

Lecture critique de la littérature clinique professionnelle

Partie II: Etudes randomisées, méta-analyses et problèmes de publication

M. F. Fey^a, A. Bühner^b

Introduction

Les rapports de cas et études cliniques sur un seul groupe ont une portée limitée; c'est ce que nous avons présenté dans notre précédent article, à l'aide d'exemples. Ils ne sont pas suffisants pour établir de nouvelles recommandations thérapeutiques, lorsqu'il est question de prouver de manière irréfutable qu'il y a de petites différences dans l'effet d'un nouveau traitement *par rapport* aux traitements établis. Dans ce cas, les études randomisées sont l'une des manières, et même parfois la seule, de fournir la preuve scientifique nécessaire. Il est donc important de connaître les éventuels pièges et traquenards de ces études pour leur interprétation. Nous présentons certaines réflexions à ce sujet, parsemées d'exemples tirés de l'oncologie.

Etudes randomisées

Les études randomisées comparent directement les résultats d'un nouveau traitement à ceux d'un groupe témoin, et les conséquences (p. ex. survie prolongée, toxicité plus importante ou moins), peuvent être directement imputées à ce traitement. La randomisation, soit l'attribution des patients au hasard, avant la mise en route du traitement, à un groupe «ancien» traitement ou témoin (standard en vigueur) ou à un groupe «nouveau» traitement (en phase «d'expérimentation» clinique) résout ainsi certains des problèmes d'interprétation et d'analyse évoqués dans notre premier article à propos des études réalisées selon des plans différents. La randomisation garantit que les groupes à comparer ne diffèrent que par le traitement, et que d'autres facteurs, influençant eux aussi l'évolution, sont uniformément répartis dans les deux groupes. Dont des facteurs pronostiques connus, mais aussi encore inconnus. Les premiers peuvent éventuellement

intervenir dans la formation des groupes témoins «historiques», mais en aucun cas les autres. Des facteurs de biologie moléculaire par exemple (Tumor-Suppressor-Gene inactifs, oncogène activés et beaucoup d'autres) ne peuvent intervenir sans autres dans le diagnostic de routine, ou même être (encore) parfaitement inconnus lors d'une étude.

Les études randomisées doivent garantir que les patients recevant un traitement établi sont aussi consciencieusement traités et documentés que ceux qui reçoivent un nouveau traitement. Les patients sous un nouveau traitement (moins connu) sont souvent pris en charge de manière particulièrement soignée, ce qui peut donner de meilleurs résultats, indépendamment du traitement à proprement parler. Malgré leurs avantages, les études randomisées ne sont pas immunisées contre tout problème.

Critères des études randomisées. Les critères («endpoints») sont des unités de mesure de l'effet d'un traitement. A la lecture d'un article avec des données d'une étude randomisée, la question se pose de savoir si les critères choisis sont cliniquement corrects (Tableau 1). Dans le traitement adjuvant, médicamenteux, du cancer du sein censé améliorer le pronostic des patientes dans un concept curatif, des critères tels que survie sans récurrence et survie globale sont particulièrement importants. Dans une étude sur le traitement palliatif d'un cancer métastatique, des critères comme qualité de vie sont souvent plus importants que la survie globale. L'amélioration de la proportion de rémissions du cancer métastatique n'a de sens clinique pour les patients que si elle va de pair avec une prolongation de la survie et/ou une amélioration de paramètres subjectifs. Au cas où le traitement ne fait que régresser radiographiquement des métastases asymptomatiques, mais au prix d'une toxicité considérablement plus marquée et d'une évolution globale identique, un véritable progrès thérapeutique est sujet à caution.

Groupe standard ou témoin dans les études randomisées. Du fait que le groupe standard ou témoin sert de référence dans l'évaluation d'un nouveau traitement, il faut en examiner les résultats avec un soin tout particulier.

- Le traitement du groupe standard est-il un standard actuel? Se référer aux recommandations («guidelines») publiées et articles de revue récents (de préférence d'autres auteurs).
- Les résultats du groupe standard correspondent-ils aux attentes? S'ils sont moins bons que généralement admis, c'est en faveur du groupe «expérimental», même si le nouveau traitement n'est en fait pas une amélioration notable. Et s'ils dépassent les

^a Institut d'Oncologie médicale, Hôpital de l'Île, Berne

^b Médecine interne FMH, Novazzano

Tableau 1.
Exemples de sélection de critères dans les études cliniques.

	Critères principaux, c.-à-d. les plus importants	Critères secondaires
Concept curatif chimiothérapie adjuvante dans le cancer du sein	Survie globale	Survie sans récurrence
Concept palliatif chimiothérapie dans le cancer du poumon métastatique	Qualité de vie Survie sans progression	Proportion de rémission Survie globale
Mesure d'appui facteurs de croissance hématopoïétiques après chimiothérapie	Diminution du risque d'infection en cas de neutropénie, et diminution de la consommation d'antibiotiques	Amélioration des examens de laboratoire hématologiques Survie

attentes, il est parfois impossible de prouver une différence entre traitement témoin et expérimental.

- Le traitement et le suivi clinique ont-ils été contrôlés aussi consciencieusement dans le groupe standard que dans le groupe expérimental (nombre des consultations, examens de laboratoire, et contrôles radiologiques)?

Le Groupe Suisse de Recherche Clinique sur le Cancer (SAKK) a poursuivi pendant plusieurs années des études sur la chimiothérapie adjuvante dans le carcinome colorectal après opération radicale. Les patients opérés ont été soit uniquement suivis (groupe standard), soit traités par chimiothérapie (groupe expérimental). Dans la deuxième génération d'études, les résultats du groupe standard, traitement chirurgical uniquement, ont été très nettement meilleurs que ceux des études de première génération. Il a donc été très difficile de prouver un quelconque effet additionnel d'une chimiothérapie adjuvante (U. Laffer, Bienne – communication personnelle).

Une étude randomisée a examiné l'intérêt de la chimiothérapie régionale dans les métastases hépatiques du cancer colorectal [1]. Les patients du groupe expérimental ont reçu une chimiothérapie adéquate; et la prise en charge clinique a été largement définie. Dans le groupe standard, les patients ont bénéficié d'un traitement de soutien léger, et certains d'une chimiothérapie systémique, arbitrairement. Les résultats ont été en faveur d'un avantage de la chimiothérapie régionale. Mais comme le suivi des patients du groupe témoin a manifestement été nettement moins bon et moins bien con-

trôlé, le bénéfice clinique du groupe expérimental ne peut être attribué à la chimiothérapie régionale exclusivement. Malheureusement dans cette étude (publiée dans une revue réputée), un nouveau concept thérapeutique cliniquement parfaitement contrôlé a été comparé à une politique désordonnée dans le groupe témoin, et sa portée s'en trouve donc amputée.

Analyses de sous-groupes dans les études randomisées. Malgré le fait que les critères d'admission dans les études randomisées déterminent la sélection des patients, le collectif traité n'est en fin de compte absolument pas homogène. Par exemple, une étude sur le traitement adjuvant de patientes préménopausées ayant un cancer du sein avec métastases ganglionnaires, recrute selon ses critères d'admission toute une gamme de patientes dont les facteurs pronostiques sont très variables. Pour une patiente de 35 ans ayant plus de 4 ganglions lymphatiques atteints et des récepteurs hormonaux négatifs, le pronostic est beaucoup plus défavorable que celui d'une femme de 15 ans plus âgée, mais ayant des récepteurs hormonaux positifs et une seule métastase ganglionnaire axillaire; toutes deux remplissent pourtant les critères de sélection cités. Pour de telles études, il faut donc sélectionner des sous-groupes et les analyser séparément; des sous-groupes différents quant aux facteurs pronostiques les plus importants par exemple (status des récepteurs hormonaux et des ganglions dans le cancer du sein). Ces analyses sont faciles à faire, souvent très utiles, mais sont parfois aussi problématiques.

L'étude ISIS-2, randomisée, a comparé chez 17 187 patients victimes d'un infarctus du myocarde les possibilités thérapeutiques streptokinase, Aspirine, streptokinase et Aspirine, à l'absence de telles options spécifiques dans un groupe témoin [2]. Les biostatisticiens de cette étude ont effectué une analyse des résultats de sous-groupes spécifiques. Ils ont choisi des critères cliniquement intéressants (sexe, anamnèse de diabète, âge), mais aussi des critères supposés n'avoir aucune importance pour les patients post-infarctus (signe du Zodiaque), avec une certaine marge de sécurité (Tableau 2). Il semble que l'Aspirine ne protège pas de la mort par infarctus les patients nés sous le signe de la Balance ou des Gémeaux. Malgré cela, personne ne renoncerait à donner un antiagrégant plaquettaire à ces sous-groupes de patients. L'argument que les analyses de sous-groupes se font généralement d'après des critères de sélection cliniquement importants ne met pas à l'abri de telles erreurs avec toute la sécurité voulue.

Ergo: «Don't do subgroup analyses – if you do, don't believe the results.» Nous ne sommes pas obligés de suivre certains biostatisticiens dans cette optique très stricte. De nombreux résultats d'analyses de sous-groupes sont effectivement très utiles en clinique. Le fait actuellement incontesté actuellement que les patientes ayant un cancer du sein négatif pour les récepteurs hormonaux posent un problème clinique particulier, a été découvert à l'origine dans des analyses de sous-groupes. La meilleure manière de résoudre ce problème consiste à vérifier les résultats des analyses de sous-groupes dans des études prospectives.

«Power» statistique d'une étude randomisée.

Pour le profane, les réflexions et calculs statistiques de résultats d'études sont souvent incompréhensibles. Les paragraphes «Méthode» d'un article par exemple sont flous si nous recherchons la signification par le test exact de «Fischer». La quadrature du cercle aurait-elle été plus facilement résolue avec le test du «chi carré»? Le simple clinicien est souvent dans l'impossibilité de savoir si les tests statistiques présentés étaient «indiqués» ou non. Et même les profanes en mathématiques-statistiques ne peuvent savoir si une étude a inclus suffisamment de patients, ni si les différences de résultats entre les traitements sont mathématiquement inattaquables. Mais quelques considérations accessibles au lecteur ignorant tout de la statistique (dont font partie les auteurs de cet article) peuvent être d'une certaine utilité.

- Réfléchir si un nouveau traitement offre *a priori* le potentiel d'un progrès considérable ou plutôt modeste. Plus l'effet thérapeutique attendu est modeste, plus il faudra de patients pour le confirmer dans une étude.
- Les études démontrant un progrès considérable avec un petit nombre de patients peuvent être
 - a) un authentique progrès;
 - b) statistiquement faibles, surestimant ainsi l'effet; ou
 - c) faussement positives.

Les options B et C sont (malheureusement) plus fréquentes que la A, en oncologie tout au moins.

- La publication rend-elle compte de tous les patients incorporés? N'importe quelle étude peut incorporer quelques patients qui ne remplissent pas les critères d'admission. Mais ils doivent être le moins nombreux possible (pas plus de quelques % du collectif global).
- Tous les patients sont-ils «analysables»? Combien ne le sont pas? Combien de patients ont-ils reçu le traitement prévu, et combien l'ont interrompu avant terme? Combien de patients ont-ils été perdus pour le suivi («lost to follow-up»)? Plus le nombre de patients dont les données sont inanalysables ou incomplètes est grand, plus les résultats de l'étude seront sujets à caution.

Méta-analyses

Les méta-analyses sont à la mode. Si certains y voient un progrès décisif pour l'analyse des données d'études cliniques, d'autres affirment que ce sont des «lots of fancy statistics on lots of garbage data». Les méta-analyses prennent les données d'un grand nombre d'études randomisées sur le même sujet, dans le but d'obtenir une «power» statistique très rarement atteignable dans une seule étude [3]. Les effets d'un traitement peuvent être quantitativement moins importants pour être significatifs. Par exemple, il est possible de prendre toutes les études sur l'utilité d'une chimiothérapie adjuvante dans le cancer du sein sans métastase ganglionnaire. Pour une telle méta-analyse, il faut des études qui comparent la chimiothérapie à un groupe témoin; ce n'est pas le type de chimiothérapie qui compte ici. Les méta-analyses ont surtout pour but d'examiner des principes de traitement en négligeant délibérément certains détails, comme les différentes chimiothérapies. L'intérêt d'une méta-analyse est donné ou non par la sélection des études (Figure 1). Une méta-analyse idéale sur un certain sujet doit prendre toutes les études, publiées ou non, indépendamment de leurs résultats. La stratégie suivie par les auteurs pour identifier les articles qu'ils jugent adéquats doit ressortir de la publication.

Problèmes de la publication des études

Abstracts. Les «abstracts», ou résumés d'études, sont la manière habituelle de se donner la possibilité de présenter des données lors de congrès. Des résultats d'études en cours, précoces ou «immatures» après une durée

Tableau 2. Analyse de sous-groupes dans l'étude ISIS-2.

Signe zodiacal	Effet de l'Aspirine sur la mortalité entre 0 et 30 jours après infarctus du myocarde	
	Nombre de décès groupe Aspirine contre groupe placebo	Signification statistique
Balance ou Gémeaux	150 / 8587 contre 147 / 8600	non significatif
Tous les autres signes	564 / 8587 contre 869 / 8600	p < 0,000 0001

d'observation trop brève, sont souvent annoncés pour être présentés dans des congrès. La question se pose de savoir combien de ces données présentées sous forme d'abstracts, et publiées dans des comptes-rendus de congrès, entreront ultérieurement dans la littérature sous la forme de publications originales, formelles et propres. Moins que nous l'aimerions!

Le compte-rendu du congrès de 1984 de l'American Society of Clinical Oncology (ASCO) compte environ 200 abstracts [5]. 60% seulement ont été publiés en version définitive dans un article original. Les trois quarts de tous les abstracts ayant pu être présentés à ce congrès ont été publiés, alors qu'à peine la moitié des abstracts figurant dans le compte-rendu du congrès, mais jamais présentés, n'ont jamais vu le jour sous la forme d'un véritable article. Si un abstract donne des résultats positifs (p. ex. progrès thérapeutique avec un nouveau médicament), sa chance d'être présenté verbalement à un congrès, avec publication ultérieure, est de 60–75%; si les résultats sont négatifs, elle tombe à environ 30% seulement. 80% de tous les abstracts sur des études cliniques ont été publiés par la suite avec les

mêmes résultats et les mêmes conclusions; dans 20% des cas, les divergences de résultats et de conclusions entre l'abstract et la publication ultérieure étaient importantes.

Les abstracts doivent être interprétés avec tout le recul critique nécessaire. Ils sont la plupart du temps insuffisants pour établir de nouveaux standards thérapeutiques.

Chances de publication des résultats positifs et négatifs. Les résultats positifs d'études cliniques sont plus faciles à publier que les négatifs [6]. De plus, la motivation d'entreprendre les démarches parfois difficiles jusqu'à la publication de résultats positifs est plus grande que pour les négatifs. Ce principe en soi bien compréhensible présente le risque que des résultats négatifs, mais cliniquement importants, ne passent pas dans la littérature. Il serait important de confirmer qu'un nouveau médicament ne présente aucun avantage sur un traitement bien établi, connu et souvent meilleur marché. Ce qui est particulièrement critique dans la sélection des études pour les méta-analyses (v. plus haut).

«Biais» linguistiques dans les revues scientifiques. Publier est un travail académique essentiel à une brillante carrière («publish or perish»). Les articles publiés dans des journaux reconnus au niveau international, soigneusement sélectionnés, disposant d'un système de «peer-review», ont un poids tout particulier. Ceux par contre qui ne paraissent que dans des revues locales, généralement dans la langue du pays, sont moins importants dans un curriculum vitae ou une liste de publications. Une étude intéressante a examiné la question de savoir dans quelle mesure les travaux sur les études cliniques publiés par des auteurs germaniques sont différents s'ils sont publiés en anglais au niveau international ou en allemand au niveau régional [7].

Cette étude a recherché des directeurs d'étude en Suisse, en Allemagne et en Autriche, ayant publié leurs études cliniques randomisées aussi bien en allemand qu'en anglais. Pour les études avec des résultats positifs (ayant pu démontrer p.ex. une différence statistiquement significative en faveur d'un nouveau traitement), la chance d'être publiées dans une revue anglophone reconnue était 3–4× plus grande que pour les études avec des résultats négatifs, qui (probablement après refus de l'un ou l'autre journal international) ont finalement été publiées en allemand (en partie «enterrées») [7].

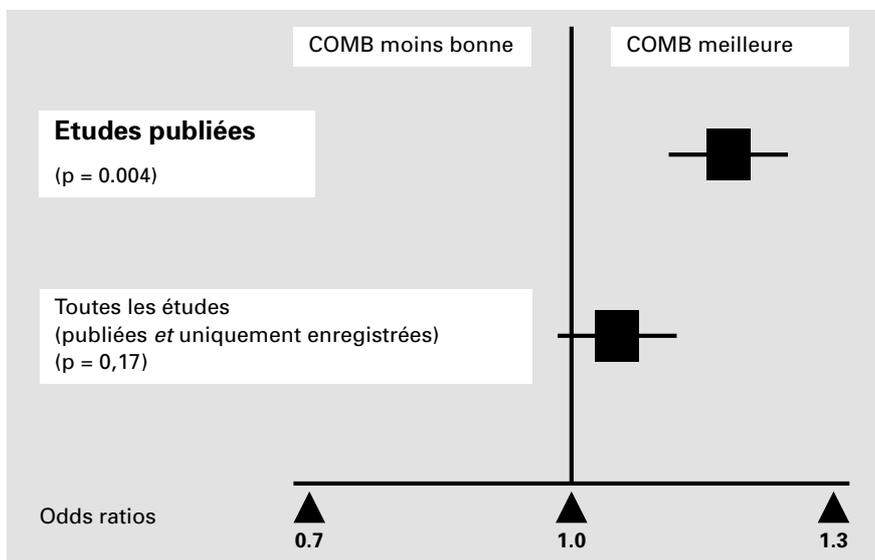


Figure 1.

Méta-analyse sur la question: les chimiothérapies combinées (COMB) sont-elles meilleures que les monochimiothérapies cytotostatiques dans le cancer de l'ovaire inopérable? Une «odds ratio» de 1,0 implique que ces deux options thérapeutiques sont équivalentes. Les carrés noirs indiquent les moyennes des effets thérapeutiques de toutes les études examinées: les valeurs >1,0 indiquent un avantage en faveur de la chimiothérapie combinée, et celles <1,0 en faveur de la monochimiothérapie. Les rectangles horizontaux indiquent la dispersion [4].

L'analyse de toutes les études *publiées* sur ce sujet parle en faveur d'un avantage des chimiothérapies combinées. Mais en prenant toutes les études enregistrées (publiées ou non), le résultat est moins évident. L'explication principale de cet effet est qu'il y a certaines études enregistrées qui n'ont trouvé aucun avantage à une chimiothérapie par plusieurs cytotostatiques, et qui n'ont pas été publiées, ou n'ont (pour cette raison?) pas pu l'être (v. également paragraphe «Chances de publication des résultats positifs et négatifs»).

Il est parfaitement possible que des résultats négatifs soient importants; il se pourrait par

exemple qu'un nouveau traitement (souvent plus cher que l'ancien) ne parvienne pas à fournir la preuve de sa meilleure efficacité. Le risque est que de tels articles sont sous-représentés dans la littérature anglo-saxonne. En outre dans les études cliniques, pour différentes raisons statistiques (qui ne seront pas abordées en détail ici), le risque de résultats faux positifs est très nettement supérieur à celui de résultats faux négatifs. La littérature anglo-saxonne, avec sa préférence pour la publication de données cliniques positives, prend donc moins de risque de publier des résultats faux négatifs que les revues en allemand moins estimées, qui acceptent de publier les «trials» négatifs.

Problèmes particulier des articles de revues (reviews). Les revues sont elles aussi moins immunisées contre les problèmes que les articles originaux. Le lecteur attend généralement un résumé pondéré de la littérature sur le thème choisi, et des conclusions bien pesées, motivées. Les revues peuvent être rédigées sur le mode «narratif» (le présent ouvrage en est un exemple), ou sur celui des méta-analyses.

Les méta-analyses sont intéressantes pour débattre de questions thérapeutiques, car leur méthode exclut dans une large mesure qu'un auteur de revue reprenne arbitrairement certains résultats pour en banaliser d'autres, ou même totalement les exclure. Elles garantissent la transparence du choix des études examinées et la reproductibilité du traitement des don-

nées. C'est pour cela que le style narratif est préférable dans les revues pour commenter les résultats d'études randomisées [8].

L'appréciation de la qualité d'une revue de style narratif («narrative review article») n'est pas simple, et il est surprenant que la discussion de ce problème n'ait été mise en route qu'au cours de ces dernières années. Les revues, comme le présent texte, ne peuvent pas être standardisées sans autre de la même manière que le résumé d'études randomisées, avec la méta-analyse comme instrument. La possibilité existe d'exploiter les données dans un sens ou dans l'autre («bias»), consciente ou non, pour une interprétation arbitraire et pour une sélection dirigée des travaux examinés (et parfois avec scotomisation des travaux du groupe opposé). Mais il est possible de formuler quelques exigences à la qualité d'un article. Les questions doivent être présentées clairement en introduction, et la technique de recherche et de sélection de la littérature doit être brièvement présentée.

Conclusions pratiques

Tout ce qui est écrit n'est pas forcément juste! Même des revues renommées, malgré un «peer-review» très critique dans la sélection des articles acceptés pour publication, se trompent. Même pour le spécialiste dans son domaine, il est impossible de distiller avec sécurité dans chaque publication les données exactes, et de savoir s'il s'agit d'éphémères ou tout simplement de choses fausses, ce qui peut avoir des conséquences à long terme en pratique. Il est tout de même possible de donner quelques indices permettant d'évaluer les articles sur des études clinico-thérapeutiques.

Nous n'avons pas la prétention que notre travail de revue lève tous les doutes, ni dans son contenu ni dans sa forme, ni que nous avons évité tous les pièges que nous avons trouvés dans d'autres travaux, avec l'élégance dont nous sommes capables. Nous ne disons pas non plus que presque tout ce qui est écrit est en principe sujet à caution. Mais nous espérons être parvenus, même de manière lacunaire, à transmettre quelques suggestions pour une technique de lecture critique de la littérature médicale, à l'exemple d'études oncologiques et d'autres travaux. Et pourquoi n'utiliserez-vous pas nos suggestions de lecture critique de la littérature scientifique pour notre article, comme exercice?

Quintessence

- Les **abstracts** et comptes-rendus de congrès ne suffisent pas pour établir de nouveaux standards cliniques.
- Attention au «**level of evidence**» d'une étude. Plus il est élevé, plus les données publiées sont fiables.
- S'assurer que les résultats ont pu être confirmés par des groupes d'études **indépendants** auprès d'un autre collectif.
- Attention au **collectif témoin** contre lequel les nouveaux résultats cliniques sont mesurés. Ces résultats correspondent-ils au groupe standard, au traitement standard, aux expériences et aux prévisions?
- Vérifier dans les **articles de revue** («reviews») que la méthode de traitement des données discutées (p.ex. technique de recherche dans la littérature) est bien présentée dans l'article.

Références

- 1 Allen-Mersh TG, Earlam S, Fordy C, Abrams K, Houghton J. Quality of life and survival with continuous hepatic-artery floxuridine infusion for colorectal liver metastases. *Lancet* 1994;344:1255-60.
- 2 ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988;II:349-60.
- 3 Egger M, Davey Smith G, Andrew N Phillips. Meta-analysis: principles and procedures. *Br Med J* 1997;315: 1533-53.
- 4 Symes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11-29.
- 5 De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Ann Oncol* 1992;3:187-91.
- 6 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267: 374-8.
- 7 Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997; 350:326-9.
- 8 Williams CJ. The pitfalls of narrative reviews in clinical medicine. *Ann Oncol* 1998;9:601-5.