Workshop 2 — General approach to critical appraisal of a medical journal paper

A critical appraisal or review is a complete and systematic evaluation of a paper, and consists of the following tasks:

* Determination of the objectives and rationale of the study
* Detailed and critical examination of the methods
* Assessment of the data and their method of presentation
* Evaluation of the conclusions and relevance to the stated objectives
* Consideration of possible errors, and how they could have been avoided

Important questions

Several important questions have to be answered during the process of critical appraisal:

(1) Are the results of clinical and/or public health importance?

(2) Are the results basically valid?

(3) Are the results reasonably reliable?

(4) Can the results be applied in another setting, especially that of relevance to me?

Are the results of clinical and/or public health importance?

The answer may vary according to the reader's background, including specialty and practice setting. The question is usually answered by reviewing the **Introduction** and **Discussion**, but good judgement would usually require up-to-date knowledge in the specific area—a geriatrician may not be able to judge the importance of a paper on an unusual paediatric condition. Peer reviewers have a very important role to play here, especially for a general medical journal covering a wide range of medical specialties. They can inform the editors whether publishing a particular submission is worthwhile and/or likely to generate interest among readers.

Are the results basically valid?

As mentioned in the introduction to this series,[1] before we attempt to apply results that may look very interesting and important, we need to ensure that the results are basically correct or valid. Results being valid mean that they are likely to be free from biases or systematic errors and therefore expected to reflect the truth. Validity is assessed by reviewing the methods, as well as how the results are derived and presented. This is the most important part of the critical appraisal process and is discussed in more detail below.

Are the results reasonably reliable?

In other words, will the results be similar (repeatable) if the study is carried out with another representative sample of subjects from the same potential pool? Could similar results have occurred purely by chance? This is of relevance to sample size or statistical power of the study, and the answer can be obtained by reviewing the confidence intervals (CIs) and/or the P values. An apparently big effect of a certain intervention could have occurred by chance alone, if the estimated true effect could vary between no or even harmful effect to marvellous beneficial effect (indicating a wide CI covering the point of no effect). Results with high precision (narrow CI) that minimise the uncertainties are more likely to be of value in making clinical and/or public health decisions.

Can the results be applied in another setting, especially that of relevance to me?

This question is only relevant after we have satisfied ourselves that the results are basically valid and reasonably reliable, and that they have clinical or public health importance. The answer to this question is context-specific, requiring some understanding of the setting where the study was conducted (time, place, referral or primary care setting, etc) and the characteristics of the study subjects (persons—age, gender, ethnicity, co-morbidity, etc), as well as a defined practice setting (by time, place, and person) in which the reported valid results are to be applied. If a good match is not apparent, one should then consider whether the internally valid results of the study could have external validity. Results from a high-quality randomised controlled trial (RCT) conducted in a tertiary referral hospital may not be equally applicable in a community health setting, as the subjects involved in the RCT are usually less heterogeneous (due to strict inclusion and exclusion criteria), and the interventions/treatments are closely supervised, ensuring good compliance. It has sometimes been argued that if a study has been conducted in a Caucasian or African population, the results may not be applicable to Chinese subjects. In the practice of evidence-based medicine, we adopt the best *available* evidence approach, as the 'best' or perfect evidence may not be forthcoming in the foreseeable future for many clinical or public health decisions that we have to make.

On validity of results and sources of bias

Validity of the results is assessed by examining whether they are free from bias and reflect the truth. A handy way to examine bias is to group the sources under three categories: selection bias, information

TABLE. Sources of bias and sub-grouping

Sources	Sub-grouping	Points to note in appraisal

Selection bias	• Selection by the investigator(s)
	  • Inclusion and exclusion criteria not clear or not followed
	  • Inappropriate sampling methods
	• Selection by the study subjects
	  • Non-response, rejection, or refusal
	  • Loss to follow-up
		• Review methods to see whether inclusion and exclusion criteria were stated clearly
		• Determine whether all eligible subjects were included or an appropriate method of sampling was used to select a representative sample
		• Check whether the participation rate, response rate, follow-up rate, etc, were reported as well as reasonably high; if not, check whether the investigators had compared the groups and discussed possible bias and limitations

Information/ measurement bias	• Assessment of exposure(s) • Assessment of outcome(s) • Assessment of confounder(s)	• Look for standard protocols for data collection, and be aware of unequal efforts in ascertaining exposures and/or outcomes in different groups (eg more frequent follow-up or more investigations in a certain group) • Check if data collection was blinded • Routine documentations available in past (eg staff rosters, medical records) are generally regarded as less subjected to bias • Objective measurements (eg laboratory results directly from machines without human interpretation) are less subject to bias • Death is the most objective outcome, but the causes of death may not be entirely objective
Confounding (bias)*	• From literature and/or knowledge in specific area • From current data	• All major known factors associated with the outcome should be mentioned and taken care of in the data analysis • Reported associations in the current study should also be taken care of in examining other associations

* Some epidemiologists do not agree with the term 'confounding bias', but the presence of confounding, if not controlled for, would distort or bias the results of associations. Hence, it is included as a source of bias here

or measurement bias, and bias from confounding (Table).

Selection bias

Selection bias can result from the acts of the investigator(s) or the study subjects. Bias resulting from selection of non-representative study subjects by the investigator(s) is regarded as an unforgivable error in the conduct of good medical research. Clear inclusion and exclusion criteria—in terms of time, place, person and other specific characteristics—are required to define subject eligibility. If not all eligible subjects are included in the study, the selected subjects should be representative of those not selected. The best way to achieve a representative sample is through random sampling, with each eligible subject having an equal and known chance of being selected. One should be aware that in some studies, study subjects were picked by the investigators by convenience (convenient sampling), but reported as being randomly sampled.

Selection bias resulting from the acts of the study subjects is also referred to as self-selection bias. In a free society, eligible subjects can self-select to participate or not in any study or stay on with the follow-up assessments or not. If those participating or staying on for the follow-up are systematically different from those who do not, they may not be representative of all eligible subjects, thus resulting in a bias.

Information or measurements

Information or measurements may not be accurate and can bias the results, especially when human responses or assessments (subjectivity) are involved, eg questionnaire, grading of severity. Information in most medical studies can be categorised into three groups: exposures or interventions, outcomes, and confounding. Exposure assessment is usually more subject to bias when collected retrospectively, eg case-referent studies and historical cohort studies, whereas outcome ascertainment is more subject to bias if performed prospectively, eg cohort, prognosis and intervention studies. One must also consider information on confounding factors, as inaccurately collected information used to adjust the association between exposure/intervention and outcome may also result in bias.

Confounding

Confounding can distort or bias the result of an association between an exposure/intervention and an outcome. As confounders are independently associated with the outcome, a good starting point would be to examine whether the prevalence of major known risk factors (from literature and/ or knowledge in the specific area) are similar in the groups being compared. In a sense, this tests the effectiveness of randomisation in RCTs, and if different, one must consider whether they have been taken into account adequately, eg by adjustment in the course of statistical analysis.

Ignatius TS Yu
Shelly LA Tse
Clinical Epidemiology Group
Hong Kong Medical Journal

Reference

1. Yu IT, Tse SL. Clinical Epidemiology Workshop—Introduction. Hong Kong Med J 2011;17:315-6.