

CHAPITRE XII

LES POINTS CLE DE LA VALIDITE INTERNE D'UNE ETUDE BIAS ET FACTEURS CONFONDANTS.

P Duhaut, J Schmidt

Toutes les méthodes de recherche en épidémiologie clinique (ou en médecine basée sur les faits) consistent à essayer d'établir les principes gouvernant notre pratique clinique sur des faits observés, quantifiables, mesurés, vérifiés, plutôt que sur des affirmations théoriques, de l'empirisme simple, voire des croyances médicales répandues ayant, d'une façon ou d'une autre, pris racine dans notre savoir ou nos attitudes. Cet effort a sans doute permis des progrès notables. Aucune méthode d'investigation ou d'observation n'est cependant parfaite, et la littérature regorge de résultats contradictoires ou différents. L'épidémiologie clinique, en essayant de dégager des tendances, n'établit pas de vérité 'dure', comme la découverte de l'ADN et son rôle dans la transmission du génome. Ce chapitre passe en revue quelques unes des limites d'interprétation des études, qui en fait sont aussi nombreuses que les sujets étudiés.

L'Epidémiologie Clinique, dont les méthodes sont en partie inspirées des méthodes d'Epidémiologie classique, a été développée ces dernières années afin d'aider le clinicien à répondre aux multiples questions relevant de sa pratique quotidienne dans les divers domaines de la prévention, du diagnostic, de la thérapeutique, de l'évaluation du pronostic, de la nosologie... :

Quels sont les facteurs de risque du cancer du colon ? Quelle est la meilleure façon de prévenir les complications de l'hypertension artérielle ? Quelle en est l'incidence ou la prévalence au sein de la population âgée ? Quel est le meilleur traitement de l'infarctus du myocarde pris dans les premières heures ? Quelle est la meilleure stratégie thérapeutique d'un cancer du sein stade II ? Quelle est la démarche diagnostique optimale face à une fièvre au long cours ? Quelle est la sensibilité et la spécificité des tests sérologiques de deuxième génération de l'hépatite C ?...

Chaque médecin, en fonction de son expérience, de ses habitudes thérapeutiques, de son école de formation, peut avoir une opinion fondée sur des bases lui paraissant raisonnables... Opinion souvent différente de, voire parfois diamétralement opposée à, celle d'autres médecins expérimentés.

Les études cliniques cherchent à résoudre ces interrogations ou ces contradictions en quantifiant les faits sur des bases reproductibles à partir d'un nombre de patients suffisamment important pour obtenir une réponse statistiquement significative. Le résultat doit -en principe- être applicable au prochain patient atteint de la même pathologie.

Les outils méthodologiques de l'analyse épidémiologique et statistique ont été développés dans ce but. Force nous est de constater, cependant, qu'il existe entre de nombreuses études apparemment bien conduites des différences de conclusion ou des contradictions aussi criantes que celles existant entre les opinions individuelles.

Les origines de ces différences sont multiples.

Elles relèvent tout d'abord d'un possible manque de précision d'une étude correctement construite.

Elles peuvent correspondre ensuite à la présence de vices cachés, appelés biais, plus ou moins inhérents à la structure des études en question. Les biais conditionnent la validité interne de l'étude.

Elles dépendent en troisième lieu des caractéristiques mêmes des populations étudiées : une même maladie peut avoir des causes différentes, et un même traitement peut avoir une efficacité variable, dans différentes populations. Ceci introduit la notion de validité externe, ou capacité à être généralisée, de l'étude.

Elles font appel enfin à la connaissance des limites d'applicabilité de l'outil épidémiologique et statistique à l'analyse et à la compréhension d'un phénomène.

Ces questions résolues -s'il est possible- laissent entier le problème de l'application à un individu de résultats obtenus au niveau d'un groupe. Problème particulièrement important pour le clinicien, qui souvent doit se référer aux études de groupe pour déterminer la cause de la maladie d'un individu, en choisir le traitement ou en inférer le pronostic... et paradoxe des études de groupe, dont le but premier et ultime est souvent d'aider le clinicien à répondre aux questions qu'il se pose face à l'individu

II- NOTION DE PRECISION. ERREUR LIEE AU HASARD :

Imaginons un essai randomisé cherchant à comparer un protocole A et un protocole B de chimiothérapie dans les cancers épidermoïdes du pharynx (excluant les cancers peu différenciés liés au Virus d'Epstein-Bar du nasopharynx), conduit dans deux unités de cancérologie différentes I et II.

Supposons ces deux unités de cancérologie ouvertes à tout patient susceptible de présenter un cancer ORL, sans distinction aucune de sexe, d'origine, d'âge, de milieu social, d'exposition à un facteur de risque particulier ou de tout autre facteur. Tous les nouveaux patients présentant un carcinome épidermoïde hospitalisés durant la période de réalisation de l'étude ont été inclus et randomisés entre les bras A et B en double aveugle, éliminant en principe tout biais de sélection. L'unité I a inclus 23 malades, et l'unité II 19. Les données, exposées ci-dessous, sont analysées dans chaque service séparément.

Protocole	Unité 1		Unité 2	
	A	B	A	B
Patients inclus	12	11	8	11
Survie à 1 an	6	4	3	5
Survie à 1 an (%)	50%	36%	37%	45%

Ces deux essais, identiques, correctement conçus, sans biais évident, portant sur le même type de malades diagnostiqués de la même façon, semblent donner des résultats contradictoires. Quelles peuvent en être les raisons ?

Une première explication peut être donnée par le hasard d'échantillonnage: quoiqu'il n'y ait pas de biais de sélection au niveau de l'inclusion dans l'étude (puisque tous les patients ont été inclus, et que les deux unités sont ouvertes à tous les patients) ou dans un bras (puisque l'allocation a été randomisée), les patients issus de l'Unité I ou de l'Unité II ne représentent que deux échantillons tirés au hasard de l'ensemble de la population des patients atteints de cancer ORL épidermoïde dans le monde. Ces deux échantillons peuvent par hasard, ne pas

être équivalents entre eux, ou ne pas être également représentatifs de la population globale. Ils peuvent par conséquent, répondre de façon différente à un traitement A ou B sans que cela soit lié au traitement lui-même : la disparité, réelle, observée entre l'Unité I et II dans l'efficacité du traitement correspondra en fait à la différence non connue, survenue au hasard d'échantillonnage, entre les deux groupes.

On peut supposer maintenant que la différence observée ne soit pas réelle, ne soit pas significative. Un petit nombre seulement de patients ont été inclus, et les différences de 14 % dans un sens et de 12% dans l'autre observée dans l'Unité I et II ne tiennent chaque fois qu'à un individu : une survie supplémentaire survenue par hasard dans le bras B de l'unité I ramènerait le succès de B à 45%, et une survie supplémentaire survenue par hasard dans le bras A de l'unité II le succès de A à 50%, homogénéisant ainsi les résultats des deux unités... et faisant disparaître toute différence entre les traitement A et B, de façon concordante dans les deux centres!

Des résultats obtenus à partir d'un petit nombre de patients éveillent un certain scepticisme. Il existe cependant des "garde-fous" statistiques, exprimés sous la forme d'intervalle de confiance, permettant de "quantifier" le hasard dans l'obtention des résultats et par conséquent leur significativité. On admet en général que des résultats sont significatifs lorsqu'ils ont moins de 5% de chance d'être dus au hasard, et on appelle l'erreur toujours possible de conclure à une différence significative n'existant pas réellement, alpha.

En d'autres termes, on demande le plus souvent que l'erreur alpha soit inférieure à 5%.

Oublions maintenant l'unité II, et admettons que 5 patients soient vivants à un an dans le bras B de l'unité I, ramenant ainsi le succès de ce bras à 45%. Il n'y a plus de différence notable entre les traitements A et B. Est-ce à dire que les deux traitements sont d'efficacité (ou d'inefficacité) comparables ? Ou que l'étude a failli à montrer une différence existant réellement ?

Une étude doit inclure suffisamment de patients, être suffisamment puissante, pour avoir une chance acceptable d'arriver à montrer une différence existant réellement. Il existe là encore, un "garde-fou" statistique permettant de quantifier la puissance d'une étude, c'est-à-dire sa chance d'arriver à une conclusion reflétant la réalité que l'on cherche à estimer : une bonne étude doit avoir au moins 80% de chances d'arriver à résoudre la question qu'elle s'était posée, soit une puissance de 80%. L'erreur, toujours possible, de conclure à tort qu'il n'y a pas de différence entre les groupes comparés parce que l'étude a failli à montrer une différence existant réellement, est appelée l'erreur bêta.

Autrement dit, on demande en général que l'erreur bêta soit inférieure à 20%, ou que la puissance soit supérieure à 80%.

Puissance = 1 - bêta.

Hasard d'échantillonnage, et résultats obtenus au hasard de l'erreur alpha ou bêta représentent donc les sources majeures d'erreur liées au manque de précision d'une étude par ailleurs correctement construite et menée. Il est possible, pour diminuer ce risque d'erreur liée au hasard, d'augmenter la taille de l'échantillon étudié : on augmentera ainsi la puissance de l'étude, on augmentera la significativité des résultats, et on améliorera les chances que l'échantillon soit véritablement représentatif de la population de patients atteints de la maladie que l'on se propose d'étudier.

Autrement dit, augmenter la taille de l'échantillon étudié diminue l'erreur alpha, diminue l'erreur bêta, et diminue les erreurs liées au hasard d'échantillonnage.

Une revue systématique des essais effectués en anesthésie a ainsi montré qu'en 2000, seuls 56 % des essais randomisés avaient une puissance suffisante. Le score s'est amélioré à 86 % en 2006, mais seuls 18% des essais à résultats négatifs analysaient leur risque d'erreur beta¹.

Cette étude reflète assez bien la tendance générale, et il est rare que la puissance d'une étude soit considérée avant de conclure, peut-être à tort, à l'absence d'effet.

II- LES BIAIS :

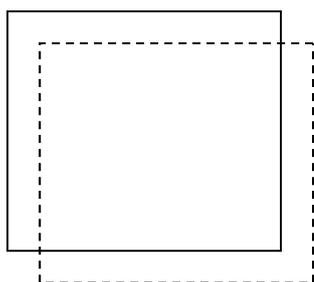
Contrairement aux erreurs précédentes, les biais sont à l'origine d'erreurs relatives à de véritables vices de forme de l'étude, apparaissant à l'une quelconque des étapes de sa conception. Un biais est - en principe - un travers que l'on doit s'efforcer de réduire au minimum ou de faire disparaître - cela n'est pas toujours possible -, avant de débiter l'étude sur le terrain. Un biais n'est pas dû au hasard.

Les vices de forme sont nombreux, et certains types d'études prédisposent plus particulièrement à tel ou tel type de biais. Olli Miettinen a proposé une classification en trois grandes catégories :

- 1- Les biais de sélection.
- 2- Les biais de mesure, ou d'information, ou de mauvaise classification.
- 3- Les éléments confondants.

A- Les biais de sélection :

Les biais de sélection surviennent lorsque la population effectivement étudiée n'est pas représentative de la population que l'on voulait étudier, et à laquelle on voudrait pouvoir appliquer les résultats.



Trait plein : population que l'on veut étudier
Trait pointillé : population effectivement atteinte

Plusieurs facteurs peuvent concourir à produire un biais de sélection, qui peut être suffisamment important pour entacher la validité des résultats :

Les critères d'inclusion et d'exclusion peuvent ne pas avoir été adéquats. Imprécis, laissant une trop grande incertitude quant à leur interprétation et permettant ainsi, dans le cadre d'une étude multicentrique, l'inclusion d'une population hétérogène et différente selon les centres participants. Trop restrictifs ou trop précis au contraire, excluant de l'étude un grand nombre

de sujets et mettant ainsi en péril, la validité externe de l'étude : les résultats ne pourront être appliqués au patient tout venant atteint de la pathologie étudiée, car trop différent du "modèle idéal" sur lequel l'étude a été réellement basée.

Dans une étude cas-contrôle, la sensibilité et la spécificité des tests ou des critères diagnostiques de la maladie étudiée sont déterminants pour la sélection des cas et des témoins: trop sensibles, ils aboutiront à l'inclusion, dans le groupe des cas, de patients non atteints de la maladie étudiée. Trop spécifiques, ils aboutiront à l'exclusion de patients réellement atteints de la maladie. Une bonne spécificité est cependant préférable à une bonne sensibilité pour la validité de l'étude ².

La sélection des témoins dans une étude cas-contrôle représente un élément très important de la construction de l'étude, dont les résultats sont basés sur la comparaison du groupe de patients et du groupe de témoins. Le choix idéal des témoins est réalisé par un tirage au sort dans la population dont sont issus les cas (Cf. Etudes cas-témoins, p. x-y). Ceci n'est pas toujours possible, et de nombreuses études recrutent les témoins parmi la famille, les amis ou le voisinage des cas.

Le biais de non-réponse peut représenter un biais de sélection important, alors même que les étapes précédentes (choix des critères d'inclusion et d'exclusion, choix de la population témoin dans une étude cas-contrôle) ont évité les embûches. Il peut survenir lorsqu'un certain nombre de patients pressentis pour entrer dans l'étude ne répondent pas au courrier ou à l'appel téléphonique les y invitant, ou refusent de participer à l'étude après avoir été informés.

Les sujets non-répondeurs ou refusant de participer à l'étude peuvent être différents des sujets finalement inclus, qui dès lors ne seront plus représentatifs de l'ensemble de la population initialement ciblée. Certains auteurs se sont attachés à étudier les sujets non-répondeurs, et ont effectivement mis en évidence des différences socio-démographiques suffisamment importantes pour modifier les résultats de l'étude entreprise ³.

Les biais de sélection possibles peuvent être évidents ou plus cachés :

Exemples :

- Le taux de réponse au questionnaire de Richard Doll sur la prévalence de la consommation excessive d'alcool et ses complications potentielles dans la cohorte des médecins britanniques a été de 73%, versus 95% pour les questionnaires s'intéressant aux complications du tabagisme. La réduction de 22 % correspond sans doute à un biais de sélection important, les non-répondeurs pour cette question socialement sensible étant sans doute différents des répondeurs. Dès lors, la mesure des complications du facteur de risque est biaisée, et les résultats de l'étude difficiles à interpréter, du moins sur le plan quantitatif ⁴.

- Le biais de survie sélective constitue un autre biais de sélection possible : un traitement ne pourra s'appliquer, bien sûr, qu'à des patients vivants. Ceux-ci, dans la mesure où le temps 0 du début d'une maladie est rarement connu, peuvent ne constituer que la fraction survivante de la population initialement atteinte, et donc la fraction de meilleur pronostic. Une étude récente a cherché à quantifier ce biais dans l'expression des résultats de la prise en charge de patients admis à l'hôpital pour infarctus du myocarde : elle a pu montrer que le biais de survie augmentait avec le délai d'admission à l'hôpital, et que 'l'efficacité' mesurée d'un traitement objectivement inefficace pouvait, en réduction de mortalité, passer de 4 à 27 % si le caractère changeant de la maladie en fonction du temps passé était ignoré ⁵ : les survivants, par définition, ont un meilleur pronostic que patients décédés rapidement, et ce meilleur pronostic peut à tort, être mis sur le compte du traitement. Ce biais peut survenir aisément dans toute étude cas-témoins ou de cohorte : ainsi, la mortalité dans la 'Cardiovascular Health Study' (CHS), pourtant basée sur la population, était inférieure de 40 % à celle observée dans plusieurs cohortes de patients enregistrés dans des programmes

d'assurance-maladie de type Medicare aux Etats-Unis, *alors que les patients de la CHS étaient en moyenne plus âgés et que la proportion d'hommes y était plus importante* ⁶. La randomisation d'une population particulière pour un essai thérapeutique ne le fait pas disparaître, et divers biais de sélection peuvent toujours influencer les résultats de l'essai ⁷.

Lorsque l'inclusion de témoins par tirage au sort n'est pas possible, il vaut mieux, d'une manière générale, préférer les contrôles recrutés dans le voisinage du patient aux amis directs : ils sont plus susceptibles d'être issus de la population dont sont issus les cas, sans pour autant partager toutes les habitudes ou facteurs de risque des cas comme pourraient le faire les amis (notamment en terme de consommation tabagique ou alcoolique)⁸.

Le biais de détection est un autre exemple de biais de sélection, souvent très difficile à mettre en évidence ou à quantifier.

Exemples :

- Un biais de détection possible a été longuement discuté dans les diverses études cas-témoins s'intéressant à la relation entre oestrogénothérapie substitutive de la ménopause et survenue d'un cancer de l'endomètre (Cf. Chapitre Etudes cas-témoins, pages XX-XY)⁹, avec des résultats contradictoires entre études montrant pour le moins qu'une étude construite pour éviter un biais peut être prise à son propre piège ¹⁰.
- L'utilité du dépistage du cancer du poumon a été longtemps débattue, et les principales études menées en Europe ou aux Etats-Unis basées sur la réalisation d'une radiographie pulmonaire simple par an n'ont pas montré d'avantage net en terme de survie des patients dépistés. Une étude récente effectuée au Japon, avec réalisation de scanners thoraciques systématique, semble mettre en évidence une amélioration de la survie des patients dépistés. Deux questions non résolues restent posées : celle d'un biais de détection précoce du cancer allongeant tout simplement la durée d'observation des patients, et donc la survie observée, et celle d'un biais de détection tout court de petits cancers moins agressifs non diagnostiqués sur la radiographie pulmonaire simple, et dont le pronostic, potentiellement meilleur, modifie les données de survie auparavant collectées sur des patients plus graves ¹¹.
- L'administration de finastéride, traitement de l'hypertrophie bénigne de la prostate, a été associée dans un essai randomisé à une diminution de 25% par rapport au placebo de la prévalence de cancer prostatique objectivé par biopsie. Par contre, la prévalence de cancer de haut grade sous finastéride était supérieure à celle sous placebo. En essayant d'en comprendre les raisons, les auteurs ont montré que la distribution du PSA dépendait du volume prostatique avec une diminution de la surface sous la courbe ROC (voir chapitre 'Evaluation d'un test diagnostique') associée à l'augmentation du volume prostatique, et ceci en fonction du grade du cancer : autrement dit, les PSA étaient plus performants dans le diagnostic de cancer de haut grade sur des prostates de petit volume, par comparaison aux prostates augmentées de volume, et l'augmentation apparente de cancer de la prostate de haut grade dans le groupe finastéride était en partie au moins, liée à leur détection facilitée par la diminution de volume de la prostate...elle-même en relation avec la prise de finastéride¹².

Un biais de détection peut donc, en fonction de la pathologie étudiée et des conditions d'étude, jouer dans le sens du diagnostic de pathologies moins graves avec amélioration apparente, mais artificielle, du pronostic, ou du diagnostic de pathologies plus graves avec aggravation apparente du pronostic.

Chaque étude peut, en fonction de sa structure et de la pathologie impliquée, générer ses propres biais de sélection et en produire de nouveaux, non encore catalogués. Leur recherche doit être systématique dans la construction de l'étude, et systématique encore lors de la lecture des résultats.

B- Les biais de mesure :

La sélection des patients (et des contrôles le cas échéant) étant réalisée de la façon la moins biaisée possible, peuvent survenir des biais de mesure encore appelés biais d'information ou de mauvaise classification : l'objet d'étude (le facteur de risque dans une étude cas-témoins, l'incidence d'une maladie dans une étude prospective en cohorte, l'effet d'un traitement dans un essai randomisé, la prévalence d'un facteur de risque ou d'une maladie dans une étude horizontale) a été mal mesuré. Les raisons peuvent en être multiples.

Le diagnostic d'une maladie, la mesure d'un effet, la détermination d'un facteur de risque dépendent étroitement de la sensibilité et de la spécificité de la méthode mise en oeuvre pour les reconnaître (test diagnostique, examen complémentaire, échelle qualitative ou quantitative, qualité de la question posée à l'interrogatoire).

Exemples:

- Le cancer colorectal compte parmi les premiers cancers affectant les deux sexes, et le dépistage de polypes ou de lésions de petite taille, si possible par des moyens non invasifs, devrait permettre d'éviter l'évolution vers les cancers invasifs. La vidéocapsule dans un essai récent a été comparée à la colonoscopie conventionnelle. La Se pour des polypes de plus de 6 mm de diamètre n'a pas excédé 64 %, et la Sp, 84 %, et des chiffres similaires ont été retrouvés pour les adénomes avancés. La vidéocapsule appliquée sur de larges populations conduirait pour l'instant à de nombreux faux positifs et faux négatifs, et donc, si l'on voulait comparer le pronostic d'une population dépistée et non dépistée, à un biais vers le nul par 'mixage' des deux populations¹³.
- L'examen anatomo-pathologique est habituellement considéré comme le 'gold standard', ou étalon-or, pour le diagnostic de lésions bénignes ou malignes : nous n'avons pas mieux pour l'instant que le microscope, éventuellement aidé de méthodes de marquage ou de biologie moléculaire, pour le diagnostic étiologique d'une lésion. Quatre anatomo-pathologistes spécialistes en pathologie gastro-intestinale se sont intéressés à leur reproductibilité inter-observateurs sur l'examen de polypes colorectaux, pour trouver que le coefficient de reproductibilité kappa, prenant en compte les diagnostics concordants par simple hasard, ne dépassait pas... 49%¹⁴.

Il faut bien sûr s'entourer d'un maximum de précautions dans l'utilisation des tests diagnostiques, ou des mesures d'exposition. Malgré cela, l'étude précise des tests, y compris des tests étalon-or, met en évidence leurs limites et les résultats des études les utilisant doivent être interprétés avec ces limites présentes à l'esprit. L'étude de la reproductibilité notamment met en évidence les limites de validité d'un test, même lorsqu'il est appliqué par un médecin expérimenté.

Le biais de souvenir auquel sont particulièrement exposées les études rétrospectives est un autre inducteur fréquent de biais d'information.

Exemple : Il a été rapporté que les avortements légaux pouvaient représenter un facteur de risque du cancer du sein. Lindefors-Harris et coll ont comparé deux études cas-témoins, la

première utilisant comme source de données des interviews de patientes et de leur témoin, la seconde recueillant l'information de façon plus objective dans le registre national suédois des avortements légaux. Le risque relatif de cancer du sein en regard des antécédents d'avortement s'est avéré être 1,5 fois plus important dans l'étude basée sur les interviews que dans l'étude utilisant les registres (différence significative). De plus, le rapport entre la sous-estimation des avortements chez les témoins et la surestimation chez les cas (données d'interrogatoire comparées aux données objectives du registre) s'est avéré être égal à 22,4 !

¹⁵.

Les sujets malades ont tendance à vouloir trouver une explication à leur maladie, alors que les témoins sains ont tendance à oublier leur passé médical... Le biais de souvenir peut donc expliquer des odds ratio anormalement élevés, reflétant mal ou peu la réalité et inducteur de conclusions faussées. L'importance de l'erreur au niveau du résultat dépend cependant des conditions des facteurs examinés (facteur de risque dans une étude cas-témoins) et peut être minime lorsque la prévalence du facteur de risque est faible ¹⁶.

Les biais de mesure peuvent se rencontrer sur tous les types de variables étudiées, que ce soit des variables quantitatives de type dosage biochimique (les techniques de dosage ont leur imprécision), des variables semi-quantitatives comme l'appréciation d'un stade de tumeur (I, II, III, ou IV), car il existe des limites de détectabilité des localisations secondaires viscérales ou ganglionnaires, aux variables qualitatives comme certaines données d'interrogatoire (antécédents personnels ou familiaux 'oubliés) ou d'examen anatomo-pathologique ou radiologique (en faveur, ou non, de tel ou tel diagnostic).

Il existe enfin des biais de mauvaise classification liés aux erreurs de manipulation des données, aux erreurs de remplissage de questionnaires, voire aux données fausses fournies sciemment par des investigateurs peu scrupuleux...

Lorsque ce biais de mauvaise classification s'effectue de façon bi-directionnelle (patients du groupe A malencontreusement inclus dans le groupe B, et patients du groupe B malencontreusement inclus dans le groupe A), les deux groupes que l'on cherchait à comparer, que l'on supposait initialement différents, s'uniformisent. La différence disparaît alors artificiellement et peut à l'extrême s'annuler. Il s'agit du "bias toward the null" anglo-saxon. Cette vérité générale peut souffrir quelques exceptions, en cas d'exposition multiple en particulier ¹⁷.

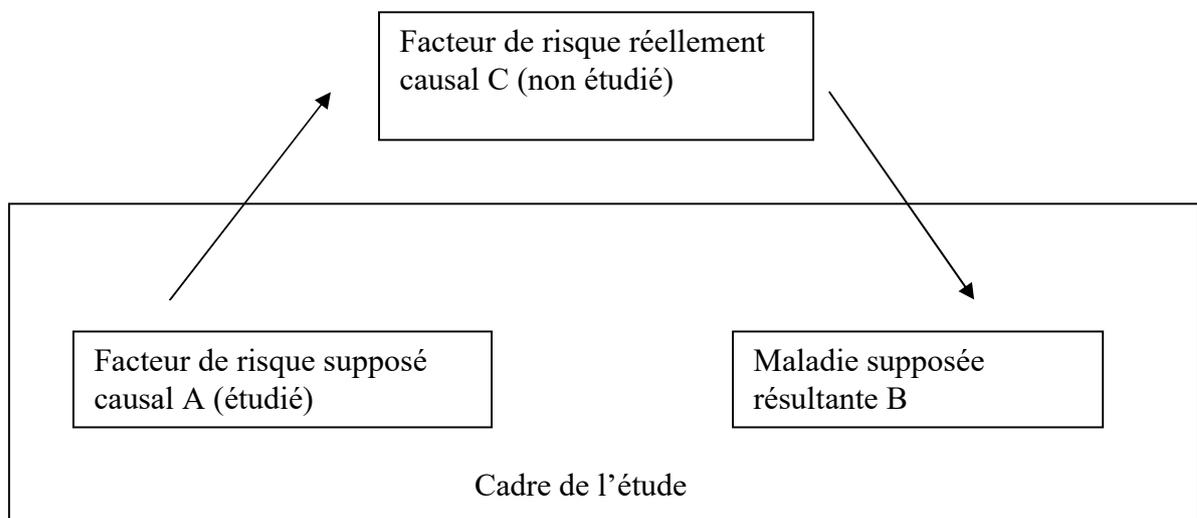
Lorsque le biais s'effectue de façon uni-directionnelle, le résultat dépendra du sens du biais ! Dans une étude rétrospective avec surestimation du facteur de risque dans le groupe des patients et sous-estimation dans le groupe contrôle (cas de figure le plus souvent rencontré), il y aura accroissement de l'odds ratio et par conséquent surestimation du risque... Dans une étude prospective de cohorte où les perdus de vue se rencontreraient essentiellement chez les sujets exposés (exemple : employés exposés d'une entreprise ayant fait faillite, obligés de quitter la région pour des raisons professionnelles), le risque relatif peut être artificiellement abaissé par défaut d'information uni-directionnel sur les perdus de vue (cas de figure en fait inhabituel en pratique).

C- Les éléments confondants :

Indépendants des deux premiers types de biais, le biais de confusion se produit lorsque l'association observée sur le plan statistique ne correspond pas à une réalité biologique, pathologique ou étio-pathogénique, mais s'explique en fait par un troisième facteur, réellement impliqué dans la physiopathologie de la maladie étudiée.

Imaginons une étude de cohorte étudiant l'effet du facteur de risque A sur la survenue de la maladie B. Une étude de cohorte ne peut pas tester de nombreuses hypothèses étiologiques (contrairement à une étude cas-témoins rétrospective), et le facteur de risque A doit donc être fortement suspecté d'être causal avant le début de l'étude prospective. Imaginons maintenant que l'on observe, dans la part de la population soumise au facteur de risque A, une incidence de la maladie B plus élevée que chez les sujets non exposés. En conclure que A est véritablement un agent étiologique semble logique : Le facteur de risque était présent avant que la maladie ne se déclare, et représente justement le facteur distinctif entre les sujets devenus malades et les sujets restés sains.

Ceci est vrai à une condition : qu'il n'existe pas un ou plusieurs facteurs C, ou éléments confondants, non mesurés ans l'étude, présentant d'une part une association statistiquement significative (correspondant à une réalité biologique ou non) avec le facteur de risque A étudié, et réellement impliqués d'autre part dans la genèse de la maladie B.



Le facteur C, réellement causal mais ignoré dans le cadre de l'étude, fait donc apparaître du fait de son association avec le facteur A, ce dernier comme responsable de la maladie B dans une association fallacieuse, statistiquement significative mais biologiquement non fondée.

Conséquence immédiate :

- Le véritable agent étiologique n'aura pas été reconnu malgré le travail de recherche entrepris.

Conséquences possibles :

- Une action d'éradication du facteur A dans le but de prévenir l'apparition de la maladie B peut s'avérer tout à fait inefficace pour peu que l'association entre A et C ne soit que statistique et non pas biologique. Autrement dit, l'élimination du facteur A, si A n'est pas physiquement lié à C, n'entraînera pas l'élimination du véritable facteur étiologique et n'aura pas d'effet sur l'incidence de la maladie.
- Un agent thérapeutique élaboré contre le facteur A dans le but de guérir la maladie B n'atteindra pas l'effet espéré.
- Des programmes de recherche peuvent être lancés sur une fausse piste (situation non exceptionnelle permettant parfois de faire avancer la connaissance!).

Exemple : Le SIDA est une maladie dont l'étiologie est connue, mais la transmission du virus est plus fréquente dans des groupes dits à risque. La prévalence élevée dans la maladie dans le Bronx New-Yorkais a motivé la réalisation d'une étude visant à déterminer les facteurs de risque particuliers des sujets résidents utilisateurs de drogues. La prévalence du SIDA s'est avérée trois fois plus élevée chez les patients de race noire comparés aux malades de type caucasien, et la tentation de conclure à une susceptibilité augmentée à l'infection chez les premiers facile...

En fait, le type de drogue employé n'était pas le même dans les deux groupes pour des raisons économiques. La cocaïne était utilisée préférentiellement par les Noirs, et l'héroïne principalement par les Blancs. La cocaïne du fait de sa durée d'action plus brève nécessitait trois à quatre fois plus d'injections que l'héroïne et le risque potentiel de transmission du virus se trouvait augmenté d'autant¹⁸. Une analyse des données contrôlée par le facteur confondant (type de drogue) aurait permis une conclusion plus précise.

Le biais de confusion peut se glisser insidieusement dans les meilleures études, à différentes étapes de sa réalisation. Ainsi, Vach et Blettner décrivent comment l'utilisation correcte de méthodes de traitement des données manquantes dans les études cas-contrôle peut amener à créer un biais de confusion indépendant de la construction même de l'étude, mais apparaissant lors de la phase d'analyse. Le remède consiste à étudier de façon plus précise, la répartition et le pourquoi des données manquantes dans les différents sous-groupes¹⁹.

Les études réalisées à partir de bases de données informatisées constituées sont sujettes tout particulièrement au biais de confusion. Les bases de données représentent une source potentielle d'information souvent très riche, mais elles n'ont pas été construites en vue d'une étude précise. Elles ne peuvent donc pas prendre en compte les différentes possibilités de biais inhérentes à chaque étude, et n'offrent pas la possibilité d'y remédier.

Exemple : Le rôle des bêta-bloqueurs dans la prévention de la maladie coronarienne chez les patients hypertendus peut être analysé dans le cadre de bases de données déjà constituées rassemblant les prescriptions et la délivrance des médicaments, comme elles existent dans les pays scandinaves. Les bêta-bloqueurs cependant, sont prescrits dans différentes indications y compris le traitement de l'angine de poitrine, et l'intrication des pathologies (HTA, angor, post-infarctus) chez un même patient rend l'analyse particulièrement difficile.

L'utilisation d'informations rassemblées dans un but différent de celui de l'étude que l'on désire réaliser (but économique, évaluation de la consommation médicamenteuse et de son évolution par exemple) est toujours sujette à caution, et les conclusions de l'étude doivent être interprétées avec autant de prudence²⁰.

Correction du biais de confusion :

Il faut, pour que l'élément confondant puisse être éliminé lors de l'analyse des données, qu'il ait été prévu de collecter les informations le concernant lors de la construction de l'étude de telle sorte à ce qu'il puisse être contrôlé. Le contrôle consiste à établir des sous-groupes de patients en fonction du facteur confondant suspecté, et à calculer l'association entre la maladie et le facteur de risque étudié à l'intérieur de chaque sous-groupe.

Exemple, à partir de données fictives inspirées d'articles parus : le risque d'infection par le virus HIV semble augmenté chez les drogués d'un groupe ethnique A, par rapport à un autre (B)

	HIV +	HIV -	
Groupe A	287	163	450
Groupe B	163	287	450
	450	450	

$$OR = \frac{287 \times 287}{163 \times 163} = 3,1$$

Si l'on admet que les sujets du groupe A utilisaient préférentiellement la cocaïne et se piquaient quatre fois plus souvent, le type de drogue utilisé représente alors un élément confondant qu'il faut contrôler : on construit des sous-groupes sur la base de la drogue utilisée :

	Cocaïne		Héroïne		
	Groupe A	Groupe B	Groupe A	Groupe B	
HIV+	275	138	12	25	450
HIV-	25	12	138	275	450
	300	150	150	300	

Puis on calcule les odds ratio pour chaque sous-groupe :

1- Sous-groupe cocaïne : 2- Sous-groupe Héroïne :

$$OR = \frac{275 \times 12}{25 \times 138} = 0,95 \qquad OR = \frac{12 \times 275}{138 \times 25} = 0,95$$

et l'on s'aperçoit que la différence observée initialement disparaît et était à mettre sur le compte, en fait, de la drogue utilisée et du nombre d'injections quotidiennes.

Le contrôle par l'élément confondant lors de l'analyse statistique n'est bien sûr possible que si l'élément confondant potentiel est déjà suspecté lors de la construction de l'étude, et si les données sont recueillies en conséquence !

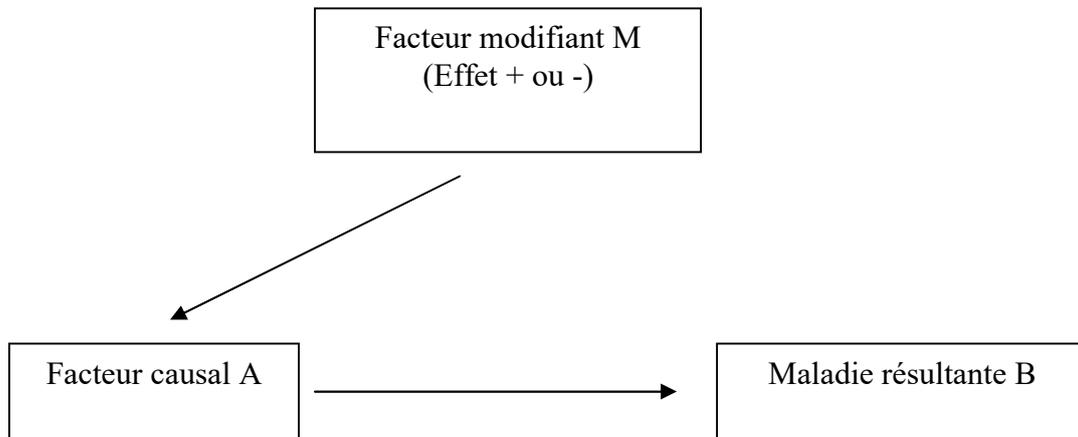
Une autre méthode de contrôle statistique de l'élément confondant, peut se montrer plus puissante car elle ne nécessite pas la division de l'échantillon initial en sous-groupe : il s'agit de l'analyse multivariée en régression logistique, dans laquelle l'élément confondant potentiel peut être introduit comme variable indépendante. Si la plus grande part du pouvoir explicatif revient à l'élément confondant potentiel au détriment du facteur de risque supposé initial, l'élément confondant sera retenu dans le modèle comme significatif, et le facteur de risque supposé initial en verra sa significativité amoindrie, voire annulée.

D- MODIFICATEURS D'EFFET, OU FACTEURS MODIFIANTS :

Nous avons vu qu'un facteur confondant représente un facteur causal véritable, dont l'ignorance dans une étude fait attribuer à tort, la responsabilité causale à un autre facteur mesuré dans l'étude et statistiquement mais non biologiquement associé à la survenue de la maladie.

Il faut bien différencier facteurs confondants et facteurs modifiants : un facteur modifiant est, par définition, un agent qui va modifier la relation existant entre un facteur réellement causal

et la maladie qui en résulte. Il y a interaction entre le facteur causal et le facteur modifiant. Plusieurs cas de figure peuvent se présenter :



1- Le facteur modifiant M ne joue pas de rôle pathogène s'il est isolé, mais accentue le rôle du facteur causal A lorsque A et M sont présents simultanément. Par exemple, le virus delta ne peut à lui seul, provoquer une hépatite. En association avec le virus de l'hépatite B, il renforce le rôle pathogène de ce dernier avec aggravation de l'hépatite et évolution plus fréquente vers la chronicité et la cirrhose. *M seul, ne peut pas provoquer la maladie B.*

2- Le facteur modifiant M ne joue pas de rôle lorsqu'il est isolé, mais atténue le rôle du facteur causal A lorsqu'ils sont associés. C'est le cas des mécanismes de protection génétiques ou acquis contre une affection : le gène de la drépanocytose ne joue aucun rôle dans l'infection palustre. Cependant, sa présence à l'état hétérozygote diminue la sévérité de la maladie due au Plasmodium par diminution de la capacité de ce dernier à parasiter les globules rouges.

3- Le facteur modifiant M joue un rôle dans la genèse de la maladie, mais sa présence associée au facteur causal A ne résulte pas en l'addition de leurs effets séparés, mais en une multiplication d'effets: M et A sont synergiques. Le tabac et l'hypertension sont deux facteurs de risque bien connus des maladies cardio-vasculaires. Leur association aboutit cependant à une multiplication, et non pas une addition, du risque.

4- Le facteur modifiant M joue un rôle dans la genèse de la maladie, mais sa présence associée au facteur causal A diminue son pouvoir pathogène : M et A sont antagonistes. Cette situation est malheureusement assez rare en Médecine, et pourrait illustrer la vieille théorie d'Hippocrate du combat du mal par le mal ? Il semblerait ainsi que l'augmentation du cholestérol au niveau du cerveau observée chez des patients alcooliques antagoniserait en fait, l'action de l'alcool sur les connections neuronales ²¹.

5- L'effet antagoniste peut parfois s'exercer... au détriment d'un bénéfice final. Plusieurs études ont montré une diminution du risque de cancer colo-rectal chez les patientes prenant du Calcium et de la Vitamine D en prévention de l'ostéoporose. Cet effet bénéfique n'a pas été retrouvé dans un essai randomisé, dont un bras comprenait également un traitement par oestrogènes. Il s'avère que dans cet essai, les oestrogènes ont agi comme modificateur d'effet antagoniste sur l'action de calcium et de la vitamine D, pour en annuler l'effet préventif bénéfique ²².

Comment, en pratique épidémiologique, différencier facteur de confusion et facteur modificateur d'effet ?

L'analyse contrôlée par le facteur de confusion (le type de drogue dans l'exemple précédent) permettait d'éliminer la responsabilité d'un facteur présumé causal à tort (le groupe ethnique). L'analyse contrôlée par le facteur modificateur d'effet permettra, de même, d'établir le rôle de chaque facteur causal. Imaginons qu'une étude de cohorte sur les méfaits du tabac en pathologie ORL donne les résultats suivants après un suivi de 10 ans :

	Tabac +	Tabac -
Cancer ORL +	200	60
Cancer ORL -	600	740
	800	800

S'agissant d'une étude de cohorte, nous calculons les incidences cumulées sur 10 ans des cancers ORL dans le groupe de patients tabagiques et le groupe de patients non tabagiques :

Groupe tabac : Incidence = $200/800 = 25\%$

Groupe sans tabac : Incidence = $60/800 = 7,5\%$

L'alcool étant connu comme facteur de risque des cancers ORL, les coordinateurs de l'étude avaient prévu d'en mesurer la consommation, de pair avec celle du tabac, chez tous les membres de la cohorte. Une analyse contrôlée par le facteur alcool est effectuée, qui donne les résultats suivants:

	Alcool +		Alcool -	
	Tabac +	Tabac -	Tabac +	Tabac -
Cancer ORL +	160	40	40	20
Cancer ORL -	240	360	360	380
Total	400	400	400	400
Incidence	40%	10%	10%	5%

Nous voyons tout d'abord que de contrôler par la variable alcool ne fait disparaître la relation entre tabac et cancer : l'incidence des cancers ORL chez les sujets tabagiques est toujours plus élevée que chez les sujets non tabagiques. L'alcool n'est donc pas un facteur confondant de la relation tabac-cancer, qui existe réellement.

Nous voyons ensuite que le groupe (tabac - alcool -) a un risque de cancer ORL de 5%, les groupes (tabac + alcool -) et (tabac - alcool +) un risque équivalent de 10%, et le groupe (tabac + alcool +) un risque non pas de 20% (somme des risques séparés de l'alcool et du tabac), mais de 40%. Nous sommes dans la situation 3, où l'alcool est à lui seul un facteur de risque de la maladie (car le groupe sans alcool a un risque plus faible que le groupe avec alcool) et un facteur modifiant le risque toxique du tabac en l'augmentant.

Nous remarquons enfin que l'incidence brute de 25% dans le groupe (tabac+) est comprise entre l'incidence de 40% dans le groupe (Tabac+ Alcool+) et l'incidence de 10% dans le groupe (tabac+ Alcool-). De même, l'incidence brute de 7,5% dans le groupe (Tabac -) est comprise entre l'incidence de 10% dans le groupe (Tabac - Alcool +) et l'incidence de 5% du

groupe (tabac - alcool -). Cette relation, que l'on peut résumer comme suit, est caractéristique de rôle de facteur modifiant de la variable M par laquelle est faite le contrôle dans l'analyse :

Incidence groupe M - < incidence brute < Incidence groupe M +

Lorsque le rôle exact (facteur confondant ou facteur modifiant ?) d'une variable dans la genèse d'une maladie est inconnu, l'analyse contrôlée par cette variable permet d'une part de reconnaître son caractère, et d'autre part, dans le cas d'un facteur modifiant, de quantifier le sens (synergie ou antagonisme ?) et l'ampleur exacte de la modification d'effet d'un facteur causal.

Dans notre exemple, tabagisme ou alcoolisme pris isolément multiplient le risque de cancer ORL par 2 (10% versus 5%), et, associés, par 8 (40% versus 5%).

E- COMMENT INTERPRETER UNE ETUDE EPIDEMIOLOGIQUE ?

Les biais structurels des études cliniques que nous venons de détailler expliquent sans doute nombre de résultats divergents observés dans la littérature. D'autres facteurs cependant ne doivent pas être oubliés :

E1- La qualité de l'analyse statistique :

Le type d'analyse statistique, les tests employés, l'application aux questions posées doivent être définis dès la construction de l'étude. Cela permet d'une part de recueillir les données sous la forme adéquate à l'analyse, évitant ainsi les manipulations inopportunes, et d'autre part de tester de façon rigoureuse les hypothèses initiales en évitant la multiplication des calculs effectués au hasard des données. En d'autres termes, la multiplication des comparaisons, des calculs de test d'association expose à découvrir et à décrire comme statistiquement significatives des différences ou des associations pour lesquelles le coefficient de significativité n'est significatif que par hasard.

Exemples :

- Certains logiciels de statistique permettent d'établir des matrices de corrélations entre toutes, ou une partie, des données collectées. Il s'agit de corrélations statistiques, ne correspondant pas obligatoirement à une réalité biologique ou épidémiologique. Si l'on accepte un coefficient d'erreur alpha de 5% et si l'on effectue au hasard 100 tests de corrélation, 5 de ces tests peuvent s'avérer faussement significatifs car correspondant aux 5% d'erreur admise. Le risque est grand ensuite, si l'on ne publie que les 5 résultats revenus positifs sans indiquer qu'ils ont été observés du fait du hasard dans un lot de 100 tests, de ne porter à la connaissance des lecteurs qu'un artefact statistique. Il s'agit de la publication de résultats ne répondant pas à une hypothèse de recherche pré-établie.
- Tous les logiciels rendent l'analyse statistique très facile et rapide, par rapport au temps nécessaire aux calculs effectués manuellement... Et donc, ils exposent tous au risque de tests multiples sans rationnel pré-établi, et au risque de valeurs de p significatives juste... par hasard.

L'erreur alpha communément admise à 5% relève d'un choix arbitraire. Le risque au terme d'une analyse statistique bien conduite d'obtenir un résultat positif par chance ou hasard subsiste toujours, et il suffit parfois de l'adjonction de quelques cas supplémentaires à une

série pour qu'alpha passe à 6 ou 7% et que les résultats soient tenus comme non significatifs...
Abaisser alpha augmente la fiabilité des résultats.

E2- Les études non publiées :

Une étude à résultats négatifs a généralement moins de chance d'être publiée qu'une étude à résultats positifs (les grandes études multicentriques mises à part). Il existe donc dans la littérature un biais vers les résultats positifs, particulièrement important sur le plan pratique pour les études de thérapeutique alors qu'il serait important de savoir que l'efficacité d'un traitement n'est pas retrouvée de façon régulière dans différents centres investigateurs. Le biais de publication est un biais invisible s'il n'existe pas de registre de l'ensemble des études réalisées, publiées ou non. Lorsqu'un tel registre existe, comparer les résultats publiés aux résultats effectivement obtenus, mais parfois non publiés, met en évidence une différence significative et une influence manifeste du biais de non-publication²³.

E3- Toute étude doit être interprétée dans le contexte plus large des connaissances déjà acquises :

B. A. Hill proposait en 1965, 9 critères d'interprétation d'une étude épidémiologique²⁴ :

1- Force de l'association observée : une association a d'autant plus de chances d'être réelle que son amplitude mesurée est plus importante. Des risques relatifs inférieurs à 2 sont faibles.

2- Reproductibilité : un résultat sera d'autant plus fiable qu'il est retrouvé, comme en biologie, par différents investigateurs.

3- Spécificité : une relation de causalité spécifique entre un facteur de risque et une maladie donnée est un argument supplémentaire pour admettre la responsabilité du facteur. Exemple : BK et tuberculose, alors que l'on hésitera à reconnaître comme pathogénique d'une lésion pulmonaire des germes retrouvés dans les crachats... Un facteur de risque peut cependant intervenir dans différentes maladies.

4- Temporalité : établir que la cause supposée précède l'effet constaté représente un argument de poids !

5- Gradient biologique : l'augmentation ou la diminution de l'incidence d'une maladie sous l'influence de l'augmentation ou la diminution de la fréquence ou de l'intensité de l'exposition, rend la responsabilité pathogénique de l'exposition plus probable.

6- Plausibilité : sa définition souffre de toute évidence de nombreux biais, mais le bon sens reste utile même dans les démarches scientifiques les plus avancées !

7- Cohérence : des résultats seront plus vraisemblables s'ils s'intègrent dans le cadre des connaissances acquises comme solides sans trop les bouleverser... Il y a bien sûr des exceptions historiques notables.

8- Expérimentation : elle n'est pas toujours possible en Epidémiologie, mais peut venir conforter des données observées sur le terrain.

9- Analogie : ce critère a lui aussi ses limites, mais la ressemblance d'un phénomène observé avec d'autres phénomènes déjà connus peut accroître sa crédibilité.

D4- Application des résultats d'une étude à un individu :

Les études de groupes sont nécessaires afin de mieux cerner la réalité d'ensemble d'une pathologie ou d'un traitement, impossible à appréhender à partir d'un seul ou de quelques patients. Elles permettent d'observer au niveau du groupe des tendances générales, et l'on appliquera les conclusions au niveau de groupes avec des résultats raisonnablement prévisibles *car escomptés sur les données recueillies au niveau d'un groupe* :

Exemple : La vaccination anti-variolique a permis d'éradiquer la maladie de la surface du globe, donc de réduire l'incidence à 0/an/million d'habitants.

Corollaire: les individus semblent actuellement protégés de façon définitive contre la maladie au prix cependant d'un certain nombre d'encéphalites vaccinales graves, survenues chez des patients qui peut-être n'auraient jamais été atteints de la maladie : *le bénéfice individuel peut être dissocié du bénéfice de groupe*.

En pratique, le clinicien est souvent confronté à des situations où le poids respectif du bénéfice et du risque pour un individu précis est beaucoup plus difficile à apprécier, malgré la réalisation d'études extensives :

Exemple : La plupart des études de la thrombolyse à la phase aiguë de l'infarctus du myocarde ont montré une réduction de la mortalité hospitalière chez les patients traités. Il s'agit donc de résultats confirmés, et d'autant plus fiables que l'amplitude de la réduction est semblable d'une étude à l'autre (environ 25%). Dans l'étude GISSI, la mortalité globale passe de 13% dans le groupe non traité, à 10,7% dans le groupe traité (réduction de 2,3% en valeur absolue, de 18% = $[(10,3 - 13)/13] \times 100$ en valeur relative)²⁵.

Les accidents neurologiques sous thrombolyse surviennent cependant de façon non exceptionnelle dans 0,94 à 1,33% des cas²⁶. Traiter 1000 patients par thrombolyse à la phase aiguë revient à éviter 23 décès durant la période d'hospitalisation, mais à provoquer 10 à 13 accidents neurologiques présumés hémorragiques. Le bénéfice est supérieur au risque et le traitement donc licite, mais le clinicien face à son patient n'a que très peu de moyens de savoir, même en respectant les contre-indications des thrombolytiques, si :

- 1- Son malade fait partie des 23 patients courant un risque mortel en cas de non-administration du traitement;
- 2- Son malade ne fait pas partie des 13 patients courant un risque d'hémorragie cérébrale sous traitement;
- 3- Son malade ne fait pas partie des 977 patients pour lequel le bénéfice du traitement en terme de survie est nul.

L'épidémiologiste peut chercher à mieux définir les facteurs pronostiques, les indications du traitement, les caractéristiques des patients non améliorés. Cependant, son outil méthodologique n'est pas adapté à l'étude particulière de l'individu et il ne peut fournir qu'une orientation générale (précieuse) à moduler par le clinicien en fonction du cas présent. Les arbres de décision cherchent à combler cette lacune en intégrant les particularités propres à chaque patient, mais ils restent d'application difficile et ne peuvent pas être employés de façon routinière.

Les techniques épidémiologiques appliquées à la recherche épidémiologique pure ou à la recherche clinique ont donc, comme toute techniques, leurs limitations tant dans leur champ exploratoire que dans leurs possibilités d'interprétation. Elles n'en demeurent pas moins

indispensables et inégalées jusqu'à présent chaque fois que l'on veut mesurer un phénomène affectant le vivant dans son environnement normal, avec l'ensemble des interactions le caractérisant, en-dehors des conditions artificielles d'isolement d'un laboratoire. Elles restent le seul outil capable de vérifier sur le terrain, les hypothèses engendrées in vitro ou par l'expérimentation. Elles permettent enfin d'attribuer à chaque facteur le poids qui lui revient dans la réalité dans son action pathogénique, physiopathologique ou thérapeutique... mais elles doivent, du fait de la complexité des phénomènes qu'elles survolent et tentent de décrire de façon globale, être maniées avec la connaissance de leurs défauts. Elles sont à la médecine, ce que le solfège est à la musique : nécessaire, mais insuffisant pour interpréter la partition et l'adapter au moment présent.

-
- ¹ - Greenfield ML, Mhyre JM, Mashour GA, Blum JM, Yen EC, Rosenberg AL. Improvement in the quality of randomized controlled trials among general anesthesiology journals 2000 to 2006: a 6-year follow-up. *Anesth Analg* 2009;108:1916-1921.
 - ² - Brenner H, Savitz DA. The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies. *AM J Epid* 1990;132:181-192.
 - ³ - Holt VL, Daling JR, Stergachis A, Voigt LF, Weiss NS. Results and effect of refusal recontact in a case-control study of ectopic pregnancy. *Epidemiology* 1991;2:375-379.
 - ⁴ - Doll R, Peto R, Hall E, Wheatley K, Gray R. Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors. *BMJ* 1994; 309: 911–918.
 - ⁵ - Austin PC, Mamdani MM, van Walraven C, Tu JV. Quantifying the impact of survivor treatment bias in observational studies. *J Eval Clin Pract* 2006;12:601-612.
 - ⁶ - Di Martino LD, Hammill BG, Curtis LH, Gottdiener JS, Manolio TA, Powe NR, Schulman KA. External validity of the cardiovascular health study: a comparison with the Medicare population. *Med Care* 2009;47:916-923.
 - ⁷ - Ioannidis JP, Polycarpou A, Ntais C, Pavlidis N. Randomised trials comparing chemotherapy regimens for advanced non-small cell lung cancer: biases and evolution over time. *Eur J Cancer* 2003;39:2278-2287.
 - ⁸ - Robins J, Pike M. The validity of case-control studies with non-random selection of controls. *Epidemiology* 1990;1:273-284.
 - ⁹ - Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *NEJM* 1978; 299:1089-1094.
 - ¹⁰ - Hulka B, Grimson RC, Greenberg BG, Kaufman DG, Fowler WC, Hogue CJR, Berger GS, Pulliam CC. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. *Am J Epidemiol* 1980;112:376-387.
 - ¹¹ - Sagawa M, Usuda K, Aikawa H, Machida Y, Tanaka M, Ueno M, Sakuma T. Lung cancer screening and its efficacy. *Gen Thorac Cardiovasc Surg* 2009;57:519-27.
 - ¹² - Elliott CS, Shinghal R, Presti JC Jr. The influence of prostate volume on prostate-specific antigen performance: implications for the prostate cancer prevention trial outcomes. *Clin Cancer Res* 2009;15:4694-4699.
 - ¹³ - Van Gossum A, Munoz-Navas M, Fernandez-Urien I, Carretero C, Gay G, Delvaux M, et al. Capsule endoscopy versus colonoscopy for the detection of polyps and cancer. *NEJM* 2009;361:264-270.
 - ¹⁴ - Wong NA, Hunt LP, Novelli MR, Shepherd NA, Warren BF. Observer agreement in the diagnosis of serrated polyps of the large bowel. *Histopathology* 2009;55:63-66.
 - ¹⁵ - Lindefors-Harris BM, Eklund G, Adami HO, Meirik O. Response bias in a case-control study : analysis utilizing comparative data concerning legal abortions from two independent Swedish studies. *Am J Epid* 1991;134:1003-1008.

-
- ¹⁶ - Drews CD, Greenland S. The impact of differential recall on the results of case-control studies. *Int J Epidemiol* 1990;19:1107-1112.
- ¹⁷ - Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746-748.
- ¹⁸ - Schoenbaum EE, Hartel D, Selwyn PA, Klein RS, Davenny K, Rogers M, Feiner C, Friedland G. Risk factors for human immunodeficiency in intravenous drug users. *NEJM* 1989; 321:874-879.
- ¹⁹ - Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad-hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991;134:895-907.
- ²⁰ - Psaty BM, Koepsell TD, Siscovick D, Wahl P, Logerfo JP, Inui TS, Wagner EH. An approach to several problems in using large databases for population-based case-control studies of the therapeutic efficacy and safety of anti-hypertensive medicines. *Statistics in Medicine* 1991, 10, 653-662.
- ²¹ - Crowley JJ, Treistman SN, Dopico AM. Cholesterol antagonizes ethanol potentiation of human brain BKCa channels reconstituted into phospholipid bilayers. *Mol Pharmacol* 2003;64:365-372.
- ²² - Ding EL, Mehta S, Fawzi WW, Giovannucci EL. Interaction of estrogen therapy with calcium and vitamin D supplementation on colorectal cancer risk: reanalysis of Women's Health Initiative randomized trial. *Int J Cancer* 2008;122:1690-1604.
- ²³ - Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977-984.
- ²⁴ - Hill BA. The environment and disease : association or causation? *Proc Royal Soc Med* 1965;58:295-300.
- ²⁵ - Gruppo Italiano per studio della Streptochinasi nell' infarto miocardio (GISSI): Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. *Lancet* 1986;1:397-402.
- ²⁶ - Gruppo Italiano per lo Studio della Sopravvivenza nell' Infarto miocardico : GISSI 2 : a factorial randomized trial of alteplase versus streptokinase and heparin versus non heparin among 12 490 patients with acute myocardial infarction. *Lancet* 1990;336:65-71.