How to read and evaluate a scientific paper?

Mohey Eldin Ragab Elbanna, MD
Department of General Surgery, Ain-Shams University, Cairo, Egypt

Introduction:

To write a scientific paper, you need to be capable of reading scientific papers. Additionally, it is essential to master the process of critical evaluation of a scientific paper, in order to decide whether its content of information or conclusion can be used in solving a clinical problem. In this regard there are two categories of clinical problems; namely: How to handle an individual patient, and how to decide on clinical guidelines or standard treatment options for groups of patients. For both categories, a critical appraisal of published studies provides a good basis for decisions.

Critical appraisal starts with a well-formulated question. This typically has four parts and the mnemonic is PICO: Patient, Intervention, Comparator, and Outcome. You need to define the patient's health problem, the interventions you want to evaluate or compare, and the important clinical changes you expect the intervention to provide. This applies to questions about individual patients as well as to populations.

For individual patients, the question is usually rather easy to formulate; so clearly, it can be answered by study results: "In this 40-year-old woman that has been diagnosed with acute cholecystitis, how well would a laparoscopic cholecystectomy relieve the patient (Outcome) as compared to conservative treatment and interval cholecystectomy (Comparator)?"

When considering groups of patients, the process can be quite similar. The question above would then translate to a parallel format: "In adults (aged 20 to 50 years), how well would ..." But for a patient population, other aspects may also apply. It may be necessary not only to consider a certain drug or treatment program, but also to think of whether the health service organization can provide a defined treatment to all patients who have a certain health problem. For instance, would there be sufficiently enough experienced surgeons and facilities to perform laparoscopic cholecystectomy for all new patients with acute cholecystitis?

Sometimes, the need to answer a question arises because new and promising treatments have become available, or a new diagnostic method has been developed. Questions may have to do with screening for possible disease or prognosis of a chronic disease, and in these cases, you need to work in the time perspective. The reason for asking often involves budgeting: You have to decide which ones of the available treatments are the most cost-effective.

For each type of question, certain typical elements need to be considered, and these differ according to the type of study that best answers this kind of question.

The effectiveness question is best answered with data from randomized trials. Diagnostic studies must be designed so that both types of tests to be compared are taken from each study participant, and results from the 'new' test must be interpreted without knowledge about the results of the 'old' test.

Prognostic studies must fulfill their own quality criteria which again are different.

Economic evaluations have rules of their own, plus a challenge of transferring the cost information reliably between different healthcare systems.

If your question has to do with interventions, the Cochrane library is the place to start with. The Guidelines International Network (G-I-N) provides comprehensive information about guidelines for member organizations, and basic information for anyone who visits the website (www.g-i-n.net). An instrument for guidelines evaluation is (AGREE) and it is also available on the G-I-N website.

The article:What? who? and where?

The title of the study can be used to indicate what the study is about. But be skeptical. It is correct that most authors try to give sufficient information about the content of the article in the title, yet some market their work with titles that promise more than is provided by the text.

Titles, even those claiming a definite result, should be considered simply as "eye-catchers".

Next, the institution a paper originates from can give weight to credibility. Honorable and well-known institutions usually have some internal reviewing process to guarantee a certain standard of scientific work. The authors' credentials are as important. Some investigators stand for solid, carefully performed work.

The money that financed the research is to be considered. Look for the sponsoring organization. If the scientific project is supported by a private company or a private granting agency, the reader has to be more careful than when the money for the project has come from public funds.

Publication bias has to be scrutinized. Publication bias is multifactorial. The most common bias is underreporting. It means avoiding reporting some of the results, if they compromise the impression about investigated diagnostic or treatment method.

Publication in a high quality journal helps to build a trust in this piece of research. Some words of warning are, however, needed here. A core idea of using the evidence-based approach is that users can appraise the quality of the information on their own. Although good journals in general publish better studies than bad journals, mistakes occur even in top clinical papers. And although experienced researchers generally do better research than novices, it happens that well-known scientists make mistakes. So the impact factor of the journal, the names on the authors' list, or even the fact that a good and trustworthy friend of yours has published the article – these are no guarantee that the results of the study can be trusted or used.

Credentials do work the other way around: Sometimes certain information – or a lack of it – is a warning sign about the quality of the study.

Research funded by sources that have a vested interest in the topic may be well done, but disagreeable results may remain unpublished. If the team of authors does not include members who have clinical experience, the entire question may be irrelevant for practicing professionals. And the more obscure the journal, the less likely it is that manuscripts have been through a thorough quality check. Luckily, you can look at the studies on your own, using checklists developed to help appraise them.

Critical appraisal:

After the title, authors and institution, it is natural to take a look at the abstract. It is a common belief that the answer to your question is there in a concise form. Experience shows, however, that it is risky to consider the abstract a true summary of the content of the article. Use it to decide whether the subject is really what you are interested in, and to avoid reading a study that you already can see here to be based on a poor design and not well carried out.

Instead of reading the article from beginning to end, in critical appraisal you do it in a structured fashion, answering questions about its quality.

These are listed in **Table(1)**. Bear in mind that if you find serious flaws, there is no reason to go on –you cannot trust the results in the end, and there are plenty of other good articles to be read. This speeds up your reading and helps you discard bad studies early, without having to spend too much time on them.

Start by looking for a description of the problem the study addresses: is there a well-defined research question? This allows you to decide whether the study design is the best to answer the type of question. The research question is most often found in the last sentence of introduction, but you may need to look in the methods section for more details. The description of the interventions is an important issue.

Table (1): Critical appraisal checklist.

I. Study validity:

1. Study question:
   Is there a well-defined research question that can be answered using this study design?
2. Randomization:
   Were the patients randomized to the intervention and control groups by a method that ensured the assignment was random?
   Was the randomization list concealed from patients, clinicians and researchers?
   Were the patients in the groups similar at the start of the study?
3. Blinding:
   Were the patients and the clinicians kept blinded (masked) to which treatment was being given?
   Were they kept blinded until the end of the study?
4. Follow-up:
   Were all patients accounted for at the end of the study?
   If not, how many patients were lost to follow-up and for what reasons?
   Were the patients analyzed in the groups they originally were randomized to?
5. Interventions and co-interventions:
   Were the interventions described in sufficient detail to be repeatable by others?
   Were the two groups cared for in a similar way except for the study intervention?

II. Results:

6. Selection of outcomes:
   Does the article report all relevant outcomes including side effects?
7. Effect size:
   Was there a difference between the outcomes of the treatments, and how big was the difference?
   How reliable is the estimate: what are the confidence intervals?

III. Applicability:

8. Using results in your own setting:
   Are your patients so different from those studied that the results may not apply to them?
   Is your working environment so different from the one in the study that the methods could not be used there?

Study validity:

Randomization:

You look for the validity of the study by checking the way it was carried out. The first question is about randomization and how it was performed. The reason for this is to get two groups that are as similar as possible. Age, gender, stage of disease, etc. may influence the outcome of a treatment, and there may be factors that cannot be taken into consideration because they are unknown.

Ideally all these are distributed equally between the groups. The randomization procedure makes it possible for the statistician to assess the effect of chance.

Your next concern is whether patients or their caregivers knew beforehand who would get what treatment. Allocation concealment is essential, because otherwise professionals or patients can (knowingly or subconsciously) select which treatment would be best for a certain patient. This can be estimated by looking at the description of randomization procedure.

There must be a description of the patient characteristics at the start of the study, so that you may judge whether the patients in the intervention and control groups were sufficiently similar. The smaller the number of patients, the more likely it is that the groups differ in some important background

characteristic, such as gender or age distribution. This may happen even in well-randomized studies, so you need to judge whether the difference could have an effect on the results.

Blinding (masking) and follow-up:

The next question to ask is: Were the patients and the clinicians kept blinded (or masked, which is the less aggressive term nowadays!) to which treatment was being given, until the end of the study? If this is not the case, the attitudes of patients and clinicians may in subtle ways have introduced bias and so led to false conclusions.

Sometimes, it is not possible to mask the participants entirely to what kind of therapy is given. You need to think to which extent the authors actually could have managed to keep patients and researchers blinded to the treatment, and to what effect this would have on the results.

You then look for the table or flow chart that shows the proportion of patients who were followed up until the end of the study, how many were lost and for what reasons. Ideally, all patients should be accounted for at the end of the study. If the number of dropouts is high, you can calculate a 'worst-case scenario' result by assuming that all dropouts in the intervention group did not improve at all, and all those in the control group were perfectly well after the treatment. This may dramatically change the results.

Interventions and co-interventions:

The treatments must be described in sufficient detail for you to understand what has been done, and to make sure that the two groups were cared for in completely the same way except for the study intervention.

Remember to look both in the methods and the discussion sections of the article. The authors should tell you about the strengths and weaknesses of their procedures.

The last validity question has to do with co-interventions. Optimally, the only difference in treatment is the intervention to be studied. Sometimes, however, the patients are allowed to take certain medications or receive some postoperative measures or use other home therapies, in addition to the study interventions. These should be recorded and reported. What does this study tell about patients' use of other treatments?

Evaluating the results:

If you are satisfied that the study is likely to have produced valid results, you are ready to look at the results section of the article. If you think the study has serious flaws, then forget about the results!

A low quality study cannot give useful results. On the other hand a good study is done well enough and reported openly enough so that you have to look at the results.

Start by looking at the tables. Newcomers to critical appraisal of articles often look for the results in the text only. Actually, most of the information and the real fun of reading are in the tables.

From tables and figures try to find out if there are the outcomes of the treatment, and how big the difference is.

Any difference may be due to the treatment or to chance. Here comes the value of the magic P.

The magic P:

Most conclusions given in biomedical papers are strengthened by inferential statistics yielding a p-value. The p-value has a very simple meaning; $p<0.05$ tells you that there is only a less than 5 % chance that the result is only a product of chance. In statistical terms, it means the null hypothesis is rejected.

This magic bullet of biomedical analysis has several important requisites. In addition to the usual requirement of random sampling and random allocation of treatments, there are many methods by which this p-value can be obtained. The precisely defined question of research will decide whether the method used is appropriate or not.

Even if appropriate, the magnitude of the p-value does not give any indication whether the result reported is clinically relevant. In practice, neither the presence nor the absence of statistical significance has a structurally reliable relationship to the extent of clinical importance.

Applying the information:

If you have decided that this valid study has shown an important effect, the next step is to consider if the results may be applied in your setting. Are your patients different from those in the study groups? Is the health gain large enough for your patients? Is the effect more important than the risk of side effects and worth the economical cost?

Even if there is no difference in clinical outcomes, you may be able to make a clear choice if one of the treatments is cheaper.

Look for data to compare the study patients with your own population.

What kind of patients is not included? In some studies, you may find that the study population after exclusions and drop-outs is too special, that your kind of patients is not included at all.

Before making up your mind about the quality of the study, you may go back to the journal's online version. These often publish responses from their readers right after publication. You may find it useful to compare your conclusions with this type of comments. Some of the commentators have special knowledge that may confirm or challenge your conclusions.

The ten rules:

Yancy (1990) recommended ten rules for reading clinical research reports. In short they are:

1- Be skeptical.
2- Look for the data.
3- Differentiate between descriptive and inferential statistics.
4- Question the validity of all descriptive statistics.
5- Question the validity of all inferential statistics.
6- Be wary of correlation and regression analysis.
7- Identify the population sampled.
8- Identify the type of the study.
9- Look for indices of magnitude of treatment effects.
10- Draw your own conclusions.

In conclusion, these are some of the rules the surgeon needs to follow to critically appraise a scientific paper. With more reading and a deeper surgical experience, the surgeon becomes more capable of weighing the evidence and accurately evaluating publishable surgical studies.

References:

1- Makela M, Witt K: How to read a paper: Critical appraisal of studies for application in healthcare. *Singapore Med J* 2005; 46(3): 108.
2- Rothmund M, Stinner B: How I read and assess a scientific paper. In: Surgical research: Basic principles and clinical practice. Troidl H, McKenneally MF, Mulder DS, Wechsler AS, McPeek B, Spitzer WD (Editors); Springer Verlag (Publisher); 3rd edn. 1998; p.69-74.
3- Yancy JM: Ten Rules for reading clinical research reports. *Am J Surg* 1990; 159: 533-539.