

ÉVALUATION DE LA QUALITÉ D'UNE ÉTUDE ÉPIDÉMIOLOGIQUE : MÉTHODOLOGIE GÉNÉRALE POUR DÉBUTANTS

Bénet Jean-Jacques¹, Marsot Maud², Garin Emmanuel³ et Rivière Julie¹



RÉSUMÉ

Cet article propose une méthode d'analyse de la validité des résultats des études épidémiologiques, reposant sur plusieurs étapes : définition de la qualité d'une étude épidémiologique, évaluation de l'adéquation des buts de l'étude aux besoins, identification du type d'étude (descriptive ou explicative), évaluation de la validité interne du protocole (échantillonnage, mesures, analyse et interprétation des données), évaluation de la validité externe, bilan global.

Il s'appuie également sur des études de cas présentées sous forme d'enseignement programmé (problèmes à résoudre à l'aide de questions à choix multiples et organisées selon un itinéraire permettant l'apprentissage de la méthode d'évaluation), qui pourront ultérieurement être téléchargées à partir du site de l'AEEMA (<http://aeema.vet-alfort.fr/>).

Mots-clés : méthode, évaluation, étude épidémiologique, qualité, validité interne, validité externe.

ABSTRACT

In this paper, a method of analysis of the validity of the results of epidemiological studies is suggested. The method relies on several steps: definition of quality of epidemiological studies, evaluation of pertinence of the goals of the study to the needs, determination of the type of study (descriptive or explanatory), evaluation of internal validity (study design: sampling, measures, data analysis and interpretation), evaluation of external validity and global assessment.

This paper is completed by several problem based learning cases studies (through succession of questions/answers organized in order to use the different steps of the evaluation method), which will be downloaded at a later time from the AEEMA website (<http://aeema.vet-alfort.fr/>).

Keywords: Method, Evaluation, Epidemiological studies, Performance, Internal validity, External validity.



¹ Université Paris Est, École nationale vétérinaire d'Alfort, USC ENVA-Anses EpiMAI, 94704 Maisons-Alfort, France

² Université Paris Est, ANSES, Laboratoire de santé animale, Unité Epidémiologie, Maisons-Alfort, France

³ Coop de France, service sanitaire, 75011 Paris

I - INTRODUCTION

Chacun est en permanence sollicité par des informations nouvelles, y compris dans le domaine scientifique. La sélection de la source d'information est le premier moyen de faire le tri pour distinguer « l'info » de « l'into ». Mais, cette précaution n'est pas suffisante : il faut aussi être capable de juger par soi-même de la validité d'une information. En effet, certains auteurs, même scientifiques, peuvent diffuser ou analyser une information erronée sans le savoir et sans volonté de diffuser de « l'into ».

La qualité des résultats d'une étude épidémiologique dépend directement de la qualité de conception et de réalisation de l'étude elle-même. C'est pourquoi, pour s'assurer de leur qualité, il faut procéder à l'évaluation de la validité du protocole de l'étude.

Mais, si l'on vient de débiter en épidémiologie, et même si l'on maîtrise les connaissances de base, on a bien conscience que ce n'est pas suffisant pour être capable d'évaluer la qualité d'une étude épidémiologique. Il est en effet nécessaire de renouveler en permanence ses connaissances pour en être capable, et les articles utilisés pour ce besoin sont accessibles surtout dans des revues à comité de lecture. Ce besoin d'évaluation de la qualité peut aussi concerner un projet d'étude, car nombreuses sont les personnes qui ne sont pas épidémiologistes de métier mais qui peuvent avoir l'opportunité de réaliser une étude épidémiologique.

Cet article est centré sur l'épidémiologie animale, dont les méthodes, adaptées aux populations animales, peuvent être spécifiques, sous certains aspects. Il est destiné à un public déjà au fait des connaissances épidémiologiques de base, mais non encore aguerri à leur mise en pratique, l'épidémiologie n'étant pas leur métier. Il vise à présenter une approche pour apprendre à maîtriser progressivement les points clés essentiels à appliquer dans l'analyse critique d'une étude, dont la connaissance résulte d'un savoir-faire qui s'acquiert normalement grâce à une longue expérience.

Il aborde les trois questions auxquelles cette évaluation va permettre de répondre :

- Les résultats obtenus sur la population étudiée sont-ils fiables ?

- Si ces résultats obtenus sur la population étudiée sont fiables, sont-ils généralisables à une population plus large ?
- L'étude répond-elle aux attentes du demandeur ?

Pour cela, l'article fournit des clés méthodologiques, qui abordent successivement les points suivants :

1. Diverses approches de la qualité d'une étude,
2. Évaluation de l'adéquation des buts de l'étude aux besoins du demandeur,
3. Évaluation de la validité interne de l'étude,
4. Évaluation de la validité externe de l'étude,
5. Avis sur la qualité globale de l'étude,
6. Discussion sur la manière d'utiliser ces outils.

L'apprentissage de ces clés méthodologiques est conforté par la résolution de quelques cas que nous proposons, sous un mode d'enseignement programmé (problèmes à résoudre à l'aide de questions à choix multiples et organisées selon un itinéraire permettant l'apprentissage de la méthode d'évaluation) que nous espérons ludique et qui peut être consulté en ligne sur le site de l'AEEMA (<http://aeema.vet-alfort.fr/>, rubrique « *Épidémiologie pour tous* »). À terme, nous souhaitons mettre à disposition en ligne d'autres études de cas résolus, cette offre de formation restant réservée aux adhérents de l'AEEMA.

Cet article n'aborde pas les connaissances épidémiologiques autrement que sous forme d'allusions, le lecteur devant, si besoin, se reporter à d'autres sources, comme le livre édité par l'AEEMA [Toma *et al.*, 2010] ou le site de l'AEEMA (rubriques « Annexes », « Epidémiologie pour tous » ou « Terminologie » auxquelles les indications ^A, ^{EPT} ou ^T apparaissant dans le texte renvoient respectivement). Si cette méthode peut aussi être utilisée par tout auteur d'un projet ou d'un compte-rendu pour en vérifier lui-même la validité avant de le soumettre à un évaluateur, elle ne constitue pas la méthode de référence pour l'évaluation d'une étude épidémiologique par un évaluateur.

II - DIVERSES APPROCHES DE LA QUALITÉ D'UNE ÉTUDE

Pour commencer, il faut d'abord définir ce qu'on entend par « *qualité* ». Les définitions de l'Association Française de NORmalisation (AFNOR) sont pour cela extrêmement précieuses :

« *La qualité d'un produit ou d'un service peut se définir par les propriétés qui vont satisfaire les besoins exprimés et/ou implicites des utilisateurs* » (AFNOR).

Dans le cadre de l'épidémiologie, le « *produit* » ou le « *service* » est constitué par les résultats d'une étude épidémiologique et les « *utilisateurs* » sont des décisionnaires qui vont se servir de ces résultats pour prendre les décisions de santé qui sont de leur ressort.

Cette définition a l'avantage de mettre l'accent sur deux composantes : « *propriétés* » et « *utilisateurs* ». Le terme de « *propriétés* » d'un produit ou d'un service, évoque, pour une étude épidémiologique, les caractéristiques de son protocole. Le terme « *utilisateurs* » étend le champ de ces propriétés à leurs attentes, pour lesquelles il faut savoir tenir compte, selon cette définition, des besoins non seulement *exprimés*, mais aussi *implicites*. Il sera remplacé dans la suite de cet exposé par celui de *demandeurs*.

Ces commentaires nous amènent, par conséquent, à distinguer deux dimensions : la **qualité interne**, qui concerne les propriétés de l'étude, et la **qualité externe**, en relation avec les besoins des demandeurs.

□ « *La qualité interne désigne la manière dont sont élaborés les produits ou services pour l'utilisateur* » (AFNOR).

Pour une étude épidémiologique, c'est le respect des normes relatives à l'échantillonnage, la mesure, le traitement statistique des données, l'interprétation des résultats. Cette qualité interne est aussi appelée **validité interne**, qui va conditionner la validité des résultats à l'échelle de la population étudiée.

« *La validité d'un test, d'un examen ou d'une étude épidémiologique est sa capacité à mesurer ce qu'il est censé mesurer, sans être influencé par des sources d'erreurs aléatoires et systématiques* » [Toma et al., 1991].

La **validité interne** recouvre différents aspects (fiabilité, exactitude^T, fidélité, répétabilité^T,

reproductibilité^T) qui doivent pouvoir être identifiés dans la lecture d'un compte-rendu pour les évaluer. Ces aspects seront présentés plus en détail dans la suite de cet article.

□ La **qualité externe** comporte deux aspects, l'un **objectif**, l'autre **subjectif**.

L'aspect d'ordre **objectif** résulte de **critères** permettant de vérifier si les résultats de l'étude sont aptes à correspondre aux besoins de santé de la population pour laquelle le demandeur doit prendre une décision : les résultats obtenus sur l'échantillon (les échantillons) et « *inférés* » à la population d'où ils sont issus, doivent pouvoir être projetés à une population, qui peut « *déborder* » la population qui a fait l'objet de l'étude. C'est ce qu'on appelle la **validité externe**.

L'aspect d'ordre **subjectif** correspond aux **attentes, explicites** ou **implicites**, du demandeur. Or, il est important d'*anticiper* sur ces attentes pour pouvoir les prendre en compte de façon appropriée dans la conception de l'étude : par conséquent, c'est la richesse des échanges avec le demandeur de l'étude *avant* la conception de l'étude qui permettra de correctement circonscrire ces attentes explicites et surtout implicites, et ainsi adapter le protocole aux attentes du demandeur. Cet aspect constitue, par conséquent, le point de départ de la conception d'une étude épidémiologique.

Pour pouvoir être considérée de qualité, une étude doit donc cumuler les aspects de :

- validité interne qui relève de la démarche scientifique,
- validité externe, qui vise à étendre les conclusions de l'étude à une population plus large que celle de l'étude tout en étant adaptées aux attentes du demandeur.

Ce qui concerne l'adaptation aux attentes du demandeur peut faire l'objet d'une double évaluation : *a priori*, puisque la conception de l'étude doit en tenir compte et *a posteriori*, car il faut vérifier qu'au final, compte tenu de la validité interne, l'étude permet bien de répondre aux attentes du demandeur.

Les différents volets de cette démarche sont présentés successivement et sont résumés dans le tableau 1.

III - ÉVALUATION DE L'ADAPTATION AUX BESOINS

Pour pouvoir aborder l'évaluation de l'adaptation d'une étude épidémiologique aux besoins du demandeur, il faut d'abord évoquer la manière de procéder pour adapter une étude aux besoins d'un demandeur.

1. INTERDÉPENDANCE DE LA CONCEPTION DE L'ÉTUDE ET DE LA DÉCISION DE SANTÉ QUI EN DÉCOULE

La demande d'étude épidémiologique est formulée pour aider à prendre une décision dans le domaine de la santé. Par conséquent, l'étude épidémiologique se doit d'être adaptée à la décision qu'elle aide à prendre.

La représentation de la « *décision* » selon Lemaitre [Lemaitre, 1981] permet de préciser la relation entre étude épidémiologique et décision :

Si je fais ceci (idée d'action), et que les conditions sont cela (situation), alors j'obtiendrai probablement (risque) le résultat suivant, qui m'apportera telle utilité finale, compte tenu de mes aspirations traduites en critères de choix [Lemaitre, 1981].

S'agissant d'une décision dans le domaine de la santé et selon la conception présentée, les cinq premiers aspects ont à voir avec l'épidémiologie :

- L'**idée d'action** renvoie à un **programme de santé**, dont le suivi du déroulement fait appel aux méthodes de l'épidémiologie.
- La **situation** correspond à l'**état de santé de la population**, dont la mesure relève de l'*épidémiologie descriptive*, mais aussi aux mécanismes de propagation de la maladie dans la population, dont l'étude procède de l'*épidémiologie analytique*.
- La composante de **risque** fait référence aux probabilités, dont l'estimation est également réalisée par des méthodes épidémiologiques (**analyse de risque**).
- L'appréciation du **résultat** escompté du programme de santé utilise des **indicateurs épidémiologiques** (épidémiologie descriptive). Si on veut, dans une évaluation ultérieure, rapporter ces résultats à l'application du programme de santé, il faut prévoir un dispositif d'observation relevant de l'épidémiologie analytique.

- Quant à l'**utilité finale**, elle peut être appréciée par des outils qu'utilise l'épidémiologie pour l'analyse d'utilité (**arbres de probabilités, arbres de décision**).

Le fait que ces différentes composantes de la décision aient à voir avec l'épidémiologie signifie qu'il vaut mieux que le concepteur de l'étude en soit prévenu, de façon à anticiper une attente ultérieure du demandeur d'utiliser les résultats de l'étude pour l'un ou l'autre des aspects de la décision une fois celle-ci mise en œuvre. Cette anticipation peut conduire à penser le protocole sur n'importe lequel de ses aspects (échantillonnage, mesure, analyse statistique) de façon à ce que les résultats puissent servir non seulement à la prise de décision, mais aussi au-delà, lors de sa mise en œuvre.

Un seul aspect, le dernier, n'a rien à voir avec l'épidémiologie, puisqu'il concerne les aspirations personnelles du décisionnaire : la décision est bien l'affaire d'une *personne*, qui choisit en fonction de ses préférences, qui reflètent, implicitement, une échelle de valeurs, même si la décision devait être prise de façon collégiale par plusieurs personnes dans une instance *ad hoc*. Cela permet de comprendre pourquoi deux décisionnaires, confrontés au même ensemble d'informations, peuvent prendre une décision différente.

Par exemple, lorsqu'autrefois la lecture des tuberculinations était systématiquement subjective, on pouvait mettre en évidence une différence objective dans les pratiques des vétérinaires praticiens selon leur ancienneté professionnelle. Dans une même clientèle vétérinaire d'élevages bovins indemnes de tuberculose [Collin, 1987], les vétérinaires expérimentés constataient un taux de bovins et d'élevages présentant des réactions faussement positives bien plus faible que le jeune confrère débutant sa carrière : les anciens se basaient sur la connaissance intuitive de l'exploitation leur permettant d'apprécier le degré de risque, le plus souvent inexistant, que l'élevage soit confronté à l'infection par le bacille tuberculeux ; pour eux, le risque d'erreur qu'ils encouraient en classant ces réactions, qu'ils considéraient comme des erreurs par excès, comme des résultats négatifs, était facile à supporter, ce choix étant également attendu par leur client comme la manifestation de leur compétence à savoir faire le tri entre le vrai du faux, ce qui les confortait dans ce choix. Le débutant, ne

disposant pas de cette connaissance fruit d'une longue expérience, n'avait d'autre ressource que de recourir à un test de contrôle, car, pour lui, le risque d'erreur, même s'il pouvait être minime étant donné la très faible fréquence de l'infection dans cette clientèle, était incompatible avec sa situation de début de carrière, car comportant des conséquences potentielles (civiles, administratives voire pénales) totalement insupportables, du fait qu'il ne pourrait pas justifier de son autorité professionnelle pour se défendre, et que si le risque était réalisé, il serait beaucoup plus simple pour la communauté de lui faire assumer la responsabilité de cette erreur, plutôt que de reconnaître publiquement des pratiques non-conformes à la réglementation.

Par conséquent, compte tenu de cette composante personnelle, le concepteur de l'étude doit se limiter à apporter l'information nécessaire, sans interférer dans le processus de décision.

Par la suite, nous supposons que le concepteur de l'étude et le demandeur sont deux personnes différentes (ce qui n'est pas forcément toujours le cas), car, dans le cas de cumul de ces deux fonctions, le processus d'évaluation est alors rendu plus difficile du fait de l'implication personnelle.

2. MODALITÉS PRATIQUES DE L'ADAPTATION DE L'ÉTUDE AUX BESOINS

2.1 PROBLÈME POSÉ

Il faut prendre connaissance de la nature du besoin de santé et de la question posée.

Par exemple, telle maladie parasitaire paraît fréquente dans les élevages d'une région, d'après l'expérience des professionnels de l'élevage. La question posée est d'estimer plus objectivement cette fréquence dans la population concernée.

2.2 IDENTIFICATION DES ATTENTES DU DEMANDEUR

Il convient de se méfier d'une demande formulée par le demandeur de façon trop précise. En effet, lorsqu'un non spécialiste d'une discipline pose une question très précise à un spécialiste de cette discipline, la réponse, très précise, peut être exacte, mais non pertinente, car la question ne l'était pas.

Par exemple, la formulation d'une demande recourant au terme d'incidence peut être comprise de façon littérale par un

épidémiologiste, comme une étude permettant de suivre l'évolution du nombre de cas dans le temps, alors que le demandeur a utilisé ce terme dans un sens courant, pour désigner les répercussions d'une maladie sur l'état de santé global de la population, ou sur son potentiel de production.

À défaut de méthode proprement dite, on doit se contenter d'une approche, tournée vers la décision et l'action. Le concepteur d'étude épidémiologique doit prendre connaissance du type de décision (choix entre telle et telle orientation) et de l'action qui en résulte (impact attendu de santé sur la population), afin d'identifier les éléments d'information qui seront nécessaires et utiles au décisionnaire.

Les professionnels aguerris savent qu'il faut questionner le demandeur pour identifier plus précisément ses besoins véritables, au-delà de ceux qui ont été explicités, pour révéler les attentes implicites. Deux questions permettent de parvenir à ce résultat : « *Pourquoi ?* », en un seul mot et « *Pour quoi ?* », en deux mots. Le premier « pourquoi » amène l'interlocuteur à évoquer son raisonnement l'ayant conduit à la formulation de sa demande. Cette évocation peut être source d'informations très utiles pour compléter ou rectifier la demande. Le deuxième « pour quoi » repose sur une simulation : *si le demandeur disposait maintenant des résultats de l'étude, quelle décision prendrait-il ?* Cette évocation de l'utilisation des résultats de l'étude permet de mieux révéler l'information véritablement utile au demandeur.

2.3 FORMULATION DES BUTS ET OBJECTIFS

En réponse à la demande, le concepteur d'une étude va proposer un projet qui articule un plan de collecte de données tourné vers des buts à atteindre. Buts et objectifs sont deux termes utilisés souvent l'un pour l'autre, mais, s'ils désignent la même chose, à savoir le « *résultat escompté d'une action* », un objectif est formulé de façon plus précise que le but et nous allons voir quel rôle y joue l'épidémiologie.

Prenons l'exemple d'une étude dont le but serait de connaître la fréquence d'atteinte des élevages par une maladie donnée en vue de décider de l'opportunité de mise en place de mesures collectives de lutte. Cette formulation générale du but d'une étude est beaucoup trop imprécise pour garantir l'adéquation aux attentes du décisionnaire. Le recours à des **indicateurs épidémiologiques** permet d'introduire la précision nécessaire.

Un indicateur épidémiologique se présente sous forme d'un rapport, comportant un numérateur et un dénominateur. Le numérateur est le résultat d'un dénombrement dont on doit préciser l'unité épidémiologique (élevage, animaux, lots, etc.), la nature de l'événement observé (par exemple, cas de maladie qui doit être défini, résultat d'un dépistage), et la manière de prendre en compte le temps (prévalence, incidence). Le dénominateur dénombre les individus sur lesquels l'observation est conduite.

Dans notre exemple, il conviendrait de mesurer la fréquence des élevages atteints et, dans les élevages atteints, la fréquence des animaux atteints. Ce serait bien de disposer d'éléments permettant de chiffrer la gravité clinique, ce qui permettrait de mettre en relation les coûts de la lutte et le bénéfice qui pourrait en résulter.

Ces indicateurs vont constituer autant d'objectifs précis, qui seront utilisés pour la conception de l'étude afin d'explicitier des buts dont la formulation est nécessairement générale.

3. ÉVALUATION DE L'ADÉQUATION AUX BESOINS

Ne disposant pas de norme pour l'évaluation de l'adéquation de l'étude aux besoins du demandeur, on peut tout d'abord rechercher les traces de la

mise en relation des buts de l'étude épidémiologique avec la prise de décision.

L'absence d'évocation de la position d'un demandeur qui utilisera le résultat de l'étude dans le cadre d'une décision, de ses attentes, de l'usage futur de ces résultats est suspecte d'une mauvaise adéquation.

Au contraire, la formulation des buts d'une étude évoquant le demandeur et précisant ses attentes, sous forme d'une décision de santé ultérieure, est le gage d'une présomption d'adéquation de l'étude à ses attentes, sans preuve formelle toutefois.

Ensuite, il faut vérifier l'adéquation des buts formulés par rapport aux besoins d'information.

Si le besoin d'information est descriptif, on doit pouvoir reconnaître dans les buts des formulations renvoyant, au moins en partie, à la trilogie « population, temps, espace ».

Si le besoin d'information nécessite l'étude d'une relation causale, les buts doivent être de type explicatif et faire apparaître l'intention non seulement de mettre en évidence une association entre une ou des variables d'exposition et une maladie, mais aussi de mesurer l'intensité de cette relation dans des conditions permettant de neutraliser des facteurs de confusion éventuels.

IV - ÉVALUATION DE LA VALIDITÉ INTERNE

L'évaluation de la validité interne consiste à établir si les résultats sont valides à l'échelle de la **population étudiée** [Dohoo *et al.*, 2009] en recherchant l'absence de **biais**^T éventuels, c'est-à-dire une erreur systématique qui peut fausser les résultats et leur interprétation et qui peut porter sur l'échantillonnage (biais de sélection), la qualité des données (biais de classement) et l'analyse, l'interprétation des données (biais de confusion). La reconnaissance de l'absence de biais dans les résultats de l'étude rend possible l'*inférence*, c'est-à-dire la projection de ces résultats à la population étudiée (à ne pas confondre avec l'*extrapolation* qui relève de la validité externe).

Les critères de validité interne sont complètement différents selon que l'étude est descriptive ou explicative. Il est donc indispensable de savoir

rapidement repérer à laquelle de ces deux catégories l'étude appartient.

1. IDENTIFICATION DU TYPE D'ÉTUDE

Simple dans son principe, l'identification du type d'étude n'est pas aussi facile qu'il y paraît, en raison d'insuffisances fréquentes dans les formulations ou les choix opérés.

Si l'on veut déterminer un **pourcentage de prévalence** dans une **population**, dans une région, ou une **évolution de l'incidence** dans le temps, il s'agit d'une étude **descriptive**. En épidémiologie animale, ce type d'étude recourt *le plus souvent* à un seul échantillon (mais il peut aussi y en avoir plusieurs si l'étude porte sur plusieurs ensembles, comme des zones géographiques différentes).

Si l'on veut aborder la question de la **relation entre des facteurs d'exposition et la fréquence de l'apparition de la maladie**, cela caractérise une étude **explicative**. En épidémiologie animale, ce type d'étude recourt *le plus souvent* à deux échantillons, les études ne comportant qu'une seule cohorte étant beaucoup moins fréquentes.

En supposant que la demande soit à la fois descriptive et explicative, on considèrera le protocole comme descriptif pour son analyse de qualité : en effet, si une étude descriptive peut permettre, dans une certaine mesure, d'avoir une approche explicative, l'inverse n'est pas vrai. Mais la préoccupation explicative sera traduite, dans ce cas, en besoin d'analyse des données en vue d'aboutir à la formulation d'hypothèses explicatives, ce qui veut dire qu'il faudra s'en donner les moyens. En effet, la formulation d'un tel besoin combinant le descriptif et l'explicatif, conduira inéluctablement à augmenter d'autant les exigences du protocole.

Un cas particulier mérite d'être signalé, en raison de sa fréquence et de la difficulté pour le classer, celle d'une étude de prévalence comportant un seul échantillon, ce qui devrait permettre de la considérer comme descriptive ; mais l'analyse des données a conduit à comparer deux groupes, à l'intérieur de cet échantillon, en les distinguant selon leur exposition, en utilisant les ressources statistiques de l'épidémiologie analytique, ce qui amènerait à la considérer comme explicative.

La résolution de cette contradiction repose sur deux approches complémentaires, la manière dont l'échantillon a été constitué, et la position de la formulation d'hypothèse dans le cadre de la logique de raisonnement hypothético-déductif. L'échantillonnage, à bien y regarder, ne comporte en réalité qu'un seul échantillon constitué à partir de la population source ; c'est dans la phase ultérieure d'analyse des données que des groupes (exposés/non exposés ou cas/témoins) ont été constitués à partir de cet échantillon initial et non de la population source. Par conséquent, du fait qu'il n'y a en réalité qu'un seul échantillon sélectionné à partir de la population source, les critères retenus pour l'analyse de cette étude de prévalence devront être descriptifs.

Du point de vue de la logique hypothético-déductive, rappelons tout d'abord que la construction de connaissances nouvelles repose sur deux phases, la première de production de faits qui auront le statut d'hypothèse, la deuxième qui vise à soumettre cette hypothèse à l'épreuve des faits à l'aide d'un protocole choisi en fonction de la nature

de l'hypothèse : pour une hypothèse explicative, le protocole devra être conçu selon les modèles d'études analytiques. Ces deux phases portent sur des jeux de données différents, récoltés selon des modalités spécifiques et différentes pour chacune d'elles.

Pour ces deux raisons, dans une telle étude, les résultats à vocation explicative auront le statut d'hypothèse.

En résumé :

Si le but est descriptif (« population, temps, espace ») ainsi que le protocole (un échantillon, éventuellement plus), l'évaluation doit utiliser les critères convenant pour une étude descriptive. Si le but est explicatif ainsi que le protocole (au moins deux groupes), l'évaluation doit utiliser les critères convenant pour une étude analytique ;

Si but et/ou protocole sont « mixtes », l'étude peut être analysée successivement selon les grilles descriptive et analytique, afin de retenir celle qui conviendra finalement le mieux. Mais il est très probable qu'on devra finalement lui appliquer des critères descriptifs.

La suite de la démarche porte sur les trois composantes méthodologiques d'un protocole d'étude épidémiologique : échantillonnage, mesures, traitements statistiques, auxquels nous ajouterons l'interprétation.

2. ÉCHANTILLONNAGE

Quel que soit le type d'étude, l'échantillonnage doit toujours être **maîtrisé par l'investigateur**, afin d'écartier les risques de **biais d'échantillonnage**^T (dits aussi **biais de sélection**^T) (tableau 1). L'investigateur détermine la population source à partir de laquelle il sélectionnera l'échantillon, il définit le mode de sélection des individus de l'échantillon, il s'assure de la bonne réalisation de ce mode de sélection. Toute atteinte à ce principe fondamental, par un processus d'**auto-sélection** ayant échappé, partiellement ou en totalité, à l'investigateur, devra être prise en compte lors de l'interprétation des résultats. Les méthodes d'exploitation de grandes bases de données ne relèvent pas de notre propos.

L'échantillonnage porte sur deux aspects, le mode de sélection, que nous développerons, et le nombre de sujets nécessaire, que nous ne développerons pas, car une insuffisance dans ce domaine se traduit

par un manque de puissance ou de précision, ce qui n'induit pas une erreur dans le résultat.

Comme indiqué précédemment, les modalités d'échantillonnage sont différentes pour les études descriptives et explicatives.

2.1 ÉTUDES DESCRIPTIVES

L'échantillonnage, centré sur la population, comporte un échantillon, ou plus, sélectionné(s) par l'investigateur (tableau 1).

2.1.1 Deux types d'études

Les études descriptives peuvent viser à estimer un indicateur de santé (études quantitatives) dans une population, ou à classer une population par rapport à un seuil de prévalence (études qualitatives). Le critère de qualité majeur est cependant le même.

2.1.2 Critère de qualité majeur : la représentativité

L'échantillon doit ressembler suffisamment à la population étudiée de façon à en refléter fidèlement la situation épidémiologique. C'est pourquoi le critère de qualité *majeur* est la **représentativité**^{EPT} de l'échantillon, qui ne peut être atteinte que par les méthodes probabilistes, comme par exemple, le **tirage au sort** des individus qui constitueront l'échantillon, à partir d'une **base de sondage**^T. D'autres méthodes, non probabilistes peuvent être utilisées dans le cas où on ne dispose pas de base de sondage, comme la méthode des quotas par exemple, mais elles ont l'inconvénient de ne pas permettre le calcul de la marge d'incertitude, à la différence des méthodes probabilistes.

Seul le respect de ce critère de qualité, la **représentativité**, garantit que les résultats de l'échantillon pourront faire l'objet de l'**inférence statistique**, c'est-à-dire être appliqués à la **population source**^T, d'où provient l'échantillon.

La procédure permettant le respect de ce critère de qualité doit être exposée avec suffisamment de détails pour qu'on puisse en vérifier la validité. Toute formulation allusive, simplement déclarative d'une intention, sans précision suffisante pour en établir la validité, sera considérée comme suspecte d'un défaut de représentativité.

Exemple : Une enquête sur la tremblante du mouton a été réalisée lors d'un salon agricole sur

la fréquence des symptômes suspects. Un questionnaire était mis à disposition sur le stand de la Fédération des éleveurs. Une personne était chargée de solliciter les personnes passant à proximité du stand. Les organisateurs ont été satisfaits du grand nombre de réponses (plus de 600), grand nombre qui, à leurs yeux, garantissait la qualité des résultats. En fait, sur la seule question de la représentativité, l'échantillon constitué spontanément ne représentait aucunement la population des éleveurs qui constituaient la cible, puisque seuls ceux ayant assisté à cette manifestation pouvaient répondre, à condition d'être passés à proximité du stand et d'avoir accepté (et été en mesure) de remplir le questionnaire.

Les **critères d'inclusion, de non inclusion** ou d'**exclusion** des unités composant la population source et les unités de l'échantillon doivent être explicités et cohérents par rapport aux objectifs poursuivis.

Les procédures appliquées dans le cas de **refus** ou de manques divers (questionnaire mal rempli, prélèvements inexploitable) doivent également être explicitées.

Exemple : Une enquête sur la fréquence de la paratuberculose, présumée de l'ordre de 20 à 30 p. cent, a été réalisée par tirage au sort à partir de la liste départementale des éleveurs, mise à jour quelques mois auparavant pour les besoins de l'étude. Un échantillon de 200 individus avait été jugé nécessaire pour la précision attendue. Anticipant sur les refus possibles, la liste de sujets tirés au sort était de 300, afin de permettre de remplacer un sujet qui n'aurait pu être enquêté par un autre également tiré au sort, ce qui a été le cas pour 65 individus. Au final, les 200 sujets prévus ayant bien participé à l'enquête après avoir été tirés au sort, les organisateurs ont été satisfaits. Pourtant, cet échantillon n'avait pas la représentativité attendue du fait de la possibilité offerte à 65 sujets de ne pas participer à l'enquête. Même s'ils ont été remplacés par 65 autres sujets eux-mêmes tirés au sort, permettant de constituer l'ensemble prévu des 200 sujets tirés au sort, leur retrait de l'échantillon peut constituer un biais si la raison pour laquelle ils ont fait ce choix de se soustraire à l'enquête était liée à l'objet de l'étude. Leur proportion élevée, de même ordre de grandeur que la proportion attendue d'élevages atteints, est de nature à jeter une lourde suspicion sur la validité du résultat final :

la discussion devra aborder les raisons de ce refus et leur répercussion sur l'estimation.

La représentativité de l'échantillon doit être vérifiée par la comparaison des caractéristiques de l'échantillon et de la population étudiée. Pour cela, il faut avoir pris soin de procéder aux enregistrements d'informations nécessaires, en définissant les variables, généralement d'ordre démographique connues (type de production, race, taille de l'élevage, âge, sexe, etc.), qui permettront ce contrôle *a posteriori*.

Exemple : Une enquête a été réalisée (en 1984) dans une vingtaine de départements sur la fréquence d'accidents présumés liés à la vaccination contre la fièvre aphteuse (cette vaccination était obligatoire en France pour les bovins jusqu'en 1991). Un vétérinaire volontaire par département devait envoyer un questionnaire simple et rapide à remplir à 15 vétérinaires de son département qu'il aurait tirés au sort à partir de l'annuaire professionnel. Pour éviter les refus, la règle était de relancer par téléphone les confrères en insistant sur la nécessité de participer à cette enquête. Pour chaque département, il a été possible d'obtenir les 15 questionnaires, sauf pour quelques départements, pour lesquels il avait pu en manquer un ou deux. Ces manques n'ont pas été considérés comme susceptibles de fausser les résultats en raison de leur faible nombre. Le questionnaire permettant de récolter le nombre total de bovins et d'élevages par vétérinaire enquêté, ces chiffres ont pu être comparés aux états du recensement annuel effectué par département par les Directions départementales des Services vétérinaires pour le suivi de la prophylaxie de la tuberculose bovine afin d'en vérifier la représentativité. Aucune différence significative n'a été mise en évidence à l'aide d'un test statistique approprié, sauf pour deux départements (élevages d'effectifs nettement moins nombreux). Lors de la réunion de restitution des résultats, le coordonnateur de l'étude a interpellé individuellement les vétérinaires responsables de l'échantillonnage à l'échelle des départements, en leur rendant compte, sur la base de l'analyse des données, du respect ou non par chacun du protocole prévu : les deux confrères repérés, décontenancés par son extrême clairvoyance, ont rapidement reconnu qu'ils avaient préféré faire appel à des confrères de leurs amis, plutôt que tirés au sort, afin d'être sûrs d'obtenir leurs réponses, certains qu'ils étaient de ne pas les obtenir avec un échantillon tiré au sort.

Cette comparaison de l'échantillon à la population source ne permet pas de conclure de façon formelle sur la représentativité, qui ne peut être garantie que par une méthode d'échantillonnage probabiliste, l'absence d'écart constaté ne signifiant pas validation de la représentativité. En revanche, le constat d'un écart attire l'attention et sa portée sur la validité des résultats doit être abordée dans la discussion.

2.2 ÉTUDES EXPLICATIVES

L'échantillonnage des études explicatives est centré non pas sur la population mais sur la relation entre les facteurs étudiés et la survenue de la maladie. Le plus souvent en épidémiologie animale, il comporte **deux** échantillons, ou plus, sélectionnés chacun par l'investigateur *avant* le début des opérations de collecte en vue de comparaison (tableau 1).

2.2.1 Deux types d'études

Les échantillons peuvent être sélectionnés selon leur exposition, *exposés/non exposés* (étude de cohortes) ou selon leur état vis-à-vis de la maladie, *atteints/non atteints* (étude *cas/témoins*).

Exemple : Le GDS (Groupement de défense sanitaire) veut réaliser une enquête sur la relation entre les pratiques de traite et la fréquence d'apparition des mammites chez les vaches laitières. Si l'on choisit un protocole de type exposés/non exposés, on peut sélectionner des élevages selon la qualité de leurs pratiques de traite (à risque pour les uns, pas à risque pour les autres) en s'appuyant sur un réseau approprié de professionnels visitant de façon systématique les élevages (techniciens d'élevage, vétérinaires) ; dans les élevages ainsi repérés selon leur catégorie, à risque et non à risque, l'enquête conduira à noter l'apparition des cas de mammites pendant la durée de l'étude. Dans le cas d'une étude cas/témoins, on repère des élevages atteints et non atteints de mammites grâce aux données du contrôle laitier ; dans chacun des groupes ainsi constitués, on peut récolter les informations sur les pratiques de traite.

La réalisation de deux échantillons à partir d'un seul échantillon de départ, constitués *spontanément* selon leur état (atteints/non atteints) et/ou selon leur exposition (exposés/non exposés), ne répond pas à ces exigences et correspond toujours à un protocole de type descriptif, puisque ce n'est pas l'investigateur qui a effectué la sélection en ces

sous-catégories avant la constitution de l'échantillon.

Exemple : Dans le cadre de l'enquête sur la relation entre les pratiques de traite et la fréquence d'apparition des mammites, un échantillon de 100 élevages a été tiré au sort dans le fichier des adhérents. Les données de l'enquête ont permis de classer les élevages selon la qualité des pratiques de traite (à risque ou non) et la survenue de mammites (présence, absence). Les organisateurs de l'étude, du fait que la totalité des éleveurs sélectionnés avaient accepté de participer, ont validé la qualité de l'échantillonnage. Effectivement, celui-ci garantissait la représentativité de l'échantillon. Pourtant, les groupes avaient été constitués spontanément selon les caractéristiques de chaque individu du groupe (pratiques de traite, mammites), aboutissant à des effectifs déséquilibrés subis par les organisateurs de l'étude.

2.2.2 Critère de qualité majeur : la comparabilité

Le critère *majeur* de qualité d'une étude explicative est la **comparabilité** de ces échantillons, c'est-à-dire « la meilleure ressemblance possible des échantillons au facteur étudié près » (« Épidémiologie descriptive versus analytique »^{EPT}).

Elle vise à maîtriser un *facteur de confusion* (ou plusieurs), c'est-à-dire un facteur, autre que celui étudié, qui fausse la mesure de la relation entre le facteur d'exposition et la fréquence d'apparition de la maladie [Ancelle, 2002]. Il faut pour cela définir des critères, généralement d'ordre démographique, qui permettront de satisfaire à cette exigence. Diverses procédures peuvent être utilisées pour la maîtrise des facteurs de confusion, selon qu'elles sont utilisées *a priori* ou *a posteriori*.

La procédure *a priori* repose sur une sélection appropriée des sujets des échantillons :

- Par **appariement** (constitution de « paires »), qui consiste à choisir un ou plusieurs sujets qui correspondront à chacun de ceux choisis en première instance (exposés ou cas selon le type d'étude) sur la base de critères pour lesquels ils seront considérés comparables ;
Exemple : Dans l'enquête sur les pratiques de traite, on commence par définir les critères de sélection des individus. On écarte les élevages de races laitières ou de pratiques de traite (pas de salle de traite) faiblement représentées dans le

département, voire de cas particuliers (élevages en cessation d'activité). On définit ensuite les critères qui permettront d'assurer la comparabilité : nombre de vaches laitières en production par élevage, en acceptant une marge de fluctuation, en s'appuyant sur les statistiques départementales (par exemple, le seuil de 15 animaux près). Puis, on considère en premier lieu le groupe le moins représenté dans la population, par exemple, celui des exposés ou celui des cas, selon le type de protocole choisi. Pour chaque individu appartenant à ce premier groupe d'une taille donnée, il faudra parvenir à sélectionner au moins un autre appartenant au deuxième groupe de nombre de vaches laitières en production correspondant à la fourchette retenue. Si l'on avait choisi de prendre en compte d'autres facteurs d'appariement (l'âge de l'éleveur, la zone géographique), il aurait été préférable de retenir deux, voire trois sujets considérés comme similaires, car du fait du plus grand nombre de critères d'appariement, il aurait été plus difficile de constituer des paires rigoureusement identique.

- Par **tirage au sort**, dans un essai terrain, qui permet d'affecter les individus selon leur lot (exposés ou non exposés).
Exemple : Dans un essai thérapeutique, un certain nombre de patients sont affectés par tirage au sort, en nombre équivalent, aux différents groupes de l'essai.

La procédure *a posteriori* consiste à effectuer les **ajustements** statistiques, nécessaires dans le calcul des indicateurs d'intensité de relation entre facteur d'exposition et fréquence d'apparition de la maladie en prenant en compte une variable qui peut jouer un rôle de facteur de confusion. L'ajustement consiste à calculer l'indicateur (risque relatif, RR ou odds ratio, OR) pour chaque strate de la variable pour laquelle on veut réaliser l'ajustement, puis à en calculer la valeur pondérée globale à l'aide d'un logiciel statistique approprié. Mais il faut pour cela que les effectifs soient suffisants dans chaque strate pour que les calculs soient possibles et fournis avec la précision nécessaire, ce qui doit avoir été pris en compte lors de la conception du protocole. Lorsque plusieurs variables causales et/ou de confusion peuvent intervenir, les techniques d'analyse multivariée peuvent être utilisées, comme la régression logistique si la variable étudiée est binaire. Elles permettent de calculer des OR ajustés sur l'ensemble des facteurs inclus dans l'analyse.

Le choix entre appariement et ajustement dépend du type de problème étudié, chaque méthode ayant ses avantages et ses inconvénients [Bouyer *et al.*, 2009] : l'appariement nécessite un bien moins grand nombre de sujets, mais il n'est pas possible de prendre en compte le rôle éventuel de la variable qui a servi à constituer la comparabilité, et le nombre de variables servant à l'appariement doit être limité.

3. MESURES

Par mesure, il faut entendre l'ensemble des informations récoltées permettant d'effectuer les enregistrements nécessaires à l'estimation d'un indicateur de santé, comme incidence ou prévalence, ou d'une relation entre une variable et l'apparition d'une maladie.

Pour l'état vis-à-vis d'une maladie, il peut s'agir d'un diagnostic clinique (malade, non malade), d'un résultat à un test de dépistage ou à un test d'isolement ou de détection d'un agent pathogène. Les variables associées peuvent être descriptives (âge, sexe, race, type de production, par exemple), ou avoir une vocation analytique (les mêmes que précédemment, ou tout autres variables explicatives ou facteurs de confusion).

Ces mesures doivent être *fiabes*, c'est-à-dire qu'elles doivent refléter la réalité de façon fidèle et ne pas comporter de biais (**biais de classement**). Les qualités intrinsèques des outils de mesure interviennent (sensibilité, spécificité) et doivent être précisées (tableau 1).

Par exemple, une technique de dépistage de spécificité limitée (0,9 voire moins) peut conduire à une qualité de résultat (valeur prédictive positive) médiocre en situation de fréquence faible de la maladie, conduisant à une surestimation de la fréquence de la maladie en raison de nombreux résultats faussement positifs.

Mais les conditions d'application doivent aussi être connues, en particulier les enquêteurs doivent appliquer des procédures permettant de standardiser les informations qu'ils collectent afin de réduire la variabilité propre aux enquêteurs.

L'utilisation d'un questionnaire est un bon moyen de standardisation, à la condition que les enquêteurs respectent scrupuleusement le libellé des questions posées et les recommandations données lors de la réalisation du questionnaire : dans le cas où la personne

enquêtée n'a pas compris, il est généralement recommandé de simplement répéter la question et non de la reformuler en utilisant d'autres termes, ce qui pourrait introduire un biais selon le talent d'improvisation de l'enquêteur.

La récolte des données reposant sur un travail d'équipe, des dispositions, le plus souvent la formation des intervenants, doivent avoir été prises ou prévues pour maîtriser les questions de *répétabilité*^T, de *reproductibilité*^T et en vérifier la validité. Si ces aspects ne sont pas évoqués, cela veut dire qu'une part d'incertitude inconnue est potentiellement présente dans les résultats.

Après une formation initiale des enquêteurs à l'utilisation d'un questionnaire, ou d'une grille de codification visuelle des observations, il peut être utile de vérifier à la fin de l'enquête le niveau de performance des enquêteurs. En effet, les enquêteurs placés en situation extrême (faible fréquence de niveau de la variable observée, par exemple) peuvent avoir progressivement « dérivé », pour compenser cette faible fréquence qu'inconsciemment ils jugeaient suspecte.

Après leur récolte, les questionnaires doivent systématiquement être relus par une personne en charge du traitement des données, pour vérifier l'absence d'oublis et permettre une correction éventuelle.

Les tris à plat peuvent conduire à révéler des anomalies. Par exemple, dans une enquête sur les boiteries des vaches laitières, un enquêteur avait cumulé à lui seul une proportion considérable de l'ensemble des observations pour une variable décrivant le logement des vaches (longueur suffisante, ou non, de l'espace individuel dans une stabulation entravée) ; la vérification par téléphone a permis de révéler qu'il avait mal rempli le questionnaire, qui, de fait, est apparu mal conçu : les réponses au questionnaire se faisaient en cochant des cases, sauf pour la question litigieuse, où l'instruction avait été donnée de « rayer la mention inutile », ce que l'enquêteur n'avait pas vu ; ayant coché « longueur suffisante », cette information avait été interprétée par l'opérateur de saisie conformément à l'instruction comme une mention rayée et par conséquent codée comme une « longueur insuffisante ». La correction a permis de faire disparaître l'anomalie.

4. ANALYSE ET INTERPRÉTATION DES DONNÉES

4.1 RÈGLES COMMUNES AUX DIFFÉRENTS TYPES D'ÉTUDES

De la même façon, la description des traitements statistiques doit permettre de vérifier leur fiabilité, en cherchant les risques de **biais de confusion**. Par exemple, le logiciel de saisie et d'enregistrement doit garantir la fiabilité des enregistrements et des opérations de calcul. Un tableur de bureautique n'offre pas spontanément ces garanties et peut être la source d'erreurs d'autant plus pernicieuses qu'elles sont cachées (ces logiciels sont conçus pour faciliter les tâches des utilisateurs de bureautique, en anticipant sur leurs intentions présumées à l'aide de formules « macro » qui peuvent facilement introduire le désordre dans la gestion de données à finalité statistique) et demandent des procédures de vérification diverses pour les révéler. Les logiciels statistiques, comme Epi-Info™, ne comportent pas ce type de risque et fournissent les garanties nécessaires pour la gestion des données.

Les traitements statistiques doivent être annoncés, ainsi que le seuil de probabilité retenu pour le risque d'erreur de première espèce (risque alpha). Les conditions d'indication et de validité des tests doivent être respectées (tableau 1). Or, en santé animale, les effectifs animaux sont toujours regroupés du fait des conditions d'élevage : réaliser un test statistique sur l'ensemble des animaux sans tenir compte de ces regroupements constitue une infraction au principe d'*indépendance des données* (c'est-à-dire indépendance des unités d'un échantillon entre elles), conduisant à remettre en question la validité des résultats. Le traitement statistique peut prévoir d'utiliser des procédures statistiques appropriées pour en tenir compte. Les résultats doivent être rapportés conformément aux usages statistiques, en particulier en ce qui concerne l'incertitude associée (intervalle de confiance, cf. « *Fluctuations d'échantillonnage* »^A) lorsqu'il s'agit d'estimations par sondage.

Les résultats doivent répondre à la question posée !

Par exemple, un objectif annoncé d'estimation d'un pourcentage d'élevages atteints par une maladie ne peut être satisfait par un pourcentage d'animaux atteints par cette même maladie, quelle que soit l'explication donnée pour justifier cette discordance.

4.2 CONDITIONS PROPRES AUX ÉTUDES DESCRIPTIVES

L'échantillon doit être décrit : nombre total de sujets et ventilation des effectifs selon les variables

descriptives essentielles (sexe, âge, etc.) permettant de vérifier la conformité des résultats à ce nombre, sinon les indications d'un écart éventuel doivent être données.

Cette description doit permettre la comparaison à la population étudiée pour vérifier que l'échantillon lui ressemble suffisamment (ou n'en diffère pas de façon significative), car la sélection des échantillons selon une méthode valide peut avoir été altérée pour quelque raison que ce soit et il faut disposer d'éléments attestant que la représentativité souhaitée a été effectivement atteinte. Les résultats doivent être présentés en respectant les normes statistiques (intervalle de confiance ; mention de significativité des écarts lors de comparaisons éventuelles ; utilisation d'une méthode de standardisation dans le cas de comparaisons entre groupes).

Le non-respect de ces normes est suspect, de même qu'un nombre excessif de chiffres indicatifs après la virgule.

Ils doivent porter sur des aspects strictement descriptifs, c'est-à-dire des pourcentages selon les grands axes « *population, espace, temps* ». Ils ne doivent pas comporter d'interprétation à vocation explicative.

Par exemple, dans une étude sur la fréquence d'atteinte de chiens par la leishmaniose, le constat d'un pourcentage d'atteinte plus élevé pour une race de chiens ne permet pas d'en déduire que cette race est plus « sensible » que les autres, cet adjectif étant délibérément explicatif, d'autres facteurs (de confusion) pouvant expliquer cet écart (plus grande taille des chiens les exposant à un plus grand risque de piqûre par les phlébotomes ; choix par les propriétaires de ce type de race résidant dans une zone pavillonnaire permettant l'entretien de ces animaux mais correspondant à une zone de plus grande exposition, etc.).

4.3 CONDITIONS PROPRES AUX ÉTUDES EXPLICATIVES

Il faut vérifier que les conditions d'interprétation des résultats portant sur la relation entre le facteur d'exposition et la fréquence d'apparition de la maladie sont réunies, tout en tenant compte du fait qu'il s'agit d'une étude d'*observation*, qui n'a pas la même puissance d'interprétation causale qu'une expérience de laboratoire.

Les critères d'interprétation causale sont nombreux, ce qui permet d'en déduire que l'interprétation causale d'une étude d'observation

n'est pas facile ! Dans le cadre de cet article, nous sommes dans une démarche inverse, consistant à rechercher des défauts méthodologiques, ce qui est beaucoup plus facile : on peut, en effet, dégager quelques grandes lignes simples, très robustes, qui permettent de détecter les défauts les plus fréquents limitant toute velléité d'interprétation causale. On peut les résumer, de façon pédagogique, par la formule mnémotechnique de « règle des trois A », correspondant à la première lettre du terme clé correspondant (tableau 1).

Association suffisante : ce critère est le plus facile à vérifier. Il faut non seulement qu'une association ait été établie (écart significatif, par exemple à la suite d'un χ^2), mais aussi que l'intensité de la relation, mesurée par un risque relatif (RR) ou un odds ratio (OR), soit « suffisante ». Ce terme de « suffisant » vise d'abord à souligner qu'il faut disposer de ce type d'indicateur et pas simplement se contenter d'un résultat de χ^2 : toute interprétation, dans cet esprit, d'une valeur de « p » (la « p value » des logiciels statistiques), aussi faible soit-elle, est totalement illusoire du fait que le χ^2 est sensible aux effectifs. Ce terme de « suffisant » vise ensuite à souligner qu'il faut une valeur de RR ou d'OR qui ait « du sens », d'un point de vue épidémiologique. Un OR juste légèrement supérieur à 1, même s'il est significatif, n'emportera pas franchement la conviction.

Antériorité : du fait que la cause précède l'effet, le protocole doit permettre d'établir que le facteur d'exposition a bien précédé l'apparition de la maladie. Une étude de cohortes apporte normalement ce type de garantie du fait qu'elle consiste à mesurer l'incidence. Pour une étude cas/témoin, il faut vérifier que les cas utilisés étaient

bien des cas récents, de façon à établir que l'exposition au facteur lui était antérieure. À l'inverse, une étude de prévalence n'apporte pas ce type de garantie, sauf circonstances particulières. Par exemple, pour des réactions sérologiques qui ont pu permettre de classer les individus selon leur atteinte, certaines catégories d'anticorps dont la durée de vie est relativement courte permettent d'interpréter les cas détectés comme des cas incidents.

Absence de facteur de confusion : le protocole doit avoir pris en compte les facteurs de confusion éventuels, de façon appropriée, que ce soit par appariement et/ou par ajustement des indicateurs correspondants. L'absence de l'indication de l'ajustement des valeurs de ces indicateurs par rapport aux facteurs de confusion pris en compte signale la non satisfaction à ce critère.

Si un seul de ces trois critères n'est pas satisfait, il n'est pas possible d'interpréter les résultats dans une perspective explicative de façon fiable. S'ils sont satisfaits, il n'est toutefois pas possible de conclure de façon définitive sur la validité de l'interprétation causale : le fait d'avoir réussi à franchir ces étapes de validation permet de réduire indéniablement l'incertitude que l'on pouvait avoir avant de réaliser l'étude et, par exemple, de légitimer une prise de décision ; mais ces résultats ne permettent pas d'établir à eux seuls la validité des conclusions en tant que *connaissance* scientifiquement indiscutable sur la relation causale. Le processus de validation de connaissances nouvelles est un processus complexe de la démarche scientifique, qui sort très nettement des intentions de notre article, qui est focalisé sur l'analyse de la validité d'une étude.

V - ÉVALUATION DE LA VALIDITÉ EXTERNE

L'évaluation de la validité externe consiste à vérifier s'il est possible de transposer les résultats observés sur le (ou les) échantillon(s) à la population cible, qui peut être plus large que la population étudiée, voire d'extrapoler à une population beaucoup plus vaste, dans la perspective d'une généralisation dans le cadre de la production de nouvelles connaissances scientifiques (tableau 1).

Elle n'est envisageable bien entendu que si la validité interne a été satisfaite.

Les résultats de l'étude doivent être confrontés aux travaux d'autres auteurs. Si la cohérence peut conforter, la discordance doit susciter l'interrogation sur la validité des résultats.

1. ÉTUDE DESCRIPTIVE

Une étude descriptive peut être reconnue valide sur le plan de sa validité interne, du fait que l'échantillon est représentatif de la population *source*. En ce qui concerne la validité externe, encore faut-il que la *population source* d'où est tiré

l'échantillon correspond bien à la **population cible**^T. Or, par exemple, comme pour effectuer un tirage au sort il faut une **liste de sondage**, celle-ci peut ne pas correspondre à la totalité de la population cible. Tout écart éventuel sur ce plan devra être étudié lors de l'interprétation des résultats.

Exemple : Une enquête a été réalisée par téléphone, après tirage au sort par un ordinateur à partir d'un annuaire afin d'obtenir un échantillon représentatif. Toutefois, à l'heure actuelle, de nombreuses personnes n'ont plus de téléphone filaire et nombre d'abonnés se sont inscrits sur des listes leur permettant d'éviter de faire l'objet de démarchage téléphonique. Il faut en plus que les correspondants soient présents chez eux au moment où l'appel est effectué, ce qui exclut les personnes qui travaillent et renforce la probabilité de joindre une personne inactive (retraités, chômeurs, etc.). Bien que tiré au sort, l'échantillon ne sera pas représentatif de la population cible.

Autant la validité interne est généralement facilement vérifiable (respect ou non des procédures de sélection permettant d'exclure un risque de biais), autant l'évaluation de l'adéquation de la population source à la population cible peut être délicate.

La comparaison de la description de l'échantillon à la population cible fournit une information complémentaire qui peut être utile pour détecter un défaut dans la qualité d'échantillonnage qui ne serait pas apparu à l'examen du protocole. Un écart significatif, certes, peut être le simple fait du hasard, mais il doit malgré tout attirer l'attention et inciter à vérifier la validité des informations sur les conditions d'échantillonnage qui n'étaient peut-être pas si conformes qu'il n'était initialement apparu.

2. ÉTUDE EXPLICATIVE

L'évaluation de la validité externe, dans le cas des études explicatives, revient à poser la question : « *est-ce que la relation entre facteur d'exposition et fréquence d'apparition de la maladie observée dans cette étude, et dont la validité interne a été établie, reflète bien cette relation dans la population cible ?* ».

Conclure positivement est difficile, et nous n'aborderons pas cet aspect. En revanche, comme précédemment, nous soulignerons quelques clés permettant de révéler des défauts manifestes. Le plus facile est de repérer un biais de sélection sur l'échantillon clé (individus exposés ou cas) : si ces individus, pour une raison ou pour une autre, ne représentent pas l'ensemble de la population d'individus exposés ou cas, selon le type d'étude, qu'ils sont censés représenter, on dispose de la preuve qu'il ne sera pas possible d'appliquer les résultats explicatifs à l'échelle de la population cible.

Par exemple, les individus recrutés proviennent d'un établissement hospitalier réputé drainant des individus non représentatifs des cas survenant dans la population.

La conclusion relative à l'échantillon peut demeurer valide, si par exemple le critère de comparabilité est satisfait, mais l'inférence ne sera possible que vis-à-vis de la population source, et l'application des résultats à la population cible sera limitée à la population qui « ressemble » suffisamment à celle de l'échantillon... et encore, car il faudrait prendre en compte le biais de sélection résultant de l'auto-sélection des individus qui se sont rendus dans cet établissement hospitalier réputé ! Nous n'aborderons pas d'autres sources de limitation de l'interprétation à l'échelle de la population cible, qui sont nombreuses, subtiles, et dépassent le cadre de cet exposé.

VI - AVIS SUR LA QUALITÉ GLOBALE DE L'ÉTUDE

Après cette approche analytique de l'évaluation de la qualité de l'étude, il reste à en faire la synthèse. Pour cela, il convient de reprendre les points principaux de l'évaluation de l'étude au travers d'une succession d'étapes (cf. encadré 1). L'approche pour un compte-rendu d'étude épidémiologique est la plus complète du fait de la disponibilité des résultats et de la discussion,

celle d'un projet s'arrêtant à l'énoncé des méthodes, avant la présentation des résultats.

□ Préambule

Pour chacune des étapes, on doit être en mesure de procéder à l'évaluation et, pour cela, il faut au minimum disposer de l'information nécessaire : l'absence est un très mauvais point.

L'évaluation vise à faire apparaître les points forts, les points faibles, afin de juger si les critères de qualité sont satisfaits ou non. De plus, le jugement doit porter aussi sur la pertinence d'ensemble : cohérence des différents maillons de l'étude entre eux et pertinence avec les besoins.

Proposition d'une grille qualitative d'évaluation

Vu le nombre d'étapes et de points à considérer, cette entreprise peut être considérée comme difficile pour le novice. C'est pourquoi nous proposons une échelle qualitative permettant d'exprimer le jugement en y apportant des nuances, si besoin.

L'échelle que nous proposons comporte quatre degrés, le plus faible étant le degré zéro, synonyme de l'absence de conformité de la réalisation, ce qui donne le degré trois pour le plus élevé. Les indications pour aider à déterminer le jugement sont les suivantes.

Commençons par le degré de conformité optimale (degré 3) qui correspond à l'avis « totalement satisfaisant » ; en dessous nous pouvons placer un degré également favorable mais nuancé (2), comme « pas totalement satisfaisant, en raison d'insuffisances qui peuvent cependant être corrigées ou tolérées » ; en dessous encore, sont situés les degrés défavorables, le premier de ceux-ci (1) étant nuancé et exprimant l'avis « très partiellement satisfaisant, de nombreux points insuffisants, rendant l'interprétation des résultats délicate ou limitée », et le dernier sans nuance (0) « pas du tout satisfaisant, un ou des aspects franchement réhivitoire(s) rendant toute exploitation des résultats impossible ».

Cette échelle à quatre degrés a l'avantage de ne pas comporter de degré neutre, de façon à obliger à porter un jugement, favorable ou défavorable ; elle autorise la nuance pour ce qui n'est pas totalement satisfaisant ou non, sans envisager d'autres degrés intermédiaires pour lesquels un jugement plus affiné pourrait être délicat pour un débutant ; de ce fait, elle est facile à mettre en œuvre, même par une personne peu expérimentée.

L'absence d'information sur un point donné pourra être prise en compte de façon variable selon son importance : tantôt elle conduira seulement à

nuancer un avis, tantôt au contraire elle pourra aboutir à exprimer un avis résolument défavorable.

Cet avis global doit être abordé en deux parties, la première porte sur le jugement à l'échelle scientifique, la deuxième sur le degré de satisfaction des attentes.

1. JUGEMENT À L'ÉCHELLE SCIENTIFIQUE

Rappel du problème posé

Pour la partie scientifique, il faut commencer par le rappel des « commémoratifs », sous forme de la formulation initiale du problème posé auquel correspond un résultat attendu, pour servir quel besoin, de santé. Les informations nécessaires doivent être disponibles et la relation logique entre les différents éléments exposés doit être perceptible et pertinente.

Adéquation *a priori* aux attentes du demandeur

Les attentes du demandeur doivent être mentionnées, qu'elles soient explicites ou implicites : à quelle décision devront servir les résultats, pour quelle intention ultérieure éventuelle ? L'étude doit avoir été conçue de façon adaptée à ces attentes et il doit être possible de se faire une idée sur la question dès le début du document.

Jugement sur la validité interne

Il faut ensuite mettre en perspective les éléments du protocole tels qu'ils apparaissent au chapitre des « matériels et méthodes » : le type de l'étude, descriptif ou explicatif, les éléments de la validité interne (qui comporte trois maillons : échantillonnage, mesures, analyse statistique) doivent répondre, d'une part, aux critères de qualité correspondants, d'autre part, aux besoins et aux attentes. Cette partie comportant de nombreux aspects, la stratégie simplificatrice à adopter consiste à vérifier la solidité de chacun des maillons de cette chaîne de récolte et de traitement d'information, ce qui devrait être le cas le plus fréquent pour une étude correctement menée, et surtout, dans le cas contraire, à repérer non seulement les défauts, mais surtout s'il existe un maillon faible pour en apprécier la répercussion sur l'ensemble de l'étude.

Dans un compte-rendu, la poursuite de l'évaluation est possible avec la présentation des résultats et

leur interprétation, qui doivent de la même façon satisfaire aux critères de qualité qui leur sont spécifiques du point de vue de la validité interne et de la pertinence à la question posée, tout en respectant la cohérence de qualité permise par le protocole.

Un premier jugement *a priori* portant sur la validité interne des résultats peut alors être formulé : soit les trois maillons sont satisfaisants, les résultats sont fiables et l'inférence à la population étudiée est permise ; soit l'un des maillons est non satisfaisant, et l'interprétation des résultats sera à nuancer en fonction de la gravité et de l'impact des insuffisances ; enfin, si un seul maillon est totalement insatisfaisant, l'interprétation des résultats est impossible.

Ce jugement *a priori* peut être nuancé grâce aux informations qui figurent dans la discussion, qui est une partie cruciale d'un compte-rendu, car, normalement, il doit être possible d'y trouver les réponses aux interrogations qui ont pu naître à la lecture du document. En effet, aucune étude épidémiologique ne peut être parfaite, car c'est le sort des études d'observation d'être soumises aux aléas d'une réalité qui ne se conforme jamais aux attentes des investigateurs ! Par conséquent, l'absence de discussion est un caractère définitivement réhibitoire.

Le premier point abordé dans la discussion est normalement celui du jugement que les auteurs se doivent d'envisager eux-mêmes, sur la validité interne de leur étude. Les points repérés par le lecteur dans son analyse doivent pouvoir trouver des éléments de réponse, peut-être de justification ; les auteurs doivent formuler un jugement sur la répercussion éventuelle des défauts inéluctables de toute étude épidémiologique sur la validité des résultats, et ce jugement doit rejoindre celui du lecteur, voire le dépasser, car les auteurs doivent être eux-mêmes les plus rigoureux de leurs juges. L'absence de réponse à des interrogations qui ont pu naître de la lecture du compte-rendu doit être prise en compte, de façon défavorable.

A l'issue de ce complément d'information, le lecteur doit être parvenu à un jugement définitif sur la fiabilité des résultats et de leur interprétation, après avoir soit tempéré le premier jugement, soit l'avoir confirmé.

☐ Jugement sur la validité externe

A ce stade, la confrontation du protocole et des informations émanant de la présentation des

résultats permet au lecteur une première approche de la validité externe, avis qu'il laisse en suspens pour le moment.

Dans le cas qu'on peut espérer le plus fréquent où le jugement portant sur la validité interne est favorable, même s'il peut être nuancé, la première question qui se pose est de savoir si les résultats peuvent bien être appliqués à la population cible, du fait que la population source peut être légèrement, voire notablement, différente. Ensuite, les auteurs peuvent, si besoin, envisager aussi l'opportunité d'une généralisation des résultats, et pour cela la confrontation aux résultats d'autres publications est indispensable. Des résultats d'autres publications cohérents avec ceux de l'étude confortent un avis favorable ; dans le cas contraire, la suspicion sur la validité étant légitime en première intention, il faut que les auteurs puissent expliquer de façon convaincante pourquoi leurs résultats demeurent dignes de considération, malgré leur position divergente par rapport aux autres travaux de la communauté scientifique.

Un dernier point de la discussion porte sur les perspectives. Il peut s'agir, tout d'abord, d'un avis sur l'utilité de reprendre les travaux en vue de confirmation, voire de les poursuivre, car toute confrontation au réel est source d'enseignement, que ce soit au plan méthodologique (le simple fait de conduire une étude révèle toujours quelque chose qui peut être source d'enseignement, surtout si elle comportait des défauts), ou au plan de la connaissance, ce qui permet de mettre en application un principe bien connu et systématiquement appliqué en épidémiologie « *on fera mieux la prochaine fois* » : sans ce principe, il ne serait pas possible d'espérer pouvoir réaliser des études d'observation valides, chacune profitant de l'expérience de la précédente. Cette partie de discussion peut aussi conduire à sortir du cadre des résultats, en reliant l'apport de l'étude à d'autres connaissances, d'autres projets, d'autres interrogations afin d'esquisser des pistes possibles pour de nouveaux travaux.

Après cette revue des différentes étapes de l'évaluation, on est alors en mesure de prononcer un premier jugement global sur l'étude, sous son aspect scientifique, en cumulant les différents jugements émis à chacune de ces étapes, pour répondre finalement à la question : c'est de « *l'info* », ou de « *l'into* ». Mais, comment faire ? Compte tenu du fait que l'échelle utilisée est qualitative, précisons que toute tentative d'utilisation des valeurs apparaissant sous forme de chiffre pour effectuer des opérations diverses

(sommes, moyennes ou autres) serait hasardeuse, car nous aurions pu tout aussi bien utiliser des lettres, mais ces chiffres ont l'avantage de rendre plus clairement compte de l'organisation des avis selon une échelle ordinale. Par conséquent, l'avis global utilisant ces valeurs est une opération qui fait appel au « jugement » dans son sens le plus littéral, c'est-à-dire qu'il laisse une large part à la subjectivité dans cette dernière phase, les précédentes ayant au contraire utilisé davantage l'objectivité : le plus simple est d'utiliser la même échelle qualitative, mais cette fois à l'échelle globale. Avec un peu d'expérience, il est facile de vérifier que cette opération n'est pas si difficile qu'il peut y paraître : c'est pourquoi nous proposons des exercices (à télécharger sur le site de l'AEEMA) pour faciliter l'entraînement à cette capacité de jugement.

Dans le cas d'un jugement défavorable portant sur un projet d'étude, il convient de formuler des propositions d'amélioration du protocole en se servant de l'analyse étape par étape, en considérant en priorité les défauts majeurs. Il peut arriver que les insuffisances soient si importantes que la conception d'ensemble du projet doive être envisagée à nouveau.

2. JUGEMENT GLOBAL RELATIF À LA SATISFACTION DES ATTENTES DU DEMANDEUR

Il reste encore, pour un compte-rendu d'étude, à

formuler un deuxième jugement global portant sur la satisfaction des attentes du demandeur, jugée *a posteriori* en tenant compte de l'évaluation au plan scientifique : les défauts constatés peuvent altérer de façon plus ou moins importante l'atteinte de cet objectif, car il est logique qu'il faille une étude de qualité suffisante pour répondre à ces attentes, et si l'étude avait été conçue *a priori* de façon à satisfaire les attentes, les défauts constatés pourraient rendre cet objectif inatteignable. Mais il convient malgré tout d'examiner le jugement concernant la satisfaction des attentes indépendamment de l'évaluation au niveau scientifique, car le degré de tolérance aux écarts à la qualité peut ne pas être le même aux niveaux scientifique et pratique : une étude comportant des défauts jugés importants au plan scientifique peut malgré tout être utile pour la prise de décision, car elle aura quand même permis de réduire suffisamment l'incertitude. Il peut aussi advenir une situation extrême, dans laquelle, d'une façon qui peut paraître surprenante, une étude jugée invalide au plan scientifique puisse malgré tout servir les intentions du demandeur de façon satisfaisante (voir un des exercices proposés).

VII – DISCUSSION DE LA MÉTHODE D'ÉVALUATION PROPOSÉE

Les étapes de l'évaluation d'une étude épidémiologique étant présentées dans l'encadré 1, et les outils correspondants étant résumés dans le tableau 1, il reste à voir comment les mettre en œuvre d'un point de vue pratique.

1. MODE DE MISE EN ŒUVRE DE LA MÉTHODE D'ÉVALUATION

Nous avons présenté les étapes de façon linéaire, car il n'était pas possible de procéder autrement. Malgré ce que suggère l'encadré 1, cela ne veut pas dire que l'évaluation d'une étude doit impérativement suivre ces différentes étapes, dans l'ordre du document et de bout en bout.

Les informations peuvent être à rechercher à différents endroits. Le plan classique d'un compte-

rendu est pourtant bien connu : introduction, où sont données les informations générales permettant de situer le contexte (situation, besoins, question posée), les grandes orientations de la démarche et les buts retenus ; matériel et méthode ; résultats ; discussion ; conclusion. Mais, d'une part, tous les auteurs n'ont pas la rigueur de respecter cet ordonnancement, d'autre part, certaines informations complémentaires peuvent être délivrées dans la discussion.

Les lecteurs réguliers de littérature scientifique utilisent des parties clés du document, comme le résumé, la conclusion, l'introduction dans un premier temps pour décider, à chacune de ces parties, s'ils vont poursuivre la lecture. Chacun adapte sa façon de faire au gré de son intérêt et de sa pratique. Nous voulons insister sur le fait que la

lecture doit aussi prendre en considération les repères que donne la méthode d'évaluation. De ce fait, une interrogation peut naître à n'importe quel endroit du document nécessitant alors de faire un saut, en arrière ou en avant, pour obtenir la réponse correspondante... ou constater que ce besoin n'est pas satisfait.

Par conséquent, l'ordre de lecture doit être adapté aux habitudes du lecteur, mais il doit avoir aussi le souci de passer en revue les points essentiels de la méthode d'évaluation pour parvenir au jugement final, ce qui peut l'amener à ne pas respecter scrupuleusement l'ordre de rédaction du document ou l'ordre indiqué dans la présentation de la méthode.

Mais il ne faut pas perdre de vue que la préoccupation première doit être tournée vers la satisfaction des besoins, des attentes, la réponse à la question posée. C'est pourquoi il n'est pas forcément utile d'entrer dans un degré de détail du protocole afin de *tout* scruter en détail, mais il faut en priorité repérer les éléments méthodologiques cruciaux qui font que cette satisfaction sera ou non atteinte. Et si un défaut rédhibitoire est repéré, il n'est plus besoin de poursuivre la lecture, sauf pour se conforter dans ce jugement.

2. PORTÉE ET LIMITES DE CETTE MÉTHODE D'ÉVALUATION

Cette méthode est destinée à des débutants en épidémiologie. Les messages se devaient d'être par

conséquent simples pour être utilisables facilement et rapidement, mais cette simplification peut aussi être considérée comme abusive⁴.

Par exemple, nous avons associé le type d'étude avec le nombre d'échantillons, un, au moins, pour les études descriptives et deux, au moins, pour les études explicatives. Cette simplification ne prend pas en compte les études de cohortes ne comportant qu'un seul échantillon, ainsi que les grandes bases de données, pour lesquelles des analyses statistiques sophistiquées sont disponibles, mais nécessitent des compétences statistiques approfondies pour leur mise en œuvre ainsi que pour l'évaluation de leur validité.

Ce choix pédagogique de simplification a été fait conformément à l'objectif pédagogique visé, mais aussi pour éviter que les débutants en déduisent l'idée fautive que, de toute façon, la statistique peut faire des miracles et compenser des défauts de protocole.

Les protocoles utilisés en pratique sont bien plus complexes que les archétypes présentés, pour pouvoir tenir compte de circonstances contraignantes qui peuvent rendre caduques les exigences simplistes que nous avons formulées. L'habileté critique des lecteurs pourra progresser par l'entraînement afin de s'affranchir de ces règles trop simples pour pouvoir aborder des cas plus difficiles.

Cette méthode est valide aussi bien pour évaluer un compte-rendu d'étude qu'un projet, même si cela n'a pas été systématiquement souligné.

VI - CONCLUSION

La méthode exposée donne des clés pour simplifier la complexité apparente de l'ensemble des connaissances nécessaires à mobiliser simultanément pour évaluer la qualité d'une étude épidémiologique, en vue de savoir si on peut accorder foi aux résultats et à leur interprétation, ou au contraire en douter, et jusqu'à quel point. Elle peut aussi être utilisée pour évaluer un projet d'étude. L'encadré 1 et le tableau 1 en résument les aspects essentiels de façon synthétique.. Elle s'appuie sur une série de jugements qualitatifs émis à chacune des étapes de la méthode, puis sur un

jugement global portant sur la validité scientifique de l'étude et le degré de satisfaction des attentes du demandeur.

L'articulation entre ces différents éléments sera mieux perçue par la résolution des problèmes que nous proposons en ligne sur le site de l'AEEMA. Il est certain que c'est l'entraînement sur ce type d'exercices qui permettra de renforcer cet apprentissage ce qui peut être fait de manière très régulière à la lecture de la littérature scientifique, voire même grand public.

⁴ « Ce qui est simple est toujours faux. Ce qui ne l'est pas est inutilisable » [Valéry, 1960, p. 864]

BIBLIOGRAPHIE

Ancelle T. - Statistique. Épidémiologie. Éditions Maloine, Paris, 2002, 300 p.

Bouyer J., Hémon D., Cordier S., Derriennic F., Stücker I., Stengel B., Clavel J. - Épidémiologie. Principes et méthodes quantitatives. Inserm Ed., Paris, 2009, 498 p.

Collin C. - Contribution à l'étude épidémiologique de la tuberculose bovine dans le département du Puy-de-Dôme. Thèse de doctorat vétérinaire, Alfort, 1987, 108 p.

Dohoo I., Martin W., Stryhn H. - Veterinary epidemiologic research. VER Ed, Charlottetown, 2009, 865 p.

Lemaitre P. - La décision, Éditions d'organisation, Paris, 1981, 176 p.

Toma B., Bénet J.J., Dufour B., Eloit M., Moutou F., Sanaa M. - Glossaire d'épidémiologie animale. Éditions du Point vétérinaire, Maisons-Alfort, 1991, 365 p.

Toma B., Dufour B., Bénet J.-J., Sanaa M., Shaw A., Moutou F. - Épidémiologie appliquée à la lutte collective contre les maladies transmissibles majeures. 3^{ème} éd., AEEMA Eds, Maisons-Alfort, 2010. 600 p.

Valéry P. - Œuvres II, Galimard, Paris, 1960, 1 726 p.

SITES INTERNET

AEEMA : <http://aeema.vet-alfort.fr/> ; rubriques « Annexes », « *Epidémiologie pour tous* », « Terminologie » pour les rappels de connaissances en épidémiologie ;

« *Epidémiologie pour tous* » pour les études de cas résolus sous forme d'enseignement programmé.

AFNOR : <http://www.definition-qualite.com>

EPI INFO : <https://www.cdc.gov/epiinfo/index.html>



ENCADRÉ 1

Étapes de l'évaluation d'une étude épidémiologique

PREMIÈRE ÉTAPE : RAPPEL DU PROBLÈME

Vérification de l'adéquation entre :

- La situation de santé, les besoins de santé,
- La question posée,
- Le but de l'étude.

DEUXIÈME ÉTAPE : ADÉQUATION AUX ATTENTES DU DEMANDEUR

Vérification de l'adéquation *a priori* avec :

- Attentes explicites/implicites,
- « pourquoi » ?, « pour quoi » ?

TROISIÈME ÉTAPE : ÉVALUATION DE LA VALIDITÉ INTERNE

- Identification du type d'étude :
 - Un groupe (ou plus), « *population, temps, espaces* » : **Descriptif**,
 - Deux groupes (au moins), **relation causale : explicatif**.
- Vérification du **critère majeur de qualité d'échantillonnage**
 - Descriptif : **Représentativité**,
 - Explicatif : **Comparabilité**.
- Mesures : **fiabilité**,
- Analyse et interprétation des **résultats**
 - Conformité aux usages statistiques,
 - Absence de biais ; inférence à population étudiée,
 - Réponse à la question posée.

QUATRIÈME ÉTAPE : ÉVALUATION DE LA VALIDITÉ EXTERNE

- Généralisation** à population cible, à population plus large.

CINQUIÈME ÉTAPE : BILAN GLOBAL

- Compte-rendu** :
 - Jugement sur la **fiabilité** des résultats,
 - Vérification *a posteriori* de l'**adéquation** aux attentes.
- Projet** : propositions d'**amélioration**

Tableau 1
Critères de qualité d'une étude épidémiologique

Critère de qualité	Étude DESCRIPTIVE	Étude EXPLICATIVE (ou analytique)
ADAPTATION (a priori) AUX BESOINS	Connaître la fréquence, pour prendre la décision d'opportunité d'action	Connaître les mécanismes, pour décider des modalités de lutte et de leur adaptation éventuelle
BUTS	Déterminer la distribution de la maladie dans une triade {Population, Temps, Espace}, son évolution Réduire l'incertitude	Établir une relation de cause à effet entre un (ou des) facteur(s) et une maladie Réduire l'incertitude
VALIDITÉ INTERNE	ABSENCE DE BIAIS	
ECHANTILLON	LA SÉLECTION DOIT ÊTRE MAÎTRISÉE PAR L'INVESTIGATEUR	
Nombre	<u>Un</u> échantillon peut suffire	Au moins DEUX échantillons sont nécessaires (pour les comparer)
Mode de sélection	Sélection à partir de la population source (reflétant une population cible)	Sélection sur la base de l' exposition à la maladie (exposé/non) ou de l' état par rapport à la maladie (atteint/non)
	« Autosélection » = risque de biais	
Critère de qualité principal	Représentativité (tirage au sort) Prévision de vérification de représentativité : caractéristiques démographiques	Comparabilité Pour maîtrise des facteurs de confusion (FC) <i>A priori</i> : appariement
Taille	Selon hypothèse de fréquence de la maladie dans la population	Selon l'écart entre les lots à mesurer <i>Risques alpha et bêta (puissance suffisante)</i>
MESURE	Mesures non biaisées Sensibilité, spécificité ; valeurs prédictives...	
	Concerne aussi bien les tests que les questionnaires, d'une façon générale le classement des individus dans la bonne catégorie	
	RESPECT DES CONDITIONS DE VALIDITÉ STATISTIQUE : TAILLE D'ÉCHANTILLON, INDÉPENDANCE DES INDIVIDUS. LE « SENS », LA RÉPONSE AUX QUESTIONS POSÉES.	
ANALYSE DES DONNÉES	Estimation : %, intervalle de confiance Seuil de signification fixé avant récolte des données Si comparaison, standardisation Inférence à la population source (étudiée)	« Règle des 3A » : 1. Association suffisante E+/E-* : Risque relatif (RR + Intervalle de confiance) KT/*- : Odds Ratio (OR + Intervalle de confiance) 2. <u>Absence de facteur de confusion</u> : RR ou OR ajusté 3. Antériorité
VALIDITÉ EXTERNE	Généralisation à population cible : Possible si <i>représentativité</i> de	
	population source à population cible	relation facteur/maladie dans population étudiée à celle de population cible.
Adaptation (a posteriori) aux besoins	Utilisation des résultats pour les décisions correspondant aux besoins exprimés.	

* E+/E- : enquête exposés/non exposés ; K/T- : enquête cas/témoins.

EXERCICES

Pour la résolution des problèmes suivants, vous devrez ultérieurement télécharger le support sur le site de l'AEEMA (<http://aeema.vet-alfort.fr/>). Deux exercices sont reproduits ici à titre d'exemple.

ENQUÊTE SUR LA PARAMPHISTOSE

D'après les constatations faites par les vétérinaires sur le terrain, la paramphistose est fréquente dans le département du Yadupour, mais trop peu d'éleveurs prennent des mesures de prévention.

C'est pourquoi le GDS (Groupement de défense sanitaire) de ce département a conduit une étude à l'abattoir de Nipourniconre, car la mise en place d'une étude auprès des élevages a été considérée comme hors d'atteinte des moyens disponibles : seuls les animaux en provenance du département ont été pris en considération, et la proportion des élevages atteints de paramphistose du département devait être estimé après redressement de l'échantillon grâce à l'indication de l'élevage de provenance des animaux. Pendant un mois (novembre), les 1 817 bovins du département abattus dans cet abattoir ont fait l'objet d'un examen de leurs estomacs : après nettoyage, l'examen des viscères a été réalisé par une étudiante de l'École vétérinaire dans le cadre de son stage en épidémiologie de façon à les classer visuellement selon cinq degrés d'infestation (de 0 à 4), en utilisant une grille inspirée de la grille INRA de détermination de l'état de propreté des bovins laitiers ayant servi pour une enquête portant sur l'hygiène en élevage.

Soixante-sept pour cent des bovins ont été trouvés infestés de paramphistomes. Les animaux de race à viande étaient deux fois plus souvent atteints que ceux de races laitières. Ce fort pourcentage d'atteinte a conduit le GDS à utiliser ces résultats dans le cadre d'une campagne de sensibilisation des éleveurs.

Évaluez la qualité de cette étude.



ÉTUDE D'UN ÉPISODE D'ACCIDENTS SURVENUS APRÈS VACCINATION DE CHATS

Le service commercial d'un laboratoire français distribuant en Grande-Bretagne un vaccin pour chats a reçu des plaintes téléphoniques de plusieurs vétérinaires britanniques : dans les jours suivant l'injection du vaccin, un certain nombre de chats avaient été ramenés chez leur vétérinaire. Ces chats présentaient de la fièvre, un abattement et de la diarrhée. Il s'agissait de jeunes chats, de race siamoise pour la plupart.

Le chef de produit a été d'autant plus alerté que la firme était en train de conquérir le marché britannique par une politique commerciale très différente de celle des autres firmes. Leur politique commerciale était centrée sur la qualité de la relation aux vétérinaires, grâce à un réseau d'agents commerciaux démarchant les vétérinaires praticiens, alors que les entreprises déjà implantées, dans une logique toute différente, s'appuyaient sur des plateformes de vente à distance. Il n'a pu obtenir aucune information auprès de la maison mère, en raison de la léthargie estivale d'un pont du 14 juillet.

Il a décidé, en urgence le mercredi, de réaliser une enquête afin de savoir si les chats vaccinés avec le vaccin de la marque avaient bien connu un taux d'accident supérieur à ceux des chats vaccinés avec d'autres vaccins, et s'il fallait ou non rappeler le vaccin suspect. Deux échantillons de 100 vétérinaires chacun ont été constitués par tirage au sort, l'un parmi le fichier des clients de la firme, l'autre à partir de l'annuaire national des vétérinaires dont les clients de la firme avaient été retirés : ce nombre était compatible avec le temps disponible et le nombre d'enquêteurs. Les commerciaux disposaient de 48 heures pour collecter l'information auprès d'un échantillon de vétérinaires : nombre de chats vaccinés au cours des 15 derniers jours, nombre de chats présentés à nouveau dans les jours suivants la vaccination, type de symptômes, race, âge et sexe des animaux.

Les données étant envoyées par voie électronique dès qu'elles étaient disponibles, les résultats ont été analysés en urgence le vendredi en début d'après-midi, en vue de prendre la décision d'envoyer ou non un avis d'alerte au Veterinary Record, pour 18 heures au plus tard. 92 vétérinaires ayant utilisé le vaccin suspect sur 937 chats ont pu être contactés, et 85 vétérinaires ayant utilisé un des vaccins de la concurrence sur 842 chats.

Ceux qui avaient utilisé le vaccin suspect ont constaté 6 % de troubles divers ayant conduit le propriétaire à ramener l'animal chez le vétérinaire dans les jours suivant la vaccination, tandis que ce taux était de 3,8 % chez ceux qui ont utilisé le vaccin B (Chi 2, $p = 0,035$).

Si vous aviez été à la place du chef marketing, qu'auriez-vous décidé, sachant que toute mauvaise décision a pour sanction le licenciement pour faute grave ?

