



Epidémiologie analytique



Tryptophane



Et coucou tout le monde ! Voici ma dernière fiche, plutôt courte, avec quelques petits calculs mignons à savoir faire ! Vous verrez ce cours c'est surtout des informations à savoir, des définitions mais pas tant de maths que ça vous inquiète pas. Bon courage et toujours pareil, pour les questions, ça se passe sur le forum ! (mes remarques seront de cette couleur)

Objectifs du cours

- Savoir définir les termes suivants : épidémiologie analytique, risque relatif, facteur 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/inconvénients des enquêtes de cohortes/cas-témoins
 - Expliquer le principe de puissance épidémiologique

I- Définition de l'épidémiologie analytique

Aussi appelée 'épidémiologie **étiologique**' ou 'épidémiologie **explicative**', elle a pour but de : rechercher les **causes** des maladies, identifier, quantifier et interpréter un lien de cause à effet entre une **exposition** (facteur de risque) et un **état de santé** (la présence ou l'absence d'une maladie). Elle réalise donc des **mesures d'association** entre ces 2 notions. L'épidémiologie analytique répond en fait à la question '**Comment/pourquoi** une personne est atteinte d'une maladie donnée ?'

	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 ? (<i>Hommes ? Femmes ?</i>) Quand ? (<i>Depuis longtemps ? Récemment ?</i>) Où ? (<i>Uniformément réparti ? Concentré ?</i>)	Comment ? Pourquoi ?
But	GENERER des hypothèses	TESTER des hypothèses

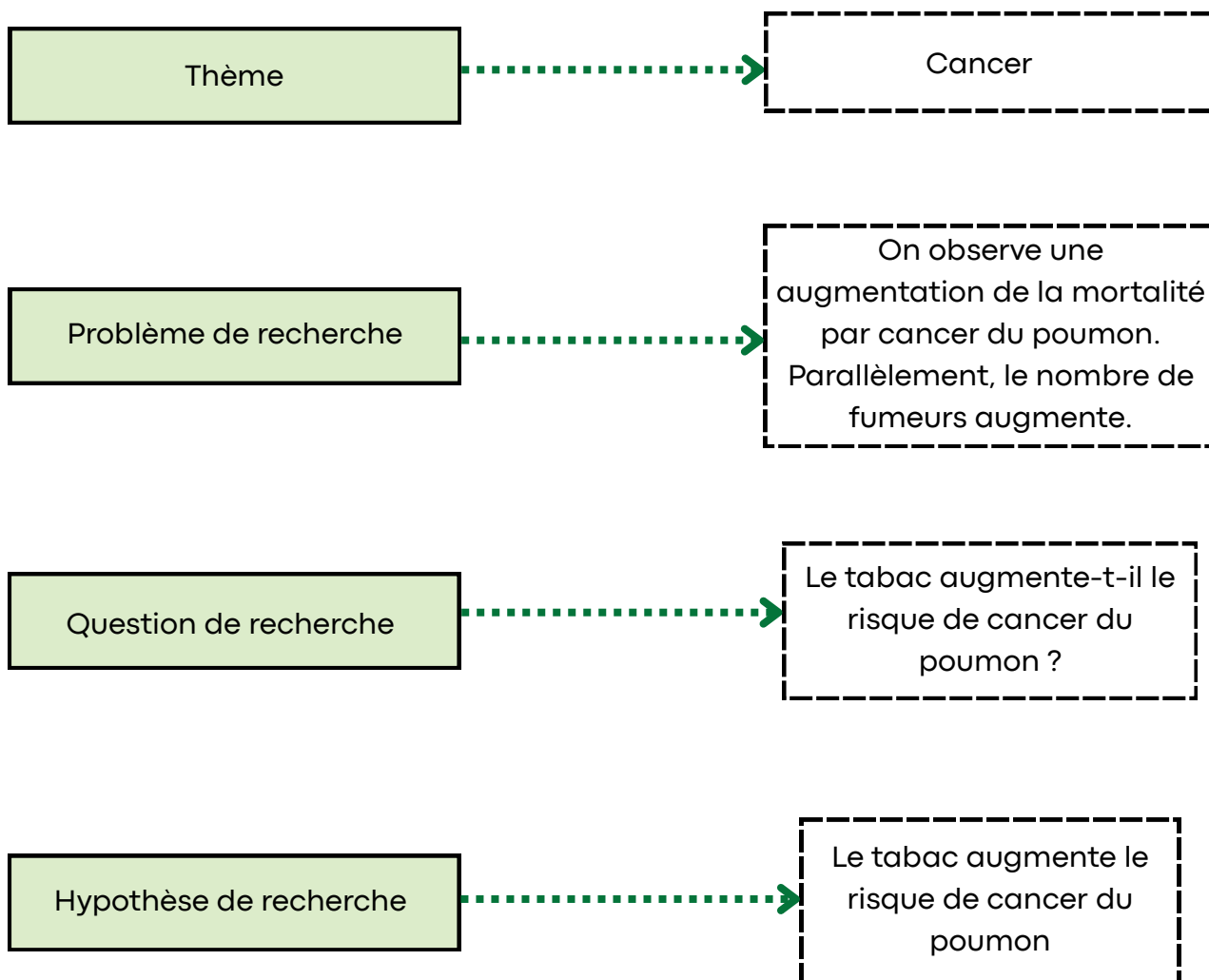
Petit tableau de ma vieille Lou pour bien comprendre :)

II- Hypothèse

Elle est le **point de départ** de toutes les enquêtes épidémiologiques :

- elle prédit une **relation** entre deux **variables** → l'exposition à un facteur et la présence ou l'absence de la maladie (ex : si je fume, ai-je une maladie ou pas par rapport à une personne qui ne fume pas ?)
- elle prolonge une **problématique** de recherche
- elle est l'élément de **base** de toute recherche : il est donc très important qu'elle soit claire, précise et courte
- elle est toujours rédigée **avant** même d'exécuter l'expérience puisque cette dernière est faite dans le but de la **vérifier**

Petit exemple :



C'est à partir de cette **hypothèse** que va être construit le **protocole** de recherche qui va viser à **valider ou invalider** cette hypothèse.

III- Définitions et concepts importants

- **Risque** : c'est la probabilité d'être atteint d'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'est pas exposé à ce facteur.
- **Facteur de risque** : ce sont tous les facteurs influant de façon favorable ou défavorable sur la survenue ou sur l'évolution d'un problème de santé.

1) Risque Relatif (RR)

Il correspond au **rapport** entre l'**incidence** de la maladie chez les sujets **exposés** (c'est-à-dire le nombre de nouveaux cas qui apparaissent chez les personnes exposées au facteur étudié) et l'**incidence** de la maladie chez les **non-exposés** (c'est-à-dire le nombre de nouveaux cas de la maladie chez les sujets qui ne sont pas exposés aux facteurs de risque). *Bon ça c'est du blabla on retient la formule :*

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

La valeur du **RR** permet d'évaluer la **force d'association**, c'est-à-dire la force du lien qui existe 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, 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.
- Si le **RR < 1**, cela signifie que le **facteur** entraîne une **diminution** de la probabilité d'apparition de la maladie.

Exemple : Un RR = 3 signifie que les sujets exposés au facteur de risque ont 3 fois plus de chances d'avoir la maladie que les non-exposés.

2) Intervalle de confiance du RR

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. (En gros, vu qu'on travaille à petite échelle, ce risque relatif sera probablement différent bien que proche entre plusieurs échantillons issues d'une même population ++).

Ainsi, cela nous amène à utiliser des **intervalles de confiance** (voir mon cours sur les stats descriptives) pour essayer de préciser dans quel intervalle se trouve la vraie valeur du risque relatif (donc celui de la population qui est inaccessible dans son entièreté on se souvient ++).

L'intervalle de confiance à 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 un IC-95% [2.5-8.31].

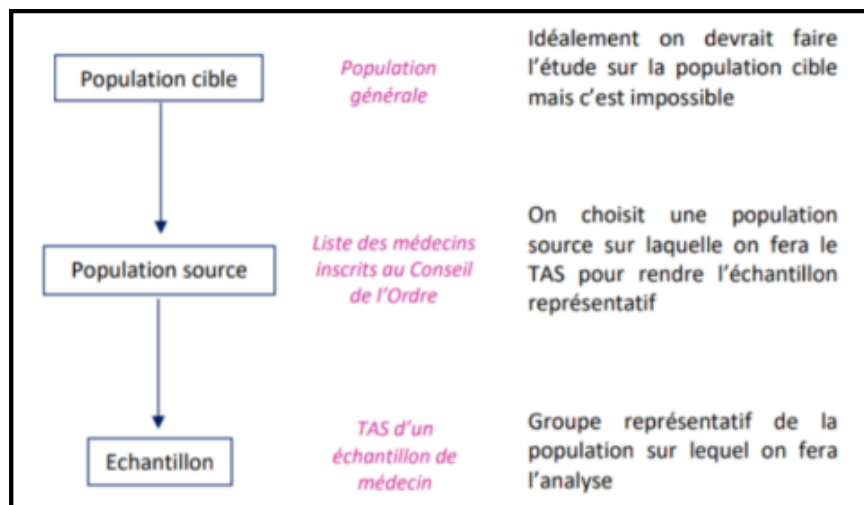
Cela signifie que le RR a 95% de chances de se trouver entre 2,5 et 8,31.

Or, dans tous les cas, ce risque est supérieur à 1. 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. Ainsi, 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 ++**

3) Inférence statistique (petite aide pour comprendre la suite)



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.

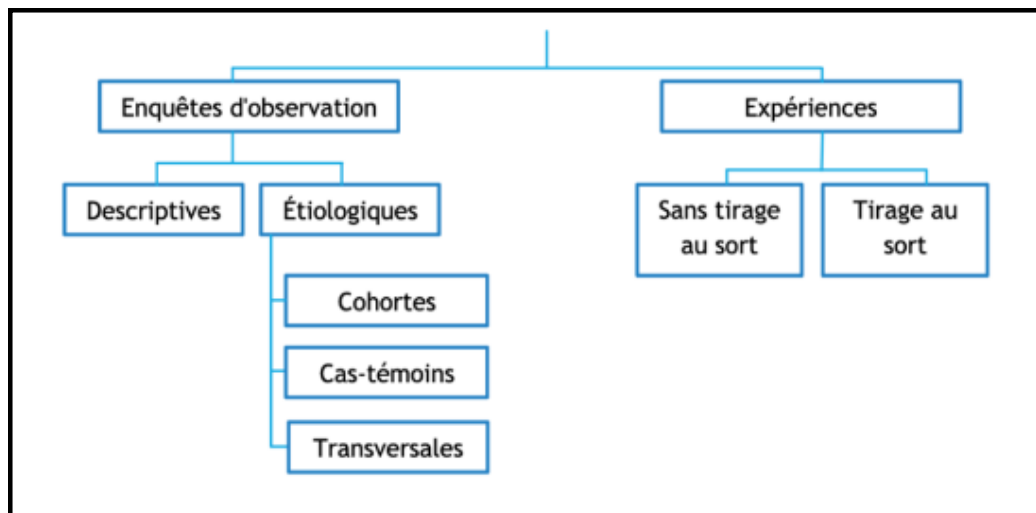
IV- Types d'études (partie importante ++)

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 observe 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.



Normalement, en recherche, les études sont contrôlées par l'**investigateur**. Tous les paramètres, les **conditions** dans lesquelles l'étude est réalisée et 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**'.

Dans les enquêtes **épidémiologiques**, on n'utilise **pas de TAS** 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 fumeurs et aussi demander 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.

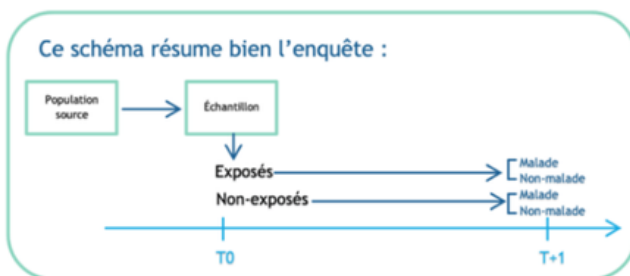
Dans le cours d'aujourd'hui on s'intéresse aux **enquêtes d'observation étiologiques** et dans ce cadre-là nous avons 3 grands types d'enquêtes : **les enquêtes de cohortes, les enquêtes cas-témoins et les enquêtes transversales.**

1) 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**. Donc on a au départ une population qui n'est **pas malade** et on va suivre cette population dans le **temps** pour voir si elle va **développer ou non** la maladie donnée.

Il s'agit d'une enquête **prospective**, donc on regarde ce qu'il se passe dans le temps, 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.



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

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

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

Ici, on réalise donc le calcul du RR avec les données du tableau en utilisant la formule : $RR = \frac{0,47}{0,07} = 6,7$

Les personnes fumeuses ont donc 6,7 fois plus de chances de développer un cancer du poumon que les personnes non fumeuses.

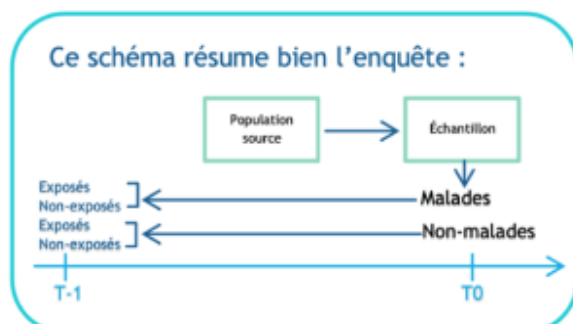
Avantages	<ul style="list-style-type: none"> • Estimation directe du risque relatif • Meilleure adaptation de la planification de l'étude aux objectifs recherchés • Bon contrôle des biais • Évaluation de l'influence du facteur étudié sur l'apparition de plusieurs pathologies
Inconvénients	<ul style="list-style-type: none"> • Etude longue et coûteuse donc difficilement reproductible • Possibilité de perdus de vue (= les patients qui quittent l'étude) • Effectif important au départ (pour limiter les pertes de gens) • Nécessite une population stable et une maladie à incubation courte (il ne faut pas attendre 30 à 40 ans avant que la maladie survienne)

2) 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.

Il s'agit d'une enquête **rétrospective** (pensez à un objet rétro = vieux = on regarde l'exposition passée).

Elle compare la fréquence de l'exposition **antérieure** chez des sujets atteints d'une pathologie (les cas) et chez les sujets non atteints pris comme témoins (c'est juste de la big répétition les bg alors c'est important ++).



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

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

⚠ **ALORS LA ATTENTION ARCHI IMPORTANT ⚠ : on ne peut pas calculer de risque relatif dans une enquête cas-témoins** car puisque l'on a choisi nous-mêmes un effectif de sujets 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 de ce risque relatif : c'est **l'Odd Ratio ++**

Si la maladie est rare, (<5%) alors on peut faire une **approximation** à partir du tableau ci dessus. Elle nous permet d'avoir une méthode de calcul de **l'Odd ratio (OR)** assez simple pour avoir une **approximation du risque relatif ++** :

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

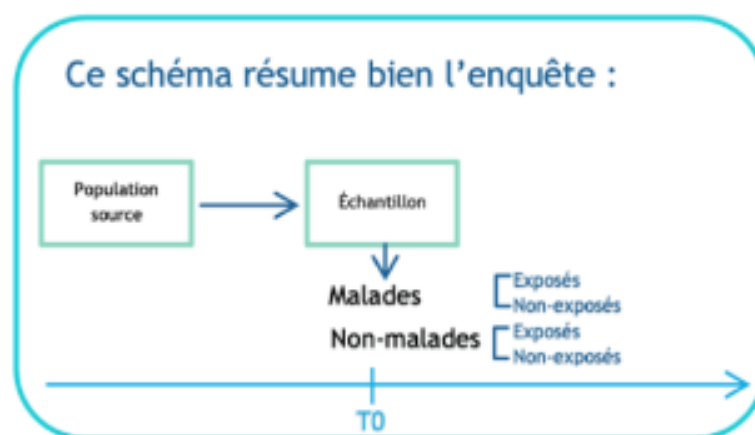
Avantages	<ul style="list-style-type: none"> • Enquête de courte durée donc coût modéré et possibilité d'itération (= la refaire) • Indiquée pour l'étude maladies rares • Possibilité d'évaluer plusieurs facteurs de risque pour une seule maladie
Inconvénients	<ul style="list-style-type: none"> • Évaluation indirecte du risque relatif (on utilise l'OR++) • Biais plus difficile à contrôler (ex : biais de mémorisation des patients -> parfois les souvenirs des expositions sont floues...) • 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)

3) 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

Les **biais** sont des **problèmes** posés par les enquêtes d'observation. Ces derniers peuvent être rencontrés à **trois moments** :

- Au moment du choix de la **population** → **BIAIS DE SELECTION**
- Au moment de la **mesure** de l'exposition et de la maladie → **BIAIS DE MESURE**
- 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**

On va voir en détail ces 3 biais : quand peut-on les rencontrer, quelles sont les conséquences sur l'enquête et comment peut-on les éviter. On s'accroche les gars c'est bientôt fini !

BIAIS DE SELECTION	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	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 dans la mesure de l'exposition pour en avoir une de qualité

BIAIS DE CONFUSION	Quand ?	<p>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)</p>
	Quelles conséquences ?	<p>On risque de produire une association statistique erronée même si le facteur étudié initialement et la maladie sont indépendants : il n'y a <u>pas de réel lien entre la maladie et le facteur</u> 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</p>
	Comment l'éviter ?	<p>°<u>Avant l'enquête</u> : appariement (exemple : dans une enquête cas témoins on prend à chaque fois un cas et un témoin du même âge)</p> <p>°<u>Au moment de l'analyse statistique</u> : 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.</p>

Alors alors on ne s'affole pas, je vous ai mis les infos les plus importantes en gras, faites attention ces notions tombent en qru !

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

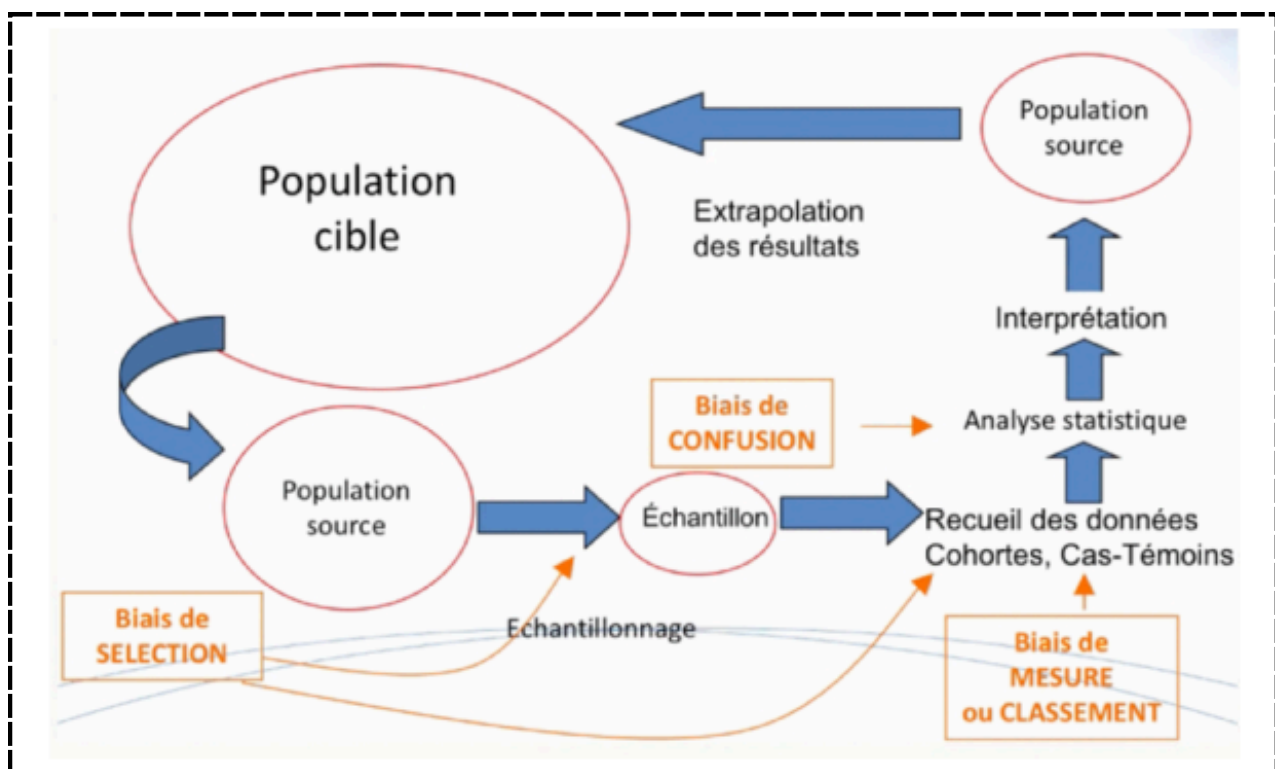
Ce n'est pas parce qu'on ne met pas en évidence un effet dans un échantillon qu'aucun effet n'existe en réalité ++

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 sujets, l'étude n'est pas assez puissante pour mettre en évidence cette relation et on va conclure à tort qu'il n'y a pas de relation entre les 2 variables.*

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



Fin du cours les loulous ! Alors beaucoup de texte, peu de maths mais du coup surtout de la compréhension ! Le plus important ici c'est les différents types d'enquêtes, le RR/OR et les différents biais. Vous inquiétez pas, on va vous concocter plein de qru pour vous entraîner, c'est important 😊 Bon courage pour la suite !

Grosse dédicace à mes fillots d'amour : Irsa, Morgane, Hélène, Valentine, Alexandre et Sarra, vous êtes vraiment les meilleurs !

Dédi à Mathys et Léa du bad je vous l'avez promis ;)

Dédi à TOI qui lis cette fiche, accroche toi peu important tes résultats du S1 tu as laaaagement le temps de te rattraper pendant ce magnifique ☆ S2 ☆

Si t'as des questions bg surtout n'hésite pas !!