How to Read and Interpret a Scientific Paper

Özer Makay and Kerstin Lorenz

Contents

A. L. Shifrin et al. (eds.), *Endocrine Surgery Comprehensive Board Exam Guide*,
https://doi.org/10.1007/978-3-030-84737-1_45

45

45.1 Introduction

» The art of knowing is knowing what to ignore. – Jalaleddin Rumi[1]

In the aim to provide optimal patient care, we are appointed to continuously update our medical knowledge. The increasing number of publications, the accelerating new data accessible, and the claim to conform to the standards of evidence-based medicine make it challenging to keep oneself up to date [1]. All the more it is important to select and direct the reading to valuable scientific reports. To invest the necessary time into rewarding reports, some skills of critical appraisal of scientific publications can be refined by applying some basic principles. Next to grasping the medical content of a scientific paper, it is important to interpret and evaluate the data presented; thus, the statics used need to be regarded [2]. However, specifics of statistical analysis will not be addressed in detail in this outline. In the following, the basic principles for evaluation of scientific publications that can be generalized for the fields of experimental, clinical (text and video), and epidemiological medical research are outlined with emphasis on clinical research reports.

45.2 What Scientific Report to Read?

The decision which paper to read is closely linked to the intention behind the reading. In case one keeps up to date by continuous periodical perusal of publications in journals dedicated to the field of one's interest, it is most likely a choice made by a catching title, indication of a novelty, change of established proceedings, and personal interest. For a rough orientation, reading the abstract helps to decide whether to read the full article. Generally, the abstracts mirror the articles' structures (aims, methods, results, conclusions, and keywords) and provide a brief account on the primary objective investigated and the content, so the reader has an idea of what to expect. In case a certain aspect in a given field is to be explored or a quick orientation in a subject is warranted, reading most recent review article(s), ideally by renowned experts in the field, are advisable. For a more in-depth exploration of the chosen topic, the reader can primarily refer to the article's references and from there in the references listed in the cited publications. For fur-

1 Also known as "Mevlana," Rumi was a thirteenth-century poet and Sufi mystic, originally from Greater Khorasan in Greater Iran. Rumi's influence transcends national borders and ethnic divisions. His poems have been widely translated into many of the world's languages and transposed into various formats. Like other mystic and Sufi poets of Persian literature, Rumi's poetry speaks of love which infuses the world. Rumi has been described as the "most popular poet" and the "best selling poet" in the United States (source: ► www.en.wikipedia.org).

ther exploration and search of the most recent update in the topic as in case there are no current reviews available, search of medical databases (PubMed, Cochrane, EBSCOhost, Medscape, Ovid, Embase, etc.) is advised. The given keywords can also be used to search the databases for corresponding publications.

45.3 How to Structure the Reading?

By implication, the reading follows the article's structure. As stated above, the abstract provides a rough orientation of the full article. In case the reader looks for information referring strictly to a certain group of patients, efficient selection of articles will be directed to find inclusion or exclusion criteria firstly in the title and secondly in the methods part of the abstract, and if not found there, consult the methods section in the full article. Whenever the article is valued worthwhile thorough reading by perusal of the abstract, the reader follows the given structure [3, 4]. Most scientific medical publications adhere to this structure:

1. Abstract (see above)
2. Introduction
3. Methods
4. Results
5. Discussion
6. Conclusions
7. References
8. Amendments

2. In the introduction, the reader can expect to be briefed about the objective of the current investigation in the light of previous or ongoing research and basic preconditions that entailed the specific question that is addressed now. The clinical or scientific background presented here is usually backed by the most important and recent accompanying findings and statements from other researchers, and the references are adequately provided.
3. The methods section can be referred to as an indicator for the quality and validity of the study presented. Generally, all procedures and the course of actions as well as the choice of study design are described and argued. Importantly, the study design must be aligned to the aim targeted. Thus, the nature of the study can vary from description to exploration or confirmation. Irrespective of the study type, a precise definition of the endpoint(s) and precise data on the study's details should be described to the degree that a reproduction study according to the information provided can be designed. The methods description correlates with the nature of

the investigation: for laboratory and experimental studies, more details addressing the model used or the execution of the procedural steps are warranted compared to clinical studies. The statistical methods applied should be accurately described and be suitable for the study design and endpoint chosen [2]. According to the nature of the study, some elements are required and indicate the quality of the investigation:

- Experimental/laboratory study: study plan, study protocol, pilot study, study cohort (patient/animal/cell line), control cohort, study location(s), study period, ethics approval, study design, study execution, and accuracy.
- Clinical study: prospective/retrospective, unicentric/bicentric/multicentric, sample size estimation (power calculation), control group (historical, actual, placebo), randomized/non-randomized, and blinded/unblended. Review of the Consolidated Standards of Reporting Trials (CONSORT) diagram typically illustrates the subject recruitment.
- Epidemiological study (intervention, cohort, case control, cross-sectional, ecological): rare diseases/tumors, environmental influences/exposure/agents, single/multiple agents, multiple endpoints, cause/effect and cofactors, standardized conditions/measurement (reliability/validity), scale type, choice of statistical test, description measures (table/graph), mean values, confidence intervals, p-values, and appropriate sample size [5].

4. In the results section, account of all findings is presented ideally without interpretation. The reader should be able to objectively receive the data structured according to the aim of the study and detailed study cohort. Thus, the findings are firstly described followed by all necessary statistical parameters (case number, mean value, confidence interval, variation, statistical significance, effect size), and for clarity, the main findings may be depicted in tables or graphs. It is commonplace that study results with statistical significance will be more likely to be published and recognized; however, this harbors a publication bias as the clinical impact of insignificant results may be important [5]. In some journals, extensive additional data may be presented in a supplement. This data may be positively confirmatory and provide very detailed information for the dedicated reader. Perusal of these supplements is recommended to elucidate the findings and to exclude masking of contradictory data.
5. In the discussion section, the outlined basis initially presented in the introduction is picked up again, and the results of the current study are clearly stated and mir-

rored against previous or actual comparable studies. The reader should be aware if the results presented are plausible and in accord with the methods applied and data provided. Furthermore, it should be made clear if the current data add to the state of knowledge in the field and if the data are reliable enough so that the conclusions will influence or change the presently established practice. Possibly, the current findings need to initiate further investigations before any change on established procedures may be considered or new and unanticipated questions arise. In case the findings of the current study have strong implications to influence the established practice, comparison to other or previous studies should positively correlate in line with the adjustments suggested. Whenever there result contradictory conclusions to current practice or the main corpus of corresponding publications, the discussion section should provide plausible and clear arguments. Any vague or imprecise explanation or missing comment should alert the reader that weaknesses regarding the evidence or the study design may be missed, underreported, or masked.

High standard articles address in the discussion section the limitations identified and clarify whether these will affect the results [3, 4]. Generally, the more open and direct the limitations are detailed, the more sustainable the data presented appear. In critical discussion, possible bias with systematic or random effect will be checked. Selection bias and group discrepancies are particularly detailed. Special attention can be directed to the completeness or lacking of data in the follow-up, e.g., how many patients were lost or dropped out of a study. Good-quality studies describe the reasons for and characteristics of dropouts and estimate the effect of missing data. Furthermore, important factors that can produce errors in clinical studies are confounders. These confounders should ideally be identified and evaluated by the authors, or in lack thereof, the error margin and potential bias should be addressed. Whenever dependent variables are closely associated, the precise impact of a factor may not be discriminated, and this needs to be adjusted for the specific confounder in question. In the discussion section, the prevalent study's findings are to be compared to equivalent investigations. The apparent and important confirmatory as well as differing results are debated by the authors, and the strengths and potential weaknesses of the study introduced are weighed against the prevalent literature.

6. In the conclusions, the authors highlight the most relevant findings of the study presented. Ideally, the potential conclusions are firstly indicated when the results are presented since they must be deduced from the results

and backed by the trial's data. Again, as prepared for in the discussion section, the authors should indicate how the limitations of the study may have affected the results and weigh if the conclusions formulated withstand these arguably. In case the presented study confirms other study's findings, this should be mentioned, as well as authors should explore possible reasons and potential errors in case they arrive at conflicting conclusions. Even studies that state their findings as outlined and do not reach significant results support their credibility and harbor clinical impact.

7. The references section should be complete with regard to the citations used and represent an adequate selection of the important and recent publication in the topic addressed. Sometimes historical references will be essential to include.
8. Bonus information that can help to reflect the credibility of an article can be found in the following:
   - Author's list: are all expected contributors (e.g., all specialties involved in a clinical trial) listed?
   - Institutions: are the authors and institutions by accord likely to be credible?
   - Funding and conflict of interest statement: is any form of funding involved and does this harbor potential conflict of interest to the conduct of the study; is a conflict of interest statement provided and is it plausible with regard to the study?
   - Acknowledgments: some journals restrict the number of authors listed, then contributions by other researchers may be acknowledged in an amendment statement, or the contribution provided is judged to not comply with an authorship.

45.4 What About Evidence and Its Grading in Medicine?

In prevalent clinical questions, especially when contradictory practice patterns and recommendations exist, evidence is desired to choose wisely. In such scenario, one can consult the most recent professional society's guidelines that specify the problem in question or turn to systematic reviews and meta-analyses [6]. With regard to clinical research articles, the lowest level of evidence is represented by case reports and the highest level by data from multicentric randomized studies (RCT). Research articles that present data of a clinical study will usually represent only low grades of evidence and strength of recommendation as its findings will most likely need to be confirmed by others and further studies. Guidelines are ranked by the level of evidence research and the corresponding grad-

ing of recommendations they are based on. In meta-analyses, the level of evidence is rarely stated but can be derived from the quality of the primary studies included. Depending on the adequate and comprehensible statistics applied, meta-analyses can also provide improved levels of evidence compared to primary studies. Systematic reviews are based on a specific clinical problem that can be structured by, e.g., the PICO model (P = patient/population/problem, I = intervention, C = comparison/control, O = outcome/effect).

To quickly find and determine the quality of the best available evidence, it is possible to label the existing evidence. The use of best available evidence in making decisions and the use of levels of evidence and grades of recommendations will improve clinical practice. Several evidence rating scales are available. One of these is the GRADE: Grading of Recommendations Assessment, Development and Evaluation Working Group (modified by the EBM Guidelines Editorial Team) (■ Table 45.1) [7].

Levels of evidence and grades of recommendations according to the SIGN grading system [8–10]:

■ **Table 45.1** Level of evidence in medicine

Code	Quality of evidence	Definition
A	High	Further research is very unlikely to change our confidence in the estimate of effect
		Several high-quality studies with consistent results
		In special cases: one large, high-quality multicenter trial
B	Moderate	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
		One high-quality study
		Several studies with some limitations
C	Low	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
		One or more studies with severe limitations
D	Very Low	Any estimate of effect is very uncertain
		Expert opinion
		No direct research evidence
		One or more studies with very severe limitations

Levels of Evidence

1++ – High-quality meta-analyses, systematic reviews of RCTs, or RCTs with a very low risk of bias

1+ – Well-conducted meta-analyses, systematic reviews, or RCTs with a low risk of bias

1– – Meta-analyses, systematic reviews, or RCTs with a high risk of bias

2++ – High-quality systematic reviews of case control or cohort or studies
High-quality case control or cohort studies with a very low risk of confounding or bias and a high probability that the relationship is causal

2+ – Well-conducted case control or cohort studies with a low risk of confounding or bias and a moderate probability that the relationship is causal

2– – Case control or cohort studies with a high risk of confounding or bias and a significant risk that the relationship is not causal

3 – Non-analytic studies, e.g., case reports, case series

4 – Expert opinion

Grades of Recommendations

A – At least one meta-analysis, systematic review, or RCT rated as 1++ and directly applicable to the target population
A body of evidence consisting principally of studies rated as 1+, directly applicable to the target population, and demonstrating overall consistency of results

B – A body of evidence including studies rated as 2++, directly applicable to the target population, and demonstrating overall consistency of results
Extrapolated evidence from studies rated as 1++ or 1+

C – A body of evidence including studies rated as 2+, directly applicable to the target population, and demonstrating overall consistency of results
Extrapolated evidence from studies rated as 2++

D – Evidence level 3 or 4
Extrapolated evidence from studies rated as 2+

In the following, a choice of questions and aspects is provided to roughly assess research publications like clinical trial data (Table 45.2).

Table 45.2 Quick checkbox on quality and credibility of articles

Question/aspect	Benefit/intention	Check for/caveat	*In favor for* quality/ credibility of article	*Against* quality/ credibility of article
Study design				
Aim of the study	Primary endpoint clearly defined	Primary/ secondary endpoints maintained	Primary/secondary endpoints maintained	Change in endpoints
Question of clinical value/interest/ relevance/innovative			+++	
Comparative statements/data	Research environment	Citation correctness/ comprehensive-ness	Support of state-ments/data	
Data support conclusions			+++	
Conflict of author statement	Credibility, independence of data	Imprecise/ missing statements		
Study methods				
Type of study determined	Confirmatory vs. exploratory vs. descriptive study design	Study type in line with aim	Type of study clearly stated; defined goals are followed	Missing declara-tion of study type
Study population	Area, period, recruitment, power calcula-tion	Imprecise/ missing statements	Adequate number included, power adequate	Cohort too small/ dropouts or exits too high
Dropouts/exits	Traceability of study subjects	Imprecise/ missing statement, only exits	Coherent numbers and reasons for dropouts/crossovers	Contradictory numbers, no dropout informa-tion or only exit
Monitoring	Data complete-ness/confounders	Missing data/ lack of identifica-tion of confounders		

(continued)

Table 45.2 (continued)

Question/aspect	Benefit/intention	Check for/caveat	*In favor for* quality/ credibility of article	*Against* quality/ credibility of article
Study statistics				
Multicenter trial	Larger/diverse sample size, rare disease/event	Firm regulation/ adherence protocol Outcome validated/ reproducible	Compromise standardization intervention + outcome	Sites are covariates
Dichotomous outcome measure	Treatment comparison	Clear definition of endpoint		
Continuous outcome measure	More power to detect differences between groups			
Correlation and p-value	Differences distinct	p-value may be very sensitive for weak correlations	95% confidence interval (CI) provided	No 95% confidence interval (CI) provided

45.5 Checkbox Basic Questions on Study Quality and Credibility

- Sample size adequate?
- Recruitment method and study population adequate?
- Difference of samples in population relevant?
- Number of subjects in study arms and number of and reasons for dropouts or crossover.
- Bias in study design? Subject in treatment study arm, blinding (single/double, process successful/unsuccessful?), and comparator.
- Randomized study: intention-to-treat (ITT) analysis provided?
- Study type: what is tested for – superiority, equivalence, non-inferiority? Are the definitions of the treatment arms different at the start? Check sample-size calculation; larger samples are needed for equivalence/non-inferiority. Inappropriate subsequent change of study type?
- Definition of success? How is it measured? Is it validated? Precise and reproducible? Is it clinically relevant?
- Primary outcome: focus kept?
- Surrogate (secondary) outcomes: strong independent association of surrogate and desired outcome? Surrogate and clinical outcome improve concordantly.
- Dichotomous and continuous outcomes, correlations, time-to-event endpoints: are endpoints defined clearly?
- Confidence interval 95% provided?

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Scott RW. Evidence base medicine: what is and what isn't. BMJ. 1996;312:71–2.
2. Govani SM, Higgins PD. How to read a clinical trial paper: a lesson in basic trial statistics. Gastroenterol Hepatol. 2012;8:241–8.
3. Du Prel JB, Röhrig B, Blettner M. Critical appraisal of scientific articles. Dtsch Arztebl Int. 2009;106:100–5.
4. Röhrig B, du Prel JB, Blettner M. Study design in medical research. Dtsch Arztebl Int. 2009;106:184–9.
5. Amrhein V, Greenland S, MCShane B. Retire statistical significance. Nature. 2019;567:305–7.
6. Mak A, Cheung MW, Fu EY, Ho RC. Meta-analysis in medicine: an introduction. Int J Rheumat Dis. 2010;13:101–4.
7. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ (Clin Res ed). 2008;336(7650):924–6.
8. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ. 2001;323(7308):334–6.
9. The GRADE working group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1–8.
10. Siglo BA, Murad MH, Schünemann HJ, Kunz R, Vigersky R, Guyatt GH, Montori VM. A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the Grading of recommendations, Assessment, Development, and Evaluation system. J Clin Endocrinol Metab. 2008;93:666–7.