

Épidémiologie Analytique

Le cours n'a pas changé par rapport à l'année dernière mais j'ai rajouté qlq détails et explications :)

Notions à acquérir 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



I. 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 »

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	<u>But</u> : 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

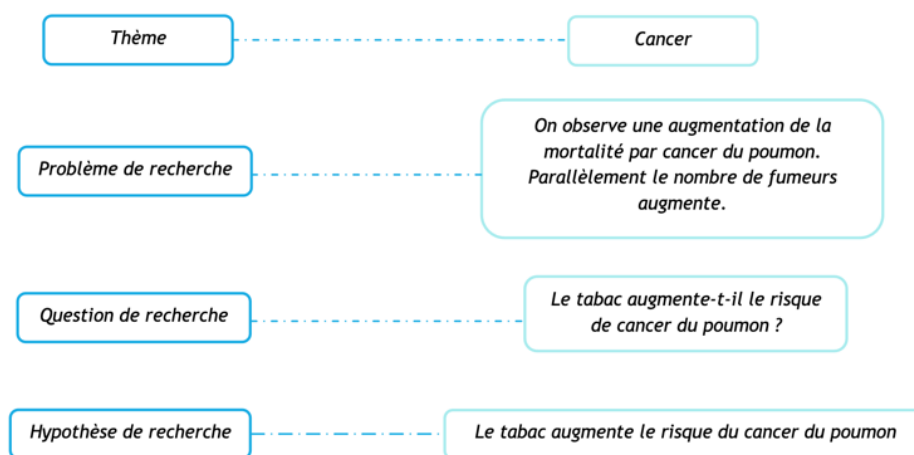
II. L'hypothèse

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

- l'exposition à 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 :



III. Principes importants en statistiques

Tut'def

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.

Échantillon : 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

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

Je rappelle juste que l'incidence d'une maladie c'est le nombre de nouveaux cas de cette maladie observés sur une période donnée

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.
- ➔ Si le $RR < 1$, cela signifie que le facteur entraîne une diminution de la probabilité d'apparition de la maladie

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

Intervalle de confiance du risque relatif

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.

⚠ On retient bien que l'intervalle de confiance 95 % est l'intervalle dans lequel la vraie valeur du paramètre étudié à 95% de chance de se trouver.

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

Cela signifie que le risque relatif a 95% de chances de se trouver entre 2,5 et 8,3.

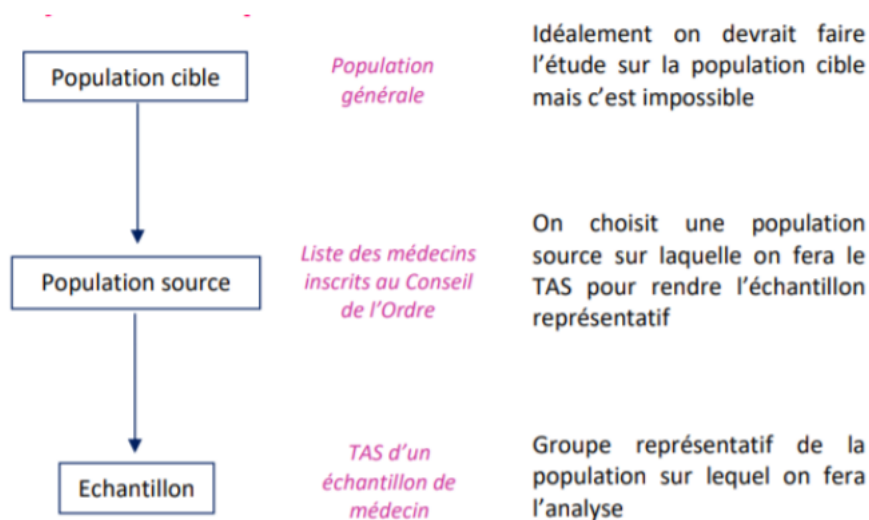
Or, dans tous les cas, ce risque est supérieur à 1 : la borne inférieure (ici 2,5, mais selon les situations on peut s'intéresser à la borne supérieure) est supérieure à la valeur 1, qui n'est pas comprise dans l'intervalle. L'association entre l'exposition et la maladie est dite « significative », avec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

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.

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

IV. Types d'études

Qu'est-ce qu'une étude statistique ?

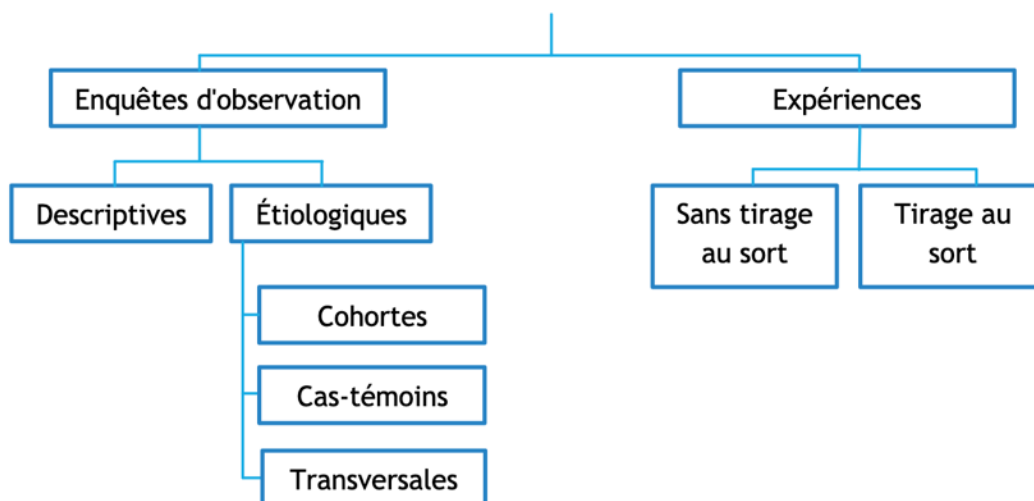
Les études analytiques permettent de **tester des hypothèses** (ex : le tabac augmente le risque de développer un cancer du poumon). Elles cherchent à mettre en évidence la relation de cause à effet entre l'exposition à un facteur de risque et la survenue de la maladie.

✍ 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**.

Les différents types d'étude



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 appelé **randomisation**. *(En gros randomisation = tirage au sort)*

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** (non-fumeurs) 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 :

On rappelle la formule : $RR = \frac{\text{Incidence chez les exposés}}{\text{Incidence chez les non-exposés}}$

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

Ici, on réalise donc le calcul suivant grâce aux données du tableau :

$RR = \frac{0,47}{0,07} = 6,7$. C'est important de savoir faire ces petits calculs car ils ne sont pas très compliqués et ils peuvent tomber le jour j, je vous en mettrai dans mes qcm 😊

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** (= les témoins) quant à leur exposition passée à un facteur de risque. C'est une enquête **rétrospective** (= concernant une exposition dans le passé). (*Attention on ne confond pas prospective=futur et rétrospective=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 :

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

*Hyper important de comprendre que l'odd ratio est une **approximation** du RR.*

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)

On retient bien que :

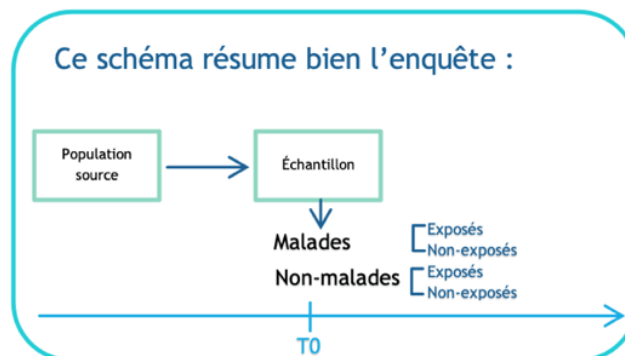
- *Enquête de cohorte → On peut **directement** calculer le RR (d'ailleurs ça peut vous être demandé donc à savoir !)*
- *Enquête cas témoin → On ne peut pas calculer directement le RR donc on calcule uniquement une **approximation** de celui-ci grâce à l'odd ratio (calcul à savoir aussi).*

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 a un temps T₀. Du coup on ne mesure pas d'incidence (*le prof compare l'étude a une photo instantanée*), et on ne peut pas faire d'études dynamiques.

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



V. Les biais

On peut rencontrer des problèmes dans les enquêtes d'observation à trois moments :

- 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 (par exemple s'il n'y a 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...).

Quelles conséquences ?

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.

Quelles conséquences ?

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** » (*Exemple : l'âge*)

Quelles conséquences ?

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 (*Exemple : 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 (*Exemple : modèle logistique, modèle de Cox...*) qui vont prendre en compte les facteurs de confusion et produire des risques relatifs et des Odd ratio ajustés.

Ici on fait bien attention à différencier les biais, quand est ce qu'ils tombent, les conséquences etc car c'est souvent demandé

VI. Puissance d'un test et nombre de sujets nécessaires

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

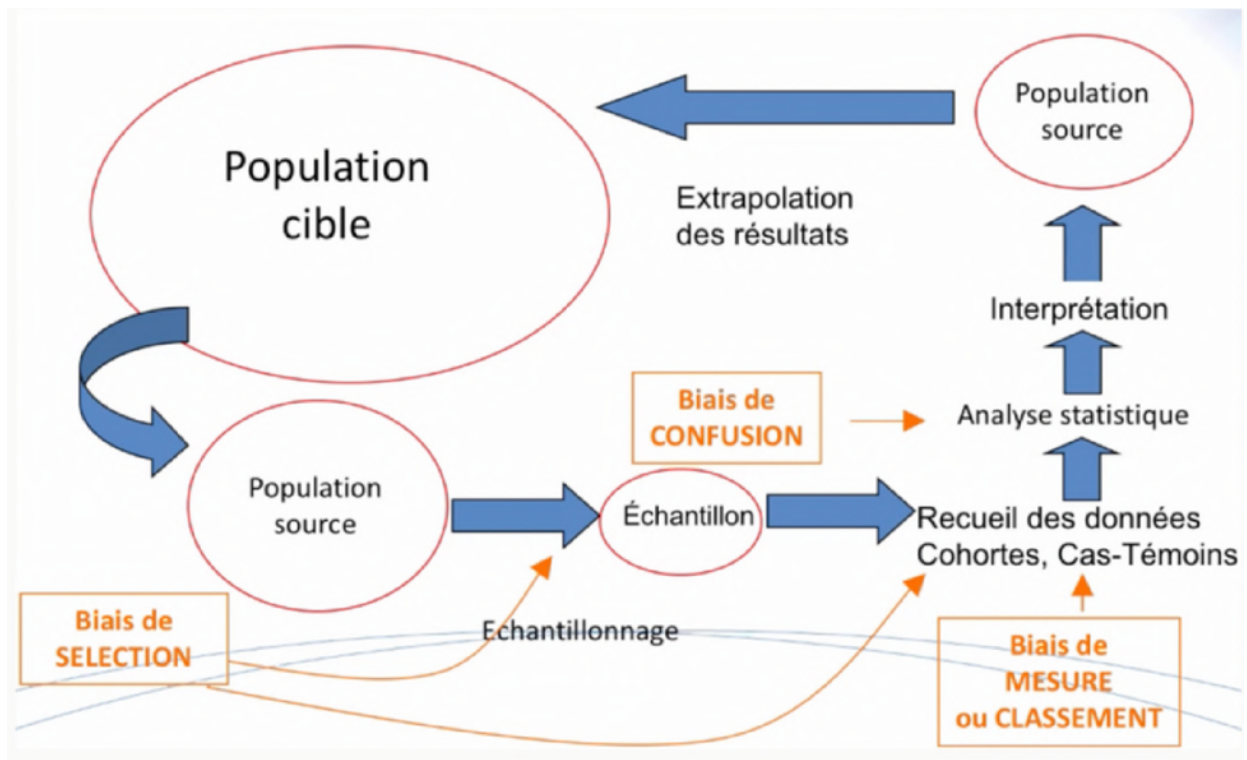
La probabilité de détecter une différence qui existe entre deux groupes est appelée la « **puissance du test** ». Il faut avoir une population suffisamment importante pour avoir la capacité de mettre en évidence une relation entre un facteur et une maladie si cette association existe réellement. En gros 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.

Normalement la notion de « puissance du test » vous allez la retrouver dans le cours sur les essais cliniques avec les hypothèses etc... elle sera plus détaillée, si ça ne reste pas très clair n'hésitez pas à me demander sur le forum !

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**.

VII. Synthèse



Dédis : *(C'est ma dernière fiche donc biggg dédi à tout le monde)*

Dédi à vous pour tout le chemin parcouru, ne lâchez rien vous verrez que ça en vaut la peine

Dédi à mes fillots : Iwan, Emma, Antoine, Ninon et Stella, vous allez réussir je crois en vous !!

Dédi à Aless et à nos délires que personne ne comprend (« *BBL gratuit svppp* »)

Dédi à Sarah qui trouve des sosies improbables (et très avantageux) à tout le monde

#tomcruise

Dédi à Margot qui trouve que je suis une super co-pilote *(c'est faux je nous ai perdu dans valrose)*

Dédi *(ou pas)* à Antoine qui me fait des doigts d'honneur h24 *(meme pendant les partiels ???)*

Dédi à Fefe et son expression « de malade » que je réutilise tout le temps

Dédi à Anouck et nos petites aventures en forêt quand on skie *(rip on a trauma aless...)*

Dédi à Osman la petite chipie qui connaît toujours tous les dramas *(et qui fonce dans des poteaux en ski wtf)*

Dédi à Bastien et nos pauses de 2h pendant le S2 *(très productifs oui)* à la fontaine de valrose *(elle est sacrée mtn, allez-y et vous réussirez votre p1)*

Dédi à Alexandra et notre passion commune pour le shatta

Dédi à Camilia *(qui ose pas monter dans ma voiture bref)* et nos tea sur les gens de la fac

Dédi à Fayrouz pcq on est tout le temps matchy *(très girl to girlll)*

Dédi à Anna qui fait comme partie de la famille pour moi. Bientôt nos 10 ans d'amitié, ça en fait beaucoup des souvenirs ensemble *(et des big plans boubiers...)*

Et surtout dédi à mon meilleur ami qui m'a toujours soutenu, qui me fait mourir de rire et qui m'a emmené en Espagne après la p1, je t'aime fort t'es une personne incroyable <3