

Méthodologie des principaux types d'études médicales et glossaire épidémiologique

1. Méthodologie des principaux types d'études médicales

Les principaux types d'études médicales se définissent comme suit :

1.1. Selon le type de l'étude (ou l'attitude de l'investigateur)

1.1.1. Les études expérimentales (ou d'intervention)

Elles sont toujours comparatives et prospectives.

Le chercheur intervient sur le statut d'exposition des sujets. Il peut intervenir sur :

- le ou les facteurs d'exposition
- le moment d'exposition
- les personnes exposées

Dans un premier temps, les groupes sont suivis puis sont comparés entre eux. La réalisation d'études expérimentales est souvent limitée en raison de contraintes éthiques.

1.1.2. Les études quasi-expérimentales

Elles sont comparatives et non randomisées, comme l'évaluation d'une action ou d'un programme de santé publique (programme de prévention ou de dépistage) : études « avant et après » ou « ici et ailleurs ».

1.1.3. Les études d'observation ou observationnelles

Elles reposent sur l'observation de phénomènes de santé et des facteurs qui les sous-tendent. Il n'y a pas d'intervention de l'investigateur.

1.2. Selon l'objectif de l'étude

1.2.1. Les études descriptives de prévalence

Elles étudient la présence d'un facteur donné ou d'une maladie dans une population à un moment précis. Elles permettent d'apprécier l'importance d'une maladie, de connaître les groupes à risque, et ainsi d'orienter des programmes de santé publique. Il s'agit d'enquêtes transversales, uniques ou répétées.

1.2.2. Les études de cohorte descriptives (ou études d'incidence)

Elles ont pour but de suivre l'évolution d'une population (ou cohorte) sur une période donnée. Durant cette période, on observe l'apparition d'une ou de plusieurs maladies. Le taux d'incidence permet d'évaluer l'évolution d'une maladie et l'impact des mesures préventives.

1.2.3. La série chronologique ou série temporelle ou chronique

Il s'agit d'une suite finie de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps. C'est une série statistique à deux variables dont l'une, le temps, est reporté sur l'axe des abscisses. Suivant la nature du problème étudié, le temps peut être exprimé en minutes, en heures, en jours, en mois, en trimestres ou en années.

1.2.4. Les études étiologiques d'observation

Elles sont toujours comparatives. Elles cherchent à mettre en évidence l'association entre l'exposition à un facteur de risque et la survenue d'une maladie. On distingue trois principaux types d'étude :

- les études de cohortes toujours prospectives (sur échantillon représentatif, de type exposés/non exposés ou cohorte historique)
- les études cas-témoins, rétrospectives
- les études transversales

Le choix du type d'étude dépend du problème posé.

1.2.4.1. L'étude de cohorte sur échantillon représentatif de la population à étudier, c'est-à-dire d'un sous-groupe de sujets obtenu par tirage au sort. La séparation entre sujets exposés et non exposés se fait après leur sélection.

1.2.4.2. L'étude de cohorte de type exposés/non exposés porte sur deux cohortes de sujets constituées de façon distincte. Une cohorte est exposée au facteur de risque, l'autre n'est pas exposée à ce facteur. Ces deux cohortes sont suivies pendant une période de temps au cours de laquelle est notée l'apparition de la maladie.

1.2.4.3. L'étude de cohorte historique associe la recherche rétrospective de l'exposition à un facteur de risque et le suivi de la cohorte pour l'apparition de la maladie.

1.2.4.4. L'étude cas-témoins dans laquelle l'enquêteur choisit les groupes étudiés sur la base de leur statut de sujets malades (cas) ou non malades (témoins). Les facteurs de risque sont recherchés dans le passé par l'interrogatoire des individus ou la collecte de données dans les dossiers médicaux.

1.2.4.5. L'étude transversale où les données sur les éventuels facteurs de risque sont recueillies en même temps que l'information sur le statut de la maladie. Une enquête étiologique transversale n'apporte qu'une simple indication sur une éventuelle liaison entre le facteur étudié et la maladie. Elle permet cependant d'apporter des arguments lorsqu'une liaison est suspectée, afin de mener une nouvelle enquête étiologique de type cohorte ou cas-témoins.

1.2.5. Les études pronostiques

Elles sont menées pour répondre à un objectif concernant l'évolution d'un phénomène de santé. Elles portent sur des sujets porteurs d'une affection précise dont on veut connaître le devenir ou le pronostic : guérison, rechute, survenue d'une complication, décès...

L'objectif principal d'une étude pronostique peut être :

- Descriptif : **analyse de survie descriptive** sur un seul groupe suivi
- Étiologique, étude toujours comparative s'appuyant sur :
 - . **des analyses de survie** portant sur deux ou plusieurs groupes suivis (par exemple : entre traitements ou entre facteurs de risque)
 - . **des études de cohorte de type exposés-non exposés**

1.2.6. Les études évaluatives d'observation

Elles sont toujours comparatives. L'évaluation porte sur :

- les pratiques professionnelles, c'est-à-dire le bien fondé d'un acte, sur sa conformité avec les règles de l'art et sur la qualité des soins prodigués
- la validité de méthodes diagnostiques, thérapeutiques ou de dépistage
- l'efficacité d'une procédure ou d'une technique

1.2.7. L'enquête CAP

Il s'agit d'une enquête spécifique des connaissances, attitudes et pratiques (CAP) des bénéficiaires visés, par rapport à certains éléments spécifiques à leurs compétences de santé. Elle évalue trois points : le niveau de la connaissance complète, les attitudes motivant les comportements et les pratiques préventives et de prise en charge des populations cibles. Il s'agit d'une étude mixte, qualitative et quantitative, menée dans un but analytique. Les résultats sont catégorisés, qualifiés ou quantifiés, selon l'objectif poursuivi.

En effet, cette étude CAP peut conduire à :

- collecter des informations sur une population particulière ou une thématique peu connue qui débouchent sur des activités d'éducation et de prévention (**CAP informative** qui est l'approche la plus utilisée en première intention par les professionnels de santé)
- tester des hypothèses (**CAP évaluative**, utilisée pour les stratégies d'intervention ou d'acceptabilité des messages reçus)
- établir une base de données pour mesurer et évaluer les changements d'attitudes et de comportement à deux moments différents (**CAP comparative** qui nécessite un échantillonnage rigoureux)

Une recherche informative (qualitative ou quantitative) supplémentaire peut être conduite avant l'étude CAP ou après celle-ci. Parmi les méthodes de recherche qualitatives les plus largement utilisées, on distingue les discussions de groupe "**Focus Group Opinion**". Ces discussions sont particulièrement intéressantes pour répondre à des questions de type "comment" et "pourquoi" par une dynamique de groupe et d'échanges entre les participants, à la différence des études CAP "classiques".

1.2.8. Les séries de cas

Elles correspondent à l'observation détaillée d'un certain nombre de patients *semblables entre eux* : « *les cas* », dans un but descriptif ou ayant comme objectif préliminaire de générer des hypothèses et de faire une étude plus structurée dans un deuxième temps.

Elles ne permettent pas de tirer des conclusions généralisables à d'autres cas. Ce type d'étude correspond **à l'analyse d'une série de cas sans pour autant avoir un objectif de comparaison avec un groupe témoin ou un autre groupe de patients**. Dans ce type d'étude, on ne considère que le nombre de cas de la maladie, qui représente le numérateur, si bien qu'il ne doit pas servir pour faire une évaluation de risques. Parfois, pour mettre en valeur ses résultats, le chercheur présente le calcul d'une distribution proportionnelle pour une telle série, c'est-à-dire les pourcentages du nombre total de cas qui appartiennent à telle tranche d'âge, à tel genre ou à tel groupe ethnique. Ces valeurs ne sont pas des taux (incidence ou prévalence) car le dénominateur dans le calcul est toujours le nombre de cas de la série et non pas la population à risque ou la population générale.

Le recueil des patients peut être rétrospectif (recenser les patients vus dans le service dans le passé par retour au dossier médical) ou prospectif (à chaque fois qu'un nouveau patient arrive dans le service, recueil par interrogatoire, des examens...).

1.3. Selon la chronologie du recueil des données

Le recueil des données peut être :

1.3.1. Permanent : pour la morbidité ou la mortalité, type registre, intéressant un ou plusieurs événements de santé dans une population géographiquement définie.

Des informations homogènes et standardisées concernant les patients sont recueillies selon un protocole déjà établi.

1.3.2. Longitudinal et prospectif : le recueil de données a lieu lors d'un **suivi** dans le temps (dans le futur), sur un ou plusieurs groupes de sujets dans le but d'étudier les phénomènes de santé qui affectent le ou les groupes au cours du temps.

1.3.3. Longitudinal et rétrospectif : le recueil de données a lieu dans le **passé**. Souvent, il est possible de choisir une période d'une année ou plus et de répartir les cas étudiés tout au long de cette période, ce qui réduit les risques de biais liés aux variations saisonnières.

1.3.4. Transversal : le recueil d'informations a lieu à un moment donné (**dans le présent**) dans le groupe de population auquel on s'intéresse et donne une image instantanée d'une situation épidémiologique. C'est un recueil d'information **unique** de tous les individus permettant de mesurer à un moment donné, sur une période brève, la fréquence des cas présents, c'est à dire la prévalence de la maladie. Le recueil transversal peut être **répété dans le temps** et permet d'avoir une vision longitudinale des phénomènes étudiés.

1.4. Selon le mode de sélection de l'échantillon

1.4.1. Dans le cadre d'une étude expérimentale ou d'intervention

L'étude est dite **randomisée** si l'affectation des sujets à un groupe ou à un autre est **tirée au sort**. C'est le hasard qui détermine donc l'appartenance du sujet à un groupe. La randomisation permet de limiter les biais de sélection (groupes homogènes) et les biais de confusion (répartition équilibrée des facteurs de confusion connus ou non).

Les études ou essais cliniques randomisés sont qualifiés :

- d'ouverts quand le traitement est connu de tous
- en simple aveugle quand seuls les patients ignorent la nature du traitement
- en double aveugle quand patients et médecins ignorent la nature du traitement

Le **plan expérimental en groupes parallèles** « parallel design » consiste à comparer deux ou plusieurs groupes simultanément. Après la randomisation, les patients demeurent dans le même groupe et reçoivent le même traitement tout au long de l'étude.

Le **plan expérimental croisé** « cross-over design » consiste en l'administration à tous les sujets qui participent à l'étude, d'un traitement A et d'un traitement B, alternativement.

Le **plan expérimental factoriel** « factorial design » répond à deux questions différentes avec le même essai. Il permet un gain de temps et une « économie en patients ». Par exemple, étudier l'efficacité « propre » de deux traitements différents (A et B) et de leur administration en association (A+B).

1.4.2. Dans le cadre d'une étude quasi-expérimentale

Les sujets sont répartis sans tirage au sort dans les groupes recevant ou non une action ou un programme de prévention par exemple, dans le but d'évaluer l'efficacité de cette action ou de ce programme. Ces essais se font chez des personnes saines (prévention primaire) ou malades (prévention secondaire).

L'étude est réalisée par comparaison simultanée « ici-ailleurs » de deux communautés distinctes géographiquement, l'une recevant l'intervention et l'autre pas. D'une façon générale, les sujets sont appariés selon l'âge, le genre et la catégorie socioprofessionnelle.

L'étude est réalisée par comparaison successive « avant et après » de deux groupes distincts dans le temps. On compare selon le critère de santé retenu, des sujets avant la mise en place de l'intervention et après. La situation « avant » sert de référence pour évaluer l'efficacité de l'intervention.

Ces études permettent de donner des pistes de réflexion pour un moindre investissement.

1.4.3. Dans le cadre d'une étude d'observation étiologique

Les sujets peuvent être sélectionnés :

- sur la base de l'exposition au risque étudié. On compare alors la proportion de malades (ou de décès) observée entre deux ou plusieurs groupes de sujets initialement indemnes de la pathologie et qui diffèrent par leur niveau d'exposition à un ou plusieurs facteurs de risque soupçonnés de cette maladie. Il s'agit d'une étude étiologique de cohorte qui peut être prospective ou rétrospective.
- selon le statut vis-à-vis de la maladie. On compare alors la fréquence de l'exposition antérieure à un (ou plusieurs) facteur(s) de risque dans un groupe de « cas » atteints de la maladie étudiée et dans un groupe de « témoins » indemnes de la maladie étudiée. La mesure du facteur de risque étudié est effectuée rétrospectivement, en général par entretien ou par auto-questionnaire. Le recrutement des cas peut porter sur les cas présents au début de l'étude (cas prévalents) ou sur les nouveaux cas (cas incidents). L'étude cas-témoin est alors qualifiée à tort de « prospective » car le recrutement est prospectif mais l'évaluation des facteurs de risque demeure rétrospective.
- Dans le présent au moment de l'enquête et on mesure simultanément l'exposition actuelle (ou antérieure) à un facteur de risque et l'état de santé.

1.4.4. Dans le cadre d'une étude d'observation descriptive

L'échantillon aléatoire par tirage au sort d'une population (sondage) peut être :

- **Simple** : le tirage au sort des unités statistiques (qu'on utilise pour calculer les mesures de fréquence) est réalisé à partir de la base de sondage. L'inconvénient de ce mode de sondage est son caractère fastidieux lorsque la taille de la population et/ou de l'échantillon est grande, donnant lieu à un échantillon très dispersé qui alourdit le travail de collecte des données.
- **Stratifié** : on commence à choisir le ou les critère(s) de stratification qui doit (doivent) être lié(s) au phénomène étudié et figurer au niveau des informations contenues dans la base de sondage. La base de sondage est triée selon ce critère donnant des sous-groupes de la population. Pour chaque sous-groupe, on applique les mêmes modalités de tirage au sort que pour un sondage élémentaire.
- **En grappe** : le tirage au sort concerne des groupes d'individus qu'on appelle grappes. Pour qu'un sondage en grappe soit efficace, il faut que les grappes se ressemblent le plus possible. Il arrive que les grappes soient de tailles inégales. Pour remédier à ce problème, les grappes doivent être stratifiées selon leur taille.
- **À deux ou à plusieurs degrés** : on procède par étapes successives.

L'échantillon pseudo-aléatoire peut être :

- **empirique ou méthode des quotas**. L'exemple type est celui des sondages d'opinion. On connaît avec précision les facteurs qui influencent le choix (l'opinion) des individus sur un événement ou un état donné : profession, genre, niveau de vie, origine sociale, région géographique par exemple. On détermine la répartition en pourcentages des individus de la population en fonction de ces facteurs. On calcule en fonction de ces pourcentages et sur la base de la dimension souhaitée de l'échantillon (100 à 1500 individus par exemple), le nombre des individus ayant tel et tel caractère devant constituer l'échantillon. Ces nombres sont appelés quotas.
- **Systématique**. La disposition d'une base de sondage de tous les individus de la population (de taille N) est exigée. Les unités d'échantillons (de taille n) sont désignées dans la population suivant une suite arithmétique de raison $r=N/n$, calculée de façon à couvrir toute la population cible et de base choisie au hasard entre 1 et r. Cette méthode d'échantillonnage permet de répartir l'échantillon dans l'ensemble de la liste et comporte certains avantages : la simplicité dans la désignation des unités de l'échantillon (une suite arithmétique de raison N/n), le choix au hasard d'une seule unité et l'obtention d'une bonne

précision. Par contre, elle présente certains inconvénients, parmi lesquels le biais éventuel causé par la périodicité et le classement des unités dans le fichier.

1.4.5. La revue systématique de la littérature « Systematic Review »

C'est une synthèse de la littérature scientifique qui consiste en un travail de collecte, d'évaluation critique et de synthèse des connaissances existantes sur une question donnée et qui adopte une méthode rigoureuse minimisant les biais et les erreurs. La méthodologie comprend une question de recherche spécifique, un protocole de recherche, une stratégie de recherche bien développée utilisable dans diverses bases de données préalablement sélectionnées, des critères d'inclusion et d'exclusion, un processus de sélection, des méthodes d'analyse des données et d'évaluation des articles avec ou sans recours aux méthodes statistiques dans la synthèse des résultats des études. L'élaboration d'une revue systématique doit suivre une approche explicite selon les normes PRISMA statement, laquelle propose une checklist de 27 items et un flowchart permettant d'encadrer et faciliter la réalisation de la revue et d'en assurer la qualité.

2. Glossaire épidémiologique

Ce glossaire est non exhaustif.

Aléatoire : il relève du seul hasard.

Échantillon aléatoire : l'échantillon est obtenu par un sondage aléatoire.

Appariement : la procédure de constitution d'un groupe **témoin** de non malades (ou non exposés) est identique au groupe des **cas** (ou exposés) vis-à-vis d'un ou de plusieurs critères, le plus souvent de l'âge et du genre.

Attaque (taux d') : c'est le nombre de personnes ayant présenté une maladie, rapporté au nombre de personnes susceptibles de l'avoir. Un taux d'attaque est un **risque**.

Exemple : parmi les 500 personnes qui ont déjeuné au restaurant d'entreprise le même jour, 25 ont été intoxiquées ; le taux d'attaque est 25/500.

Biais : c'est tout effet qui tend à produire une **estimation** (de la fréquence d'une maladie, de l'association entre une maladie et un facteur de risque...) qui diffère systématiquement, en plus ou en moins, de la vraie valeur. Certains biais peuvent être contrôlés ou limités. Leur existence doit être prise en compte dans l'interprétation des résultats. On distingue, selon leur origine, de multiples biais que l'on peut regrouper en trois catégories :

- **biais de sélection**, intervenant à la conception de l'étude dans la constitution des échantillons : groupe étudié restreint à des volontaires, groupe témoin mal adapté (témoins hospitaliers par exemple). Dans les études en milieu de travail, un biais particulier intervient « le biais de sélection professionnelle ou healthy worker effect ». Dans ce cas, les personnes en activité (ou exposées à un risque) sont (presque) toujours en meilleure santé que l'ensemble de la population (ou non exposés). Les biais liés aux « perdus de vue » nécessitent également une attention particulière.

- **biais d'information**, intervenant dans le recueil des données : c'est une erreur de **mesure** (biais d'observation lié à l'enquêteur qui, par exemple, interroge plus soigneusement les cas que les témoins ; biais de déclaration ou de mémorisation, lié à la personne interrogée).

- **biais dus aux facteurs de confusion**, dus à l'existence de **facteurs de confusion qui perturbent l'association entre l'exposition étudiée et la maladie**. Il est possible de contrôler un facteur de confusion au niveau de la planification de l'étude et au niveau de l'analyse.

Cas-témoins (étude, enquête) : c'est une méthode épidémiologique consistant à comparer les caractéristiques de sujets « atteints » (cas) à celles de sujets « non atteints » (témoins).

Causalité : un facteur est dit en cause de la maladie lorsqu'une modification de sa fréquence (ou de sa valeur moyenne s'il est quantitatif) dans un groupe de sujets entraîne une modification de la fréquence de la maladie. Seule la méthode expérimentale pourrait apporter la certitude de la causalité d'une association. Pour des raisons éthiques et pratiques évidentes, une telle démonstration n'est généralement pas possible.

L'hypothèse d'une relation causale peut cependant être renforcée par un certain nombre de conditions, jusqu'à aboutir, dans certains cas, à une quasi-certitude. Les principaux arguments en faveur d'une relation causale sont les suivants :

- **séquence dans le temps** : l'exposition au facteur présumé causal doit précéder l'apparition de la maladie
- **élimination d'explications par des facteurs de confusion**, susceptibles de créer l'association constatée
- **cohérence avec les connaissances actuelles** : la causalité doit être plausible du point de vue biologique
- **force de l'association statistique** : une relation très forte a peu de chances de s'expliquer par la présence d'un facteur de confusion
- **existence d'une relation dose-effet**: le risque doit augmenter avec le niveau d'exposition au facteur
- **constance de l'association et reproductibilité** : la même relation facteur-maladie doit se retrouver dans différentes populations et différentes conditions.

Cohorte : de façon large, c'est l'ensemble de sujets ayant vécu une même expérience et suivis depuis la date de cette expérience (qui peut différer d'un sujet à l'autre). Le suivi de la cohorte est organisé de façon à repérer un ou plusieurs événements donnés, très souvent le décès avec sa cause. En pathologie professionnelle, une cohorte est le plus souvent un ensemble de personnes ayant été exposé professionnellement à une substance donnée, à une date connue.

Exemples : l'ensemble des agents ayant travaillé au moins 5 ans dans un établissement donné entre le 1er janvier 1955 et le 31 décembre 1996 ; l'ensemble de sujets nés à une même période et suivis dans le temps...

Contingence : c'est un tableau classant des observations selon les valeurs conjointes de deux (ou plusieurs) variables qualitatives. Chaque case du tableau contient un effectif. Dans le cas de deux variables, les lignes représentent les modalités de l'une des variables, les colonnes les modalités de l'autre. La constitution d'un tel tableau permet l'étude de la liaison entre les deux variables concernées ; cette liaison est testée le plus souvent par un test du Khi-2.

Dépistage (test de) : c'est un examen dont l'objectif est d'identifier rapidement une maladie ou une anomalie, si possible à un stade précoce. La **sensibilité** d'un test de dépistage est la proportion de personnes réellement malades identifiées comme malades d'après le test. La **spécificité** est la proportion de personnes réellement non malades, considérées comme non malades d'après le test. Les situations possibles peuvent être résumées par un tableau à quatre cases.

Échantillon : c'est un sous-ensemble de la population, obtenu par **sondage**. À partir des valeurs observées sur un échantillon, il est possible d'obtenir des renseignements sur la population entière par une **estimation** de paramètres ou de trancher entre des hypothèses.

Échantillon représentatif : le plus souvent, un échantillon est dit représentatif d'une population définie s'il a été obtenu par tirage **aléatoire** ou par une procédure équivalente (voir **sondage**). De façon plus générale, un « bon » échantillon est un échantillon permettant de répondre au mieux aux questions posées.

Épidémiologie : c'est l'étude de la distribution et des déterminants des états de santé et des maladies dans les populations humaines. On distingue classiquement l'**épidémiologie descriptive**, l'**épidémiologie étiologique** et l'**épidémiologie évaluative**.

D'autres types d'étude peuvent être considérés comme faisant partie de l'épidémiologie : certaines études cliniques (épidémiologie clinique) ou les essais thérapeutiques (épidémiologie expérimentale).

Épidémiologie étiologique : c'est la mise en évidence et l'analyse des relations qui existent entre les maladies et différents facteurs (**facteurs de risque**) qui pourraient avoir un rôle étiologique. Le point de départ d'un travail d'épidémiologie étiologique est le test d'une hypothèse (existence d'une relation entre tel facteur et telle maladie).

Les outils privilégiés de l'épidémiologie étiologique sont les enquêtes cas-témoins et les enquêtes de cohortes (exposés-non exposés). Les résultats sont présentés le plus souvent sous forme de mesure des relations entre la probabilité de la survenue d'une maladie et un ou plusieurs facteurs de risque.

Exemple : les études de cohortes en milieu professionnel dont l'objectif est de tester l'existence d'une relation entre une exposition professionnelle et une maladie, ou de préciser cette relation.

Épidémiologie descriptive : c'est la description de la fréquence des maladies ou des états de santé dans des **populations définies** et de ses variations en fonction des caractéristiques des personnes (âge, genre...), du temps ou de l'espace. Les outils privilégiés sont, ici, les enregistrements systématiques, les registres, les enquêtes transversales. Les résultats sont des mesures de fréquence **brutes** ou **spécifiques (incidence, prévalence, mortalité)**.

Épidémiologie évaluative : le troisième axe de l'épidémiologie est l'évaluation des interventions ou des programmes destinés à prévenir les maladies ou leurs complications. En milieu de travail, les conditions idéales d'une évaluation sont la constitution par **randomisation** de deux groupes (avec et sans intervention) et la comparaison dans les deux groupes de la situation (avant intervention et après l'intervention). L'évaluation porte sur les effets de l'intervention en termes de maladies, décès ou handicaps évités. L'évaluation peut porter aussi sur les procédures et, éventuellement, sur les coûts (comparaison coût-efficacité).

Exemple : l'évaluation d'une intervention en milieu hospitalier comportant la généralisation d'aides à la manutention des malades peut porter sur les effets sur la santé du personnel hospitalier, et aussi sur l'acceptabilité et les conditions réelles d'utilisation du nouveau matériel.

La mise en place des instruments d'observation permettant l'évaluation doit se faire avant le début de l'intervention, de façon à ce que la comparaison « avant-après » soit possible.

Facteur de confusion : c'est une variable liée à la fois à la maladie et au **facteur de risque**, pouvant être à l'origine d'une association observée entre la maladie et ce facteur de risque. La démarche fondamentale consiste à rechercher si la relation constatée entre la maladie et le facteur de risque peut s'expliquer par la présence d'un facteur de confusion. S'il y a confusion, un **ajustement** sur le facteur de confusion est nécessaire pour obtenir une estimation de cette association sans **biais**.

Facteur de risque : c'est une variable associée statistiquement à la survenue d'une maladie ou d'un phénomène de santé (relation généralement mais pas obligatoirement de nature étiologique). Exemple :

une pression artérielle élevée est un facteur de risque de la maladie ischémique; le risque est accru en présence du facteur.

Parmi les facteurs de risque, on peut distinguer les facteurs **endogènes** (propres à l'individu, notamment génétiques), les facteurs **exogènes** (liés à l'environnement), les facteurs **prédisposants** (qui rendent le sujet vulnérable) et les facteurs **déclenchants** ou **précipitants** (qui déclenchent le phénomène pathologique). Ces distinctions sont parfois contestables.

Groupe à risque : c'est le groupe dont le **risque** est accru pour une maladie donnée, du fait de ses caractéristiques en termes de facteur de risque.

Exemple : les vendeurs, travaillant debout, forment un groupe à risque pour les troubles de la circulation veineuse.

Incidence : c'est la mesure de la survenue de nouveaux cas. Les événements comptés comme cas sont le plus souvent des maladies. S'il s'agit de décès, on parle de mortalité. Le terme général d'incidence recouvre deux notions différentes, pas toujours clairement distinguées : le **risque** ou incidence cumulée et le **taux** d'incidence. Le **risque**, pour une période de temps donnée, est la probabilité qu'a un sujet indemne au début de la période d'être atteint par la maladie au cours de la période. Il se calcule soit directement, soit indirectement, à partir des **taux d'incidence**.

Le **taux d'incidence** est une mesure de la fréquence de survenue de nouveaux cas par unité de temps dans une population donnée.

Létalité (taux de) : c'est le pourcentage de décès parmi les malades.

Morbidité : c'est l'écart subjectif ou objectif à un état de bien-être physiologique ou psychologique. Cet écart est apprécié différemment selon la personne ou l'institution.

On distingue classiquement :

- **la morbidité ressentie** : c'est l'appréhension par l'individu des phénomènes pathologiques ou des gênes le concernant
- **la morbidité diagnostiquée** qui fait l'objet d'un recours au système de soins (recours au médecin en particulier). Elle est classée en maladies selon la nosologie en vigueur
- **la morbidité objective**, appréciée par un examen de santé standardisé, tenant compte des connaissances, des techniques et des **normes** actuelles.

Il peut exister des différences très importantes entre ces approches de la morbidité. La plupart des gênes mineures ressenties, par exemple, ne font pas l'objet d'un recours au système de soins. Toute interprétation des données de morbidité doit aussi tenir compte de l'environnement social : le nombre de **maladies professionnelles** ou d'accidents de travail déclarés dépend de la législation et de la façon dont celle-ci est appliquée. En Tunisie, la morbidité est connue, entre autres, par des registres de morbidité pour certaines affections, notamment pour le cancer et les maladies cardiovasculaires.

Mortalité : l'analyse des décès occupe une part importante de la description de la mesure de l'état de santé d'une population. Le taux **brut** de mortalité dans une population n'est pas un indice très informatif. On préfère les taux par âge ou d'autres indices développés en démographie.

Personne-année : c'est une unité de mesure combinant les personnes et le temps pendant lequel celles-ci sont suivies ou exposées au risque ; chaque sujet contribue pour le nombre d'années (entier ou pas) pendant lequel il est observé.

Exemple : pour une étude portant sur les travailleurs d'une entreprise, entre 1955 et 2000, la contribution au nombre de personnes-années sera de :

- 45 pour un travailleur présent dans l'entreprise en 1955, suivi jusqu'en 2000 et non décédé à cette date;
- 30 pour un travailleur embauché en 1965, décédé en 1995.

Le nombre de personnes-années peut être utilisé comme dénominateur dans le calcul d'un **taux** annuel d'**incidence** par exemple. Le numérateur est alors le nombre d'événements (maladies ou décès) survenus pendant la période d'observation.

Population étudiée : elle doit être définie de façon précise.

Exemple : si la population étudiée est celle des salariés d'une entreprise, il faut préciser : salarié à telle date, le type de contrat, inclusion ou non des travailleurs employés de façon précaire, etc.

Il est indispensable, pour calculer des **taux**, de connaître l'effectif de la population.

Exemple : l'observation de cinq cas de tuberculose pulmonaire lors d'un dépistage systématique ne peut être interprétée que si ces cinq cas peuvent être rapportés à une population précise.

Enquête dans la population : c'est une enquête dans un ensemble large de personnes, malades ou non. Exemple : l'ensemble des tunisiens.

Prévalence : Rapport du nombre de personnes affectées par une maladie à l'effectif de la **population** susceptible de présenter la maladie, à un instant donné (prévalence instantanée). La prévalence exprime la situation épidémiologique à un moment donné : on parle parfois de « taux de prévalence ».

Relation entre l'incidence et la prévalence : dans le cas où l'incidence et la prévalence d'une maladie sont constantes au cours du temps dans une population, on a : $P = I \times D$ où : P = prévalence ; I = taux d'incidence ; D = durée moyenne de la maladie.

Dans ces conditions, cette relation permet d'estimer l'incidence à partir de la mesure de la prévalence (facile à obtenir à moindre coût) et de la durée moyenne de la maladie.

Probabilité : c'est une valeur numérique, entre 0 et 1. Elle est attachée à un événement qui résulte d'une épreuve (tirage...) caractérisant la fréquence avec laquelle l'événement se produirait si l'épreuve était répétée une infinité de fois.

Proportion : c'est le rapport de deux effectifs dans lequel le numérateur est inclus dans le dénominateur. Exemple : proportion de décès dus à une cause médicale donnée.

Protocole : c'est la description de l'ensemble des étapes correspondant à une étude, y compris les documents de recueil de données (questionnaire...). Dans une étude épidémiologique, la mise au point du protocole est importante et peut nécessiter autant de temps que le recueil des données.

Ratio : c'est un rapport de deux quantités ou indices, exprimés le plus souvent dans la même unité. Exemple : sex-ratio.

Registre : c'est une enquête permanente et exhaustive. Le nom de registre est généralement réservé à des enquêtes concernant une pathologie bien spécifiée dans une zone géographique donnée.

Exemple : registre du cancer dans un département.

Rétrospective (étude, enquête) : ce terme est le plus souvent utilisé à la place de « cas-témoins ». La confusion possible avec un mode de recueil rétrospectif des données (dans une étude de cohorte historique, par exemple), sans qu'il s'agisse d'étude cas-témoins, incite à utiliser préférentiellement les termes « cas-témoins » ou « cas-contrôle ».

Risque : c'est la probabilité qu'un sujet développe une maladie donnée pendant une période déterminée. Le terme « risque » est utilisé dans beaucoup d'expressions : **facteur de risque, groupe à risque**, etc. « Risque » est aussi utilisé dans un sens proche du sens courant : « toute activité comporte des risques ». Ainsi, on parle de « risque chimique », de « risque physique », de « risque biologique ». Il s'agit ici d'une exposition à un **facteur de risque** et non de la probabilité de survenue d'une affection.

Sondage : c'est le synonyme de l'échantillonnage. Il s'agit d'une sélection d'un **échantillon** à partir d'une **population** constituée d'**unités statistiques**.

Plan de sondage : c'est la description complète du processus de sélection.

On distingue différents types de sondage selon le mode de sélection :

- **sondage aléatoire** : toutes les unités statistiques ont une probabilité non nulle d'être retenues et font partie ou non de l'échantillon selon le résultat d'un tirage au sort. Dans un **sondage aléatoire simple**, les unités statistiques ont toutes la même probabilité d'être sélectionnées. Dans un **sondage aléatoire stratifié**, la population est subdivisée en sous-groupes, le tirage s'effectue dans chaque sous-groupe avec un taux de sondage propre au sous-groupe. Exemple : dans un sondage stratifié par âge, le taux de sondage est 1/10 pour les moins de 40 ans, 1/20 au-delà.

- **sondage systématique** : une procédure systématique remplace le tirage au sort, systématique étant parfois synonyme d'**exhaustif**. Exemple : toutes les personnes se présentant à une visite de médecine du travail, nées entre le 10 et le 15 du mois font partie de l'échantillon.

- **sondage par quota** : ce n'est pas un tirage au sort, mais un tirage qui doit respecter des contraintes prédéfinies sur les effectifs par sous-groupes, le mode de sélection étant par ailleurs laissé à la liberté de l'enquêteur.

Exemple : l'échantillon doit comporter 30 hommes (cadres supérieurs de moins de 40 ans) ; 50 agriculteurs (anciens agriculteurs de plus de 60 ans).

- **sondage en grappe** : les unités statistiques constituent des groupes, ou grappes. Le tirage s'effectue sur ces groupes, dont tous les éléments appartiennent à l'échantillon. Exemple : les unités statistiques sont des familles ; on tire un ensemble de familles chez lesquelles on interroge tous les membres.

- **sondage à deux degrés** : un premier tirage est effectué dans une liste d'unités statistiques dites « unités primaires ». Pour chaque unité primaire, un tirage d'unités secondaires est effectué.

Exemple : tirage au sort d'entreprises, puis tirage au sort de sujets interrogés dans chaque entreprise.

- **sondage raisonné** : la sélection s'effectue selon des critères de choix justifiés. Ceci concerne plus souvent le choix des unités primaires.

Exemple : on peut inclure dans l'enquête telle entreprise, ou exclure telle autre au niveau de laquelle le recueil des données ne pourra pas être effectué dans des conditions satisfaisantes.

Standardisation : les mesures de mortalité ou de morbidité (incidence, prévalence) doivent être standardisées afin de permettre des comparaisons entre les populations en éliminant l'effet des **facteurs de confusion**, tels que l'âge ou le genre.

Les méthodes de standardisation les plus classiques sont la standardisation directe et la standardisation indirecte. Exemple : standardisation par rapport à l'âge :

- dans la **standardisation directe**, on calcule le taux que l'on observerait dans la population étudiée si elle avait la même structure d'âge qu'une **population de référence**

- dans la **standardisation indirecte**, on calcule un nombre d'événements (décès ou maladies) attendu dans la population étudiée, en appliquant à chaque classe d'âge les taux spécifiques d'une population de référence.

Le rapport : Nombre d'événements observé / Nombre d'événements attendu est un **rapport comparatif de mortalité (SMR)** ou un rapport comparatif **d'incidence (SIR)** selon qu'il s'agit de mortalité ou de morbidité.

Taux : c'est, le plus souvent, la mesure de fréquence d'un phénomène dans une population par unité de temps. Exemple : taux d'**incidence** annuel.

Le terme de taux est cependant parfois utilisé dans un sens différent. Exemple : taux de mortalité périnatale ou le taux d'**attaque** ; il s'agit dans ces cas de **proportion** et non de taux, au sens strict. Voir aussi : **brut, spécifique, standardisation**. Taux comparatif (de mortalité, d'incidence).

Témoin (groupe témoin ; enquête cas-témoins) : dans une étude cas-témoins, on compare un groupe de cas à un groupe de témoins, du point de vue de l'exposition passée à des **facteurs de risque**. Les cas sont des malades, les témoins des personnes non malades. Les témoins doivent être représentatifs de la **population** d'où sont issus les cas. Pratiquement, il est souvent difficile de définir le groupe témoin idéal. Une étude cas-témoins est parfois le complément d'une étude de **cohorte**. On parle alors d'étude cas-témoins au sein d'une cohorte. On examine, par exemple, l'exposition professionnelle, non pas de tous les travailleurs de la cohorte, mais seulement de ceux qui sont décédés d'une cause donnée et d'un échantillon témoin.

Transversale (étude, enquête) : une enquête transversale permet d'étudier la **morbidité** et les **facteurs de risque** à une date donnée. Une telle enquête est mal adaptée à l'étude des liens entre une maladie et des facteurs de risque car l'interprétation des liaisons observées est souvent difficile.

Validité prédictive (ou valeur prédictive, ou valeur diagnostique) :

La valeur prédictive positive est la probabilité que la maladie soit présente lorsque le test est positif : $VPP = VP / (VP + FP)$.

La valeur prédictive négative est la probabilité que la maladie ne soit pas présente lorsque le test est négatif : $VPN = VN / (VN + FN)$.

Un tel mode de calcul n'est valide que lorsque l'échantillon extrait, sur lequel l'on étudie le test ou l'examen, est représentatif de la population.

Pour plus d'information et de détails, vous pouvez consulter :

- 1- Leclerc A, Lellouch J. Glossaire statistique et épidémiologique. Encycl Med Chir. (Elsevier Masson, Paris), Toxicologie-Pathologie professionnelle, 16-905-A-10, 2000, 15 p.
- 2- Bouyer J, Hémon D, Cordier S, Derriennic F, Stücker I, Stengel B, et al. Épidémiologie : Principes et méthodes quantitatives. INSERM Paris, France (ISBN 10 : 274301167x) : Tec et Doc - EM Inter-Lavoisier ; juillet 2010. 498 p.
- 3- Bouyer J. Méthodes statistiques : Médecine-Biologie. INSERM Paris, France (ISBN 10 : 2909455742): De boeck Estem, juillet 2010. 351 p.
- 4- Goldberg M, Charpak Y, Chevalier A, Godard C, Gottot S, Guenel P, et al. L'épidémiologie sans peine. 2^{ème} éd. Paris, France (ISBN : 2-87671-039-0): Editions Frison-Roche, 20 mars 1998. 194 p.