



Alison Hill BSC
FFPHM FRCP Director,
and **Claire
Spittlehouse** BSc
Business Manager,
Critical Appraisal
Skills Programme,
Institute of Health
Sciences, Oxford

What is critical appraisal?

Sponsored by an educational grant from AVENTIS Pharma

- **Critical appraisal** is the process of **systematically** examining research evidence to assess its **validity, results** and **relevance** before using it to inform a decision.
- Critical appraisal is an essential part of **evidence-based clinical practice** that includes the process of systematically **finding, appraising** and **acting** on evidence of effectiveness.
- Critical appraisal allows us to make sense of research evidence and thus begins to close the gap between research and practice.
- **Randomised controlled trials** can minimise bias and use the most appropriate design for studying the effectiveness of a specific intervention or treatment.
- **Systematic reviews** are particularly useful because they usually contain an explicit statement of the objectives, materials and methods, and should be conducted according to explicit and reproducible methodology.
- Randomised controlled trials and systematic reviews are not automatically of good quality and should be appraised critically.

What is critical appraisal?

Critical appraisal is one step in the process of evidence-based clinical practice. Evidence-based clinical practice is 'an approach to decision making in which the clinician uses the best evidence available, in consultation with the patient, to decide the option which suits the patient best'.¹ To determine what is the 'best' evidence, we need critical appraisal skills that will help us to understand the methods and results of research and to assess the quality of the research. Most research is not perfect, and critical appraisal is not an exact science – it will not give us the 'right' answer. But it can help us to decide whether we think a reported piece of research is good enough to be used in decision making.

There are many factors that come into play when making healthcare decisions – research evidence is just one of them. If research has flaws, it is up to readers to use their critical appraisal skills to decide whether this affects the usefulness of the paper in influencing their decision.

Pros of critical appraisal in practice

- Critical appraisal provides a systematic way of assessing the validity, results and usefulness of published research papers.
- Together with skills in finding research evidence and changing practice as a result of research, critical appraisal is the route to closing the gap between research and practice¹ and as such makes an essential contribution to improving healthcare quality.
- Critical appraisal encourages objective assessment of the usefulness of information – critical appraisal skills are applied to published research, but all evidence should be appraised to weigh up its usefulness.
- Critical appraisal skills are not difficult to develop. Critical appraisal is a common sense approach to reading, and user-friendly tools are available to help anyone develop these skills.

Cons of critical appraisal in practice

- Critical appraisal can be time-consuming initially, although with time it becomes the automatic way to look at research papers.
- Critical appraisal does not always provide the reader with the 'easy' answer or the answer one might have hoped for; it may highlight that a favoured intervention is in fact ineffective.
- Critical appraisal can be dispiriting if it highlights a lack of good evidence – it may take determination to persist with an area of interest when access to good research in the area is limited.

Appraising randomised controlled trials

Box 1³ (opposite) provides a checklist of questions for critically appraising *randomised controlled trials (RCTs)*. The RCT is the most appropriate research design for studying the effectiveness of a specific intervention or treatment.² In an RCT, participants are randomly assigned to two (or more) groups: one (or more) experimental group(s) receiving the intervention that is being tested, and a comparison or control group receiving a placebo or an alternative treatment. The two (or more) groups are then followed up to see what differences result. Randomisation ensures that the groups differ only in the intervention given, so any difference between the outcomes in each group can be attributed to the intervention.

RCTs' methodology *can* minimise accidental or intentional bias, but this does not automatically mean that every RCT is of good quality. We must *critically appraise* individual studies to assess the validity of their methods. Once we are happy that the methods were sound, then we can look at what the results tell us and consider whether we can apply them to our own population.

This method of critically appraising an RCT can be applied to the paper about the Heart Outcomes Prevention Evaluation

Box 1. 12 questions to help you make sense of a trial. Adapted from Guyatt *et al*³

A. Are the results of the study valid?

Screening questions

1. *Did the trial address a clearly focused research question?*

Tip: a research question should be 'focused' in terms of:

- The population studied
- The intervention given
- The outcomes considered.

2. *Did the authors use the right type of study?*

Tip: the right type of study would:

- Address the research question
- Have an appropriate study design.

Is it worth continuing?

Detailed questions

3. *Was the assignment of patients to treatments randomised?*

Tip: consider if this was done appropriately.

4. *Were all of the patients who entered the trial properly accounted for at its conclusion?*

Tip: look for:

- The completion of follow-up
- Whether patients were analysed in the groups to which they were randomised.

5. *Were patients, health workers and study personnel 'blind' to treatment?*

Tip: this is not always possible, but consider if it was possible – was every effort made to ensure 'blinding'?

6. *Were the groups similar at the start of the study?*

Tip: think about other factors that might effect the outcome such as age, sex, social class.

7. *Aside from the experimental intervention, were the groups treated equally?*

Tip: for example, were they reviewed at the same time intervals.

B. What are the results?

8. *How large was the treatment effect?*

9. *How precise was the estimate of the treatment effect?*

Tip: look for the confidence limits.

C. Will the results help locally?

10. Can the results be applied to the local population?

Tip: consider whether the patients covered by the trial are likely to be very different from your population.

11. Were all clinically important outcomes considered?

12. Are the benefits worth the harms and costs?

© Critical Appraisal Skills Programme

(HOPE) study,⁴ which investigates the effects of ramipril on cardiovascular events in high-risk patients. A brief critical appraisal of the methods of the HOPE paper (addressed by questions 1–7) is included to demonstrate how critical appraisal tools can be used.

How should these questions be used?

Questions one and two are screening questions, and any research paper should give the reader the answers to these two questions on the first page, ideally in the abstract. If it is not possible to answer 'yes' to both these

questions, we would normally not continue to read the paper as it is unlikely to be helpful. All questions in sections A and C should be answered with either 'yes', 'no' or 'can't tell', which should be justified with evidence from the text; section B requires text answers.

Section A: are the results of the study valid?

The first seven questions address the methods used in the trial. If the methods are robust, we can then look at the results of the trial and decide if the results can be useful for us. If the research methods are flawed, then this may invalidate the trial.

Screening questions:

1. *Did the trial address a clearly focused research question?*

It is important to be clear about the focus of the question – does the research set out to answer a question that will be useful to you and can help to inform your decision?

Unfocused research cannot provide you with reliable answers. If it is not focused it is not worth spending time appraising it.

2. *Did the authors use the right type of study?*

Having determined the research question in question 1 above, you now need to assess if the paper uses an appropriate type of research method that is likely to answer that question. If the research question is 'does intervention x produce the desired measurable outcome in a specific population?', then the best type of single research study to address this is an RCT. Alternatively, if the research question were 'what are the views, beliefs and perceptions of a specific population?', then the best type of research study to address this would be a qualitative study.

For example, the HOPE paper⁴ asks the question: *does ramipril [the intervention] prevent myocardial infarction, stroke or death from cardiovascular causes [the outcomes] in patients who were at high risk for cardiovascular events but who did not have left ventricular dysfunction or heart disease [the population]?* This question is best addressed by an RCT.

Is it worth continuing?

If we have established that the paper has a clear research question that can be addressed

by an RCT, we need to continue through this checklist of questions and consider the quality of the methods used.

Detailed questions:

3. *Was the assignment of patients to treatments randomised?*

The randomisation method should be robust to reduce the possibility of bias within the allocation of participants to treatment and control groups. The most robust types of randomisation are by computer-generated numbers or tables of random numbers: the participants have a 50:50 chance of being in either group, and both the participants and the clinicians are completely unaware before randomisation which group the participants will be in. Quasi randomisation is less robust and often involves randomising by alternate participants, day of the week or alternate months.

In the HOPE paper we are told that the study participants were randomly assigned to groups, but we are not given any details of the randomisation method, which would have given us more confidence in the methods. A previous paper is referenced for details of the study design.⁵ (This previous paper tells us that patients were randomised internationally by a telephone call to a central office and assigned to groups in a 2x2 factorial design.)

4. *Were all of the patients who entered the trial properly accounted for at its conclusion?*

It is important to account for the participants in a trial who were **lost to follow up** – that is, those who withdrew from the trial before its conclusion. Those who are lost are not necessarily representative of the whole group, so their loss may distort the results. It is also important to consider how the participants were analysed – was **intention-to-treat analysis** used? That is, analysing the participants in the groups to which they were randomised – this is important because if participants move between groups, the random generation of groups will have been affected and, again, this may distort the results.

At the end of the HOPE study, 99.9% of randomised patients were accounted for – this is a very good level of follow-up. We must also bear in mind that 32% of patients had discontinued treatment at some stage during

the trial and 28% permanently discontinued. There was also a run-in phase before the RCT began, which resulted in the exclusion of 10% of patients before randomisation due to non-compliance, side-effects, abnormal serum creatinine or potassium levels, or withdrawal of consent. The implications of this would need to be addressed in **section C**, which addresses the generalisability or external validity of the study.

5. Were patients, health workers and study personnel 'blind' to treatment?

If the people involved in the study are **blind** (that is, they are not aware of who is in the treatment or control groups), this reduces the possibility of bias. To blind patients and health workers, the intervention and placebo must appear to be identical, which is not always possible. However, the study personnel can always be blinded; they do not need to be aware of whether the patients they assess or data that they are working with is from the treatment or control group.

In the HOPE study, we could assume that the patients were blinded as we are told that they were given the drug or a 'matching placebo'. It would have been straightforward to have blinded the health workers administering the drug and the study personnel who analysed the data as well, but this is not discussed in this paper. (In the previous paper⁵ it is stated that 'emergency unblinding is available centrally and locally but will only be done when absolutely necessary and after a check list is completed by telephone call to the project office', which implies that there was blinding.)

6. Were the groups similar at the start of the study?

If the sample is large enough and the randomisation method robust, then the make-up of the treatment and control groups should be very similar and the only difference should be the intervention or control or alternative treatment. This is important if we are to be reasonably sure that the outcome of the trial is due to the intervention and not any other factors.

In the HOPE study, the randomisation process appears to have been successful, as the baseline characteristics of the patients in the two groups are very similar. This is clearly displayed in a table.

7. Aside from the experimental intervention, were the groups treated equally?

Again, it is important to be confident that the groups were treated equally, so that we can attribute the outcome of the trial to the intervention. It is possible for groups to be treated differently if the health workers or study personnel are not blinded.

In the HOPE study, we know that the groups were followed up over the same time period and we have no reason to believe that the groups were *not* treated equally. If we assume that everyone involved in the trial was blinded, then the groups must have been treated equally.

Section B: what are the results?

Questions 8 and 9 address the results presented in the paper.

8. How large was the treatment effect?

The outcome measures used differ between papers. Often a measure of **relative risk** is used. The chances of a particular outcome being observed in an individual is called the risk (risk can refer to something good or bad). By comparing the risk in the experimental and control groups, a measure of relative risk can be calculated.⁶ A relative risk of 1 occurs when the incidences are the same in the two groups.⁷ If we hope that the intervention would lead to more of the outcome measured (for example, reduction of symptoms) then we want a relative risk of more than 1; if we hope that the intervention would lead to less of the outcome measured (for example, myocardial infarction, stroke or death from cardiovascular causes, which were the primary outcomes considered in the HOPE trial) then we want a relative risk of less than 1. Results may also be presented as NNTs (numbers needed to treat).⁸

9. How precise was the estimate of the treatment effect?

There will always be some doubt about the result (or best estimate), as a trial only looks at a sample of the population. The **confidence interval (CI)** indicates the range of doubt around the best estimate.⁶ Results are often also presented with **p values**. These describe the probability that a particular result has happened by chance.⁶ If the p value is less

than 0.05, then this is usually described as being **statistically significant** and means that the results are unlikely to be due to chance.

Section C: will the results help locally?

Any weaknesses in the methods or in the presentation of the results should be borne in mind when considering how useful the results are locally.

10. *Can the results be applied to the local population?*

It is important to consider if there are any differences between the participants in the trial and the local population that would make it impossible to apply the results locally. Participants dropping out of the study must be considered as their loss may distort the results. The patients lost from the HOPE study are discussed in question 4.

11. *Were all clinically important outcomes considered?*

A single trial cannot address all the important outcomes that we are interested in, but consider if the paper has answered the original research question and whether any other important outcomes have been highlighted or missed out.

12. *Are the benefits worth the harms and costs?*

Financial information is not normally given in a trial, but we need to consider what the negative effects could be and whether these are outweighed by the benefits. Other research, such as an economic evaluation, might help with the cost implications.

Healthcare decisions are not usually made purely on the basis of one trial. There are many other factors that influence decisions, and there may be many trials that together can provide more conclusive evidence. Systematic reviews can provide valuable information.

Appraising systematic reviews

Box 2⁹ (opposite) provides a checklist for appraising systematic reviews. Reviews collect together primary research and summarise their results and conclusions. Systematic reviews are particularly useful because they usually contain an explicit statement of objectives, materials and methods, and

should have been conducted according to explicit and reproducible methodology. But, as with RCTs, systematic reviews should be critically appraised by users so they can decide for themselves whether their methods are valid, assess what the results are saying and decide whether these results can be applied locally. Further information on systematic reviews can be found in 'What is a systematic review?'.¹⁰

Appraising other types of studies

In addition to RCTs and systematic reviews, it is increasingly being recognised that other types of studies contribute to evidence-based decision-making. These include cohort and case-control studies, economic evaluations, studies on diagnostic tests and qualitative studies. Checklists for these types of studies are available from the Critical Appraisal Skills Programme (CASP).^{2,6,11} The JAMA user guides and the Centre for Evidence-based Medicine are other important sources of appraisal checklists.^{12,13}

References

1. Gray JAM. *Evidence-based healthcare: how to make health policy and management decisions*. Edinburgh: Churchill Livingstone, 1997.
2. Critical Appraisal Skills Programme and HCLU. *Evidence-based health care: an open learning resource for healthcare professionals*. Oxford: Update Software, 1999.
3. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II: how to use an article about therapy or prevention. *JAMA* 1993; **270**: 2598–2601 and **271**: 59–63.
4. The Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med* 2000; **342**: 145–153.
5. The HOPE Study Investigators. The HOPE (Heart Outcomes Prevention Evaluation) study: the design of a large, simple randomized trial of an angiotensin-converting enzyme inhibitor (ramipril) and vitamin E in patients at high-risk of cardiovascular events. *Can J Cardiol* 1996; **12**(2): 127–137.
6. Critical Appraisal Skills Programme and HCLU. *Evidence-based health care. A computer aided learning resource*. Oxford: Update Software, 1999.
7. Kirkwood BR. *Essentials of medical statistics*. Oxford: Blackwell Science, 1998.
8. Moore A, McQuay HJ. *What is an NNT?* London: Hayward Medical Communications, 1997.
9. Oxman AD *et al*. Users' guides to the medical literature. VI: how to use an overview. *JAMA* 1994; **272**(17): 1367–1371.
10. Davies HT, Crombie IK. *What is a systematic review?* London: Hayward Medical Communications, 1998.
11. CASP. <http://www.casp.org.uk>
12. Centre for Evidence-based Medicine. <http://cebmr2.ox.ac.uk/>
13. Canadian Centers for Health Evidence. <http://www.cche.net/>

Box 2. Ten questions to help you make sense of a systematic review. Adapted from Oxman *et al*⁹

Three broad issues need to be considered when appraising research:

- A Are the results of the study valid?
- B What are the results?
- C Will the results help locally?

The questions below are designed to help you think about these issues systematically

- The first two questions are screening questions and can be answered quickly. If the answer to both is 'yes', it is worth continuing.
- There is a fair degree of overlap between several of the questions.
- You are asked to record a 'yes', 'no' or 'can't tell' to most of the questions.
- A number of tips are given after each question. These are designed to remind you why the question is important.

A. Are the results of the review valid?

Screening questions

1. *Did the review address a clearly focused research question?*

Tip: a research question should be 'focused' in terms of:

- The population studied
- The intervention given or exposure
- The outcomes considered.

2. *Did the review include the right type of studies?*

Tip: these would:

- Address the review's research question
- Have an appropriate study design.

- The results were similar from study to study (look for tests of heterogeneity)

- The reasons for any variations in results are discussed.

Tip: think about other factors that might effect the outcome such as age, sex, social class.

B. What are the results?

6. *What are the main results of the review?*

Tip: consider:

- How the results are expressed (for example, odds ratio, relative risk and so on)
- What the results are.

7. *Could the results be due to chance?*

Tip: look for tests of statistical significance (p values) and confidence intervals (CIs).

C. Will the results help locally?

8. *Can the results be applied to the local population?*

Tip: consider whether:

- The population sample covered by the review could be sufficiently different from your population to cause concern
- Your local setting is likely to differ much from that of the review.

9. *Were all important outcomes considered?*

Tip: consider outcomes from the point of view of the:

- Individual
- Policy makers and practitioners
- Family/carers
- Wider community.

10. *Should policy of practice change as a result of the evidence contained in this review?*

Tip: consider whether the benefits are worth the harms and costs.

Is it worth continuing?

Detailed questions

3. *Did the reviewers try to identify all relevant studies?*

Tip: look for:

- Which bibliographic databases were used
- Follow-up from reference lists
- Personal contact with experts
- Search for unpublished studies
- Search for non-English language studies.

4. *Did the reviewers assess the quality of the included studies?*

Tip: a clear predetermined strategy should be used to determine which studies are included. Look for:

- A scoring system
- More than one assessor.

5. *If the results of the studies have been combined, was it reasonable to do so?*

Tip: consider whether:

- The results of each study are clearly displayed

Tritace prescribing information

Presentation: Capsules containing 1.25mg, 2.5mg, 5mg or 10mg ramipril. **Indications:** Reducing the risk of myocardial infarction, stroke, cardiovascular death or need for revascularisation procedures in patients of 55 years or more who have clinical evidence of cardiovascular disease (previous MI, unstable angina or multivessel CABG or multivessel PTCA), stroke or peripheral vascular disease. Reducing the risk of myocardial infarction, stroke, cardiovascular death or need for revascularisation procedures in diabetic patients of 55 years or more who have one or more of the following clinical findings: hypertension (systolic blood pressure >160mmHg or diastolic blood pressure >90mmHg); high total cholesterol (>5.2mmol/l); low HDL (<0.9mmol/l); current smoker; known microalbuminuria; clinical evidence of previous vascular disease. Mild to moderate hypertension. Congestive heart failure as adjunctive therapy to diuretics with or without cardiac glycosides. Reduction in mortality in patients surviving acute MI with clinical evidence of heart failure. **Dosage and administration:** Reduction in risk of MI, stroke, cardiovascular death or need for revascularisation procedure: The initial dose is 2.5mg Tritace o.d.. Depending on tolerability, the dose should be gradually increased. It is recommended that this dose is doubled after about 1 week of treatment then, after a further 3 weeks, increased to 10mg. The usual maintenance dose is 10mg Tritace o.d.. Patients stabilised on lower doses of Tritace for other indications where possible should be titrated to 10mg Tritace o.d.. Hypertension: Initial dose 1.25mg titrated up to a maximum of 10mg o.d. according to response. Usual dose 2.5mg or 5mg o.d.. Stop diuretic therapy 2 - 3 days before starting Tritace and resume later if required. Congestive heart failure: Initial dose 1.25mg o.d. titrated up to a maximum of 10mg daily according to response. Doses above 2.5mg daily can be given o.d. or b.d.. Post Myocardial Infarction: Initiate treatment between day 3 and day 10 following AMI. Initially 2.5mg b.d. increasing to 5mg b.d. after 2 days. Assessment of renal function is recommended prior to initiation. Reduced maintenance dose may be required in impaired renal function. Monitor patients with impaired liver function. In the elderly the dose should be titrated according to need. Not recommended for children. **Contraindications:** Hypersensitivity to ramipril or excipients, history of angioneurotic oedema, haemodynamically relevant renal artery stenosis, hypotensive or haemodynamically unstable patients, pregnancy, lactation. **Precautions:** Do not use in aortic or mitral valve stenosis or outflow obstruction. Assess renal function before and during use, as there is a risk of impairment of renal function. Use with caution during surgery or anaesthesia. Hyperkalaemia. Do not use in patients using polyacrylonitrile (AN69) dialysis membranes or during low-density lipoprotein apheresis with dextran sulphate. Agranulocytosis and bone marrow depression seen rarely with ACE inhibitors as well as a reduction in red cell count, haemoglobin and platelet content. Symptomatic hypotension may occur after initial dose or increase in dose, especially in salt/volume depleted patients. **Drug interactions:** Combination with diuretics, NSAIDs, adrenergic blocking drugs or other antihypertensive agents may potentiate antihypertensive effect. Risk of hyperkalaemia when used with agents increasing serum potassium. May enhance the effect of antidiabetic agents. May increase serum lithium concentrations. **Side effects:** Dizziness, headache, weakness, disturbed balance, nervousness, restlessness, tremor, sleep disorders, confusion, loss of appetite, depressed mood, anxiety, paraesthesiae, taste changes, muscle cramps & joint pain, erectile impotence, reduced sexual desire, fatigue, cough, hypersensitivity reactions; pruritus, rash, shortness of breath, fever, cutaneous and mucosal reactions, Raynauds phenomenon, gastrointestinal disturbances, jaundice, hepatitis, impaired renal function, angioneurotic oedema, pancreatitis, eosinophilia and vasculitis. Symptomatic hypotension, myocardial infarction or cerebrovascular accident possibly secondary to severe hypotension in high-risk patients, chest pain, palpitations, rhythm disturbances, angina pectoris may occur. Use with caution and closely monitor patients with impaired liver function. Reduced serum sodium levels, elevated blood urea nitrogen and serum creatinine. Pre-existing proteinuria may deteriorate.

Basic NHS cost: 28 x 1.25mg capsules £5.30; 28 x 2.5mg capsules £7.51; 28 x 5mg capsules £9.55; 28 x 10mg capsules £13.00

Marketing authorisation numbers: 1.25mg 13402/0021, 2.5mg 13402/0022, 5mg 13402/0023, 10mg 13402/0024

Legal category: POM

Marketing authorisation holder: Aventis Pharma, 50 Kings Hill Avenue, Kings Hill, West Malling, Kent, ME19 4AH

Date of preparation: July 2000

This publication, along with the others in the series, is available on the internet at www.evidence-based-medicine.co.uk



Sponsored by an educational grant from
AVENTIS Pharma

Published by Hayward Medical Communications, a division of Hayward Group plc.

Copyright © 2001 Hayward Group plc.

All rights reserved.