

# Statistiques inférentielles et épidémiologie analytique : mesure des risques et puissance en épidémiologie

**Objectifs :** en gros ce qu'il faut savoir pour le cc

1. Savoir définir les termes suivants : Epidémiologie analytique, hypothèse, inférence statistique, Risque relatif, facteurs de risque, intervalle de confiance du risque relatif
2. Savoir décrire les différents types d'enquêtes analytiques
3. Expliquer les inconvénients des enquêtes d'observation (citer les trois grands types de biais)
4. Définir le principe des enquêtes de cohortes
5. Savoir comment calculer un risque relatif
6. Définir le principe des enquêtes cas-témoins
7. Savoir comment calculer un Odds-Ratio et citer ses conditions d'utilisation
8. Citer les avantages et les inconvénients des enquêtes de cohortes et des enquêtes cas-témoins
9. Expliquer le principe de puissance en épidémiologie
10. Citer les critères de jugement de causalité en épidémiologie

## I. Epidémiologie

### A. Définition

Epidémiologie : Etude de la fréquence des pathologies. Plus généralement on va étudier l'étude de la distribution des états de santé et de leurs déterminants.

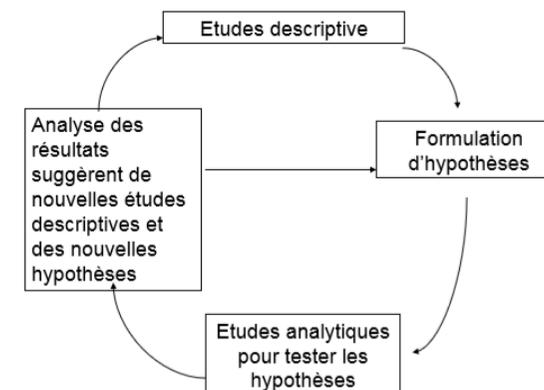
## B. Epidémiologie descriptive et analytique

	Epidémiologie Descriptive	Epidémiologie Analytique (=Etiologique=Explicative)
Définition	Description de la distribution spatiale ou temporelle des états de santé dans les populations humaines. → Mesure de la <b>prévalence</b> et de l' <b>incidence</b>	Identification, quantification, et interprétation du <b>lien de cause à effet</b> entre une exposition (facteur de risque) et un état de santé (maladie) → Mesure d'association
Répond aux questions	<b>Qui ?</b> (hommes ?, femmes ?, ouvriers ?, cadres ?,...) <b>Quand ?</b> (Depuis longtemps ?, récemment ?, ...) <b>Où ?</b> (Zone urbaine ?, campagne ?, ...).	<b>Comment ?</b> <b>Pourquoi ?</b>
	Permet de générer des hypothèses	Permet de tester des hypothèses

Prévalence : Nombre de malades à un moment donné dans une population (comme une photo faite à un **instant t**).

Incidence : Nombre de nouveaux cas dans un laps de temps, on observe sur la **durée**.

## C. Récap



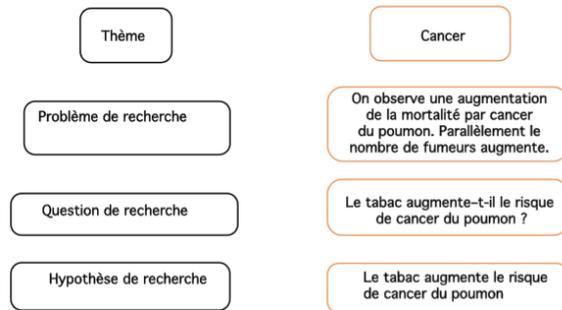
## II. Principes importants en Statistiques

### A. Hypothèse

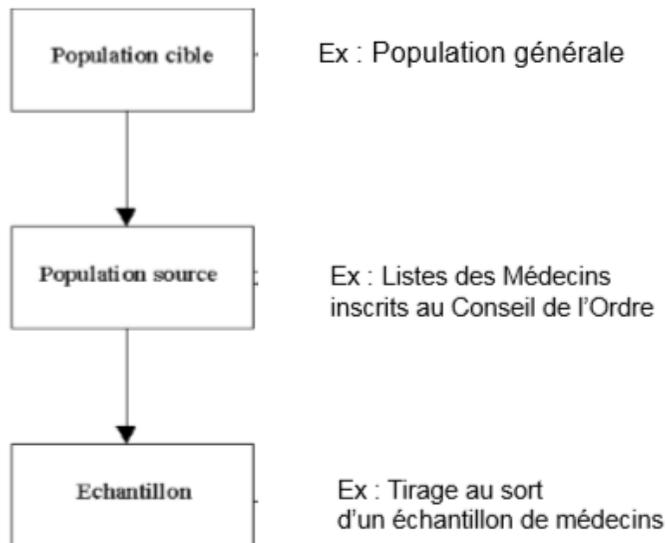
**Définition** : **Élément de base de toute recherche**. Elle doit donc être **claire, précise** et **courte**. Elle prédit une relation entre deux variables et prolonge une problématique de recherche.

⚠ l'hypothèse est **toujours rédigée avant d'exécuter l'expérience** puisque cette dernière est faite dans le but de vérifier l'hypothèse ! De plus on ne fait pas d'étude analytique sans avoir posé d'hypothèse (c'est indispensable). ⚠

**Exemple** :



### B. Inférence en statistiques



➤ **Inférence statistique** : Tirer des conclusions sur une population à partir des données d'un échantillon.

↳ Sous certaines conditions on fait l'hypothèse que le résultat obtenu à partir d'un échantillon représentatif d'une population source est généralisable à cette population. Ainsi, il est possible d'estimer la vraie valeur inconnue d'un paramètre d'une population à partir d'un échantillon.

➤ **Echantillon** : Partie de la population constituée de manière encadrée (*La manière dont un échantillon a été constitué doit toujours pouvoir être décrite*) qui permettra de réaliser des **test** d'hypothèse (*On tirera des conclusions sur une population à partir des données d'un échantillon*).

↳ Le **nombre d'individus** et certaines de leurs caractéristiques sont à préciser pour justifier de la **représentativité de l'échantillon**.

### C. Risques Relatifs

➤ **Risque** : Probabilité d'être atteint d'une maladie

➤ **Facteurs de risque** : Tous facteurs influant de façon péjorative ou favorable sur la survenue ou sur l'évolution d'un problème de santé.

➤ **Risque Relatif (RR)** : Risque d'être malade quand on est exposé au facteur étudié, par rapport au risque d'être malade quand on n'est pas exposé à ce facteur.

↳ Le RR est le rapport entre l'incidence de la maladie chez les sujets exposés et l'incidence de la maladie chez les non-exposés.

$$RR = \frac{\text{Incidence de la maladie chez les sujets exposés}}{\text{Incidence de la maladie chez les non-exposés}}$$

$$\text{Ex : } RR = \frac{\text{Incidence du cancer du poumon chez les fumeurs}}{\text{Incidence du cancer du poumon chez les non-fumeurs}}$$

↳ La valeur du RR permet d'évaluer **la force d'association**.

↳ Si le facteur étudié ne joue pas un rôle causal, il ne doit pas exister de différence d'incidence entre les sujets exposés et les non exposés. Dans ce cas, le RR doit être égal à 1. Si le RR > 1, cela signifie que la présence du facteur entraîne une augmentation de la probabilité d'apparition de la maladie.

Ex : Un risque relatif de 3 (RR=3) signifie que les sujets exposés au facteur de risque ont une probabilité 3 fois plus élevée d'avoir la maladie que les non-exposés.

➤ **Intervalle de confiance du Risque Relatif** : Intervalle à x% dans lequel la vraie valeur du paramètre étudié à x% de chance de se trouver. L'intervalle de confiance à 95 % est l'intervalle le plus couramment utilisé.

Ex : 1) RR= 5,2 ; IC-95% [2,5-8,3] -> la borne (supérieure ici mais peut être inférieure selon les situations) est éloignée de la valeur 1, l'association entre l'exposition et la maladie est dite "significative", avec  $p < 0.05$

2) RR= 1,5 ; IC-95% [0,5-5,3] -> l'intervalle de confiance contient la valeur 1, l'association entre l'exposition et la maladie est dite "non significative", ce qui veut dire que la valeur de « p » est supérieure à 5 %.

↪ Dans la mesure où les enquêtes analytiques sont réalisées à partir d'échantillons théoriquement représentatifs des populations concernées, on comparera non **pas** l'incidence véritable dans les groupes mais une **estimation de cette incidence**. Aussi, l'analyse peut-elle être complétée par des méthodes comme l'estimation des limites de confiance.

La notion d'IC repose sur l'idée suivante: si la même étude était réalisée sur un échantillon différent de patients, les résultats ne seraient pas identiques, mais seraient eux aussi proches du résultat véritable, qui reste inconnu. L'IC fournit la fourchette de valeurs à l'intérieur de laquelle "nous sommes certains à 95% de trouver la valeur étudiée pour la population considérée".

↪ L'IC donne une mesure de la précision des résultats d'une étude pour pouvoir les **extrapoler** à tous les patients du même type (en théorie la valeur exacte est contenue dans 95% de la fourchette).

### III. Les études statistiques

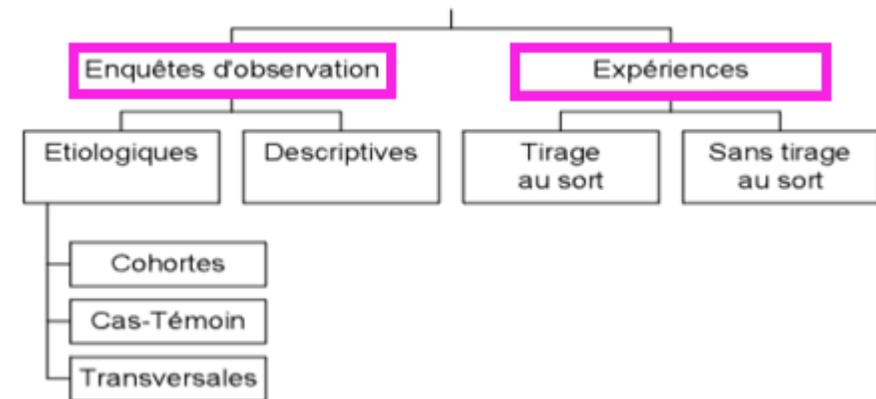
#### A. Différents types d'étude

**Rappel!** : Les études analytiques permettent de **tester des hypothèses** (ex : *le tabac augmente le risque de développer un cancer du poumon*). Elles cherchent à mettre en évidence la relation de cause à effet entre l'exposition à un **facteur de risque** et la **survenue d'une maladie**.

1) on compare : Un groupe exposé au facteur étudié (ex : *tabac*) et un groupe non-exposé à ce facteur (*non-fumeurs*)

2) On estime le Risque Relatif

**Il existe deux grands types d'études analytiques : les Etudes Expérimentales et les Etudes Observationnelles.**



#### B. Problèmes des enquêtes

1) **Choix de la population** : Biais de sélection

2) **Mesure de l'exposition et de la maladie** : Biais de mesure

3) **Pas de tirage au sort de l'exposition au facteur étudié** : Problème de comparabilité des groupes

**Les Biais** : (ou erreur systématique) Distorsion de l'estimation de la mesure d'une association entre exposition à un facteur de risque et la survenue d'une maladie.

↪ Les biais entraînent : Une surestimation de l'effet ; Une sous-estimation de l'effet ; Même un effet inverse (facteur de risque au lieu d'un facteur protecteur).

Il est donc important de prévenir les biais !

**Les différents types de biais :**

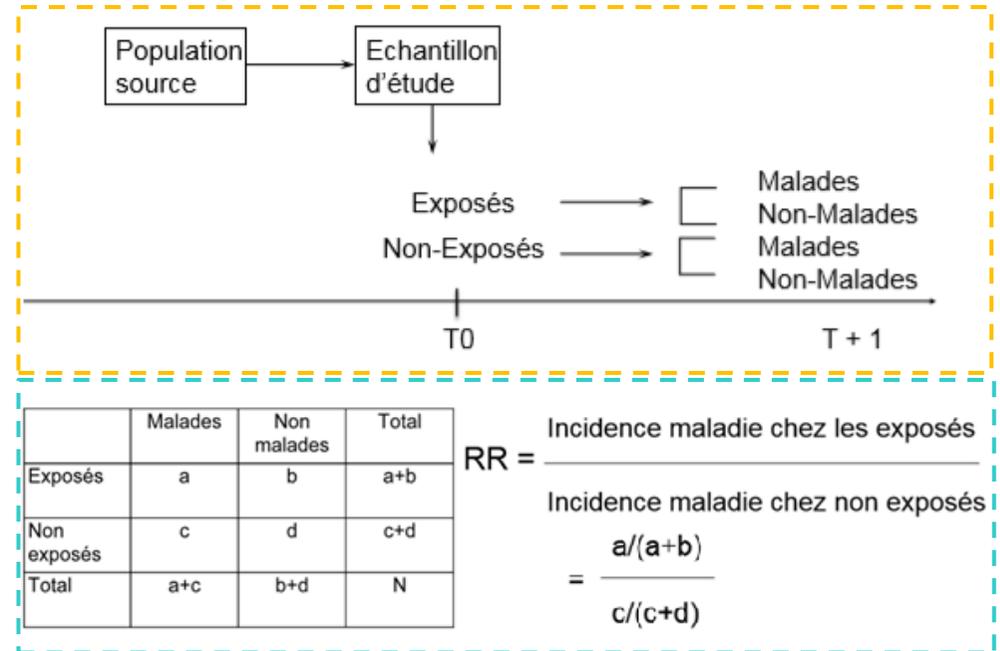
- **Biais de sélection** : Survient lors de la constitution des groupes. Il vient de la sélection préférentielle liée au statut exposé ou malade.
  - ↳ **Quand ?** Survient lors de la constitution de l'échantillon ou des groupes ou pendant la réalisation de l'enquête (non-répondants, ...)
  - ↳ **Quelle conséquence ?** Echantillon non représentatif de la population source. On ne pourra pas « extrapoler » les résultats.
  - ↳ **Comment l'éviter ?** Tirage au sort de l'échantillon. Bien choisir sa population source.
- **Biais de Mesure** : (ou d'information ou de classement) Survient lors de la mesure de l'exposition et de la maladie.
  - ↳ **Quand ?** Lors de la mesure de l'exposition et/ou de la maladie.
  - ↳ **Quelle conséquence ?** Sur ou sous-estimation du risque relatif.
  - ↳ **Comment l'éviter ?** Bien distinguer malades et non malades. Qualité de la mesure de l'exposition
- **Biais de confusion** : Concerne la non comparabilité des groupes pour différents facteurs autres que le facteur étudié (ex : âge, sexe). Liés à un tiers facteur ou facteur de confusion.
  - ↳ **Quand ?** Au moment de l'analyse statistique des données. Quand les groupes ne sont pas comparables (Ex: pour l'âge). Présence d'un facteur dit « de confusion » (Ex: l'âge).
  - ↳ **Quelle conséquence ?** Peut produire une association statistique même si le facteur étudié initialement et la maladie sont indépendants. Abouti à une « fausse » association.
  - ↳ **Comment l'éviter ?** Avant l'enquête : Appariement (Ex: Population de même âge). Au moment de l'analyse statistique : Standardisation, Ajustement (analyses multivariées).

**C. Enquête de cohorte**

**Définition** : L'enquête de cohorte mesure l'exposition à un facteur de risque chez **des individus sains** au cours du temps et s'intéresse à la **survenue ou non de la maladie**.

**Caractéristique** : Enquête **prospective**.

**Objectif** : Comparer l'incidence d'une pathologie chez des sujets exposés à sa valeur chez des sujets non exposés pris comme témoins.



Avantage	Inconvénients
<ul style="list-style-type: none"> <li>👉 Estimation directe du risque relatif</li> <li>👉 Meilleure adaptation de la planification de l'étude aux objectifs recherchés</li> <li>👉 Bon contrôle des biais</li> <li>👉 Evaluation de l'influence du facteur étudié sur l'apparition de plusieurs pathologies</li> </ul>	<ul style="list-style-type: none"> <li>👉 Etude longue</li> <li>👉 Etude coûteuse</li> <li>👉 Etude difficilement reproductible</li> <li>👉 Possibilité de perdus de vue</li> <li>👉 Effectif important au départ</li> <li>👉 Nécessité population stable</li> <li>👉 Nécessité maladie incubation courte</li> </ul>

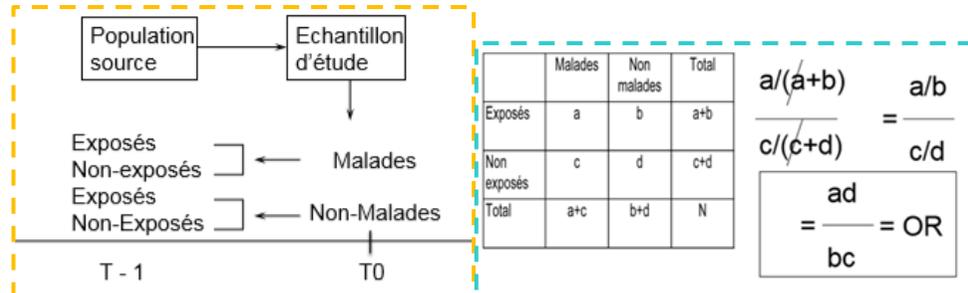
Ex : La fameuse étude de Doll et Hill #LeTabacC'estmaaaaaal

## D. Enquêtes cas-témoins

**Définition** : L'enquête cas-témoins compare des **sujets atteints de la maladie** étudiée (les cas) à des sujets indemnes de cette maladie (Témoins) quant à leur exposition passée à un facteur de risque.

**Caractéristique** : Enquête **rétrospective**.

**Objectif** : Comparer la fréquence de l'exposition antérieure chez des sujets atteints par une pathologie ("les cas") et chez des sujets non atteints pris comme témoins.



**Remarque** : On ne peut pas calculer le RR car on ne peut pas calculer le taux d'incidence. Mais on peut calculer l'Odds-Ratio. Bonne approximation du Risque Relatif si la maladie est rare.

Avantage	Inconvénients
<ul style="list-style-type: none"> <li>👉 Enquête de courte durée</li> <li>👉 Coût modéré</li> <li>👉 Possibilité d'itération</li> <li>👉 Indiquée pour l'étude maladies rares</li> <li>👉 Possibilité d'évaluer plusieurs facteurs de risque pour une seule maladie</li> </ul>	<ul style="list-style-type: none"> <li>👉 Evaluation indirecte du risque relatif</li> <li>👉 Biais plus difficile à contrôler</li> <li>👉 Sélection délicate des témoins</li> <li>👉 Nécessité d'une bonne information diagnostique au départ</li> </ul>

Ex : enquête de Doll et Hill à l'hôpital

## E. Enquêtes transversales

**Définition** : L'information recueillie concerne l'existence de l'exposition et la maladie au moment de l'enquête.

**Caractéristique** : Seulement pour un moment donné.

**Remarques** : Pas de mesures de l'incidence, pas d'études dynamiques. Mal adaptée à des études analytiques (ou explicatives)

## IV. Paramètres d'une étude

### A. Risques

- ✦ **Risque  $\alpha$**  = risque de **première espèce** = on **rejette H0 vraie** : c'est la probabilité de conclure à une relation entre l'exposition (à un facteur donné) et la maladie, alors qu'en fait cette relation n'existe pas. En général,  **$\alpha=5\%$** . C'est le seuil au-delà duquel on convient de rejeter H0
- ✦ **Risque  $\beta$**  = risque de **deuxième espèce** = **rejeter H1 vraie** : c'est la probabilité de conclure à l'absence de relation entre l'exposition (à un facteur donné) et la maladie alors qu'en fait cette relation existe. En général,  **$\beta=20\%$**

### B. Puissance et NNT

**Puissance du test** : **probabilité de détecter une différence qui existe entre deux groupes**. C'est aussi  **$(1-\beta)$**  n supposant que le risque relatif réel dans la population-cible soit différent de 1, la puissance  $(1 - \beta)$  est la probabilité de trouver un estimateur de ce risque significativement différent de 1. En général, puissance=80%.

⚠ Ne pas mettre en évidence un effet dans un échantillon ne signifie pas qu'aucun effet n'existe en réalité.

⚠ Toute étude doit faire état du calcul préalable du nombre de patients nécessaire et de la puissance de l'essai

### C. Déterminer la taille de l'échantillon

Il est important de spécifier les 5 points suivants :

- 1) Le niveau de signification exigé  **$\alpha$**  ( $=0,05$ )
- 2) La **puissance** du test  $(1-\beta)$  ( $\beta$  se situe entre 0,10 et 0,20)
- 3) La **fréquence relative d'exposition** parmi les non-exposé, dans la population cible ( $P_0$ ).
- 4) Le **risque relatif** attendu

## D. Critère de jugement d'une relation de cause à effet

### 1) L'évaluation de la séquence dans le temps

> L'exposition à un agent causal précède la maladie.

Ex : *On admet la séquence, dans le temps, du tabagisme et du cancer du poumon*

### 2) La force de l'association

> Précision avec laquelle une variable peut permettre de prédire l'autre

Ex : *tous les sujets exposés tombent malades*

### 3) La spécificité de la cause et de l'effet

> Présence de la cause étudiée chez tous les malades

Ex : *La relation tabac et cancer pulmonaire est spécifique ; dans neuf cas sur dix, le malade est un fumeur.*

### 4) La relation de type "dose-effet"

Plus l'exposition au facteur est importante, plus le risque de la maladie augmente.

Ex : *tabac et cancer du poumon : le risque de cancer augmente avec la durée du tabagisme et le nombre de cigarettes fumées.*

### 5) L'évaluation de la constance de l'association et de la reproductibilité

> Les diverses méthodes et approches (rétrospectives et prospectives) conduisent aux mêmes conclusions

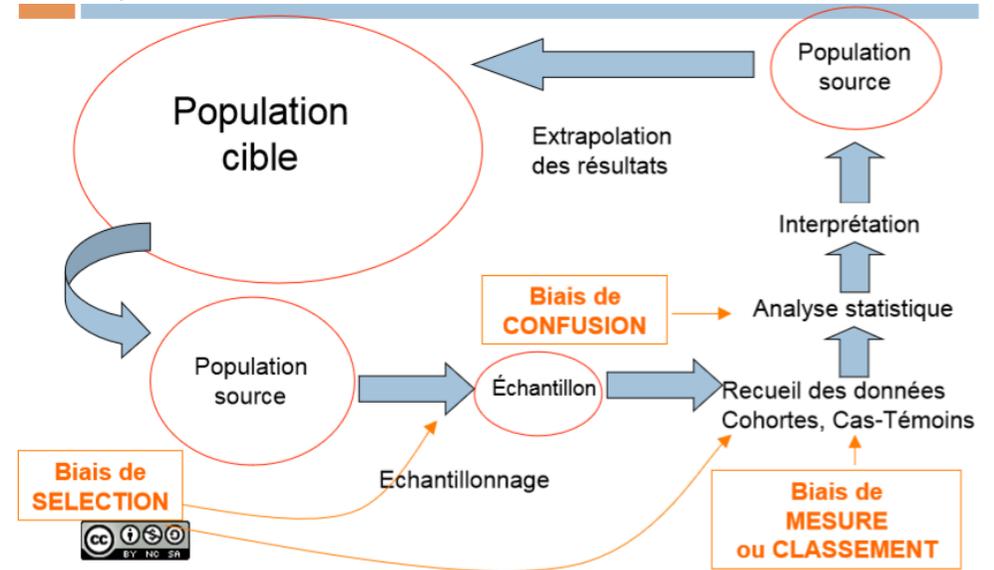
Ex : *tabac et cancer du poumon : plusieurs études différentes ont confirmé la constance de la relation tabagisme-cancer du poumon.*

### 6) Plausibilité biologique

> Cohérences du mécanisme d'action du facteur par rapport aux connaissances

Ex : *tabac et cancer du poumon: la relation entre tabac et cancer du poumon est plausible, explicable.*

## E. Synthèse



*Voilou ! © La fiche est complète par rapport au diapo ! Il y a tout pour les plus paniqués. Mais vraiment ne vous prenez pas la tête le Pr. Pradier dit en cours « n'apprenez que ce qui est souligné dans le diapo » et sur la fiche j'ai mis en couleurs tout ce qui est souligné de sa diapo. Je vous sors en même temps une fiche récap avec que les infos à retenir. Bossez bien ces qcm sont offerts au concours :P (dédicace à Emma lpb <3 sois fière de toi un peu)*