

# Méthodes épidémiologiques destinées à mettre en évidence une relation de causalité - problèmes particuliers dans le cas de facteurs de faible importance

Marcel GOLDBERG \*

---

*(Manuscrit reçu le 8 juin 1989)*

## RESUMÉ

La recherche de facteurs étiologiques repose sur des méthodes épidémiologiques bien réglées. Les principaux modèles d'observation sont l'enquête de cohorte (contemporaine ou historique) et l'enquête cas-témoins (en population générale ou au sein d'une cohorte).

L'application de ces modèles à l'étude de facteurs pour lesquelles un effet faible est attendu soulève des difficultés qui peuvent être considérables. Elle impose une évaluation très précise de l'exposition au facteur étudié (souvent pendant de très longues périodes pouvant atteindre plusieurs décennies), la prise en compte de facteurs de confusion qui peuvent être nombreux notamment en milieu professionnel, et la mise en place d'enquêtes où l'effectif des sujets observés est habituellement très élevé.

## ABSTRACT

Well established epidemiological methods are used in the search for etiological factors. The main models of observation are the cohort study (concurrent or historical) and the case-control survey population - based or within a cohort.

Considerable difficulties can be encountered when applying these models to study factors resulting in small effects. It is, then, necessary to very accurately evaluate the exposure factor under study (often during several decades), to take into account confounding factors, which can be multiple especially in the occupational environment, and to set up surveys having a very large number of observed individuals.

---

\* INSERM, Unité 88, 91, bd de l'hôpital, 75634 Paris Cedex 13.

## 1. NOTION DE CAUSALITÉ EN ÉPIDÉMIOLOGIE

### 1.1. Généralités

**Définition :** On dit qu'un facteur est cause d'une maladie si une modification de sa fréquence (ou de sa moyenne) **entraîne** une modification de la fréquence de la maladie (définition **probabiliste**).

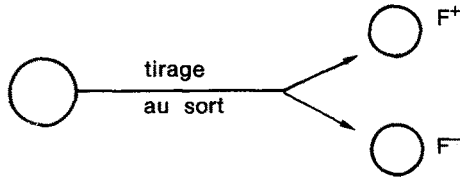
Pour **prouver** l'existence d'une " relation causale " :

$$F \rightarrow M$$

(F = facteur soupçonné ; M = maladie étudiée)

il faut utiliser la **méthode expérimentale**. L'expérimentateur introduit, dans le cours naturel d'un phénomène, un facteur, **toutes choses étant égales par ailleurs** (" *ceteris paribus* "). Si le déroulement du phénomène est modifié, on peut alors affirmer que le facteur en est la cause.

Quand il s'agit d'expérimenter sur l'homme, la seule façon d'être sûr (en probabilité) que " toutes choses sont égales par ailleurs " est le **tirage au sort** :



En épidémiologie à visée étiologique, la question de base est : l'exposition au facteur F entraîne-t-elle l'augmentation de la fréquence de M ? Pour expérimenter, il faudrait :

- soit tirer au sort des sujets  $F^+$  et  $F^-$  ( $F^+$  : exposé au facteur F ;  $F^-$  : non exposé au facteur F) ;
- soit tirer au sort, parmi les  $F^+$ , ceux qu'on va soustraire à F (études " d'intervention ").

Exposer (avec tirage au sort) des sujets à un facteur potentiellement pathogène n'est pas possible ; les études d'intervention randomisées sont moins impossibles.

Mais, dans l'ensemble, l'épidémiologiste n'est presque jamais en situation expérimentale, il est pratiquement toujours en **situation d'observation**.

### 1.2. Situation expérimentale

En pratique, la situation expérimentale ne se rencontre en épidémiologie que dans le cadre d'études d'intervention, quand on veut évaluer l'effet de l'intervention.

**Définitions :**

– **Étude expérimentale** : les **individus** soumis à l'intervention sont **tirés au sort** pour obtenir 2 (ou plusieurs) groupes, avec et sans l'intervention à évaluer ; le tirage au sort peut ne pas se faire à l'échelle des individus (sondage en grappe).

*Ex* : efficacité du dépistage du cancer colo-rectal par Hémocult en médecine du travail ; on tire au sort les médecins qui donnent et ne donnent pas le test Hémocult.

– **Étude quasi expérimentale** : l'investigateur **choisit** 2 (ou plusieurs) populations **sans randomisation** ; l'intervention est faite dans une population et pas dans l'autre.

*Ex* : efficacité de l'Hémocult dans 2 départements voisins ayant un registre du cancer ; dans un département, les médecins distribuent le test et pas dans l'autre.

### 1.3. Situation d'observation

#### CRITÈRES DE CAUSALITÉ

Pour **conclure** que l'exposition à F **entraîne** une modification de la fréquence de M, en l'absence d'un protocole expérimental, il faut **réunir un ensemble d'arguments** en faveur de la causalité.

#### a) Arguments négatifs

On est en présence de données concernant des sujets exposés ( $E^+$ ) et non-exposés ( $E^-$ ), malades ( $M^+$ ) et non-malades ( $M^-$ ) et on trouve une association statistique entre E et M. Il faut éliminer des associations statistiques non causales :

– **association fallacieuse**, due à un " biais " (erreur systématique sur le choix des sujets ou le recueil de données),

*Ex* : on constate que des sujets atteints d'une grave maladie chronique ont consommé dans le passé plus de médicaments que des sujets du même âge non malades ; ce ne sont pas ces médicaments qui entraînent l'apparition de la maladie, mais les premiers épisodes de l'installation de M ;

– **association indirecte** : une variable est liée à M, mais joue en fait par l'intermédiaire d'une autre ;

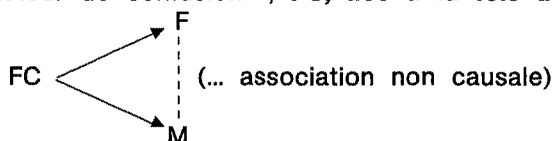
*Ex* : obésité/infarctus, statistiquement liés.

En fait, on a :

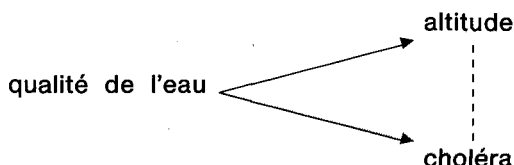
obésité → HTA → infarctus

L'obésité isolée n'entraîne pas une augmentation du risque d'infarctus; c'est l'HTA, qui est très liée à l'obésité, qui provoque cette augmentation.

– **association due à un “ tiers-facteur ”** (= “ facteur de confusion ”). Il s'agit d'un cas particulier de l'association indirecte, très fréquent ; on observe une association entre F et M, mais elle s'explique par une variable, “ facteur de confusion ”, **FC**, liée à la fois à F et M.



Ex : choléra et altitude à Londres (Snow, XIX<sup>e</sup> siècle)



On avait remarqué que la fréquence du choléra était liée à l'altitude : il était d'autant plus fréquent que l'altitude était basse. Snow a montré que la thèse de l'air salubre en altitude était fautive, mais s'expliquait par la qualité de l'eau, elle-même liée à l'altitude.

**b) Arguments positifs : critères de causalité**

Ex : tabac et cancer pulmonaire :

- \* séquence temporelle : le cancer survient après un temps de latence suivant l'exposition ;
- \* relation dose-effet : plus on fume, plus le risque de cancer est élevé ;  
“ gradient biologique ”
- \* puissance de l'effet : les fumeurs ont 15-20 fois plus de chances d'avoir un cancer pulmonaire ; plus l'effet est élevé, moins il peut être expliqué par un facteur inconnu ;
- \* constance et reproductibilité : on retrouve l'effet dans de nombreuses enquêtes, dans des populations diverses, avec des méthodes d'investigation variées ;
- \* plausibilité ; cohérence avec les données scientifiques : le tabac contient des substances irritantes pour le tractus respiratoire ; nombreuses expérimentations *in vitro* et chez l'animal ;
- \* effet de l'intervention : le risque de cancer du poumon diminue après cessation du tabagisme.

## CONCLUSION

Le jugement de causalité se fait en deux étapes :

1) **étape statistique** : il faut observer une association entre F et M. Le jugement statistique repose sur le **risque relatif (RR)** (ou d'autres mesures d'association ayant la même signification) :

$$RR = \frac{\text{incidence } E^+}{\text{incidence } E^-} = \frac{IE^+}{IE^-}$$

lorsque RR est significativement supérieur à 1, il indique que les E<sup>+</sup> ont RR fois plus de chances d'avoir M que les E<sup>-</sup> ;

2) **étape logique** : il faut éliminer les biais, contrôler les facteurs de confusion et réunir le maximum d'arguments positifs.

Certains critères sont **internes à une étude**. D'autres doivent être vérifiés à **l'échelle de plusieurs études** (reproductibilité, cohérence...).

**Au total**, conclure à la causalité peut être très difficile (et très long !).

## 2. PRINCIPAUX MODÈLES D'ENQUÊTES D'OBSERVATION À VISÉE ÉTIOLOGIQUE

### 2.1. Principes de base

Les enquêtes épidémiologiques à visée étiologique reposent sur un "protocole" visant à simuler la situation expérimentale, afin :

- d'éliminer les biais ;
- de tenir compte des facteurs de confusion ;
- de vérifier explicitement les critères positifs de causalité.

Nous allons voir comment les modèles d'enquêtes utilisés s'efforcent de reproduire une situation sans biais, permettant de vérifier les critères de causalité. Pour ce qui concerne les **facteurs de confusion**, dont le rôle est fondamental, il n'est possible d'en tenir compte que grâce à des **observations individuelles** où les données concernant les expositions, les facteurs de confusion et la maladie sont **recueillies simultanément sur chaque sujet** inclus dans l'étude. Ceci est la **seule façon de pouvoir tenir compte des facteurs de confusion dans l'analyse statistique**.

En effet, les études basées sur la recherche de liaisons statistiques entre variables dans la population, si elles peuvent être utiles pour soulever des hypothèses, ne **permettent jamais d'être formel dans le jugement de causalité**. Elles posent, en effet, de très nombreux **problèmes d'interprétation**.

*Exemple* : imaginons que dans une certaine ville, on compte habituellement en moyenne 4 cas annuels de leucémie chez l'enfant. L'année suivant le passage du nuage de Tchernobyl, on a compté 6 cas. Peut-on dire que le nuage a provoqué une augmentation de 50 % des cas de leucémie ?

Pour répondre à cette question, il faut envisager différents problèmes. Tout d'abord, le **hasard** peut expliquer de telles variations d'une année à l'autre. Dans cet exemple, le calcul montre que la probabilité d'observer au moins 6 leucémies une année quelconque alors même que la moyenne de 4 par an est inchangée, est d'environ 20 % : nuage ou pas, on aura en moyenne 6 cas ou plus tous les 5 ans.

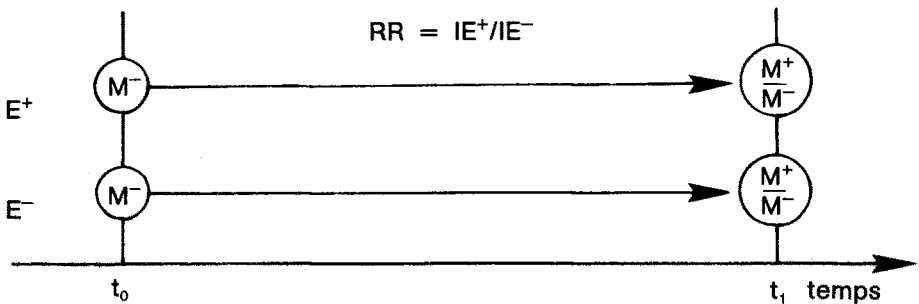
Mais même si les fluctuations dues au hasard peuvent être éliminées, reste à **attribuer au nuage** l'augmentation observée. Beaucoup d'éléments autres que le nuage peuvent intervenir : plus forte attention au problème et meilleur enregistrement des cas ; moyens diagnostiques améliorés ; modifications de population dans la période ; augmentation de la prévalence d'autres nuisances leucémogènes, etc. Généralement ces éléments ne sont pas connus ; ils constituent donc des biais. Mais même s'ils sont connus, il n'est, de toute façon, pas possible d'en tenir compte dans une analyse statistique : ils restent des biais, même s'ils sont identifiés.

Seules les enquêtes faites au niveau individuel peuvent permettre de " contrôler " ces variations, extérieures au problème à résoudre, mais qui peuvent interférer.

## 2.2. Modèles de référence : l'enquête de cohorte

### a) Principe :

On observe de façon longitudinale (" prospective ") deux groupes de sujets  $E^+$  et  $E^-$ , tous indemnes de M au départ de l'observation ; on compare l'incidence de M dans les deux groupes en fin d'observation. On peut donc estimer  $I^+$  et  $I^-$  et faire leur rapport :

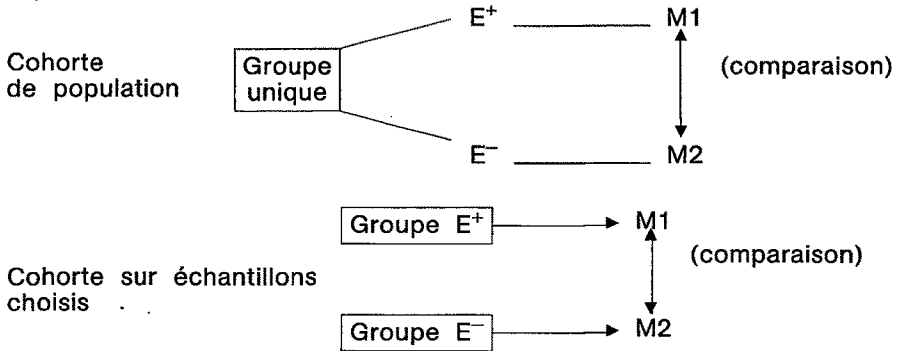


*Remarque importante : il est possible de distinguer, au sein du groupe exposé des **niveaux d'expositions différents** (fréquence, intensité, durée), ce qui permet d'étudier l'existence d'une **relation dose-effet**.*

**b) Choix des groupes  $E^+$  et  $E^-$**

**Cohorte de population** : on a un groupe unique (population ou échantillon) qui se scinde en  $E^+$  et  $E^-$  (par randomisation ou spontanément ; ex: fumeurs/non fumeurs).

**Cohorte sur échantillon choisis en fonction de l'exposition** : on choisit deux groupes différents  $E^+$  et  $E^-$  (ex : des travailleurs d'une entreprise soumis à une nuisance spécifique ( $E^+$ ) et un autre groupe non soumis à  $E^-$ ).



**Remarques :**

– il faut toujours veiller à ce que les groupes  $E^+$  et  $E^-$  soient le plus comparables possible pour les principaux facteurs de confusion (âge, sexe et souvent catégorie socio-professionnelle) ;

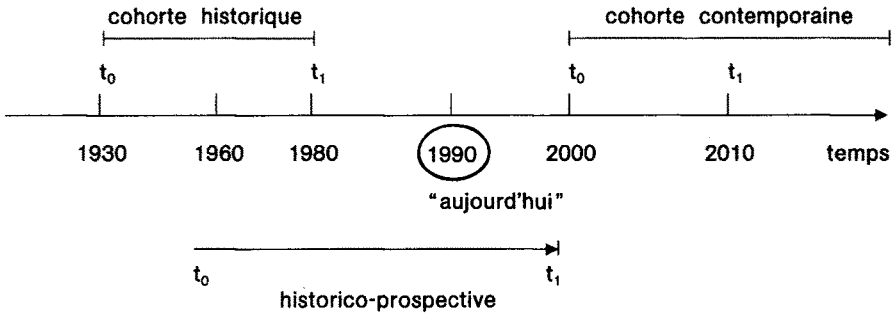
– la variante “ sur échantillons choisis ” est surtout utilisée quand l'exposition au facteur de risque (FR) est rare dans la population générale (ex : certaines expositions professionnelles) ;

– dans la variante “ sur échantillons ”, le groupe non exposé peut être l'ensemble de la population générale (pays, région) si on connaît l'incidence de M dans la population générale (mortalité ou incidence d'une maladie quand existe un registre).

**c) Choix de  $t_0$  et  $t_1$**

L'investigateur peut choisir de commencer l'observation :

- “aujourd'hui” : étude “contemporaine” (concurrente) ;
- dans le passé : cohorte “historique” ;
- dans le passé et continuer : étude “historico-prospective”.



#### Remarques :

– le choix de  $t_0$  et  $t_1$  est surtout lié au temps de latence entre exposition et occurrence de  $M$  ainsi qu'à la fréquence de  $M$ , le but étant de ne pas attendre trop longtemps les résultats !

– pour qu'une cohorte historique soit possible, il faut disposer d'une liste du groupe  $E^+$  à  $t_0$ , être capable de revoir quels sont les sujets ayant eu  $M$  entre  $t_0$  et  $t_1$  dans le groupe  $E^+$ , ainsi que dans la population  $E^-$  de référence. En pratique, cela n'est possible que lorsque  $M$  est une maladie entraînant un décès, car les données de mortalité sont les seules enregistrées systématiquement, sauf exception (registre du cancer par exemple). En France, la législation actuelle interdit les cohortes historiques (interdiction d'accès aux certificats de décès).

#### d) Avantages des études de cohorte

- pas de biais sur la mesure de  $E$  (mesure antérieure à  $M$ ) ;
- mesure précise de  $E$  et des tiers-facteurs (planifiée, mieux observable).

#### e) Inconvénients des études de cohorte

- durée (cohorte contemporaine) ;
- effectifs élevés (cf. plus loin) ;
- perdus de vue ;
- logistique et coût.

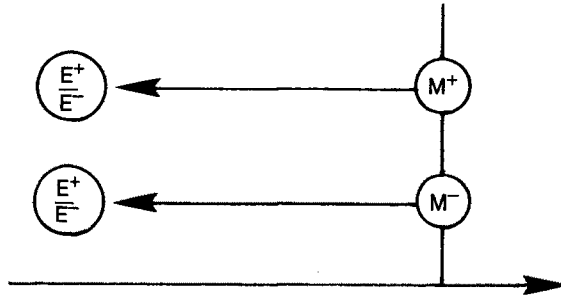
#### f) Indications

- intervalle libre  $E \rightarrow M$  bref ;
- $M$  fréquent ;
- effet fort de  $E$  (RR élevé)



### 2.3. Enquêtes cas-témoins

**a) Principe :** on choisit deux groupes de sujets, comparables pour les principaux tiers-facteurs, et qui, **au moment de leur inclusion dans l'enquête**, diffèrent par leur statut vis-à-vis de la maladie M : un groupe comporte des malades ("cas" :  $M^+$ ), l'autre groupe des indemmes ("témoins" :  $M^-$ ). On compare, entre les deux groupes, l'exposition au FR (si FR joue un rôle étiologique, il doit être plus important parmi les  $M^+$ ).



Sous certaines conditions (presque toujours vérifiées), on peut estimer le RR entre  $E^+$  et  $E^-$ .

*Remarques :*

–  $t_0$  ici réfère au moment où M est survenu chez les cas ; on cherche de façon rétrospective l'exposition au FR dans le passé des sujets  $M^+$  et  $M^-$  ;

– les résultats d'une étude cas-témoins sont exprimés par le tableau

	$M^+$	$M^-$	
$E^+$	a	b	a, b, c, d = effectifs
$E^-$	c	d	

On admet que :

$$\frac{a d}{b c} = \text{" odds-ratio " } \neq \text{RR}$$

– il est nécessaire que le groupe  $M^+$  soit représentatif des malades et le groupe  $M^-$  issu de la même population que les  $M^+$  pour ce qui concerne l'exposition au FR ;

– on peut distinguer des niveaux d'exposition différents et faire autant de tableaux  $2 \times 2$  qu'on a fait de niveaux d'exposition (relation dose-effet).

## b) Choix des groupes $M^+$ et $M^-$

### 1. Assurer la comparabilité des groupes pour les facteurs de confusion connus. Deux méthodes :

– **contrôle des facteurs de confusion a priori** = “ appariement individuel ”. Pour chaque cas inclus, on choisit un (plusieurs) témoin(s) équivalent(s) pour les facteurs de confusion ;

*Ex* : cas = homme, ouvrier, 43 ans, non fumeur  
témoin = homme, ouvrier, 43 ans, non fumeur

Les deux groupes sont alors, par construction, identiques pour les facteurs de confusion.

– **contrôle des tiers-facteurs a posteriori** = groupes “ stratifiés ” sur les facteurs de confusion ; *globalement* les deux groupes sont comparables.

*Ex* : si on a comme facteurs de confusion, âge, sexe et CPS, on sélectionnera un groupe témoin où la répartition des modalités des facteurs de confusion sera voisine de celle du groupe des cas.

*Remarque* :

*L'appariement individuel peut être difficile à réaliser (trouver le(s) témoin(s) correspondant au cas ; il peut aussi diminuer, dans certains cas, la “puissance” de l'étude (cf plus loin). Sauf exception, on a intérêt à préférer un groupe témoin stratifié.*

### 2. Assurer la représentativité des groupes $M^+$ et $M^-$ vis-à-vis de l'exposition au FR.

C'est la plus grosse difficulté des études cas-témoins.

**a) Choix des cas** : l'idéal est d'avoir une **liste de tous les cas incidents dans la population étudiée** : on peut alors les prendre tous ou un échantillon aléatoire. Ceci peut se produire dans deux cas :

1. existence d'un “ registre ” de la maladie  $M$  ;

2. à l'intérieur d'une cohorte, qui représente la population étudiée, on a tous les cas incidents pendant la période d'observation (cas des expositions rares).

Quand cela n'est pas possible (très fréquemment), on prendra les cas dans des sources habituellement médicales : services hospitaliers spécialisés, etc... Il faut alors bien veiller au **problème de recrutement des institutions concernées** (biais fréquent).

*Remarque* :

*Quand les cas sont recrutés dans des registres ou des institutions médicales, on parle d'étude “cas-témoin” en population générale (“population based”) ; quand les cas sont recrutés dans une cohorte, on parle d'étude “cas-témoins au sein d'une cohorte”.*

**b) Choix des témoins** : *a priori* dans les mêmes sources que les cas :

– étude cas-témoins au sein d'une cohorte : les témoins sont extraits de la cohorte (tous ou un échantillon aléatoire) ;

– étude cas-témoins en population générale : quand les cas sont issus de la population générale, il faut que les témoins le soient aussi ; cela peut être, de fait, **très difficile à réaliser** ;

– recrutement dans la population (annuaire, voisinage, etc.) ;

– recrutement en institution médicale (hôpital) le plus fréquent, mais **attention aux biais** : les malades hospitalisés pour une autre maladie que M pouvant :

● présenter **plus** fréquemment une exposition à des FR communs avec M ;

● présenter **moins** fréquemment ces expositions.

**c) Avantages des études cas-témoins** :

– rapidité (durée non liée au temps de développement de M) ;

– effectifs faibles ;

– coût moindre.

**d) Inconvénients des études de cas-témoins** :

– possibilité de biais sur le FR (M déjà réalisé pour les cas ; M non réalisé pour les témoins) ;

– difficultés pour assurer la représentativité.

**e) Indications** :

– maladie rare ;

– long terme de latence E → M.

### 3. PROBLÈMES PARTICULIERS À L'ÉTUDE DES FACTEURS CANCÉRIGÈNES D'ORIGINE PROFESSIONNELLE DE FAIBLE IMPORTANCE (FAIBLES EFFETS ET/OU FAIBLES DOSES)

Dans les deux premières parties de cet exposé, on a vu les méthodes et les problèmes généraux de l'épidémiologie à visée étiologique.

On va examiner ici les problèmes plus particuliers que l'on rencontre lors de l'étude épidémiologique des facteurs cancérigènes d'origine professionnelle, spécialement lorsque ces facteurs sont de faible importance, soit du fait d'un **effet intrinsèque attendu faible** (ex : champs électro-magnétiques, formaldéhyde), soit parce que les sujets ne sont exposés qu'à de **faibles doses** (rayonnements ionisants). On peut regrouper

ces problèmes spécifiques en deux grandes catégories, liées aux caractéristiques de la maladie et des expositions aux facteurs pris en compte. Selon ces caractéristiques, on verra que les conséquences sur les effectifs des sujets nécessaires sont très importants.

### 3.1. Difficultés liées aux caractéristiques de la maladie

S'agissant des cancers, deux caractéristiques essentielles peuvent rendre le travail de l'épidémiologiste difficile :

— la **rareté** de la maladie : même les cancers les plus fréquents sont, pour le statisticien, des événements "rares" ; les taux d'incidence (nombre de nouveaux cas) annuelle sont au mieux de 1 pour 1 000, et souvent nettement inférieurs, de l'ordre de 1 pour 10 000 ou pour 100 000, on voit immédiatement les difficultés de réunir un nombre de cas suffisant pour qu'une enquête soit statistiquement correcte (voir plus loin la notion de "puissance" statistique) ;

— la **durée des "temps de latence"** : pratiquement sans exception, une période d'au moins 10 ans (et plus souvent de 20 ou 30 ans) s'écoule entre la première exposition à un agent carcinogène et la survenue d'un cancer imputable à cet agent ; pour pouvoir étudier une relation de cause à effet, il faut donc organiser des enquêtes prenant en compte **une observation se déroulant sur plusieurs décennies** ; ceci ne signifie pas qu'il faudra toujours attendre plusieurs dizaines d'années pour avoir des résultats, mais que les enquêtes doivent être conçues pour inclure de longues périodes d'observation de façon plus ou moins rétrospective (enquête cas-témoins ou études de cohortes "historiques"). Ceci ajoute de nouvelles difficultés pour la mesure des expositions, comme on va le voir.

### 3.2. Difficultés liées aux caractéristiques des expositions

— **Rareté des expositions professionnelles** : certaines expositions ne sont **présentes que dans des entreprises particulières**, aux effectifs qui (pour le statisticien et en tenant compte de ce qui a été dit concernant la rareté des cancers) doivent être considérés comme faibles. Dans d'autres cas, qui correspondent à des produits ou des procédés d'usage courant (soudage, formaldéhyde, solvants, par exemple), **l'exposition est très diffuse**, ne concernant à chaque fois qu'un petit nombre de personnes dans l'entreprise. De plus, même dans les "meilleurs" cas, on n'a qu'un **très faible pourcentage de la population qui est susceptible d'être exposé**, et la fréquence des expositions n'atteint jamais celles qu'on peut observer pour des facteurs comme le tabac, l'alcool ou d'autres agents d'origine alimentaire ou environnementale. Concrètement, la première situation (exposition "concentrée" dans quelques entreprises) oblige à des enquêtes de type "cohorte professionnelle", où seuls les travailleurs des entreprises concernées sont pris en compte, ce qui pose souvent des problèmes d'effectifs très difficiles. La seconde situation (exposition "diffuse" dans de nombreux secteurs d'activité) oblige à des enquêtes de

type "cas-témoins" au sein de la population générale, aux effectifs très importants par rapport à d'autres études, en raison de la rareté des expositions professionnelles dans la population générale, et de la nécessité de disposer d'un nombre suffisant de sujets exposés, pour des raisons de "puissance" statistique.

— **Ancienneté des expositions** : en raison de l'importance des temps de latence évoquée plus haut, il est nécessaire (si on ne veut pas attendre 20 ou 30 ans pour avoir les résultats d'une investigation) de travailler de façon rétrospective, et donc de prendre en compte et d'essayer de mesurer de la façon la plus précise, quantifiée autant que faire se peut, des **expositions professionnelles très anciennes pour lesquelles on ne dispose habituellement d'aucune mesure objective**. Il faut donc recourir à diverses formes d'"expertises" (auprès d'hygiénistes industriels d'entreprises, de spécialistes de la métrologie des ambiances professionnelles, etc.) pour tenter d'estimer les expositions des sujets observés dans les enquêtes épidémiologiques. Selon qu'il s'agit d'études se déroulant au sein d'une entreprise ou en population générale, les modalités de ce travail de reconstruction *a posteriori* des expositions sont sensiblement différentes. Mais il s'agit toujours d'un travail très long, difficile, pour lequel les experts sont rares, et dont les résultats valent ce qu'ils valent !

— **Multiplicité des expositions professionnelles** : l'expérience montre qu'il est très rare, voire exceptionnel, que des travailleurs exposés à un agent cancérigène ne soient exposés qu'à celui-ci. Bien au contraire, la **règle générale est l'exposition multiple, simultanée et/ou au long de la carrière professionnelle**. De façon **simultanée**, on observe des associations très fréquentes, par exemple : soudure, amiante, solvants, huiles de coupe, etc. ; dans les industries de transformation métallurgique, on trouve à côté des multiples composés provenant de la matière transformée elle-même, des produits adjuvants, etc. Au cours de la **carrière des sujets**, il faut considérer que les différents postes de travail et/ou emplois différents occupés ont beaucoup de chance d'être similaires : certaines professions sont évidemment plus exposées que d'autres, quelle que soit l'entreprise ou le poste occupé. Concrètement, ce phénomène de cumul des expositions pose de difficiles problèmes : pratiques, car il faudra s'efforcer de reconstituer l'ensemble des carrières professionnelles des sujets et des différentes expositions susceptibles d'y être attachées ; méthodologiques, car l'analyse statistique devra prendre en compte toutes les expositions afin d'essayer d'isoler le rôle du facteur dont on souhaite étudier la cancérogénicité.

— **Existence de tiers-facteurs non professionnels** : à côté des expositions d'origine professionnelle diverses que nous venons d'évoquer, il faut également considérer des facteurs cancérigènes d'origine extra-professionnelle qui peuvent obscurcir la relation entre l'agent professionnel étudié et la maladie. Ainsi, pour ce qui concerne les cancers respiratoires (les plus fréquents, et de loin, des cancers d'origine professionnelle), il faut tenir compte du tabac, ainsi que de l'alcool pour certains sites. Quand on connaît la fréquence de l'exposition à ces facteurs, et leur pouvoir cancérigène souvent bien supérieur à celui des nuisances professionnelles étudiées, on conçoit la nature des problèmes qui se

posent à l'épidémiologiste : pratiques, en raison des difficultés du recueil des données d'exposition individuelle ; méthodologiques pour ce qui concerne l'analyse des expositions simultanées.

Ces difficultés sont communes à l'étude de tous les facteurs professionnels cancérigènes. Mais lorsque **l'effet attendu est faible, les conséquences de ces difficultés ont une importance particulière**. Pour comprendre cela, il faut faire appel à certaines considérations de nature statistique.

— **Précision de la mesure et puissance de l'étude** : d'une façon générale, une mesure imprécise de l'exposition au facteur étudié fait tendre le risque relatif observé vers la valeur 1 (c'est-à-dire signifiant absence de l'effet). L'imprécision peut porter sur les valeurs quantitatives mesurées chez chaque sujet, ou sur le classement des sujets exposés et non exposés. Si l'effet est fort (risque relatif élevé), cette sous-estimation peut n'être pas trop gênante ; mais s'il est faible, l'effet peut ne pas être détecté, le risque n'étant plus significativement différent de 1 (perte de "puissance" de l'étude, cf plus loin).

— **Valeur du risque relatif et rôle éventuel d'un facteur de confusion inconnu** : pour qu'un facteur de confusion inconnu (non pris en compte par sa mesure auprès de chaque sujet) puisse produire, de façon artéfactuelle, un risque relatif significatif, il faut que ce tiers-facteur soit lui-même plus fréquemment présent chez les exposés au facteur étudié que chez les non-exposés. On démontre que plus l'effet propre du facteur étudié est lui-même élevé, plus le tiers-facteur doit être fréquent chez les exposés sans l'être chez les non-exposés : on conçoit qu'à partir de certaines valeurs du risque relatif, il est très improbable qu'un tel facteur inconnu puisse exister sans qu'on le sache. Par contre, **si l'effet du facteur étudié est faible, la probabilité qu'un tiers-facteur non pris en compte puisse exister peut être importante**, rendant très incertain le jugement de causalité qu'on pourrait porter au vu d'un risque relatif peu élevé, mais significatif.

On voit donc que lorsqu'on étudie des facteurs de risque dont l'effet attendu est faible, en plus de la nécessité d'un effectif important des sujets, un soin tout particulier doit être apporté à la mesure de l'exposition au facteur étudié, ainsi qu'aux tiers-facteurs éventuels, d'origine professionnelle ou non.

Or, nous avons vu que les expositions multiples sont la règle en milieu du travail ; il est donc **très important de mesurer l'ensemble des agents potentiellement cancérigènes**, et pas seulement celui qui fait l'objet explicite de l'étude.

### 3.3. 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 cohorte requièrent des effectifs

plus importants que les enquêtes cas-témoins. Il ne s'agit nullement d'un problème académique, et les **conséquences pratiques sont très importantes**. Pour illustrer cette affirmation, et donner une idée des ordres de grandeur 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 et l'on se demande s'il vaut mieux faire une enquête de cohorte ou une enquête cas-témoins ; la décision va reposer en grande partie sur le nombre de sujets nécessaires, car ce critère a une implication évidente sur la "**faisabilité**" de l'enquête.

Le nombre de sujets nécessaires **est lié à la "puissance" de l'étude**. La puissance est un concept statistique : elle mesure la **probabilité de détecter l'effet pathogène soupçonné**. On la fixe à l'avance (ex : puissance de 80 % = 80 % de chances de détecter l'effet s'il existe réellement). Une fois la puissance choisie, le nombre de sujets nécessaires dépend de plusieurs paramètres :

- la fréquence de l'exposition au facteur de risque dans la population (pourcentage de personnes exposées) ;
- le risque relatif, en réalité lié au facteur, c'est-à-dire l'augmentation de la probabilité d'avoir la maladie quand on est exposé au facteur ;
- 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 de cette incidence.

Il existe des formules et des tables pour calculer les effectifs en fonction de ces paramètres. Le tableau I 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 la puissance souhaitée (fixée à 90 %), ni la proportion des exposés au facteur de risque dans la population (fixée à 50 %).

TABLEAU I

Incidence spontanée (parmi les non-exposés)	Pourcentages exposés dans la population	Risque relatif	Effectif minimal dans chaque groupe	
			Enquête de cohorte	Enquête cas- témoins
1 pour 1 000	50 %	1,2	576 732	2 535
		2	31 449	177
		4	5 815	48
1 pour 100	50 %	1,2	57 000	2 535
		2	3 100	177
		4	567	48
1 pour 10	50 %	1,2	5 137	2 535
		2	266	177
		4	42	48

(Extrait de H.A. KAHN, An introduction to epidemiologic methods. Oxford : University Press, 1983).

Deux phénomènes remarquables méritent d'être soulignés.

Alors que l'effectif de chaque groupe (il faut 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 de cohorte** (puisqu'on passe, pour un risque relatif de 1,2 de plus d'un million de sujets au total quand l'incidence est de 1 pour 1 000, à environ 10 000 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.

Le nombre des 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 de cohorte 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 de cohorte, et pratiquement par 15 pour une enquête cas-témoins. Ceci est encore plus spectaculaire quand on a un risque-relatif de 4. On voit donc que plus l'effet pathogène (mesuré par le risque relatif) est faible, plus les effectifs nécessaires croissent vite.

On comprend, en voyant ces chiffres pourquoi des **enquêtes épidémiologiques dont l'objectif et le protocole général peuvent paraître très semblables, peuvent en réalité être des entreprises n'ayant concrètement pratiquement rien à voir**. 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 !

On conçoit donc pourquoi les travaux épidémiologiques sont souvent si longs, et leurs résultats rarement très assurés. D'autant plus qu'à ces difficultés "communes", s'en surajoutent fréquemment d'autres plus conjoncturelles, qui tiennent parfois au **peu d'empressement des partenaires sociaux** concernés de collaborer à de telles études ou de les faciliter.