GUIDE:
Critically Appraising Evidence

Contents

Purpose of Resource

The purpose of this resource is to provide additional information for learners to support their understanding of critically appraising evidence and research and to provide a step-by-step guide to apply these principles when undertaking research for the planning, development and implementation of population health programs.

Units of Competency

This resource supports learning and completion of assessments for the following units of competency:
- HLTPOP502C Work with the community to identify health needs
- HLTPOP503C Plan a population health project

Acknowledgement

This tip sheet has been adapted from Evidence-Based Answers to Clinical Questions for Busy Clinicians (2009), The Centre for Clinical Effectiveness, Southern Health, Melbourne, Australia.

What is Critical Appraisal?

Critical Appraisal is the process you use to determine if the research you have identified is accurate, reliable and relevant. In your determination you need to consider if:

- it is a reliable piece of research, and can therefore use its contents
- it is not a reliable piece of research and therefore should not use the results
- the research is OK but has limitations therefore should use the results with discretion

There are three basic aspects to appraising research
1. Is it worth looking at the results of this study?
2. What are the results?
3. Are the results relevant for my purpose?

Appraising Research

It is difficult to design a generic appraisal process that addresses all the potential issues in different study designs, however the following information should help to assess the validity of the study or article you are interested in.

When reviewing the results of a paper, it is important to ask yourself the following questions:
1. **Why was the study done?**
2. **What was the research question?**
3. **What type of study design was used? Was this design the most appropriate for the research question posed?** (see table* below)

If your question is about...	Look for a...
Intervention or Therapy	➢ Randomised Controlled Trial
Diagnosis/Screening To assess the accuracy of the test:	 ➢ Cohort study where all subjects receive **both** the study test and gold standard reference test
To assess effect of test on health outcomes:	➢ Randomised Controlled Trial
Prognosis	➢ Longitudinal cohort
Aetiology/Risk factors	➢ Randomised controlled trial ➢ Cohort for rare exposure with common outcome ➢ Case-control for rare outcome with common exposure

*Please refer to **Appendix A** for a summary of the approaches to determining levels of evidence and an explanation of the trial types listed in the table above.

4. **Using the PICO (Population, Intervention, Comparison, Outcomes) template below, review the research you are appraising and answer the following questions about the study characteristics?**

P – Patient or Population	**What us the patient or population characteristics of the research/study?** (i.e. age, gender etc.)	
I – Interventions or Indicator	**What type of intervention is it?** (e.g. therapy, diagnostic test or exposure)	
C – Comparison	**Does the study include a comparison component and can the study be compared to other interventions?**	
O - Outcomes	**What were the outcomes of the study?**	

5. **Are these characteristics compatible with my question?**

☐Yes   ☐Maybe   ☐No - Stop reading as this research won't assist.

6. **Are the results valid?**

This question aims to determine whether the study you have found was carried out in an appropriate way and whether the study design has minimised the opportunity for bias to affect the results.

Bias in health research is systematic error in the design, conduct or analysis of a study that means the results of the study are distorted away from the truth. Bias may

produce either underestimation or overestimation of the effect of an intervention or exposure, or the extent of a relationship.

There are two main types of bias including:
- Selection bias – the impact of how subjects are selected or allocated to the study, or groups within the study
- Information bias – the impact of inaccurate or incomplete measurement of the data about the subjects, their exposure or the effects of the intervention

Minimising opportunity for bias is the aim of good research design.

To answer question six you need to consider the following:
- What weaknesses (opportunities for bias) exist in this study?
- What effect would this have on outcomes?

7. Has the study been carried out in a sufficiently careful way so that bias is minimised and we can be relatively confident that the results are close to the truth?

☐Yes ☐Maybe ☐No - Stop reading as this research won't assist.

8. What are the results?

The following will assist in answering this question:
- Are the outcome measures used relevant and comprehensive?
- What is the size of the effect? (i.e is there a clinical significance – is this an important effect for patients?)
- What is the precision of the effect? (i.e. statistical significance – is it likely that this effect is not just due to chance?)

9. Are the results relevant to my clinical or research situation?

For example, is there a similar patient population, similar definitions used, similar protocols followed or within a similar health system?

Additional Reading and Links on Critical Appraisal

The Centre for Clinical Effectiveness - www.mihsr.monash.org/cce
Clinical Evidence - www.clinicalevidence.com
Centre for Evidence Based Medicine - www.cebm.net
Netting the evidence - www.shef.ac.uk/scharr/ir/netting/

Appendix A: Levels of Evidence Approaches and Trial Types Explained

There are several examples of levels of evidence. The summary below provides three explanations of levels of evidence.

1. A summary of how evidence can be graded.

In simple terms, one way of looking at levels of evidence is as follows (the higher the level, the better the quality; the lower, the greater the bias):

I. Strong evidence from at least one systematic review of multiple well-designed randomised controlled trials.
II. Strong evidence from at least one properly designed randomised controlled trial of appropriate size.
III. Evidence from well-designed trials such as pseudo-randomised or non-randomised trials, cohort studies, time series or matched case-controlled studies.
IV. Evidence from well-designed non-experimental studies from more than one centre or research group or from case reports.
V. Opinions of respected authorities, based on clinical evidence, descriptive studies or reports of expert committees.

or...

Category I: Evidence from at least one properly randomized controlled trial.

Category II-1: Evidence from well-designed controlled trials without randomization.

Category II-2: Evidence from well-designed cohort or case-control analytic studies, preferably from more than one center or research group.

Category II-3: Evidence from multiple times series with or without intervention or dramatic results in uncontrolled experiments such as the results of the introduction of penicillin treatment in the 1940s.

Category III: Opinions of respected authorities, based on clinical experience, descriptive studies and case reports, or reports of expert committees.

Source: *Harris, R.P. et al. (2001). Current methods of the U.S. Preventive Services Task Force: a review of the process. American Journal of Preventive Medicine. April 20 (3 Supplement): 21-35.*

2. A Diagrammatic representation of the levels

The EVIDENCE PYRAMID is used to illustrate the development of evidence. At the base of the pyramid is Background Evidence and Expert Opinion - this is where ideas are first developed. As you progress up the pyramid the amount of information available decreases in volume, but increases in relevance and quality of evidence.
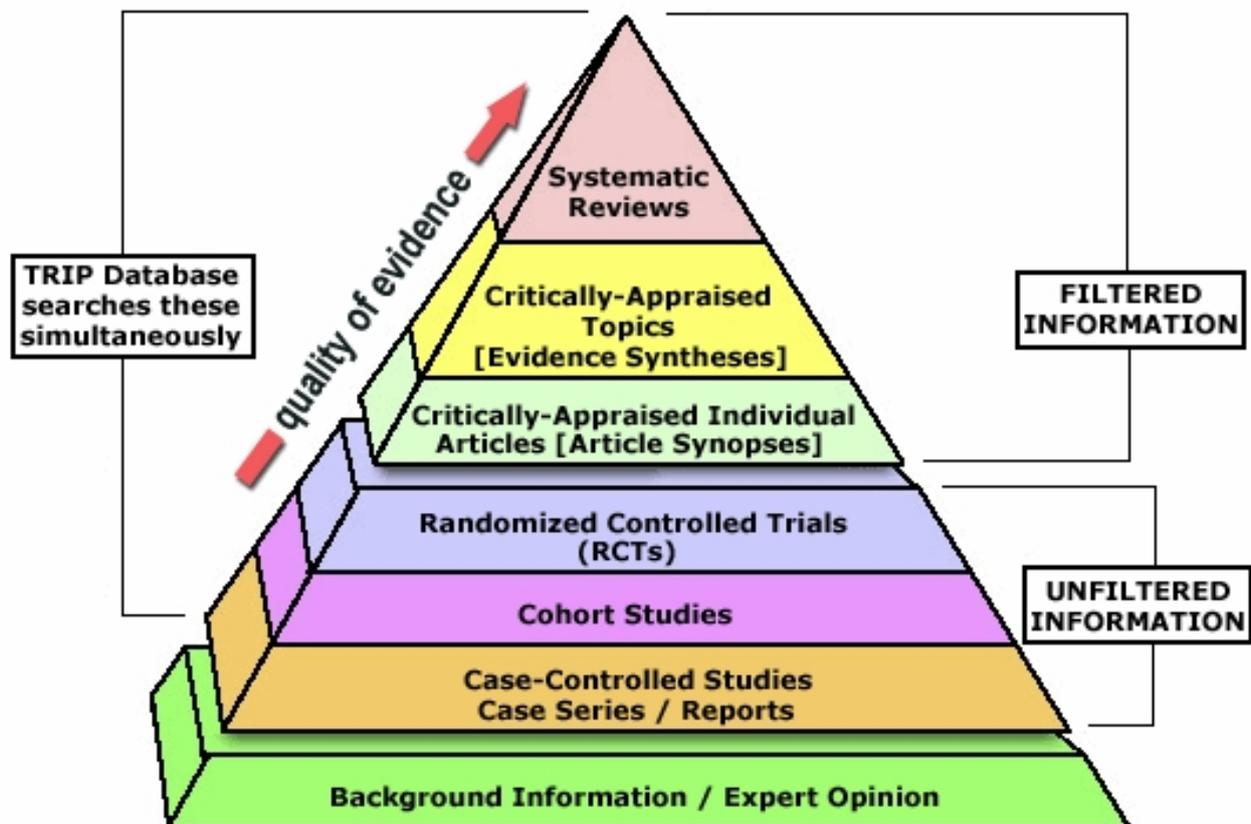
Source: *Evidence-Based Clinical Practice Resources:*
http://guides.library.yale.edu/content.php?pid=9786&sid=73113

3. National Health and Medical Research Council Levels of Evidence

Level	Intervention 1	Diagnostic accuracy 2	Prognosis	Aetiology 3	Screening Intervention
I 4	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies	A systematic review of level II studies
II	A randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,5among consecutive persons with a defined clinical presentation	A prospective cohort study	A prospective cohort study	A randomised controlled trial
III-1	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard,5among non-consecutive persons with a defined clinical presentation	All or none	All or none	A pseudorandomised controlled trial (i.e. alternate allocation or some other method)

III-2	A comparative study with concurrent controls: • Non-randomised, experimental trial • Cohort study • Case-control study • Interrupted time series with a control group	A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	Analysis of prognostic factors amongst persons in a single arm of a randomised controlled trial	A retrospective cohort study	A comparative study with concurrent controls: • Non-randomised, experimental trial • Cohort study • Case-control study
III-3	A comparative study without concurrent controls: • Historical control study • Two or more single arm study • Interrupted time series without a parallel control group	Diagnostic case-control study	A retrospective cohort study	A case-control study	A comparative study without concurrent controls: • Historical control study • Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard)	Case series, or cohort study of persons at different stages of disease	A cross-sectional study or case series	Case series

4. Oxford (CEBM) levels of evidence

Level	Therapy / Prevention, Aetiology / Harm	Prognosis	Diagnosis	Differential diagnosis / symptom prevalence study	Economic and decision analyses
1a	SR (Systematic Review), with homogeneity, of RCTs (Randomised Control Trials)	SR (with homogeneity) of inception cohort studies; CDR (Clinical Decision Rule) validated in different populations	SR (with homogeneity) of Level 1 diagnostic studies; CDR with 1b studies from different clinical centres	SR (with homogeneity) of prospective cohort studies	SR (with homogeneity) of Level 1 economic studies
1b	Individual RCT (with narrow Confidence Interval)	Individual inception cohort study with > 80% follow-up; CDR validated in a single population	Validating cohort study with good reference standards; or CDR tested within one clinical centre	Prospective cohort study with good follow-up	Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses
1c	All or none	All or none case-series	Absolute SpPins and SnNouts (types of diagnostic findings)	All or none case-series	Absolute better-value or worse-value analyses

2a	SR (with homogeneity) of cohort studies	SR (with homogeneity) of either retrospective cohort studies or untreated control groups in RCTs	SR (with homogeneity) of Level >2 diagnostic studies	SR (with homogeneity) of 2b and better studies	SR (with homogeneity) of Level >2 economic studies
2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR or validated on split-sample only	Exploratory cohort study with good reference standards; CDR after derivation, or validated only on split-sample or databases	Retrospective cohort study, or poor follow-up	Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses
2c	"Outcomes" Research; Ecological studies	"Outcomes" Research		Ecological studies	Audit or outcomes research
3a	SR (with homogeneity) of case-control studies		SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies	SR (with homogeneity) of 3b and better studies

3b	Individual Case-Control Study		Non-consecutive study; or without consistently applied reference standards	Non-consecutive cohort study, or very limited population	Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.
4	Case-series (and poor quality cohort and case-control studies)	Case-series (and poor quality prognostic cohort studies)	Case-control study, poor or non-independent reference standard	Case-series or superseded reference standards	Analysis with no sensitivity analysis
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory or "first principles"

Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998. Updated by Jeremy Howick March 2009.

Source: Oxford Centre for Evidence-Based Medicine: http://www.cebm.net/levels_of_evidence.asp