

Statistiques Inférentielles & Épidémiologie Analytique : mesure des risques et puissance



Objectifs

- Savoir définir les termes : épidémiologie analytique, risque relatif, facteurs de risque, intervalle de confiance du risque relatif
- Savoir décrire les différents types d'enquêtes épidémiologiques analytiques
- Expliquer les inconvénients des enquêtes d'observation (citer les trois grands types de biais)
- Définir le principe des enquêtes de cohorte et des enquêtes cas-témoin
- Expliquer le principe de puissance en épidémiologie

I) Épidémiologie

Définition de l'épidémiologie : C'est l'étude de la fréquence des pathologies (Qui et combien ?). Plus généralement, on va étudier la distribution des états de santé et de leurs déterminants (facteurs pronostics).

Epidémiologie descriptive / analytique :

	Epidémiologie descriptive	Epidémiologie analytique = étiologique = explicative
Définition	Description de la distribution spatiale et temporelle des états de santé dans les populations humaines → Mesure prévalence et incidence	But : rechercher les causes Identification, quantification, interprétation du lien de cause à effet entre une exposition (facteur de risque) et un état de santé (maladie) → Mesure association (difficile à interpréter)
Répond aux questions...	Qui ? (Hommes ? Femmes ?) Quand ? (Depuis longtemps ? Récemment ?) Où ? (Uniformément réparti ? Concentré ?)	Comment ? Pourquoi ?
But	GENERER des hypothèses	TESTER des hypothèses

(A propos du schéma précédent) On décrit : au cours du temps...

- Le nombre de cancers de l'estomac diminue

- Le nombre de cancers des poumons et des VADS augmente

Qui ? Les hommes

Quand ? entre 1950 et 1989

Où ? En France

On voit que la consommation de cigarettes augmente aussi.

Hypothèse : il y a une relation entre l'augmentation de la consommation de cigarettes et l'augmentation du nombre de cancers des poumons

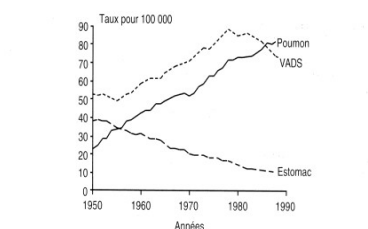
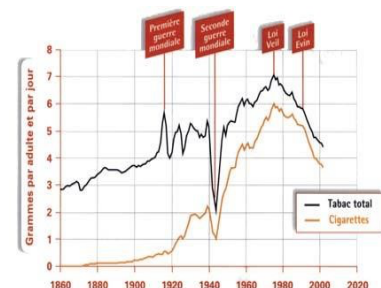
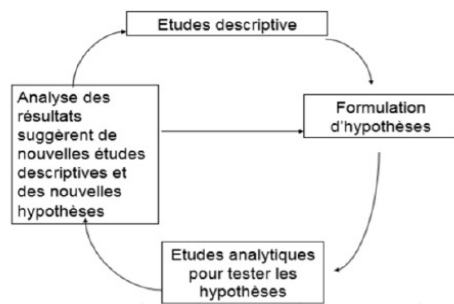


Figure 1-3 Variations de la mortalité masculine par cancer de l'estomac, du poumon et des voies aérodigestives supérieures (VADS) en France de 1950 à 1989. (Hill et al. 1992)





Indicateurs de fréquence utilisés : +++

Prévalence : nombre de malades à un instant t dans une population

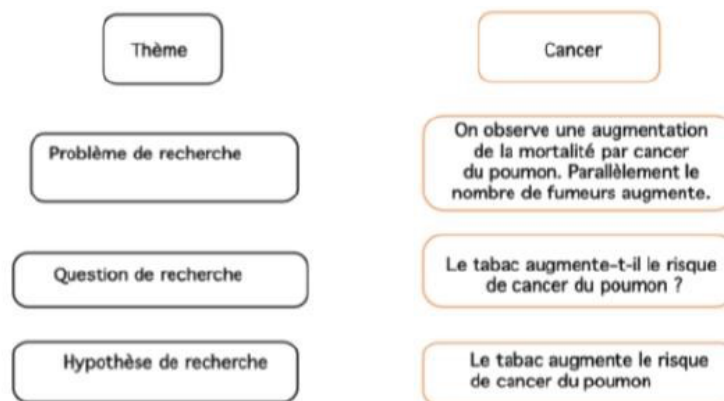
Incidence : nombre de nouveaux cas dans un laps de temps, observé sur une durée (dynamique)

II) Principes importants en statistiques

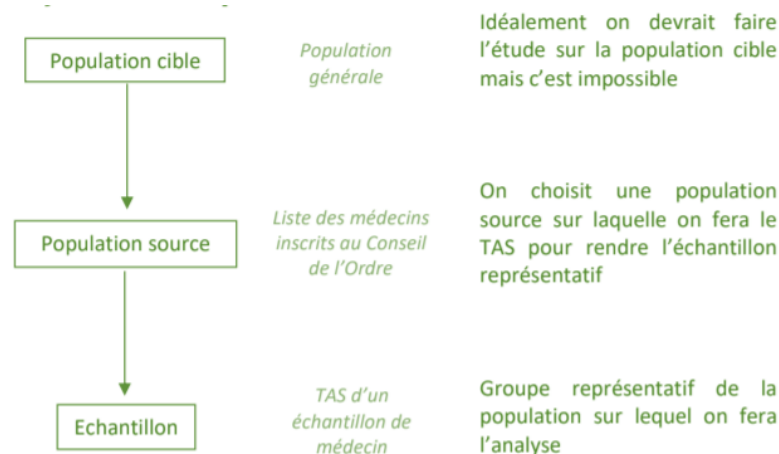
L'hypothèse :

C'est l'**élément de base** de toute recherche. Elle doit 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. En effet, **l'expérience sert à vérifier l'hypothèse**. Il est indispensable de poser l'hypothèse pour faire l'analyse analytique. Ce n'est pas une question, c'est une prédiction.

Exemple :



L'inférence statistique :



L'inférence statistique est le fait de tirer des conclusions sur une population à partir des données d'un échantillon.

→ On fait l'hypothèse que le résultat obtenu à partir d'un échantillon représentatif d'une population source soit généralisable à cette population source

Ainsi, on estime la vraie valeur inconnue d'un paramètre dans une population.

Echantillon : partie de la population constituée de manière **encadrée** (cette constitution doit toujours pouvoir être décrite) qui permettra de réaliser des tests d'hypothèse (puis de tirer des conclusions sur une population).

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

Le risque relatif :

Risque : probabilité d'être atteint d'une maladie

Facteur de risque : facteur influant de façon péjorative ou favorable sur la survenue ou l'évolution d'un problème de santé. Il y en a des protecteurs mais en général ils augmentent la probabilité d'avoir une maladie.

Risque relatif (RR) +++ : risque d'être malade quand on est exposé au facteur étudié, par rapport au risque d'être malade quand on n'y est pas exposé. C'est le rapport de l'incidence de la maladie chez les sujets exposés sur l'incidence chez les non-exposés.

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

La valeur du risque relatif 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. On doit alors avoir un rapport qui donne **RR = 1**
- ⇒ Si **RR > 1**, le facteur entraîne une **augmentation de la probabilité** d'apparition de la maladie
- ⇒ Si **RR < 1**, le facteur entraîne une **diminution de la probabilité** d'apparition de la maladie

Ex : RR = 5 signifie que les sujets exposés au facteur ont 5 fois plus de risque de développer la maladie

Intervalle de confiance du risque relatif ++ : intervalle à x% dans lequel la vraie valeur du risque relatif à x% de chance de se trouver. L'intervalle de confiance à 95% est le plus couramment utilisé. La largeur de l'IC détermine la précision de l'estimation du risque relatif.

Ex : On a RR = 5,2 et IC-95% = [2,5 ; 8,3]. Cela signifie que le risque relatif a 95% de chances de se trouver entre 2,5 et 8,3.

Or, dans tous les cas, ce risque est supérieur à 1 : la borne inférieure (ici 2,5, mais selon les situations on peut s'intéresser à la borne supérieure) est supérieure à la valeur 1, qui n'est pas comprise dans l'intervalle. L'association entre l'exposition et la maladie est dite « significative », avec le degré de signification p < 0,05.

Si maintenant on a RR = 1,5 et IC-95% = [0,5 ; 2,5], l'intervalle de confiance contient la valeur 1. L'association entre l'exposition et la maladie est dite « non significative » et p > 0,05.

Dans la mesure où les enquêtes analytiques sont réalisées à partir d'échantillons théoriquement représentatifs, on compare non pas l'incidence véritable dans les groupes mais une estimation de cette incidence. Aussi, il est logique de compléter l'analyse par une estimation des limites de confiance.

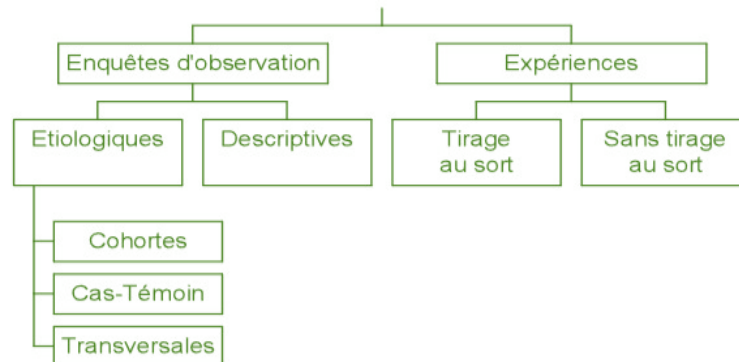
La notion de l'IC repose sur l'idée suivante : si la même étude était réalisée à partir d'un échantillon différent, les résultats ne seraient pas identiques, mais également proches du résultat véritable, qui reste inconnu. L'IC fournit la fourchette de valeurs à l'intérieur de laquelle nous sommes certains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

III) Les études statistiques

Qu'est-ce qu'une étude statistique ?

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 de la maladie.

- ⇒ On compare un groupe exposé au facteur étudié (ex : fumeurs) et un groupe non-exposé à ce facteur (ex : non-fumeurs).
- ⇒ On estime le risque relatif On distingue 2 grands types d'études analytiques : les études **expérimentales** et les études **observationnelles** ++



Idéalement on fera des études expérimentales car les groupes sont **comparables (TAS)**. Pour les études observationnelles, les groupes ne sont pas comparables donc il y aura des **problèmes de jugement de causalité**. Une étude observationnelle permet cependant d'étudier l'exposition à des facteurs de risque (ex : tabac), cela étant plus compliqué en expériences avec TAS. *Ex : un non-fumeur se retrouve dans le groupe tabac.*

Les 3 grands types de biais :

Biais (ou erreur systématique) : distorsion de l'estimation de la mesure d'une association entre l'exposition à un facteur de risque et la survenue d'une maladie. Il faut les prévenir car ils peuvent entraîner une sur/sous-estimation de l'effet ou même un effet inverse.

Biais de sélection : lors du choix de la population :

Quand ? Lors de la constitution de l'échantillon ou des groupes (sélection préférentielle liée au statut exposé ou malade) ou pendant la réalisation (non-répondants...)

Quelle conséquence ? L'échantillon devient non-représentatif de la population source ; on ne pourra plus extrapoler. Si on extrapole, il y aura des erreurs.

Comment l'éviter ? TAS de l'échantillon + bien choisir la population source

Biais de mesure (= d'information = de classement) : lors de la mesure :

Quand ? Lors de la mesure de l'exposition et/ou de la maladie

Quelle conséquence ? Sur/sous-estimation du risque relatif

Comment l'éviter ? Bien distinguer malades et non-malades : qualité de la mesure de la maladie et de l'exposition

Biais de confusion : lors de l'analyse statistique des données :

Quand ? Lors de l'analyse statistique, alors que les groupes ne sont pas comparables pour différents facteurs autres que le facteur étudié : les facteurs de « confusion »

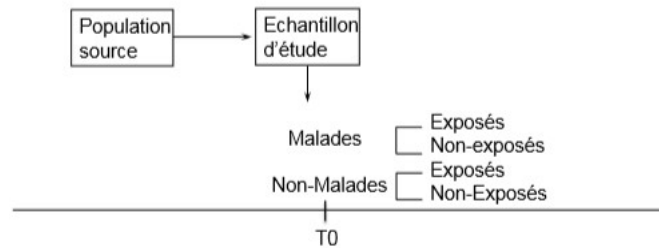
Quelle conséquence ? Une « fausse association » : association statistique même si le facteur étudié initialement et la maladie sont indépendants

Comment l'éviter ? Avant l'enquête par un appariement + au moment de l'analyse : standardisation, ajustement (utilisation d'analyses multivariées)

Enquête de cohorte +++ :

Elle mesure l'exposition à un facteur de risque chez les individus sains au cours du temps.

Elle s'intéresse à la survenue ou non de la maladie → **prospective** : elle compare l'incidence chez les sujets exposés et chez les sujets non exposés pris comme témoins.

**Avantages : +++**

- Estimation directe du risque relatif (puisque l'on compare les incidences)
- Meilleure adaptation de la planification de l'étude au cours du temps
- Bon contrôle des biais
- Evaluation de l'influence du facteur sur plusieurs pathologies

Inconvénients : +++

- Longue
- Couteuse
- Possibilité de perte de vue
- Nécessite un effectif important au départ
- Nécessite une population stable
- Nécessite une maladie à incubation courte
- Non reproductibles

	Malades	Non malades	Total
Exposés	a	b	a+b
Non exposés	c	d	c+d
Total	a+c	b+d	N

$$RR = \frac{\text{Incidence maladie chez les exposés}}{\text{Incidence maladie chez non exposés}} = \frac{a/(a+b)}{c/(c+d)}$$

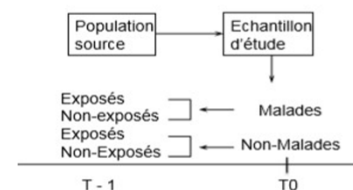
Enquête cas-témoins +++ :

Elle compare des sujets atteints de la maladie (cas) étudiée à des indemnes (témoins). Elle s'intéresse à la fréquence de leur exposition passée → **rétrospective**

Cependant, comme on ne mesure pas l'incidence, on ne peut pas calculer le risque relatif. On calcule donc **l'Odds Ratio**, qui est une bonne approximation si la maladie est rare ++

Avantages : +++

- Courte durée
- Coût modéré
- Possibilité d'itération (=reproduction)
- Indiquée pour l'étude des maladies rares
- Possibilité d'évaluer plusieurs facteurs de risque pour une maladie

**Inconvénients : +++**

- Evaluation indirecte du risque relatif
- Biais plus difficiles à contrôler (ex : biais de mémoire)
- Sélection délicate des témoins
- Nécessité d'une bonne information diagnostique au départ

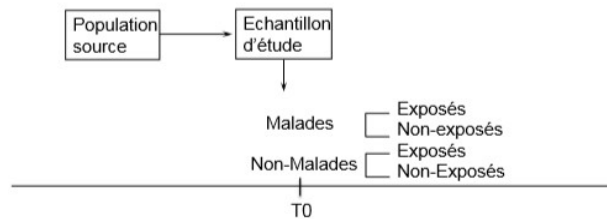
$$\frac{a/(a+b)}{c/(c+d)} = \frac{a/b}{c/d} = \frac{ad}{bc} = OR$$

	Malades	Non malades	Total
Exposés	a	b	a+b
Non exposés	c	d	c+d
Total	a+c	b+d	N

$$\text{Odds-Ratio} = OR = \frac{ad}{bc}$$

Enquête transversale :

Elle s'effectue uniquement à un **instant t**. L'information recueillie concerne l'existence ou non de l'exposition au facteur étudié, au moment même de l'enquête.

**Inconvénients :**

- Pas de mesure de l'incidence ++ : ce n'est pas une étude dynamique
- Mal adaptée à des études analytiques (ou explicatives)

IV) Paramètres d'une étude

Ne pas mettre en évidence un effet ne veut pas dire qu'il n'y a pas d'effets ++

Risques α et β , puissance :**Risque α , dit risque de première espèce :**

- Risque de rejeter H_0 alors qu'elle est vraie
- Risque de conclure à une relation entre l'exposition (à un facteur donné) et la maladie, alors que cette relation n'existe pas
- En général, $\alpha = 5\%$: c'est le seuil au-delà duquel on convient de rejeter H_0

Risque β , dit risque de deuxième espèce :

- Risque de rejeter H_1 alors qu'elle est vraie
- Risque de conclure à l'absence de relation entre l'exposition (à un facteur donné) et la maladie, alors que cette relation existe
- En général, $\beta = 20\%$

Puissance du test, $1 - \beta$:

- Elle se définit comme la probabilité de détecter une différence qui existe entre deux groupes
- On peut le traduire ainsi : en supposant que le RR réel dans la population cible soit différent de 1, la puissance est la probabilité de trouver un estimateur de ce risque significativement différent de 1
- En effet, elle est l'inverse de β , qui correspond à une absence de différence entre les 2 groupes, ou encore à trouver une différence non significative avec 1 compris dans l'IC du RR
- En général, $1 - \beta = 80\%$

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

Taille de l'échantillon :

Il faut spécifier les points suivants :

- ⇒ Le niveau de signification exigé : α
- ⇒ La puissance du test : $1 - \beta$
- ⇒ La fréquence relative d'exposition parmi les non-exposés, dans la population cible
- ⇒ Le risque relatif attendu

Critères de jugement d'une relation de cause à effet :

Evaluation de la séquence dans le temps : l'exposition à un agent causal précède la maladie.

Ex : le fait de fumer précède et provoque le fait d'avoir un cancer du poumon

Force de l'association : précision avec laquelle une variable peut permettre de prédire l'autre

Ex : tous les sujets exposés tombent forcément malades

Spécificité de la cause et de l'effet : présence de la cause étudiée chez tous les malades

Ex : le risque de cancer du poumon est spécifique : dans 9 cas sur 10, le malade fume

La relation du types « dose-effet » : plus l'exposition au facteur est importante, plus le risque de maladie augmente

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

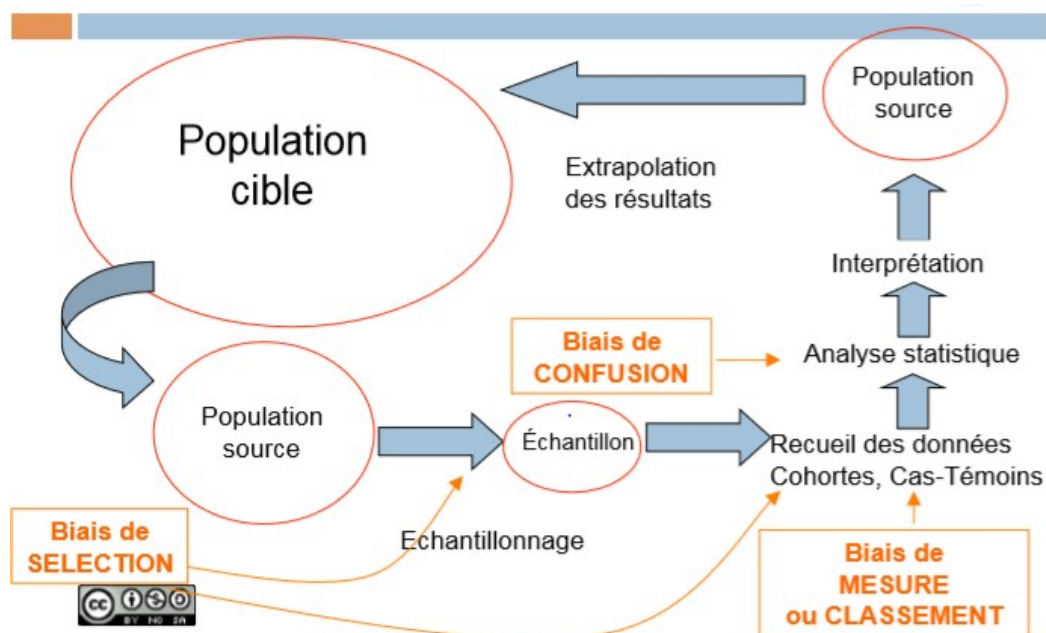
L'évaluation de la constance de l'association et de la reproductibilité : les diverses méthodes et approches (rétrospectives et prospectives) conduisent aux mêmes conclusions

Ex : plusieurs études différentes ont confirmé la relation tabagisme-cancer du poumon

Plausibilité biologique : cohérence du mécanisme d'action du facteur par rapport aux connaissances

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

V) Synthèse



😊 Ptite dédi à toi qui vient de finir cette fiche, laches rien champion(ne) ! 😊

😘 La biostat vous aime 😘

