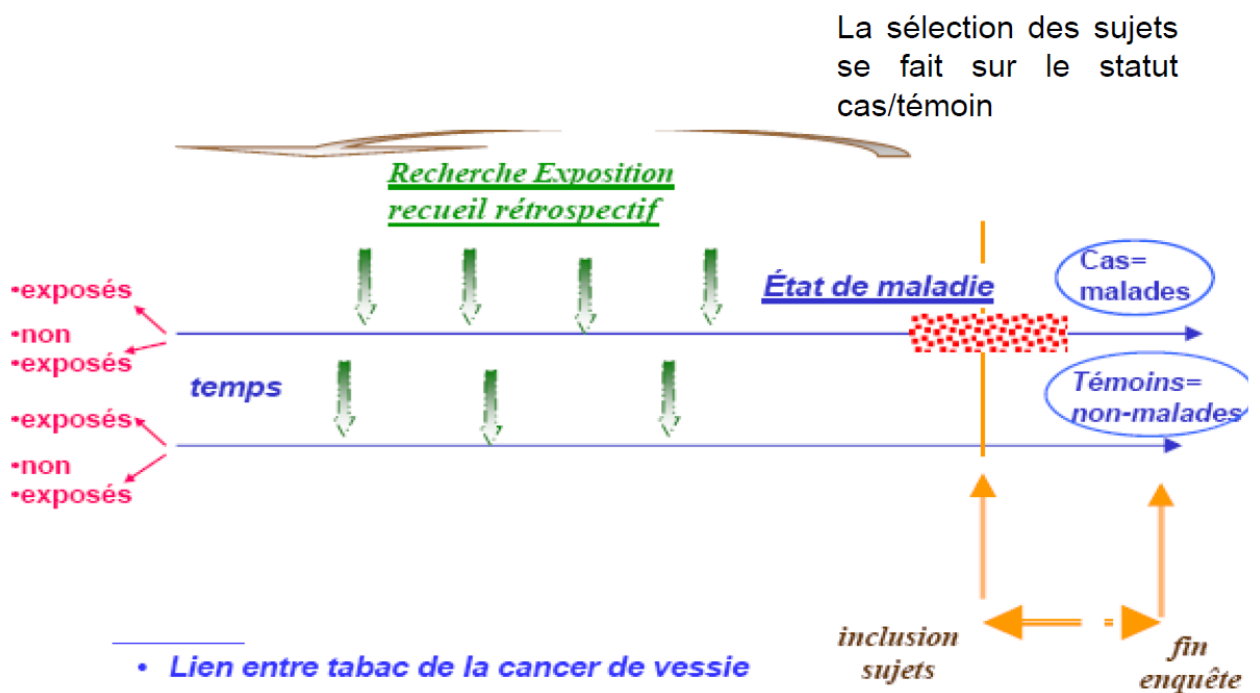
Avantages	Inconvénients
Bien adaptée pour étudier : <ul style="list-style-type: none"><li>- Les risques (incidences)</li><li>- Les expositions rares</li><li>- Plusieurs maladies</li><li>- La séquence exposition-maladie</li></ul>	Non adaptée pour étudier : <ul style="list-style-type: none"><li>- Les maladies rares</li><li>- Plusieurs expositions</li></ul>
Peu de biais de sélection (sauf perdus de vue)	Coût élevé lié à la durée de l'étude
Peu de biais d'information	Logistique lourde
	Longue période de latence

## Enquête cas-témoins

### Objectifs

- Vérifier l'hypothèse d'une **relation causale entre l'exposition à un facteur de risque et la survenue d'une maladie**
- Par comparaison de la fréquence d'exposition au FR étudié dans deux groupes de sujets : les cas atteints de la maladie et les témoins exempts de la maladie

### Schéma de l'enquête



## Principe

- **Choix des sujets**
  - o Recrutement des cas : sujets atteints de la maladie
  - o Recrutement des témoins : sujets indemnes de la maladie
  - o Recueil de **données rétrospectives** souvent par interrogatoire sur l'exposition au FR dans le passé (recherche de la chronologie exposition – maladie)
- **Durée de l'étude**
  - o Les sujets ne sont vus qu'une fois pour recueillir les données : pas de suivi longitudinal
  - o Durée nécessaire au recrutement et au recueil de données de l'ensemble des sujets
- A la fin de l'étude, **on dispose des données suivantes** :

	Cas	Témoins
E+	a	b
E-	c	d
	Nc	Nt

On ne peut pas calculer d'incidence donc pas de risque relatif.

Mais on peut calculer les fréquences d'exposition chez les cas et chez les témoins

⇒ **Calcul de l'Odds ratio** (bonne estimation du risque relatif)

$$OR = \frac{a.d}{b.c}$$

## Indicateurs obtenus

Nécessité de connaître les incidences de la maladie dans le groupe exposé et dans le groupe non exposé pour estimer le risque relatif et l'excès de risque

- Dans une enquête cas-témoins, on ne connaît pas ces incidences du fait que le point de départ est le diagnostic de la maladie : dans l'absolu, on ne peut estimer le RR et l'excès de risque
- ⇒ **On estime l'odds ratio (OR)** qui est une estimation du risque relatif quand la maladie est rare (<1%)

On fournit toujours son intervalle de confiance (en général à 95%) pour estimer la précision et la significativité :

- o S'il contient la valeur 1: FR non significatif
- o S'il ne contient pas la valeur 1: FR significatif

## Biais

### Biais de sélection des groupes

- Cas : doit être représentatif de l'ensemble des cas → diversification des lieux de recrutement
- Témoins : veiller à ce que l'exposition du groupe témoin soit représentatif de celle de la population générale dont sont issus les cas (ex : témoins hospitaliers, témoins de population)

### Biais de mesure de l'exposition

Dû au caractère rétrospectif du recueil de l'information

- Biais de mémorisation de l'enquêté
- Subjectivité
  - o De l'enquêteur : tendance inconsciente à classer plus facilement comme exposé un cas qu'un témoin
  - o De l'enquêté : tendance inconsciente à rechercher une cause à sa maladie donc à trouver des expositions par excès

### Biais de confusion

Avantages	Inconvénients
Bien adaptée pour étudier : <ul style="list-style-type: none"><li>- Les maladies rares</li><li>- Plusieurs facteurs de risque</li></ul>	Non adaptée pour étudier : <ul style="list-style-type: none"><li>- Les expositions rares</li><li>- La séquence temporelle exposition-maladie</li></ul>
Coût faible Logistique moins lourde Rapidité d'exécution Échantillons de taille modérée	Pas de calcul direct du RR OR biaisé si maladie fréquente dans la population Biais de sélection et d'information +++

## Etude expérimentale

### Cas particulier de l'Essai Thérapeutique Comparatif

#### But

Évaluer une **procédure thérapeutique** appliquée en clinique humaine pour une pathologie donnée

- En termes d'**efficacité**
- En termes de **tolérance**

Chez des **patients atteints de cette pathologie** : domaine de la Recherche Clinique

**Procédure expérimentale** : car c'est l'expérimentateur qui choisit l'intervention à tester

#### Les essais thérapeutiques des médicaments

Les phases d'expérimentation d'un médicament (très encadré par la loi)

- Essais précliniques (chez l'animal)
- Essais cliniques (chez l'homme)
  - o Phase I
  - o Phase II
  - o Phase III
  - o Phase IV

#### Phase III

Phase de **l'essai comparatif** :

- Pour étudier les propriétés du médicament par rapport
    - o Soit à un placebo (dénué d'activité pharmacologique)
    - o Soit à ceux déjà existants (traitement de référence)
  - Doit apporter la preuve
    - o De l'efficacité du médicament
    - o Et/ou de sa supériorité par rapport au médicament de référence
- ⇒ Imputation causale

Attribution ou non de l'Autorisation de Mise sur le Marché (AMM) si rapport bénéfice/risque acceptable.

Les essais comparatifs peuvent aussi comparer :

- Des techniques chirurgicales
- Des techniques médicales
- Des stratégies +/- complexes de prise en charge
- Des mesures préventives (essai de prévention)
- Des mesures d'éducation pour la Santé

## Objectif

- Vérifier l'hypothèse d'une **relation causale entre une «intervention thérapeutique» et la survenue d'un critère de jugement** (critère d'efficacité)
- Dans une population bien définie de sujets atteints d'une maladie donnée
- Par comparaison du critère entre deux groupes de sujets :
  - o Les patients traités avec le traitement à évaluer
  - o Les patients du groupe contrôle soit traités par le traitement de référence, soit sous placebo

## Choix des sujets

Obtenir un **échantillon représentatif** de la population cible :

- Population à laquelle on veut appliquer les résultats de l'étude
- N'utiliser que les critères de sélection indispensables pour définir cette population
- Comme ce sont les médecins qui vont inclure, il faut diversifier les lieux de recrutement : Essai multicentrique +++

### Préciser les critères d'inclusion

- Critères liés au patient : âge, sexe, compliance, signature du consentement ...
- Critères liés à la pathologie : forme clinique, gravité ...

### Préciser les critères de non-inclusion

- Contre indications à l'un des deux traitements (ex : allergie)
- Toujours respecter la clause d'ambivalence : un patient inclus doit pouvoir recevoir indifféremment les deux traitements

## Choix et définition des traitements

### Traitement à évaluer

- Traitement dont on a de bonnes raisons de penser qu'il est efficace
- Préciser les Modalités d'administration, la durée

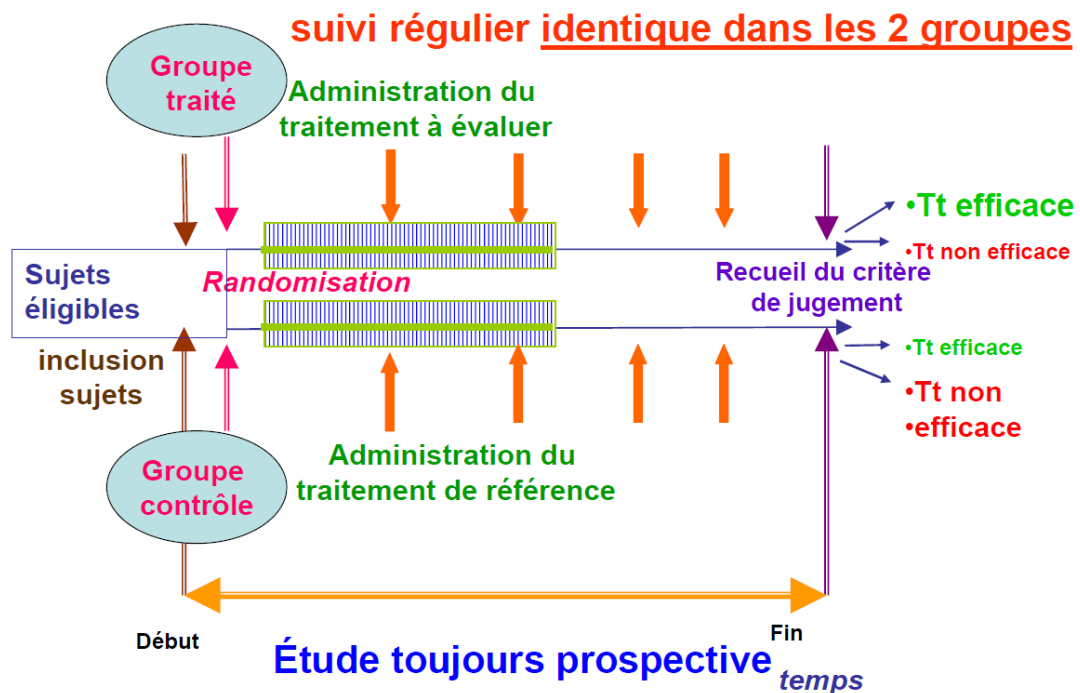
### Traitement du groupe contrôle

- Soit **traitement de référence reconnu** efficace et communément administré (bibliographie)
- Soit **placebo** sous réserve que ce soit éthique : le placebo est une substance dénuée de tout effet pharmacologique mais qui a l'aspect d'un médicament
- Soit exceptionnellement l'absence de traitement

**Prévoir les traitements complémentaires autorisés** : doivent être les mêmes dans les deux groupes



## Schéma d'un essai thérapeutique comparatif



### Principe

#### Toujours prospectif

#### Choix des sujets

- Tous atteints de la maladie et répondant strictement aux critères d'inclusion
- Répartition au hasard (randomisation) des sujets en deux groupes : traités par le traitement à évaluer versus traités par le traitement de référence

#### Période d'observation

Durée liée au délai susceptible de s'écouler entre le traitement et la mesure du critère d'efficacité : très variable d'un essai à l'autre

#### Suivi répété des sujets pendant toute la période d'observation

- Notification d'éventuels écarts au protocole : modifications de traitement, arrêt de traitement, prise d'autres traitements
- Recueil du critère de jugement dans chacun des deux groupes : comparaison statistique du critère de jugement entre les deux groupes (test du chi 2 si critère qualitatif, test de comparaison de moyennes si critère quantitatif)

#### Doit permettre l'imputation causale

- C'est bien le nouveau traitement qui a entraîné l'amélioration : jugement de causalité
- Avec le minimum d'erreur sur les valeurs observées : jugement de signification (statistique)
  - o Supériorité observée mais due au hasard (risque  $\alpha$ )
  - o Ou supériorité non mise en évidence (risque  $\beta$ )

## Comment obtenir l'imputation causale ?

### Pour le jugement de causalité :

- En comparant deux groupes qui ne diffèrent que par le traitement : le seul moyen d'obtenir deux groupes de patients aussi semblables que possible est d'attribuer le traitement par tirage au sort à chaque patient inclus
- ⇒ C'est la randomisation

### Pour le jugement de signification (statistique) :

- En incluant un nombre suffisant de patients dans l'essai de façon à minimiser les risques d'erreur statistique  $\alpha$  et  $\beta$
- ⇒ C'est le calcul du Nombre de Sujets Nécessaire

## Essais comparatifs : points clés

- **Comparabilité initiale des groupes** : la randomisation ou TAS
- **Maintien de la comparabilité des groupes au cours de l'essai** : l'aveugle ou l'insu et l'analyse en Intention de Traiter (ITT)
- **Pertinence de la mesure d'efficacité** : choix et moment de mesure du critère de jugement principal
- **Avoir une puissance statistique suffisante** : calcul du NSN

## Comparabilité initiale des groupes

### Pourquoi le tirage au sort ?

- Chaque patient, quelles que soient ses caractéristiques, doit avoir la **même probabilité** de recevoir un traitement ou l'autre
- Garantie d'imprévisibilité du traitement

### Que fait le tirage au sort ?

- Il évite les **biais de sélection et de confusion**
- Donc il **équilibre la répartition des facteurs pronostiques (connus ou inconnus) entre les groupes comparés**

### Le tirage au sort est-il éthique ?

- Un essai comparatif représente la façon la plus éthique de juger un nouveau traitement à deux conditions :
  - Les connaissances actuelles ne permettent pas de déterminer quel est le traitement le plus efficace : le médecin ne peut prendre de meilleure décision thérapeutique que celle de persuader le malade de participer à l'essai
  - L'essai réalisé permettra, par la valeur de la méthodologie employée, de répondre de façon aussi convaincante que possible au problème thérapeutique posé

## Maintien de la comparabilité des groupes au cours de l'essai

Au cours du déroulement de l'essai, de nombreux phénomènes peuvent modifier unilatéralement les groupes à comparer :

- Modification de la prise du traitement
- Thérapeutiques complémentaires
- Surveillance irrégulière
- Disparition des sujets : perdus de vue

### Perte de la comparabilité des groupes lors de l'analyse

La différence finale observée ne pourra être attribuée à la seule différence de traitement

Moyens pour limiter ces problèmes :

- Essai en aveugle (ou en insu)
- Analyse en intention de traiter (ITT)

## Double aveugle (ou double insu)

Les traitements à comparer sont indiscernables pour le patient comme pour le médecin et l'équipe soignante

Cette technique permet :

- Un comportement identique vis à vis des malades tout au long de l'essai
  - Une évaluation neutre donc objective de l'efficacité des traitements et des effets secondaires
- ⇒ Contrôle certains biais d'information

## Alternatives au double aveugle

- Simple aveugle : le patient est en aveugle et le médecin connaît le groupe de traitement du patient
- Quand ce n'est pas possible : l'essai est dit « ouvert »

Le double aveugle est indispensable

- Lorsque le critère de jugement est subjectif
- Lorsque le traitement de référence est un placebo

## Comment obtenir l'aveugle ?

Méthodes +/- complexes

- Pour des traitements médicamenteux : formes identiques dans les groupes de traitement (forme, couleur, odeur, goût, fréquence et mode d'administration)
- Pour des traitements non médicamenteux : dispositif ou appareils médicaux inactifs (appareil éteint ...), simulation de l'intervention chirurgicale

## Analyse en intention de traiter

Si les patients qui présentent des écarts au protocole (inclusion à tort, mauvaise observance, arrêt de traitement, ...) sont exclus de l'analyse, on risque d'introduire un biais par destruction de la comparabilité initiale des groupes issus de la randomisation surtout si les motifs d'arrêt sont liés au traitement

⇒ **Biais d'attrition**

Pour éviter ce biais : Analyse en intention de traiter (ou ITT)

### Principe

Tous les patients randomisés doivent être suivis :

- Jusqu'à la fin de l'essai ou jusqu'à l'observation du critère de jugement
  - Et analysés dans leur groupe de randomisation quels que soient les écarts au protocole
- ⇒ *C'est un choix volontaire d'analyse a priori*

Permet de se rapprocher de la «vraie vie».

Cependant plus il y aura d'écarts au protocole, plus la validité des résultats sera contestable.

## Pertinence de la mesure d'efficacité

### Critères de jugement

Un seul critère de jugement principal

- Critère sur lequel est jugée l'efficacité du traitement
  - Il doit être pertinent cliniquement : ce qu'il mesure a une réelle importance pour la santé du malade
  - Il doit être fiable : il évalue toujours le même état de la même façon quelles que soient les conditions de mesure et celui qui le mesure
- Plus facile quand il est objectif mais ce n'est pas toujours possible (ex: douleur) : on évaluera alors le niveau de douleur sur une échelle visuelle analogique

## Avoir une puissance statistique suffisante

Lorsqu'on analyse l'essai, on fait un test statistique pour comparer les deux groupes

Si différence significative, rejet  $H_0$  avec un risque  $\alpha$  défini *a priori*, indépendamment de  $H_1$

⇒ On conclut donc à l'efficacité du traitement

Si différence non significative, on ne conclut pas qu'on accepte  $H_0$  car Le risque  $\beta$  que l'on encourrait n'est pas connu : il varie selon la « valeur de  $H_1$  »

⇒ On préfère conclure qu'on ne rejette pas  $H_0$

- On conclut donc qu'on ne peut pas dire que le traitement est efficace
- Mais on ne conclut pas à son inefficacité

## Notion de Nombre de Sujets Nécessaire (NSN)

Dans un essai thérapeutique la taille de ce qu'on veut être capable de voir est la différence minimale  $\Delta$  entre le résultat de deux traitements par exemple que l'on veut être capable de mettre en évidence : c'est la différence jugée cliniquement intéressante

Cette capacité est mesurée par la **puissance**

En pratique, on se donne les moyens de montrer ce qu'on veut montrer:

- En calculant au préalable le NSN pour mettre en évidence une différence  $\Delta$  avec un risque  $\alpha$  et une puissance  $1-\beta$  choisis *a priori*
- Le NSN varie en fonction de ces paramètres :
  - Quand  $\downarrow$ ,  $N \uparrow$
  - Quand  $\downarrow$ ,  $N \uparrow$
  - Quand  $\downarrow$ ,  $N \uparrow$
  - Quand  $\uparrow$ ,  $N \uparrow$
  - Quand le test est bilatéral (vs unilatéral),  $N \uparrow$

La variabilité ( $\sigma$ ) et la différence attendue ( $\Delta$ ) sont celles du critère de jugement principal et «connues» à partir d'études antérieures

Le NSN est d'autant plus grand que :

- La différence que l'on veut déceler est petite
- La dispersion (ou variabilité) des résultats est grande
- Les risques d'erreur choisis sont faibles
  - En principe :  $\alpha = 5\%$  bilatéral sauf si essai contre placebo : unilatéral
  - Puissance = au moins 80%

## Considérations Éthiques

Loi Huriet-Serusclet du 20 décembre 1988 et la loi de santé publique d'août 2004 (Transposition des directives européennes) sur la protection des personnes qui se prêtent à des recherches biomédicales

Une étude peut être mise en œuvre seulement après :

- Avis favorable d'un comité de protection des personnes (CPP)
- Autorisation de l'autorité compétente (AFSSAPS) (Agence Française de Sécurité Sanitaire et des Produits de Santé)

Qui sont chargés de s'assurer que :

- La Balance Bénéfice-Risque est favorable pour tout patient inclus dans l'essai
- Elle est conduite selon une méthodologie scientifiquement rigoureuse dans un milieu médical compétent et bien équipé
- Le consentement libre, éclairé et exprès du malade est obtenu

# Evaluation des procédures diagnostiques

## Position du problème

### Décision médicale

- **Cascade de décisions** : diagnostiques (quelle démarche suivre ?), pronostiques (quelles sont les informations pertinentes ?), thérapeutiques (quel est le meilleur traitement ?)
- **Prises en situation d'incertitude**

On appelle **test** tout moyen d'obtenir une information utile au médecin pour l'assister dans sa démarche diagnostique face à un patient

- Peut être un signe clinique (ex: douleur abdominale)
- Le résultat d'un examen complémentaire (ex: ECG, scanner thoracique, dosage biologique...)

On mesure la **valeur informationnelle** (ou les qualités diagnostiques) d'un test :

- Pour diminuer l'incertitude
- Dans un but diagnostique

Toute prescription d'examen complémentaire a un **coût**:

- Financier
- Lié au désagrément ou au risque qu'il occasionne
- Du préjudice subi par le malade si le résultat de l'examen : décision inadéquate

Ce coût doit se justifier par l'espérance d'un **bénéfice**

- Appréciation plus exacte du diagnostic
- Meilleure décision thérapeutique

Donc **nécessité d'une évaluation rigoureuse des avantages et des inconvénients** du test par rapport aux tests de référence (gold standard\*)

⇒ Etude de la valeur diagnostique d'un test pour une maladie

\*Le gold standard fournit l'information sur le vrai statut pathologique des patients

## Les indices de mesure de la qualité diagnostique d'un test

### Matrice de décision

	M+ Présence de la maladie	M- Absence de la maladie	
T+ Test anormal	VP	FP	VP+FP
T- Test normal	FN	VN	FN+VN
	VP+FN	FP+VN	N

VP = vrais négatifs VP = vrais positifs

Un test parfait :

- Détecte tous les cas de la maladie : FN=0, pas de Faux Négatifs
- Ne détecte que les cas de maladie : FP=0, pas de Faux Positifs

### Caractéristiques intrinsèques du test pour la maladie

- **Sensibilité**

Probabilité que le test soit positif si le sujet est atteint de la maladie  
Taux de vrais positifs parmi les malades

$$Se = P_{M+}(T+) = \frac{VP}{VP + FN}$$

Varie entre 0 et 1

**Exprime l'aptitude du test à détecter la maladie**

Si Se = 1 :

- ⇒ Le test est toujours + dans la maladie
- ⇒ La négativité du test exclut la maladie

- **Spécificité**

Probabilité que le test soit négatif si le sujet n'est pas atteint de la maladie  
Taux de vrais négatifs parmi les non malades

$$Sp = P_{M-}(T-) = \frac{VN}{VN + FP}$$

Varie entre 0 et 1

**Exprime l'aptitude du test à ne diagnostiquer que la maladie**

Si Sp = 1 :

- ⇒ Le test n'est + qu'en présence de la maladie : pathognomonique de la maladie (signe qu'on ne retrouve que dans cette maladie)



## Indices globaux

Se et Sp varient en sens inverse

Donc difficile de choisir entre deux examens

- L'un sensible et peu spécifique
- L'autre peu sensible et spécifique

Nécessité d'utiliser des indices globaux : les rapports de vraisemblance Positif et Négatif

- **Le rapport de vraisemblance Positif**

Rapport de la probabilité conditionnelle du test positif quand la maladie est présente sur la probabilité conditionnelle du test positif quand la maladie est absente

$$RV+ = \frac{Se}{1 - Sp}$$

Varie de 0 à + l'infini

La valeur diagnostique d'un résultat positif est d'autant plus grande que RV+ est grand.

- **Le rapport de vraisemblance Négatif**

Rapport de la probabilité conditionnelle du test négatif quand la maladie est présente sur la probabilité conditionnelle du test négatif quand la maladie est absente

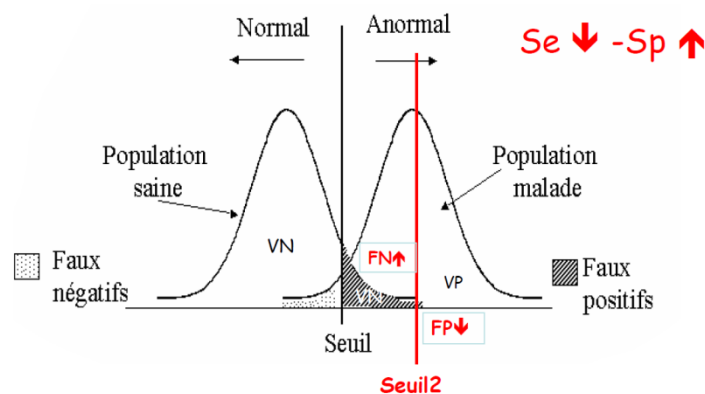
$$RV- = \frac{1 - Se}{Sp}$$

Vaut 0 si Se = 1

La valeur diagnostique d'un résultat négatif est d'autant plus grande que RV- est petit et proche de 0.

## Définition de la positivité du test

Les critères de positivité du test lorsque son résultat n'est pas dichotomique sont fondamentaux pour interpréter les valeurs de Se et de Sp.



## Caractéristiques extrinsèques du test pour la maladie

- **Valeur prédictive positive**

Probabilité que le sujet soit atteint de la maladie si le test est positif

$$VPP = P_{T+}(M+)$$

% de sujets malades parmi les sujets T+

- **Valeur prédictive négative**

Probabilité que le sujet ne soit pas atteint de la maladie si le test est négatif

$$VPN = P_{T-}(M-)$$

% de sujets non malades parmi les sujets T-

## Relation Caractéristiques intrinsèques / Valeurs prédictives

### Valeur Prédictive Positive = VPP

$$VPP = \frac{P(M+).Se}{P(M+).Se + (1 - P(M+)).(1 - Sp)}$$

### Valeur Prédictive Négative = VPN

$$VPN = \frac{(1 - P(M+)).Sp}{(1 - P(M+)).Sp + P(M+).(1 - Se)}$$

Les valeurs prédictives peuvent donc s'exprimer en fonction du couple sensibilité / spécificité et de la fréquence de la maladie dans la population ( $P(M)$ =prévalence de la maladie).

Exemple de l'utilisation de la VPN et de la VPP

- En banque du sang, il est fondamental d'avoir une bonne VPN : on ne cherche pas à diagnostiquer la maladie mais à l'éliminer pour éviter la contamination
- Dans un service spécialisé, le but est d'aboutir au diagnostic donc il est nécessaire d'avoir une bonne VPP

## Influence de la prévalence sur les VP

A Se et Sp constantes, les VP (ou probabilité a posteriori ou post-test) varient selon la prévalence de la maladie

Pb (M+) = Prévalence maladie (ou probabilité a priori ou pré-test)

Si la prévalence augmente de 0 à 1 (avec Se et Sp élevées)

- ⇒ VPP augmente régulièrement de 0 à 1 (très faible dans une population où la maladie est rare)
- ⇒ VPN diminue de 1 à 0 mais très faiblement

## Terminologie

Pas de condition	Probabilité de maladie <i>a priori</i> ou pré-test	Prévalence dans le contexte de population où est vu le malade	P(M)
Réalisation du test (Se – Sp)		Résultat +	
		Résultat -	
Conditionnellement à la réalisation du test	Probabilité de maladie <i>a posteriori</i> ou post-test	Valeur Prédictive positive	P(M/T)

Nomogramme de Fagan permettant d'interpréter les résultats des tests diagnostiques

