STV

Note méthodologique

Sang Thrombose Vaisseaux 2021 ;
33, n° 1 : 7-17

Épidémiologie étiologique : du facteur de risque à la cause*

Brigitte Ranque[1], Laura Mechtouff[2], Sophie Grabar[3]

[1] Université Paris Descartes, AP-HP, Service de médecine interne, Hôpital européen Georges Pompidou, 20 rue Leblanc, 75015 Paris, France
<brigitte.ranque@club-internet.fr>
[2] Service de neurologie vasculaire, Hospices civils de Lyon, Hôpital Pierre Wertheimer, Bron-Lyon, France
[3] Université Paris Descartes, AP-HP, Hôpital Cochin, Unité de biostatistique et épidémiologie et INSERM U943, Paris, France

La série d'articles coordonnée par Emmanuel Touzé, publiée en 2011, avait soulevé l'enthousiasme de nos lecteurs. Il nous a donc paru intéressant de proposer à notre lectorat une actualisation de ces articles, auxquels s'ajouteront de nouveaux articles. Emmanuel Touzé s'est impliqué à nouveau dans cette deuxième édition, qui, nous l'espérons, continuera à passionner notre lectorat.

Résumé. L'épidémiologie étiologique est la discipline qui étudie les facteurs responsables des maladies (ou facteurs causaux). Pour des raisons éthiques et logistiques, les études étiologiques sont principalement des études observationnelles de type cas/témoins ou cohortes exposés/non exposés. Elles présentent de ce fait de nombreuses limites méthodologiques, qu'il faut avoir à l'esprit pour être en mesure d'évaluer leur validité. Il faudra, en effet, s'assurer de la pertinence du choix des populations étudiées, de la qualité de l'exploration des biais, de l'objectivité, de l'exhaustivité et de la précision du recueil des informations concernant les expositions et les maladies étudiées, et de l'adéquation des analyses statistiques. D'une façon générale, le lecteur se doit de toujours rester prudent dans son interprétation des résultats. En particulier, une association statistique valide entre un facteur d'exposition et une maladie ne permet pas en soi de garantir l'existence d'un lien de causalité entre ce facteur et cette maladie : plusieurs autres conditions doivent être présentes pour pouvoir évoquer ce lien, comme une séquence temporelle compatible, une relation dose-effet, une plausibilité biologique ou des études expérimentales et épidémiologiques compatibles.

Mots clés : épidémiologie, étiologie, cas/témoins, cohorte, biais, causalité

Abstract

Etiological epidemiology: from risk factors to causes

Etiological epidemiology is a discipline which studies the causal factors of disease. For ethical and logistical reasons, most etiological studies are observational case/control studies or exposed/non exposed cohorts. Therefore, they share many methodological limits which have to be taken into consideration to evaluate their validity. The choice of the study population sample, the quality of bias exploration, the objectivity, completeness and precision of both disease and exposition data collection and the quality of statistical analyses must be verified. Overall, the reader needs to interpret results cautiously. In particular, a valid statistical association between exposure to a risk factor and the onset of disease does not guarantee the presence of a causality link between this factor and this disease. Several other conditions have to be reviewed before suggesting a causality link, such as a consistent temporal sequence, dose effect relationship, biological plausibility and consistent experimental and epidemiological studies.

Key words: epidemiology, etiology, causality, bias, cohort, case-control

Tirés à part :
B. Ranque

* Actualisation de l'article paru dans *STV* en 2011 : Ranque B, Mechtouff L, Grabar S. Épidémiologie étiologique : du facteur de risque à la cause. *Sang Thrombose Vaisseaux* 2011 ; 23 (5) : 242-52. doi:10.1684/stv.2011.0607

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

doi:10.1684/stv.2020.1150

L'épidémiologie étiologique est la discipline qui cherche à mettre en évidence et à quantifier l'effet de facteurs responsables d'une maladie (ou facteurs causaux). Le terme de « facteur de risque », qui sous-entend un lien de causalité entre le facteur et la maladie, doit être interprété avec prudence. Il s'agit, en effet, d'un abus de langage, car le facteur étudié n'est le plus souvent qu'un « marqueur » de risque, qui est associé à la survenue de la maladie mais n'en est pas responsable.

De nombreux schémas d'études sont théoriquement envisageables pour rechercher la cause d'une maladie, mais seule une étude dans laquelle l'exposition à un facteur potentiellement causal serait tirée au sort permettrait d'établir un lien de causalité entre le facteur et la maladie observée. Toutefois, ce type d'étude paraît, dans la grande majorité des cas, impossible à réaliser ou éthiquement non acceptable. De ce fait, les études étiologiques sont principalement des études observationnelles de cohortes exposés/non exposés ou cas/témoins, ne permettant pas l'attribution d'un fort niveau de preuve. Elles présentent de nombreuses limites intrinsèques, qu'il faut avoir à l'esprit pour être en mesure d'évaluer leur validité.

L'objectif de cet article est de présenter les différents types d'études, de donner les outils permettant d'interpréter leurs résultats, et de présenter les arguments en faveur d'une relation causale lorsqu'une association entre un facteur et une maladie est mise en évidence.

Les différents types d'études étiologiques [1-3]

Études de cohorte exposés/non exposés (*figure 1*)

Leur principe est de comparer l'incidence de la maladie chez des individus exposés au facteur de risque supposé et chez des individus non exposés. Les études de cohorte exposés/non exposés sont le plus souvent observationnelles. Les individus sont issus de la population générale (si le facteur d'exposition est fréquent) ou sélectionnés selon leur niveau d'exposition (si le facteur d'exposition est rare), et suivis prospectivement. Le critère de jugement est l'incidence de la maladie étudiée, mesurée dans les deux groupes. Les résultats sont présentés en termes de risque relatif (*Relative Risk,* RR) ou de rapport de risque instantané (*Hazard Ratio,* HR). Les principaux avantages de ce type d'étude sont de pouvoir décrire la séquence exposition/maladie, mesurer l'exposition de façon fiable, estimer l'incidence de la maladie et étudier l'effet d'une exposition sur plusieurs maladies. Toutefois, elles nécessitent un long suivi (plusieurs années sont généralement nécessaires) et une logistique complexe et coûteuse.

La plus célèbre étude étiologique de cohorte est la *Framingham Heart Study*, débutée en 1948. Initialement, 5 209 habitants de Framingham, une petite ville située près de Boston, ont été recrutés. Ces individus ont eu un suivi médical et paramédical systématique tous les deux ans. Environ 5 000 individus supplémentaires ont été recrutés en 1971 puis en 2001 et l'étude se poursuit toujours avec environ 14 000 individus actuellement suivis [4]. Cette cohorte a permis d'identifier les principaux facteurs de risque vasculaires actuellement connus (tabagisme, hypercholestérolémie, obésité, absence d'exercice, etc.).

Études cas-témoins (*figure 2*)

Le principe des études cas-témoins (*Case-Control,* CC) est de comparer la fréquence d'exposition au facteur de risque chez les patients atteints de la maladie et chez les individus non atteints. La sélection des individus se fait en fonction de leur statut malade/sain et on recherche rétrospectivement une exposition au facteur de risque étudié. Les résultats sont exposés en termes de rapports de cotes (*Odds Ratio,* OR). Ce type d'étude est le seul envisageable si la maladie étudiée est très rare ou avec un délai de survenue très long. Par rapport aux études de cohortes, les principaux avantages des études cas-témoins sont le gain de temps et le faible coût, mais, comme nous le verrons plus loin, ce type d'étude est le plus à risque de biais de sélection et d'information [5].

Par exemple, l'étude MEGA a étudié le rôle des contraceptifs oraux de 3^{e} génération dans la survenue de la maladie thromboembolique veineuse [6]. Cette étude néerlandaise a inclus 1 524 femmes ayant une phlébite ou une embolie pulmonaire et 1 760 témoins tirés au sort dans la population locale et a comparé dans les deux populations la fréquence d'utilisation d'une contraception orale. Un risque accru de maladie thromboembolique veineuse a été observé quelle que soit la contraception orale utilisée, avec un risque croissant avec la dose d'œstrogène et qui variait selon le type de progestatif (risque moindre en cas de lévonorgestrel).

Une variante de plus en plus utilisée est l'étude cas-témoins nichée dans une étude de cohorte (*Nested Case-Control,* NCC), avec une mesure de l'exposition *a posteriori* (mais qui peut se faire sur des échantillons collectés prospectivement). Les patients sont issus d'une même cohorte, dans laquelle la maladie a été dépistée de façon prospective. Ce type d'étude combine les avantages d'une cohorte (mesure fiable de l'exposition, meilleure représentativité des témoins car ils sont issus de la même cohorte) et ceux d'une étude cas-témoins : gain de temps et

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

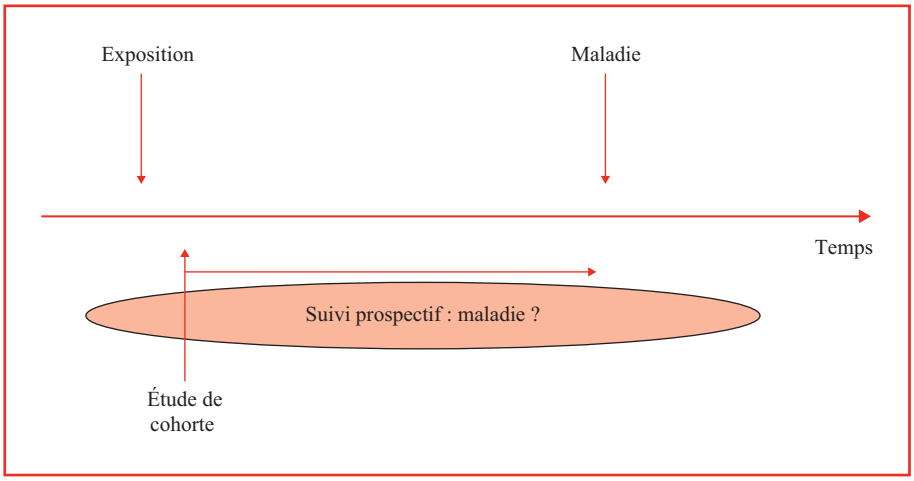

Figure 1. Étude de cohorte exposés/non exposés.

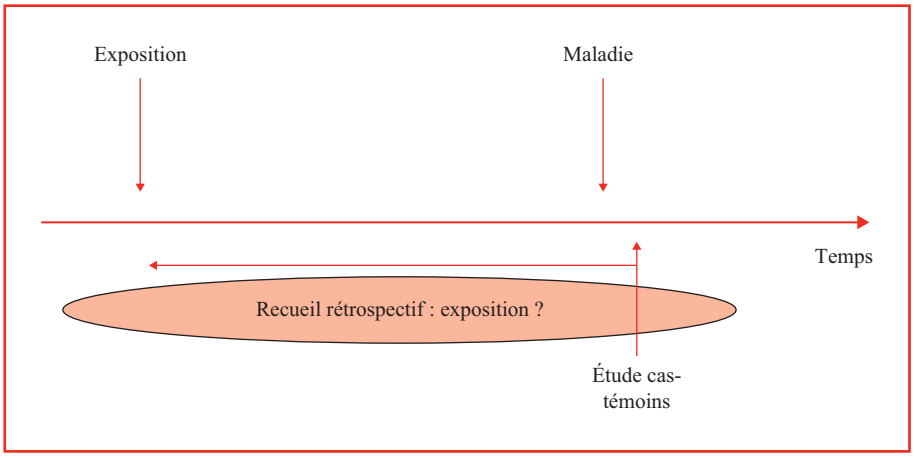

Figure 2. Étude cas/témoins.

réduction du coût, puisque le suivi des individus a déjà eu lieu et qu'il est possible de ne mesurer l'exposition que chez un sous-groupe de patients (par exemple, si la maladie est rare, chez tous les cas et un échantillon des témoins).

Dans une étude sur l'effet de polymorphismes du gène de la CRP (protéine C-réactive) (modifiant le taux de CRP ultrasensible), sur les maladies ischémiques artérielles [7], les auteurs ont repris trois grandes cohortes de plusieurs milliers d'individus, toutes menées à Copenhague. Les événements cardiovasculaires étaient recensés de façon prospective. Les auteurs ont *a posteriori* génotypé les patients pour les polymorphismes du gène de la CRP et mesuré les taux initiaux de CRPus. Ils ont ensuite comparé ces valeurs chez les patients avec et sans événement cardiovasculaire survenu au terme du suivi. L'étude confirme que la CRPus est un bon marqueur prédictif de risque cardiovasculaire mais que le fait d'être porteur de polymorphismes du gène de la CRP induisant une CRPus plus élevée ne modifie pas ce risque (voir commentaire sur la randomisation mendélienne dans l'*encadré 3*).

Les avantages et inconvénients respectifs des études cas/témoins et des cohortes exposés/non exposés sont présentés dans le *tableau 1*.

La méthode cas-croisés (*figure 3*)

Cette approche est basée sur le modèle d'une étude cas-témoin où chaque cas est son propre témoin. Le principe est de comparer l'exposition d'un sujet à un facteur dans un délai proche de la survenue de la maladie (période cas) à l'exposition de ce même sujet au cours d'une (ou plusieurs) période(s) à distance de la maladie (période(s) témoin(s)) [8]. Cette méthode a été notamment appliquée à l'analyse des effets à court terme de facteurs environnementaux tels que la pollution atmosphérique, représentant ainsi

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

Tableau 1. Caractéristiques comparées des études de cohorte exposés/non exposés et des études cas/témoins.

Type	Cohorte	Cas-témoins
Mode de suivi	Prospectif	Rétrospectif
Sélection des patients	Selon l'exposition ou non au facteur de risque	Selon le statut vis-à-vis de la maladie (atteint ou non)
Critère de jugement	Taux d'incidence de la maladie	Prévalence de l'exposition
Indice d'effet	Risque relatif, risque attribuable, *hazard ratio*	*Odds ratio*
Avantages	- Calcul du taux d'incidence, du risque relatif et du risque attribuable - Description de la séquence exposition-maladie - Étude possible de plusieurs maladies et de facteurs d'exposition rares en même temps	- Étude possible des maladies rares et de plusieurs facteurs de risque en même temps - Courte durée d'étude - Moindre coût - Échantillons plus petits - Pas de perdus de vue - Pas de problème éthique
Inconvénients	- Grande taille d'échantillon - Ne permet pas d'étudier facilement les maladies rares - Longue durée d'enquête - Logistique coûteuse et complexe - Les patients peuvent être perdus de vue ou changer de statut exposés/non exposés - Problèmes éthiques	- Pas de calcul possible du taux d'incidence absolue - Problème de comparabilité des témoins et des cas - Biais de sélection (biais de survie, choix des témoins difficile) - Importants biais de mesure (mémorisation, données manquantes)

une méthode alternative aux séries temporelles[1]. Chaque cas étant son propre témoin, les facteurs de confusion individuels ne variant pas à court terme sont ainsi pris en compte. Le choix des périodes témoins est déterminant dans le contrôle des facteurs de confusion temporels. Par exemple, sachant qu'il peut exister un effet « jour de la semaine » ou de la saison, il vaut mieux choisir une période témoin le même jour et le même mois que la période cas.

Principales difficultés méthodologiques

Quel que soit le type d'étude observationnelle, de nombreux écueils méthodologiques peuvent altérer la validité des résultats, en produisant de faux positifs (association factice ou surestimée) ou de faux négatifs (sous-estimation ou négation de l'effet réel d'un facteur d'exposition). Ces résultats invalides résultent majoritairement d'erreurs systématiques, c'est-à-dire non aléatoires, que l'on appelle « biais ». Les biais sont classiquement regroupés en trois types : sélection (ou recrutement), classification (ou mesure ou information) et confusion. L'*encadré 1* donne une définition et quelques exemples pour chacun de ces biais. Le biais sera différentiel lorsque l'erreur survient différemment (plus ou moins souvent) dans les deux groupes d'individus étudiés (cas contre témoins ou exposés contre non exposés). Il modifie alors la mesure de l'association entre la maladie et le facteur de risque en l'augmentant ou en la diminuant. Le biais est non différentiel si l'erreur touche également les deux groupes, il induira alors une sous-estimation de l'association par imprécision (manque de puissance).

Lors de la conception d'une étude étiologique, tout doit être mis en œuvre pour tenter d'éviter ces biais ou de les contrôler : les deux étapes les plus cruciales seront le choix des populations étudiées et celui des méthodes de mesure de l'exposition et de la maladie.

[1] La série temporelle consiste à réaliser une mesure (par exemple, un dosage biologique) avant et après une intervention (par exemple, l'administration d'une substance) pour chaque individu et à comparer ces deux mesures pour estimer l'effet de l'intervention. Ce type d'étude a de nombreuses limites, dont l'absence de comparaison avec l'évolution « naturelle » de la mesure (notamment, régression vers la moyenne et tendance saisonnière). Elle est rarement utilisée pour les études étiologiques, pour les mêmes raisons éthiques que les essais randomisés.

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

Encadré 1. Les grands types de biais

Sélection (recrutement)
= non représentativité des cas et/ou des témoins (ou des exposés/non exposés).
Exemple 1
Étude cas-témoins hospitalière sur l'embolie pulmonaire: les patients les plus graves seront morts avant d'être inclus et les facteurs favorisant les embolies sévères ne seront pas étudiés (*biais de survie*).
Exemple 2
Étude cas-témoins sur les tumeurs gastriques : si les cas et les témoins sont recrutés dans un service de gastro-entérologie, les témoins auront des pathologies digestives (ulcère gastrique, Crohn, etc.) leur interdisant de prendre des anti-inflammatoires non stéroïdiens (AINS). Ils y seront donc moins exposés que la population générale, ce qui entraînera une surestimation de l'effet des AINS sur le risque de tumeur gastrique (*biais de recrutement).*
Exemple 3
Étude cas-témoins sur le risque de sclérodermie étudiant l'exposition à un produit d'entretien, où les témoins sont sélectionnés par téléphone. Cette méthode risque de sélectionner beaucoup de femmes au foyer, plus exposées à ce type de produit que les femmes travaillant hors foyer. L'effet du produit sur le risque de sclérodermie sera alors sous-estimé.

Classification (mesure/information)
= recherche de l'exposition ou de la maladie mise en œuvre différemment selon le groupe d'étude (cas contre témoins ou exposés contre non exposés).
Exemple 1
Étude cas-témoins d'une exposition alimentaire dans la survenue d'un cancer digestif, l'exposition étant recherchée par un autoquestionnaire. Les malades auront tendance à mieux se rappeler avoir consommé l'aliment en question, car ils auront déjà réfléchi aux causes externes possibles de leur cancer, ce qui aura pour effet de surestimer le risque lié à l'exposition (*biais de mémorisation*).
Exemple 2
Étude de cohorte étudiant le risque de cancer du poumon lié à l'exposition à l'amiante, où les exposés sont suivis de façon plus rapprochée et avec des méthodes plus sensibles que les non exposés (par exemple scanner contre radiographie). Ici les cancers survenant chez les témoins seront moins bien dépistés, et le risque lié à l'exposition sera surestimé (*biais de mesure*).
Exemple 3
Étude cas-témoins sur dossiers hospitaliers de l'effet la possession d'un animal domestique sur le risque d'asthme. Le mode de vie sera mieux renseigné dans les dossiers des patients asthmatiques que chez les autres, donc les données manquantes le seront de façon non aléatoire *(biais d'information*).

Confusion (méprise)
= présence d'un tiers facteur influençant à la fois l'exposition et la maladie, mais sans appartenir à la chaîne de causalité de l'exposition.
Exemple 1
Étude de cohorte sur le risque de cancer du poumon et l'exposition au charbon. Les exposés sont des mineurs ayant un tabagisme plus important que la population non exposée. Le tabagisme étant un facteur de risque important de cancer du poumon, on observerait en analyse univariée une association entre l'exposition au charbon et le cancer, même en l'absence d'effet causal du charbon.
Exemple 2
Étude cas-témoins sur le risque d'infarctus lié à la prise d'œstroprogestatifs. Les femmes prenant un contraceptif oral sont plus souvent fumeuses que les autres. Par ailleurs, le tabagisme est un facteur de risque d'infarctus, on retrouve donc une association entre la prise d'œstroprogestatifs et le risque d'infarctus.
NB: Ce type de biais peut être prévenu par la stratification préalable en fonction du facteur de confusion ou corrigé par l'analyse statistique multivariée en ajustant sur ce facteur.

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

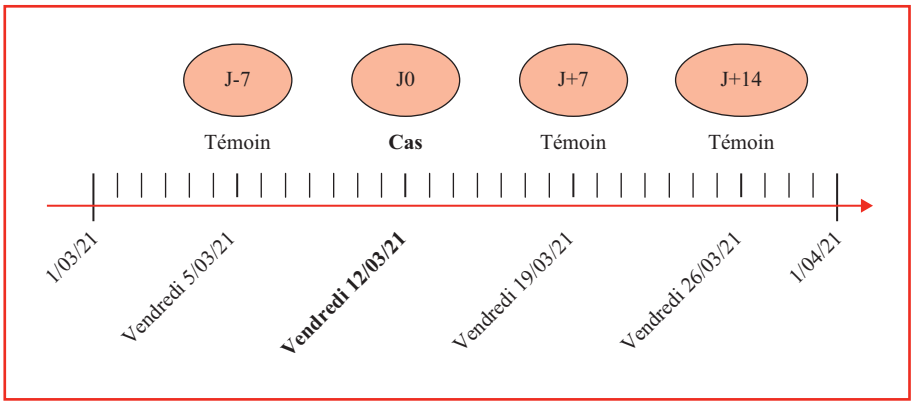

Figure 3. Étude cas-croisés. Approche stratifiée sur le temps (ici un mois).

Choix des populations étudiées

Pour être valide, l'association entre un facteur et une maladie implique que les individus malades et non malades ne diffèrent que par leur exposition au facteur étudié. Ce problème de comparabilité des deux groupes est fondamental, comme dans toute étude comparative, et constitue une des plus grandes difficultés des études étiologiques. En l'absence de possibilité de tirage au sort, le choix des individus exposés et non exposés (étude de cohorte), ou des malades et des non-malades (étude cas-témoins), doit être extrêmement réfléchi. Il peut en effet facilement engendrer deux types de biais : biais de sélection et biais de confusion *(encadré 1)*.

Études cas-témoins

Pour limiter les biais de confusion, une solution est de sélectionner les cas et les témoins dans un environnement géographique, professionnel et/ou familial commun. Toutefois, les résultats pourraient alors ne pas être généralisables à d'autres populations et cette stratégie peut entraîner un biais dans l'exposition à d'autres facteurs de risque que celui étudié (biais de sélection ou recrutement). Par exemple, pour une étude concernant l'association entre le cancer pulmonaire et le tabagisme, si les cas et les témoins sont des patients hospitalisés dans le même service de pneumologie d'un hôpital, les témoins auront probablement une exposition au tabac plus importante que celle de la population générale, car le tabagisme est un facteur de risque de nombreuses autres maladies pulmonaires, ce qui induira une sous-estimation de l'association entre tabagisme et cancer pulmonaire.

L'alternative est de rechercher des témoins parmi la population générale dont sont issus les cas. Par exemple, dans l'étude MEGA, les témoins ont été sélectionnés par tirage au sort de numéros de téléphone dans la région et appariés selon le sexe et l'âge aux cas recrutés au préalable dans une clinique d'anticoagulant [6]. Les non-réponses peuvent cependant constituer un biais de sélection si le fait de ne pas répondre est lié à la maladie ou à l'exposition étudiée. Cette méthode a en outre l'inconvénient d'être très coûteuse.

Un problème de sélection particulier aux études cas-témoins est le biais de survie : lorsque la maladie étudiée est potentiellement rapidement létale, les patients les plus graves ne sont jamais inclus car ils meurent avant d'être identifiés. Une solution permettant d'éviter ce biais de survie et d'avoir des cas et des témoins plus comparables est la réalisation d'une étude cas/témoins nichée dans une étude de cohorte réalisée pour autre objectif, car les patients ont été recrutés indépendamment de la maladie (qui est survenue plus tard dans le suivi).

Études de cohorte

Il est souhaitable d'utiliser une population non exposée la plus proche possible des patients exposés, pour que seule l'exposition varie. Par exemple, pour étudier l'exposition à un toxique, on peut inclure tous les ouvriers travaillant dans une même usine utilisant ce toxique. Toutefois, il peut être alors difficile de recruter suffisamment de personnes non exposées. Il est alors possible d'utiliser des individus contrôles issus de la même population que les exposés (voisins, famille).

Dans certains cas, c'est la population nationale ou régionale qui est utilisée comme contrôle (groupe non exposé), lorsque des registres contenant les informations d'incidence de la maladie étudiée sont disponibles. Cela est notamment le cas lorsqu'il s'agit d'une maladie à déclaration obligatoire, mais il existe un fort risque de sous-déclaration et donc de biais de classification (ou information).

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

Enfin, un problème particulier des études prospectives est le nombre de perdus de vue car le fait d'être perdu de vue peut être lié à la maladie et/ou à l'exposition. Le nombre de perdus de vue doit être le plus faible possible pour ne pas diminuer la représentativité initiale des individus. Une étude comparative des caractéristiques des patients perdus de vue et de ceux qui ont achevé leur suivi est souhaitable pour évaluer le risque de biais de sélection.

Mesure de l'exposition et diagnostic de la maladie

À l'instar de toute autre étude clinique, les critères de jugement doivent être définis précisément et leur mesure doit être reproductible pour éviter les biais de classification (ou information). Il est donc indispensable que, dans une étude de cohorte (respectivement une étude cas-témoins), la maladie (respectivement l'exposition) soit évaluée de la même façon chez les exposés et les non-exposés (respectivement les cas et les témoins). En pratique, l'investigateur d'une cohorte sachant qu'un patient est exposé à un facteur potentiellement dangereux aura tendance à rechercher plus souvent et de façon plus approfondie la maladie étudiée et donc à la diagnostiquer par excès chez les patients exposés. De ce fait, le meilleur moyen d'obtenir une égalité de mesure est que celle-ci soit réalisée par un investigateur en aveugle du statut exposé/non exposé pour le diagnostic d'une maladie ou du statut malade/non malade pour la mesure d'une exposition.

Dans le cadre d'une cohorte, le suivi doit être suffisamment long pour que la maladie puisse éventuellement se développer, ce qui suppose d'avoir une idée du temps de latence de survenue de la maladie après l'exposition. Un temps de suivi trop court induira un biais - pas forcément différentiel - de classification (certains individus considérés comme sains au terme de l'étude développeront en fait la maladie plus tard) et un manque de puissance (peu de cas). Pour diminuer ce temps de latence, qui peut être de plusieurs dizaines d'années, on est souvent amené à utiliser des marqueurs intermédiaires, dont le caractère prédictif de la maladie doit être bien établi. Par exemple, pour étudier les facteurs de risque vasculaire durant l'enfance, les auteurs de l'étude [9] ont retenu comme critère de jugement l'épaisseur de l'intima media artérielle, qui est validée comme un bon marqueur prédictif d'événement cardiovasculaire, sachant que les éventuels infarctus du myocarde ou accidents ischémiques cérébraux n'auraient lieu que plusieurs dizaines d'années plus tard.

Dans le cas des études cas-témoins, il existe souvent un biais de mesure lié au caractère rétrospectif (informations anciennes plus souvent incomplètes ou mesurées avec des méthodes différentes), qui ne peut être contrôlé. Notamment, lorsque la caractérisation de l'exposition repose sur l'interrogatoire des individus, le biais de mémorisation est extrêmement fréquent, et le plus souvent différentiel. En effet, un malade se souviendra plus facilement d'avoir été exposé à un facteur de risque, car il s'est déjà interrogé sur l'origine de sa maladie et est désireux d'en identifier une cause extérieure. Un moyen d'atténuer ce biais est d'utiliser des aide-mémoires (photos, calendrier, etc.) et de questionner les patients sur d'autres expositions n'étant pas censées avoir de lien avec la maladie.

Méthodes statistiques

La mesure du risque conféré par le facteur de risque potentiel diffère en fonction du type d'étude.

Dans le cas des études de cohorte, la mesure du risque lié à l'exposition se fait en termes de risque relatif (RR) et doit être accompagné de son intervalle de confiance à 95 %. Son calcul est rappelé dans la *figure 4*. L'impact pratique du RR lié à l'exposition dépendra de l'incidence de la maladie dans la population témoin. Par exemple, un facteur qui est associé avec un RR élevé à une maladie extrêmement rare, ne sera en réalité responsable que de peu de cas. Une façon d'intégrer les deux informations (RR et risque de base) en un seul indice est de calculer le risque attribuable chez les exposés (RA), qui estime la réduction du risque de maladie que l'on observerait chez les individus exposés si ces individus n'étaient pas exposés. Il s'agit de la différence entre le risque chez les exposés et le risque chez les non exposés (*figure 4*), il ne dépend donc pas de la fréquence de l'exposition mais de celle de la maladie.

Un autre indice est le calcul du NNH (*Number Needed to Harm*), qui s'interprète exactement comme le NNT (*Number Needed to Treat*) des essais thérapeutiques. Il s'agit du nombre d'individus qui doivent être exposés sur une période donnée pour que l'un d'eux développe la maladie alors qu'il ne l'aurait pas développée sans l'exposition. Mathématiquement, il s'agit de l'inverse du risque attribuable chez les exposés (1/RA).

Dans les études cas-témoins, l'incidence de la maladie ne peut être estimée car le nombre de malades et de non-malades est fixé d'avance. On présente alors le risque en termes d'*odds ratio* (OR) ou rapport de cotes. Son calcul est rappelé dans la *figure 4*. Lorsque la maladie est rare (prévalence < 5 %), l'OR est une bonne approximation du RR. Sous réserve d'avoir de bons arguments pour penser que la relation entre l'exposition et la maladie est causale, l'impact du risque lié à l'exposition au niveau de la population peut

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

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

Risque absolu (études de cohorte) : R+ = a/(a+b) , R-= c/(c+d)
Risque relatif : RR= R+/R- = a (c+d)/c (a+b)
Risque attribuable chez les exposés : RA= R+ - R- = a/(a+b) – c/(c+d)
Odds ratio : OR= ad/bc

Figure 4. Tableau de contingence et calcul des indices d'effet (voir aussi la note sur les mesures de risque [4]).

être apprécié par la proportion de risque attribuable (PCA) qui dépend du RR et de la fréquence de l'exposition. Il s'agit du risque attribuable dans la population (risque chez les exposés - risque dans la population) divisé par le risque dans la population. Il peut également se calculer à partir du RR et de la fréquence de l'exposition Pe dans la population, comme suit : PCA = Pe (RR-1)/(1 + Pe (RR-1)).

Dans tous les cas, une analyse multivariée doit être réalisée, pour prendre en compte au mieux les biais de confusion. Elle permettra de donner une estimation du risque propre au facteur étudié, après ajustement sur les facteurs déjà connus pour influencer le risque de maladie. L'analyse se fait généralement par le biais d'une régression logistique dans le cas des études cas-témoins. Dans le cadre d'une étude de cohorte, une méthode de prise en compte des données censurées doit être utilisée, par exemple un modèle de Cox.

Le résultat est alors présenté en termes de rapport de risque instantané ou *hazard ratio* (HR) (se référer aux notes méthodologiques spécifiques pour plus de détail) [4]. Il est important de noter que l'analyse multivariée ne peut corriger que les biais de confusion liés aux facteurs de risque déjà connus (et dont les données ont été recueillies) et ne saurait donc totalement remplacer un tirage au sort. De plus, elle ne peut en aucun cas corriger les biais de sélection et de mesure.

Interprétation des résultats [2, 3, 10-12]

Évaluation de la validité interne et externe de l'étude

Rappelons qu'il existe des recommandations de publications propres aux études observationnelles, baptisées STROBE (*STrengthening the Reporting of OBservational studies in Epidemiology*) [15]. Ces recommandations ont été élaborées par un comité de méthodologistes avant tout afin d'améliorer la qualité des publications d'études observationnelles, mais elles permettent également de disposer des informations nécessaires à l'évaluation de la validité de ce type d'étude.

En accord avec ces recommandations, nous proposons, pour évaluer la validité des résultats d'une étude étiologique, de contrôler systématiquement les items ci-après.

Force de l'association et significativité statistique

La force de l'association est évaluée par la valeur du RR ou de l'OR. Un RR supérieur à 3 peut être considéré comme très fort (par exemple, le tabagisme confère un RR de 3 à 17 pour le cancer du poumon et de 2 à 3 pour le cancer de la vessie). Pour être valide, la force de l'association doit s'assortir d'une certaine précision : en pratique, le résultat sera statistiquement significatif si l'intervalle de confiance à 95 % de l'OR ou du RR exclut 1 (ce qui est équivalent à une *p value* < 0,05). En cas de résultat non significatif mais avec un RR (ou OR) élevé, il faut se demander si la taille de l'échantillon est suffisante pour garantir une bonne puissance à l'étude. Si l'étude a une faible puissance, il s'agit possiblement d'un résultat faussement négatif.

Exploration des biais potentiels (encadré 1)

– Sélection/recrutement : les cas et les témoins (ou les exposés et les non exposés) sont-ils similaires en dehors de leur exposition au facteur testé (ou de la maladie étudiée) ?

– Mesure/classification/information : la mesure de l'exposition et le diagnostic de la maladie ont-ils été réalisés

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

selon des méthodes fiables et de façon identique dans les deux groupes d'individus (notamment en aveugle) ?
– Confusion : tous les facteurs de risques connus ont-ils été mesurés et pris en compte dans une analyse stratifiée ou multivariée ? Rappelons que cela n'exclut pas l'effet de facteurs de risque encore inconnus.

Validité externe

Comme pour toute autre étude, les résultats seront directement applicables à nos patients si la population de l'étude est représentative de ces patients. D'autre part, le fait que les associations mises en évidence aient déjà été trouvées dans la littérature, *a fortiori* par d'autres groupes d'investigateurs, avec d'autres méthodes et dans des populations différentes, constitue un bon argument de validité externe.

Discussion du lien de causalité [10-12]

Un risque relatif supérieur à 1 de façon significative est synonyme d'association statistique entre un marqueur de risque R et la maladie M. Cette association peut correspondre à trois cas de figure : soit une causalité (R cause M), soit une causalité inverse (M cause R), soit une confusion (un facteur X cause R et M). Seule l'expérimentation pourrait permettre de façon formelle d'établir la causalité entre un marqueur de risque et une maladie. En cas d'étude observationnelle, il faudra disposer d'un faisceau d'arguments avant de conclure à une probable causalité. Une liste de ces arguments a été établie par Bradford Hill dès 1965 [14] (*cf encadré 2*) et constitue encore la référence dans le domaine. On peut les regrouper en arguments internes et externes à l'étude.
– Arguments internes : force de l'association (risque > 3 ou < 0,3) et degré de significativité, temporalité compatible (l'exposition doit forcément précéder la maladie), existence d'une relation dose/effet (augmentation du RR lorsque la quantité d'exposition augmente), réversibilité du risque à l'arrêt de l'exposition, récidive à la reprise de l'exposition.
– Arguments externes : cohérence avec la littérature épidémiologique (résultats similaires dans d'autres études épidémiologiques) et/ou fondamentale (modèles expérimentaux animaux), mécanisme physiopathologique compatible. Il faut toutefois se méfier particulièrement des hypothèses physiopathologiques formulées au vu des résultats, car, en raison de la complexité de la physiologie humaine, il est presque toujours possible de trouver une justification physiopathologique *a posteriori* à une association, quel que soit le sens du résultat...

Encadré 2. Critères d'Austin Bradford Hill (1965)

1. Preuve expérimentale : y a-t-il une preuve provenant d'une véritable expérience scientifique chez l'homme ?
2. Force de l'association : quelle est la force de l'association entre le facteur de risque putatif et la maladie étudiée ? (risque relatif > 3-4)
3. Concordance : est-ce que les résultats ont été répliqués dans d'autres études, différentes populations, par des investigateurs différents et sous des conditions différentes ?
4. Temporalité : l'exposition précède-t-elle bien la maladie ?
5. Gradient biologique : une exposition plus importante (en dose et/ou en durée) est-elle associée à un risque accru de maladie ?
6. Cohérence avec l'histoire naturelle et l'épidémiologie de la maladie (gradient géographique entre le niveau d'exposition et la prévalence de la maladie) ?
7. Spécificité : l'exposition est-elle associée à une maladie spécifique plutôt qu'à un grand nombre de maladies ?
8. Plausibilité : y a-t-il un mécanisme crédible biologique ou physique pouvant expliquer l'association ? Des expérimentations animales ont- elles permis de faire développer la maladie à des animaux exposés ?
9. Analogie : existe-t-il une relation causale connue entre une cause similaire et une maladie proche ?

– Noter que le critère de spécificité (un facteur d'exposition donne un seul type de maladie) n'est plus utilisé car il est rarement observé, en dehors de certaines maladies infectieuses. Par exemple, le tabagisme est un facteur de risque prouvé de cancer du poumon, et pourtant il prédispose aussi à de nombreuses autres maladies (cancers oropharyngés ou de vessie, bronchite chronique obstructive, maladies cardiovasculaires, etc.).
Un autre type d'argument a été utilisé plus récemment pour différencier causalité vraie de causalité inverse ou d'un biais de confusion, baptisé randomisation mendélienne [14]. Le principe est de regarder si les patients porteurs d'un génotype les prédisposant à être davantage exposés à un marqueur de risque (par exemple, mutation du gène *apoB* conférant une hypercholestérolémie familiale) ont effectivement un risque accru de maladie. L'idée est que l'attribution des génotypes se fait au hasard lors de

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

Encadré 3. Randomisation mendélienne

Lors de la fabrication des gamètes, l'attribution du génotype à un locus, c'est-à-dire la combinaison des deux allèles d'un gène, se fait au hasard sur quatre combinaisons possibles (un allèle tiré au sort parmi les deux du père et l'autre allèle tiré au sort parmi les deux de la mère). Or certains génotypes prédisposent leur porteur à être exposé à un facteur de risque potentiel (exemple : hypercholestérolémie familiale due à un polymorphisme du gène *apoB*) ou au contraire le protège d'une exposition (exemple : intolérance à l'alcool due à un polymorphisme du gène de l'aldéhyde deshydrogénase). On peut donc considérer que ce facteur de risque a été tiré au sort et il est alors intéressant de regarder si les porteurs du génotype en question ont effectivement un risque accru de maladie.

Si une association est observée entre la maladie et l'allèle majorant l'exposition, cela constitue un fort argument en faveur d'une relation causale entre le facteur d'exposition et la maladie. Par exemple, il existe une association entre le risque de maladies cardiovasculaires et certains polymorphismes du gène *apoE* augmentant le taux de cholestérol, ce qui est évocateur d'un rôle causal de l'hypercholestérolémie dans son association connue avec les maladies cardiovasculaires [7].

À l'inverse, si l'association avec la maladie n'est pas retrouvée lorsqu'on étudie l'effet de génotypes augmentant l'exposition, la causalité est peu probable. Par exemple, alors qu'il existe une claire association entre le risque de maladie cardiovasculaire et le taux de CRPus, il n'a pas été retrouvé d'association avec les génotypes du gène de la CRP (pourtant fortement associés au taux de CRPus) dans l'étude [7]. L'élévation du taux de CRPus serait donc plutôt une conséquence de l'athérosclérose (causalité inverse) ou résulterait d'autres facteurs de risque de maladie cardiovasculaire (biais de confusion).

Cette approche a toutefois plusieurs limites. Tout d'abord, on ne connaît pas (encore) de gène polymorphique contrôlant le niveau de tous les marqueurs de risque. Ensuite, la population étudiée doit être la plus homogène possible sur le plan ethnique, sous peine de stratification génétique à l'origine de fausses associations.Il faut également se méfier de la pléiotropie des gènes dont certains effets peuvent en annuler d'autres, et du déséquilibre de liaison avec d'autres gènes à effet inverse, rendant alors l'association faussement négative [10].

la fabrication des gamètes et ne peut donc être soumis à des biais environnementaux. Cela revient donc en théorie à randomiser l'attribution du facteur de risque (pour plus de détails voir l'*encadré 3*).

Conclusion

Les études observationnelles étiologiques sont particulièrement difficiles à mener dans de bonnes conditions méthodologiques, car sujettes à de nombreux biais souvent non contrôlables. Elles ne pourront probablement jamais remplacer totalement les essais expérimentaux bien conduits. Cependant dans bien des cas, les approches sont complémentaires et il convient de connaître les intérêts et les écueils des différents types d'études, ainsi que nous venons de les présenter brièvement. De la qualité de l'exploration des biais, du recueil et des analyses statistiques dépendront également les inférences qui pourront être portées. D'une facon générale, le lecteur se doit de toujours rester prudent dans son interprétation des résultats, notamment en termes de causalité. ■

Liens d'intérêts : les auteurs déclarent ne pas avoir de lien d'intérêt.

Références

1. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. *Clinical Epidemiology. How to do clinical practice research*. Third Edition. 2006. Lippincott Williams and Wilkins.

2. Bouyer J, Hemon D, Cordier S, *et al*. Épidémiologie : principes et méthodes quantitatives. Paris : INSERM, 1995.

3. Straus SE, Richardson SW, Glasziou P, Haynes RB. *Evidence-based medicine. How to practice and teach EBM*. Third edition. 2005. Elsevier.

4. Mendis S. The contribution of the Framingham Heart Study to the prevention of cardiovascular disease: a global perspective. *Prog Cardiovasc Dis* 2010 ; 53 : 10-4.

5. Schulz K, Grimes DA. Case-control studies: research in reverse. *Lancet* 2002 ; 359 : 231-4.

6. van Hylckama Vlieg A, Helmerhorst FM, Vandenbroucke JP, Doggen CJ, Rosendaal FR. The venous thrombotic risk of oral contraceptives, effects of oestrogen dose and progestogen type: results of the MEGA case-control study. *BMJ* 2009 ; 339 : b2921.

7. Zacho J, Tybjaerg-Hansen A, Jensen JS, Grande P, Sillesen H, Nordestgaard BG. Genetically elevated C-reactive protein and ischemic vascular disease. *N Engl J Med* 2008 ; 359 : 1897-908.

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.

8. Chardon B, Host S, Pedrono G, Gremy I. Apport de la méthode cas-croisé à l'analyse des effets sanitaires à court terme de la pollution atmosphérique: réanalyse de données du programme Erpurs. *Rev Epidemiol Santé Publique* 2008 ; 56 : 31-40.

9. Raitakari OT, Juonala M, Kâhonen M, *et al*. Cardiovascular risk factors in childhood and carotid artery intima-media thickness in adulthood: the cardiovascular risk in young Finns study. *JAMA* 2003 ; 290 : 2277-83.

10. Weed DL. Epidemiologic evidence and causal inference. *Hematol Oncol Clin North Am* 2000 ; 14 : 797-807.

11. Hill AB. The environment and disease: association or causation? *Proc Roy Soc Med* 1965 ; 58 : 295.

12. Grimes D, Schulz K. Bias and causal associations in observational research. *Lancet* 2002 ; 359 : 248-52.

13. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The - Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007 ; 4 : e296.

14. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004 ; 33 : 30-42.

Copyright © 2021 John Libbey Eurotext. Téléchargé par MONSIEUR BERNARD LEVY le 16/05/2021.