

**LES DIFFERENTS TYPES D'ENQUETES
EN EPIDEMIOLOGIE ANALYTIQUE
MODELES STATISTIQUES
ET INDICATEURS DE RISQUE ASSOCIES***

M. SANAA (1)

RESUME : Cet article présente les principaux types d'enquêtes utilisés en épidémiologie analytique et les mesures de risque associées. Les trois principaux types d'enquêtes sont les enquêtes cohorte, cas/témoins et de prévalence. Les enquêtes prospectives ou du type cohorte sont les mieux adaptées pour tester les hypothèses étiologiques. En effet, ces dernières ressemblent le plus aux dispositifs expérimentaux. Les enquêtes cas/témoins présentent d'autres avantages et sont parfois faciles à mettre en oeuvre sous certaines conditions.

Les mesures d'association reflètent le degré de la relation statistique entre le facteur étudié et la maladie. Ces mesures peuvent utiliser l'incidence (taux ou risque), la prévalence ou la mortalité. Le choix de la mesure et son interprétation dépendent du type d'enquête.

SUMMARY : This paper presents the principal designs of analytic epidemiologic surveys and the measures of association between the study factor and disease.

The three basic designs of epidemiologic surveys are cohort, cross-sectional and case controls.

The prospective or cohort survey is preferred, in theory, for testing an etiologic hypothesis, because this type of observational study most closely resembles an experiment.

Other studies type like case-controls present certain practical advantages that make them attractive under special conditions.

Measures of association reflect the strength of statistical relationship between the study factor and the disease. The measures association may involve incidence (rate or risk), prevalence, or mortality. The choice of an association measures and its interpretation depend on the survey design.

*
* *

(1) I.T.E.B., 149 rue de Bercy, 75595 Paris cedex 12.

* Texte de l'exposé présenté lors de la réunion du 27 mars 1991.

INTRODUCTION

L'épidémiologie se définit comme l'étude des maladies et des facteurs de santé dans une population [Toma et al., 1991]. Les principaux objectifs de la recherche épidémiologique sont :

- De décrire le statut sanitaire des populations en énumérant l'occurrence des maladies, obtenant ainsi des fréquences relatives dans des groupes et déterminant des tendances temporelles ou spatiales de l'évolution des maladies.
- D'expliquer l'étiologie des maladies en déterminant les facteurs causant ces maladies spécifiques ou des tendances et en découvrant les modes de transmission.
- De prédire le nombre de cas de maladies à apparaître et la distribution des fréquences des maladies dans les populations.
- De contrôler les distributions des maladies en mettant en oeuvre des mesures préventives, d'éradication des cas existants, voire prolonger la durée de vie des malades.

Au regard de ces différents objectifs de la recherche épidémiologique, on peut dégager deux buts ou niveaux différents : comprendre et agir. Ce qui correspond aussi à la dichotomie présentée par Schwartz et al. [1970] qui réfèrent à l'attitude explicative et à l'attitude pragmatique.

Pour comprendre, on procède à partir d'observations vers l'inférence qui nécessite une accumulation de connaissances relatives à l'apparition des maladies et leurs étiologies.

Pour agir, on collecte les informations empiriques qui aident à prendre des décisions concernant la santé des populations.

La recherche épidémiologique est donc une recherche du type empirique car elle nécessite systématiquement une collecte d'observations du phénomène étudié dans une population donnée.

L'objet de cet exposé est de présenter les principaux types d'enquête analytique, les différents indices de risque et les modèles statistiques associés. Ces techniques sont de précieux outils dans la recherche des facteurs de risque et des mesures prophylactiques contre les maladies.

LES DIFFERENTS TYPES D'ENQUETES

Les méthodes de recherche en épidémiologie analytique se basent sur des comparaisons inter-groupes. Ces comparaisons sont accomplies en estimant et en testant le degré d'association entre un facteur supposé causal et son effet (la maladie).

Pour étudier l'association entre un facteur supposé causal (F) et la maladie (M), différents types d'enquêtes sont possibles, se ramenant à 3 modes d'échantillonnage : les distributions marginales de M et F sont toutes deux aléatoires, il s'agit des enquêtes de prévalence (cross-sectional survey) ; la distribution marginale de F est fixée par le chercheur, celle de M reste aléatoire, il s'agit des enquêtes exposés/non exposés ou de cohorte (cohort survey) ; la distribution marginale de M est fixée par le chercheur, celle de F reste aléatoire, il s'agit des enquêtes cas/témoins (case-control survey).

ENQUETE DE PREVALENCE

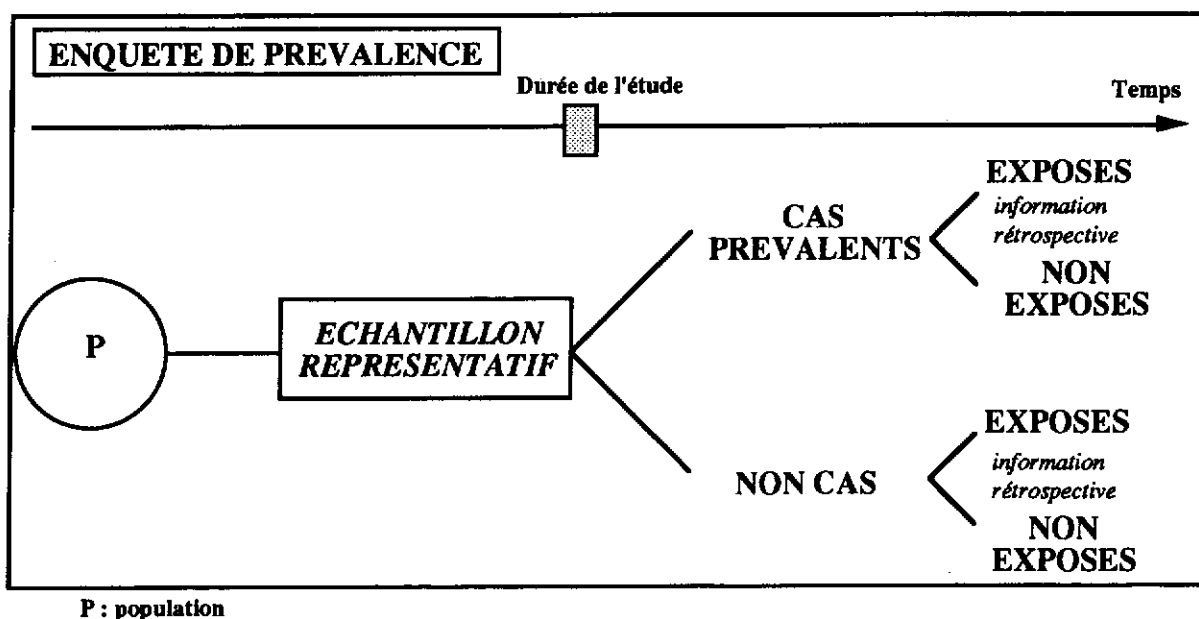
C'est une enquête transversale. L'échantillon d'étude est tiré au sort à partir d'une population cible définie par l'investigateur. Comme le montre la figure 1, dans ce type d'enquête, les cas (malades) appartiennent à la population des cas prévalents.

Après la sélection des animaux ou groupes d'animaux (bandes, élevages, ...) tous les individus sont examinés ou observés, leur statut d'exposé ou non-exposé est déterminé et d'autres variables sont parfois relevées.

Dans certains cas, l'échantillon n'est pas aléatoire. Il faut noter que sans échantillonnage aléatoire, une étude de prévalence a une valeur très limitée si elle a comme objectif de déduire la fréquence de la maladie ou autres caractéristiques d'une population et présente ainsi de très sérieuses limitations pour faire de l'inférence causale.

Quand la maladie est rare, ce type d'étude reste peu puissant pour détecter d'éventuelles relations entre un facteur et la maladie.

Figure 1 : Enquête de prévalence.



ENQUETE EXPOSES/NON-EXPOSES (figure 2)

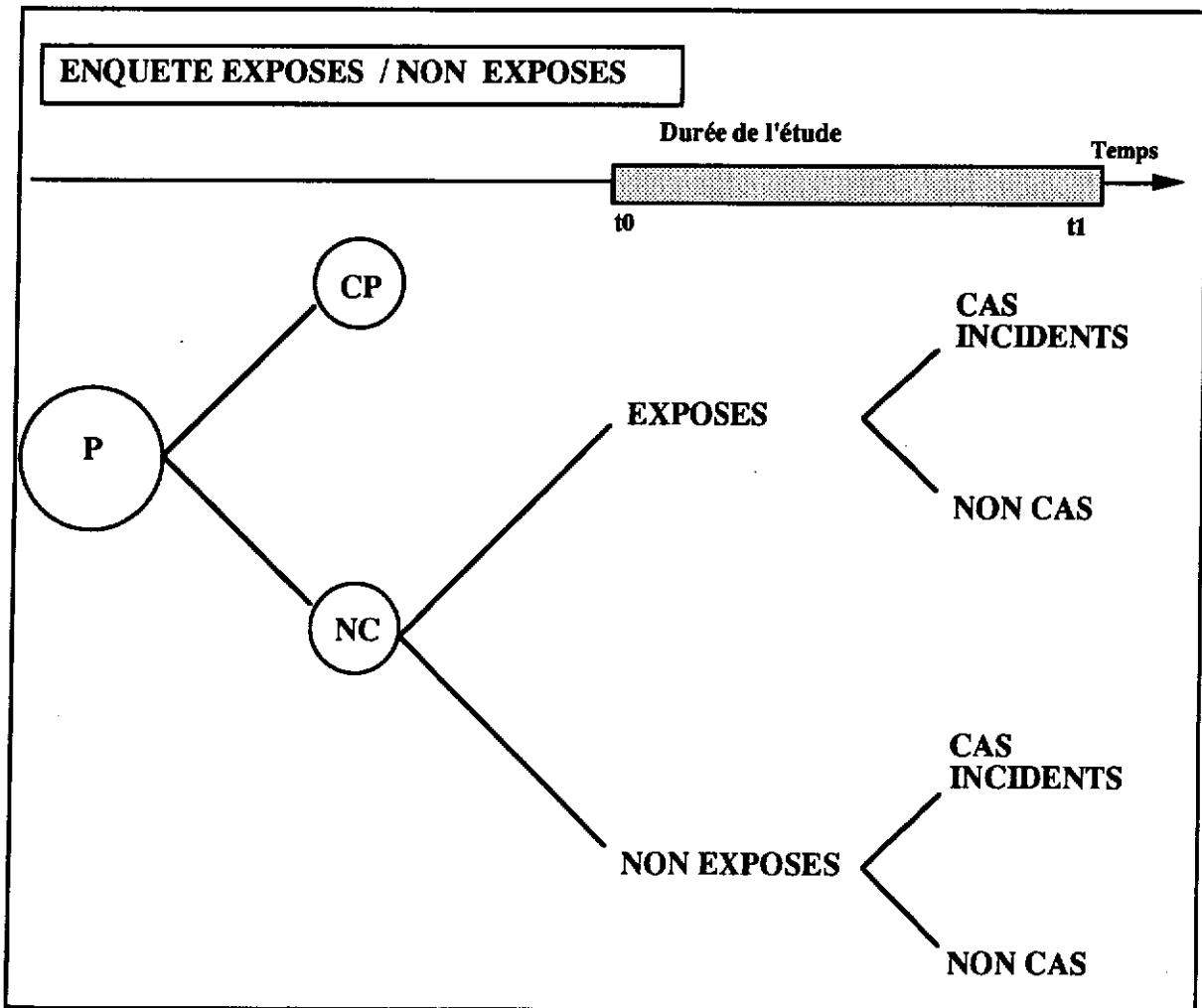
Appelée aussi enquête prospective ou enquête de cohorte, c'est une enquête dans laquelle l'information concernant le facteur d'exposition étudié (exposés ou non-exposés) est connue sur tous les individus au début de la période de suivi avant même que la maladie apparaisse. La population à risque (exposée ou non-exposée) est suivie pendant une période de temps comprise entre t_0 et t .

L'enquête exposés/non-exposés permet ainsi d'énumérer les cas incidents apparus pendant la période de suivi (t_0 et t) chez les exposés et chez les non-exposés.

Du fait que ce type d'enquête d'observation ressemble le plus à un dispositif expérimental, l'enquête exposés/non-exposés est généralement préférée pour faire de l'inférence causale.

Cependant, dans le cas d'une maladie rare ou d'une période de latence (délai entre l'exposition et l'apparition de la maladie) trop longue, l'étude prospective nécessite un nombre de sujets trop élevé et des périodes d'observation très longues. Ces deux paramètres sont difficiles à gérer sur le terrain (abandon, vente d'animaux, cessation d'activité, etc.).

Figure 2 : Enquête exposés/non-exposés.



P : population CP : cas prévalents
NC : non cas

ENQUETE CAS/TEMOINS

Elle est connue aussi sous le nom d'enquête rétrospective. Elle consiste à comparer un groupe de sujets malades (cas) et un groupe de sujets non malades (témoins ou non-cas), grâce à la récupération d'informations sur l'exposition antérieure à un ou plusieurs facteurs (figure 3).

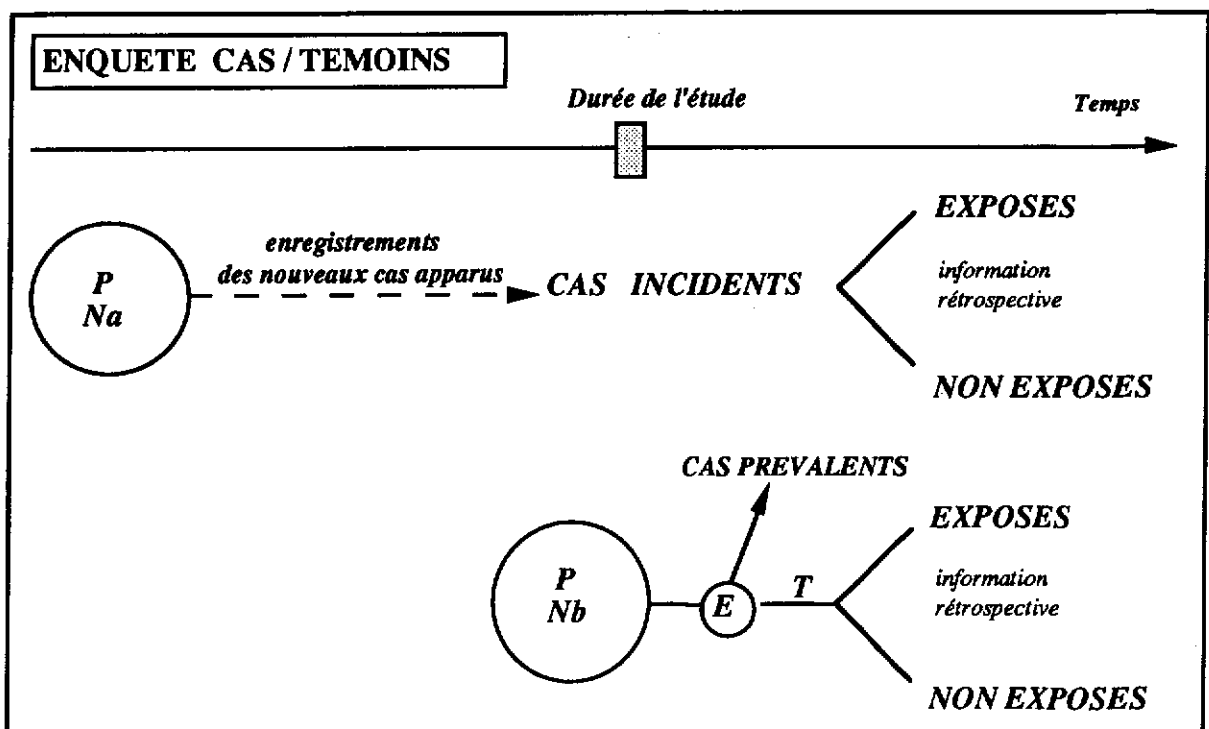
L'existence d'un système de suivi de l'état sanitaire de la population (P_{Na}) permet d'enregistrer les cas incidents. Les cas sont recrutés au fur et à mesure de leur apparition. Les témoins sont tirés à partir de la population générale (P_{Nb}) après la détermination des cas.

Les deux populations (P_{Na} et P_{Nb}) ne sont pas systématiquement équivalentes. En effet, le système de déclaration des cas peut exclure une catégorie de la population et le délai entre la détermination des cas et des témoins peut rendre les deux populations non comparables.

Pour pouvoir faire de l'inférence causale, il faut vérifier ou supposer que les non-cas (témoins) tirés de la population (P_{Nb}) sont représentatifs de la même population (P_{Na}) à partir de laquelle les cas se sont développés i.e. les deux populations sont équivalentes. Afin d'assurer la comparabilité, des techniques d'appariement peuvent être utilisées.

Dans certaines études, il n'existe pas de système de déclaration de la maladie permettant de répertorier les cas incidents. Le recrutement des cas et des non-cas se fera alors à partir de la population générale et les cas deviendront des cas prévalents. Nous verrons plus tard les répercussions d'un tel choix sur l'estimation de l'association entre le facteur F et la maladie M .

Figure 3 : Enquête cas/témoins.



P_{Na} : population à partir de laquelle les nouveaux cas enregistrés sont apparus

P_{Nb} : population à partir de laquelle on tire au sort les témoins

E : échantillonnage, les cas prévalents sont exclus

T : témoins (non cas)

MESURES DE LA RELATION ENTRE LE FACTEUR F ET LA MALADIE M

FREQUENCE DE LA MALADIE

En épidémiologie, le premier objectif est de décrire la fréquence des maladies dans une population et d'exprimer la ou les différences qui existeraient entre deux ou plusieurs populations.

La mesure de la fréquence des maladies peut se faire par le calcul d'incidence exprimée sous forme de taux ou sous forme de risque. Un taux mesure l'occurrence de nouveaux cas par unité de temps au temps t par rapport au nombre de susceptibles au temps t . Le risque est la probabilité conditionnelle d'être malade entre t_0 et t sachant que l'animal est non malade à t_0 . Le taux et les risques ne sont estimables que dans les enquêtes du type exposés/non-exposés.

Dans une enquête exposés/non-exposés, on peut estimer les taux par la Densité d'Incidence : ID (incidence density) :

$$DI(t_0, t) = I/AT$$

où I est le nombre de cas apparus entre t_0 et t et AT le nombre animaux-temps.

Si on connaît les dates d'entrée et de sortie de chaque individu dans l'étude, AT n'est autre que la somme des durées de participation de chaque individu à l'étude. Un dénominateur exact n'est pas toujours nécessaire et n'est pas souvent connu. On peut faire l'approximation suivante :

$$AT = \frac{\text{Nombre au début} + \text{nombre à la fin}}{2} \times (t - t_0) T$$

Le risque peut être estimé dans les enquêtes du type exposés/non-exposés par l'incidence cumulative :

$$R(t_0, t) = I/n$$

où I est le nombre de cas apparus entre t_0 et t et n est le nombre d'animaux entrant dans l'étude. Pour avoir plus de détails sur l'estimation des fréquences des maladies, le lecteur peut consulter l'article de Goldberg et Leclerc [1990].

La comparaison directe des incidences de la maladie des deux populations exposée et non-exposée au facteur F n'est possible que dans le cas de l'enquête du type exposés/non-exposés.

MESURE DE LA RELATION ENTRE F ET M

La mesure de la relation entre F et M dépendra du choix d'un modèle qui lui-même peut être lié au type de l'étude entreprise.

La mesure de l'association exprime la force ou l'intensité de la relation statistique entre le facteur étudié et la maladie. Rappelons que les méthodes statistiques en elles-mêmes ne peuvent établir la preuve de la relation causale dans une association.

Cas d'une enquête exposés/non-exposés

Deux cas de figure sont possibles :

Cas où on peut estimer des taux d'incidence

A la fin de l'étude, nous obtiendrons l'un des tableaux suivants (I et II).

Tableau I : Etude prospective permettant l'estimation des taux d'incidence (facteur à plusieurs classes).

	k + 1 catégories d'exposition				Total
	E_k	...	E_1	E_0	
Nombre de nouveaux cas	a_k	...	a_1	a_0	m_1
Animaux-Temps	AT_k	...	AT_1	AT_0	AT

Tableau II : Etude prospective permettant l'estimation des taux d'incidence (facteurs à deux classes).

	Exposition		Total
	E_1	E_0	
Nombre de nouveaux cas	a_1	a_0	m_1
Animaux-Temps	AT_1	AT_0	AT

Pour chaque catégorie d'exposition, nous pouvons calculer un taux d'incidence (ou plus exactement taux de densité d'incidence : Incidence Density : ID) :

$$ID_0 = a_0/AT_0, ID_1 = a_1/AT_1 \dots ID_k = a_k/AT_k$$

Le taux d'incidence de chaque catégorie peut être rapporté au taux d'incidence du groupe de référence (ID_0) :

$$IDR_1 : \text{incidence density ratio} : ID_1/ID_0$$

$$IDR_k : \text{incidence density ratio} : ID_k/ID_0$$

La mesure d'association dans ce cas est une comparaison directe des mesures de fréquence pour différentes valeurs ou catégories du facteur étudié. Dans le cas où le facteur est quantitatif (poids, nombre de jours d'exposition à un produit toxique...), il est possible de faire une mise en classes des valeurs de la variable, constituant deux ou plusieurs groupes ou catégories. Chaque catégorie peut être alors comparée à une catégorie désignée comme **groupe de référence**. Le choix de ce dernier est d'importance capitale dans les études analytiques.

Cas où on peut estimer des taux de risque

A la fin de l'étude, nous obtiendrons l'un des tableaux suivants (III et IV).

Tableau III : Etude prospective permettant l'estimation des taux de risque (facteur à plusieurs classes).

	k + 1 catégories d'exposition				Total
	E_k	...	E_1	E_0	
Nombre de nouveaux cas	a_k	...	a_1	a_0	m_1
Nombre de non-cas	c_k	...	c_1	c_0	m_0
Total	n_k	...	n_1	n_0	n

Tableau IV : Etude prospective permettant l'estimation des taux de risque (facteur à deux classes).

	Exposition		Total
	E_1	E_0	
Nombre de nouveaux cas	a_1	a_0	m_0
Nombre total d'animaux suivis	n_1	n_0	n

Pour chaque catégorie d'exposition, nous pouvons calculer un taux de risque :

$$R_0 = a_0/n_0, R_1 = a_1/n_1 \dots R_k = a_k/n_k$$

Le risque dans chaque catégorie peut être rapporté au risque dans le groupe de référence (R_0) :

$$RR_1 : \text{risque ratio : risque relatif : } R_1/R_0$$

$$RR_k : \text{risque ratio : risque relatif : } R_k/R_0$$

Cas d'une enquête cas/témoins

Dans une étude cas/témoins, l'incidence n'est pas estimable. Seules les probabilités suivantes peuvent être estimées :

$$P(E_1/\text{cas}) \text{ et } P(E_1/\text{témoin})$$

qui sont les probabilités conditionnelles d'être exposé au facteur selon que le sujet est un cas (incident ou prévalent, en fonction du protocole) ou un témoin (représentatif ou non de la population des non-cas, en fonction du protocole).

Nous obtiendrons les tableaux suivants :

Tableau V : Données d'une enquête cas/témoins
(facteur à deux classes).

	Exposition		Total
	E ₁	E ₀	
Nombre de nouveaux cas	a	b	m ₁
Témoins	c	d	m ₀
Total	n ₁	n ₀	n

Pour chaque catégorie, on peut calculer un odds (chance : pari)

- odds d'être exposé parmi les nouveaux cas = a/b
- odds d'être exposé parmi les témoins = c/d

le quotient de ces deux odds s'appelle Odds Ratio : OR

$$OR = (a/b)/(c/d) = (ad)/(bc)$$

Quand le facteur d'exposition (F) est à plusieurs classes (k + 1), on calculera k odds ratio différents. Si les cas sont recrutés parmi les cas prévalents, on parlera plutôt de **Prévalence Odds Ratio : POR**.

Cas d'une enquête de prévalence

Dans ce type d'étude, deux indices sont estimés :

- Le ratio de prévalence = PR = P₁/P₀
où P_i est la prévalence de la maladie dans le groupe d'exposition i et P₀ la prévalence de la maladie dans le groupe de référence.
- La prévalence odds ratio.

MESURE DE LA RELATION ENTRE F ET M PAR DIFFERENCE

Trois indices sont utilisés :

- La Différence de Densité d'Indice (IDD),
- La Différence de Risque (RD), appelé aussi risque en excès,
- La Différence de Prévalence (PD).

Les deux premiers indices ne sont estimables que dans les enquêtes du type exposés/non-exposés ou prospective et le dernier seulement dans les enquêtes du type prévalence.

RELATION ENTRE LE RISQUE RELATIF (RR) ET L'ODDS RATIO (OR)

Le risque relatif est un indice d'interprétation facile : c'est un coefficient multiplicateur de la probabilité d'être malade en passant du groupe des non-exposés au groupe des exposés. Mais cet indice n'est estimable que dans les enquêtes exposés/non-exposés.

L'odds ratio qui représente le rapport des "chances" relatives d'être malade ou non malade n'est pas directement interprétable, mais peut être estimé aussi bien dans les enquêtes exposés/non-exposés que dans les enquêtes cas/témoins.

Dans le cas où la maladie est rare, on montre par le calcul que la valeur de l'odds ratio se rapproche beaucoup de celle du risque relatif (tableau VI). Cette propriété de l'odds ratio est indépendante du type d'enquête.

Tableau VI : Valeur de l'odds ratio en fonction de la prévalence de la maladie et du risque relatif.

PREVALENCE DE LA MALADIE	RISQUE RELATIF	
	RR = 2	RR = 10
10 ⁻³	OR = 2	OR = 10,05
10 ⁻²	OR = 2,02	OR = 10,5
10 ⁻¹	OR = 2,22	OR = 20,00

Dans une étude cas/témoins où les malades et les non malades sont triés à partir d'une grande population en "état stationnaire", l'odds ratio est égal au risque relatif si les cas sont des cas incidents et les témoins sont représentatifs des non malades. Si dans les mêmes conditions, les cas sont des cas prévalents, l'odds ratio calculé (prevalence odds ratio) dépendra du risque relatif et des durées de la maladie chez les exposés et chez les non-exposés :

$$POR = RR (D_1/D_0)$$

où D_1 est la durée moyenne de la maladie chez les exposés et D_0 la durée de la maladie chez les non-exposés.

Si on suppose que les malades exposés et les malades non-exposés ont en moyenne la même durée de maladie, la prévalence odds ratio est égale au risque relatif.

L'odds ratio peut être estimé aussi bien dans une enquête cas/témoins que dans une enquête cohorte. Son estimation ne dépend pas du type d'enquête.

MODELES STATISTIQUES ASSOCIES AUX DIFFERENTES MESURES

Les modèles statistiques en épidémiologie analytique sont utilisés pour tester des hypothèses étiologiques ou pour prédire l'occurrence de la maladie. Pour simplifier leur présentation, prenons le cas le plus simple où le facteur d'exposition F est dichotomique (absence, présence).

Notons pour X la variable exposition :

- X = 0 si le sujet n'est pas exposé à F
- X = 1 si le sujet est exposé à F

MODELE LINEAIRE SIMPLE

On peut modéliser la prévalence ou le risque de la maladie (Y).

$$Y = b_0 + b_1X$$

b_0 correspond à la prévalence ou risque dans le groupe de référence et b_1 la mesure d'association par différence.

Y = prévalence alors $b_1 = PD =$ différence de prévalence

Y = risque alors $b_1 = RD =$ différence de risque

Ce modèle ne peut être appliqué que dans le cas où l'enquête est du type prévalence ou exposés/non-exposés.

Si le facteur F est quantitatif, sur le plan mathématique, ce modèle ne permet pas à X de prendre n'importe quelle valeur.

MODELE LOG-LINEAIRE

$$\log(Y) = b_0 + b_1X$$

b_0 correspond au log de la prévalence ou au log du risque dans le groupe de référence et b_1 correspondra alors :

au logarithme du Risque Relatif (log RR) si Y = risque

ou

au logarithme du ratio de prévalence si Y = prévalence

comme le modèle linéaire simple, le modèle log-linéaire ne peut être appliqué que dans les enquêtes exposés/non-exposés ou de prévalence.

MODELE LOGISTIQUE

$$\text{logit}(Y) = \log(Y/1 - Y) = b_0 + b_1X$$

Ce modèle peut être utilisé dans tous les types d'enquêtes. Dans le cas d'une enquête exposés/non-exposés le modèle s'écrit :

$$\text{logit}(R) = b_0 + b_1X$$

ou

$$\text{logit}(P(\text{malade}/X)) = b_0 + b_1X$$

b_0 correspond alors au logarithme de l'odds du risque dans le groupe de référence et b_1 au logarithme de l'odds ratio.

Dans le cas d'une enquête cas/témoins, on ne peut pas modéliser le risque, mais on modélise la probabilité d'être exposé sachant qu'on est malade :

$$\text{logit}(P(X/\text{malade})) = b_0^* + b_1X$$

b_1 correspond au logarithme de l'odds ratio, par contre, b_0^* devient un paramètre de nuisance.

Dans le cas d'une enquête de prévalence, b_0 sera interprété comme le logarithme de l'odds de la prévalence dans le groupe de référence et b_1 comme le logarithme de POR : Prévalence Odds Ratio. Sur le plan mathématique, le modèle logistique est le seul modèle qui permet de bien prendre en compte à la fois des facteurs quantitatifs et des facteurs qualitatifs.

CONCLUSION

Les méthodes décrites dans cet article ont été largement utilisées dans les enquêtes sur les facteurs de risque des maladies humaines. Nul doute qu'elles ne rendent de grands services dans le domaine de la pathologie animale. Des adaptations sont nécessaires, notamment au niveau du choix de l'unité d'observation, l'intégration des facteurs d'élevage et des facteurs individuels dans un même modèle, le choix et le recrutement des cas et des témoins dans les enquêtes cas/témoins...

Le modèle logistique apparaît comme le modèle général utile pour tester des hypothèses étiologiques ou prédire l'occurrence de l'évènement pathologique étudié.

L'intérêt est essentiellement pragmatique : détecter les groupes à haut risque sur lesquels on ciblerait des actions de prévention ou des thérapeutiques nouvelles... Toutefois, la recherche et le test du rôle spécifique de chaque facteur dans l'apparition de la maladie, en utilisant des modèles logistiques multifactoriels, peuvent fournir des éléments de réponse nécessaire à l'inférence causale.

BIBLIOGRAPHIE

- GOLDBERG M. et LECLERC A.- Epidémiologie à visée étiologique. *Epidémiol. et santé anim.*, 1990, 18, 1-30.
- KLEINBAUM D.G., KUPPER L.L. and MORGENSTERN H.- *Epidemiologic Research. Principals and quantitative methods.* Van Nostrand Reinhold Company, New York, 1982, 530 p.
- MARTIN S.W., MEEK A.H. and WILLEBERG P.- *Veterinary epidemiology. Principes and methods*, 1987, Iowa State University Press, Ames, Iowa, 344 p.
- RUMEAU-ROUQUETTE C., BREART G. et PADIEU R.- *Méthodes en épidémiologie.* Flammarion, Paris, 1985, 398 p.
- SCHWARTZ D., FLAMANT R. et LELLOUCH J.- *L'essai thérapeutique chez l'Homme.* Flammarion Médecine Sciences, 1970, 298 p.
- TOMA B., BENET J.J., DUFOUR B., ELOIT M. MOUTOU F. et SANAA M.- *Glossaire d'épidémiologie animale.* Point vétérinaire, 1991, 368 p.