REVIEW ARTICLE

Critical Appraisal of Scientific Articles

Part 1 of a Series on Evaluation of Scientific Publications

Jean-Baptist du Prel, Bernd Röhrig, Maria Blettner

SUMMARY

Introduction: In the era of evidence-based medicine, one of the most important skills a physician needs is the ability to analyze scientific literature critically. This is necessary to keep medical knowledge up to date and to ensure optimal patient care. The aim of this paper is to present an accessible introduction into critical appraisal of scientific articles.

Methods: Using a selection of international literature, the reader is introduced to the principles of critical reading of scientific articles in medicine. For the sake of conciseness, detailed description of statistical methods is omitted.

Results: Widely accepted principles for critically appraising scientific articles are outlined. Basic knowledge of study design, structuring of an article, the role of different sections, of statistical presentations as well as sources of error and limitation are presented. The reader does not require extensive methodological knowledge. As far as necessary for critical appraisal of scientific articles, differences in research areas like epidemiology, clinical, and basic research are outlined. Further useful references are presented.

Conclusion: Basic methodological knowledge is required to select and interpret scientific articles correctly.

Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Johannes Gutenberg-Universität, Mainz: Dr. med. du Prel, MPH, Dr. rer. nat. Röhrig, Prof. Dr. rer. nat. Blettner

Despite the increasing number of scientific publications, many physicians find themselves with less and less time to read what others have written. Selection, reading, and critical appraisal of publications is, however, necessary to stay up to date in one's field. This is also demanded by the precepts of evidence-based medicine (1, 2).

Besides the medical content of a publication, its interpretation and evaluation also require understanding of the statistical methodology. Sadly, not even in science are all terms always used correctly. The word "significance," for example, has been overused because significant (or positive) results are easier to get published (3, 4).

The aim of this article is to present the essential principles of the evaluation of scientific publications. With the exception of a few specific features, these principles apply equally to experimental, clinical, and epidemiological studies. References to more detailed literature are provided.

Decision making

Before starting a scientific article, the reader must be clear as to his/her intentions. For quick information on a given subject, he/she is advised to read a recent review of some sort, whether a (simple) review article, a systematic review, or a meta-analysis.

The references in review articles point the reader towards more detailed information on the topic concerned. In the absence of any recent reviews on the desired theme, databases such as PubMed have to be consulted.

Regular perusal of specialist journals is an obvious way of keeping up to date. The article title and abstract help the reader to decide whether the article merits closer attention. The title gives the potential reader a concise, accurate first impression of the article's content. The abstract has the same basic structure as the article and renders the essential points of the publication in greatly shortened form. Reading the abstract is no substitute for critically reading the whole article, but shows whether the authors have succeeded in summarizing aims, methods, results, and conclusions.

The structure of scientific publications

The structure of scientific articles is essentially always the same. The title, summary and key words are followed by the main text. This is divided into Introduction, Methods, Results and Discussion (IMRAD), ending when appropriate with Conclusions and References.

The content and purpose of the individual sections are described in detail below.

Introduction

The Introduction sets out to familiarize the reader with the subject matter of the investigation. The current state of knowledge should be presented with reference to the recent literature and the necessity of the study should be clearly laid out. The findings of the studies cited should be given in detail, quoting numerical results. Inexact phrases such as "inconsistent findings," "somewhat better" and so on are to be avoided. Overall, the text should give the impression that the author has read the articles cited. In case of doubt the reader is recommended to consult these publications him-/herself. A good publication backs up its central statements with references to the literature.

Ideally, this section should progress from the general to the specific. The introduction explains clearly what question the study is intended to answer and why the chosen design is appropriate for this end.

Methods

This important section bears a certain resemblance to a cookbook. The description of the procedures should give the reader "recipes" that can be followed to repeat the study. Here are found the essential data that permit appraisal of the study's validity (6). The methods section can be divided into subsections with their own headings; for example, laboratory techniques can be described separately from statistical methods.

The methods section should describe all stages of planning, the composition of the study sample (e.g., patients, animals, cell lines), the execution of the study, and the statistical methods: Was a study protocol written before the study commenced? Was the investigation preceded by a pilot study? Are location and study period specified? It should be stated in this section that the study was carried out with the approval of the appropriate ethics committee. The most important element of a scientific investigation is the study design. If for some reason the design is unacceptable, then so is the article, regardless of how the data were analyzed (7).

The choice of study design should be explained and depicted in clear terms. If important aspects of the methodology are left undescribed, the reader is advised to be wary. If, for example, the method of randomization is not specified, as is often the case (8), one ought not to assume that randomization took place at all (7). The statistical methods should be lucidly portrayed and complex statistical parameters and procedures described clearly, with references to the specialist literature. *Box 1* contains further questions that may be helpful in evaluation of the Methods section.

Study design and implementation are described by Altman (7), Trampisch and Windeler (9), and Klug et al. (10). In experimental studies, precise depiction of the design and execution is vital. The accuracy of a method, i.e. its reliability (precision) and validity (correctness), must be stated. The explanatory power of the results of a clinical study is improved by the inclusion of a control

---

BOX 1

Questions on methodology

- Is the study design suited to fulfill the aims of the study?
- Is it stated whether the study is confirmatory, exploratory or descriptive in nature?
- What type of study was chosen, and does it permit the aims of the study to be addressed?
- Is the study's endpoint precisely defined?
- What statistical measure is employed to characterize the endpoint? Do epidemiological studies, for instance, give the incidence (rate of new cases), prevalence (current number of cases), mortality (proportion of the population that dies of the disease concerned), lethality (proportion of those with the disease who die of it) or the hospital admission rate (proportion of the population admitted to hospital because of the disease)?
- Are the geographical area, the population, the study period (including duration of follow-up), and the intervals between investigations described in detail?

---

group (active, historical, or placebo controls) and by the randomized assignment of patients to the different arms of the study. The quality can also be raised by blinding of the investigators, which guarantees identical treatment and observation of all study participants. A clinical study should as a rule include an estimation of the required number of patients (case number planning) before the beginning of the study. More detail on clinical studies can be found, for instance, in the book by Schumacher and Schulgen (11). International recommendations specially formulated for the reporting of randomized, controlled clinical trials are presented in the most recent version of the CONSORT Statement (Consolidated Standards of Reporting Trials) (12).

Epidemiological investigations can be divided into intervention studies, cohort studies, case-control studies, cross-sectional studies, and ecological studies. *Table 1* outlines what type of study is best suited to what situation (13). One characteristic of a good publication is a precise account of inclusion and exclusion criteria. How high was the response rate ($\geq$80% is good, $\leq$30% means no or only slight power), and how high was the rate of loss to follow-up, e.g. when participants move away or withdraw their cooperation? To determine whether participants differ from nonparticipants, data on the latter should be included. The selection criteria and the rates of loss to follow-up permit conclusions as to whether the study sample is representative of the target population. A good study description includes information on missing values. Particularly in case-control studies, but also in nonrandomized clinical studies and cohort studies, the choice of the controls must be described precisely. Only then can one be sure that the control group is comparable with the study group and shows no systematic discrepancies that can lead to misinterpretation (confounding) or other problems (13).

Is it explained how measurements were conducted? Are the instruments and techniques, e.g. measuring devices, scale of measured values, laboratory data, and

TABLE 1

The best type of study for epidemiological investigations (from 13)

Purpose of investigation	Study type
Investigation of rare diseases, e.g. tumors	Case-control studies
Investigation of exposure to rare environmental factors, e.g. industrial chemicals	Cohort study in an exposed population
Investigation of exposure to multiple agents, e.g. the joint effect of oral contraceptive intake and smoking	Case-control studies
Investigation of multiple endpoints, e.g. the risk of death from various causes	Cohort studies
Estimation of incidence in exposed populations	Exclusively cohort studies
Investigation of cofactors that vary over time	Preferably cohort studies
Investigation of cause and effect	Intervention studies

time point, described in sufficient detail? Were the measurements made under standardized—and thus comparable—conditions in all patients? Details of measurement procedures are important for assessment of accuracy (reliability, validity). The reader must see on what kind of scale the variables are being measured (e.g. eye color, nominal; tumor stage, ordinal; bodyweight, metric), because the type of scale determines what kind of analysis is possible. Descriptive analysis employs descriptive measured values and graphic and/or tabular presentations, whereas in statistical analysis the choice of test has to be taken into consideration. The interpretation and power of the results is also influenced by the scale type. For example, data on an ordinal scale should not be expressed in terms of mean values.

Was there a careful power calculation before the study started? If the number of cases is too low, a real difference, e.g. between the effects of two medications or in the risk of disease in the presence vs. absence of a given environmental factor, may not be detected. One then speaks of insufficient power.

Results
In this section the findings should be presented clearly and objectively, i.e. without interpretation. The interpretation of the results belongs in the ensuing discussion. The results section should address directly the aims of the study and be presented in a well-structured, readily understandable and consistent manner. The findings should first be formulated descriptively, stating statistical parameters such as case numbers, mean values, measures of variation, and confidence intervals. This section should include a comprehensive description of the study population. A second, analytic subsection describes the relationship between characteristics, or estimates the effect of a risk factor, say smoking behavior, on a dependent variable, say lung cancer, and may include calculation of appropriate statistical models.

Besides information on statistical significance in the form of p values, comprehensive description of the data

and details on confidence intervals and effect sizes are strongly recommended (14, 15, 16). Tables and figures may improve the clarity, and the data therein should be self-explanatory.

Discussion
In this section the author should discuss his/her results frankly and openly. Regardless of the study type, there are essentially two goals:

Comparison of the findings with the status quo— The Discussion should answer the following questions: How has the study added to the body of knowledge on the given topic? What conclusions can be drawn from the results? Will the findings of the study lead the author to reconsider or change his/her own professional behavior, e.g. to modify a treatment or take previously unconsidered factors into account? Do the findings suggest further investigations? Does the study raise new, hitherto unanswered questions? What are the implications of the results for science, clinical routine, patient care, and medical practice? Are the findings in accord with those of the majority of earlier studies? If not, why might that be? Do the results appear plausible from the biological or medical viewpoint?

Critical analysis of the study's limitations—Might sources of bias, whether random or systematic in nature, have affected the results? Even with painstaking planning and execution of the study, errors cannot be wholly excluded. There may, for instance, be an unexpectedly high rate of loss to follow-up (e.g. through patients moving away or refusing to participate further in the study). When comparing groups one should establish whether there is any intergroup difference in the composition of participants lost to follow-up. Such a discrepancy could potentially conceal a true difference between the groups, e.g. in a case-control study with regard to a risk factor. A difference may also result from positive selection of the study population. The Discussion must draw attention to any such differences and describe the patients who do not complete the study. Possible distortion of the study results by missing values should also be discussed.

Systematic errors are particularly common in epidemiological studies, because these are mostly observational rather than experimental in nature. In case-control studies, a typical source of error is the retrospective determination of the study participants' exposure. Their memories may not be accurate (recall bias). A frequent source of error in cohort studies is confounding. This occurs when two closely connected risk factors are both associated with the dependent variable. Errors of this type can be corrected and revealed by adjustment for the confounding factor. For instance, the fact that smokers drink more coffee than average could lead to the erroneous assumption that drinking coffee causes lung cancer. If potential confounders are not mentioned in the publication, the critical reader should wonder whether the results might not be invalidated by this type of error. If possible confounding factors were not included in the analysis, the potential sources of error should at least be critically

debated. Detailed discussion of sources of error and means of correction can be found in the books by Beaglehole and Webb (17, 18).

Results that do not attain statistical significance must also be published. Unfortunately, greater importance is still often attached to significant results, so that they are more likely to be published than nonsignificant findings. This publication bias leads to systematic distortions in the body of scientific knowledge. According to a recent review this is particularly true for clinical studies (3). Only when all valid results of a well-planned and correctly conducted study are published can useful conclusions be drawn regarding the effect of a risk factor on the occurrence of a disease, the value of a diagnostic procedure, the properties of a substance, or the success of an intervention, e.g. a treatment. The investigator and the journal publishing the article are thus obliged to ensure that decisions on important issues can be taken in full knowledge of all valid, scientifically substantiated findings.

It should not be forgotten that statistical significance, i.e. the minimization of the likelihood of a chance result, is not the same as clinical relevance. With a large enough sample, even minuscule differences can become statistically significant, but the findings are not automatically relevant (13, 19). This is true both for epidemiological studies, from the public health perspective, and for clinical studies, from the clinical perspective. In both cases, careful economic evaluation is required to decide whether to modify or retain existing practices. At the population level one must ask how often the investigated risk factor really occurs and whether a slight increase in risk justifies wide-ranging public health interventions. From the clinical viewpoint, it must be carefully considered whether, for example, the slightly greater efficacy of a new preparation justifies increased costs and possibly a higher incidence of side effects. The reader has to appreciate the difference between statistical significance and clinical relevance in order to evaluate the results properly.

Conclusions

The authors should concentrate on the most important findings. A crucial question is whether the interpretations follow logically from the results. One should avoid conclusions that are supported neither by one's own data nor by the findings of others. It is wrong to refer to an exploratory data analysis as a proof. Even in confirmatory studies, one's own results should, for the sake of consistency, always be considered in light of other investigators' findings. When assessing the results and formulating the conclusions, the weaknesses of the study must be given due consideration. The study can attain objectivity only if the possibility of erroneous or chance results is admitted. The inclusion of nonsignificant results contributes to the credibility of the study. "Not significant" should not be confused with "no association." Significant results should be considered from the viewpoint of biological and medical plausibility.

So-called levels of evidence scales, as used in some American journals, can help the reader decide to what

---

BOX 2

Critical questions

- Does the study pose scientifically interesting questions?
- Are statements and numerical data supported by literature citations?
- Is the topic of the study medically relevant?
- Is the study innovative?
- Does the study investigate the predefined study goals?
- Is the study design apt to address the aims and/or hypotheses?
- Did practical difficulties (e.g. in recruitment or loss to follow-up) lead to major compromises in study implementation compared with the study protocol?
- Was the number of missing values too large to permit meaningful analysis?
- Was the number of cases too small and thus the statistical power of the study too low?
- Was the course of the study poorly or inadequately monitored (missing values, confounding, time infringements)?
- Do the data support the authors' conclusions?
- Do the authors and/or the sponsor of the study have irreconcilable financial or ideological conflicts of interest?

---

extent his/her practice should be affected by the content of a given publication (20). Until all journals offer recommendations of this kind, the individual physician's ability to read scientific texts critically will continue to play a decisive role in determining whether diagnostic and therapeutic practice are based on up-to-date medical knowledge.

References

The references are to be presented in the journal's standard style. The reference list must include all sources cited in the text, tables and figures of the article. It is important to ensure that the references are up to date, in order to make it clear whether the publication incorporates new knowledge. The references cited should help the reader to explore the topic further.

Acknowledgements and conflict of interest statement

This important section must provide information on any sponsors of the study. Any potential conflicts of interest, financial or otherwise, must be revealed in full (21).

Table 2 and *Box 2* summarize the essential questions which, when answered, will reveal the quality of an article. Not all of these questions apply to every publication or every type of study. Further information on the writing of scientific publications is supplied by Gardner et al. (19), Altman (7), and Altman et al. (22). Gardner et al. (23), Altman (7), and the CONSORT Statement (12) provide checklists to assist the evaluation of the statistical content of medical studies.

TABLE 2

Checklist to evaluate the quality of scientific publications

	Yes	Unclear	No
Design			
Is the aim of the study clearly described?	☐	☐	☐
Are the study population(s) and the inclusion and exclusion criteria described in detail?	☐	☐	☐
Were the patients allocated randomly to the different arms of the study?	☐	☐	☐
If yes: Is the method of randomization described?	☐	☐	☐
a) Is the number of cases discussed?	☐	☐	☐
b) Were sufficient cases enrolled (e.g. Power ≥50%)?	☐	☐	☐
Are the methods of measurement (e.g. laboratory examination, questionnaire, diagnostic test) suitable for determination of the target variable (with regard to scale, time of investigation, standardization)?	☐	☐	☐
Is there information regarding data loss (response rates, loss to follow-up, missing values)?	☐	☐	☐
Study inception and implementation			
Are treatment and control groups matched with regard to major relevant characteristics (age, sex, smoking habits etc.)?	☐	☐	☐
Are the drop-outs analyzed for differences between the treatment and control groups?	☐	☐	☐
How many cases were observed over the whole study period?	☐	☐	☐
Are side effects and adverse events during the study period described?	☐	☐	☐
Analysis and evaluation			
Have the correct statistical parameters and methods been selected, and are they clearly described?	☐	☐	☐
Are the statistical analyses clearly described?	☐	☐	☐
Are the important parameters (prognostic factors) included in the analysis or at least discussed?	☐	☐	☐
Is the presentation of the statistical parameters appropriate, comprehensive, and clear?	☐	☐	☐
Are the effect sizes and confidence intervals stated for the principal findings?	☐	☐	☐
Is it apparent why the given study design/statistical methods were chosen?	☐	☐	☐
Are all conclusions supported by the study's findings?	☐	☐	☐

By using a checklist such as this, the statistical and methodological soundness of a study can be assessed and improvements considered.
Not all of the points in this checklist can be used to evaluate all study types; for example, randomization is particularly applicable to clinical studies.

REFERENCES

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, RW Scott: Evidence based medicine: what it is and what it isn't. Editorial. BMJ 1996; 312: 71–2.
2. Albert DA: Deciding whether the conclusion of studies are justified: a review. Med Decision Making 1981; 1: 265–75.
3. Gluud LL: Bias in clinical intervention research. Am J Epidemiol 2006; 163: 493–501.
4. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR: Publication bias in clinical research. Lancet 1991; 337: 867–72.
5. Greenhalgh T: How to read a paper: getting your bearings (deciding what the paper is about). BMJ 1997; 315: 243–6.
6. Kallet H: How to write the methods section of a research paper. Respir Care 2004; 49: 1229–32.
7. Altman DG: Practical statistics for medical research. London: Chapman and Hall 1991.
8. De Simonian R, Charette LJ, Mc Peek B, Mosteller F: Reporting on methods in clinical trials. N Engl J Med 1982; 306: 1332–7.
9. Trampisch HJ, Windeler J, Ehle B: Medizinische Statistik. Berlin, Heidelberg, New York: Springer 2000, 2nd Revised Edition.
10. Klug SJ, Bender R, Blettner M, Lange S: Wichtige epidemiologische Studientypen. Dtsch Med Wochenschr, 2004; 129: T7–T10.
11. Schumacher M, Schulgen G: Methodik klinischer Studien. Methodische Grundlagen der Planung, Durchführung und Auswertung. Berlin: Springer 2008, 3rd Revised Edition.
12. Moher D, Schulz KF, Altman DG für die CONSORT Gruppe: Das CONSORT Statement: Überarbeitete Empfehlungen zur Qualitätsverbesserung von Reports randomisierter Studien im Parallel-Design. Dtsch Med Wochenschr 2004; 129: T16–T20.
13. Blettner M, Heuer C, Razum O: Critical reading of epidemiological papers. A guide. J Public Health 2001; 11: 97–101.
14. Borenstein M: The case for confidence intervals in controlled clinical trials. Control Clin Trias 1994; 15: 411–28.
15. Gardner MJ, Altman DG: Confidence intervals rather than P values: estimation rather than hypothesis testing. BMJ 1986; 292: 746–50.
16. Bortz J, Lienert GA: Kurzgefaßte Statistik für die Klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben. Berlin: Springer 2003; 2nd Edition, 39–45.
17. Beaglehole R, Bonita R, Kjellström T: Einführung in die Epidemiologie. Bern, Göttingen, Toronto, Seattle: Huber 1997.
18. Webb P, Bain C, Pirozzo S: Essential epidemiology. An introduction for students and health professionals. New York: Cambridge University Press 2005.
19. Gardner MJ, Machin D, Brynant TN, Altman DG: Statistics with confidence. Confidence intervals and statistical guidelines. London: BMJ books 2002.

20. Ebell MH, Siwek J, Weiss BD et al.: Simplifying the language of evidence to improve patient care. J Fam Pract 2004; 53: 111–20.

21. Bero LA, Rennie D: Influences on the quality of published drug studies. Int J Technol Assess Health Care 1996; 12: 209–37.

22. Altman DG, Gore SM, Gardner MJ, Pocock SJ: Statistical guidelines for contributers to medical journals. BMJ 1983; 286: 1489–93.

23. Gardner MJ, Machin D, Campbell MJ: Use of check lists in assessing the statistical content of medical studies. BMJ (Clin Res Ed) 1986; 292: 810–2.

Corresponding author
Dr. med. Jean-Baptist du Prel, MPH
Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
Universitätsklinikum Mainz
Obere Zahlbacher Str. 69
55131 Mainz, Germany
duprel@imbei.uni-mainz.de