

Épidémiologie Analytique

Liste des notions à avoir acquis à la fin du cours :

- Savoir définir les termes suivants : 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 cohortes
- Définir le principe des enquêtes cas-témoins
- Citer les avantages et les inconvénients des enquêtes de cohortes et des enquêtes cas-témoins
- Expliquer le principe de puissance en épidémiologie

Définition de l'épidémiologie analytique

Aussi appelée épidémiologie étiologique ou explicative, elle a pour but de **rechercher les causes de maladie** en identifiant, en quantifiant et en interprétant un **lien** de cause à effet entre une exposition (=facteur de risque) et un état de santé (maladie). Elle réalise donc des **mesures d'association** entre ce facteur de risque et la présence éventuelle d'une maladie

L'épidémiologie analytique répond à la question : « Comment et pourquoi une personne est atteinte d'une maladie donnée »

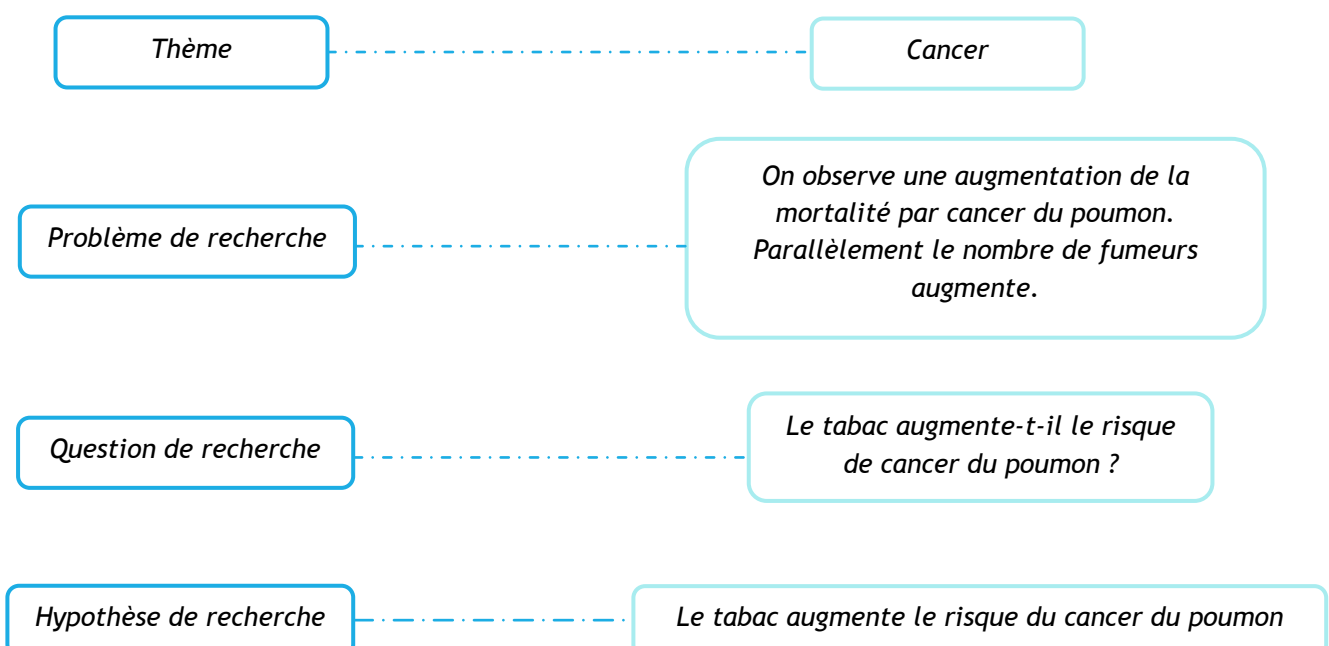
L'hypothèse

Elle est le point de départ de toutes les enquêtes épidémiologiques, **elle prédit une relation** entre deux variables qui sont :

- l'exposition a un facteur (en général le facteur de risque)
- la présence ou l'absence de maladie

Elle prolonge une problématique de recherche. C'est l'élément de base de toute recherche, il faut donc qu'elle soit **claire**, **précise** et **courte**. Elle est rédigée **avant** même d'exécuter l'expérience puisque le but de celle-ci est de vérifier ou non l'hypothèse.

Exemple du professeur :



Risque, Risque relatif et facteurs de risque

Risque :

C'est la probabilité d'être atteint d'une maladie

Facteur de risque :

Ce sont tous facteurs qui influents de façon péjorative ou favorable sur la survenue ou sur l'évolution d'un problème de santé.

Risque relatif (RR) :

C'est le 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. Ainsi il est le rapport entre l'incidence de la maladie chez les sujets exposés et l'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 risque relatif permet d'évaluer la force de l'association (= force du lien entre le facteur de risque étudié et la survenue de la maladie).

Si le $RR = 1$, le facteur étudié ne joue pas un rôle causal et il ne doit pas exister de différence d'incidence entre les sujets exposés et les non-exposés.

Si le $RR > 1$, cela signifie que la présence du facteur entraîne une augmentation de la probabilité d'apparition de la maladie. 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

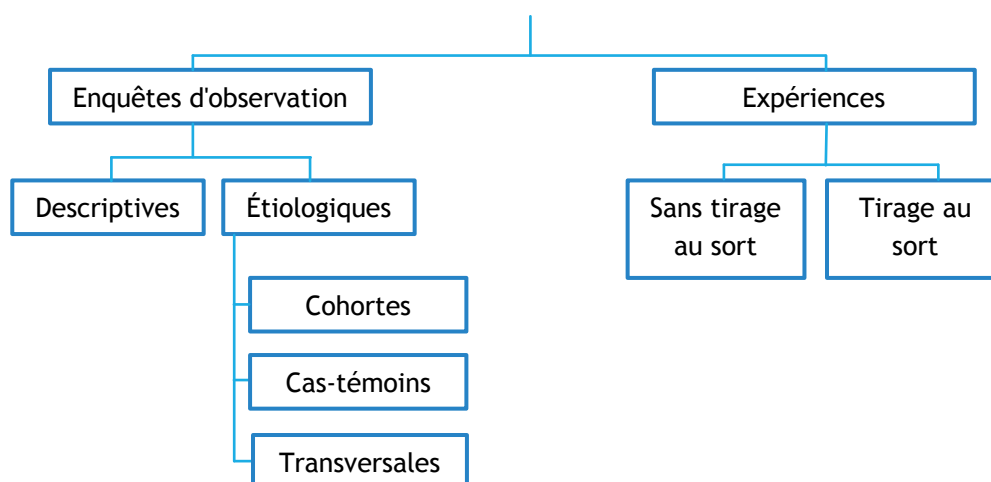
Dans la mesure où les enquêtes analytiques sont réalisées à partir d'échantillons théoriquement représentatifs des populations concernées, on comparera non pas l'incidence réelle dans les groupes mais une estimation de cette incidence puisqu'on travaille sur des échantillons.

Ceci nous amène à utiliser des intervalles de confiance pour essayer de préciser dans quel intervalle se trouve la vraie valeur du risque relatif.

A retenir ++ : L'intervalle de confiance à 95 % est l'intervalle dans lequel la vraie valeur du paramètre étudié à 95% de chance de se trouver.

Par exemple. : un $RR = 5.2$ peut être précisé avec IC-95% [2.5-8.31].

Les différents types d'études



Normalement en recherche les études sont contrôlées par l'instigateur, c'est-à-dire que tous les paramètres, les conditions dans lesquelles l'étude est réalisée et notamment quelles sont les personnes qui vont recevoir ou non les facteurs de risques sont contrôlés. Dans les expériences il y a un élément qui est très important qui est le tirage au sort, on va pouvoir déterminer par TAS les personnes qui vont être exposées ou non à un facteur. Dans la plupart des cas, comme pour les essais cliniques, le facteur étudié est l'exposition à un traitement donné. Ce traitement est donc alloué au patient par TAS, aussi appelée randomisation.

Dans les enquêtes épidémiologiques on n'utilise **pas de tirage au sort** pour la simple raison qu'on étudie des facteurs liés à l'environnement ainsi il ne serait **pas éthique** d'exposer certaines personnes pour des raisons de recherche alors qu'elles ne le souhaitent pas. Si on reprend notre exemple, il faudrait demander à des gens de fumer alors qu'ils ne sont pas fumeur et aussi demandé à des fumeurs de ne pas fumer.

Dans ces cas-là on utilise plutôt des enquêtes d'**observation** : on va avoir dans la population des personnes exposées (fumeurs) et des personnes non exposées et on va essayer d'étudier chez elles la présence ou la survenue de la maladie. Dans le cours d'aujourd'hui on s'intéresse aux enquêtes d'observation étiologiques et dans ce cadre-là nous avons 3 grands types d'enquêtes : les enquêtes de cohortes, les enquêtes cas témoins et les enquêtes transversales.

L'enquête de cohorte

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. En gros on part d'une population pas malade qu'on va suivre sur un plus ou moins long terme pour voir si elle développera une maladie donnée. C'est une enquête **prospective** (= concernant l'évolution future).

Elle compare l'incidence d'une pathologie chez des sujets exposés à sa valeur chez des sujets non exposés pris comme témoins.

Ce schéma résume bien l'enquête :



	Malades	Non-Malades	Total
Exposés	a	b	a+b
Non-Exposés	c	d	c+d
Total	a+c	b+d	N

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

Avantages	<ul style="list-style-type: none"> - Estimation directe du risque relatif - Meilleure adaptation de la planification de l'étude aux objectifs recherchés - Bon contrôle des biais - Évaluation de l'influence du facteur étudié sur l'apparition de plusieurs pathologies
Inconvénients	<ul style="list-style-type: none"> - Étude longue et coûteuse donc difficilement reproductible - Possibilité de perdus de vue (= les gens qui quittent l'étude) - Effectif important au départ (pour limiter les pertes de gens) - Nécessite une population stable et une maladie à incubation courte

→ Exemple de calcul du risque relatif

Tableau 1-1 Taux de mortalité par cancer du poumon standardisé sur l'âge, chez les hommes de 45 à 74 ans en fonction de la consommation quotidienne de tabac la plus récente. (Doll et Hill 1956)

Facteur de risque : usage des cigarettes	Non-fumeurs	Fumeurs		
		1-14 g/jour	15-24 g/jour	25 g/jour et plus
Taux d'incidence ^(a)	0,07	0,47	0,86	1,66
Risque relatif	1,0 ^(b)	6,7	12,3	23,7

(a) Nombre de nouveaux cas pour 1 000 personnes et par an

(b) Référence

$$RR = \frac{\text{Incidence maladie chez les exposés}}{\text{Incidence maladie chez les non-exposés}} = \frac{0,47}{0,07} = 6,7$$

L'enquête cas-témoins

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. C'est une enquête **rétrospective** (= *concernant une exposition dans le passé*).

Elle compare la fréquence de l'exposition antérieure chez des sujets atteints par une pathologie et chez des sujets non atteints pris comme témoins.

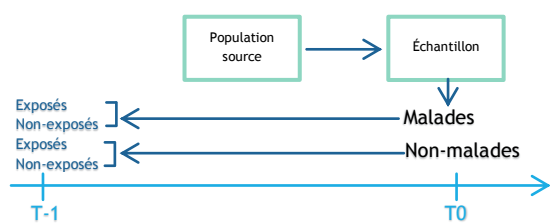
Attention : On ne peut pas calculer un risque relatif ici puisque comme on a choisi nous-même l'effectif de malades et non malades (équivalent entre eux souvent) on ne peut pas calculer des taux d'incidence.

Ainsi on utilise une autre technique d'approximation pour calculer le risque relatif : c'est l'Odd Ratio

→ Si la maladie est rare (>5%) alors on peut faire une approximation à partir du tableau ci-dessous. Elle nous permet d'avoir une méthode de calcul de l'Odd ratio assez simple pour avoir une approximation du risque relatif :

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

Ce schéma résume bien l'enquête :



	Malades	Non-Malades	Total
Exposés	a	b	a+b
Non-Exposés	c	d	c+d
Total	a+c	b+d	N

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

Avantages	<ul style="list-style-type: none"> - Enquête de courte durée donc coût modéré et possibilité d'itération (=la refaire) - Indiquée pour l'étude maladies rares - Possibilité d'évaluer plusieurs facteurs de risque pour une seule maladie
Inconvénients	<ul style="list-style-type: none"> - Évaluation indirecte du risque relatif - Biais plus difficile à contrôler (ex : biais de mémorisation des patients -> parfois les souvenirs des expositions sont flous) - Sélection délicate des témoins - Nécessité d'une bonne information diagnostique au départ (= on doit être sûr que les patients sont bien malades pour éviter les erreurs de classements)

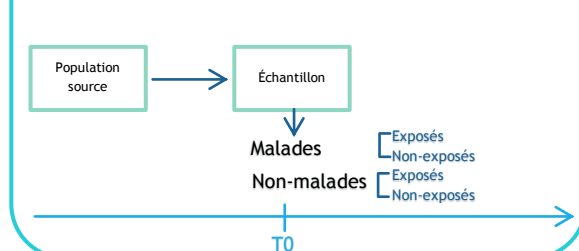
L'enquête transversale

L'information recueillie concerne l'existence de l'exposition et la maladie **au même moment** (= moment de l'enquête).

On interroge le sujet sur son exposition et sur s'il est malade mais à un temps T0. Du coup on ne mesure pas d'incidence (le prof compare l'étude à une photo instantanée), et on ne peut pas faire d'études dynamiques.

Finalement ces enquêtes sont mal adaptées à des études analytiques (ou explicatives)

Ce schéma résume bien l'enquête :



Problèmes posés par les enquêtes d'observation

Ces problèmes concernent 3 moments des enquêtes :

- Au moment du choix de la population → **Biais de sélection**
- Au moment de la mesure de l'exposition et de la maladie → **Biais de mesure (ou classement)**
- Au moment de l'analyse des données (Pas de tirage au sort de l'exposition au facteur étudié) → Problème de comparabilité des groupes et **biais de Confusion**

BIAIS DE SÉLECTION

Quand ?

Survient lors de la **constitution de l'échantillon** ou des groupes ou pendant la réalisation de l'enquête (non-répondants, refus de participation ...)

Quelle conséquence ?

L'échantillon n'est pas représentatif de la population source et on ne pourra donc pas « extrapoler » les résultats.

Comment l'éviter ?

En réalisant un tirage au sort de l'échantillon quand c'est possible, si on ne peut pas il faut bien choisir sa population source et s'assurer qu'elle ne soit pas trop biaisée

BIAIS DE MESURE (OU D'INFORMATION OU DE CLASSEMENT)

Quand ?

Lors de la **mesure de l'exposition** et/ou de la maladie

Quelle conséquence ?

On va sur ou sous-estimer le risque relatif ou l'Odd Ratio.

Comment l'éviter ?

Il faut bien distinguer malades et non malades (= avoir une bonne information diagnostique) et être très précis (limite obsessionnel) dans la mesure de l'exposition pour en avoir une de qualité.

BIAIS DE CONFUSION

Quand ?

Au moment de l'**analyse statistique** des données quand les groupes ne sont pas comparables : c'est très souvent le cas en épidémiologie puisqu'on prend des sujets exposés et non exposés de façon **non-aléatoire** qui ne sont donc pas comparables.

Cela entraîne la présence d'un facteur qu'on appelle « facteur de confusion » (*Ex : l'âge*)

Quelle conséquence ?

On risque de produire une association statistique erronée même si le facteur étudié initialement et la maladie sont indépendants : il n'y a pas de réel lien entre la maladie et le facteur mais l'analyse nous donne une association statistiquement significative.

Cela aboutit à une « fausse » association et a une désinformation de la population en déclarant un facteur dangereux alors qu'il ne l'est pas ou le contraire.

Comment l'éviter ?

- Avant l'enquête : Appariement (*Ex : dans une enquête cas témoins on prend à chaque fois un cas et un témoin du même âge*)
- Au moment de l'analyse statistique : Standardisation ou Ajustement grâce à des analyses multivariées (modèle logistique, modèle de Cox...) qui vont prendre en compte les facteurs de confusion et produire des risques relatifs et des Odd ratio ajustés/

Puissance d'un test et nombre de sujets nécessaire

Point ++ :

Ce n'est pas parce qu'on ne met pas en évidence un effet dans un échantillon qu'aucun effet n'existe en réalité.

La probabilité de détecter une différence qui existe entre deux groupes est appelée la « **puissance du test** ».

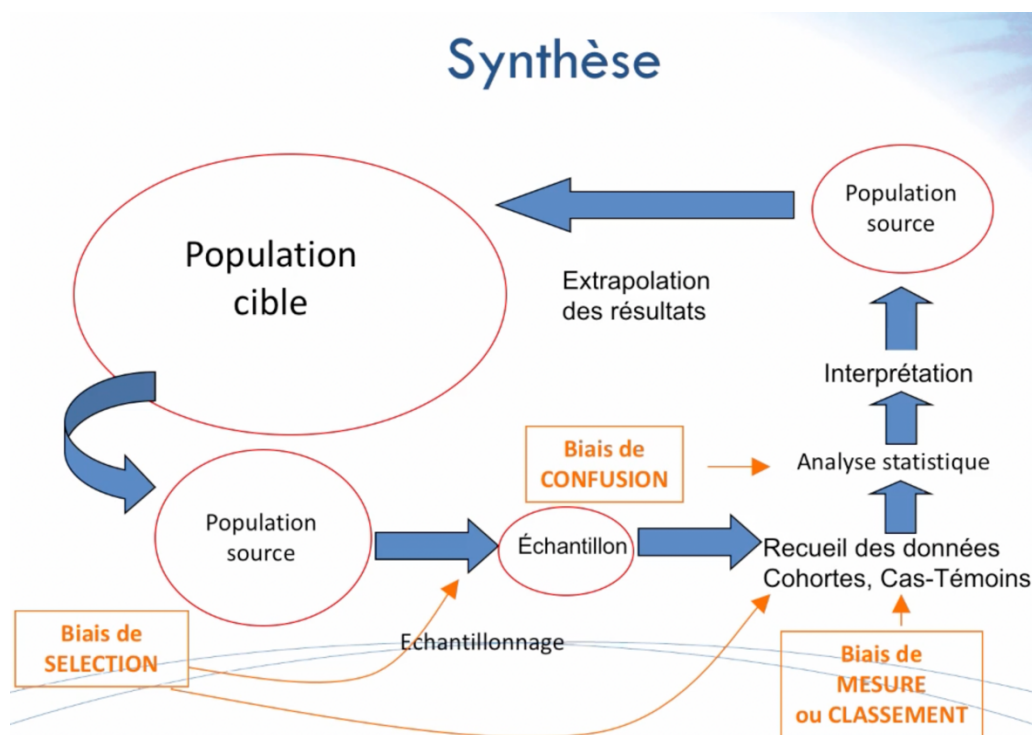
Il faut avoir une population suffisamment importante pour avoir la capacité de mettre en évidence une relation entre un facteur et une maladie si cette association existe réellement. En gros si on n'a pas assez de sujet l'étude n'est pas assez puissante pour mettre en évidence cette relation et on va conclure à tort qu'il n'y en pas.

Point ++ n°2 :

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

Donc si un jour vous devez réaliser ou analyser une étude épidémiologique il faudra toujours regarder si dans le chapitre « matériel et méthode » il existe une information sur le calcul du nombre de sujets nécessaires et si on a bien calculé la puissance. S'il n'y a pas ces informations vous pouvez émettre des doutes sur la qualité de l'étude.

Schéma de synthèse



Instant dédié :

Dédi à ma Valou, t'es trop forte je crois en toi à fond <3
 Dédi aux Manon² : Manix et Manose
 Dédi aux Camille² : Votre CT trop cool et votre tut de chimie
 Dédi à tous les vieux de biostat
 Dédi à Eléa le petit bonbon du tutorat et à ses talents en informatique
 Dédi à Baptiste aka Tonystonks
 Dédi toujours à Oscar et aussi Anis ce Vp ronéo incroyable
 Dédi à Elly et Emma <3
 Dédi à Noa, Sana et Siouar du discord et à tous ceux qui viennent aux perms de biostat
 Dédi au big daron / papy biostat avec ses 47 ans j'ai nommé Glyc'oliv