

How to critically appraise an article

Jane M Young* and Michael J Solomon

SUMMARY

Critical appraisal is a systematic process used to identify the strengths and weaknesses of a research article in order to assess the usefulness and validity of research findings. The most important components of a critical appraisal are an evaluation of the appropriateness of the study design for the research question and a careful assessment of the key methodological features of this design. Other factors that also should be considered include the suitability of the statistical methods used and their subsequent interpretation, potential conflicts of interest and the relevance of the research to one's own practice. This Review presents a 10-step guide to critical appraisal that aims to assist clinicians to identify the most relevant high-quality studies available to guide their clinical practice.

KEYWORDS bias, conflict of interest, critical appraisal, study design

REVIEW CRITERIA

MEDLINE and Google Scholar were searched in October 2008 for English language articles published between 1996 and 2008. The keywords "critical appraisal" were combined with the following keywords "skills", "medical literature", "research", "randomized trial", "cohort", "cross-sectional", "case series", "diagnostic test", "systematic review", "meta-analysis", "economic", "validity", "bias", "generalizability". Additional searches were conducted using the terms "methodological quality", "conflict of interest" and "responsible conduct". Relevant review articles and original articles were retrieved and their reference lists searched for additional articles. Critical-appraisal tools and checklists were accessed from the web pages of organizations identified in the search. Textbooks on clinical epidemiology were referred to for definitions.

JM Young is an Associate Professor of Public Health and the Executive Director of the Surgical Outcomes Research Centre at the University of Sydney and Sydney South-West Area Health Service, Sydney, and MJ Solomon is Head of the Surgical Outcomes Research Centre and Director of Colorectal Research at the University of Sydney and Sydney South-West Area Health Service, Sydney, Australia.

Correspondence

*Surgical Outcomes Research Centre (SOuRCe), Royal Prince Alfred Hospital, PO Box M157, Missenden Road, NSW 2050, Australia
jyoung@email.cs.nsw.gov.au

Received 10 August 2008 Accepted 3 November 2008 Published online 20 January 2009

www.nature.com/clinicalpractice
doi:10.1038/npcgasthep1331

INTRODUCTION

To practice evidence-based medicine, clinicians need to apply the findings of scientific research to the circumstances of individual patients as part of their clinical decision-making process. Clinicians, therefore, must be able to select and appraise scientific literature that is relevant to their field, understand the implications of research findings for individual patients, elicit patients' own preferences and develop an appropriate management plan based on the combination of this information. Each of these tasks presents its own challenges, but the sheer volume of medical literature means that the first step (that of selecting and appraising scientific evidence) can be daunting. The number of new medical research articles published each year continually increases, and more than 12,000 new articles, including papers on in excess of 300 randomized controlled trials (RCTs), are added to the MEDLINE database each week.^{1,2} One practical way that clinicians can manage this 'information overload'² is to develop efficient skills in critical appraisal, which enable them focus on only the highest-quality studies that will guide their clinical practice and to extrapolate information when necessary from studies of less rigorous design if high-quality trials are unavailable.

Critical appraisal has been defined as the "...application of rules of evidence to a study to assess the validity of the data, completeness of reporting, methods and procedures, conclusions, compliance with ethical standards, etc. The rules of evidence vary with circumstances."³ Although the methodological criteria by which the validity of a study is assessed will vary according to its design, some general principles underpin the evaluation of any research study. Various guidelines and assessment tools have been developed to provide a structured approach to the process of critical appraisal for clinicians.⁴⁻¹⁴

Despite the plethora of documents available to guide the process, no 'gold-standard' instrument for critical appraisal exists. The criteria used to assess the validity and relevance of scientific

literature are not static; they must evolve with improvements in understanding of the important sources of bias inherent in different study designs, and increased awareness of the potential influence of other nonmethodological factors, such as conflicts of interest.¹⁵ A structured approach to critical appraisal could potentially improve the quality of this process, and simple checklists can be useful to screen out research that is of low quality or of little relevance.¹⁶ This Review presents a guide to the critical-appraisal process.

SELECTION AND CRITICAL APPRAISAL OF RESEARCH LITERATURE

Ten key questions (Box 1) can be used to assess the validity and relevance of a research article. These questions can assist clinicians to identify the most relevant, high-quality studies that are available to guide their clinical practice.

Is the study's research question relevant?

Even if a study is of the highest methodological rigor, it is of little value unless it addresses an important topic and adds to what is already known about that subject.¹⁷ The assessment of whether the research question is relevant is inevitably based on subjective opinion, as what might be crucial to some will be irrelevant to others. Nonetheless, the first question to ask of any research article is whether its topic is relevant to one's own field of work.

Does the study add anything new?

Scientific-research endeavor is often likened to 'standing on the shoulders of giants', because new ideas and knowledge are developed on the basis of previous work.¹⁸ Seminal research papers that make a substantive new contribution to knowledge are a relative rarity, but research that makes an incremental advance can also be of value. For example, a study might increase confidence in the validity of previous research by replicating its findings, or might enhance the ability to generalize a study by extending the original research findings to a new population of patients or clinical context.¹⁷

What type of research question does the study pose?

The most fundamental task of critical appraisal is to identify the specific research question that an article addresses, as this process will determine the optimal study design and have a major bearing on the importance and relevance of the findings. A

Box 1 Ten questions to ask when critically appraising a research article.

- Is the study question relevant?
- Does the study add anything new?
- What type of research question is being asked?
- Was the study design appropriate for the research question?
- Did the study methods address the most important potential sources of bias?
- Was the study performed according to the original protocol?
- Does the study test a stated hypothesis?
- Were the statistical analyses performed correctly?
- Do the data justify the conclusions?
- Are there any conflicts of interest?

well-developed research question usually identifies three components: the group or population of patients, the studied parameter (e.g. a therapy or clinical intervention) and the outcomes of interest.¹⁰ In general, clinical research questions fall into two distinct categories, below.

Questions about the effectiveness of treatment

These types of questions relate to whether one treatment is better than another in terms of clinical effectiveness (benefit and harm) or cost-effectiveness.

Questions about the frequency of events

Such questions refer to the incidence or prevalence of disease or other clinical phenomena, risk factors, diagnosis, prognosis or prediction of specific clinical outcomes and investigations on the quality of health care.

Was the study design appropriate for the research question?

Studies that answer questions about effectiveness have a well-established hierarchy of study designs based on the degree to which the design protects against bias. Meta-analyses of well-conducted RCTs and individual RCTs provide the most robust evidence followed by nonrandomized controlled trials, cohort studies, case-control studies, and other observational study designs.^{19,20} However, in some circumstances, RCTs are either not feasible or considered ethically inappropriate. These issues are more common in nonpharmaceutical trials, such as those of surgical procedures. One review of gastrointestinal surgical research found that only 40% of research questions could have been

Box 2 Key methodological points to consider in the appraisal of systematic reviews and meta-analyses.

Were all relevant studies included (i.e. was the search comprehensive, did it exclude articles on the basis of publication status or language and was the potential for publication bias assessed)?

Were selected articles appraised and data extracted by two independent reviewers?

Was sufficient detail provided about the primary studies, including descriptions of the patients, interventions and outcomes?

Was the quality of the primary studies assessed?

Did the researchers assess the appropriateness of combining results to calculate a summary measure?

answered by an RCT, even when funding was not an impediment. Patients' preferences, the rarity of some conditions, and the absence of equipoise among surgeons proved to be the major obstacles to performing RCTs of gastrointestinal surgery in this setting.²¹ When an RCT is not feasible, the specific reasons that preclude its use will determine the type of alternate study design that can be used.²¹ Observational studies, rather than RCTs, are the most appropriate study design for research questions on the frequency of events.

Did the study methods address the key potential sources of bias?

In epidemiological terms, the presence of bias does not imply a preconception on the part of the researcher, but rather means that the results of a study have deviated from the truth.³ Bias can be attributed to chance (e.g. a random error) or to the study methods (systematic bias). Random error does not influence the results in any particular direction, but it will affect the precision of the study;²² by contrast, systematic bias has a direction and results in the overestimation or underestimation of the 'truth'. Systematic biases arise from the way in which the study is conducted, be it how study participants were selected, how data was collected, or through the researchers' analysis or interpretation.²³

Different study designs are prone to varying sources of systematic bias. Once the study design of a given article has been identified, we recommend that clinicians use one of the available design-specific critical-appraisal checklists to decide whether the study in question is of high quality. The Critical Appraisal Skills Programme (CASP) includes such tools and the program

coordinators have developed separate checklists for the appraisal of systematic reviews, RCTs, cohort studies, case-control studies, diagnostic test studies, economic evaluations and qualitative research that each comprise 10 questions.⁹ They have been developed from the Users' guides to the medical literature series of articles that were originally published in the *Journal of the American Medical Association*. These articles are now available in book form⁵ and are readily accessible on the internet.⁹

Systematic reviews and meta-analyses

A meticulous, standardized protocol is used in a systematic review to identify, critically appraise and synthesize all the relevant studies on a particular topic. Some systematic reviews may then proceed to a meta-analysis, in which the results from individual studies are combined statistically to produce a single pooled result.³ Although planning to undertake a systematic review or a meta-analysis prospectively is possible,²⁴ the majority of these types of article are retrospective and a risk of bias exists, which arises from the selection of studies and the quality of these primary sources.²⁵ Publication bias, which results from the selective publication of studies with positive findings, is of particular concern, as it distorts overall perceptions of the findings on a particular topic.^{26,27}

The QUORUM (Quality of Reporting of Meta-Analyses) statement provides a comprehensive framework for assessments of the quality of reporting in meta-analyses and systematic reviews.^{25,28} In addition, the AMSTAR²⁹ assessment tool, which comprises 11 questions, has been developed for the appraisal of systematic reviews, and this tool or the CASP checklist⁹ could be more useful than the QUORUM statement for clinicians who wish to undertake a rapid appraisal of these types of articles. Key methodological points to consider in the appraisal of systematic reviews and meta-analyses are listed in Box 2.

Systematic reviews and meta-analyses are not restricted to RCTs alone. The MOOSE (Meta-Analysis Of Observational Studies in Epidemiology) guidelines have been developed as a corollary of the QUORUM statement for meta-analyses of non-RCTs.³⁰

Randomized controlled trials

In an RCT, the random allocation of participants should ensure that treatment groups are equivalent in terms of both known and

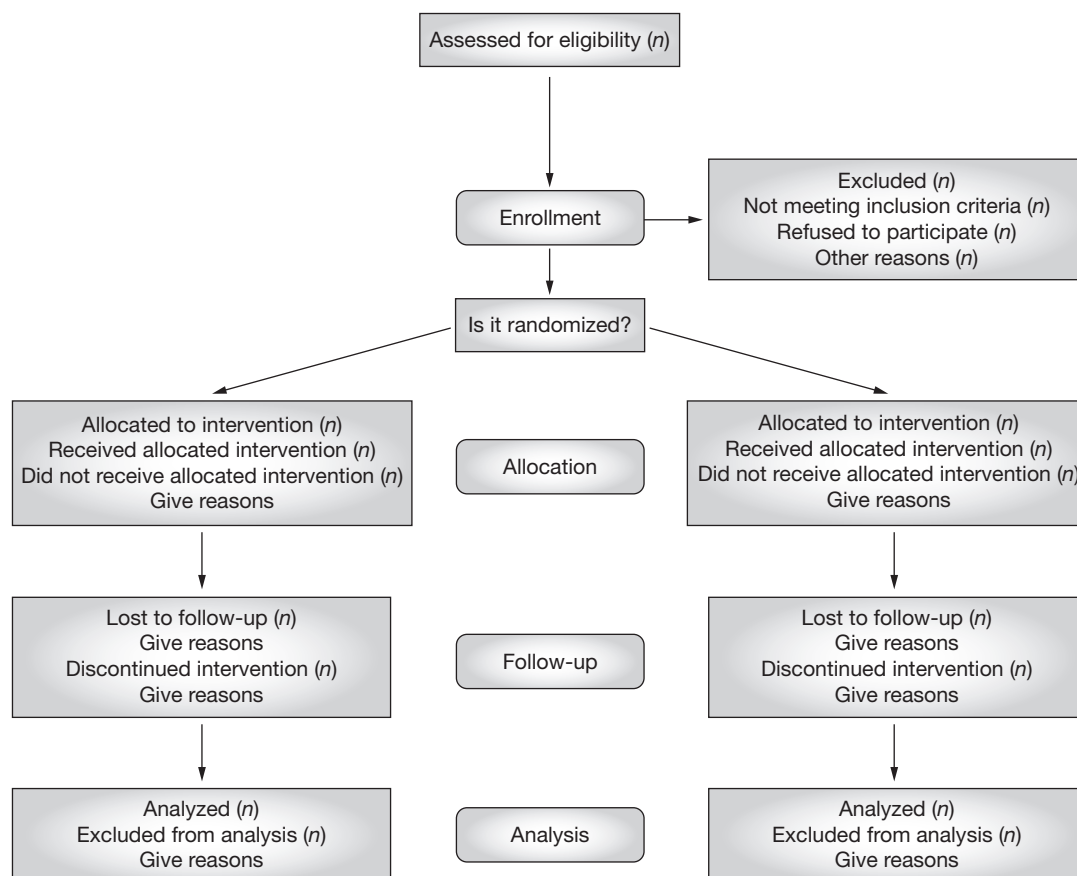


Figure 1 Consolidated standards of reporting trials (CONSORT) statement flowchart for the standard reporting and appraisal of randomized controlled trials. With permission from CONSORT.

unknown confounding factors; any differences in outcomes between groups can, therefore, be ascribed to the effect of treatment.³¹ Study design alone, however, will not guard against bias if crucial aspects of the study protocol are suboptimal. The potential for selective enrollment of patients into the study can be one important source of bias if the group to which individuals will be allocated is known or can be guessed.³² Centralized methods of randomization, for example a computer-generated allocation, are preferable to less concealed methods, such as use of color-coded forms or pseudo-random sequences based on medical record numbers or days of the week.³¹ Failure to conceal the allocation sequence has been shown to result in a greater distortion of the results than lack of double-blinding—another major source of bias in RCTs.³³

The CONSORT (Consolidated Standards of Reporting Trials) statement flow chart (Figure 1)

is functionally equivalent to the QUORUM statement for systematic reviews, and provides a comprehensive tool with which to assess the standard of reporting in randomized trials.³⁴ Key points to consider in the appraisal of an RCT are listed in Box 3.

Cohort studies

Cohort, or longitudinal, studies involve following up two or more groups of patients to observe who develops the outcome of interest. Prospective cohort studies have been likened to natural experiments, as outcomes are measured in large groups of individuals over extended periods of time in the real world.³⁵ Cohort studies can also be performed retrospectively; such studies usually involve identifying a group of patients and following up their progress by examining records that have been collected routinely or for another purpose, such as medical data, death registry records and hospital admission databases.

Box 3 Key methodological points to consider in the appraisal of an randomized controlled trials.

Was the process of treatment allocation truly random?

Would participants have been able to know or guess their treatment allocation?

Were participants and researchers 'blinded' to participants' treatment group?

Were outcomes assessed objectively?

Were all participants who were randomly allocated a treatment accounted for in the final analysis?

Were all participants' data analyzed in the group to which they were randomly allocated?^a

^aSee section on intention-to-treat analysis under 'Were the statistical analyses performed correctly?'

Box 4 Key methodological points to consider in the appraisal of a cohort study.

Is the study prospective or retrospective?

Is the cohort representative of a defined group or population?

Were all important confounding factors identified?

Were all important exposures and/or treatments, potential confounding factors and outcomes measured accurately and objectively in all members of the cohort?

Were there important losses to follow-up?

Were participants followed up for a sufficient length of time?

The major methodological concern with cohort studies is their high potential for selection bias and confounding factors. These problems are particularly relevant when cohort studies (or non-RCTs) are used to evaluate therapeutic interventions. In this situation, the treatment that someone receives is determined by the patient's or clinician's preferences, referral patterns, current treatment paradigms or local policy.³⁶ Important differences are likely to exist between patients who receive disparate treatments and these differences, rather than the treatment itself, might be responsible for the observed outcomes. Although some potential confounding factors can be measured and accounted for in the analysis,³⁷ such adjustments are more difficult in retrospective than prospective studies, as data on important potential confounders might not have been collected, or might be of poor quality.

The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology)

statement is the corollary of the QUORUM and CONSORT statements for observational studies, including cohort, case-control and cross-sectional studies.³⁸ Key methodological features to consider in the appraisal of cohort studies are listed in Box 4.

Case-control studies

Case-control studies are always retrospective by their very nature—the case patients are selected because they have already developed the outcome of interest (e.g. a disease). Data are then collected about factors that might have influenced this outcome, and these exposures are compared with those of a group of people who differ from the case patients only in that they have not developed the outcome of interest. Case-control studies are ideal for the investigation of risk factors when the outcome of interest is rare, as it would take too long to recruit a prospective cohort.

Major methodological difficulties with case-control studies are the selection of appropriate control individuals and the possibility of 'recall bias' (a patient's subjective interpretation of what caused their condition can alter their recall of certain events or experiences). Controls should be drawn from exactly the same population as the cases, and the only difference between controls and cases should be that the controls have not developed the condition of interest. Although objective measures of possible causative factors are preferable, case-control studies often rely on participants' recall, and patients might be more likely to remember certain events or experiences than controls.³⁹ Key aspects to consider when assessing a case-control study are listed in Box 5.

Cross-sectional analyses

Cross-sectional studies provide a 'snapshot' in which all parameters (exposures and outcomes) are assessed at the same time; examples of cross-sectional designs include one-off surveys and audits of practice. Key methodological points to consider in the appraisal of a cross-sectional study are listed in Box 6.

Case series

Case series provide low-level evidence about therapeutic effectiveness; however, these articles are very common in medical literature. Key methodological issues to consider when assessing such articles are listed in Box 7.

Studies that assess the accuracy of diagnostic tests

These studies are usually cross-sectional in design, but possess a number of specific methodological issues that should be considered in addition to those noted above.⁴⁰ To investigate the accuracy of a diagnostic test, it is performed on a sample of patients and the results are compared with those of a reference or gold-standard diagnostic test.⁴¹ The level of agreement between the investigated test and the gold-standard diagnostic test can then be reported either in terms of the sensitivity and specificity, or likelihood ratio.^{4,41}

The STARD (Standards for the Reporting of Diagnostic Accuracy Studies) website provides a detailed flowchart (Figure 2) and 25-item checklist for standardized reporting and appraisal of studies that assess the accuracy of diagnostic tests.^{42,43} The CASP also provides a similar, but more simple, tool for this type of study.⁹ Important features to consider when appraising a study of diagnostic accuracy are listed in Box 8.

Economic evaluations

Economic-evaluation studies focus on cost-efficiency, or which treatment can provide the greatest benefit for the least cost.⁴⁴ Several types of economic-evaluation studies exist, including cost-benefit, cost-effectiveness and cost-utility analyses, all of which differ in how they measure health benefits.⁴⁵ An important feature of critical appraisal of any cost analysis is an assessment of how well the various costs and consequences of individual treatments have been identified and measured. The CASP has developed a checklist to aid with the appraisal of economic evaluation studies.⁹

Was the study performed in line with the original protocol?

Deviations from the planned protocol can affect the validity or relevance of a study. One of the most common problems encountered in clinical research is the failure to recruit the planned number of participants. An estimate suggests that more than a third of RCTs recruit less than 75% of their planned sample.⁴⁶ This deviation from the study plan not only potentially reduces the extent to which the results of the study can be generalized to real-world situations, because those who actually were recruited might be different from those who weren't for some reason, but also reduces the power of the

Box 5 Key methodological points to consider in the appraisal of a case-control study.

- Were the cases clearly defined?
- Were the cases representative of a defined population?
- How were the controls selected and were they drawn from the same population as the cases?
- Were study measures identical for cases and controls?
- Were study measures objective or subjective and is recall bias likely if they were subjective?

Box 6 Key methodological points to consider in the appraisal of a cross-sectional study.

- Was the study sample clearly defined?
- Was a representative sample achieved (e.g. was the response rate sufficiently high)?
- Were all relevant exposures, potential confounding factors and outcomes measured accurately?
- Were patients with a wide range of severity of disease assessed?

Box 7 Key methodological points to consider in the appraisal of a case study.

- Were cases identified prospectively or retrospectively?
- Are the cases a representative sample (e.g. a consecutive series of individuals recruited from multiple centers) and similar to patients in your practice?
- Were all relevant exposures, potential confounding factors and outcomes measured accurately?

study to demonstrate significant findings. Other differences to the original protocol might include changes to the inclusion and exclusion criteria, variation in the provided treatments or interventions, changes to the employed techniques or technologies, and changes to the duration of follow-up.

Does the study test a stated hypothesis?

A hypothesis is a clear statement of what the investigators expect the study to find and is central to any research as it states the research question in a form that can be tested and refuted.³ A null hypothesis states that the findings of a study are no different to those that would have been expected to occur by chance. Statistical hypothesis testing involves calculating

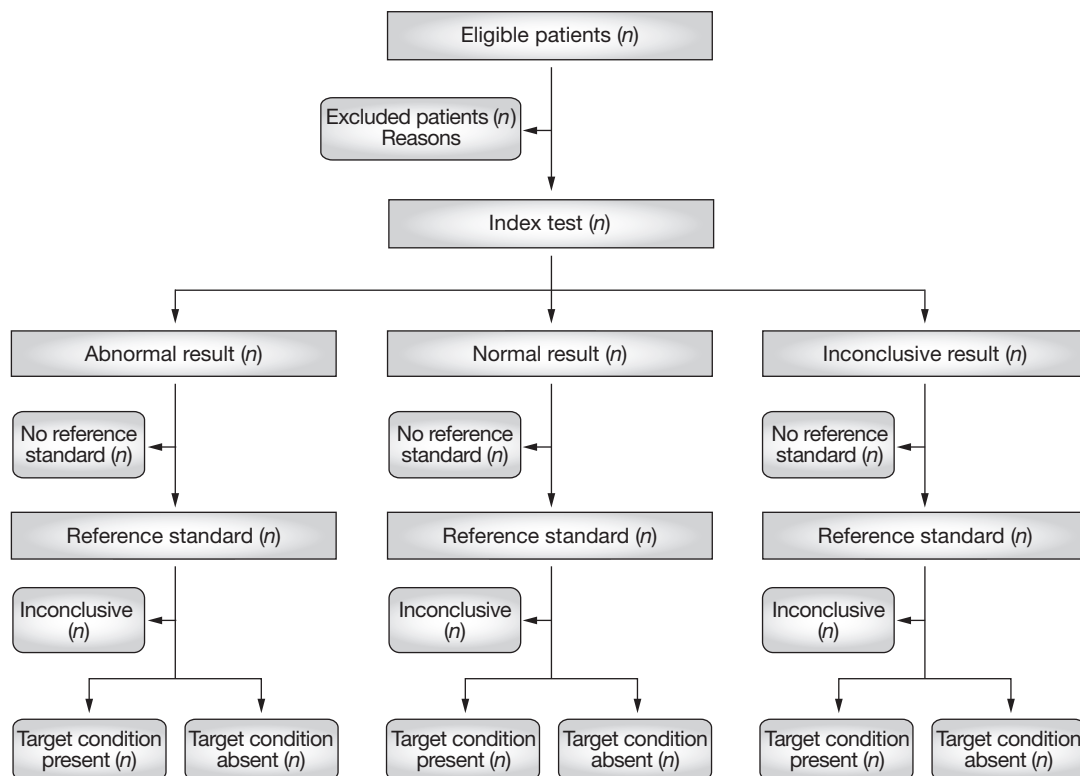


Figure 2 Standards for the reporting of diagnostic accuracy studies (STARD) statement flowchart for the standard reporting and appraisal of studies examining the accuracy of diagnostic tests. With permission from STARD.

the probability of achieving the observed results if the null hypothesis were true. If this probability is low (conventionally less than 1:20 or $P < 0.05$), the null hypothesis is rejected and the findings are said to be 'statistically significant' at that accepted level.

Study hypotheses must crucially be identified *a priori* (that is, before the study is conducted, and are developed from theory or previous experience). If the study investigates the statistical significance of associations that were not prespecified in the original hypothesis (post-hoc analysis), such analyses are prone to false-positive findings because, at a significance level of 5% ($P = 0.05$), 1 in 20 associations tested will be significant (positive) by chance alone. When a large number of such tests are conducted some false-positive results are highly likely to occur. Another important consideration it to check that all data relevant to the stated study objectives have been reported, and that selected outcomes have not been omitted.

Where treatments for a medical condition already exist, trials can be designed to test

whether a new therapy has similar efficacy to an existing one. This type of trial is called an equivalence or noninferiority trial, as its purpose is to establish that the new treatment is no worse than the existing one.⁴⁷ Equivalence studies require that the degree of outcome difference at which the two treatments will not be considered equivalent be determined in advance.⁴⁸ For example, researchers might decide that if the primary outcome for a new treatment is no greater than 5% worse than that of the existing treatment, the two treatments will be considered to be equivalent. Equivalence studies determine whether a new treatment is at least as good as an existing treatment so that decisions about which treatment to administer to a given patient can be made on the basis of criteria, such as cost or ease of administration.^{47,48}

The CONSORT statement for randomized trials has been extended to incorporate guidelines for reporting equivalence studies.⁴⁹ A key question when appraising this type of study is whether the trial results were analyzed appropriately for an equivalence study. If a study

is designed to show that a new treatment is at least as good as an existing treatment, statistical methods, for conventional testing of a hypothesis that one treatment is superior to another should not be used. Appropriate analysis of the results in an equivalence study often involves calculating confidence intervals for the treatment effect, and determining whether these limits are within the predetermined margin of noninferiority.⁴⁸ Another key question is whether the sample size was calculated correctly for an equivalence study, as these types of study usually require a larger sample size than a corresponding superiority trial.⁴⁹

Were the statistical analyses performed correctly?

Assessing the appropriateness of statistical analyses can be difficult for nonstatisticians. However, all quantitative research articles should include a segment within their 'Method' section that explains the tools used in the statistical analysis and the rationale for this approach, which should be written in terms that are appropriate for the journal's readership. In particular, the approach to dealing with missing data and the statistical techniques that have been applied should be specified; patients who are lost in follow-up and missing data should be clearly identified in the 'Results' section. Original data should be presented in such a way that readers can check the statistical accuracy of the paper.

An important consideration in the statistical analysis of RCTs is whether intention-to-treat (ITT) or per-protocol analyses were conducted. According to the ITT principle, participants' data are analyzed with reference to the group to which they were randomly allocated, regardless of whether they actually received the allocated treatment. ITT analyses are preferred, because they maintain the randomization and ensure that the two treatment groups are comparable at baseline.⁵⁰ However, if a lot of participants are nonadherent or a large proportion cross over to other treatments, an ITT analysis will be somewhat conservative and the results might be difficult to interpret. In this situation, a per-protocol analysis that includes only those patients who complied with the trial protocol can be used to supplement the ITT analysis. As per-protocol analyses are at increased risk of selection bias, they should not usually be used as the primary method of analysis unless a compelling reason exists to justify this approach.⁵⁰ The CONSORT

Box 8 Key methodological points to consider in the appraisal of a study of diagnostic accuracy.

Does the sample of patients represent the full spectrum of patients with and without the diagnosis of interest?

Was there a comparison with an appropriate 'gold-standard' test?

Did all patients receive both the test under evaluation and the same 'gold-standard' test?

Were the tests performed independently with blinding of assessors to the results of the 'gold-standard' test?

Were the cut-offs that were used to classify patients as having a positive test result clearly described?

flowchart (Figure 1) enables the flow of participants and the groups used in the analysis of the trial to be clearly identified.³⁴

Do the data justify the conclusions?

The next consideration is whether the conclusions that the authors present are reasonable on the basis of the accumulated data. Sometimes an overemphasis is placed on statistically significant findings that invoke differences that are too small to be of clinical value; alternatively, some researchers might dismiss large and potentially important differences between groups that are not statistically significant, often because sample sizes were small. Other issues to be wary of are whether the authors generalized their findings to broader groups of patients or contexts than was reasonable given their study sample, and whether statistically significant associations have been misinterpreted to imply a cause and effect.

Are there any conflicts of interest?

Conflicts of interest occur when personal factors have the potential to influence professional roles or responsibilities.⁵¹ Members of a research team must make judgments that have the potential to affect the safety of the participants and the validity of the research findings. Researchers are in a position to decide which studies will be conducted in their unit, which patients will be invited to participate in a study and whether certain clinical occurrences should be reported as adverse events.⁵² These decisions require researchers to act with integrity and not for personal or institutional gain.

Potential financial conflicts of interest include the receipt of salary and consultation fees from

the company that has sponsored the research and ownership of stocks and shares or other pecuniary interests, such as patents related to the research.⁵² Units that recruit research participants might be paid a per-capita fee for every patient enrolled, which can be greater than the expenses involved.⁵³ Many potential financial sources of conflicts of interest, such as industry funding for educational events, travel or gifts, are increasingly recognized both within the context of daily clinical practice and research.⁵⁴ However, other potential conflicts are inherent to the research setting. An example is that medical researchers' status and future research income is dependent on the success of their research.⁵⁵

Identification of a potential conflict of interest is not synonymous with having an actual conflict of interest or poor research practice. Potential conflicts of interest are extremely common, and the most important questions are whether they have been recognized and how they have been dealt with.⁵⁶ A main mechanism for dealing with potential conflicts of interest is open disclosure.⁵⁶ In the process of critically appraising a research article, one important step is to check for a declaration about the source of funding for the study and, if a potential conflict of interest had been identified for a statement about how this conflict was managed. For example, the researchers might state specifically that the sponsoring agency had no input into the research protocol, data analysis or interpretation of the findings. Many journals now routinely require authors to declare any potential financial or other conflicts of interest when an article is submitted. The reader must then decide whether the declared factors are important and might have influenced the validity of the study's findings.

CONCLUSIONS

Critical appraisal is a systematic process through which the strengths and weaknesses of a research study can be identified. This process enables the reader to assess the study's usefulness and whether its findings are trustworthy. The most important component of critical appraisal is careful assessment of the study design; however, other steps, such as evaluation of the statistical methods used, interpretation of the findings and potential conflicts of interest are also essential. Finally, consideration of the importance of the research to one's own patients will help clinicians identify the most relevant, high-quality studies available to guide their clinical practice.

KEY POINTS

- Critical appraisal is a systematic process used to identify the strengths and weaknesses of a research article
- Critical appraisal provides a basis for decisions on whether to use the results of a study in clinical practice
- Different study designs are prone to various sources of systematic bias
- Design-specific, critical-appraisal checklists are useful tools to help assess study quality
- Assessments of other factors, including the importance of the research question, the appropriateness of statistical analysis, the legitimacy of conclusions and potential conflicts of interest are an important part of the critical appraisal process

References

- 1 Druss BG and Marcus SC (2005) Growth and decentralisation of the medical literature: implications for evidence-based medicine. *J Med Libr Assoc* **93**: 499–501
- 2 Glasziou PP (2008) Information overload: what's behind it, what's beyond it? *Med J Aust* **189**: 84–85
- 3 Last JE (Ed.; 2001) *A Dictionary of Epidemiology (4th Edn)*. New York: Oxford University Press
- 4 Sackett DL et al. (2000) *Evidence-based Medicine. How to Practice and Teach EBM*. London: Churchill Livingstone
- 5 Guyatt G and Rennie D (Eds; 2002) *Users' Guides to the Medical Literature: a Manual for Evidence-based Clinical Practice*. Chicago: American Medical Association
- 6 Greenhalgh T (2000) *How to Read a Paper: the Basics of Evidence-based Medicine*. London: Blackwell Medicine Books
- 7 MacAuley D (1994) READER: an acronym to aid critical reading by general practitioners. *Br J Gen Pract* **44**: 83–85
- 8 Hill A and Spittlehouse C (2001) What is critical appraisal? *Evidence-based Medicine* **3**: 1–8 [www.evidence-based-medicine.co.uk] (accessed 25 November 2008)
- 9 Public Health Resource Unit (2008) *Critical Appraisal Skills Programme (CASP)*. [http://www.phru.nhs.uk/Pages/PHD/CASP.htm] (accessed 8 August 2008)
- 10 National Health and Medical Research Council (2000) *How to Review the Evidence: Systematic Identification and Review of the Scientific Literature*. Canberra: NHMRC
- 11 Elwood JM (1998) *Critical Appraisal of Epidemiological Studies and Clinical Trials (2nd Edn)*. Oxford: Oxford University Press
- 12 Agency for Healthcare Research and Quality (2002) Systems to rate the strength of scientific evidence. *Evidence Report/Technology Assessment No 47, Publication No 02-E019* Rockville: Agency for Healthcare Research and Quality
- 13 Crombie IK (1996) *The Pocket Guide to Critical Appraisal: a Handbook for Health Care Professionals*. London: Blackwell Medicine Publishing Group
- 14 Heller RF et al. (2008) Critical appraisal for public health: a new checklist. *Public Health* **122**: 92–98
- 15 MacAuley D et al. (1998) Randomised controlled trial of the READER method of critical appraisal in general practice. *BMJ* **316**: 1134–37

- 16 Parkes J *et al.* Teaching critical appraisal skills in health care settings (Review). *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: cd001270. doi:10.1002/14651858.cd001270
- 17 Mays N and Pope C (2000) Assessing quality in qualitative research. *BMJ* **320**: 50–52
- 18 Hawking SW (2003) *On the Shoulders of Giants: the Great Works of Physics and Astronomy*. Philadelphia, PN: Penguin
- 19 National Health and Medical Research Council (1999) *A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines*. Canberra: National Health and Medical Research Council
- 20 US Preventive Services Taskforce (1996) *Guide to clinical preventive services (2nd Edn)*. Baltimore, MD: Williams & Wilkins
- 21 Solomon MJ and McLeod RS (1995) Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery* **118**: 456–467
- 22 Rothman KJ (2002) *Epidemiology: an Introduction*. Oxford: Oxford University Press
- 23 Young JM and Solomon MJ (2003) Improving the evidence-base in surgery: sources of bias in surgical studies. *ANZ J Surg* **73**: 504–506
- 24 Margitic SE *et al.* (1995) Lessons learned from a prospective meta-analysis. *J Am Geriatr Soc* **43**: 435–439
- 25 Shea B *et al.* (2001) Assessing the quality of reports of systematic reviews: the QUORUM statement compared to other tools. In *Systematic Reviews in Health Care: Meta-analysis in Context 2nd Edition*, 122–139 (Eds Egger M *et al.*) London: BMJ Books
- 26 Easterbrook PH *et al.* (1991) Publication bias in clinical research. *Lancet* **337**: 867–872
- 27 Begg CB and Berlin JA (1989) Publication bias and dissemination of clinical research. *J Natl Cancer Inst* **81**: 107–115
- 28 Moher D *et al.* (2000) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. *Br J Surg* **87**: 1448–1454
- 29 Shea BJ *et al.* (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology* **7**: 10 [doi:10.1186/1471-2288-7-10]
- 30 Stroup DF *et al.* (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* **283**: 2008–2012
- 31 Young JM and Solomon MJ (2003) Improving the evidence-base in surgery: evaluating surgical effectiveness. *ANZ J Surg* **73**: 507–510
- 32 Schulz KF (1995) Subverting randomization in controlled trials. *JAMA* **274**: 1456–1458
- 33 Schulz KF *et al.* (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **273**: 408–412
- 34 Moher D *et al.* (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology* **1**: 2 [www.biomedcentral.com/1471-2288/1/2] (accessed 25 November 2008)
- 35 Rochon PA *et al.* (2005) Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ* **330**: 895–897
- 36 Mamdani M *et al.* (2005) Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. *BMJ* **330**: 960–962
- 37 Normand S *et al.* (2005) Reader's guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ* **330**: 1021–1023
- 38 von Elm E *et al.* (2007) Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* **335**: 806–808
- 39 Sutton-Tyrrell K (1991) Assessing bias in case-control studies: proper selection of cases and controls. *Stroke* **22**: 938–942
- 40 Knottnerus J (2003) Assessment of the accuracy of diagnostic tests: the cross-sectional study. *J Clin Epidemiol* **56**: 1118–1128
- 41 Furukawa TA and Guyatt GH (2006) Sources of bias in diagnostic accuracy studies and the diagnostic process. *CMAJ* **174**: 481–482
- 42 Bossuyt PM *et al.* (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* **138**: W1–W12
- 43 STARD statement (Standards for the Reporting of Diagnostic Accuracy Studies). [http://www.stard-statement.org/] (accessed 10 September 2008)
- 44 Raftery J (1998) Economic evaluation: an introduction. *BMJ* **316**: 1013–1014
- 45 Palmer S *et al.* (1999) Economics notes: types of economic evaluation. *BMJ* **318**: 1349
- 46 Russ S *et al.* (1999) Barriers to participation in randomized controlled trials: a systematic review. *J Clin Epidemiol* **52**: 1143–1156
- 47 Timmouth JM *et al.* (2004) Are claims of equivalency in digestive diseases trials supported by the evidence? *Gastroenterology* **126**: 1700–1710
- 48 Kaul S and Diamond GA (2006) Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* **145**: 62–69
- 49 Piaggio G *et al.* (2006) Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* **295**: 1152–1160
- 50 Heritier SR *et al.* (2007) Inclusion of patients in clinical trial analysis: the intention to treat principle. In *Interpreting and Reporting Clinical Trials: a Guide to the CONSORT Statement and the Principles of Randomized Controlled Trials*, 92–98 (Eds Keech A *et al.*) Strawberry Hills, NSW: Australian Medical Publishing Company
- 51 National Health and Medical Research Council (2007) *National Statement on Ethical Conduct in Human Research 89–90* Canberra: NHMRC
- 52 Lo B *et al.* (2000) Conflict-of-interest policies for investigators in clinical trials. *N Engl J Med* **343**: 1616–1620
- 53 Kim SYH *et al.* (2004) Potential research participants' views regarding researcher and institutional financial conflicts of interests. *J Med Ethics* **30**: 73–79
- 54 Komesaroff PA and Kerridge IH (2002) Ethical issues concerning the relationships between medical practitioners and the pharmaceutical industry. *Med J Aust* **176**: 118–121
- 55 Little M (1999) Research, ethics and conflicts of interest. *J Med Ethics* **25**: 259–262
- 56 Lemmens T and Singer PA (1998) Bioethics for clinicians: 17. Conflict of interest in research, education and patient care. *CMAJ* **159**: 960–965

Competing interests

The authors declared no competing interests.