

Principes de la lecture critique de la littérature médicale : Illustration urologique

G. LANDRIVON*, F. DELAHAYE*, R. ECOCHARD*, R.F. HELLER**

* Centre d'Information Médicale des Hospices Civils de Lyon, Hôtel-Dieu, Lyon ; ** Center for Clinical Epidemiology and Biostatistics, University of Newcastle, N.S.W., Australie

RESUME

La littérature scientifique s'est développée de façon considérable, spécialement dans le domaine de la santé et de la médecine. En parallèle, la recherche médicale est devenue de plus en plus complexe. Les médecins ont de ce fait besoin d'avoir à leur disposition des méthodes de lecture critique. Le principe de la lecture critique de la littérature médicale est d'évaluer la valeur des publications qui sont susceptibles d'influencer la pratique médicale du lecteur. Le lecteur a un double objectif : évaluer la crédibilité de ces publications et définir l'applicabilité des informations qu'elles contiennent. Grâce à une méthode de lecture en huit étapes, le lecteur est capable d'éliminer rapidement ce qui n'est pas valide, et d'avoir un regard objectif sur la qualité des résultats qui lui sont proposés. Ces huit étapes sont l'analyse des objectifs de l'étude, du type de l'étude, du facteur étudié, du critère de jugement, de la population étudiée et de son échantillon, des biais et facteurs de confusion, des résultats avec leur signification statistique et clinique ainsi que la puissance de l'étude, et enfin la synthèse des jugements du lecteur. Cette méthode a le mérite de s'adapter à tout les types de publication. A travers cette approche le dernier mot appartient au clinicien face à son malade.

Mots-Clés : lecture critique - littérature médicale - évaluation - méthodologie - validité

INTRODUCTION

Les publications médicales représentent chaque année une quantité d'informations colossale, et qui ne cesse de croître. A travers la lecture de la littérature médicale qui le concerne, le médecin est à la recherche d'une information utile. Mais lire les journaux médicaux est une activité très prenante et la plupart des médecins se contentent, pour leur revue habituelle de la littérature, des titres et des résumés. De plus, la qualité et la pertinence des publications sont variables et le lecteur doit être capable de "séparer le bon grain de l'ivraie". La lecture de la littérature nécessite donc une sélection et une évaluation. Pour cela s'est développé le concept de "lecture critique" [1].

Le principe de la lecture critique de la littérature est de juger la valeur des publications qui vont avoir une influence sur la pratique médicale du lecteur. Cela concerne la qualité de la recherche entreprise et la pertinence des résultats publiés.

L'OBJECTIF DU LECTEUR EST DOUBLE

1. Apprécier la crédibilité de la publication

Les résultats rapportés par l'auteur correspondent-ils vraiment à la réalité ? Le lecteur peut-il avoir confiance dans les conclusions proposées par l'auteur [2] ?

Pour ce faire, il doit apprécier la **validité** de l'étude en étant capable d'identifier rapidement, en fonction du type de l'étude, les différentes étapes et leurs composants, en les évaluant de façon à définir le **niveau de crédibilité** des informations fournies.

Il s'agit d'une enquête sur la méthodologie utilisée, et cette investigation repose sur la connaissance des concepts de base de l'épidémiologie.

2. Définir l'applicabilité des informations contenues dans la publication

Si les conclusions retenues sont considérées comme valides, celles-ci sont-elles applicables à la pratique médicale du lecteur [3] ?

L'objectif pour le lecteur est en effet d'améliorer, à travers les informations fournies par la littérature, sa propre pratique médicale, et par là le bien-être de ses patients. Le lecteur doit donc être capable d'identifier la population étudiée à sa propre population de malades, et d'évaluer le processus de sélection et de recrutement des sujets de l'étude, pour pouvoir accepter de généraliser les conclusions de l'étude et les appliquer à sa propre pratique.

Pour ce deuxième objectif, il s'agit également d'une enquête méthodologique sur le type et la conception de l'étude publiée.

LA METHODE

1. Description générale (Tableau 1)

Elle repose sur l'utilisation d'une grille de lecture. Le plan d'analyse proposé a le mérite d'être applicable à tous les types de publications. Il s'agit d'une adaptation du "Critical Appraisal Worksheet" du Centre for Clinical Epidemiology and Biostatistics (Pr R. F. Heller) de l'Université de Newcastle (New South Wales, Australia) [4, 5].

Cette grille est constituée de 8 lignes, correspondant à 8 critères d'évaluation. Chacun de ces 8 critères appelle les 3 mêmes types de questions, ce qui correspond aux 3

colonnes de la grille :

- 1) est-il possible de trouver dans l'article l'information pour le critère en question ?
- 2) la façon dont le critère en question a été abordé est-elle correcte ?
- 3) si la façon d'aborder le critère en question est incorrecte, cela menace-t-il la validité de l'étude ?

Ces 8 lignes correspondent aux différentes étapes de la conception d'un protocole.

A travers cette lecture en 8 étapes, le lecteur a donc la capacité d'écarter très vite ce qui n'est pas valide, et peut porter un regard objectif sur la qualité des résultats qui lui sont proposés [6].

2. Illustration

Le propos n'est pas de faire ici une lecture critique d'un article médical mais d'illustrer brièvement et schématiquement les définitions et concepts rencontrés à travers les 8 étapes proposées.

Imaginons une recherche bibliographique sur le thème fictif "consommation d'alcool et cancer de vessie", et retenons une étude de type cas-témoins pour traiter ce sujet pour illustrer les 8 étapes qui suivent.

Il est bien entendu que sur un tel sujet tout ne peut pas être dit en quelques paragraphes et sans le support d'un véritable article. Il ne s'agit que d'un exemple limité et schématique et les commentaires faits à ce propos sont loin d'être exhaustifs.

3. Les 8 étapes

◆ Il y a de nombreuses raisons de lire les articles médicaux. En dehors de la "gazette" médicale et des informations concernant la vie et l'organisation de la spécialité, le médecin est **à la recherche d'informations scientifiques concernant ses quatre grandes préoccupations** : [1]

- l'histoire naturelle d'une maladie, son évolution et son pronostic,

Tableau 1 :

L'information existe-t-elle pour chacune des 8 questions ?	La façon d'aborder la question est-elle correcte ?	Si non, cela menace-t-il la validité de l'étude ?
1 - Objectif - pronostic - évolution - test diagnostique - impact d'une intervention - étiologie - causalité	* Y-a-t-il une hypothèse ?	
2 - Type d'étude - rapport de cas - série de cas - étude transversale - étude cas-témoin - étude de cohorte - essai contrôlé	* Le type de l'étude est-il approprié à la question posée ?	* Si non, les résultats de l'étude sont-ils totalement inutiles ?
3 - Facteur(s) étudié(s) - exposition - intervention - test diagnostique	Sont-ils bien décrits ? Comment sont-ils mesurés ? * Même méthode de mesure chez tous les sujets ? dans tous les groupes ? * Méthode à l'aveugle ? Y-a-t-il une comparaison indépendante avec le gold standard?	* Sinon ce biais de mesure menace-t-il la validité de l'étude ? * Idem Sinon ce biais menace-t-il la validité de l'étude ?
4 - Critère(s) de jugement	Comment sont-ils mesurés ? * Idem question 3 Tous les critères de jugement pertinents sont-ils évalués ?	* Idem question 3 Si non, ceux qui ont été oubliés sont-ils importants ?
5 - Population source et sujets étudiés	* La sélection est-elle correcte ? * Y-a-t-il randomisation ? * Les groupes diffèrent-ils par des caractéristiques autres que les facteurs étudiés ? * Quelle est la proportion de sujets atteignant la fin du suivi ? * Y-a-t-il, pour le test, un large éventail de patients ?	* Si non, ce biais menace-t-il la validité externe? * Si non, ce biais menace-t-il la validité interne ? * Si elle n'est pas optimale, la validité interne est-elle menacée? * Si non, ce biais menace-t-il la validité externe?
6 - Facteurs de confusion potentiels et biais	* Sont-ils tous envisagés ? * Sont-ils bien contrôlés ?	* Si non, cela invalide-t-il l'étude ?
7 - Analyses statistiques Intervalle de confiance ? Test statistique ? * si résultats positifs * si résultats négatifs Force de l'association Calcul des likelihood ratios	* Taille de l'échantillon suffisante ? * Cliniquement intéressant ? * Puissance du test, taille de l'échantillon ?	* Si non, les résultats sont-ils inutiles ? * Si non, l'étude est-elle utile ? * Si insuffisant, l'étude est-elle utile ou non concluante ?
8 - Conclusions des auteurs? Réponses aux questions ? Vérification de l'hypothèse ? Objectif atteint ?	* Les conclusions répondent-elles à l'objectif ?	<u>En somme :</u> * <u>Les résultats sont-ils acceptables appliqués à la population-source ?</u> = VALIDITE * <u>Les résultats sont-ils acceptables pour votre propre pratique ?</u> = APPLICABILITE

- la validité et l'utilisation d'un nouveau test diagnostique,
- l'impact d'une intervention, le plus souvent thérapeutique, pour distinguer l'utile de l'inutile voire de la dangereuse, mais aussi de dépistage, d'intervention ou d'éducation.
- la détermination d'une étiologie ou d'une causalité.

Il peut aussi être intéressé par :

- des explications physiopathologiques,
- des recettes ou des conseils sur la façon d'appréhender un problème particulier et rare par un auteur (ou un groupe) performant et renommé.

La lecture critique consiste d'abord à replacer l'article examiné dans l'un de ces cadres afin d'avoir une idée claire sur les **objectifs** du sujet traité, son intérêt, ainsi que sur la ou les questions posées et l'hypothèse testée.

Ainsi, dans notre exemple, il s'agit de la détermination d'une étiologie : le cancer de la vessie est-il associé à la consommation d'alcool ?

Attention, " association " ne veut pas dire " relation de cause à effet "

◆ Ensuite, la **conception de l'étude** rapportée dans l'article doit être clarifiée.

La plupart des études sont conçues selon les 6 modèles suivants [6] :

- "à propos de 1 cas" :

un rapport de cas est une description d'un cas intéressant et inhabituel.

- série de cas :

une série de cas est une description de plusieurs cas comparables. Dans cette conception on ne peut pas prétendre répondre à une question spécifique ou comparer les résultats avec un autre groupe de cas ou de témoins.

- étude transversale :

une étude transversale est une description de la fréquence d'une maladie, de ses fac-

teurs de risque ou de ses autres caractéristiques dans une population déterminée à un temps déterminé.

- étude cas-témoin :

une étude cas-témoin est une étude d'observation, rétrospective, dans laquelle les caractéristiques de patients atteints d'une maladie (les cas) sont comparées avec celles de patients indemnes de la maladie (les témoins).

- étude de cohorte :

une étude de cohorte est une étude d'observation, prospective, dans laquelle un groupe de sujets exposés à des facteurs de risque d'une maladie est suivi pendant une période de temps donné. Le taux d'incidence de la maladie dans ce groupe exposé est comparé à celui d'un groupe contrôle, suivi pendant le même temps, mais non exposé aux facteurs de risque.

- essai contrôlé :

un essai contrôlé est une étude expérimentale dans laquelle une intervention est pratiquée à un groupe de sujets et le résultat de cette intervention est comparé à un groupe semblable, témoin, qui ne bénéficie pas de l'intervention étudiée.

Le lecteur doit reconnaître le modèle de l'étude pour vérifier si celui-ci est le plus approprié pour la question posée. En outre, il existe une "hiérarchie" parmi ces modèles et le niveau de preuve des résultats d'une étude (et la confiance du lecteur) est variable d'un modèle à l'autre. Il va croissant, du cas ou de la série de cas à l'étude transversale, puis à l'étude cas témoin, à l'étude de cohorte, pour être maximal avec l'essai contrôlé.

Une conception inadaptée à la question posée, comme un niveau de preuve insuffisant, ne rendent pas nécessairement l'étude complètement inutile. Celle-ci peut en effet être utilisée pour démontrer la pertinence d'une question, et pour soulever des hypothèses à tester par des études de conception différente.

Pour le sujet de notre exemple, c'est l'étude cas-témoins la plus pertinente.

En effet, dans ce modèle, on part d'un groupe de sujets atteints de cancer de la vessie chez qui l'on étudie les antécédents de consommation alcoolique, que l'on compare à ceux d'un groupe de sujets témoins, indemnes de cancer de la vessie.

Seul le modèle retrospectif est concevable. On imagine la difficulté de conception d'une étude prospective sur ce sujet : partir d'un groupe de sujets "alcooliques" et les suivre dans le futur de nombreuses années, pour collecter les cas incidents de cancer de la vessie qui seraient comparés à ceux qui surviennent chez des sujets suivis en parallèle, mais "non alcooliques".

L'essai contrôlé est bien entendu impensable.

Quant aux séries de cas et études transversales, elles seraient nécessairement non concluantes du fait de l'absence de groupe témoin.

Ce type d'étude, cas-témoins, est donc bien adapté à la question posée.

◆ Le type d'étude étant défini, quel est le **facteur étudié** ?

Le facteur étudié est l'exposition ou l'intervention supposée avoir des conséquences sur un problème de santé, une maladie ou un état clinique.

Le lecteur doit pouvoir savoir comment le ou les facteurs ont été mesurés, si tous les facteurs pertinents ont été pris en compte, et si la même méthode de mesure a été appliquée à tous les sujets, ainsi que d'un groupe à l'autre. Il doit aussi pouvoir apprécier la qualité de cette mesure (variabilité, mesure en "insu") [7].

Ainsi, à ce stade comme au précédent, il est possible de savoir si la validité de l'étude est menacée.

Le facteur étudié est la consommation d'alcool. Comment celle-ci est-elle mesurée ?

(Par questionnaire ou interrogatoire direct ?) L'intervieweur peut-il inconsciemment poser les questions de façon différente selon qu'il a en face de lui un cas de cancer de la vessie ou un sujet témoin ? Comment évalue-t-il la quantité d'alcool ingérée en fonction des différents types de boisson ? Comment pondère-t-il les différents types de boisson ?

◆ L'étape suivante est l'identification du **facteur résultant (ou critère de jugement)**.

Le facteur résultant est défini comme l'événement ou la situation supposé être le résultat de l'influence du facteur étudié (mort, maladie, inconfort, insatisfaction...) [7].

Le lecteur doit avoir à disposition dans l'article tous les éléments pour répondre aux mêmes questions que celles qui concernaient le facteur étudié afin qu'à ce stade aussi il puisse évaluer la validité des informations.

Le critère de jugement est le cancer de la vessie. Est-ce le diagnostic de cancer de la vessie ou est-ce la mortalité par cancer de la vessie ?

Si l'anatomo-pathologie définit de façon irréfutable un cas, comment peut-on être sûr qu'un témoin, dans cette étude, est indemne de cancer de la vessie ?

◆ la cinquième étape de ce processus d'évaluation est la recherche des informations permettant au lecteur une analyse de la **population étudiée et de l'échantillon utilisé** [7,8].

Population est un terme pour tout ensemble d'individualités (individus, cas, événements) qui peut être défini dans le temps et dans l'espace.

La population de référence, ou population à étudier, est le groupe auquel les résultats de l'étude, s'ils sont valides, vont s'appliquer.

L'échantillon est un sous groupe de la population étudiée, sélectionné de façon aléatoi-

re ou non, pour représenter l'ensemble de la population étudiée, lorsqu'il n'est pas possible pour des raisons pratiques d'étudier celle-ci dans sa globalité.

Il s'agit donc, à cette étape, de juger la validité externe de l'étude, c'est-à-dire de savoir si les conclusions, en admettant qu'elles soient valides, peuvent s'appliquer à une population plus large que le simple échantillon étudié.

Le comportement des sujets de l'échantillon étudié (suivi, perdus de vue) intéresse la validité interne de l'étude.

Dans l'exemple, la population de référence est celle des sujets qui consomment de l'alcool.

Dans la population étudiée, les contrôles sont identifiés par courrier ou par téléphone. Si des sujets ne sont pas accessibles par ce moyen, ou si des sujets refusent de participer à l'étude, est-ce pour des raisons socio-économiques ou psychologiques, qui pourraient expliquer aussi une consommation alcoolique supérieure à la normale ? Dans ce cas, cela ne risque-t-il pas d'augmenter artificiellement le taux des "alcooliques" chez les cas par rapport aux témoins, et d'aller dans le sens d'une fausse association entre alcool et cancer de la vessie ? Le même phénomène peut se produire si les cas sont recrutés dans un hôpital drainant une population d'un niveau socio-économique particulier, et différente de celle des contrôles.

◆ La sixième étape est celle de la discussion **des facteurs de confusion et des biais**.

Un biais est une erreur systématique qui s'introduit dans une étude et qui contribue à produire des estimations systématiquement plus élevées ou plus basses que la valeur réelle des paramètres à estimer. Il intervient par exemple au niveau de la sélection des patients, ou de la mesure des paramètres étudiés.

Un facteur de confusion est un facteur qui modifie les effets du facteur étudié sur le

critère de jugement, du fait de son lien à la fois avec le facteur étudié et avec le critère de jugement.

Sont-ils tous envisagés et pris en compte ? Si tel n'est pas le cas, la validité de l'étude peut être mise en doute. A cette étape particulièrement, l'expérience clinique du lecteur peut se révéler pertinente.

Interroger sur leur consommation d'alcool les sujets atteints de cancer avec plus de soin que les témoins constituerait un biais de mesure. Cela pourrait tendre à mettre en évidence une différence entre les deux groupes alors qu'elle n'existe pas. C'est ce qui se produirait également si les sujets ininterrogeables du groupe témoin étaient systématiquement plus alcooliques. Ce serait un biais de sélection.

De même, il faut prendre en compte dans une telle étude tous les autres facteurs connus comme étant des facteurs de risque de cancer de la vessie (tabagisme, travail dans l'industrie des colorants, etc...). En effet, s'il sont significativement plus fréquents dans le groupe des cancers, par exemple, on ne peut pas savoir si une éventuelle différence en taux de cancer entre les deux groupes est à mettre sur le compte de ces facteurs de risque ou sur le compte de la consommation alcoolique elle-même. Ces facteurs sont des facteurs de confusion.

◆ A la septième étape, le lecteur veut savoir si les résultats ne sont pas dus à la chance seule, autrement dit s'ils sont **statistiquement significatifs**. C'est le rôle magique de la valeur de p , probabilité que le résultat obtenu ne soit dû qu'à la chance. En dessous de 5 %, ou mieux de 1 % (en anglais, p value $< 0,05$ ou $< 0,01$), on considère que cette probabilité est trop faible pour que le hasard seul soit intervenu pour expliquer le résultat. Celle-ci n'a de sens que si le test statistique correct a été utilisé, et seulement si le lecteur, aux étapes précédentes, a considéré que la validité de l'étude n'était pas menacée.

Un test statistique est un outil mathématique, construit à partir des données collectées, qui permet de calculer cette probabilité [9].

- Si les résultats sont statistiquement significatifs, sont-ils **cliniquement pertinents** ?

Autrement dit, un résultat observé, une différence entre deux groupes, dus à un autre facteur (en l'occurrence le facteur étudié) qu'à la chance seule, ont-ils un sens en terme de pratique clinique ? Même si l'on observe un effet, celui-ci change-t-il quelque chose pour le malade ?

- Si les résultats sont négatifs, le sont-ils réellement, ou bien les auteurs avaient-ils les moyens de mettre en évidence des résultats positifs ?

La puissance du test est la probabilité de conclure, à partir de l'échantillon, à une différence ou à un effet statistiquement significatif lorsque la différence ou l'effet existe réellement dans la population.

On dit qu'une étude manque de puissance quand les conditions de sa réalisation rendent peu probables la mise en évidence de l'effet ou de la différence, lorsque ceux-ci existent réellement dans la population.

S'il existe des informations sur la puissance de l'étude (sur la taille de l'échantillon, et sur le risque α ou risque d'affirmer qu'il y a une différence ou un effet alors qu'en réalité il n'y en a pas), alors le lecteur est capable de dire si l'étude est concluante ou non, utile ou non.

◆ Enfin, arrivé au terme de cette évaluation, la huitième étape est celle de **la synthèse des jugements du lecteur**.

A chacune des étapes précédentes, la validité de l'étude était-elle menacée, gravement ou faiblement ?

Quelles sont les conclusions des auteurs ?

Répondent-ils aux questions ?

Les résultats sont-ils applicables à la population étudiée ? (c'est la validité externe de l'étude).

Et surtout, les résultats sont-ils acceptables pour la propre pratique du lecteur ? Vont-ils changer son comportement et améliorer l'état de ses malades ?

Dans l'exemple d'une étude cas-témoins sur la consommation alcoolique et le cancer de la vessie, c'est à la troisième étape que la validité interne de l'étude semble la plus menacée, du fait de la difficulté à mesurer le facteur étudié dans ce cas particulier. En ce qui concerne la validité externe d'une telle étude, on a vu que celle-ci était menacée par la proportion et la nature des femmes inaccessibles à l'étude.

Enfin, une étude de ce type réalisée sur une population urbaine pauvre d'Amérique du Nord, ou bien chez des Scandinaves, est-elle pertinente pour l'ensemble de la population française ?

CONCLUSION

Si cette approche nécessite du lecteur la connaissance des notions de base en méthodologie, il s'agit cependant d'une démarche essentiellement clinique. Le mot de la fin est pour le clinicien face à son malade.

REFERENCES

1. SACKETT D. L., HAYNES R. B., TUGWELL P. : Clinical Epidemiology. A basic science for clinical medicine. First ed. Boston : Little, Brown and Company, 1985.
2. MAYES L., HORWITZ R., FEINSTEIN A. : A collection of 56 Topics with Contradictory Results in Case-Control Research. In. J. Epidemiol 1988, 17 : 680-85
3. GRENIER B., LORETTE G., NIVET H. : De l'analyse critique d'un article médical à la rédaction d'une publication scientifique. La nouvelle Presse Médicale 1981; 10 : 995-96
4. DARZINS PJ, SMITH BJ, HELLER RF. : How to read a journal article. The Medical Journal of Australia 1992; 57 : 389-394.
5. LANDRIVON G, ECOCHARD R. : Principles of the critical appraisal of medical literature. Health Information and Libraries 1992; 3 : 29-34.
6. FOWKES F.G.R., FULTON P.M. : Critical appraisal of published research : introductory guidelines. B M J 1991 ; 302 : 1136-1140.

7. Fletcher R.H., Fletcher S. W., Wagner E. H. Clinical Epidemiology. The essentials. 2nd. ed. Baltimore: Williams & Wilkins, 1988.
8. LAST J.M., ED. : A dictionary of epidemiology. 2nd ed. Oxford : Oxford University Press, 1988.
9. MORTON R. F., HEBEL J. R. : A Study Guide to Epidemiology and Biostatistics including 100 multiple-choice questions. Baltimore : University Park Press, 1979.

ABSTRACT

Principles of the Critical Appraisal of Medical Literature : An Illustration for Urology

**G. LANDRIVON, F. DELAHAYE,
R. ECOCHARD, R.F. HELLER**

There has been an enormous expansion in scientific literature, especially in the area of health and medicine. At the same time the complexity of the research has increased. Through the reading of medical literature of his field of interest, the physician is in search of useful information. However, reading medical papers is very time consuming and most physicians, for their usual literature review, content themselves with the titles and the abstracts. Moreover, quality and relevancy of the publications vary greatly and

the reader must be able to "separate the wheat from the chaff". Thus the reading of literature needs selection and evaluation. As a result, physicians require critical appraisal skills. The principle of appraisal of medical literature is to assess the value of publications which will influence the medical practice of the reader. The reader has a double objective : to assess the credibility of the publications, and to define the applicability of the publication information. By following eight steps the reader is able to quickly rule out from his reading what is not valid and have an objective look at the quality of the results which are proposed to him. These eight steps comprise the assessment of the study objectives, the study design, the study factor(s), the outcome factor(s), the study population and sample, the confounding factor(s) and bias, the results with their statistical and clinical significance and the power of the study, and, lastly, the synthesis of the reader's judgment. The last word belongs to the clinician facing his patient.

Key Words : critical appraisal - medical literature - evaluation - methodology - validity.