

Statistiques inférentielles et épidémiologie : Mesure des risques et puissance en épidémiologie

I. Notion d'épidémiologie

EPIDEMIOLOGIE (= science de l'observation)	
Epidémiologie <u>Descriptive</u>	Epidémiologie <u>Analytique</u>
<p>- Etude de la distribution des pathologies et de certains états de santé dans les populations humaines ⇒ Mesure de la <u>prévalence*</u> et de <u>l'incidence*</u>.</p> <p>- Répond aux questions : « Qui ? » (Homme ? Femmes ? etc) « Quand ? » (depuis longtemps ?, récemment ?, etc) « Où ? » zone urbaine ?, campagne ? etc)</p> <p>⇒ Permet de générer des hypothèses ♥</p> <p><u>Exemple :</u> On étudie en parallèle : ⇒ La variation de la mortalité masculine (Qui ?) par cancer de l'estomac, du poumon et des voies aérodigestives supérieures (VADS) en France (où ?) de 1950 à 1990 (Quand?). ⇒ La variation de la consommation de tabac depuis plus d'un siècle.</p> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Mortalité masculine par cancer</p> </div> <div style="text-align: center;"> <p>Consommation de tabac</p> </div> </div> <p>Pourquoi l'incidence des cancers de l'appareil respiratoire augmente-t-elle ? Hypothèse générée : Il existe un lien entre l'exposition au tabac et la survenue d'un cancer de ce type.</p>	<p>- Epidémiologie analytique, appelée également épidémiologie étiologique ou épidémiologie explicative.</p> <p>- A pour but de rechercher les causes.</p> <p>- Il s'agit de l'identification, la quantification, et l'interprétation du lien de cause à effet entre une exposition à un facteur de risque (<i>ex</i> : tabac, médicament) et un état de santé (maladie, <i>ex</i> : cancer des poumons) ⇒ mesure d'association ♥</p> <p>⇒ Répond à la question « Pourquoi ? »</p> <p>⇒ Permet de tester des hypothèses ♥: On étudie si la survenue du cancer de l'appareil respiratoire est plus fréquente chez les personnes exposées au tabac que chez les personnes non exposées.</p> <p>Lien entre étude descriptive et étude analytique :</p> <div style="text-align: center;"> <pre> graph TD A[Etudes descriptive] --> B[Formulation d'hypothèses] B --> C[Etudes analytiques pour tester les hypothèses] C --> D[Analyse des résultats suggèrent de nouvelles études descriptives et des nouvelles hypothèses] D --> A </pre> </div>

* Prévalence : Nombre de malades présents à un moment donné dans la population

* Incidence : Nombre de nouveaux cas (malades) dans un intervalle de temps

II. Notion d'Hypothèse

Une hypothèse :

⇒ Prédit une relation entre deux variables

⇒ Est la base de toute recherche. L'hypothèse doit être **claire**, **précise** et **courte**.

⇒ Doit être **rédigée avant d'exécuter l'expérience**. En effet, l'expérience est menée dans le but de vérifier l'hypothèse.

Pr Pradier

Exemple :

Thème	Problème de recherche	Question de recherche	Hypothèse de recherche
⇒ Cancer	⇒ On observe une augmentation de la mortalité par cancer du poumon. ⇒ Parallèlement le nombre de fumeurs augmente (voir tableau ci-avant)	⇒ Le tabac augmente-t-il le risque de cancer du poumon ?	⇒ Le tabac augmente le risque de cancer du poumon

III. Notions d'inférence en statistique et d'échantillon

Inférence (statistique) :

Sous certaines conditions, un résultat obtenu à partir d'un **échantillon représentatif** d'une **population source**, peut être généralisé à la **population cible** intéressée par l'étude → **inférence statistique**. Une inférence statistique permet d'estimer la vraie valeur (inconnue) d'un paramètre à partir d'un échantillon.

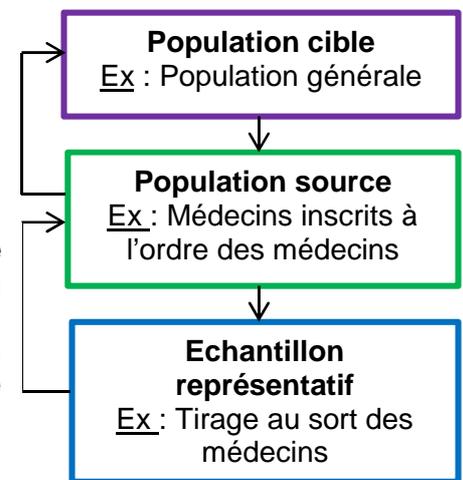
Echantillon :

⇒ Il est indispensable de toujours pouvoir décrire la façon dont est constitué un échantillon pour pouvoir justifier de **la représentativité de l'échantillon** → On doit connaître le nombre d'individus et leurs caractéristiques.

⇒ Les tests d'hypothèses sont réalisés sur l'échantillon (ex : test du X^2 , test de comparaison de moyenne, etc...). A partir de ces tests, on tirera des conclusions sur la population → **Inférence statistique**.

On **Infère** (= on fait l'hypothèse) que les résultats obtenus sur la **population source** sont « applicables » à la **population cible**

On établit une **inférence statistique** entre l'**échantillon représentatif** de la population source et la **population source**



IV. Notion de Risque, Risque Relatif, Facteur de risque

A. Définitions

Risque : 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.

Facteurs de risque : Tous facteurs influant de façon péjorative ou favorable sur la survenue ou sur l'évolution d'un problème de santé. **Exemple** : le tabac ou l'amiante peuvent agir de façon défavorable sur la santé, un traitement chimique ou chirurgical aura plutôt une action favorable sur la santé.

B. Interprétation du Risque relatif (RR) ♥

<p>R1 = Risque d'être malade (M) si on est exposé au facteur de risque (E) = Incidence de la maladie chez les personnes <u>exposées</u>.</p> $P(M / E) = P(M \cap E) / P(E)$	<p>R0 = Risque d'être malade (M) si on n'est pas exposé au facteur de risque (NE) = Incidence de la maladie chez les personnes <u>non exposées</u>.</p> $P(M / NE) = P(M \cap NE) / P(NE)$
---	---

$$RR = \frac{R1}{R0} = \frac{\text{Incidence de la maladie chez les personnes exposées au facteur de risque}}{\text{Incidence de la maladie chez les personnes non-exposées au facteur de risque}}$$

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 risque relatif permet d'évaluer la force de l'association !**

RR	Interprétation du Risque Relatif
RR = 1	⇒ Le facteur de risque étudié n'a pas d'influence sur l'incidence de la maladie. Il n'y a pas de différence entre l'incidence de la maladie chez les sujets exposés et chez les sujets non exposés. Dans l'exemple ci-avant, un RR = 1 signifierait que l'apparition d'un cancer du poumon n'est pas associée à la consommation de tabac.
RR < 1	⇒ Le facteur de risque étudié a une influence favorable sur l'incidence de la maladie, donc sur la santé des personnes en général. C'est souvent le cas lorsque le facteur de risque est un traitement médical, par exemple.
RR > 1	⇒ Le facteur de risque étudié a une influence défavorable sur l'incidence de la maladie, donc sur la santé des personnes en général. Dans l'exemple ci-avant, un Risque Relatif de 3 (RR=3) signifierait que les fumeurs ont une probabilité 3 fois plus élevée d'avoir un cancer des poumons que les non-fumeurs.

C. Intervalle de confiance du Risque Relatif

Le Risque Relatif étant calculé à partir des incidences de la maladie dans des échantillons représentatifs, et non d'après l'incidence véritable de la maladie dans la population concernée, l'analyse du Risque Relatif pourra être complétée par l'estimation de limites de confiances → **Estimation de l'intervalle de confiance du Risque Relatif.**

On étudie le plus couramment **l'intervalle de confiance à 95%** : Intervalle dans lequel la véritable valeur du Risque Relatif a 95% de chance de se trouver (et donc un risque de 5% de ne pas se trouver).

Exemple : Imaginons un RR calculé d'après les incidences de la maladie dans les groupes Exposés et non exposés, égal à 5,2. Au risque de 5%, cette valeur serait comprise dans un intervalle, exemple : [2,5 - 8,3].

Nota: La largeur de l'intervalle de confiance (IC) représente la **précision de l'estimation** du Risque Relatif.

Interprétation de l'intervalle de confiance du Risque Relatif :

⇒ Si la valeur « 1 » appartient à l'intervalle de confiance IC-95% (ex : RR= 1,5 ; IC-95% [0,5 - 5,3]) → l'**association** entre l'exposition et la maladie est dite « **non-significative** » puisque le véritable RR peut être égal à 1 (si RR = 1 → Le facteur de risque étudié n'a pas d'influence sur l'incidence de la maladie).

⇒ Si la valeur « 1 » n'appartient pas à l'intervalle de confiance IC-95% et si les limites de l'intervalle de confiance sont éloignées de la valeur « 1 » (ex : RR = 5,2 ; IC-95% [2,5 - 8,3] → l'**association** entre l'exposition et la maladie est dite « **significative** »

D. Exemple de Risques Relatifs

Facteur de Risque	Pathologie	Risque Relatif
Café	Cancer de la vessie	1,5
Contraceptifs oraux	Mortalité cardio-vasculaire	4
Amiante	Cancer du Poumon	5
Tabac	Cancer du Poumon	10

E. Principe de l'étude Analytique ... en Bref

⇒ Des hypothèses impliquant un facteur de risque et l'incidence d'une maladie, sont générées grâce aux études descriptives.

⇒ L'étude analytique cherche à mettre en évidence une **relation entre l'exposition à un facteur de risque et la survenue d'une maladie**, en général entre une cause et un effet, en **testant ces hypothèses.**

⇒ Pour cela, on compare :

- Un **groupe exposé** au facteur étudié
- Un **groupe non-exposé** au facteur étudié

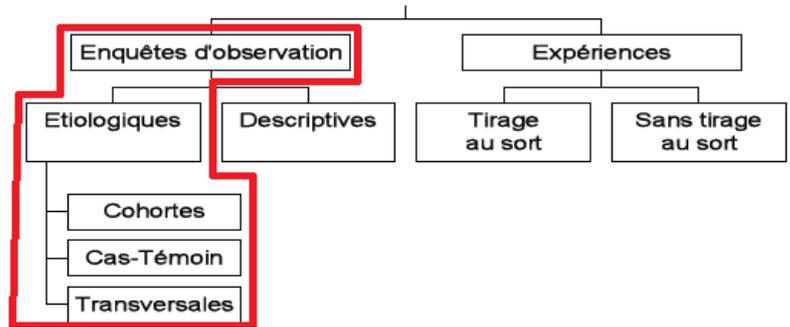
⇒ On **estime le Risque Relatif**

V. Enquêtes d'observations

A. Différents types d'étude

Les enquêtes d'observation se distinguent des expériences (essais cliniques par exemple).

Comme vu précédemment, les enquêtes d'observation (épidémiologiques) peuvent être étiologiques (= analytiques) ou descriptives. Ces deux types d'études étant liées (voir chapitre I)



B. Problèmes posés par les enquêtes d'observations : Les biais

1. Biais : Définition

Il s'agit d'une erreur systématique. C'est une distorsion de l'estimation de la mesure d'une association entre l'exposition à un facteur de risque et la survenue d'une maladie. Autrement dit, il s'agit d'un écart entre le résultat obtenu dans le cadre de l'étude et la réalité.

2. Conséquences d'un biais :

- ⇒ **Surestimation de l'effet**, exemple : dans le cadre d'une étude sur la relation entre la consommation de café et l'incidence de cancer de la vessie, on détermine un Risque Relatif de 50 (nettement supérieur à la réalité où le RR = 1,5)
- ⇒ **Sous-estimation de l'effet**, exemple : dans le cadre d'une étude sur la relation entre la consommation de tabac et l'incidence du cancer des poumons, on détermine un Risque Relatif de 1,2 (nettement inférieur à la réalité où le RR = 10)
- ⇒ **Effet inverse** : ce qui était un facteur protecteur, apparaît comme un facteur de risque. Exemple : dans le cadre d'une étude sur la relation entre la pratique sportive et l'incidence des maladies cardio-vasculaires, on détermine un RR de 2 (alors qu'en théorie, la pratique sportive préservant de ces maladie, le RR devrait être inférieur à 1)

3. Les différents types de biais ♥

Dans le cadre des études d'observation, on distingue 3 principaux types de biais. Ceux-ci interviennent à différents moments de l'enquête d'observation.

Type de biais	Quand ?	Conséquences ?	Comment l'éviter ?
Biais de Sélection	⇒ Lors de la constitution de l'échantillon ou des groupes. ⇒ Lors de la réalisation de l'enquête (non répondants)	⇒ Echantillon non représentatif de la population source ⇒ Impossibilité d' « extrapoler » les résultats → Inférence	⇒ Tirage au sort de l'échantillon ⇒ Bien choisir la population source
Biais de Mesure (ou d'information) (ou classement)	⇒ Lors de la mesure du degré de l'exposition des sujets et/ou de la de la maladie (présence ou non de la maladie)	⇒ Sur ou sous-estimation du risque relatif	⇒ Bien distinguer les malades des non-malades ⇒ Qualité de la mesure de l'exposition
Biais de confusion	⇒ Au moment de l'analyse statistique des données ⇒ Quand les groupes ne sont pas comparables (ex : différence d'âge) ⇒ Présence d'un facteur de confusion (ex : lorsque l'âge influence les résultats de l'étude)	⇒ Peut aboutir à une association statistique entre un facteur de risque et une maladie alors qu'ils sont indépendants l'un de l'autre ⇒ Aboutit à une « fausse » association	⇒ <u>Avant l'enquête</u> : Appariement, ex : population du même âge. ⇒ <u>Au moment de l'analyse statistique</u> : Standardisation, Ajustement (analyse multivariées)

C. Enquêtes Analytiques : Cohorte, Cas-Témoin, Transversale ♥

Enquête de Cohorte	Enquête Cas-Témoin																
<p>Enquête Prospective (en général): Les sujets sont suivis dans le temps.</p> <p>Principe : Compare l'incidence d'une maladie M chez des <u>sujets exposés</u>, à l'incidence de la même maladie M chez des <u>sujets non exposés</u> (=Témoins)</p> <p>1er temps ⇒ On constitue une cohorte (échantillon) de <u>sujets sains exposés</u> à un facteur de risque (ex : fumeur) et une cohorte de <u>sujets sains non-exposés</u> à ce facteur de risque (ex : non-fumeur).</p> <p>2e temps ⇒ On relève dans le temps la survenue de la maladie M (ex : cancer du poumon) chez les sujets exposés et les sujets non-exposés.</p>	<p>Enquête Rétrospective : On étudie le passé des sujets</p> <p>Principe : Compare la fréquence de l'exposition antérieure chez des <u>sujets ayant la maladie M</u> (« les cas »), à la fréquence de l'exposition chez des <u>sujets sains</u> (« les Témoins »)</p> <p>1er temps ⇒ On recrute des <u>sujets ayant la maladie M</u> (« les cas », ex : cancer du poumon) et des <u>sujets n'ayant pas cette maladie M</u> (« les témoins », ex : pas de cancer du poumon)</p> <p>2e temps ⇒ On étudie l'exposition au facteur de risque (ex : Fumeur) au cours de la vie des sujets malades et des sujets non-malades.</p>																
<table border="1" style="width: 100%; text-align: center;"> <tr> <td></td> <td style="background-color: #cccccc;">Malade</td> <td style="background-color: #cccccc;">Non Malade</td> <td style="background-color: #cccccc;">Total</td> </tr> <tr> <td style="background-color: #cccccc;">Exposés</td> <td>a</td> <td>b</td> <td>a + b</td> </tr> <tr> <td style="background-color: #cccccc;">Non Exposés</td> <td>c</td> <td>d</td> <td>c + d</td> </tr> <tr> <td style="background-color: #cccccc;">Total</td> <td>a + c</td> <td>b + d</td> <td>N = a + b + c + d</td> </tr> </table>			Malade	Non Malade	Total	Exposés	a	b	a + b	Non Exposés	c	d	c + d	Total	a + c	b + d	N = a + b + c + d
	Malade	Non Malade	Total														
Exposés	a	b	a + b														
Non Exposés	c	d	c + d														
Total	a + c	b + d	N = a + b + c + d														
<p>Objectifs :</p> <p>⇒ Etiologique = Mesurer l'association entre un facteur de risque et une maladie : on étudie si la survenue de la maladie est plus fréquente (ou moins fréquente) chez des sujets exposés que chez des sujets non-exposés</p> <p>⇒ Descriptif : Déterminer la fréquence de la maladie dans la population (calcul de l'incidence)</p> <p>Indicateur : On utilise de préférence le Risque Relatif (RR)</p> $RR = \frac{\text{Incidence de la maladie chez les exposés}}{\text{Incidence de la maladie chez les non-exposés}}$ $RR = \frac{R1}{R0} = \frac{a/(a+b)}{c/(c+d)}$ <p>Dans les <u>enquêtes de cohorte</u>, on peut considérer que :</p> <ul style="list-style-type: none"> - L'incidence de la maladie dans le groupe « Exposés » est représentative de la population générale exposée. - L'incidence de la maladie dans le groupe « Non Exposés » est représentative de la population générale non exposée. <p>⇒ Une enquête de cohorte permet donc à priori d'étudier l'incidence (fréquence) de la maladie dans la population générale.</p>	<p>Objectif :</p> <p>⇒ Etiologique = Mesurer l'association entre un facteur de risque et une maladie : on étudie si la maladie est plus fréquente (ou moins fréquente) chez des sujets exposés que chez des sujets non-exposés</p> <p>Indicateur : On utilise uniquement l'Odds-Ratio (OR)</p> $OR = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}$ <p>L'Odds-Ratio est une bonne approximation du Risque Relatif si la maladie est rare, en effet « a » et « c » → 0.</p> <p>D'où : $a / (a+b) \approx a / b$ et $c / (c+d) \approx c / d$ →</p> $RR = \frac{a/(a+b)}{c/(c+d)} \approx \frac{a/b}{c/d} \approx \frac{a \times d}{c \times b} \approx OR$ <p>Les enquêtes Cas-Témoins sont indiquées lorsque la maladie est rare dans la population générale, car elles permettent de surreprésenter dans l'étude les personnes malades → cela a un inconvénient ⇒ on ne peut pas déterminer l'incidence (fréquence) de la maladie ni le Risque Relatif dans une étude Cas-Témoins !</p>																

Enquête de Cohorte	Enquête Cas-Témoin																																
<p>Avantages : ♥</p> <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 ⇒ Evaluation de l'influence du facteur étudié sur l'apparition de plusieurs pathologies 	<p>Avantages : ♥</p> <ul style="list-style-type: none"> ⇒ Enquête de courte durée ⇒ Coût modéré ⇒ Possibilité d'itération ⇒ Indiquée pour l'étude de maladies rares ⇒ Possibilité d'évaluer plusieurs facteurs de risque pour une seule maladie 																																
<p>Inconvénients : ♥</p> <ul style="list-style-type: none"> ⇒ Etude longue ⇒ Etude coûteuse ⇒ Etude difficilement reproductible ⇒ Possibilité de perdus de vue ⇒ Effectif important au départ ⇒ Nécessité population stable ⇒ Nécessité maladie incubation courte 	<p>Inconvénients : ♥</p> <ul style="list-style-type: none"> ⇒ Evaluation indirecte du risque relatif → OR ⇒ Biais plus difficile à contrôler ⇒ Sélection délicate des témoins ⇒ Nécessité d'une bonne information diagnostique au départ 																																
<p>Exemple : Enquête de cohorte menée sur 1 an</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>Cancer des poumons</th> <th>Pas de cancer</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Fumeurs</th> <td>8</td> <td>292</td> <td>300</td> </tr> <tr> <th>Non Fumeurs</th> <td>2</td> <td>698</td> <td>700</td> </tr> <tr> <th>Total</th> <td>10</td> <td>990</td> <td>1000</td> </tr> </tbody> </table> <p> $R1 = a / (a + b) = 8 / (8 + 292) = 8 / 300 = 0,026$ $R0 = c / (c + d) = 2 / (2 + 698) = 2 / 700 = 0,0028$ $RR = \frac{R1}{R0} = \frac{a / (a+b)}{c / (c+d)} = \frac{0,026}{0,0028} = 9,28$ </p> <ul style="list-style-type: none"> ⇒ R1 = 2,6% : Incidence du cancer du poumon chez les fumeurs dans cet échantillon. ⇒ R0 = 0,28% : Incidence du cancer du poumon chez les non-fumeurs dans cet échantillon. <p>Nota : Si les échantillons de l'étude sont représentatifs, il est alors possible d'extrapoler ces résultats à la population générale → Inférence.</p> <ul style="list-style-type: none"> ⇒ RR = 9,28 : Il y a environ 9 (9,28) fois plus de risques d'avoir un cancer des poumons en étant fumeur qu'en ne fumant pas. <p>Nota : L'Odds-ratio peut être utilisé dans toutes les études (Cohorte et Cas-témoins). Dans le cas de l'étude de cohorte on utilise de préférence le RR.</p>		Cancer des poumons	Pas de cancer	Total	Fumeurs	8	292	300	Non Fumeurs	2	698	700	Total	10	990	1000	<p>Exemple : Enquête Cas-Témoins</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>Cancer des poumons</th> <th>Pas de cancer</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>Fumeurs</th> <td>80</td> <td>30</td> <td>110</td> </tr> <tr> <th>Non Fumeurs</th> <td>20</td> <td>70</td> <td>90</td> </tr> <tr> <th>Total</th> <td>100</td> <td>100</td> <td>200</td> </tr> </tbody> </table> <p> R1 et R0 : Impossibles à calculer puisque le nombre de sujets dans les groupes « Cancers des poumons » et « Pas de cancer » est déterminé de façon arbitraire. Ces groupes sont indépendants. </p> <p> $OR = \frac{a / b}{c / d} = \frac{a \times d}{c \times b} = \frac{80 \times 70}{20 \times 30} = 9,33$ </p> <ul style="list-style-type: none"> ⇒ Odds-Ratio = 9,33 : Il y a environ 9 (9,33) fois plus de risques d'avoir un cancer des poumons en étant fumeur qu'en ne fumant pas. <p>Nota : Si nous avions calculé le Risque Relatif d'après les données de cette enquête Cas-témoins, nous aurions eu le résultat aberrant suivant:</p> <p> $RR = \frac{80 / (80+30)}{20 / (20+70)} = \frac{0,72}{0,22} = 3,27$ </p> <p>→ On ne mesure pas l'association entre un facteur de risque et une maladie avec le Risque Relatif dans une étude Cas-témoin ! ♥</p>		Cancer des poumons	Pas de cancer	Total	Fumeurs	80	30	110	Non Fumeurs	20	70	90	Total	100	100	200
	Cancer des poumons	Pas de cancer	Total																														
Fumeurs	8	292	300																														
Non Fumeurs	2	698	700																														
Total	10	990	1000																														
	Cancer des poumons	Pas de cancer	Total																														
Fumeurs	80	30	110																														
Non Fumeurs	20	70	90																														
Total	100	100	200																														

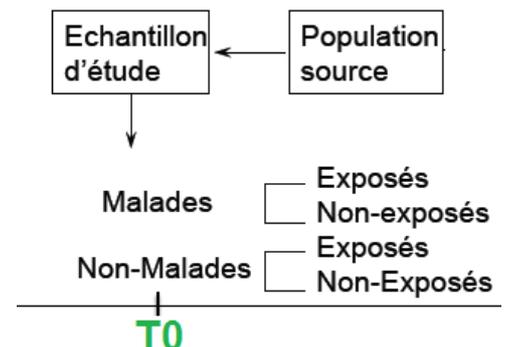
Enquêtes transversales

⇒ **Photographie d'un échantillon d'une population à un instant donné** : L'information recueillie concerne l'existence de l'exposition et de la maladie au moment de l'enquête.

Objectifs :

- ⇒ **Descriptif** (principalement) : Estimer la **prévalence** de la maladie ou celle des expositions.
- ⇒ **Etiologique** (rarement) : Ce type d'enquête est mal adapté à des études analytiques.

Nota : Ce type d'enquête ne mesure pas l'incidence d'une maladie dans la population et ne permet pas d'étude dynamique.



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

« **Puissance du test** » : Probabilité de détecter une différence qui existe entre deux groupes

Nota :

- ⇒ Ne pas mettre en évidence un effet dans un échantillon ne signifie pas qu'aucun effet n'existe en réalité.
- ⇒ Avant chaque étude, on doit calculer le nombre de patients nécessaires ainsi que la puissance de l'essai

A. Risques de première et de deuxième espèce et puissance d'un test

Risque α	<ul style="list-style-type: none"> ⇒ Risque de première espèce ⇒ Risque de rejeter l'hypothèse H0 alors qu'elle est vraie ⇒ Probabilité de conclure à une relation entre l'exposition à un facteur et la maladie, alors qu'il n'existe aucune relation. ⇒ En règle générale $\alpha = 5\%$
Risque β	<ul style="list-style-type: none"> ⇒ Risque de deuxième espèce ⇒ Risque de rejeter l'hypothèse H1 alors qu'elle est vraie ⇒ Probabilité de conclure à l'absence de relation entre l'exposition à un facteur et la maladie, alors qu'il existe bien une relation. ⇒ En règle générale $\beta = 20\%$ (ndlr : à retenir seulement dans le cadre de ce cours)
$1 - \beta$ Puissance du test	<ul style="list-style-type: none"> ⇒ Probabilité de trouver un RR significativement différent de 1, lorsque le RR réel est distinct de 1 = Probabilité de conclure à une relation entre l'exposition à un facteur et la maladie lorsqu'il existe bien une relation ⇒ En règle générale puissance = 80% (ndlr : à retenir seulement dans le cadre de ce cours)

B. Détermination de la taille de l'échantillon

La planification d'une étude épidémiologique (étude de cohorte ou cas-témoins) nécessite de déterminer l'effectif des échantillons. La spécification préalable des conditions de l'étude est nécessaire.

Les points à spécifier pour établir la taille des échantillons sont donc :

- ⇒ **α** : Niveau de signification (= 0,05 en général)
- ⇒ **Puissance du test $1 - \beta$** : La puissance se situe le plus souvent entre 0,8 et 0,9
- ⇒ **Fréquence relative d'exposition (P0)** parmi les non-exposés, dans la population cible → proportion de sujets exposés dans la population.
- ⇒ **Le risque relatif attendu** : exemple : RR (ou Odds-Ratio pour les études cas témoins) = 0,6 (on estime que le facteur à un effet protecteur)

VII. Critères de jugement d'une relation de cause à effet

Nous avons vu comment identifier une association statistique entre une exposition et une maladie (grâce aux calculs des tests d'hypothèse → accepter ou rejeter H0), comment quantifier la force de cette association (grâce au RR ou à l'Odds-ratio). **Il reste à établir la relation de cause à effet entre le facteur de risque et la maladie : « Le facteur de risque est-il bien la cause de la maladie ? »**

On distingue 6 critères de jugement : ♥♥

- ⇒ **1 - Evaluation de la séquence dans le temps** : L'exposition à un facteur de risque précède la maladie.
Exemple : La consommation de tabac doit être antérieure à l'apparition du cancer du poumon.
- ⇒ **2 - La force de l'association** : Précision avec laquelle une variable peut permettre de prédire l'autre.
Exemple : Tous les fumeurs (ou une partie significative) développent un cancer des poumons
- ⇒ **3 - La spécificité de la cause et de l'effet** : Présence de la cause étudiée chez tous les malades.
Exemple : Dans 9 cas sur 10, 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*: Le risque de cancer des poumons augmente avec 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. *Exemple* : plusieurs études différentes ont confirmé la relation entre le tabagisme et le cancer du poumon.
- ⇒ **6 - Plausibilité biologique** : Cohérences du mécanisme d'action du facteur par rapport aux connaissances. *Exemple* : La relation entre tabac et cancer du poumon est plausible, explicable.

VIII. Synthèse

