

Études de cohorte, études cas-témoins

Introduction

Depuis les débuts de la recherche épidémiologique, de nombreux types d'étude ont été et continuent d'être développés. Ils sont cependant la plupart du temps une variation autour des deux schémas essentiels pour l'épidémiologie comparative : les études cas-témoins et les études de cohortes. Les unes comme les autres représentent une avancée par rapport aux études transversales, dans le sens où elles permettent d'étudier de façon longitudinale la relation causale envisagée. Nous allons présenter chacun de ces schémas, leurs caractéristiques ainsi que les estimations qu'elles permettent de mesurer.

Les études de cohorte

Une étude de cohorte est une étude incluant une population de personnes ne présentant pas la ou les maladies d'intérêt, qui sera suivie au cours du temps et durant laquelle la ou les expositions d'intérêt seront mesurées. Cette population est appelée une **cohorte**. Ces études impliquent généralement l'inclusion d'un grand nombre de sujets avec la mise en place d'un suivi.

Caractéristiques des cohortes

Il existe plusieurs caractéristiques qui permettent de préciser le fonctionnement de l'inclusion dans la cohorte : le dynamisme de l'inclusion, le moment d'inclusion et le nombre d'étapes d'inclusion.

Concernant le dynamisme :

- la cohorte peut être **fixe**, c'est-à-dire que les sujets sont inclus à un instant donné et sont suivis dans le temps, sans inclusions de nouveaux patients en cas de perdus de vue ;
- la cohorte peut être **dynamique**, c'est-à-dire que des sujets peuvent être inclus et exclus en fonction de leur exposition ou du statut vis-à-vis de la maladie.

Les cohortes fixes sont intéressantes pour les expositions ponctuelles (ex : personnes exposées aux radiations à Nagasaki) mais la perte constante des perdus de vue peut entraîner des biais de sélection ainsi qu'une perte de puissance au cours du temps. Les cohortes dynamiques (ex : registres territoriaux) sont plus adaptées à un suivi long mais sont en revanche plus difficiles et coûteuses à maintenir.

Concernant le moment de l'inclusion des patients :

- si au moment de l'inclusion, l'ensemble des sujets est indemne de la maladie et que le début de la maladie est recherché au cours du suivi, on parle de **cohorte prospective** ;
- si au moment de l'inclusion, le statut au sujet de la maladie est connu pour tous les sujets et que l'exposition est retrouvée par des enregistrements antérieurs, on parle de **cohorte rétrospective**, ou *cohorte historique* ;
- une cohorte peut aussi comporter une phase rétrospective et une phase prospective (cohorte historico-prospective) ;

Notons que dans tous les cas, la séquence d'enregistrement des données suit bien la séquence de la causalité (exposition avant la maladie), même si cet enregistrement est parfois retrouvé *a posteriori*.

Concernant le nombre de phases d'inclusion :

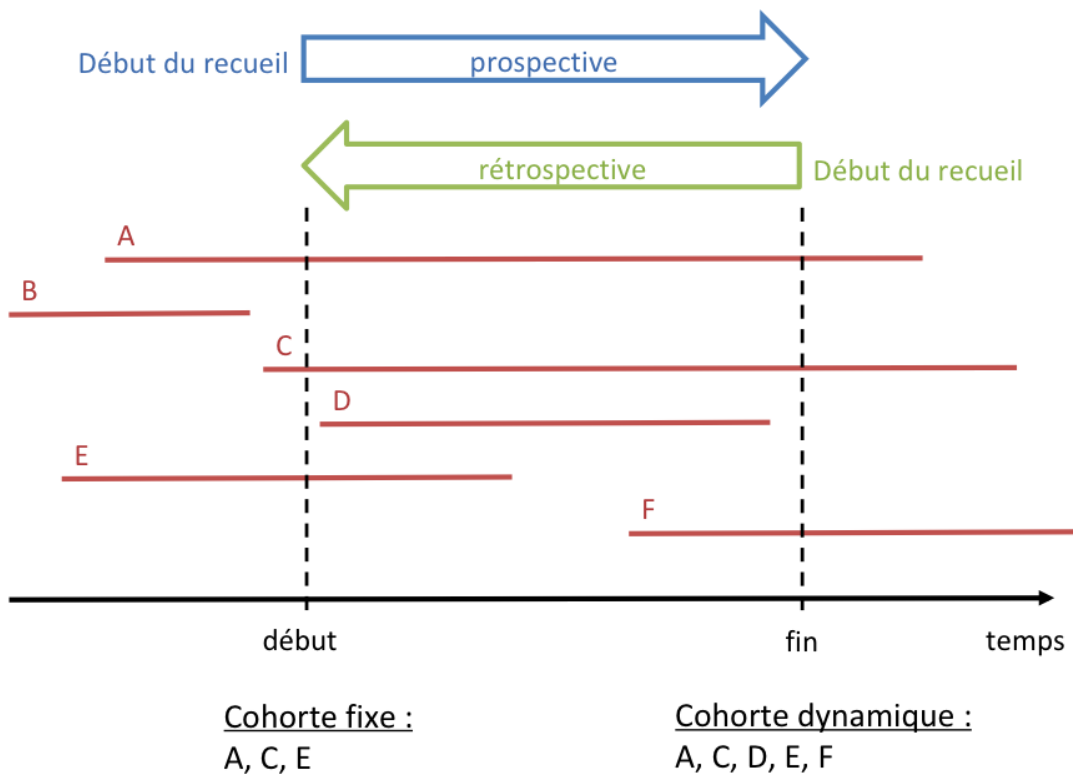


Figure 1: Représentation d'une cohorte

- dans la plupart des cohortes, la population est identifiée en une seule phase ;
- pour les expositions rares, il est parfois plus facile d'identifier dans un premier temps les personnes exposées et de constituer dans un second temps groupe de non exposés : on parle alors de **cohortes exposés/non-exposés**

Pour les études exposés/non-exposés, le risque de biais de sélection est plus grand puisqu'il faut essayer de faire en sorte que les non-exposés soient issus de la même population que les exposés (même problème que pour les cas témoins, voir plus bas).

Modalités de constitution de la cohorte :

Population source

Une fois la population éligible identifiée, il faut pouvoir identifier une source de recrutement des sujets. Pour les cohortes hospitalières, cela est assez simple puisque les patients peuvent être identifiés lors de leur venue à l'hôpital. Pour d'autres populations, l'opération peut s'avérer plus compliquée. On peut se baser sur des **bases de données administratives ou médico-administratives** (système d'assurance maladie, système d'information hospitalier, instituts nationaux de statistiques, etc.) qui ont parfois l'avantage de posséder en routine des enregistrements sur certaines expositions d'intérêt. De nombreuses études de cohorte ont été constituées pour étudier les effets de différentes expositions professionnelles. Dans ces études, la population a souvent été identifiée à partir des données des entreprises ou des syndicats. Si on s'intéresse à une population présentant des co-morbidités particulières, il peut être intéressants de se baser sur des registres de morbidité (ex : registre des cancers, registres de malformation congénitales).

Échantillonnage

Si l'on dispose d'une base de données ou d'un registre peu volumineux, il peut être intéressant de prendre en compte l'exhaustivité de la population. Cela limitera l'erreur aléatoire et les biais de sélection. Si la population est trop volumineuse, un grand nombre de méthodes d'échantillonnage aléatoire peut être mis en œuvre pour permettre des estimations valides.

Si on ne dispose pas initialement de bases de données, l'échantillonnage aléatoire est plus difficile, mais pas impossible. Dans de nombreuses études de cohorte, cependant, un échantillon de convenance est utilisé, limitant la validité des résultats.

Suivi

L'intérêt principal des études de cohorte est la possibilité de suivre dans le temps les expositions. Il est donc important d'en définir les modalités : quand et où les sujets doivent-ils se rendre au rendez-vous de suivi ? Quelles sont les variables d'intérêt et comment doit-on les mesurer ? Les réponses à ces questions doivent être définies avec précision en amont.

L'un des problèmes majeurs des études de cohorte étant le **biais lié aux perdus de vue**, un souci particulier doit être apporté aux moyens facilitant d'adhésion des sujets à l'étude.

Mesures d'occurrence et d'association

Un autre intérêt des études de cohorte est qu'elles permettent d'estimer directement la **proportion d'incidence** de la maladie (ou **risque** de la maladie). Le calcul d'un **risque relatif** et d'une différence de risque sont donc possibles, et ils permettent d'identifier directement le sur-risque lié à l'exposition.

Cependant, le risque relatif n'est pas pertinent en présence d'une attrition importante (perdus de vue, risques compétitifs) ou en cas de durées de suivi différentes entre les sujets. Dans ces cas-là, le calcul de taux d'incidence (ou densités d'incidence) est préféré, et le risque relatif sera alors remplacé par un **rapport de**

taux, et la différence de risque par une différence de taux.

Dans la plupart des cohortes, c'est donc cette dernière approche qui est généralement préférée, et le nombre de sujets à inclure va dépendre du **nombre de personne-temps nécessaire**.

Ajustement

Si les structures de population sont très différentes entre les différents groupes d'exposition, l'utilité de comparer les taux d'incidence bruts est discutable. Afin de prendre en compte ces différences, des **taux d'incidence standardisés** sont souvent utilisés (voir vidéos sur l'espace Madoc du cours sur la standardisation). Les mesures d'association utilisées sont alors des rapports de taux d'incidences standardisés.

Le principe de la standardisation permet également des **comparaisons externes** : ainsi, les données d'incidence de la cohorte pourront être comparées avec les données recueillies en population générale si une telle mesure existe via un **SMR** ou SIR (standardized mortality/incidence ratio) dont le principe est plus ou moins équivalent aux rapports de taux standardisés mentionnés plus haut.

La standardisation est souvent utilisée pour compenser un déséquilibre lié aux structures d'âge et de sexe, mais lorsqu'un grand nombre de facteurs de confusion sont envisagés, il est préférable d'utiliser un **modèle multivarié**. Théoriquement, une régression logistique est possible, mais étant donné les pertes de vue, un **modèle de Poisson** –adapté à la modélisation de l'incidence– ou un **modèle de Cox** (voir cours l'analyse de survie) est souvent privilégié.

Études cas-témoins

Lorsque la maladie d'intérêt est rare, les études de cohorte, même de grande taille, peuvent ne permettre d'identifier qu'un nombre insuffisant de cas. Il est alors plus rentable d'identifier les cas dans un premier temps et de leur trouver des témoins comparables par la suite. C'est ce qu'on appelle une **étude cas-témoins**.

Modalités de constitution d'une étude cas-témoins

Sélection de cas

Les sources de sélection des cas sont proches de celles présentées dans les études de cohortes. Ici, cependant, les registres de morbidité voire de mortalité sont des sources particulièrement intéressantes.

Sélection des témoins

La sélection des témoins est l'un des points les plus compliqués des études cas-témoins. En effet, pour que les témoins soient valides, il faut qu'ils soient issus de la même population sous-jacente que les cas : il faudrait imaginer une cohorte, dans laquelle ont été retenus l'ensemble des cas et un échantillon aléatoire de non-cas (les témoins). Cet exercice est bien souvent délicat et amène à l'apparition de **biais de sélection** par non-comparabilité des groupes.

Deux questions sont importantes à se poser : si l'un des témoins avait été malade, aurait-il été considéré comme un cas ? si l'un des cas n'avait pas été malade, aurait-il été considéré comme un témoin ? Par exemple, les cas de maladies rares dans un gros hôpital peuvent avoir été référés à ce centre alors qu'ils ne font pas partie de son bassin d'attraction classique. Ces cas ne seront donc pas forcément comparables à des témoins issus du même hôpital.

Plusieurs sources sont couramment employées pour sélectionner les témoins : - l'annuaire téléphonique pour des témoins « en population générale » ; - les témoins hospitaliers, issus du même service ou du même hôpital ; - les « proches » des cas (amis, parents, conjoints voire voisin) ; - si les cas sont issus d'une base de données ou d'une cohorte (cas-témoin niché, ou « nested cas-control study »), la sélection de témoins devient plus évidente.

Caractéristiques des témoins

Ratio témoins:cas

Si la plupart du temps un simple ratio 1:1 (un témoin par cas) est suffisant, il peut arriver que ce ratio soit modifié. Ainsi, si le nombre de cas est faible, il peut être nécessaire de compenser en augmentant le nombre de témoins. Néanmoins, l'augmentation de puissance apportée par chaque témoin supplémentaire diminue rapidement et il est souvent dérisoire d'aller au-delà d'un ratio 4:1 (4 témoins par cas).

Nombre de groupes témoins

Étant donné la difficulté de constituer un groupe témoin convenable, il peut être tentant de constituer plusieurs groupes. Ainsi, si les résultats sont cohérents dans les différents groupes, leur validité est augmentée. Cependant, cette pratique mène souvent à des résultats discordants et les conclusions deviennent alors difficiles. Il est donc recommandé de ne constituer qu'un groupe témoin, choisi avec soins.

Appariement

Pour augmenter la comparabilité des cas et des témoins, on peut réaliser un **appariement** : les témoins seront choisis pour qu'ils présentent la même proportion d'exposition à certains facteurs de confusion que les cas. Cet appariement peut être fait de façon proportionnel ou individuel (ex : chaque cas fumeur sera apparié à un témoin fumeur, etc.). L'inconvénient de l'appariement est qu'il peut rendre les groupes trop semblables : le manque de variabilité entre les groupes va alors empêcher la mise en évidence de différences (sous-estimation des effets). C'est ce qu'on appelle le **sur-appariement**. Ce phénomène survient également en l'absence d'appariement réel, lorsque témoins sont choisis parmi les proches des cas par exemple. Pour cette raison, il est préférable de ne pas utiliser d'appariement et de prendre en compte les facteurs de confusion par des méthodes d'ajustement, sauf quand un facteur de confusion est difficilement mesurable (ex : génotype dans les études sur jumeaux, où chaque paire fonctionne comme un couple cas-témoin).

Mesures

Le ratio cas/témoin étant choisi par protocole, la notion d'incidence n'est pas pertinente dans ce type d'étude. **On ne peut donc pas calculer directement les risques.** Il est toutefois montré que l'odds-ratio (OR) de la maladie en fonction de l'exposition est équivalent mathématiquement à l'OR de l'exposition en fonction de la maladie. Comme dans un cas témoin, l'incidence de l'exposition n'est pas modifiée, on peut donc calculer un OR pour nous apporter des informations sur le risque lié à l'exposition. Pour prendre en compte de potentiels facteurs de confusion, on réalise généralement une **régression logistique**.

		Maladie	
		+	-
Exposition	+	80	50
	-	20	50

$OR_{\text{exposition}} = (80:20)/(50:50) = 4$

$OR_{\text{maladie}} = (80:50)/(20:50) = 4$

Figure 2: Équivalence des odds-ratios

Conclusion

Les cas-témoins et les cohortes présentent des avantages et des inconvénients différents. Les études de cohorte permettent d'estimer directement le risque du phénomène étudié, de mieux mesurer l'exposition et de limiter le risque de biais de sélection, hormis celui lié aux perdus de vue. Les études cas-témoins sont quant à elles plus simples et économiques et sont adaptées aux maladies rares. Dans tous les cas, ces deux schémas d'études sont extrêmement courants en épidémiologie et en recherche clinique, puisqu'ils sont la base des études étiologiques mais également des études pronostiques et des études d'évaluation des stratégies diagnostiques et de dépistage.