

Oliver Senn

# Comment interpréter les études? EBM – pas le dogme, mais l'outil

Compte-rendu de l'atelier de la SFD Conference tenue les 29 et 30 août 2013 à Berne

**La médecine fondée sur les preuves (evidence based medicine – EBM) est l'intégration de la «best research evidence» dans le processus décisionnel incluant par ailleurs l'expérience personnelle et les souhaits du patient. Les compétences («skills») permettant d'apprécier l'évidence scientifique de manière critique peuvent aider à faire la part des choses dans la masse d'informations paraissant quotidiennement et à séparer le vrai du faux.**

## Introduction

Notre boîte de messagerie (électronique) déborde chaque jour d'innombrables messages vantant de « nouvelles acquisitions scientifiques » ou de « nouveaux traitements plus performants les uns que les autres ». En pratique quotidienne, nous médecins essayons d'être responsables en nous efforçant constamment de transposer les informations reçues à propos du diagnostic, des thérapies, du pronostic et de la prévention à notre prise en charge des patients. Nous avons besoin pour cela d'un certain nombre de compétences et d'une solide expérience clinique. Et au-delà de celle-ci, les échanges au sein de groupes de collègues (par ex. dans le cadre de cercles de qualité) sont absolument précieux lorsqu'il s'agit d'apprécier l'importance pratique des nouvelles données de la littérature. La médecine fondée sur les preuves (EBM) est l'intégration de la «best research evidence» dans le processus décisionnel incluant par ailleurs l'expérience personnelle et les souhaits du patient [1]. Une lecture critique de l'évidence (critical appraisal), en particulier sous l'angle de la validité, de la signification pratique et de la faisabilité, est incontournable et constitue par conséquent une condition nécessaire à la médecine fondée sur les preuves. Le présent article se propose de donner quelques conseils et astuces qui permettront au lecteur de séparer le grain de l'ivraie dans le foisonnement des publications de la littérature médicale.

## La question de la causalité

Imaginez que vous êtes tranquillement assis dans votre salon en compagnie de votre chien, tandis qu'un étranger s'approche de la

maison. Votre chien se met à aboyer. Une question que l'on pourrait se poser dans cette situation: le chien aurait-il aboyé si cet étranger ne s'était pas approché de la maison? Si on attribue à l'animal les qualités d'un chien de garde, la conclusion logique, l'évidence, serait l'existence d'une relation de causalité entre la présence de l'étranger et les aboiements du chien. Pour un esprit sceptique, ces deux événements pourraient cependant tout aussi bien être le fruit du hasard et, pour des personnes victimes précédemment de mauvaises expériences, l'aboiement du chien pourrait parfaitement dénoter d'autres traits de caractère de l'animal. Cet exemple illustre la fragilité du concept de la relation de causalité dans un enchaînement d'événements. Sir Austin Bradford Hill, un épidémiologiste et statisticien anglais, a formulé 9 critères [2] qu'il convient de tester de manière critique lorsqu'on postule une relation de causalité à partir de certaines observations (tableau 1). *On retiendra tout particulièrement pour les études d'observation qu'une seule publication ne fournit pas une évidence suffisamment robuste pour justifier un changement d'attitude en pratique clinique.*

L'appréciation, dans notre activité quotidienne, du pronostic d'un comportement (à risque) ou l'évaluation de l'intérêt d'une mesure thérapeutique nous confronte forcément à la question du rapport de causalité. Les avantages et les inconvénients des designs d'études les plus fréquents vont être brièvement discutés ci-dessous. On ne saurait cependant trop conseiller la lecture d'ouvrages d'épidémiologie plus complets à ce sujet [3, 4].

## Types d'études

Dans le contexte de la recherche quantitative, on distingue fondamentalement les essais expérimentaux des études non-expérimentales (observationnelles). Le gold standard pour la démonstration d'une relation de causalité entre une mesure thérapeutique et le devenir d'une maladie est l'étude (expérimentale) randomisée. L'étude de cohorte et l'étude de cas contrôle constituent d'autres designs fréquemment utilisés, notamment pour explorer l'étiologie et les facteurs de risque de certaines maladies.

**Tableau 1**

Critères de Bradford-Hill selon la référence 2.

Puissance (strength)	L'existence d'un faible degré d'association n'implique pas l'absence d'une relation de causalité.
Consistance (consistency)	Existe-t-il d'autres études dont les résultats concordent avec ces observations?
Spécificité (specificity)	Y a-t-il des éléments suggérant qu'une population spécifique (par ex. fumeurs, activités professionnelles particulières) souffre d'une maladie donnée?
Temporalité (temporality)	La relation temporelle devrait être telle que la conséquence est survenue après la cause supposée.
Gradient biologique (biological gradient)	Existe-t-il une relation dose-réponse (par ex. nombre de paquets-années et risque de cancer)?
Plausibilité (plausibility)	Un mécanisme (biologique) plausible pour le principe cause-conséquence est utile.
Cohérence (coherence)	Existe-t-il des données concordantes entre les résultats d'études épidémiologiques et ceux d'analyses de laboratoire?
Expériences (experiments)	Existe-t-il une évidence expérimentale pour une relation de cause à effets?
Analogie (analogy)	Existe-t-il une évidence selon laquelle les mêmes facteurs de risque produisent les mêmes effets?

### Etude randomisée

Une répartition aléatoire (au hasard - random) d'une population clairement définie, le plus souvent en 2 groupes (interventionnel et contrôle), est l'élément fondamental de l'étude randomisée. Le point essentiel de la randomisation (réalisée avec succès) est que l'attribution à l'un des groupes n'est pas prévisible et que les deux populations sont donc absolument comparables au début de l'expérience. L'idée de ce design est que, lorsqu'on observe des différences entre deux groupes sur le plan de l'évolution de la maladie ou des résultats thérapeutiques, la seule variable distinctive soit le type d'intervention (par ex. verum vs. placebo). L'étude randomisée recèle une grande validité interne, mais souffre souvent de limitations s'agissant de la généralisation dans la pratique quotidienne (validité externe). Cela tient au grand nombre de critères d'inclusion et d'exclusion qu'il faut appliquer [5]. *La recherche en médecine de famille dynamique peut contribuer à une meilleure prise en compte de ce setting particulier (patients âgés polymorbides) dans ce type d'études.* La recherche dans les soins de base contribue en effet à générer l'évidence qu'on ne peut justement pas déduire de l'expérimentation clinique (le plus souvent conduite par l'industrie pharmaceutique).

### Résultats dus au hasard et analyses de sous-groupes

Comme les maladies sont de gravité très variable, il est tentant de rechercher des effets thérapeutiques dans des sous-groupes spécifiques. Il est donc fréquent d'assister dans les études randomisées à la présentation de certains effets spécifiques à certains sous-groupes. *Les résultats se fondant sur des analyses de sous-groupes comportent toutefois un grand danger : les effets observés peuvent très bien relever du hasard, quelle que soit la taille de la population d'origine de l'étude.* L'« International Study of Infarct Survival » (ISIS-2) fournit un bel exemple d'effet aléatoire de sous-groupe [6]. ISIS-2 a démontré de manière impressionnante la supériorité de l'aspirine et de la streptokinase versus placebo en termes de survie post-infarctus aigu du myocarde dans une population de 17'187 patients randomisés. Pour attirer l'attention sur les conclusions ineptes ressortant des analyses de sous-groupes, les auteurs d'ISIS-2 ont réparti les patients en fonction de leur signe astrologique. Malgré l'énorme casuistique, il en est ressorti que les patients du signe astrologique « gémeaux » et « Balance » de la population de l'étude ne bénéficiaient en rien de l'aspirine et que cette dernière pouvait même s'avérer plus délétère que le placebo. Pour les non-astrologues, un tel effet ne peut être dû qu'au hasard.

De tels effets aléatoires sont en partie de nature statistique, puisqu'il existe une relation entre le nombre d'hypothèses testées et la probabilité d'obtenir un résultat significatif (valeur de  $p < 0.05$ ) (figure 1). *Les résultats d'analyses de sous-groupes peuvent aider à formuler des hypothèses (rationnelles), qui devront néanmoins être testées de manière prospective sur de nouveaux échantillons avant de tirer une quelconque conclusion quant à la relation de causalité.* L'évaluation critique des résultats de sous-groupes devrait tenir compte des points suivants:

- y a-t-il un effet (significatif) dans la population globale de l'étude?
- les analyses de sous-groupes avaient-elles été planifiées avant l'étude?
- existe-t-il une explication/hypothèse plausible pour les sous-groupes choisis?
- des précautions statistiques ont-elles été prises (correction de la valeur de  $p$ , analyses d'interactions) pour minimiser le risque de résultats dus au hasard?

Lorsqu'il s'agit de questions sur l'apparition d'une maladie ou les effets (causaux) de facteurs environnementaux, il est souvent difficile si

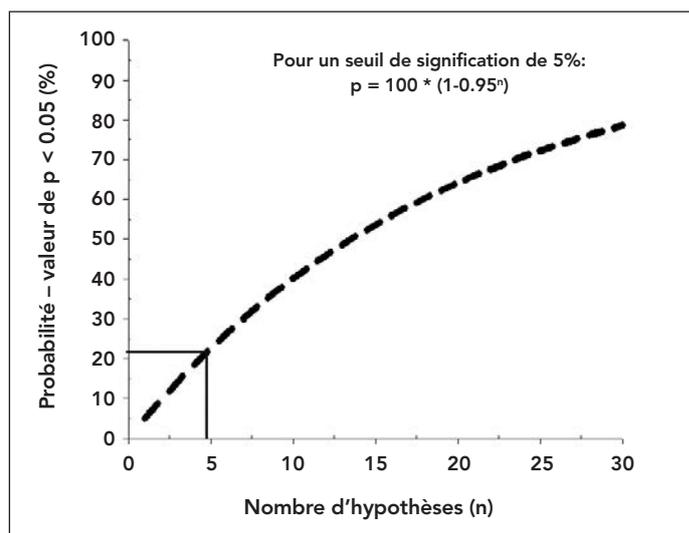


Figure 1

Cette figure illustre la relation entre le nombre de sous-groupes (n) et la probabilité d'obtenir un résultat significatif (valeur de  $p < 0.05$ ). Pour 5 sous-groupes cette probabilité qu'au moins l'un des sous-groupes présente un résultat significatif est par exemple de 22%.

ce n'est impossible de réaliser une étude randomisée. Il serait par exemple impensable, pour des raisons éthiques, de faire des études randomisées sur l'effet de la violence ou d'agents nocifs connus sur l'apparition de maladies physiques ou psychiques. Lorsqu'il est souhaitable d'obtenir des réponses à de telles questions et que l'importance clinique le justifie, il faut obligatoirement opter pour des études non-randomisées (observationnelles). Les études d'observation comportent cependant des sources d'erreurs qu'il faut connaître pour en interpréter correctement les résultats et éviter le piège de rapports de causalité seulement apparents (faux effets).

### Etude de cohorte

L'étude de Framingham est un exemple d'étude épidémiologique qui a contribué de façon remarquable à la définition des facteurs de risques cardiovasculaires [7]. Il s'agit d'une population d'étude représentative saine au moment de la collecte des données (« exposition » à des facteurs de risque cardiovasculaires suspectés à l'époque). Les nouveaux cas de maladies cardiovasculaires (incidence) ont entre-temps été documentés de manière détaillée sur une période de plus de 30 ans. Les incidences des différentes affections dans les sous-groupes de populations exposés et non-exposés ont ainsi pu être comparées et ont permis de tirer des conclusions sur les facteurs de risque causaux.

### Variables confondantes (confounding)

Lorsque nous observons, dans le cadre d'une étude observationnelle, des relations entre une exposition (variable X) et une maladie (variable Y), nous devons toujours nous poser la question de l'existence éventuelle de facteurs perturbateurs (variables confondantes - confounder) simulant une relation de causalité entre X et Y. La relation significative entre « les doigts jaunes » (exposition X) et le cancer du poumon (variable Y) est par exemple le résultat d'un facteur confondant constitué par la fumée de cigarette (en d'autres termes, la relation entre les doigts jaunes et le cancer du poumon fait suite au fait que le tabagisme est un facteur de risque de cancer et que la consommation de cigarettes est associée à des doigts jaunes). *Le lecteur critique d'études d'observation doit vérifier si une éventuelle variable confondante a fait l'objet d'un contrôle adéquat.*

Voici quelques méthodes souvent appliquées pour traiter les variables confondantes:

- distribution équilibrée des cas et des contrôles éventuellement soumis à des facteurs confondants (matching)
- stratification (autrement dit, on regarde si un effet supposé se produit en fonction de l'importance de l'expression d'un facteur confondant suspecté ; pour suggérer que le tabagisme est un facteur de risque indépendant de l'âge pour le cancer du poumon, l'incidence de ce cancer devrait par exemple être systématiquement plus élevée chez les fumeurs que chez les non-fumeurs dans toutes les classes d'âge)
- ajustement (modélisation statistique à l'aide d'une analyse de régression)

Les possibilités de contrôle ci-dessus présupposent que les éventuelles variables confondantes sont connues et documentées (mesurées). *On ne peut donc jamais totalement exclure des variables confondantes résiduelles (« residual confounding ») lors de l'interprétation des études observationnelles. Le « gold standard » pour le contrôle des variables confondantes est la randomisation, qui est seule à même de garantir la répartition équilibrée aléatoire de variables confondantes inconnues à travers les différents groupes au moment de l'étude et donc l'exclusion de distorsions lors des conclusions.*

Une étude de cohorte publiée récemment dans le New England Journal of Medicine sur l'association entre la consommation de café et la mortalité (>300 000 participants) illustre élégamment comment une relation peut se modifier en tenant compte des variables confondantes [8]. Une consommation accrue (>6 tasses/jour) était un facteur de risque de mortalité lorsque contrôlée pour l'âge en tant que principale variable confondante de la mortalité (hazard ratio 1.60; intervalle de confiance 95%: 1.51–1.69). Après la prise en compte de 19 facteurs confondants supplémentaires, l'analyse multivariée a mis en évidence une association protectrice significative entre la consommation de café et la mortalité (hazard ratio 0.90; intervalle de confiance 95%: 0.85–0.96). Les auteurs en ont conclu que la relation inverse observée entre la consommation de café et la mortalité ne permet pas de tirer une conclusion quant à la relation de causalité.

#### Etude de cas-contrôle

Contrairement à l'étude de cohorte qui compare la fréquence d'une maladie dans un collectif exposé et un collectif non-exposé, l'étude de cas contrôle compare les fréquences d'exposition entre des personnes malades (cas) et des personnes en bonne santé (contrôles). Les erreurs de sélection (biais de sélection) et les erreurs d'informations (biais de mémoire ou biais de rappel - recall bias) sont deux types d'erreurs systématiques qui faussent l'influence d'une exposition sur le risque d'une maladie et qu'il s'agit de surveiller attentivement dans les études de cas contrôles.

#### Biais de sélection

Comme dans les essais randomisés, les critères d'influence doivent être clairement définis dans les études de cas contrôles. C'est la seule façon de vérifier si la sélection des cas forme une population représentative de sujets « malades ». Il doit aussi être précisé s'il s'agit de cas incidents (nouvellement diagnostiqués) ou de cas prévalents. Les cas incidents sont en effet plus appropriés pour déterminer les facteurs étiologiques et les cas prévalents conviennent mieux pour caractériser la survie (à long terme). La sélection de contrôles strictement comparables constitue un défi considérable. Le

lecteur vérifiera en particulier aussi si les sujets contrôles recrutés étaient des patients hospitalisés ou non-hospitalisés.

#### Biais de mémoire ou biais de rappel

Un autre problème important inhérent aux études de cas contrôles est la mémoire (recall). La capacité à se remémorer peut différer entre les cas et les contrôles, ce qui est évidemment crucial. Ce genre de distorsion, également appelée « biais de rumination - rumination bias », s'explique par exemple par le fait que les mères ayant accouché d'un enfant souffrant d'une malformation ont davantage tendance à associer des événements survenus au cours de la grossesse avec l'enfant que les mères d'enfants en bonne santé. Ces différences de capacité à se remémorer pourraient expliquer pourquoi on pourrait évoquer dans l'exemple ci-dessus une relation entre les infections prénatales (maternelles) et les malformations chez l'enfant.

#### Conclusions

Si nous nous efforçons, dans notre activité quotidienne, d'intégrer la base d'évidence scientifique dans nos processus décisionnels, nous devons forcément être capables de juger de façon critique la validité et la signification effective des travaux publiés dans la littérature. Il est donc important que les « utilisateurs » connaissent les sources d'erreurs classiques des différents designs d'études, ainsi que les moyens de les contrôler. Un « screening » de ces faiblesses peut nous aider à séparer le grain de l'ivraie. Le Critical Appraisal Skills Program (CASP), développé à l'origine à Oxford, GB, a publié sur son site Internet une checklist simple concernant chacun de ces designs d'études pour faciliter l'appréciation critique systématique de l'évidence scientifique ([www.casp-uk.net](http://www.casp-uk.net)).

L'auteur n'a déclaré aucun conflit d'intérêts en relation avec cet article.

#### Références

- 1 Sackett DL, Richardson WS, Rosenberg W, et al., editors. Evidence-based medicine: how to practice and teach EBM. London: Churchill-Livingstone; 1997.
- 2 Hill AB. The Environment and Disease: Association or Causation? Proc R Soc Med 1965;58:295–300.
- 3 Gordis L, editor. Epidemiologie: Verlag im Kilian; 2001.
- 4 Bonita R, Beaglehole R, Kjellström T, editors. Basic epidemiology: World Health Organization; 2006.
- 5 Boyd CM, Vollenweider D, Puhan MA. Informing evidence-based decision-making for patients with comorbidity: availability of necessary information in clinical trials for chronic diseases. PLoS One 2012;7:e41601.
- 6 Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Lancet 1988;2:349–60.
- 7 Anderson KM, Wilson PW, Odell PM, et al. An updated coronary risk profile. A statement for health professionals. Circulation 1991;83:356–62.
- 8 Freedman ND, Park Y, Abnet CC, et al. Association of coffee drinking with total and cause-specific mortality. N Engl J Med 2012;366:1891–904.

#### Correspondance:

PD Dr Oliver Senn, MPH  
Praxis MediS, Zürich und Institut für Hausarztmedizin  
Universität Zürich  
Pestalozzistrasse 24  
8091 Zürich  
[oliver.senn\[at\]jusz.ch](mailto:oliver.senn[at]jusz.ch)