

EPIDEMIOLOGIE A VISEE ETIOLOGIQUE

Concepts, mesures et méthodes d'investigation*

M. GOLDBERG⁽¹⁾ et Arnette LECLERC⁽¹⁾

RESUME : Ce document porte sur les principaux concepts de l'épidémiologie à visée étiologique, et la mise en oeuvre de méthodes d'investigation destinées à analyser le rôle de facteurs pouvant influencer la survenue de problèmes de santé.

Dans la première partie, sont précisées des définitions : population, taux d'incidence, risque, cohorte, prévalence, facteur de risque. La question de la causalité en épidémiologie y est également développée.

La seconde partie commence par la présentation de la notion de risque relatif, mesure de comparaison entre groupes occupant un rôle central dans toute enquête étiologique. Les modes de calcul sont ensuite détaillés pour les deux grands types d'enquêtes que sont l'enquête de cohorte (ou "exposés/non-exposés") et l'enquête cas/témoins.

SUMMARY : This paper deals with the most important concepts of aetiological epidemiology and with the use of investigation methods animal at analyzing the importance of factors able to influence occurrence of health problems.

In the first part, definitions are given : population, incidence rate, risk, cohort, prevalence, risk-factor. The question of causality in epidemiology is also presented.

The second part starts with the presentation of the notion of relative risk, a measure of comparison between groups, which hold an important place in every aetiological survey. Then the calculus strategies are given for the two main types of studies : cohort study and case control study.

*
* *

INTRODUCTION

L'épidémiologie humaine se définit classiquement comme l'étude de la distribution et des déterminants des états de santé et des maladies dans les populations humaines. On distingue habituellement l'épidémiologie descriptive, l'épidémiologie évaluative, et l'épidémiologie analytique (ou à visée étiologique) qui est l'objet du présent document.

* Texte de la conférence présentée le 31 mai 1990.

(1) INSERM, Unité 88, GERS, 91 Boulevard de l'Hôpital, 75634 Paris cedex 13.

I - POPULATION, RISQUE ET MESURE DU RISQUE

A. LE CONTEXTE

L'épidémiologie étiologique (ou "explicative") a pour but de chercher les causes des problèmes de santé dans les populations. Plus précisément, il s'agit d'analyser le rôle de facteurs susceptibles d'influencer l'incidence des problèmes de santé, en l'augmentant (parfois en le diminuant). Bien des disciplines scientifiques biomédicales concourent à l'étude des causes des maladies.

Mais, aussi importants que soient leurs résultats, une confirmation épidémiologique, reposant sur l'observation chez l'Homme (ou l'animal quand il s'agit de médecine vétérinaire) avec une comparaison de sujets malades et non-malades, est toujours nécessaire (même si elle n'est pas toujours suffisante : voir plus loin la notion de causalité en épidémiologie).

D'une façon très générale, les méthodes à visée explicative procèdent en comparant des groupes de sujets diversement exposés au (x) facteur (s) dont on cherche à analyser le rôle, et parmi lesquels certains sont atteints de la maladie étudiée, et d'autres sont indemnes.

Cette comparaison repose sur l'analyse de certains indices épidémiologiques visant à mesurer l'association entre "facteurs de risque" et maladie qu'on peut établir à partir de données réunies au cours d'enquêtes réalisées auprès de groupes de sujets.

Avant de définir les principales mesures d'association utilisées en épidémiologie étiologique et les modèles d'enquêtes permettant de réunir les données permettant de calculer les indices correspondants, il faut définir certaines notions de base : population, risque (et principaux indices de mesure de risque), notion de causalité épidémiologique.

B. NOTIONS DE BASE

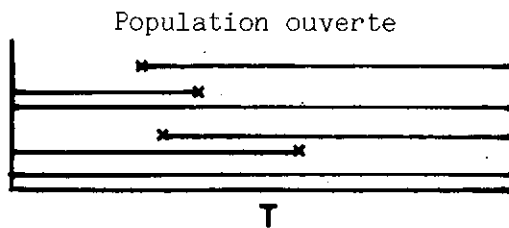
1. Population, indice, risque

On parlera ici des mesures **d'incidence** (taux d'incidence, et risque ou incidence cumulée) donc des **mesures de survenue de nouveaux cas** dans une population. La mesure de la **fréquence** des cas (nouveaux ou non) est la **prévalence**.

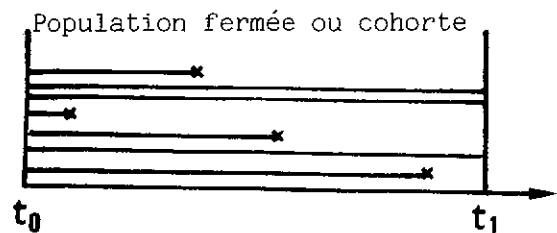
Avant de parler d'incidence, il faut définir ce que l'on appelle une population.

- Population

Il existe deux types principaux de population qui ne permettront pas d'effectuer les mêmes mesures :



T est une période d'observation dans laquelle il y a des sorties et des entrées. Il peut s'agir d'une population qui ne "vieillit pas", par exemple celle correspondant à une tranche d'âge.



Le temps t fait "vieillir" la population entre t_0 et t_1 (âges ou temps écoulé depuis un événement à t_0).
Il y a des sorties : par décès ou survenue de la maladie étudiée, mais pas d'entrée.

Exemple de population ouverte : les français de 20 ans à moins de 30 ans, période d'observation de 1 an (1/1/88 au 31/12/88).

On y entre à son anniversaire de 20 ans ; on en sort si l'on fête ses 30 ans, si l'on décède ou que l'on est atteint par la maladie étudiée (à partir de ce moment on n'est plus "soumis au risque", on ne risque plus d'être malade).

De plus, il y a des entrées et sorties par immigration et émigration.

Exemple de cohorte : un groupe de sujets à risque élevé de SIDA, volontaires pour être suivis dans le temps. Une fois défini le groupe, il n'y aura plus d'entrées, et les "sorties" sont connues (pas de perdus de vue).

Sur une population de type cohorte, il y a deux mesures possibles de survenue de nouveaux cas :

- Risque ou incidence cumulée (cependant le calcul n'est pas toujours évident)
- Taux d'incidence.

Sur une population ouverte, on calculera des taux d'incidence.

A partir de ces taux, on pourra estimer des risques mais il s'agit d'un calcul indirect.

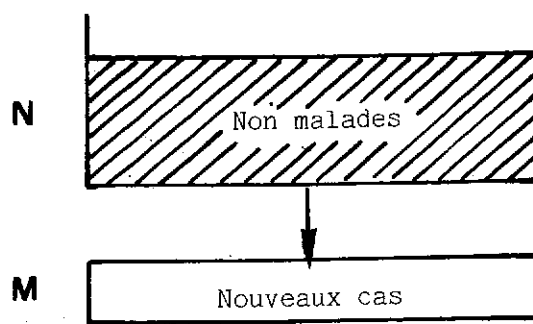
Par la suite, on supposera que la maladie étudiée ne survient qu'une fois dans la vie.

Qu'il s'agisse de survenue de maladie ou de décès, les calculs sont les mêmes, cependant on parle d'incidence dans un cas, de mortalité dans l'autre cas.

PREMIERE MESURE D'INCIDENCE : TAUX D'INCIDENCE

Soit une population non malade de taille N, observée pendant un temps T. Pendant cette période, un certain nombre de non malades deviennent malades, soit M nouveaux malades. On a :

$$M = N \times T \times \hat{\lambda}$$



Le nombre de nouveaux malades est proportionnel à la taille de la population N , à la durée de la période d'observation T , et à λ qui est une vitesse de passage d'un état dans un autre, nombre de nouveaux malades par unité de temps, rapporté à la taille de la population (analogie = débit d'un fleuve).

Le taux d'incidence est $\hat{\lambda}$:

$$\hat{\lambda} = \frac{M}{N \times T}$$

<----- nombre de nouveaux cas
<----- surface de l'aire hachurée

La présentation ci-dessus est simplifiée car on suppose que l'aire hachurée est un rectangle : même effectif N à risque (= susceptible de devenir malade) pendant toute la période d'observation. La définition générale du taux d'incidence est :

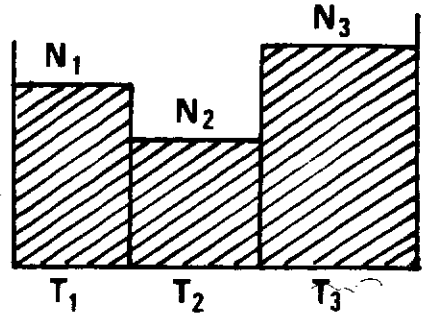
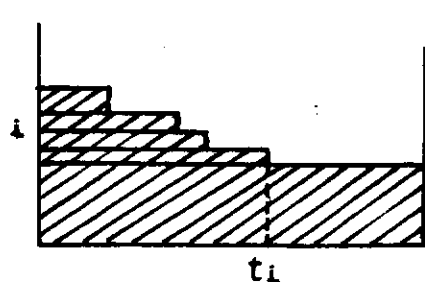
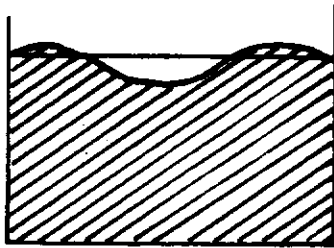
$$\hat{\lambda} = \frac{M}{PA}$$

<----- nombre de nouveaux cas
<----- nombre de personnes-années

Le dénominateur est encore "la surface de l'aire hachurée" mais dans des conditions plus complexes correspondant à une population "ouverte" ou à une cohorte, ou au sous-ensemble d'une cohorte d'un âge donné.

Le mode de calcul est plus ou moins précis selon le type d'étude et les données dont on dispose. Par exemple, si la population est "français de 20 ans à moins de 30 ans", période d'observation 1/1/88 au 31/12/88 on a $T = 1^*$, et l'effectif moyen à risque pendant la période sera la moyenne des effectifs de la classe d'âge au 1/1/88 au 31/12/88 (l'âge au 1er janvier est facilement connu, puisqu'il est déterminé uniquement par l'année de naissance). Si la maladie étudiée n'est pas rare, il faut soustraire de ces effectifs les personnes malades.

 * Si : unité de temps = année



$PA = N \times T$
 \bar{N} = effectif moyen à
 risque pendant la
 période

$PA = \sum t_i$
 t_i = le temps que
 le sujet i doit être
 compté :
 $t_i = T$ pour un sujet
 à risque toute la
 période
 Sinon : temps de suivi
 jusqu'à la maladie ou
 le décès pour autre cause

ou autre mode de
 calcul de la surface
 hachurée.

Remarque 1 : le calcul d'un taux d'incidence à partir d'une période T , tel qu'il est présenté ci-dessus, suppose que $\hat{\lambda}$ **soit homogène pour toute la population et constant pendant la période**. D'où l'intérêt d'utiliser une population ouverte qui ne vieillit pas (limité à une tranche d'âge). Dans ce cas, il ne reste qu'à vérifier que $\hat{\lambda}$ est le même quel que soit le temps calendaire (ce qui n'est pas toujours le cas, penser par exemple au taux d'incidence de la grippe, qui varie d'un mois à l'autre ou même dans un mois).

Remarque 2 : la formule donnant le taux d'incidence montre que **la mesure dépend de l'unité de temps** : la surface de l'aire hachurée sera différente selon l'unité de temps choisie. En fait, en pratique, il est rare que l'unité soit autre que l'année.

Sauf affection frappant indistinctement toutes les classes d'âge, **un taux d'incidence est propre à une classe d'âge**. De ce fait, même si les données disponibles sont **des données de cohorte**, les personnes-années concernent une classe d'âge.

SECONDE MESURE D'INCIDENCE : RISQUE OU INCIDENCE CUMULEE

Risque R_{t_0, t_1} : le risque est la probabilité qu'un sujet indemne de la maladie au temps t_0 commence à être malade entre t_0 et t_1 , conditionnellement au fait qu'il n'est pas décédé d'autre cause durant la période.

Dans quel cas peut-on calculer facilement un risque ?

D'après la définition, il suffirait de suivre une cohorte et de rapporter l'effectif de nouveaux malades à l'effectif de non malades au temps t_0 .

$$R_{t_0, t_1} = \frac{M_{t_0, t_1}}{N_{t_0}}$$

M_{t_0, t_1} <----- nouveaux malades entre t_0 et t_1
 N_{t_0} <----- non malades à t_0

Cependant, dans la définition, il y a aussi "conditionnellement au fait qu'il n'est pas décédé d'autre cause durant la période". Si cela peut être négligé (risque de décès très faible), alors la formule ci-dessus peut être utilisée. S'il faut tenir compte des décès pour autre cause, il y a un problème : que faire d'un sujet qui est décédé entre t_0 et t_1 ? S'il avait vécu jusqu'à t_1 , il aurait pu devenir malade...

Caractéristiques d'un risque par rapport au taux :

Risque	Taux
A un sens pour un sujet : probabilité de survenue de ...	Concerne une population : vitesse de survenue des maladies dans la population
Temps = le temps qui passe pour un sujet (augmentation d'âge)	Temps = temps d'observation (l'âge peut rester constant)
Nombre sans dimension : ne dépend pas de l'unité de temps	Dimension : inverse du temps
Entre 0 et 1 : c'est une probabilité	Peut être très élevé si beaucoup d'événements se passent en peu de temps (exemples : cataclysme, ...)
Difficile à calculer, sauf cas particuliers	Facile à calculer, à partir de population ouverte ou de cohorte

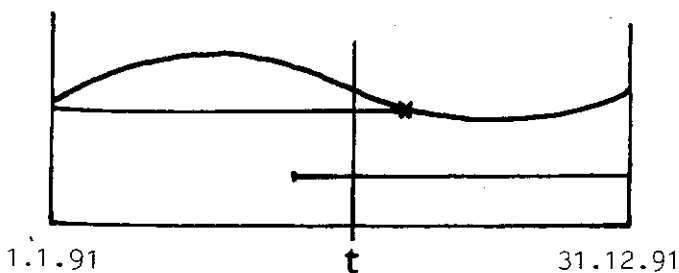
RISQUE ET TAUX. PERIODE COURTE

Prenons comme exemple :

Incidence annuelle de survenue d'un cancer ORL pour un homme français de 55 ans, à partir des données de l'année 1991.

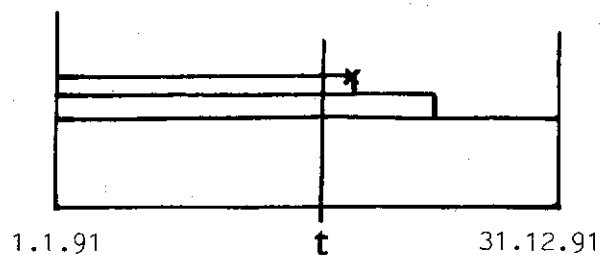
Il y a un calcul possible en termes de risque, et un autre en taux d'incidence :

Taux (pop. ouverte)



la population à t est celle qui a plus de 55 ans, moins de 56 ans à t , non malade à la date t

Risque (cohorte)



la population à t est celle qui a plus de 55 ans moins de 56 ans, au 1/1/91 non malade à t .

$$\text{Taux} = \frac{\text{Nombre de nouveaux cas (+ 55, - 56 ans à la date de diagnostic)}}{\text{Effectif moyen des 55 ans pendant l'année } (T = 1)}$$

$$\text{Risque} = \frac{\text{Nombre de nouveaux cas (+ 55, - 56 ans au 1/1/91)}}{\text{Effectif des 55 ans au 1/1/91}}$$

(on néglige ici les décès pour autre cause).

Le dénominateur est un effectif moyen pour la période s'il s'agit d'un taux, un effectif au début de la période s'il s'agit d'un risque.

Les nouveaux cas, au numérateur, sont caractérisés par l'âge au diagnostic (taux) ou par l'âge au début de la période (risque).

Numériquement, cependant, les différences seront faibles. La différence la plus importante, ici, tient à la différence moyenne de 1/2 année entre les deux populations, le risque calculé étant plus proche d'un risque entre 55 ans 1/2 et 56 ans 1/2.

Un calcul plus exact en terme de risque consisterait à suivre un groupe d'hommes de leur anniversaire de 55 ans à leur anniversaire de 56 ans, et de prendre :

$$\text{Risque} = \frac{\text{Nombre de nouveaux cas dans le groupe entre les deux anniversaires}}{\text{Effectif du groupe}}$$

Dans ce cas il n'est pas possible de se limiter à une année civile (1991, par exemple) car les dates des anniversaires sont réparties sur l'année.

RISQUE SUR UNE PERIODE A TAUX CONSTANT

Ce qui précède s'écrit :

$$R_{55 \rightarrow 56} = \hat{\lambda}_{55 \text{ ans}}$$

(taux annuel proche de risque en un an).

Même si l'on supposait que le taux d'incidence annuel reste constant sur une tranche d'âge donnée, incluant la période 55-56 ans, par exemple de 55 à 59 ans, égal à 55-59, l'expression ci-dessus resterait un peu approximative. L'expression exacte (donnée sans démonstration) serait :

$$R_{55 \rightarrow 56 \text{ ans}} = 1 - e^{-\hat{\lambda}_{55-59}}$$

(effectuer un développement limité de l'exponentielle pour retrouver la valeur approchée).

Si l'on cherche maintenant le risque de développer un cancer entre 55 et 59 ans, le taux d'incidence étant supposé constant sur la période

$$R_{55 \rightarrow 59 \text{ ans}} = 1 - e^{-[\hat{\lambda}_{55-59}] \times 5} \approx 5 \hat{\lambda}_{55-59}$$

Ce risque est donc à peu près 5 fois le risque annuel.

Il faut bien noter que dans les expressions précédentes, indiquer le risque R par 55---> 59 ans signifie que l'on s'intéresse à un risque en 5 ans. Par contre λ indicé par 55-59 reste un taux annuel, supposé constant sur un intervalle d'âge.

De façon générale, on retrouve la formule déjà donnée, s'agissant de décéder durant une période donnée :

Si le taux d'incidence est constant et égal à $\hat{\lambda}$ pendant toute la période t_0, t_1 , alors le risque de survenue de la maladie entre t_0 et t_1 est :

$$R_{t_0, t_1} = 1 - e^{-\hat{\lambda} [t_1 - t_0]}$$

RISQUE SUR UNE PERIODE, TAUX VARIABLE

Soit à calculer le risque de survenue d'un cancer ORL entre 30 et 64 ans. Le taux d'incidence est bien différent à 30 ans et à 60 ans, mais on supposera des taux λ_i constants sur des intervalles de temps de longueur D_i :

$$R_{30, 64} = 1 - e^{-\sum \hat{\lambda}_i \Delta_i}$$

2. Causalité en épidémiologie

Le risque d'avoir une maladie n'est à l'évidence pas le même pour tous les individus, ou tous les sous-groupes de la population.

En effet, le risque peut varier en fonction de certaines caractéristiques individuelles, comme le sexe, l'âge, certains facteurs socio-économiques ou certains paramètres biologiques comme la pression artérielle pour les maladies cardio-vasculaires. Ce sont ces facteurs qu'on appelle **facteurs de risque**.

On appelle facteur de risque toute variable liée statistiquement à l'évènement étudié. Il faut entendre facteur de risque au sens large, et pas seulement au sens péjoratif qu'on lui donne habituellement, bien qu'il s'agisse d'une habitude très répandue : en toute rigueur, un facteur qui diminuerait l'incidence d'une maladie dans une population (par exemple, on pense actuellement que le bêta-carotène peut jouer un rôle protection contre certains cancers) devrait être appelé un facteur de risque, puisqu'il influence la probabilité de l'évènement.

Cependant, ce n'est pas parce qu'un facteur est statistiquement lié à l'évènement étudié qu'il en est la cause.

La notion de cause en épidémiologie a un sens particulier. En effet, un facteur, pour être causal vis-à-vis d'une maladie, n'a pas obligatoirement besoin d'être présent d'une manière nécessaire, ni même suffisante.

Il en est ainsi, par exemple, pour les cardiopathies ischémiques, pour lesquelles des causes multiples ont été mises en évidence. Plus le nombre de facteurs étiologiques présents est important, plus le risque est élevé. Mais la maladie peut apparaître en l'absence d'un ou de plusieurs de ces facteurs. Autrement dit, un sujet qui fume, avec une hypercholestérolémie et une hypertension artérielle, aura un risque plus élevé qu'un individu présentant seulement un de ces facteurs. Néanmoins, il est possible qu'une cardiopathie ischémique survienne chez un individu qui, par exemple, ne fume pas. On trouvera, de même, des fumeurs qui ne seront jamais atteints.

Aussi a-t-on adopté en épidémiologie une définition de la cause plus opérationnelle, qui est de nature probabiliste : **on dit qu'un facteur est cause d'une maladie si une modification de sa fréquence (ou de sa valeur moyenne s'il est quantifiable) entraîne une modification de la fréquence de cette maladie.**

Toute la problématique de l'épidémiologie analytique est en fait, devant une association statistique entre un facteur et un évènement, de savoir si la relation ainsi observée est causale. En effet, il existe différents types d'associations statistiques qui n'ont pas le même sens.

En premier lieu, il ne faut pas oublier qu'une association statistique peut être fallacieuse, c'est-à-dire due à un **biais** (cf plus loin : les biais en épidémiologie), à une erreur quelconque ou au hasard.

Ainsi n'est-il pas rare, par exemple, de constater que, pour une même pathologie, le pronostic est nettement plus défavorable pour les patients soignés en milieu hospitalier spécialisé que pour ceux traités en ville ; mais l'explication de la liaison observée entre évolution de la maladie et type de prise en charge médicale réside dans le fait que ce sont les malades initialement les plus gravement atteints qui vont à l'hôpital, alors que les cas bénins sont soignés par leur médecin traitant habituel.

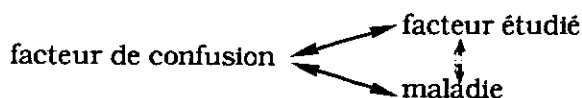
Un autre type d'erreur consiste à interpréter une association "à l'envers". Ainsi, si on observe que les fumeurs ont fréquemment l'index jaune, l'erreur consisterait à interpréter cette association dans le sens : index jaune ---> fumeur, alors que c'est parce que ces sujets sont plus souvent fumeurs qu'il ont plus fréquemment l'index jaune.

Cet exemple est caricatural, mais il existe de nombreuses situations où le sens de la relation entre maladie et facteur de risque n'est pas aussi évident (par exemple, si on étudie le rôle pathogène de certains médicaments : a-t-on eu telle maladie à la suite de la prise de ce médicament ? Ou, au contraire, a-t-on commencé à prendre ce médicament parce qu'on avait déjà les premiers signes de cette maladie ?).

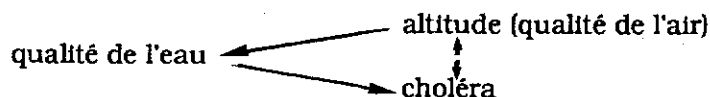
Un autre type d'association est **l'association indirecte**. Dans un tel cas, une variable est liée à l'évènement étudié, mais en fait ne joue un rôle que par l'intermédiaire d'une autre variable. Un exemple classique est celui de l'obésité et des cardiopathies ischémiques. Il existe une association forte entre obésité et incidence de la maladie, mais en fait cette association s'explique par le fait que l'obésité entraîne une augmentation de la pression artérielle : c'est l'hypertension qui est la cause de l'augmentation de l'incidence, alors que l'obésité isolée ne modifie pas celle-ci. On peut schématiser cette association indirecte de la façon suivante :

Obésité -----> Hypertension -----> Maladie

Un cas particulier de l'association indirecte est celle qui est due à un **facteur de confusion**. Ce type d'association est particulièrement fréquent en épidémiologie. Il se produit dans les circonstances suivantes. On observe une liaison entre un facteur dont on étudie le rôle et une maladie ; par ailleurs, il existe un autre facteur, le facteur de confusion, lui-même lié à la fois au facteur étudié et à la maladie. La présence du facteur de confusion peut produire l'association observée initialement, même si le facteur dont on étudie le rôle et la maladie sont indépendants, selon le schéma :



Un exemple historique de ce type d'association est celui de l'étude de Snow, qui a observé au XIX^e siècle à Londres, une relation entre la survenue du choléra et l'altitude, les quartiers les plus bas de la ville étant les plus touchés. La qualité de l'air, meilleure dans les quartiers plus élevés, a été incriminée. Or on sait qu'en fait c'est la qualité de l'eau qui doit être mise en cause. Les quartiers les plus bas de Londres étaient simplement ceux qui recevaient également de l'eau de mauvaise qualité. Le schéma était donc :



La flèche en pointillé indique une simple association statistique sans signification causale.

Dans l'exemple qu'on vient de prendre, le facteur de confusion était associé au facteur étudié et à la maladie par une liaison de même sens : la qualité de l'eau était d'autant plus mauvaise que l'altitude était faible, et le choléra fréquent, ce qui expliquerait la relation apparente entre altitude et choléra. Si les liaisons entre facteur de confusion et maladie d'une part, et facteur de confusion et facteur de risque d'autre part, sont en sens inverse, une éventuelle relation véritable entre facteur de risque et maladie peut disparaître, et on conclura à tort qu'il n'y a pas de liaison entre le facteur et la maladie.

Il existe des méthodes qui permettent de prendre en compte un ou plusieurs facteurs de confusion et les neutraliser, afin de vérifier si le facteur étudié est bien lié lui-même à l'évènement. Encore faut-il avoir pensé à l'avance à l'existence des facteurs de confusion. Il est donc très important, pour étudier le rôle causal d'un facteur, d'être très soigneux dans le choix des variables à prendre en compte et dans l'analyse des biais susceptibles de se produire.

Cependant, on n'est jamais sûr d'avoir éliminé tous les tiers-facteurs. Il peut toujours exister un facteur inconnu dont on n'a pas tenu compte et qui explique la relation observée. **Pour affirmer une liaison causale, il faudrait faire une enquête expérimentale, autrement dit disposer de deux groupes absolument comparables**, c'est-à-dire constitués par tirage au sort, un groupe étant exposé au facteur, l'autre non, et comparer dans chacun des groupes l'incidence de la maladie. Cela est impossible dans la plupart des cas pour des raisons éthiques et pratiques : en épidémiologie, il est exceptionnel d'être dans une telle situation expérimentale. On est pratiquement toujours en **situation d'observation**, c'est-à-dire où l'enquêteur ne peut que tenir compte au mieux des circonstances du terrain, sans pouvoir agir sur les conditions d'exposition aux facteurs étudiés.

En l'absence de preuve expérimentale, on peut s'entourer d'un certain nombre de garanties qui peuvent permettre d'aboutir à une quasi certitude. **Les principaux arguments** en faveur d'une relation causale sont les suivants (on prendra l'exemple classique du tabac et du cancer du poumon) :

- **Séquence dans le temps** : l'exposition au facteur présumé causal doit précéder l'apparition de la maladie (le cancer du poumon survient chez les fumeurs de longue date).
- **Constance de l'association et reproductibilité** : la relation facteur -maladie doit être retrouvée dans différentes populations et différentes conditions (de nombreuses études ont confirmé l'existence de la relation tabac-cancer du poumon).

- **Force de l'association statistique** : plus une relation entre un facteur et une maladie est forte, plus il est difficile d'imaginer qu'un facteur de confusion inconnu puisse l'expliquer (dans le cas tabac-cancer du poumon, on peut calculer qu'un facteur inconnu capable d'expliquer la relation observée devrait être environ dix fois plus fréquent chez les fumeurs que chez les non-fumeurs : on peut raisonnablement penser qu'un tel facteur aurait déjà été mis en évidence).
- **Existence d'une relation dose-effet** (le risque de cancer du poumon augmente avec la quantité de tabac fumée).
- **Effet d'une intervention supprimant le facteur de risque** (en diminuant l'exposition à ce facteur) : si le facteur joue un rôle causal pour la maladie, sa suppression dans la population (ou sa diminution) doit diminuer le risque de la maladie (on sait qu'après cessation du tabagisme, le risque de cancer du poumon diminue progressivement jusqu'à rejoindre après quelques années le risque de non-fumeurs n'ayant jamais fumé). Ce critère, quand on peut l'observer, est un des meilleurs arguments en faveur de la causalité.
- **Cohérence avec les connaissances actuelles** : plausibilité biologique, acceptabilité scientifique (le risque du cancer du poumon est plus élevé chez les sujets qui inhalent la fumée ; de plus, on a mis en évidence sur des animaux l'effet cancérigène de certaines substances composant la fumée de tabac).

Certains critères sont de nature logique : séquence temporelle (critère absolument nécessaire), existence d'une relation dose-effet, effet de l'intervention ; d'autres sont de nature statistique (force de la relation). Les autres reposent plutôt sur le "bon sens scientifique" (constance, reproductibilité, cohérence avec les connaissances).

Il est important de souligner que la plupart des critères cités ne se retrouvent que rarement dans une seule étude épidémiologique ; notamment le critère de constance et de reproductibilité implique que plusieurs études soient réalisées dans des conditions différentes. C'est la répétition d'études dont les résultats partiels vont dans le même sens et se complètent, qui finalement permettra d'obtenir un consensus en faveur de l'existence d'une relation causale. Mais, évidemment, ce processus peut être long et s'étaler sur de nombreuses années.

II - PRINCIPAUX MODELES D'ENQUETE ETIOLOGIQUE. **MESURES D'ASSOCIATION ENTRE FACTEUR DE RISQUE ET MALADIE**

A. PRINCIPES GENERAUX

Une étude épidémiologique à visée étiologique a pour objectif principal de vérifier l'hypothèse d'une relation causale entre l'exposition à un facteur de risque et la survenue d'un problème de santé. Elle se situe donc dans un contexte où on a de bonnes raisons de soupçonner l'existence d'un rôle causal du facteur étudié, généralement à la suite de premiers travaux de nature descriptive qui ont permis d'observer une association statistique, et/ou d'observations cliniques ou biologiques.

Le but d'une enquête à visée explicative est de réunir le maximum d'arguments correspondant aux critères de causalité énumérés ; l'accumulation d'arguments de ce type finira par être suffisamment convaincant.

Nous verrons que la nature du problème étudié et le contexte concret dans lequel se déroulera l'étude imposent le choix de modèles d'investigation sensiblement différents. Mais la méthode générale de tous les pays d'enquête explicative est identique : il s'agit toujours de comparer des groupes de sujets diversement exposés au facteur de risque soupçonné, et parmi lesquels certains sont atteints de la maladie étudiée et d'autres non.

Il est en effet impossible de juger du rôle causal éventuel d'un facteur en n'observant que des sujets atteints de la maladie : tel facteur peut être très fréquemment présent chez les malades mais, si on ne connaît pas sa fréquence chez les non-malades, on ne peut rien dire de son rôle (cette dernière fréquence peut en effet être aussi élevée, voire plus). Il s'agit là d'une différence fondamentale entre approche clinique et épidémiologique : la première, qui ne travaille que sur des sujets malades, ne peut que soulever des hypothèses sur le rôle causal de certains facteurs, alors que la seconde, en comparant malades et non-malades, pourra infirmer ou valider ces hypothèses.

Une autre caractéristique importante de l'épidémiologie à visée explicative est que l'observation des sujets composant les groupes à comparer se fait toujours au niveau individuel, afin de recueillir (par interrogatoire, examen, analyse de documents les concernant, etc) des informations les plus précises possible pour chacun d'eux portant, pour chaque sujet, à la fois sur la maladie et les facteurs de risque. L'avantage de disposer de données individuelles est qu'il est possible d'être beaucoup plus précis et de tenir compte simultanément de données plus nombreuses, pour chaque sujet ; surtout on dispose pour chacun d'eux simultanément de données sur la maladie, les facteurs de risque et les facteurs de confusion.

NOTION DE RISQUE RELATIF

La définition épidémiologique de la cause implique qu'une modification du niveau d'exposition à un facteur dans la population entraîne une modification de la fréquence de la maladie dans cette population, pour qu'on puisse affirmer que le facteur en question joue un rôle causal vis-à-vis de la maladie.

Une enquête explicative aura donc pour but de vérifier que tel est bien le cas. Pour cela, le meilleur modèle est de comparer l'incidence de la maladie dans des groupes de sujets diversement exposés au facteur de risque. En effet, seule l'incidence est une mesure de la fréquence de la maladie permettant de juger du rôle d'un facteur sur la survenue de la maladie, puisqu'elle mesure l'apparition des nouveaux cas dans une population pendant une certaine période (la prévalence, qui reflète les cas existants pendant la période, est influencée par la durée de la maladie, l'efficacité des soins médicaux, etc.).

Pour juger du fait que l'exposition au facteur étudié modifie l'incidence de la maladie, on utilise un indice épidémiologique appelé risque relatif (RR), qui permet de comparer des groupes de sujets soumis à des expositions différentes.

Prenons un exemple où on veut comparer deux groupes de sujets, les uns étant exposés au facteur étudié, les autres ne l'étant pas. Le risque relatif est le rapport du risque dans le groupe exposé au risque dans le groupe non-exposé.

$$RR = \frac{RE^+}{RE^-}$$

La valeur du risque relatif s'interprète de la façon suivante : si le facteur étudié ne joue pas un rôle causal, il ne doit pas exister de différence entre les sujets exposés et les non-exposés ; dans ce cas, le risque relatif doit être égal à 1 ; s'il est supérieur à 1, cela signifie que la présence du facteur entraîne une augmentation de la probabilité d'apparition de la maladie (ou une diminution de cette probabilité s'il est inférieur à 1). Un risque relatif de 3 (ou de 10) doit donc être interprété de la façon suivante : les sujets exposés au facteur de risque ont une probabilité 3 fois (10 fois) plus élevée d'avoir la maladie que les non-exposés. **Le risque relatif est donc la mesure du rôle étiologique du facteur de risque.** Nous reviendrons plus loin sur le calcul du RR (ou de mesures équivalentes) dans différents contextes.

B. LE MODELE DE BASE : L'ENQUETE DE COHORTE (OU "EXPOSES/NON-EXPOSES")

Pour pouvoir établir les risques relatifs qui permettront de juger du rôle étiologique d'un facteur de risque, il faut recueillir les données entrant dans les calculs, c'est-à-dire les données d'incidence et d'exposition au facteur de risque étudié : c'est le rôle des enquêtes épidémiologiques, qui sont des observations organisées dans la population, dont l'objectif est de recueillir les informations utiles tout en s'assurant que des biais et des facteurs de confusion ne viendront pas obscurcir l'interprétation des résultats.

Le type d'enquête le mieux adapté pour juger du rôle causal d'un facteur de risque est l'enquête de cohorte, appelée aussi exposés/non-exposés. Le principe en est simple.

On définit tout d'abord une période d'observation dont la durée est généralement liée au délai susceptible de s'écouler entre le début de l'exposition et l'apparition de la maladie (parfois plusieurs années, voire des dizaines d'années).

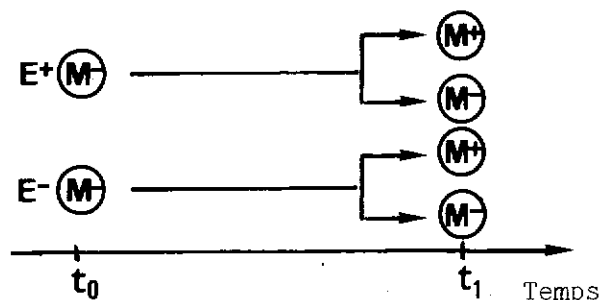
On choisit ensuite un groupe de sujets qui sont tous indemnes de la maladie étudiée au début de la période d'observation. Ce groupe est parfois composé de la totalité de la population étudiée, mais le plus souvent il s'agit d'un échantillon. Le groupe de sujets entrant sous observation fait l'objet, au début de la période, d'un recueil de données les concernant (notamment sur l'exposition au facteur de risque étudié, les facteurs de confusion possibles, l'absence de la maladie et de maladies associées).

Dans certains cas, il est possible de distinguer d'emblée au sein de l'ensemble des sujets de l'étude un groupe exposé au facteur de risque et un groupe non-exposé (par exemple, les travailleurs soumis à une nuisance professionnelle et des sujets n'ayant jamais travaillé dans la même industrie, ou bien si le facteur de risque est une caractéristique biologique que certains sujets possèdent et d'autres pas). Dans d'autres cas, on ne fait pas de distinction entre les sujets du groupe observé, et ce n'est qu'à la fin de la période d'étude, grâce aux données recueillies individuellement sur les sujets, qu'on répartira ceux-ci en fonction du niveau d'exposition.

Ce niveau d'exposition peut être exprimé de façon quantitative (nombre de kilogrammes de tabac fumé, par exemple), qualitative ou semi-qualitative (non exposés, peu exposés, moyennement exposés, etc.).

Pendant toute la durée de la période d'observation, les sujets sont systématiquement suivis afin (quand cela est nécessaire) de mesurer d'éventuelles modifications d'exposition, et d'observer l'incidence de la maladie étudiée. C'est parce que dans ce type d'enquête on attend la survenue de la maladie au fur et à mesure de l'écoulement du temps qu'on l'appelle souvent "**enquête prospective**".

A la fin de la période d'observation, on dispose alors de toutes les données utiles pour calculer les risques relatifs : niveaux d'exposition et incidence dans les groupes comparés.



Parmi les sujets, tous non-malades (M^-) au temps t_0 , certains sont exposés au facteur du risque étudié (E^+) et certains ne sont pas exposés (E^-). Dans chacun des groupes, certains deviendront malades (M^+) au temps t_1 .

1. Calcul du risque relatif

Le risque relatif RR est le rapport de deux taux d'incidence, ou de deux risques, entre deux populations dont l'une est exposée et l'autre non.

La situation la plus habituelle est celle où les calculs sont réalisés à partir de personnes-années :

	E^+	E^-	
Evénements	M_1	M_0	$\hat{\lambda}_{E^+} = M_1/PA_1$
PA	PA_1	PA_0	$\hat{\lambda}_{E^-} = M_0/PA_0$

Les deux groupes sont E^+ (exposés), E^- (non exposés), M_1 et M_0 sont les effectifs de nouveaux malades dans les deux groupes correspondant aux personnes-années PA_1 et PA_0 .

$$RR = [M_1/PA_1]/[M_0/PA_0]$$

Si l'on a pu calculer ou estimer des **risques** RE^+ et RE^- dans les deux populations, on parlera également de risque relatif :

$$RR = RE^+/RE^-$$

En l'absence de différence entre les deux groupes, exposés et non, l'espérance de RR est 1. L'interprétation, par exemple, d'un RR de 1.8 est : les personnes exposées ont un risque multiplié par 1.8 d'être malades, comparées aux non-exposées.

Remarque

La présentation ci-dessus correspond aux études de cohorte, car le calcul du risque relatif repose sur la connaissance des taux d'incidence ou des risques dans les deux populations, donc qu'il y ait eu un suivi dans le temps, permettant un comptage de nouveaux cas.

On verra que, cependant, une estimation de RR peut être obtenue dans d'autres cas (cf plus loin, enquêtes cas/témoins).

2. Intervalle de confiance de RR

On considérera que Log RR est distribué normalement, avec un écart-type :

$$s = \sqrt{\frac{1}{M_1} + \frac{1}{M_0}}$$

dans la situation présentée précédemment.

D'où l'intervalle de confiance pour le risque relatif RR au niveau

$$\left[\text{RR} e^{-Z_{\alpha/2} \sqrt{\frac{1}{M_1} + \frac{1}{M_0}}} ; \text{RR} e^{+Z_{\alpha/2} \sqrt{\frac{1}{M_1} + \frac{1}{M_0}}} \right]$$

Si tous les sujets ont été suivis pendant le même temps, on peut également calculer le risque relatif à partir du tableau des effectifs :

	E+	E-
M+	a	b
M-	c	d

$$\text{RR} = [(a/a + c)] / [(b/b + d)]$$

L'intervalle de confiance est alors différent et se calcule en considérant que Log RR est distribué normalement, avec un écart-type :

$$s = \sqrt{c/[(a + c) a] + d/[(b + d) b]}$$

Bien entendu, sur ce principe simple viennent éventuellement se greffer des méthodes de calcul du RR plus complexes : si au lieu d'avoir deux groupes (exposés, non-exposés), on en a plusieurs en fonction de niveaux d'exposition différenciés ; s'il existe des facteurs de confusion pour lesquels des données individuelles ont été recueillies ; si on étudie non pas un, mais plusieurs facteurs de risque simultanément.

Lorsque, parmi le groupe des exposés, il existe des sujets exposés à des niveaux quantitatifs différents, il est très important de calculer des risques relatifs en fonction du niveau d'exposition : on pourra ainsi comparer l'incidence entre les non-exposés et des sujets peu, moyennement, très exposés. L'existence d'une augmentation du risque relatif liée à l'augmentation de l'exposition permet de mettre en évidence une relation dose-effet, ce qui est un argument fondamental en faveur d'un rôle causal du facteur de risque.

Le principe des méthodes utilisées pour rechercher une relation dose-effet est le suivant. Supposons qu'on puisse distinguer 4 niveaux d'exposition au facteur de risque : E_0 : non exposé ; E_1 : exposition faible ; E_2 : exposition moyenne ; E_3 : exposition forte. On répartit alors les sujets de l'enquête, en fonction de leur statut vis-à-vis de la maladie et de l'exposition, non plus en un seul tableau comme on l'a vu plus haut, mais dans les 3 tableaux suivants :

	M^+	M^-		M^+	M^-		M^+	M^-
E_1	a_1	b_1	E_2	a_2	b_2	E_3	a_3	b_3
E_0	c	d	E_0	c	d	E_0	c	d

L'effectif des non exposés E_0 , malades et non-malades, est le même dans les 3 tableaux, car ces sujets servent de référence ; il vaut respectivement c et d .

Ces tableaux permettent de calculer 3 risques relatifs différents, en comparant à chaque fois l'incidence chez les exposés (à un niveau faible, moyen et fort), aux non-exposés E_0 :

$$RR_1 = \frac{a_1 d}{b_1 c} ; RR_2 = \frac{a_2 d}{b_2 c} ; RR_3 = \frac{a_3 d}{b_3 c}$$

S'il existe une relation dose-effet, on aura :

$$RR_1 < RR_2 < RR_3$$

Le principe de la méthode est le même si on travaille avec des personnes-années au lieu d'effectifs : il suffit de calculer les risques relatifs à partir du nombre de malades dans chaque catégorie d'exposition, rapporté au nombre de personnes-années de l'ensemble des sujets de cette même catégorie d'exposition.

Le principe du calcul est donc simple : il consiste à faire des comparaisons à niveau d'exposition équivalent.

3. Prise en compte des facteurs de confusion

C'est le même principe qui est utilisé pour neutraliser le rôle d'un facteur de confusion : il suffit de répartir les sujets de l'étude en "strates", c'est-à-dire en groupes où l'exposition au facteur de confusion est identique, et de calculer des risques relatifs à niveau d'exposition équivalent au facteur de confusion. S'il existe une relation entre facteur étudié et maladie, on doit logiquement retrouver un risque relatif supérieur à 1 dans chaque strate.

Par exemple, si l'âge est un facteur de confusion dans la relation entre un facteur et une maladie, on peut construire le tableau suivant, où tous les sujets ont été regroupés en 3 strates correspondant à 3 classes d'âge.

	M+	M-	
E+	a ₁	b ₁	1re classe d'âge
E-	c ₁	d ₁	
E+	a ₂	b ₂	2è classe d'âge
E-	c ₂	d ₂	
E+	a ₃	b ₃	3è classe d'âge
E-	c ₃	d ₃	

A partir de ce tableau, on pourrait calculer un risque relatif pour chacune des 3 strates. Mais ceci est gênant pour l'interprétation, car on aura trois valeurs de risque relatif. Or on voudrait en fait connaître la valeur du risque relatif pour le facteur réellement étudié, sans se soucier de l'exposition au facteur de confusion. Il existe pour cela des méthodes qui permettent d'obtenir une estimation du "risque relatif ajusté", c'est-à-dire un risque relatif qui tient compte de l'exposition au facteur de confusion. Ces méthodes sont simples si on n'a qu'un seul facteur de confusion. Mais en pratique, c'est rarement le cas, et souvent plusieurs expositions doivent être prises en compte simultanément ; on utilise alors des méthodes plus complexes (dites "multivariées" ou "multidimensionnelles"), mais qui fondamentalement reposent sur le même principe : estimation du risque relatif pour un facteur, à niveau constant des autres facteurs.

4. Effet conjoint de plusieurs facteurs

Lorsqu'on est amené à tenir compte du rôle de plusieurs facteurs simultanément, on peut se poser d'autres questions. Il ne s'agit plus d'éliminer l'effet de certains facteurs pour étudier le rôle spécifique de l'un d'entre eux, comme on vient de le voir. On cherche au contraire à étudier l'effet conjoint de l'exposition simultanée à deux ou plusieurs facteurs de risque.

Les variations de l'effet qui peuvent être dues à une exposition simultanée sont de deux types principaux :

- **Effet additif** : on dit que 2 facteurs ont un effet additif si l'effet de leur combinaison est la somme des effets de chacun d'eux isolément.
- **Effet multiplicatif** : on dit que 2 facteurs ont un effet multiplicatif si l'effet de leur combinaison est le produit des effets de chacun d'eux isolément.

Il existe des méthodes pour analyser l'effet conjoint de plusieurs facteurs, c'est-à-dire pour savoir s'il est de nature additive ou multiplicative. Elles sont relativement complexes, car l'effet conjoint n'est pas toujours aussi simple et peut prendre des formes intermédiaires.

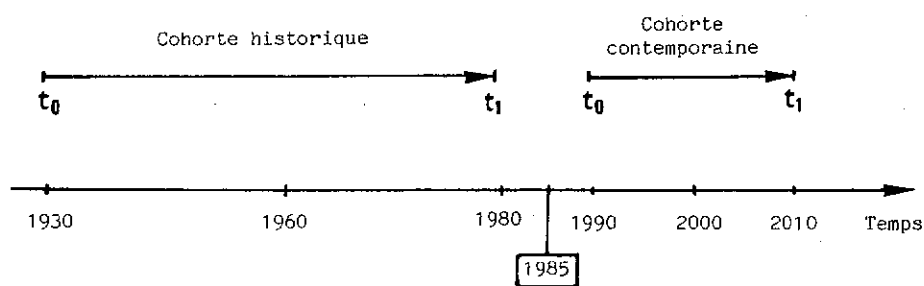
5. Principales variantes de l'enquête de cohorte

Ce que nous avons vu jusqu'ici correspond au modèle général de l'enquête exposés/non-exposés. Il existe bien entendu des variantes.

Ainsi, bien que les enquêtes exposés/non-exposés soient prospectives, en ceci qu'il faut attendre la survenue de la maladie, on peut situer le début de la période d'observation de façon très antérieure à la réalisation de l'enquête. Par exemple, si on veut étudier la mortalité des travailleurs soumis à une certaine nuisance professionnelle, on peut décider de définir comme période d'observation 1930-1980 (alors même que l'étude est réalisée en 1985) : il suffit de connaître la totalité des travailleurs présents dans l'entreprise étudiée au 1er janvier 1930 et de savoir, sujet par sujet, lesquels sont décédés et ceux qui sont encore vivants au 31 décembre 1980. Lorsqu'on travaille ainsi *a posteriori*, on dit qu'il s'agit d'une "enquête prospective dans le passé", ou encore d'une étude de "cohorte historique" (cf un peu plus loin la définition du terme cohorte). Lorsque les données nécessaires sont disponibles, on utilise volontiers cette variante pour gagner du temps, car on n'a ainsi pas besoin d'attendre que la période d'observation soit écoulée.

Habituellement, une étude de cohorte historique consiste à n'observer véritablement qu'un groupe de personnes exposées, le groupe non-exposé étant représenté par la population générale pour laquelle on possède des données comparables. Ainsi, dans l'exemple précédent, on dispose du groupe des travailleurs présents dans l'entreprise au 1er janvier 1930 et on le suit jusqu'au 31 décembre 1980 ; on a donc à la fin de la période le nombre total de sujets concernés et le nombre de décès. Or on peut parfois disposer de données équivalentes pour la population générale et pour la même période. On peut donc comparer la mortalité chez les exposés (les travailleurs concernés) et les non-exposés (la population générale). Dans un tel cas, on ne calculera habituellement pas un risque relatif, notamment en raison de la disproportion entre l'effectif du groupe exposé et celui de la population générale, mais un Ratio Standardisé de Mortalité qui compare le nombre de décès attendus dans le groupe exposé, s'il avait une mortalité identique à celle de la population générale, avec le nombre de décès effectivement observés dans le groupe exposé*.

Le schéma suivant illustre la différence entre enquêtes exposés/non exposés, selon qu'on utilise la variante "cohorte historique" ou la méthode classique, dite aussi "cohorte contemporaine".



Dans les deux cas, on a bien une étude "prospective" puisqu'on suit un groupe de personnes, indemnes au début de la période d'observation (t_0) jusqu'à la fin de cette période (t_1). Mais par rapport à la date où on décide de faire l'enquête (1985 dans cet exemple), on travaille dans le passé pour l'enquête historique ($t_0 = 1930$; $t_1 = 1980$), et dans le futur pour l'enquête contemporaine ($t_0 = 1990$; $t_1 = 2010$).

* Le Ratio Standardisé de Mortalité s'interprète comme un risque relatif ; pour son calcul cf par exemple M. Golberg et al., *L'épidémiologie sans peine*, Paris, Ed. Frison-Roche.

Il semblerait à première vue que la solution "cohorte historique" est meilleure, car elle permet d'avoir très rapidement les résultats recherchés. Elle a cependant un inconvénient majeur : la connaissance des expositions aux facteurs de risque étudiés est de moins bonne qualité, car elles sont anciennes et donc difficiles à retrouver et souvent imprécises, alors que le recueil de données pour une étude contemporaine peut être planifié soigneusement. De toute façon, la cohorte historique n'est pas toujours possible ; deux conditions sont en effet nécessaires :

- Pouvoir reconstruire *a posteriori* la cohorte, c'est-à-dire toutes les personnes concernées par l'exposition au temps t_0 (c'est le groupe exposé) ;
- Disposer des données concernant l'incidence de la maladie dans le groupe exposé et dans la population générale de référence (considérée comme le groupe non-exposé).

Ces deux conditions - tout à fait nécessaires - ne sont réunies que dans certaines circonstances. C'est habituellement en milieu professionnel (travailleurs d'une entreprise) qu'il est possible de reconstituer une cohorte à partir des fichiers du service du personnel, des archives de la Société, etc. On considère dans ce cas l'ensemble des personnes présentes dans l'entreprise au temps t_0 . Cette méthode a également été employée pour étudier les effets à long terme de certains traitements médicaux (radiothérapie massive, par exemple), car il a été possible de retrouver dans les archives médicales d'un hôpital ou d'un Service les dossiers de tous les patients ayant bénéficié du traitement.

La cohorte reconstituée, il faut identifier tous les sujets ayant présenté la maladie étudiée pendant la période écoulée entre t_0 et t_1 , et disposer des données d'incidence de cette maladie pour la population générale de référence. En pratique, ceci limite les études de cohorte historique à la prise en compte de maladies se traduisant par un décès, car seules les données de mortalité sont enregistrées de façon permanente et exhaustive à l'échelle de la population générale. A partir de cet enregistrement systématique, on peut retrouver tous les cas de la maladie survenus dans la cohorte, et disposer des taux de mortalité pour cette maladie dans la population de référence : il ne reste plus qu'à calculer les RMS (Ratios de Mortalité Standardisés). Rappelons cependant qu'en France la législation actuelle interdit l'accès nominatif aux causes de décès, et rend donc pratiquement impossible l'utilisation de cette méthode, pourtant extrêmement utile (et utilisée dans les autres pays). Signalons aussi que pour les études sur les cancers, certains pays disposent de "Registres" depuis très longtemps, et peuvent donc entreprendre des études de cohorte historique concernant l'incidence des cancers.

Enfin, il faut signaler que les enquêtes exposés/non-exposés s'inscrivent dans le cadre plus général des études de cohorte. **Une cohorte est un groupe de personnes ayant vécu une même expérience (exposition à un facteur de risque par exemple) et suivies dans le temps depuis la date de cette expérience** (qui peut éventuellement être différente d'un sujet à l'autre).

Une enquête exposés/non-exposés est donc une étude de cohorte ; mais toutes les études de cohorte n'ont pas un objectif explicatif visant à comparer l'incidence d'un problème de santé entre des exposés et des non-exposés.

Ainsi, on utilise souvent des études de cohorte pour étudier l'évolution d'un phénomène dans le temps. Par exemple, on peut utiliser cette méthode pour étudier les caractéristiques du développement staturo-pondéral d'une génération d'enfants : on choisit un groupe d'enfants à la naissance et on les suit jusqu'à l'âge adulte en effectuant régulièrement certaines mesures anthropométriques.

Ce type d'étude qui suit l'évolution d'un phénomène dans le temps est souvent appelé également "étude longitudinale", par opposition aux "enquêtes transversales"*. Les études longitudinales apportent des informations que ne peuvent fournir des observations transversales. Ainsi, dans l'exemple choisi, une enquête transversale réalisée à un moment donné, et qui inclurait des sujets de tous âges entre la naissance et l'âge adulte, ne permettrait pas de distinguer clairement ce qui revient, dans l'évolution des paramètres physiologiques, à l'effet de l'âge et à un éventuel effet de génération que seule une méthode d'observation longitudinale peut prendre en compte (à âge égal, des personnes nées à des périodes différentes peuvent avoir des caractéristiques très dissemblables. Par exemple, la taille d'une fille de 15 ans en 1990 est plus élevée que celle d'une fille ayant eu 15 ans en 1930).

6. Avantages et inconvénients des enquêtes exposés/non-exposés ; indications

Le principal avantage des enquêtes de ce type est qu'elles permettent de réunir dans de bonnes conditions des informations nécessaires au calcul direct du risque relatif, qui est par excellence l'indice épidémiologique utilisé pour porter un jugement de causalité : l'enquête exposés/non-exposés est le modèle idéalement adapté à ce type d'objectif, surtout s'il est possible de tirer au sort les sujets qui entreront dans les groupes exposés et non-exposés, réalisant alors une enquête expérimentale (mais on a déjà dit que cela est exceptionnellement réalisable dans un contexte épidémiologique).

L'exposition au facteur de risque étudié étant connue avant l'apparition de la maladie, et les sujets inclus dès le début de la période d'observation, on peut recueillir les données sur l'exposition de la meilleure façon : il n'y aura théoriquement pas de biais concernant la mesure du facteur de risque, pas d'informations manquantes concernant l'exposition, et la mesure de celle-ci peut être précise, car planifiée dès le début de l'enquête.

De plus, par définition, l'incidence de la maladie sera connue, et ceci en fonction de l'exposition au sein des groupes de sujets.

Enfin, un autre avantage de ce type d'enquêtes est qu'elles permettent, le cas échéant, de mettre en évidence le rôle du facteur de risque sur d'autres pathologies que celle qu'il a été initialement prévu d'étudier, puisque les sujets sont suivis de façon prospective (ainsi, les études prospectives sur le rôle du tabac sur les cancers broncho-pulmonaires ont mis en évidence le rôle de ce facteur sur les maladies cardio-vasculaires, le cancer de la vessie, etc.).

Les principaux inconvénients des enquêtes exposés/non-exposés sont liés à son caractère prospectif : long délai avant l'obtention des résultats, lourdeur et coût élevé. En effet, il faut attendre l'apparition de l'évènement étudié ; de plus, les maladies (même si, sur des critères médicaux, elles sont considérées comme fréquentes) ne surviennent généralement que chez un petit nombre de sujets, même parmi les exposés, surtout si le rôle étiologique du facteur de risque n'est pas très fort.

Il faut donc habituellement des groupes très importants suivis pendant longtemps pour qu'on puisse observer des effectifs de cas incidents suffisants pour faire des estimations et des tests.

.....
* Une enquête transversale consiste à observer un phénomène de façon "instantanée", sans suivi des sujets observés.

A titre d'exemple, une des premières enquêtes prospectives concernant le rôle du tabac sur les cancers broncho-pulmonaires a consisté à suivre un groupe d'un million de personnes, fumeurs et non-fumeurs, pendant 5 ans ! Outre des problèmes de coût, la logistique de telles enquêtes est difficile à assurer, et se posent de nombreux problèmes de suivi de l'exposition au facteur de risque, de perdus de vue, etc. qui peuvent occasionner autant de biais.

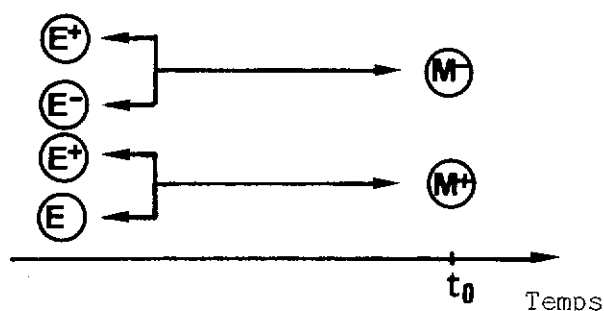
C'est pourquoi les enquêtes exposés/non-exposés sont surtout indiquées lorsque le délai entre exposition et survenue de la maladie n'est pas trop long (sauf à réaliser des études de cohorte historique ; mais, dans ces cas, on a souvent du mal à connaître *a posteriori* les expositions de façon précise), lorsque la maladie est relativement fréquente et lorsqu'on pense que le rôle étiologique du facteur de risque est fort, ces deux dernières conditions permettant de réduire la taille des groupes à comparer.

C. L'ENQUETE CAS/TEMOINS

Lorsque, pour une ou plusieurs des raisons citées plus haut, une enquête de cohorte n'est pas réalisable, on peut utiliser un modèle alternatif, théoriquement moins bien adapté à l'analyse d'une relation causale, mais qui peut être une méthode excellente quand elle est bien utilisée : **l'enquête cas/témoins**.

Le principe en est également simple. Il s'agit de sélectionner un groupe de sujets atteints de la maladie étudiée : les cas (supposés représentatifs, pour l'exposition au facteur de risque, de l'ensemble des malades ayant cette pathologie), et un (ou plusieurs) groupe (s) de sujets indemnes de cette maladie : les témoins (supposés représentatifs pour l'exposition au facteur de risque de la population dont sont issus les cas). Pour chacun des sujets de l'enquête, on va rechercher des informations concernant l'exposition aux facteurs de risque (et aux facteurs de confusion éventuels) dans leur passé. On appelle souvent pour cette raison ces enquêtes des enquêtes "rétrospectives", car l'évènement étudié - la maladie - est déjà survenu quand on cherche l'exposition antérieure au facteur de risque. On va ensuite comparer l'exposition au facteur de risque chez les cas et chez les témoins.

Le schéma suivant (à comparer avec celui d'une enquête exposés/non-exposés : cf plus haut) représente la situation pour une enquête cas/témoins.



Parmi les sujets choisis au temps t_0 de l'étude pour être déjà malades (M^+) ou indemnes (M^-), on cherche dans leur passé ceux qui ont été exposés au facteur de risque (E^+) et ceux qui ne l'ont pas été (E^-).

Dans ce type d'enquête, l'épidémiologiste fixe au départ arbitrairement l'effectif des groupes de cas et de témoins (on peut sélectionner un témoin par cas ; parfois pour des raisons qui tiennent aux tests statistiques à effectuer, on a intérêt à choisir un plus grand nombre de témoins, rarement plus de trois ou quatre pour un cas). Les effectifs sont déterminés en fonction essentiellement de l'importance du rôle étiologique soupçonné du facteur de risque et de la fréquence de l'exposition à ce facteur dans la population (voir plus loin : nombre de sujets nécessaires) ; il est donc tout à fait impossible de connaître, dans ces conditions, l'incidence de la maladie chez les exposés et chez les non-exposés.

Cependant, si certaines conditions sont vérifiées :

- Groupe des cas représentatif, pour l'exposition au facteur de risque, des personnes atteintes de la maladie ;
- Groupe témoin représentatif, pour l'exposition au facteur de risque, de la population dont sont issus les cas.

Il est alors possible d'estimer l'augmentation de la probabilité d'avoir la maladie quand on est exposé au facteur de risque (c'est-à-dire le risque relatif) sans connaître l'incidence, mais en comparant la fréquence de l'exposition chez les malades (les cas) et les non-malades (les témoins).

On démontre que, dans ces conditions, on peut estimer une valeur du risque relatif appelée "**odds-ratio**" (OR), qui est une mesure très générale d'association entre deux variables à deux modalités.

Les données réunies par l'enquête permettent de dresser le tableau suivant et de calculer OR :

	E ⁺	E ⁻
M ⁺	a	b
M ⁻	c	d

$$OR = ad/bc$$

Si $OR = 1$, il n'y a pas de relation entre M et E.

On peut remarquer que, si l'on a deux fois plus de témoins M⁻ (par exemple), les effectifs c et d sont multipliés par 2 et OR ne change pas.

Le odds-ratio sera calculable dans différents contextes, notamment dans une étude cas-témoin.

Comme pour les enquêtes de cohorte, il est possible (et très utile quand c'est possible) de calculer des OR en fonction du niveau d'exposition pour chercher à mettre en évidence une relation dose-effet, et il existe des méthodes permettant de tenir compte des facteurs de confusion éventuels pour lesquels des données individuelles ont été recueillies. Ces méthodes sont comparables à ce que nous avons vu à propos des enquêtes exposés/non-exposés. On dispose également de tests statistiques adaptés pour décider si la valeur de OR calculé est significativement différente de 1.

1. Comparabilité des sujets pour les facteurs de confusion

Il existe deux façons de l'assurer :

- *A priori*, c'est-à-dire au cours de la planification de l'enquête par le choix des témoins à inclure. Une méthode, appelée "appariement individuel", consiste à choisir, pour chaque cas inclus, un (ou plusieurs) témoin (s) identique (s) pour le facteur de confusion. Ainsi, on aura finalement des groupes "cas" et "témoins" strictement comparables, par construction, pour ce qui concerne les facteurs de confusion. Par exemple, imaginons qu'on veuille prendre en compte simultanément le sexe, l'âge, et la profession. Le premier cas inclus est un homme, âgé de 43 ans, ouvrier ; on choisira un (ou plusieurs) témoin (s) masculin (s), âgé (s) de 43 ans, ouvrier (s). Et ainsi de suite, en appariant individuellement chacun des cas à son (ses) témoin (s).

En fait, au-delà de l'apparente simplicité de cette méthode, on rencontre toutes sortes de difficultés : pratiques, car il peut être extrêmement difficile de trouver des témoins correspondant aux critères d'appariement, et ceci d'autant plus qu'ils sont nombreux ; méthodologiques, car il n'est pas rare que se produise un phénomène de "sur-appariement" : il s'agit d'un phénomène statistique qui diminue la "puissance" des tests qu'on fera et qui affaiblit donc la capacité de l'étude à mettre en évidence le rôle du facteur de risque étudié, s'il existe. De plus, les méthodes de calcul du odds-ratio ne sont pas du tout les mêmes que celles que nous avons vues plus haut. C'est pour ces raisons que la méthode de l'appariement individuel n'est généralement utilisée que de façon très partielle, sur un ou deux facteurs de confusion (âge et sexe habituellement), sauf cas particulier.

Une autre méthode consiste à effectuer *a priori* une "stratification" sur le facteur de confusion. Par exemple, on prend globalement la même proportion d'hommes et d'ouvriers chez les cas et les témoins : c'est donc la distribution des valeurs du facteur de confusion qui est comparable dans les deux groupes. Cette méthode, reposant sur des groupes stratifiés, est souvent préférée, car elle présente beaucoup moins de difficultés pratiques que l'appariement individuel.

Cependant, il faut insister sur le fait que dans les études cas/témoins, le contrôle *a priori* des facteurs de confusion ne suffit pas à annuler l'effet de ceux-ci ; son but est de faciliter l'analyse statistique ultérieure en équilibrant les deux groupes, cas et témoins.

- La méthode *a posteriori* est donc toujours nécessaire. C'est après coup, une fois l'enquête terminée (d'où le terme *a posteriori*), qu'on tiendra compte des facteurs de confusion dans l'analyse statistique, par des méthodes dont le principe a été exposé plus haut (voir enquête de cohorte).

2. Représentativité des sujets vis-à-vis de l'exposition au facteur de risque

Schématiquement, deux types de situation peuvent exister, qui donnent lieu à deux variantes distinctes de l'enquête cas/témoins.

- La prévalence de l'exposition au facteur de risque est importante, et "diffuse" dans la population. Des exemples de ce type de facteurs sont ceux qui sont liés au mode de vie ou à l'environnement général : facteurs alimentaires, pollution atmosphérique, etc. Des sujets tirés au sort dans la population ont donc une probabilité importante d'être exposés. Dans ces conditions, on réalisera une "enquête cas/témoins en population générale". Les cas seront choisis dans les sources médicales habituelles (services hospitaliers, registres, etc.), et les témoins dans des sources supposées représenter également la population générale.

Il peut être difficile d'assurer que c'est bien le cas, et le choix des sujets peut occasionner des "biais de sélection", si on n'est pas suffisamment attentif.

- La prévalence de l'exposition au facteur de risque est faible et "concentrée" dans la population. Il s'agit par exemple de nuisances d'origine professionnelle, n'existant que dans certaines entreprises. Des sujets tirés au sort dans la population générale n'ont qu'une probabilité négligeable d'être exposés, rendant pratiquement impossible l'enquête. On fera alors une "enquête cas/témoins au sein d'une cohorte". Le principe en est très simple : il suffit de constituer une cohorte de sujets appartenant à une sous-population particulièrement exposée (par exemple, les travailleurs d'une entreprise où la nuisance étudiée est présente) ; les cas et les témoins seront choisis au sein de cette cohorte, et on procédera à une étude de type cas/témoins : on cherchera dans le passé des sujets leur exposition au facteur de risque. Ainsi, si on retient l'exemple d'un facteur professionnel et d'une cohorte composée de travailleurs d'une entreprise, on comparera les différents postes de travail (qui sont plus ou moins susceptibles d'entraîner une exposition au facteur de risque) pour les cas et les témoins. Quand elle est possible, cette méthode est excellente, car elle a en outre l'avantage de réduire la variabilité pour certains facteurs de confusion entre cas et témoins, les sujets appartenant à la sous-population choisie ayant souvent beaucoup de traits communs, notamment pour ce qui concerne les modes de vie.

3. Odds-ratio et risque relatif : différences selon le type d'enquête

Nous avons dit plus haut que le OR est une estimation de RR si les conditions de représentativité des cas et des témoins sont respectées. La situation n'est cependant pas la même selon qu'il s'agit d'une enquête cas/témoins au sein d'une cohorte, ou d'une enquête cas/témoins en population générale.

a. Etude cas/témoins dans une cohorte

Tous les sujets, suivis le même temps, sont rangés en fin de période d'observation dans l'une des 4 cases, selon qu'ils étaient exposés ou non (en début de période), et que la maladie est apparue ou non au cours de la période. On peut pour les deux groupes exposés et non, calculer des risques et avoir un risque relatif, ou calculer un odds-ratio :

	E+	E-
M+	a	b
M-	c	d
	a+c	b+d

$$RR = [a/(a+c)]/[b/(b+d)]$$

	E+	E-
M+	a	b
M-	c	d

$$OR = [a/c]/[b/d]$$

On voit que dans ce cas, si le RR est supérieur à 1, le OR surestime le RR.

Cependant, **si la maladie est rare** : $OR \approx RR$, car a et b sont petits respectivement devant c et d et on peut écrire :

$$[a/(a+c)]/[b/(b+d)] \approx [a/c]/[b/d]$$

b. Etude cas/témoins en population générale (population ouverte et "stable")

Les témoins sont issus de la population d'où sont issus les cas ; ils sont indemnes de la maladie au moment du recrutement des cas (mais peuvent devenir malades avant la fin de la période d'observation). On suppose de plus que la fréquence de l'exposition ne varie pas au cours du temps dans la population. Dans ce cas l'OR est un risque relatif : OR = RR.

4. Démonstration

Le recrutement des (a+b) malades pendant la période d'observation de durée T, dans la population de taille N, sera tel que :

$$a + b = \hat{\lambda} NT$$

où $\hat{\lambda}$ est le taux d'incidence dans la population. Soit P la proportion de sujet exposés dans la population, et $\hat{\lambda}_1$ et $\hat{\lambda}_0$ les taux d'incidence parmi les exposés et non-exposés (taux supposés constants pendant toute la période T).

$$RR = \hat{\lambda}_1 / \hat{\lambda}_0$$

Les malades se recrutent parmi les exposés et non-exposés :

$$a + b = \lambda_1 PNT + \lambda (1-P)NT$$

a (exposés) b (non exposés)

La proportion d'exposés chez les malades : a/(a+b) vaut :

$$P1 = \frac{PRR}{PRR + 1 - P}$$

D'où : $RR = [P1/(1-P1)]/[P/(1-P)]$
soit (a/b) / (c/d)

Intervalle de confiance de OR

On considère que Log OR est distribué normalement avec un écart-type :

$$S = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

D'où l'intervalle de confiance pour l'odds-ratio OR au niveau $1 - \alpha$:

$$OR e^{-Z\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \quad ; \quad OR e^{+Z\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

5. Avantages et inconvénients des enquêtes cas/témoins : indications

Bien que théoriquement moins bien adaptée que l'enquête exposés/non-exposés pour porter un jugement de causalité, l'enquête cas/témoins présente de nombreux avantages.

L'évènement (la maladie étudiée) étant déjà réalisé au moment de l'enquête, on n'aura pas à attendre : il est donc possible d'avoir rapidement la réponse à la question posée. De plus, le coût de telles enquêtes est habituellement beaucoup plus faible que pour les études prospectives, car la taille des groupes observés n'est en général pas élevée (souvent quelques dizaines de cas, parfois quelques centaines ; cf plus loin : nombre de sujets).

La réalisation pratique d'une enquête cas/témoins est en général plus facile que pour les enquêtes exposés/non-exposés, car souvent c'est en milieu hospitalier (ou dans des institutions de santé d'une façon plus générale) que se trouvent les cas, ainsi que, dans beaucoup d'enquêtes, les témoins : on a donc les sujets sous la main, sans avoir à les chercher et à les suivre dans leur milieu habituel (mais ce n'est pas toujours le cas).

Les principaux inconvénients des enquêtes cas/témoins tiennent au fait qu'on ne peut pas connaître l'incidence de la maladie chez les exposés et chez les non-exposés. On ne peut, évidemment, étudier le rôle des facteurs de risque que sur la maladie prise en compte, puisque les cas sont précisément choisis sur ce critère (contrairement à l'étude exposés/non-exposés qui permet d'étudier le rôle d'un facteur sur différentes pathologies). Mais les inconvénients des enquêtes cas/témoins sont surtout dus à de nombreux biais potentiels, liés au fait que la maladie (chez les cas) et la non-maladie (chez les témoins) sont connues au moment de l'enquête et que l'investigation sur l'exposition se fait de façon rétrospective : informations biaisées par le répondant et/ou l'enquêteur du fait qu'il connaît le statut de chaque sujet observé (cas ou témoin), non-représentativité des cas et/ou des témoins.

C'est pourquoi, bien que l'enquête cas/témoins soit *a priori* séduisante et plus facile à réaliser, moins coûteuse et plus rapide que l'enquête exposés/non-exposés, elle est souvent moins fiable dans ses résultats et nécessite une grande expérience pour être menée à bien dans de bonnes conditions.

Les indications de l'enquête cas/témoins sont pratiquement opposées à celles de l'enquête exposés/non-exposés. Elle est adaptée pour les maladies rares : nous avons vu qu'il s'agit dans certains cas d'une condition nécessaire pour avoir le droit d'interpréter un odds-ratio comme un risque relatif mais, de plus, il est souvent possible de trouver en milieu médical où ils sont concentrés, un nombre de cas suffisant, alors qu'une enquête exposés/non-exposés devrait, pour une maladie rare, avoir des groupes d'effectif très important, ce qui à la limite, n'est pas réalisable. De la même façon, une maladie à long délai d'installation devra plutôt faire l'objet d'une enquête cas/témoins, car on travaille alors que la maladie est déjà présente, sans avoir à attendre sa survenue. L'enquête cas/témoins se prête également bien à l'étude simultanée de plusieurs facteurs de risque de la maladie, car il suffit de recueillir les informations correspondantes chez les mêmes sujets. Enfin, l'enquête cas/témoins peut constituer une première approche d'un problème, qui sera suivie ultérieurement par une enquête prospective.

6. Nombre de sujets nécessaires

A de nombreuses reprises, il a été fait allusion au problème de l'effectif des sujets entrant dans les enquêtes épidémiologiques. Notamment, on a souligné que les enquêtes de type exposés/non-exposés requièrent des effectifs plus importants que les enquêtes cas/témoins. Il ne s'agit nullement d'un problème académique, et les conséquences pratiques sont importantes. Pour illustrer cette affirmation, et donner une idée des ordres de grandeurs dont on parle, on va donner un exemple numérique.

Supposons qu'on souhaite étudier le rôle éventuel d'un certain facteur de risque sur une maladie. On se demande s'il vaut mieux faire une enquête exposés/non-exposés ou une enquête cas/témoins ; la décision va reposer en grande partie sur le nombre des sujets nécessaires, car ce critère a une implication évidente sur la "faisabilité" de l'enquête.

Le nombre de sujets nécessaires dépend de plusieurs paramètres

- Les risques d'erreur consentis lors des tests statistiques : risque de première espèce (α) c'est-à-dire ici la probabilité de décider que le facteur de risque a un effet alors qu'en fait il n'en a pas ; risque de deuxième espèce (β), c'est-à-dire la probabilité de ne pas détecter un effet de facteur de risque alors qu'il existe en fait.
- La fréquence de l'exposition au facteur de risque dans la population (pourcentage de personnes exposées).
- Le risque relatif qu'on pense être lié au facteur de risque, c'est-à-dire l'augmentation de la probabilité d'avoir la maladie quand on est exposé au facteur. Il faut donc, à l'avance, faire des hypothèses sur la valeur du risque relatif attendu, puisqu'on ne connaît pas celui-ci (c'est justement pour le connaître qu'on fait une enquête).
- L'incidence "spontanée" de la maladie, c'est-à-dire chez les non-exposés au facteur de risque étudié (souvent on peut considérer le taux de la maladie dans la population générale comme une estimation correcte de cette incidence, quand le facteur de risque n'est pas trop fréquent).

Il existe des formules et des tables pour calculer les effectifs en fonction de ces paramètres, selon qu'il s'agit d'une enquête de cohorte ou d'une enquête cas/témoins. Le tableau suivant montre, à titre d'exemple, l'effectif minimal de sujets qu'il faut réunir dans chaque groupe (exposés et non-exposés, ou cas et témoins) en fonction de différentes valeurs de l'incidence spontanée de la maladie et du risque relatif. Pour ne pas compliquer trop le tableau, on n'a pas fait varier les risques choisis ($\alpha = 5\%$ et $\beta = 10\%$ c'est-à-dire une "puissance", ou aptitude à détecter une augmentation du risque relatif de $[1 - \beta] = 90\%$) ni la proportion des exposés au facteur de risque dans la population qu'on a fixée ici à 50 %.

Incidence spontanée (parmi les non-exposés)	Risque relatif	Effectif minimal dans chaque groupe	
		Enquête exposés/ non-exposés	Enquête cas/témoins
1 pour mille	1.2	576.732	2.535
	2	31.443	177
	4	5.815	48
1 pour cent	1.2	57.100	2.535
	2	3.100	177
	4	567	48
1 pour dix	1.2	5.137	2.535
	2	266	577
	4	42	48

Tableau extrait de H.A. Kahn, An introduction to epidemiologic methods, Oxford University Press, 1983.

Certains phénomènes remarquables méritent d'être soulignés :

- Alors que l'effectif de chaque groupe (exposés et non-exposés, ou cas et témoins : il faut donc multiplier par 2 tous les chiffres du tableau pour avoir l'effectif total des sujets entrant dans l'étude) diminue très fortement quand l'incidence de la maladie augmente pour les enquêtes exposés/non-exposés (puisque l'on passe pour un risque relatif de 1,2 de plus d'un million de sujets au total quand l'incidence est 1 pour 1.000 à environ dix mille sujets quand l'incidence est de 1 pour 10), ce paramètre n'a aucune influence sur l'effectif d'une enquête cas/témoins qui reste identique quelle que soit l'incidence : on a là l'illustration concrète d'un des avantages pratiques des enquêtes cas/témoins, qui peut être considérable.
- Le nombre de sujets nécessaires est très sensible à la valeur du risque relatif : on voit que pour des modifications relativement peu importantes du risque relatif, l'effectif change considérablement, et ceci aussi bien pour les enquêtes exposés/non-exposés que pour les enquêtes cas/témoins. Ainsi, si on fait varier le risque relatif de 1,2 à 2 (ce qui est peu), on divise pratiquement par 20 l'effectif nécessaire pour une enquête exposés/non-exposés et pratiquement par 15 pour une enquête cas/témoins. Ceci est encore plus spectaculaire quand on a un risque relatif de 4.
- Pour simplifier le tableau on n'a pas fait varier la prévalence de l'exposition du facteur de risque dans la population (pourcentage d'exposés) ; mais il faut savoir que l'effectif de l'enquête est également influencé par la prévalence de l'exposition, dans le sens d'une augmentation de l'effectif nécessaire quand la prévalence diminue.
- En voyant ces chiffres, on comprend pourquoi des enquêtes épidémiologiques dont l'objectif et le protocole général peuvent paraître très semblables (vérifier le rôle pathogène d'un facteur de risque par une enquête de cohorte de type exposés/non-exposés par exemple), peuvent en réalité être des entreprises n'ayant concrètement rien à voir en pratique. Dans les cas les plus favorables, un épidémiologiste seul, ou avec une équipe très réduite, peut mener rapidement à bien une enquête tout à fait concluante, alors que dans d'autres circonstances, il faut réunir des équipes et des moyens énormes pour un résultat incertain.

7. La notion de risque attribuable

On a vu que le risque relatif est l'indice épidémiologique qui reflète le mieux le rôle étiologique d'un facteur de risque vis-à-vis d'une maladie. Une fois ce rôle établi, il peut être intéressant de chercher à évaluer l'impact quantitatif du facteur de risque dans la population.

En effet, le risque relatif s'il permet de dire que les sujets exposés ont une probabilité x fois plus élevée d'avoir la maladie que les non-exposés ne renseigne ni sur la proportion de cas parmi les malades qui est due au facteur de risque, ni sur le nombre total de cas qu'on peut attribuer à ce facteur dans la population.

Or, dans une optique de santé publique ces informations sont très utiles. Ainsi, un facteur de risque A qui multiplierait par 100 l'incidence d'une maladie spontanément très rare dans la population peut très bien être considéré comme moins préoccupant qu'un autre facteur B qui lui, multiplierait par 3 l'incidence d'une maladie très fréquente. Le nombre de cas de maladie dus au facteur B peut dans ces conditions être plus élevé que ceux dont est responsable le facteur A.

De même, si on considère deux facteurs de risque de la même maladie, l'un ayant un rôle étiologique très fort (RR = 100 par exemple) et l'autre un rôle plus faible (RR = 3) et que le nombre de personnes exposées au premier facteur dans la population est très faible, il peut être responsable d'une proportion de cas de la maladie plus faible que le deuxième facteur si celui-ci est très répandu.

Si le risque relatif ne permet pas de rendre compte de ces situations c'est parce qu'il ne dépend ni de la fréquence de la maladie, ni de la fréquence de l'exposition au facteur de risque étudié dans la population ; il signifie simplement que parmi les exposés la probabilité d'apparition de la maladie est RR fois plus élevée que parmi les non-exposés, et cela indépendamment de la fréquence spontanée de la maladie.

Le risque attribuable permet de remédier à ces limites de l'utilisation du risque relatif. Il s'agit en effet d'une notion épidémiologique qui permet de connaître la proportion de cas d'une maladie qu'on peut attribuer au rôle d'un facteur de risque, ainsi que le nombre de sujets atteints de son fait dans une population.

Le risque attribuable est basé sur la **différence** des risques entre exposés et non-exposés au facteur. Il reflète donc l'excès de risque dont est responsable l'exposition à ce facteur.

Deux types d'indices peut être calculés pour tenir compte du risque attribuable :

- Le "**risque attribuable pour les exposés**" (ou "fraction étiologique") (FE), qui mesure la proportion de cas de la maladie qu'on peut attribuer au facteur de risque **parmi les sujets exposés** à ce facteur (pourcentage des cas de cancer du poumon dûs au tabagisme parmi les fumeurs).

On montre que la fraction étiologique vaut :

$$FE = \frac{RR - 1}{RR}$$

où RR est le risque relatif.

Ainsi, si le risque relatif vaut 5, la fraction étiologique vaut :

$$\frac{5 - 1}{5} = 0.8$$

C'est-à-dire que 80 % des cas survenus chez des sujets exposés peuvent être attribués à l'exposition au facteur de risque et auraient donc pu être évités.

- Le "**risque attribuable dans la population**" (ou "pourcentage de risque attribuable") (PRA), qui mesure la proportion de cas de la maladie qu'on peut attribuer au facteur de risque **dans l'ensemble de la population**, incluant donc les exposés et les non-exposés (pourcentage de cas de cancer du poumon dus au tabagisme dans l'ensemble de la population). En termes de santé publique, c'est ce dernier indice qui est le plus intéressant, puisqu'il mesure l'impact global du facteur de risque sur la population, en tenant compte de la proportion de personnes exposées. Lorsqu'on parle de "risque attribuable" sans précision, c'est habituellement le risque attribuable dans la population dont il s'agit.

Le pourcentage de risque attribuable vaut :

$$PRA = \frac{p (RR - 1)}{p RR + (1-p)}$$

où RR est le risque relatif et p la proportion de personnes exposées au facteur de risque dans la population.

Par exemple si 20 % d'une population est exposée à un facteur de risque (p = 0.2) et que le risque relatif correspondant à ce facteur vaut 5, le pourcentage de risque attribuable vaut :

$$\frac{0.2 (5 - 1)}{(0.2 \times 5) + (1 - 0.2)} = 0.44$$

Ceci signifie que parmi l'ensemble des cas observés dans cette population, 44 % sont attribuables à l'exposition au facteur de risque. Si on a observé 5.000 cas une certaine année, on peut dire que 2.200 parmi ceux-ci auraient pu être évités si le facteur de risque n'avait pas existé.

Si on connaît la fréquence de l'exposition au facteur de risque et l'incidence totale de la maladie dans l'ensemble de la population étudiée, on peut donc estimer la proportion de cas parmi tous les malades qu'on peut attribuer à ce facteur, ainsi que le nombre total de malades dont il est responsable (et qu'on pourrait théoriquement éviter si ce facteur était supprimé).

CONCLUSION

Pour en savoir plus sur l'épidémiologie à visée étiologique, on pourra se référer aux ouvrages cités dans la bibliographie, ou à des ouvrages d'épidémiologie spécialisés dans un domaine d'application ou un secteur de la pathologie.

BIBLIOGRAPHIE

- GOLDBERG M.- L'épidémiologie sans peine. Editions Frison-Roche, Paris, 1990.
- LECLERC A., PAPOZ L., BREART G., LELLOUCH J.- Dictionnaire d'Epidémiologie. Editions Frison-Roche, Paris, 1990.
- LILIENFELD A.M., LILIENFELD D.- Foundations of Epidemiology. 2nd ed. New-York : Oxford University Press., 1979.
- RUMEAU-ROUQUETTE C., BREART G., PADIEUX R.- Méthodes en épidémiologie, 3è édition, Flammarion Médecine Sciences, Paris, 1985.