Lecture critique de la littérature clinique professionnelle

Partie I: Chausse-trapes dans les rapports de cas et les séries de cas

M. F. Fey[a], A. Bührer[b]

Introduction

«Evidence-based medicine» demande que nos décisions cliniques quotidiennes (et si simples qu'elles soient, comme par exemple la décision de déterminer la sédimentation chez un patient oui et chez l'autre non) se base sur l'évidence scientifique [1]. L'occupation avec la littérature professionnelle est ainsi la base indispensable de notre travail quotidien, aussi pour les confrères praticiens et à l'hôpital qui ne participent pas à des programmes de recherche. La littérature professionnelle médicale peut donc être obtenue facilement: Elle prolifère même dans une dimension qui ne peut presque plus être parcourue. Des programmes de recherche de littérature professionnelle dans l'internet, «journaux en ligne», des abstracts de congrès sur CD-ROM et autres choses semblables facilitent la recherche guidée thématique dans la littérature concernant un problème particulier. Ainsi au cabinet médicale et au travail quotidien en clinique toutes les portes seraient ouvertes a l'«evidence-based medicine» malgré le manque de temps? La lecture critique d'un article qui a été découvert en feuilletant ou par «surfing» devient d'autant plus importante à mesure que la récolte de littérature devient plus facile. On peut être consolé en partant du fait que pas tout ce qui est publié en littérature médicale ou qui peut être tiré de l'Internet satisfait aux plus hautes exigences de qualité scientifique. Thématiquement il y aurait à traiter de la littérature sur la pathogenèse de maladies différentes, sur des problèmes diagnostiques et enfin des travaux sur la valeur de mesures thérapeutiques. Nous limitons notre exposé sur la littérature concernant la thérapie surtout avec des exemples oncologiques et nous expliquons quelques principes comment parmi la littérature professionnelle l'ivraie peut être séparée du bon grain.

Tableau 1.
Paramètres associés au risque de ne pas survivre une vogue de chaleur à Chicago.

Paramètres	Victimes de la chaleur	Locataires survécus (voisins)	Odds ratio C'est-à-dire le risque de mourir par la chaleur au cas où le paramètre se trouve dans la colonne 1
Dispositif de climatisation	25%	53%	0,2 (0,2–0,4)
Appartement sous le toit	52%	32%	4,7 (1,7–12,8)
Alité	16%	4%	5,5, (2,5–12,1)
Visites à domicile par les infirmières de la santé publique	20%	5%	6,2 (2,9–13,4)

Les valeurs en % indiquent chez combien de victimes de la chaleur respectivement locataires survécus existait un certain paramètre à l'heure critique. «Odds Ratio» indique le risque de mourir par la chaleur au cas où le paramètre existe ou de survivre au cas où il n'existe pas. Le chiffre de référence est 1,0 et les valeurs de dispersion sont indiquées en parenthèses.

«Post hoc ergo propter hoc»

A l'initiation de nouvelles thérapies la question se pose quelques fois si une mesure thérapeutique influence essentiellement le cours de la maladie ou non. Ceci concerne notamment les maladies qui ont un cours spontanément changeant. Nous pensons au diabète, à l'angiocardiopathie ou au cancer du sein métastasé. Auprès de telles maladies il n'est pas du tout trivial de prouver dans quelle mesure une nouvelle thérapie influence de manière décisive (favorable ou défavorable) le cours – c'est-à-dire de produire la preuve que le cours de la maladie et la thérapie sont dans un rapport de causalité (Tableau 1).

Un exemple:
La ville américaine de Chicago est frappée en été par une vague de chaleur à laquelle succombent régulièrement quelques habitants. L'«Epidemic Intelligence Service» (sic!) du Centre for Disease Control à Atlanta, USA, et des groupes associés s'occupent des facteurs de risque de la mortalité par chaleur en été. Plusieurs paramètres qui pouvait être significatif dans ce contexte ont été comparés entre les victimes de la chaleur et leurs voisins. Les voisins avaient, comme le demandait l'étude, vécu dans les mêmes circonstances pendant le temps critique comme les habitants décédés, mais ils avaient échappé à la chaleur. Des aspects critiques concernant la question de survie étaient p. ex. si leur dispositif de climatisation était en ordre, si un habitant était mobile ou captivé dans son lit (dans une man-

[a] Institut d'oncologie médicale, Hôpital de l'Ile, Berne
[b] Spécialiste FMH en médecine interne, Novazzano

Correspondance
Pr M. Fey
Institut d'oncologie médicale
Hôpital de l'Ile
CH-3010 Berne

E-mail: martin.fey@insel.ch

sarde) à cause d'une maladie. Ces deux paramètres (dispositif de climatisation en panne, alité) suggèrent une relation causale avec la mort par chaleur (Tableau 1). Des victimes de la chaleur avaient été visités bien plus souvent par les infirmières de santé publique que leurs voisins survécus. Cette association statistiquement très significative n'implique alors (au contraire au paramètre «dispositif de climatisation en panne») pas *eo ipso* une relation causale entre la visite de l'infirmière de santé publique et la mort par chaleur (environ dans le sens que les infirmières auraient fait mourir les victimes de la chaleur!) [2].

En analogie à cet exemple une association ou corrélation d'un cours favorable ou défavorable d'une maladie avec un certain traitement ne doit pas dans chaque cas être interprétée comme preuve directe de l'utilité ou du mal du nouveau traitement.
Des effets de placebo ou suggestifs sont un problème important dans l'évaluation des effets véritables *versus* présumés. En thérapie, les effets de placebo sont bien connus. Il est intéressant que des ressorts autres que l'épidémiologie en soient également touchés.

Un registre du cancer au Canada a publié un avertissement que l'incidence du cancer aurait augmentée localement. La presse a donné un importance dramatique à cette nouvelle, et la population signalait de plus en plus de troubles qui auraient répondu à des maladies du cancer. Il se révéla que l'avertissement était inconsistant parce qu'il reposait sur un faux calcul. L'incident illustre l'effet de placebo dans la population d'une information inconsistante [3].

Il n'est pas facile de trouver la piste de tel problèmes dans la lecture de la littérature professionnelle. Nous ne promettons pas de recettes brevetées, cependant nous proposons quelques indications sur des ressources et des suggestions pour l'usage de l'imprimé.

Tableau 2. «Levels of evidence» pour étayer les décisions cliniques [4].

Level I	Grandes études, métaanalyses Grandes études cliniques randomisées avec une grande puissance statistique et avec un faible risque de résultats incorrect positif et/ou incorrect négatif
Level II	Metaanalyses moins importantes Etudes cliniques randomisées petites à moyennes avec une petite à moyenne puissance statistique (et par conséquence un risque relativement élevé surtout pour des résultats incorrects négatifs)
Level III	Etudes cliniques prospectives, mais pas randomisées
Level IV	Etudes rétrospectives
Level V	Casuistique, avis et expériences personnelles

«Levels of evidence»

Pas toutes les études ont la même importance scientifique. La force démonstrative scientifique de casuistiques est peu considérable s'il s'agit d'élaborer de nouveaux standards cliniques (Tableau 2). Par contre des grandes études randomisées – surtout des métaanalyses – qui examinent de nouvelles thérapies en comparaison à des «standards» établis dominent [4].

«Le cas» (rapport casuistique)

Le patient dont nous vivons personnellement son destin est un maître important. Des discussions de cas tiennent une place éminente dans les programmes de formation postgraduée et continue des cliniques, mais aussi lors de congrès et de cours. Ceci semble être en contradiction au faible «level of evidence» des rapports casuistiques. Cependant, il est justifié d'accorder à la majorité des rapports de cas une force probante limitée parce que la délimitation entre des cours «réguliers» et des cas exceptionnels n'est pas toujours facile. Au cas où une nouvelle thérapie obtient un effet à plein succès (par exemple la guérison rapide d'une maladie auparavant non curative), peu de cas peuvent en effet suffire comme matériel démonstratif pour déduire de l'observation clinique un nouveau standard thérapeutique. L'introduction des antibiotiques dans le traitement des maladies infectieuses autrefois mortelles (par exemple la méningite bactérienne) a amélioré leurs pronostics de telle manière impressionnant qu'il n'y avait pas besoin d'études supplémentaires (au besoin randomisées avec placebo!) pour en démontrer l'utilité. Des effets modestes (quoique cliniquement tout à fait importants) qui sont typiques pour de nombreuses nouvelles thérapies ne peuvent être mis en évidence correctement par des casuistiques.

Des séries de cas

Si un cas particulier ne suffit pas pour prouver l'utilité d'une thérapie, la confirmation de l'effet chez d'autres patients atteints de la même maladie devrait suffire comme preuve. Par conséquence des études qui informent sur des résultats thérapeutiques du médicament A pour N patients sont très fréquentes. Souvent il s'agit de séries de cas recherchés rétrospectivement dans le matériel clinique archivé dont le classement permet à beaucoup de candidats au doctorat d'acquérir leur bonnet de docteur. Des dates compilées prospectives se basent sur des protocoles dans lesquels sont déterminés (et dans l'espoir d'être observés) le diagnostic,

les critères de sélection des patients, les thérapies et les observations courantes avant le début de l'étude. Et tout de même une force démonstrative relativement modeste est accordée aux publications sur les études rétrospectives et prospectives. A quoi cela tient-il?
Dans une étude concernant l'effet d'un nouveau médicament sur une certaine maladie, il serait souhaitable de découvrir à quel point le médicament lui-même (et pas d'autres facteurs) détermine le cours. Indépendamment de la thérapie des facteurs pronostics qui aident à déterminer le cours sont en général «actif» dans la même maladie. Si la thérapie à vrai dire est efficace mais son effet modeste, l'influence des facteurs pronostics sur le cours peut s'accroître outre mesure – si le cours de la maladie est favorable la tentation est grande de l'attribuer à un effet (toutefois inexistant) d'une nouvelle thérapie «qui promet beaucoup». Dans des études rétrospectives de cas classés la standardisation du diagnostic et par cela le contrôle de ceux-ci est souvent impossible. Ainsi on discute si la substitution par les œstrogènes serait utile à l'empêchement de la maladie d'Alzheimer. La difficulté de ces études consiste dans le fait que souvent la maladie d'Alzheimer ne peut pas être délimitée de la démence vasculaire dont le traitement aux œstrogènes pourrait en effet être utile. Le cancer du sein métastasé fournit un bon exemple du problème de l'hétérogénéité des maladies entre les «groupes diagnostiques identiques»: Le cours est très différent et la palette clinique va du cours fulminant et rapidement létal jusqu'à des patientes qui survivent pendant des années avec des métastases démontrées (surtout des métastases du squelette et de la peau). Des facteurs pronostics comme l'âge lors de la manifestation de la maladie, le status des récepteurs d'hormones et autres choses encore sont des partenaires qui aident à former essentiellement le destin d'une patiente (Tableau 3).
La chimiothérapie à hautes doses pour le traitement du cancer du sein métastasé est un bon exemple de l'effet quantitatif considérable sur la sélection des patientes par des paramètres comme la survie sans récidive ou la survie dans l'ensemble.

De nombreuses publications semblent montrer des avantages d'une chimiothérapie à hautes doses pour le traitement du cancer metastasé du sein en comparaison d'une combinaison cytostatique et d'un dosage moins intensif. Quelques études ont été consacrées à la question intéressante dans quel mesure des critères de sélection pour les thérapies à hautes doses indiquent *per se* un meilleur pronostic. Des patientes qui sont en question pour des thérapies à hautes doses sont en effet sélectionnées sévèrement. Elles sont normalement jeunes, leur état général est bon, leur fonction cardiaque examinée et jugée bonne (comme supposition pour une thérapie à hautes doses d'anthracycline). Ces critères de sélection requis afin qu'une patiente puisse être considérée pour une thérapie à haute dose sont *eo ipso* des facteurs pronostics importants qui ont une influence très favorable sur le cours d'une maladie indépendamment de l'effet thérapeutique véritable [5].

Tableau 3. Les statisticiens «calmes» dans l'oncologie clinique.

L'hétérogénéité des tumeurs avec le «même diagnostic»
La sélection des patients
Des avis préconçus des médecins et/ou des patients

Une illusion ou une interprétation consciemment ou inconsciemment subjective par des chefs d'études, des médecins contrôleurs ou par les patients eux-mêmes peuvent avoir une influence indésirable sur des données cliniques, surtout si des paramètres cliniques «subjectifs» sont à l'ordre du jour.

Dans une étude anglaise concernant l'effet d'antihypertenseur, les patients, leurs parents et les médecins traitants ont été priés séparément de porter un jugement sur la qualité de vie des patients sous thérapie [6]. Les résultats ont divergé dans des proportions étonnantes: Les médecins jugeaient systématiquement excellente la qualité de vie de leurs patients, les patients prenaient une position intermédiaire et les parents étaient tous persuadés que la qualité de vie des patients était plus mal sous thérapie.

Comparer des groupes

L'examen de nouvelles thérapies se fait souvent et de façon raisonnable en comparaison avec des traitements habitués. Il se pose la question comment de telles comparaisons pourraient être réalisées pour être expressives. Dans le cas idéal, les groupes de patients devraient différer uniquement par la thérapie choisie (par exemple un nouveau médicament *versus* un ancien) pour que des résultats différents puissent être attribués à un effet thérapeutique (par exemple moins de récidives, une survie plus longues). La formation des groupes comparables devrait surtout assurer que des facteurs cliniques et biologiques essentiels qui déterminent le cours indépendant de la thérapie soient

représentés uniformément dans les deux groupes (par exemple l'âge des patients, le stade de la tumeur, la distribution des facteurs pronostic moléculaires/biologiques).

Groupe de personnes contrôles «historique»

Il est courant de comparer un groupe de patients traités dans le cadre d'une étude prospectiv avec un groupe de personnes contrôles «historique» saisis dans une banque de données et qui ont été traités avec une «ancienne» thérapie lors d'une période passée (en général avant le début d'une nouvelle étude). Ces dates archivées sont comparées avec les résultats de l'étude suivante. En règle générale les auteurs veillent à ce que les patients de l'ancienne et de la nouvelle étude ne soient pas différents par rapport aux facteurs pronostics importants et ils espèrent ainsi échapper à l'effet de sélection déjà discuté. En oncologie par exemple, on veille à ce que les groupes comparatifs n'aient pas de différences importantes quant à l'âge, le stade de la tumeur, l'étendue des métastases et d'autres caractéristiques cliniques et biologiques importantes de la tumeur. Malgré ces précautions de telles comparaisons posent des problèmes pour la raison simple que jamais *toutes* les caractéristiques importantes d'une maladie sont réparties de façon égale dans les deux groupes. En oncologie «l'effet Will Rogers» ou le phénomène de la «stage migration» est un phénomène connu [7].

A l'époque de la dépression économique américaine beaucoup de groupes de population des Etats-Unis se sentaient obligés de quitter leurs espaces vitaux. Ainsi des habitants de Oklahoma («Okies») ont émigré vers la Californie. Will Rogers, un humoriste et philosophe américain, commentait cette émigration avec le bon mot «when the Orkies left Oklahoma and moved to California, they raised the average intelligence level in both states». Apparemment Rogers avait une opinion mauvaise concernant les capacités intellectuelles des Orkies émigrés, mais encore plus mauvaise des Californiens.

L'effet «Will Rogers» a été examiné sur l'exemple du cancer des poumons. Sans que la définition des stades TNM de la tumeur aie changé les possibilités de l'examen et du «staging» du cancer des poumons ont été considérablement améliorés pendant les dernières décennies. Une étude de la Yale University School of Medicine illustre le problème du «stage migration» de manière impressionnante [7]. Deux cohortes de patients souffrants du cancer des poumons ont été comparés: une datant des années 1953–1964 et une plus tard de l'année 1977. Les patients qui ont été examinés 1977 ont profité des nouveautés techniques, surtout des méthodes d'illustration qui révèlent de manière plus sensitive les métastases dans les ganglions lymphatiques. Si lors du «staging» des patients de la seconde cohorte uniquement les résultats des techniques qui étaient aussi à la disposition de la première cohorte ont été pris en considération, il en résultait (apparemment) plus de patients dans le stade I (défini par le manque de métastases dans les ganglions lymphatiques) et (apparemment) moins dans le stade III.

Tableau 4 établit schématiquement l'effet de la «stage migration» sur le taux de survie des patients. Le pronostic des patients en stade TNM II semble être amélioré (survie générale autrefois 50% *versus* récent 70%). Cette amélioration du pronostic est due au fait que grâce à de nouvelles techniques d'examen certains patients sont classés dans un stade plus élevé (par exemple par la mise en évidence plus sensible des métastases dans les ganglions lymphatiques médiastinaux). Admettons qu'une nouvelle thérapie (mais inefficace) du cancer des poumons en stade II aie été établie entre les années 60 et la fin des années 70. Dans une comparaison des patients traités en stade II de la cohorte plus récente avec des patients de la cohorte ancienne il est presque certain que l'amélioration du pronostic de survie de 20% sera attribuée à la nouvelle thérapie.

Tableau 4.
«Stage migration» et la survie de patients malade du cancer des poumons en cas d'amélioration du diagnostique (présentation schématique; taux de survie après une période d'observation adéquate indiqué en %).

	Ancien «staging» avec des méthodes anciennes					
	Stade I		**Stade II**		**Stade III**	
	90%		50%		10%	
Sous-groupes dans les stades hétérogènes I–III	100%	**80%**	60%	**40%**	20%	0%
	Stade I		**Stade II**		**Stade III**	
	100%		70%		20%	
	Nouveau «staging» avec des méthodes plus sensibles					

Des groupes de comparaison historiques ont alors des chausse-trapes qui se mettent en travers d'une comparaison entre nouvelles et anciennes thérapies sur laquelle on peut compter. Y a-t-il d'autres possibilités de faire appel à des groupes de personnes contrôles pour l'évaluation d'une nouvelle thérapie?

«Responders» et «non-responders»

Des études oncologiques comparent quelques fois des patients qui répondent à un traitement («responders») avec des patients qui ne répondent pas («non-responders»), au lieu d'avoir recours à des groupes de comparaison. S'il se révèle que les «responders» ont un meilleur cours que les patients avec une maladie réfractaire à la thérapie (par exemple des patients avec des tumeurs résistantes aux cytostatiques) il semble que la preuve de l'efficacité d'un nouveau médicament soit produite. En est-il ainsi?

Le Coronary Drug Research Group a comparé chez des patients avec une coronaropathie le Clofibrat, un hypocholestérolémiant, avec le placébo, et cela dans une étude randomisée avec une haute «power» statistique [8]. Des patients avec une haute «compliance» où il est avéré qu'ils aient avalés plus que 80% de leurs tablettes de Clofibrat («responders») montraient une mortalité signifiquante plus basse que les patients moins disciplinés. Ainsi une bonne efficacité de la substance, dépendant du fait si le patient prend régulièrement son médicament, serait prouvée en soi (détaché de la comparaison avec le groupe placébo). Il est intéressant que le même effet se trouvait chez le groupe placébo: Des patients qui avaient pris au minimum 80% de leurs tablettes de placébo («placebo-responders») jouissaient également d'une mortalité statistiquement plus basse que les moins disciplinés. La compliance fait alors aussi son effet dans le bras placébo de l'étude. Le total des résultats de l'étude (comparaison entre le Clofibrat et le placébo) ne pouvait pas prouver un abaissement de la mortalité par Clofibrat.

Des patients avec des maladies tumorales qui ne répondent pas à la chimiothérapie se distinguent par des caractéristiques importantes des patients qui montrent une rémission de la tumeur. Ces caractéristiques déterminent de leur côté le pronostic indépendamment de l'effet thérapeutique (Tableau 5).

Est ce qu'avec notre exposé nous venons de porter la majorité des études cliniques à terre? Non! Des rapports de cas et des études rétro- et prospectives à un bras sont essentielles et précieux, à condition que leurs dates sont interprétées correctement. Dans les situations problématiques que nous avons montré d'autres dispositions d'études sont en effet nécessaires afin d'avancer la réalisation de nouveaux concepts. Ainsi dans un article suivant (voir un des prochains numéros de la SMW) nous nous occuperons d'études randomisées, des méta-analyses et notamment des problèmes des mécanismes de publication en soi.

Tableau 5.
Des raisons pourquoi des patients tumorales qui ne répondent pas à une thérapie montrent de toute façon un cours défavorable.

Des patients qui parviennent par la chimiothérapie à une rémission de la tumeur obtiennent plus de thérapie et qui dure plus longtemps, tandis qu'elle est arrêtée plus tôt chez les «non-responders»;
Les «responders» ont de toute façon un meilleur pronostic, indépendant de l'effet thérapeutique. Ainsi des paramètres biologiques qui transmettent une résistance aux cytostatiques sont souvent *eo ipso* des facteurs pronostic défavorables;
Dans le groupe des «non-responders» sont souvent inclus des patients qui meurent tôt, avant que l'effet thérapeutique puisse être jugé définitivement (par exemple des patients avec une leucémie aiguë qui pendant la phase de l'aplasie de la thérapie induisante meurent à cause des complications avant que la rémission éventuelle peut être saisie).

Références

1 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based medicine. How to practice and teach EBM. 1st ed. Churchill Livingstone; 1997.

2 Semenza J, Rubin CH, Falter KH, Selanikio JD, Flanders D, Howe HL, Wilhelm JL. Heat-related deaths during the July 1995 heat wave in Chicago. N Engl J Med 1996;335: 84-90.

3 Guidotti TL, Jacobs P. The implications of an epidemiological mistake: a community's response to a perceived excess cancer risk. Am J Publ Health 1993;83:233-9.

4 The Steering Committee on Clinical Practice Guidelines for the Care and Treatment of Breast Cancer. Clinical practice guidelines for the care and treatment of breast cancer. A Canadian consensus document. Can Med Assoc J 1998;158(3Suppl):S2.

5 Rahman ZU, Frye DK, Buzdar AU, Smith TL, Asmar L, Champlin RE, Hortobagyi GN. Impact of selection process on response rate and long-term survival of potential high-dose chemotherapy candidates treated with standard-dose doxorubicin-containing chemotherapy in patients with metastatic breast cancer. J Clin Oncol 1997;15:3171-7.

6 Jachuck SJ, Brierley H, Jachuck S, Willcox PM. The effect of hypotensive drugs on the quality of life. J R Coll Gen Pract 1982;32:103-5.

7 Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. N Engl J Med 1985;312:1604-8.

8 The Coronary Drug Project Research Group: Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. N Engl J Med 1980;303:1038-41.