

How to critically appraise a research paper

Derek Alderson

Abstract

The ability to critically analyse surgical papers is an important skill that all surgeons should possess. This article describes a simple two-step approach to the appraisal of scientific literature. It relies on a rapid review based on the abstract followed by a detailed review based on simple checklists. These can be applied to systematic reviews, randomized trials and observational studies. At the end of the process, the quality of the methods, quality of the results and applicability of the work are evaluated to formulate an opinion on overall quality. The article ends with a suggested format for writing a report based on such an appraisal.

Keywords critical appraisal; scientific writing

Introduction

An ability to read, analyse and critically appraise the surgical literature is a skill that every surgeon should have. In the same way that excellence with practical skills is acquired from a knowledge base followed by repetitive practice, so the art of reading a scientific paper and understanding its value is a combination of theoretical knowledge and practice. Without this skill, the surgeon cannot separate those contributions that say something useful from those that are worthless. It is certainly impossible to referee the unpublished work of others without a clear understanding of the critical review process. This is not just an academic exercise, but rather a fundamental attribute without which, the surgeon is unlikely to make any advances.

This article sets out a method for the critical appraisal of the surgical scientific literature and covers randomized trials, systematic reviews and non-experimental (observational) studies, including those related to prognosis and diagnosis. I have specifically avoided animal or in-vitro laboratory-based experiments, although many of the general points made below within the overall approach to critical appraisal can be applied.

There are four essentials that characterize every decent surgical paper. The background to the work must be scientifically credible, it must demonstrate originality, be written clearly and have relevance to clinical surgery. When you have finished your critical appraisal, it is always worth returning to these four fundamentals to be sure that each of these points has been met.

Whether reviewing as a journal referee, assessing the work of others in the context of your own research or reading a paper for scientific interest, a two-stage process is recommended; this involves a quick review followed by the detailed assessment.

Derek Alderson MD FRCS is the Barling Professor of Surgery at the University of Birmingham, Birmingham, UK. Conflicts of interest: none declared.

The quick review

This is done by reading the title and the abstract. A good title is succinct and should state clearly what the study is about. If the abstract is not comprehensible, and fails to give you a clear understanding of the content of the article, it is unlikely that the detailed assessment will be favourable. The ideal abstract should be able to answer the following seven questions:

- Why was the study done?
- Who was studied?
- How was the study done?
- What interventions were involved?
- What outcome measures were used?
- How many subjects were involved?
- What is the significance of the study?

Very few abstracts answer all seven points, but you should make a mental note of the ones that are not mentioned and specifically make sure that they are covered in the main text. In the context of reviewing literature relevant to your own work, you should now be able to decide if the main text is worth careful appraisal.

The detailed review

This phase of the appraisal process demands that the reader makes three assessments, of the:

- Quality of the study design
- Quality of the results
- Applicability of the results.

You should be confident that a paper is satisfactory in all three areas, in order to rely on it as reasonable evidence. This detailed review is best undertaken with a checklist of items that should be specifically sought in order to arrive at a reasonable judgment about the overall quality of an article. Irrespective of study type, many questions are common to all or simply need to be paraphrased to apply. It is easiest to complete the checklist as you read through the paper.

In order to simplify this article, the detailed review is described for a randomized trial using the checklist of questions, followed by the variations used in the evaluation of other types of study. Background details for the adapted checklists used here can be found at: <http://www.auckland.ac.nz/population-health/epidemiology-biostats/epiq> <http://www.strobe-statement.org>.

Read the introduction

A well-written introduction must outline the problem in a way that whatever is being tested seems a scientifically credible proposition. Make sure that the following four questions can be answered:

- Is it scientifically credible?
- Are the aims stated clearly?
- Is there a novel element?
- Is it all worthwhile?

If this scientific background is weak or illogical or the aims are not clear, it is unlikely that the paper will be of reasonable quality. Decisions on novelty and relevance are less easily reached at this stage, so assessing their importance is best left until the end.

Randomized controlled trials (RCTs)

An RCT is the most valid method to test effectiveness and efficacy. It relies on stratification to minimize confounding, but this does not eliminate all forms of bias. Many surgical RCTs relate to a specific subset of patients and are not always applicable to the entire population. All randomized trials should conform with CONSORT (consolidated standards for reporting randomized trials) guidelines and a flow chart depicting the groups from allocation on the basis of eligibility, through to study completion should be seen as essential. Beware the use of a clinically unimportant primary end-point or ambitious power calculation designed to create a study requiring relatively few participants. Studies from large centres dominated by tertiary referral practices may not be generalizable. Beware the poor result in the control arm that leads to a positive result in the experimental arm. These last three points are unfortunately not rare in surgical RCTs.

The RCT checklist

Read and assess the quality of the methods.

Participants

- Were the selection criteria clear, well-defined and appropriate?

Study groups (experimental and control)

- Were they well-defined and replicable?
- Was randomization concealed and successful?
- Was analysis based on randomization (intention to treat)?
- Was blinding possible?
- Were groups treated equally?
- Was compliance measured?

Outcomes

- What were the outcome measures and were they important?
- Were they well-defined and replicable?
- Was outcome assessment blind?

Time

- Was follow-up time adequate?

Rate the quality of the methods.

Read and assess the study results.

- What measures were used to describe outcomes?
- What measures of precision were used (means, medians, statistical tests and levels of significance)?
- Were numerators and denominators clear?
- If no significant effects, was power sufficient?
- If multi-centred, were results homogeneous?

Rate the quality of the results.

Read the discussion.

The discussion is probably the most variable part of an article since the writer decides how and where to place specific emphasis on the main findings. This section should, nevertheless, contain a summary of the key results with reference to the study objectives. It should be an interpretation of the results and other relevant contextual evidence including the principles and relationships learned. The limitations of the study should be

highlighted and the generalizability (external validity, applicability) of the study made clear.

Appropriate information about funding or potential conflicts of interest should appear here. The biggest problems are the unnecessary inclusion of extraneous material and repetition of introduction and results.

Assess the applicability of the study.

- Was the source population well-described?
- Were participant's representative?
- Was the study setting well-described?
- Was the management of the comparison group relevant?
- Were all important outcomes considered: benefits, harms, costs?

Rate the quality of the study applicability.

Systematic reviews

Systematic reviews involve 'standards' to ensure completeness and validity.

The evidence-based decision is made on the basis of the systematic review of evidence rather than the individual studies. A number of scoring systems have been developed to quantify the quality of the evidence within each evaluated study. It can be undertaken using non-experimental studies, but should conform to the PRISMA (preferred reporting items for systematic reviews and meta-analyses). Meta-analysis simply refers to the statistical methodology applied for analysis of the selected data. Access to original data, rather than information available from an original article, may be important.

The systematic review checklist

Read and assess the validity of the review.

Participants

- Were the selection criteria for studies clear, well-defined and appropriate?
- Was the search strategy comprehensive and complete?
- How was the validity of each study assessed?

Study groups

- What were the exposures and comparisons?
- Were they well-defined and replicable?
- Was assignment to groups randomized in all studies and was this concealed?
- Was randomization successful in all studies and if not, how was confounding dealt with?
- Was analysis based on randomization in all studies?

Outcomes

- What were the outcome measures?
- Were they well-defined, replicable and similar in all studies?
- Was follow-up sufficient in all studies?

Time

- Was follow-up time adequate?

Rate the quality of the study design.

Read and assess the results of the review.

- What measures of occurrence and exposure were reported for each study?
- What measures of precision were used?
- Could useful summary effects estimates be calculated?
- What was the precision of these effects?
- If no significant effects, was power sufficient?
- Were effects estimates consistent?

Rate the quality of the study results.

Read the discussion.

Assess the applicability of the study.

- Were the source populations well-described?
- Were participant's representative?
- Were characteristics of the study settings well-described?
- Can the relevance of the comparator groups and the applicability of exposure groups be determined?
- Were all important outcomes considered: benefits, harms, costs?

Rate the quality of the study applicability.

Systematic reviews and meta-analyses are popular in surgery, but too often the quality of the available data and heterogeneity in design or reporting methodology make these studies of limited value. A poorly conducted or inappropriate systematic review should not be viewed as superior to any other type of evidence.

Non-experimental (observational) studies about benefit, harm or causation

These studies represent the greatest volume of surgical literature; they include longitudinal cohort studies, cross-sectional studies and those involving case–controls. Most importantly, the investigator does not control allocation of the exposure (treatment) to participants. At most, the investigator categorizes patients into exposure and comparison sub-groups. The consequence is the potential for confounding. These studies are, however, appropriate when time between exposure and outcome is very long (the development of gastric cancer in patients who have undergone previous gastric surgery for benign disease) or where the outcome of interest is uncommon (the risk of recurrent nerve damage after thyroid surgery). They are appropriate for tests about diagnosis and prognosis.

The observational study checklist

Read and assess the validity of the study.

Participants

- Is the study type clear (cohort, case–control, cross-sectional)?
- Is the setting clear (locations, dates, recruitment period, exposure, follow-up, data collection)?
- Are the selection criteria clear, well-defined and appropriate (size of study, sources and methods of selection)?

Study groups

- Were they well-defined and replicable (information bias, selection bias)?
- Was the measurement of variables similar and valid (nature of data sources, measurements, missing data, statistical methods)?

- Were exposure and comparison groups similar at the start (demographic and diagnostic criteria)?
- Were participants analysed in the originally assigned groups?
- Was any blinding possible?
- Were the groups treated equally?

Outcomes

- What were the outcome measures?
- Well-defined and replicable (effect modifiers, confounders)?
- Drop-outs?
- Was outcome assessment blind?

Time

- Was follow-up time adequate?

Rate the quality of the study design.

Read and assess the study results.

- Were participants well-characterized (potential eligibility, examined, confirmed, included, completed study, analysed, reasons for non-participation)?
- What measures were used to describe outcome (descriptive data for participants, measurement of variables)?
- What measures of precision were used (unadjusted and confounder-adjusted estimates, confidence intervals, category boundaries)?
- Were numerators and denominators clear (sub-group analyses)?
- If no significant effects, was power sufficient?

Rate the quality of the study results.

Read the discussion.

Assess the applicability of the study.

- Was the source population well-described?
- Were participant's representative?
- Was the study setting well-described?
- Was the management of the comparison group relevant?
- Were all important outcomes considered: benefits, harms, costs?

Rate the quality of the study applicability.

Studies of the accuracy of diagnostic tests

These are best done in a non-experimental cross-sectional study. It is important to remember that the outcome is the test result.

The diagnostic test checklist

Read and assess the validity of the study.

Participants

- Were the selection criteria clear, well-defined and appropriate?

Study groups

- Was the reference standard of diagnosis stated clearly?
- Was it independent and valid?
- Was the reference standard applied regardless of test result?
- Did any blinding take place?

Outcomes

- What tests were used?
- Were they clearly defined and replicable?
- Was the test applied regardless of the reference standard result?
- Was the test assessment blind to the reference standard result?
- Was the test validated in a second independent group?

Rate the quality of the study design.

Read and assess study results.

- What measures of test accuracy were reported?
- What measures of precision were used?
- Could useful measures of test accuracy be calculated?
- If no significant effects, was power sufficient?

Rate the quality of the study results.

Read the discussion.

Assess the applicability of the study.

- Was the source population well-described?
- Were participant's representative?
- Was the study setting well-described?
- Can sensible estimates of pre-test probabilities be determined?
- Will post-test probabilities affect management or help patients?
- Is the test affordable, available and reproducible?

Rate the quality of the study applicability.

Studies about prognosis

These are best done in a longitudinal cohort study. The study population must be well-characterized. It is worth remembering that prognostic factors do not have to be causal and confounders can be prognostic factors. Beware of the use of the control group from an RCT; it is unlikely to be representative of the population as a whole, so any prognostic factor derived from such a group will only apply to a population subset.

The prognostic studies checklist

Read and assess the validity of the study.

Participants

- Were the selection criteria clear, well-defined and appropriate?
- Were participants at a common point in the course of their disease?

Study groups

- Were the prognostic groups well-defined and replicable?
- Was measurement of variables similar in all groups?
- Were the prognostic groups similar at the start except for the study factors?
- Were participants analysed in the originally assigned groups?
- Was blinding possible?
- Were the groups treated equally?
- Were prognostic factors re-measured during follow-up?

Outcomes

- What outcome measures were used?

- Were they clearly defined and replicable?
- Were there any drop-outs?
- Was the assessment blind?
- Was follow-up long enough to detect important prognostic factors?

Rate the quality of the study design.

Read and assess the study results.

- What measures of prognosis and differences were reported?
- What measures of precision were used?
- Could useful measures of prognosis be calculated?
- Was the precision of the prognostic estimates sufficient?
- If no significant effects, was power sufficient?

Rate the quality of the study results.

Read the discussion.

Assess study applicability.

- Was the source population well-described?
- Were participants representative?
- Was the study setting well-described?
- Can the applicability/relevance of the prognostic factors be determined?
- Were all important outcomes considered?
- Will the prognostic information impact on management?

Rate the quality of the study applicability.

Writing your critique

Start with an introductory paragraph that says what kind of study this is, what the authors set out to do, what the main findings were and what conclusions were drawn. This ensures that your understanding of an article is made clear.

Consider the need for general remarks including issues relating to language, grammar, length and style that may have contributed to clarity.

Introduce a series of numbered points with a sentence such as 'the following points merit consideration'. List these numbered points as short sentences to cover:

- Scientific credibility and originality, if appropriate
- Study design
- Study results
- Study applicability.

Indicate where defects occur and what clarification is needed. Criticism should be objective and evidence (reference) based whenever possible. The strengths and weaknesses of an article should be obvious from a well-written critique. It is often worth concluding these numbered points with a statement regarding the extent to which you feel the conclusions drawn are justified.

Summary

This short article describes a simple two-step approach to the appraisal of scientific literature. This will help you in the preparation of meaningful reviews/reports of scientific publications, whether they are systematic reviews, randomized trials or observational studies. Checklists are provided for each and when writing your final report, you should ensure that the majority, if not all, of the checklist questions have been addressed. Your written critique should be constructive, with suggestions for improvement of the manuscript and the inclusion of additional references as required. ◆