

Enquêtes de cohorte

tiré de « Enquêtes de cohorte » In *Epidémiologie* – P. Czernichow, J. Chaperon, X. Le Coutour. Masson, Paris, 2001, pages 235-254

Nantes _ 2025

Bernard BRANGER

Epi-Sûre (Association de conseil en épidémiologie et statistiques)

11 bis, rue Gabriel Luneau - 44000 NANTES

Mail : branger44@gmail.com - Tél 06 32 70 33 80

NB : Utilisation libre : référence sur le Web (<https://www.episure.fr/>)

I. Principes

Les enquêtes de cohorte (ou d'incidence) ont pour but de rechercher l'existence d'une relation (à vocation causale) entre

- la survenue d'un problème de santé (nouveaux cas) **dans une population** ;
- et la présence d'un facteur (« exposition ») suspecté de pouvoir influencer cette survenue.

Elles consistent à comparer la survenue du problème de santé dans une population initialement indemne de ce problème, selon qu'elle était exposée ou non au facteur. Il peut y avoir deux groupes (exposés et non-exposés), ou plus si on peut différencier plusieurs types et niveaux d'exposition, par exemple les différents emplois et postes de travail d'une entreprise.

Le terme « cohorte » désigne un groupe de personnes suivies au cours du temps, qui peut être défini

- par son année de naissance, la cohorte constitue alors une génération
- par la présence en son sein de personnes soumises à une exposition particulière, notamment en milieu professionnel ;
- par une autre caractéristique susceptible de représenter un facteur de risque.

L'effet du facteur étudié se manifeste après un temps d'exposition plus ou moins long : l'enquête se situe donc dans la durée ; la population est d'abord définie, puis suivie, permettant d'observer la survenue éventuelle du problème de santé. On parle aussi d'enquêtes longitudinales. Deux grands types d'études de cohorte peuvent être distingués selon qu'on peut utiliser, dans certains cas, des données déjà recueillies dans le passé (cohorte rétrospective ou historique), ou qu'il faut organiser le suivi dans le futur (cohorte prospective).

A. Cohorte prospective

La cohorte est alors constituée au début de l'étude ; on commence généralement par une enquête initiale auprès de chaque personne susceptible d'être incluse, afin de :

- s'assurer que le problème de santé à l'étude n'est pas déjà présent, même au stade infra clinique, auquel cas les personnes concernées doivent être exclues de la cohorte ;
- mesurer l'exposition actuelle (et, si nécessaire, antérieure) au(x) facteurs) qui vont être analysés.

Les personnes incluses font l'objet d'un suivi, permettant notamment d'identifier les nouveaux cas survenus. Un essai thérapeutique est une enquête de cohorte prospective particulière dans laquelle l'exposition est provoquée expérimentalement par l'investigateur, la répartition étant effectuée par tirage au sort. Il sera seulement question ici d'enquête d'observation.

Exemple : Cohorte PAQUID en Aquitaine

Une étude de cohorte prospective a été mise en place en 1988 chez 4 134 personnes âgées de 65 ans et plus, résidant en Gironde et en Dordogne.

L'objectif était d'étudier le vieillissement cérébral, en particulier la survenue d'une démence, mais aussi d'une perte d'autonomie, pour rechercher des facteurs de risque susceptibles d'expliquer ces événements. Au cours des 5 premières années du suivi, l'incidence de la maladie d'Alzheimer a été de 12 pour 1 000 personnes - années, soit 140 nouveaux cas.

Divers facteurs de risque ont été suspectés, comme l'exposition à l'aluminium dans Veau de boisson, la consommation modérée de vin (effet protecteur), le niveau d'étude, le statut marital ou les loisirs.

Principe des enquêtes de cohorte prospectives.

B. Cohorte rétrospective

Dans certaines circonstances privilégiées, on peut disposer de données recueillies antérieurement, caractérisant de manière exhaustive l'histoire épidémiologique d'une population particulière : nombre de personnes entrées et sorties de la population au cours du temps, exposition éventuelle à un ou plusieurs facteurs, et surtout survenue de nouveaux cas de certaines affections.

C'est le cas notamment en médecine du travail, l'entreprise ayant habituellement une bonne connaissance des personnes employées et de leur exposition (archives), les médecins du travail pouvant avoir consigné dans leur dossier les problèmes de santé survenus.

Les caisses d'assurance maladie, ou les mutuelles devraient elles aussi pouvoir réaliser ce type d'enquête, puisqu'elles disposent à la fois de l'historique des personnes dont elles assurent la protection sociale, et des problèmes de santé survenus ; en revanche, les expositions qui peuvent ainsi faire l'objet d'une étude de cohorte sont limitées, et l'exhaustivité des données relatives aux problèmes de santé est incertaine. Cette méthode a été surtout développée pour les maladies ayant une longue phase d'induction, par exemple les cancers.

Cohorte historique en milieu professionnel

Une étude de cohorte rétrospective a été réalisée dans l'industrie du papier² pour rechercher les effets éventuels de l'exposition professionnelle aux procédés chimiques de transformation du bois en pâte, de blanchiment de la pâte, ou aux diverses poussières.

La cohorte était constituée des employés ayant une ancienneté suffisante, et suivis de 1958 à 1992.

Le recueil des données a porté sur le sexe et la date de naissance, la carrière, les postes occupés (codage), la date et la cause de décès (< 85 ans). Les décès ont été identifiés à partir des mairies et les causes communiquées par l'INSERM après accord de la CNIL

La référence utilisée était la mortalité observée dans la population générale du département. Un excès non significatif de cancers du pancréas a été observé : 5MR = 1,3 [0,7 ; 2,3].

C. Cohorte fixe (fermée) ou dynamique (ouverte)

Cette distinction concerne le moment de l'inclusion des personnes au sein de la cohorte ; on distingue - les cohortes fixes, dites encore fermées, dans lesquelles toutes les personnes de la cohorte sont incluses au début de l'étude, sur une courte période, sans inclusion nouvelle au-delà ;

- les cohortes dynamiques, dites encore ouvertes dans lesquelles l'inclusion des personnes se poursuit au fur et à mesure que les premières personnes sont suivies, le temps de participation de chacune au suivi étant variable.

Dans une cohorte fixe, seules les personnes encore présentes dans la population peuvent être incluses dans la cohorte ; si un processus d'exclusion est intervenu auparavant, la cohorte sera un reflet de cette sélection. L'incidence du problème de santé en sera affectée, ce qui peut donc représenter un biais de sélection.

II. Constitution de la cohorte

A. Définition

Le choix d'une cohorte est avant tout déterminé en fonction de l'hypothèse étiologique qu'on veut tester par l'enquête

- on constitue un groupe caractérisé par la présence d'une exposition, souvent parce qu'elle est spécialement fréquente ou intense : ainsi la mise en évidence des effets sur la santé d'un produit est-elle plus sensible en milieu de travail, lorsqu'il existe une exposition professionnelle à ce produit ; la même relation serait plus difficile à étudier dans la population générale, l'exposition étant plus rare ou moins forte ;

- lorsqu'on souhaite une cohorte issue de la population générale, le choix se porte souvent sur une zone géographiquement définie, comme par exemple une ville; l'enquête de cohorte la plus célèbre est celle qui, depuis 1948, suit les habitants de Framingham, petite ville du Massachusetts (États-Unis), conduisant à des connaissances essentielles sur les facteurs de risque des affections coronariennes ; on attend de telles études que leurs résultats soient applicables à une population large, à commencer par les résultats descriptifs ;

- le choix est en fait aussi influencé par la facilité du recueil des données, ou bien la motivation des personnes - supposée ou réelle - pour participer : médecins britanniques, infirmières américaines, adhérents d'une mutuelle (exemple de la MGEN) volontaires.

B. Critères d'inclusion

Si la cohorte est définie par la présence d'une exposition, l'inclusion devra comporter des critères relatifs à cette exposition : nature précise, niveau, ancienneté minimum.

Dans les enquêtes prospectives, la possibilité d'un suivi ultérieur est un critère déterminant. La possibilité d'obtenir une information exhaustive et exacte sur toutes les personnes incluses est aussi une condition essentielle de validité ; ce critère intervient à la fois

- pour les données d'exposition et celles qui concernent l'état de santé

- lors de l'inclusion, et pendant le suivi.

Dans les enquêtes de cohorte rétrospectives, le choix fait intervenir la possibilité d'accès aux données recueillies, et également la qualité de ces données : archives d'entreprise, registre d'expositions professionnelles, dossiers médicaux, etc.

Un autre critère d'inclusion important est la date d'entrée dans la cohorte : c'est elle qui sera prise en compte pour calculer le temps de participation de chaque personne à la cohorte, en terme de personnes-années le cas échéant.

Enfin un critère capital de sélection est que **les personnes incluses ne soient pas atteintes de la maladie étudiée**, ce qu'il faudra vérifier avant de confirmer l'inclusion ; on peut être conduit à réaliser systématiquement des examens, dont l'acceptation par des personnes se percevant en bonne santé peut poser un problème, comme pour tout dépistage. Cette recherche préalable permet de réaliser une enquête transversale, préalablement au suivi de la cohorte ; les résultats (prévalence) d'une telle enquête peuvent être intéressants s'ils ont été obtenus à partir d'une population représentative.

C. Biais de sélection

L'inclusion dans une cohorte prospective implique toujours le volontariat ; comme pour toute enquête descriptive, celui-ci sélectionne des personnes ayant des caractéristiques particulières ainsi pour la cohorte SU.VI.MAX3 sur 15 789 candidats, 10 984 volontaires ont retourné un questionnaire de motivation, faisant apparaître la sur-représentation des femmes et des professions intermédiaires. Cette sélection ne constitue pas un biais si on réalise ultérieurement une comparaison interne à la cohorte (cf. plus loin) : la sélection de personnes en meilleure santé conduira à la survenue moins fréquente du problème de santé étudié, donc l'étude sera moins puissante pour démontrer le rôle éventuel d'un facteur.

Cette sélection de personnes volontaires peut être bénéfique pour la qualité du suivi (cf, plus loin) : Dans la cohorte EVA4, par exemple, les personnes âgées recrutées à partir de listes électorales, sans relance, avaient un niveau d'éducation, un état de santé et un niveau cognitif supérieurs à ceux de la population générale et ont eu un très bon taux de suivi.

Dans le cas d'une cohorte fixe, par exemple en milieu professionnel, les personnes présentes à un moment donné dans une entreprise sont celles qui sont en état de travailler ; elles sont donc nécessairement en meilleure santé que l'ensemble des personnes qui ont travaillé, ou travailleront dans cette entreprise, telles qu'on pourrait les recruter dans une cohorte dynamique : parmi ces dernières en effet, certaines personnes seront ultérieurement exclues du fait de leur état de santé ce biais de sélection est connu sous le nom de *healthy workers effect*. La comparaison avec une référence externe, d'une cohorte exposée, mais en meilleure santé que la population générale, peut conduire à un résultat paradoxal (incidence plus faible), au moins dans le début du suivi : ce résultat ne doit évidemment pas faire conclure à un effet protecteur de l'exposition.

Une autre source de sélection dans les cohortes professionnelles est l'exclusion des travailleurs temporaires, ou appartenant à des entreprises sous-traitantes, souvent les plus exposés au facteur étudié.

III. Mesure de l'exposition

C'est l'un des principaux atouts des enquêtes de cohorte prospectives. Lorsque la mesure des facteurs d'exposition nécessite une grande précision, par exemple pour apprécier le rôle des facteurs de l'alimentation dans la survenue de maladies chroniques, une mesure rétrospective est exclue.

Dans ces conditions, une enquête de cohorte prospective rend possible une mesure approfondie de l'exposition à différents facteurs, en associant les données de l'interrogatoire et de l'examen clinique, celles de la biologie ou d'explorations fonctionnelles, celles de prélèvements réalisés dans l'environnement de la personne. Une validation de ces mesures, par la réalisation de mesures répétées, permet d'apprécier leur variabilité à court terme.

Depuis quelques années, diverses enquêtes de cohortes ont constitué des banques de matériel biologique', par exemple des sérothèques, qui permettent l'analyse ultérieure de paramètres biologiques, parfois bien plus tard. Dans les enquêtes cas-témoins « emboîtées » dans une cohorte, les dosages peuvent être restreints à des échantillons de personnes suivies, plutôt que chez toutes.

La mesure de l'exposition doit dissocier deux dimensions :

- **la durée de l'exposition** : la date de début est en général antérieure à celle de l'entrée dans la cohorte ; un temps minimum d'exposition est parfois nécessaire pour être considéré comme exposé ;
- **l'intensité de l'exposition** : elle est souvent relativement constante au cours du temps pour un individu. La dose d'exposition est le plus souvent difficile à évaluer. La méthodologie de la relation dose-effet a largement bénéficié des travaux sur l'exposition aux radiations ionisantes : différenciation entre dose délivrée et dose absorbée, effet de seuil ; toutefois, ce modèle n'est pas toujours utilisable en pratique.

La combinaison de ces deux dimensions, dont chacune peut évoluer au cours du suivi (cf. plus loin), conduit à déterminer le niveau d'exposition de chaque personne de la cohorte, de façon plus ou moins élaborée (cf. chapitre 12) ; il est possible de constituer un éventail de différents niveaux d'exposition, dont chacun caractérise un groupe de personnes, situation idéale pour rechercher ultérieurement une relation dose-effet lorsque l'incidence de l'affection sera connue pour chaque niveau.

Lorsque l'exposition est caractérisée par une variable quantitative, une solution fréquente pour constituer un tel éventail est de répartir la cohorte en groupes de tailles égales : par exemple 5 groupes définissent des quintiles d'exposition, du premier (les 20 % les moins exposés) au se (les 20 % les plus exposés) : l'incidence pourra être calculée dans les 4 derniers quintiles relativement au premier.

La comparaison de l'exposition aux facteurs de risque lors de la constitution dans la cohorte, chez les cas prévalents, qui sont exclus de la cohorte, et chez les personnes indemnes, incluses dans la cohorte, ramène au cadre d'une étude cas-témoins .

IV. Choix du groupe de référence

Comme toujours, l'observation d'un seul groupe exposé, au sein duquel on estime une incidence, ne permet aucune conclusion ; c'est la comparaison avec l'incidence observée dans un autre groupe, qui constitue alors le groupe de référence, jugé non - ou moins - exposé, qui donne du sens à cette observation.

Deux critères sont essentiels pour le groupe choisi comme référence pour la comparaison

- il doit présenter une forte similitude avec le groupe exposé, sur des caractéristiques ayant un lien avec le problème de santé étudié : en particulier l'âge, le sexe, la catégorie socio-professionnelle, etc. Les éventuelles différences concernant ces facteurs de confusion devront être prises en compte lors de l'analyse, mais elles affaiblissent les résultats issus de la comparaison ;

2401

- il doit être caractérisé par l'absence d'exposition au facteur étudié, ou du moins un niveau d'exposition objectivement plus faible que celui du groupe exposé.

Deux possibilités sont offertes : un groupe de référence interne à la cohorte dont fait partie le groupe exposé, ou externe à elle.

A. Référence interne

C'est le choix le plus fréquent pour les études de cohortes prospectives : au sein d'une cohorte contrastée du point de vue de l'exposition au facteur étudié, on peut individualiser clairement des personnes non-exposées, qui présentent des caractéristiques de similitude suffisantes avec le groupe exposé, pour que la comparaison ait un sens. Ce critère peut être vérifié dans des cohortes issues de la population générale, il est souvent rédhibitoire dans des cohortes spécifiques, comme les groupes professionnels, où les personnes exposées et non-exposées présentent - outre l'exposition - de nombreuses différences (âge, sexe, niveau d'étude, catégorie socio-professionnelle, etc.).

L'idéal est de pouvoir identifier dans la cohorte au moins trois niveaux d'exposition (cf. plus haut), qui permettront ultérieurement de rechercher une relation dose-effet, critère important en faveur d'une relation de causalité.

En tous cas, un point essentiel est la standardisation entre les groupes de toutes les mesures réalisées, qu'elles portent sur l'exposition étudiée, les facteurs de confusion ou évidemment la survenue des problèmes de santé en question (cf. plus loin).

B. Référence externe

Dans le cas d'une étude de cohorte prospective, dont un groupe est défini par l'exposition de toutes les personnes, on peut choisir un autre groupe

- ayant des caractéristiques similaires à celles de la cohorte exposée, notamment en terme de facteurs de confusion potentiels ;
- mais qui ne soit pas exposé au facteur étudié.

Le choix de ce groupe non-exposé doit être pertinent pour être comparé au groupe exposé, mais la recherche d'un tel groupe peut être laborieuse, et de toutes façons il s'agit d'une solution lourde, car elle nécessite la mise en place d'une procédure de suivi tout à fait similaire à celle du groupe exposé, à des fins de standardisation.

Compte tenu de l'influence des facteurs de confusion dans la comparaison, on peut être amené à réaliser un appariement sur quelques-uns de ces facteurs (par exemple, catégorie socio-professionnelle, tabac, ou alcool) entre les cohortes exposées et non-exposées.

Dans le cas d'une étude de cohorte rétrospective, le choix d'une population de référence est souvent limité par les données disponibles

- si la survenue des problèmes de santé a été mesurée en terme de morbidité (nouveaux cas), on est contraint de rechercher une population de référence, définie sur une base géographique, dotée d'un registre qui enregistre la même affection ; il est rare que cette population soit celle dont est issue la cohorte exposée, ce qui fait craindre des différences pour des facteurs de confusion ;
- plus souvent, la survenue des problèmes de santé a été mesurée en terme de mortalité ; la disponibilité des données est alors plus ouverte (cf. chapitre 7). On peut souvent disposer de données locales (département) correspondant à la cohorte exposée ; il faut alors vérifier que le poids

démographique de cette cohorte par rapport à la population générale n'est pas tel que les données locales de mortalité de la population soient surtout le reflet de celles de la cohorte (par exemple ce serait observé pour la cohorte d'une importante entreprise dans un petit département, pour une cause de décès peu fréquente) ; cette difficulté peut faire opter pour une population de référence régionale ou même nationale. Dans ces conditions, la comparabilité de la population de référence, et de la cohorte exposée peut être altérée, même si, comme c'est la règle, on réalise une comparaison standardisée sur l'âge et le sexe (cf. plus loin). On recommande alors de comparer la cohorte exposée avec plusieurs populations de référence : si le résultat des différentes comparaisons possibles va dans le même sens (cohérence), on peut conclure ; si ce résultat est discordant (différence avec une référence, et pas avec une autre), on ne peut guère conclure qu'à la nécessité d'une nouvelle étude s'appuyant sur une meilleure référence, prenant mieux en compte les facteurs de confusion.

V. Suivi de la cohorte

Il permet de recueillir au cours du temps des données qui portent à la fois sur

- un numérateur lié à la survenue des problèmes de santé (cas incidents ou décès) ;
- un dénominateur qui caractérise (exposition) et quantifie la population et la durée du suivi (personnes-années).

L'objectif essentiel est d'aboutir à des mesures d'incidence (risque absolu) chez les personnes exposées et non-exposées.

Ce suivi implique d'identifier, pour chaque personne de la cohorte, la date et les circonstances de sa sortie : survenue du problème de santé, fin de suivi indemne de ce problème, suivi interrompu (décès d'une autre cause, perdu de vue). Il doit aussi préciser l'exposition des personnes : maintien ou évolution de leur statut initial.

A. Survenue des problèmes de santé

Ils peuvent être caractérisés en terme de mortalité ou de morbidité.

1. Mortalité

C'est le cas habituel dans les cohortes rétrospectives. L'information sur les décès et leur cause fait l'objet d'une attention particulière de l'INSERM qui gère ces données à l'échelle nationale (cf. chapitre] et 2). Un décret récent autorise l'accès aux données relatives au décès des personnes inscrites au Répertoire national d'identification des personnes physiques, dans le cadre des recherches dans le domaine de la santé, après accord de la Commission nationale de l'informatique et des libertés (CNIL). Le statut vital (sans la cause en cas de décès) est par ailleurs connu des mairies de naissance, des employeurs, des assurances ou des caisses de retraite.

42420

Le recueil des décès peut le plus souvent être considéré comme exhaustif, s'il a pu être recherché de façon concluante pour chaque personne incluse dans la cohorte (notamment à partir de sa commune de naissance). Il faut surtout souligner l'importance d'une standardisation du codage de la cause du décès

- c'est le cas lorsque les décès survenus dans la cohorte et ceux de la population de référence proviennent tous deux du service commun de l'INSERM ;

- en revanche, l'utilisation des données de l'INSERM pour la population de référence, et d'un codage particulier effectué pour la cohorte exposée, font craindre un biais de classement (cf. plus loin).

2. Morbidité

Elle présente l'intérêt d'une meilleure puissance statistique que le suivi des décès (les cas étant nécessairement plus nombreux que les décès). L'identification des nouveaux cas (incidence) nécessite une mesure standardisée de la morbidité incidente pour chaque personne incluse, quelle que soit son exposition

- lorsqu'on peut s'appuyer sur un registre local préexistant, consacré à la même affection, les nouveaux cas peuvent être connus au prix d'un « raccordement » des fichiers concernés (celui décrivant la cohorte, celui du registre), qui nécessite un numéro d'identification commun et peut poser des problèmes de confidentialité (avis de la CNIL) ;

- à défaut, il faut mettre en place un système de surveillance propre à la cohorte, souvent lourd et coûteux (cf. plus loin).

Les nouveaux cas » apparus très précocement après le début du suivi sont souvent considérés comme des cas prévalents passés initialement inaperçus, ce qui conduit à les exclure secondairement de l'analyse.

B. Caractériser la cohorte suivie

Pour chaque personne de la cohorte, à la date pour laquelle l'analyse sera effectuée (date de point), il faudra déterminer son statut: suivie ou perdue de vue, atteinte ou non par le problème de santé étudié. Cette détermination peut être difficile, en particulier pour les personnes de la cohorte décédées d'une autre cause. Le degré d'exhaustivité de cette recherche devra être exprimé par le pourcentage de personnes suivies, et dont ce statut est connu. Un pourcentage de perdus de vue au-delà de 5 à 10 % ouvre la possibilité de biais de sélection (cf. plus loin).

Outre la date d'entrée dans la cohorte, les données démographiques (sexe, date de naissance, et de décès éventuel) sont nécessaires pour calculer le nombre de personnes-années correspondant: un homme inclus à l'âge de 46 ans et suivi pendant 8 ans jusqu'à la date de point, contribuera au nombre de personnes-années chez les hommes de 45-49 ans pour 4 ans, puis de 50-54 ans pendant 4 autres années.

La cohorte doit évidemment être caractérisée en terme d'exposition au facteur: celle-ci peut faire l'objet d'un suivi individuel, ou bien d'enquêtes transversales répétées. Surtout, il faut être attentif aux changements d'exposition : « contamination » du groupe non-exposé par des personnes devenues exposées, arrêt d'exposition (encadré). L'analyse en personnes-années doit intégrer ces changements, une même personne pouvant successivement contribuer à des personnes-années exposées, et non-exposées. Les dates de début et fin d'exposition doivent donc être recueillies pour classer précisément le niveau d'exposition de chaque personne pour l'analyse.

Cet historique de l'exposition doit également tenir compte de l'intensité de l'exposition, qui elle aussi peut varier : la synthèse de la durée et de l'intensité de l'exposition doit être argumentée : pour le risque de cancer bronchique par exemple, la consommation de 20 cigarettes pendant 10 ans, n'est pas équivalente à celle de 10 cigarettes pendant 20 ans : le risque évolue en proportion de la quantité fumée, et avec la puissance 4 de la durée...

La même attention doit être portée au cours du suivi aux facteurs de confusion connus, comme par exemple la consommation d'alcool, ou de tabac.

Dans le cas d'une cohorte rétrospective, les problèmes sont très différents ; l'accès et la qualité des archives permettant de reconstituer la cohorte et son histoire sont les points essentiels. Les expositions successives de chaque personne peuvent avoir été mesurées de la même façon (par exemple historique des postes occupés par chacun), ou bien elle peuvent être approchées par des mesures d'exposition actuelles jugées plus ou moins « représentatives » de l'exposition passée de chacun, comme pour une enquête transversale (cf. chapitre 12).

La reconstitution d'une cohorte de la population générale est évidemment bien plus complexe, car elle n'est pas « captive » d'une entreprise. Ce travail nécessite le suivi géographique des personnes identifiées comme éligibles pour la cohorte. Le fichier du Registre National des individus affiliés à l'Assurance Maladie (RNIAM) devrait rendre possible une telle reconstitution dans l'avenir'.

Exemple de changements d'exposition dans une cohorte d'infirmières

Une étude de cohorte consacrée aux risques liés à l'usage des contraceptifs oraux (Nurses Health Study) a été mise en place chez 121 700 infirmières de 30 à 55 ans aux États-Unis.

Un questionnaire postal a recueilli en 1976 : les antécédents, les facteurs de risque des cancers et des affections cardio-vasculaires, l'usage éventuel des contraceptifs oraux.

Les femmes ayant initialement une affection coronarienne, un accident vasculaire ou un cancer, ont été exclues, et les autres suivies tous les 2 ans par questionnaire postal.

Chez 7 136 femmes qui utilisaient des contraceptifs oraux en 1976, il n'en restait que 28 % en 1978, 12 % en 1980 et 5 % en 1982.

Par ailleurs 70 ont commencé une contraception orale après 1976.

C. Modalités de suivi

Dans le cas d'une étude prospective, le suivi peut être long (une ou plusieurs décennies) si les effets recherchés sont très retardés. Dans ce cas une procédure active avec programme de convocation, avec relance si nécessaire, ou le déplacement d'un enquêteur sont souvent nécessaires à l'exhaustivité du suivi. Parfois, la forte motivation des personnes permet de s'appuyer sur un suivi par questionnaire postal, évidemment moins lourd. Les différentes méthodes peuvent d'ailleurs être alternées (encadré).

Le suivi peut parfois impliquer les professionnels de santé : en particulier en milieu professionnel, l'implication des médecins du travail peut permettre un bon suivi d'une cohorte professionnelle par les visites annuelles réglementaires.

En tous cas, le suivi d'une cohorte à moyen ou long terme pose divers problèmes :

- financiers : nécessité d'un personnel stable et qualifié ;
- éthiques : maintien de la confidentialité, conflits d'intérêt, apparition de nouvelles hypothèses pour l'analyse qui n'étaient pas envisageables initialement.

Etude PAQUID

Dans l'étude PAQUID, un premier recueil de données était réalisé initialement à l'entrée dans la cohorte, au domicile de la personne, par des enquêtrices psychologues, ayant reçu une formation spécifique. Les personnes considérées comme déjà atteintes de démence (après examen par un neurologue) ont été exclues de la cohorte. Le suivi a été organisé sur une période de 8 années, en alternant les visites à domicile par des enquêtrices (1, 3, 5 et 8 ans en Gironde, 3, 5 et 8 ans en Dordogne) et des enquêtes postales ou téléphoniques entre les visites. Les événements enregistrés étaient le décès, l'entrée en institution, l'évolution fonctionnelle et intellectuelle à l'aide d'échelles de mesure.

D. Gestion des données

Elle pose d'importants problèmes techniques, liés

- au nombre élevé de personnes suivies, de type de données recueillies, de sources utilisées, chacune ayant sa spécificité ;
- au caractère longitudinal : données successives recueillies chez les mêmes personnes au cours du temps;

- à la gestion informatique de bases de données complexes ;

à la multiplicité d'études différentes utilisant la même base de données, qui pose également des problèmes d'accès aux données et de valorisation de la participation des différents chercheurs. L'organisation de la gestion des données est donc un enjeu important ; il faut en outre respecter la confidentialité des données personnelles, garantir la protection des fichiers informatiques, intégrer des procédures de mise à jour et de validation des données.

E. Prévention des principaux biais dans le suivi

1. Biais de classement (de mesure)

Dans les enquêtes de cohorte prospective, la mesure de l'exposition initiale est nécessairement réalisée sans connaître le devenir (malade ou non) de chaque personne ; les éventuelles erreurs portant sur cette mesure ne peuvent donc être plus importantes dans un groupe que dans un autre, c'est-à-dire différentielles.

En revanche, lorsque le système choisi pour déterminer l'état de santé des personnes diffère selon le niveau d'exposition, il peut y avoir des erreurs différentielles, c'est-à-dire un biais; c'est le cas par exemple des personnes soumises à des expositions professionnelles spécifiques, qui font l'objet d'une surveillance spéciale en médecine du travail, nécessairement plus sensible que le suivi spontané d'un groupe non-exposé ici encore, une standardisation rigoureuse est indispensable

Cohorte des infirmières

Dans cette étude de cohorte prospective (cf. plus haut), l'incidence des cancers du sein a été plus forte chez les femmes ayant un traitement substitutif de la ménopause (TSM) en cours : RR = 1,36 (1,11 - 1,67).

Cet effet augmentait avec l'âge des femmes, mais non avec l'ancienneté du TSM (dose).

Chez les femmes ayant reçu un TSM interrompu ultérieurement, le risque n'était pas augmenté RR = 0,98 (0,81 - 1,18).

Ce résultat a fait évoquer un meilleur suivi des femmes ayant un TSM en cours, par exemple du fait des symptômes liés à ce traitement ; selon cette hypothèse, il n'y aurait pas plus de cas dans le groupe traité, mais un meilleur dépistage.

Le même genre d'erreur peut survenir dans les études de mortalité, lorsque le codage réalisé par l'INSERM est utilisé pour une population de référence externe, et un codage spécifique pour l'enquête dans la cohorte exposée : une meilleure sensibilité des décès liés à l'exposition est alors à craindre dans ce dernier groupe.

2. Biais de sélection : perdus de vue

La principale source potentielle de biais dans une étude de cohorte est liée aux perdus de vue. Dans les cohortes prospectives, le suivi pendant une période longue des personnes pour rechercher la survenue du problème de santé étudié doit absolument être réalisé en portant sur l'exhaustivité des deux groupes, exposés et non-exposés, jusqu'au terme de l'étude. La proportion de personnes dont l'état de santé n'a pu être déterminé à ce terme, est un critère majeur de qualité pour ce type d'étude (cf, encadré).

Dans les cohortes rétrospectives, le suivi doit être poursuivi pour chaque personne jusqu'au terme prévu (en général une date et un âge donnés), y compris lorsque la personne, ayant été ou non-exposée, a quitté la cohorte : par exemple, une cause importante de biais de sélection dans les cohortes professionnelles est représentée par l'exclusion (pré-retraite, licenciement, reclassement, etc.) des personnes ayant présenté des signes précoces de mauvaise tolérance de leur exposition à un facteur professionnel. Une recherche complémentaire est souvent nécessaire pour préciser le devenir de ces personnes. L'interprétation sera d'autant plus difficile due le fait d'être perdu de vue est plus lié à l'exposition, à la maladie ou aux deux. Un exemple caricatural est le biais de survie sélective.

3. Facteurs de confusion

Dans les enquêtes de cohorte prospective avec référence interne, ou constitution d'une cohorte non-exposée, il est évidemment indispensable de recueillir (et suivre) l'exposition aux facteurs de confusion pour les problèmes de santé étudiés.

Lorsque le groupe de référence est constitué par une population générale (par exemple, région), ces facteurs sont limités à l'âge, au sexe, voire à la catégorie socio-professionnelle.

Biais de sélection

L'étude PAQUID concernant la survenue de démence ou de perte d'autonomie en Aquitaine (cf. plus haut) a inclus 2 774 personnes âgées de 65 ans et plus (hors familles d'accueil).

Parmi elles, 10 % n'ont pas été suivies : en raison d'un décès précoce (4 %), d'un refus ou parce qu'elles n'ont pas pu être contactées (6 %).

Les personnes décédées avant le premier rendez-vous pour le suivi ont été comparées avec l'ensemble des personnes suivies, en ce qui concerne leurs caractéristiques initiales, au moment de l'inclusion. Des différences significatives ont été constatées pour l'âge, le sexe, le niveau de dépendance, le niveau de performances cognitives, plusieurs paramètres d'isolement social et types d'aide, ainsi que pour certains éléments de confort du logement (absence de salle de bains, présence de marches intérieures). Toutefois, relativement à l'ensemble de la cohorte, ces personnes décédées précocement ne représentaient qu'un faible poids (4 %), de sorte que le biais potentiel lié à l'absence de suivi pour elles peut être considéré comme négligeable.

Dans les enquêtes de cohorte rétrospective, on peut parfois s'appuyer sur un recueil de ces facteurs dans le dossier médical, par exemple en médecine du travail, si la tenue de ce dossier a été parfaite : recueil régulier de l'usage du tabac et de l'alcool, en particulier pour les cancers. En réalité les méthodes d'ajustement qui ont été proposées, sont souvent inutilisables en raison de la qualité souvent défailante de ces données.

VI. Effectif requis et puissance

L'effectif d'une cohorte est une dimension importante de la puissance de l'enquête

- pour une cohorte issue d'une population (choisie pour sa proportion élevée de personnes exposées), on parle d'échantillonnage de type 1; on fixe la taille globale de la cohorte ;
- plus souvent, on choisit un groupe pour son exposition au facteur étudié, et un autre qu'on sait ne pas être exposé : on parle d'échantillonnage de type II ou encore d'enquête « exposés - non-exposés ». On peut alors définir la taille nécessaire de chaque groupe.

En dehors du cas (peu fréquent) d'une cohorte prospective fixe, dont la durée du suivi est la même pour tous, l'unité statistique pour l'analyse est la personne-année (ou parfois une unité de temps plus petite). La puissance varie donc aussi avec la durée du suivi ; elle peut notamment être affectée par l'importance des perdus de vue, et accrue par une augmentation du temps de suivi, ou un recrutement rétrospectif plutôt que prospectif.

Le calcul de l'effectif requis pour une cohorte nécessite comme à l'habitude le choix d'un risque a et d'une puissance $(1 - P)$. L'importance du risque attendu (différence d'incidence, risque relatif ou odds ratio, Standardized Incidence Ratio ou SIR, Standardized Mortality Ratio ou SMR) est également un paramètre important. Comme ailleurs, l'effectif requis peut être obtenu à partir de tables ou par calcul à partir de formules mathématiques' ou encore en utilisant des logiciels simples '°.

Lorsque le risque de maladie est rare, on observe au terme du suivi peu de cas et beaucoup de personnes restées indemnes : dans cette situation, on réalise une enquête dite cas-cohorte (cf. chapitre 11), qui compare au sein d'une cohorte, tous les cas survenus, à un nombre limité de témoins tirés au sort parmi les non-malades (en général, pas plus de 4 par cas).

Dans le cas d'une cohorte rétrospective avec une population de référence externes (cf. plus loin), on va comparer un effectif O observé de décès ou de nouveaux cas, avec un effectif attendu E , par un ratio (SMR ou SIR) ; la comparaison, en règle unilatérale, se fait entre l'hypothèse nulle M_0 d'absence de différence (SMR = 1) et l'hypothèse alternative I_{Hi} de surmortalité dans la cohorte exposée (SMR > 1). Le nombre requis de personnes dans la cohorte, pour un risque a et une puissance $(1 - f_3)$ donnés, est celui qui conduit à un effectif attendu E approché (pour les valeurs de $E > 10$) par

$$z_{\alpha} - z_{1-p} = 2 FE (SMR - 1),$$

avec z' et $z(-,p)$ correspondant aux valeurs d'une loi normale centrée réduite (Cf. table en annexe p. 371).

Comme à l'habitude, une puissance ainsi déterminée, peut être affaiblie par des erreurs de mesure non différentielles, et naturellement des biais de sélection ou de classement.

VII. AVANTAGES DES ÉTUDES DE COHORTES

Les enquêtes de cohorte fournissent une mesure directe de l'incidence d'une affection, et donc du risque relatif associé à un facteur ; elles sont particulièrement indiquées

- lorsque l'exposition est rare, ou concerne un groupe restreint de la population ; par exemple pour le suivi (souvent à long terme) des travailleurs d'un secteur industriel ;
- pour les pathologies courantes, peu spécifiques de l'exposition ;
- pour l'étude de l'exposition à des facteurs qui nécessitent des mesures précises, ou dont les mesures rétrospectives sont impossibles ou peu fiables ; par exemple : les nutriments ;
- pour explorer systématiquement toutes les conséquences de l'exposition à un facteur: par exemple : un médicament comme avec la cohorte des infirmières (cf. plus haut) ;
- lorsque le problème de santé éventuel survient rapidement après l'exposition : par exemple pour la plupart des maladies transmissibles, ou pour les facteurs de risque au cours de la grossesse.

Ces études permettent d'établir clairement la séquence chronologique entre l'exposition et ses effets éventuels, ce qui est important si le facteur peut être une conséquence de l'affection. Ceci permet d'orienter l'interprétation des associations observées dans un sens causal_

La mesure de l'exposition initiale, lorsqu'elle est effectuée prospectivement, avant la survenue de l'affection, ne peut donc être affectée à cette dernière éventualité.

VIII. LIMITES DES ÉTUDES DE COHORTES

Elles expliquent le nombre réduit de ces études, comparativement aux études cas-témoins

- les études prospectives impliquent l'inclusion d'un nombre élevé de personnes, des moyens importants pour le suivi de la cohorte dans des conditions stables, permettant la standardisation des procédures. Leur coût est donc très élevé ; leur durée généralement longue, ne laisse espérer la réponse à la question posée qu'après un long délai ;
- les études de cohorte rétrospectives sont moins lourdes, mais leurs conditions de leur mise en oeuvre sont rarement réunies ;
- le risque d'une proportion élevée de perdus de vue constitue une source majeure de biais, difficile à limiter pour les cohortes de population, sur des longues périodes ;
- elles ne permettent pas l'étude de maladies de faible incidence (sauf cas particulier) ;
- elles sont généralement réalisées sur des populations spécifiques, ou en tous cas non représentatives de la population générale, ce qui limite la portée des résultats obtenus.

=====