Critical Appraisal on Journal of Clinical Trials

Murdani Abdullah, M. Adi Firmansyah

Department of Internal Medicine, Faculty of Medicine, University of Indonesia - Cipto Mangunkusumo Hospital. Jl. Diponegoro no. 71, Jakarta Pusat 10430, Indonesia. Correspondence mail: murdani.abdullah@ui.ac.id; murdani@hotmail.com.

ABSTRAK

Critical appraisal merupakan bagian dari kedokteran berbasis bukti (evidence-based medicine) diartikan sebagai suatu proses evaluasi secara cermat dan sistematis suatu artikel penelitian untuk menentukan reabilitas, validitas, dan kegunaannya dalam praktik klinis. Komponen utama yang dinilai dalam critical appraisal adalah validity, importancy, dan applicability. Tingkat kepercayaan hasil suatu penelitian sangat bergantung dari disain penelitian dimana uji klinis menempati urutan tertinggi. Telaah kritis meliputi semua komponen dari suatu penelitian dimulai dari komponen pendahuluan, metodologi, hasil dan diskusi. Masing-masing komponen memiliki kepentingan yang sama besarnya dalam menentukan apakah hasil penelitian tersebut layak atau tidak digunakan sebagai referensi. Kemaknaan secara statistik yang didapat hendaknya juga dibandingkan dengan kemaknaan secara klinis.

Kata kunci: *critical appraisal, telaah kritis, evidence-based medicine, uji klinis, clinical trial.*

ABSTRACT

Critical appraisal is an element of evidence-based medicine, which is defined as the process of carefully and systematically examining a research article to determine its reliability, validity and value in clinical practice. The major components evaluated in critical appraisal are validity, importancy, and applicability. The level of reliability of study results depend on the design of study, in which clinical trial has the highest rank. Critical appraisal includes all components of the research starting from the introduction, method, results and discussion. Each component has similar value to establish whether the results can realistically be applied as a referrence. The results of statistical significance should also been compared with its clinical significance.

Key words: *critical appraisal, evidence-based medicine, clinical trial.*

INTRODUCTION

Nowadays, there has been increasingly prevalent medical practice of evidence-based medicine (EBM) or evidence based practice. Principally, it is an approach trying to use current best scientific evidences obtained from scientific research in making clinical decision about treatment.

Certainly, not all information of scientific journal can become reliable knowledge with confirmed truth and subsequently suitable for being established into guidelines of clinical practice. As clinicians, we must be able to appraise the information, particularly about the clarity, accuracy, precision, reliability, relevance and other supporting evidences including argumentation for making conclusion, the depth, breadth, and considering fairness.

The rate of reliability of study results of a research extremely depends on study design, which is known as the level of evidence. The hierarchy of study design which has high evidence to low evidence are as follows: [1] randomized clinical trials (RCT), systematic reviews on RCT; [2] cohort study; [3] case-controlled study; [4] case report or case series;

and [5] expert opinions.[1]

Method of appraising information should be done sistematically to provide a good conclusion, which is the best kind of information. It includes appraisal of various sources of information as well as appraising the conclusion by providing supporting evidences. One of the methods is critical appraisal. By performing critical appraisal, we are expected to be able to evaluate results, validity and value of scientific article publications systematically. This manuscript is going to discuss about critical appraisal and focus on the appraisal of clinical trial articles in scientific journals.

CRITICAL APPRAISAL

Definition

By definition, critical appraisal is regarded as a process of evaluating a research article carefully and systematically to determine the reliability, validity and application in clinical practice.[2] In other words, through critical appraisal, we decide a research article is reliable or not. The ability to perform critical appraisal should be one of basic competence of a clinician in order to recognize and use reliable study data efficiently.

The next question is how do we know that the obtained data has good quality and reliable? How could we decide which study is more reliable when we have data of two different conclusion on the same topic? For this reason, therefore, we should performm critical appraisal.

Critical appraisal is an important element of EBM. As we have known that there are five steps in EBM when evaluating a clinical case, i.e.: 1) formulating questions with the concept of PICO (Population, Intervention, Comparation, Outcome); 2) search for the evidence or reference; 3) appraise the evidence on the component of validity and importancy; 4) decide what action to take from the findings including applying the eveidences for clinical practice; and 5) evaluate the practice performed according to the evidence-based medicine (**Figure 1**).[2,3]

Formulating question by using PICO concept can also be the element of critical appraisal itself. The following is the example of PICO concept: "In a 50-year-old female patient who is diagnosed with type 2 diabetes mellitus (Patient), how great is the role of three-month diet and exercise program (Intervention) to help normalize the fasting blood glucose (Outcome) compared to diet restriction alone (Comparator)?"

Starting Critical Appraisal of Clinical Trial

Before we are performing critical appraisal of a research article in a scientific journal, we should first comprehend the article. A good comprehension of an article will make our evaluation better. In general, the basic components of a research article are introduction, methods, results and discussion (**Figure 2**).

Usually, when we take a look at a research article, the abstract is the first part that we read. Most of us believe that the answer to our questions is there in a concise form, while it would be too risky to consider an abstract as a true summary of a research. Use it only to decide about the topic that we are interested in and to avoid further reading if by reading it we can already see a poor method of study.

In fact, we do not have to read the article word by word from beginning to end. In critical appraisal, we would evaluate the article systematically. The list of questions that we should find the answer when reading a research article are shown in **Table 1**.

Overall, components of critical appraisal for a research article are validity, importance and applicability, which is more familiar as the abbreviation of VIA as shown in **Figure 3** as the following.

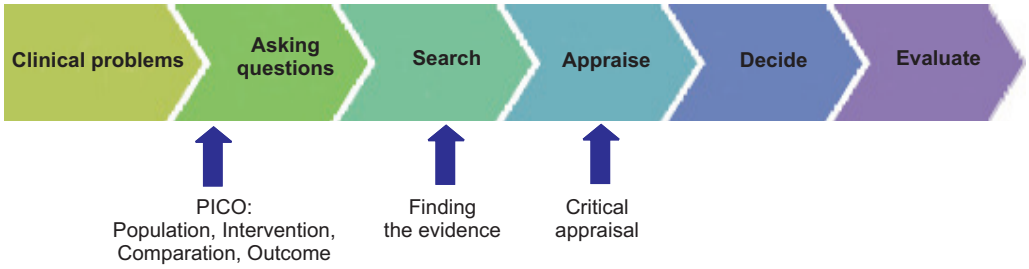

Figure 1. Steps of evidence-based medicine and the role of critical appraisal

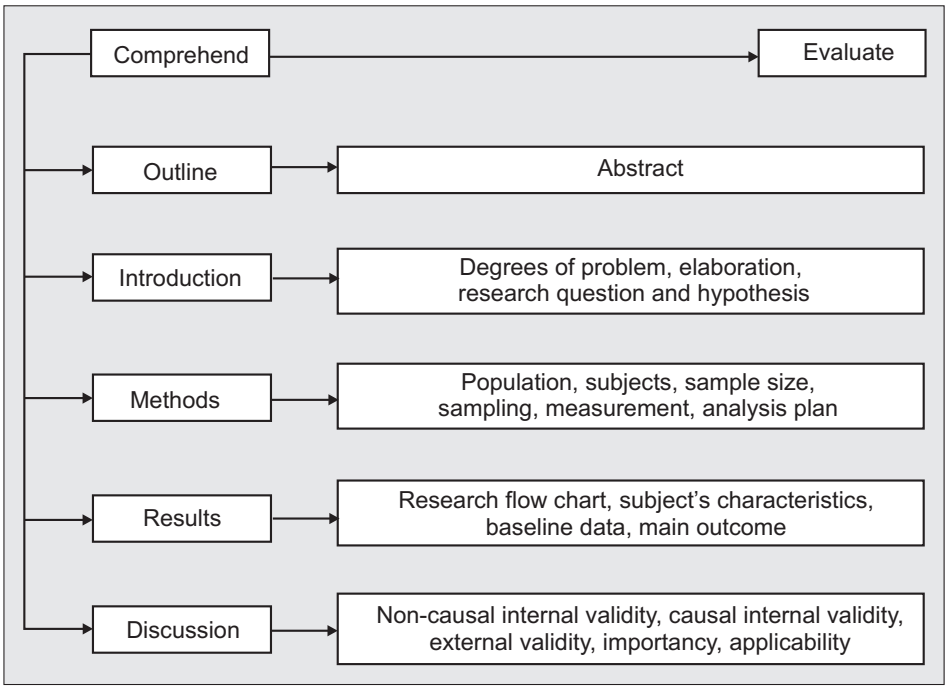

Figure 2. Components of a research article that should be comprehended before doing critical appraisal

Table 1. Critical appraisal checklist[5]

I. Study validity

Research question

- Is the research question well-defined that can be answered using this study design?

Randomization

- Were the patients randomized to the intervention and control groups by a well-defined method of randomization?
- Was the randomization list concealed from patients, clinicians and researchers?
- Do the patients in each group have similar characteristics at the beginning of the study?

Blinding

- Were the patients and clinicians kept blinded (masked) to which treatment was being given?
- Were they kept blinded until the end of the study?

Follow-up

- Were all patients counted at the end of the study?
- If not, how many patients were lost to follow up and for what reason?
- Were the patients analysed in the group they originally were randomized to?

Interventions and co-interventions

- Were the performed interventions described in sufficient detail to be followed by others?
- Other than intervention, were th two groups cared for in similar way of treatment?

II. Results

Selection of outcomes

- Does the article report all relevant outcomes including side effect?

Effect size

- Was there a difference between the outcomes of the treatments, and how big was the difference?
- How reliable is the results: what are the confidence intervals?

III. Applicability

Using results in your own setting

- Are your patient so different from those studied that the results may not apply to them?
- Is your environment so different from the one in the study that the methods could not be use there?

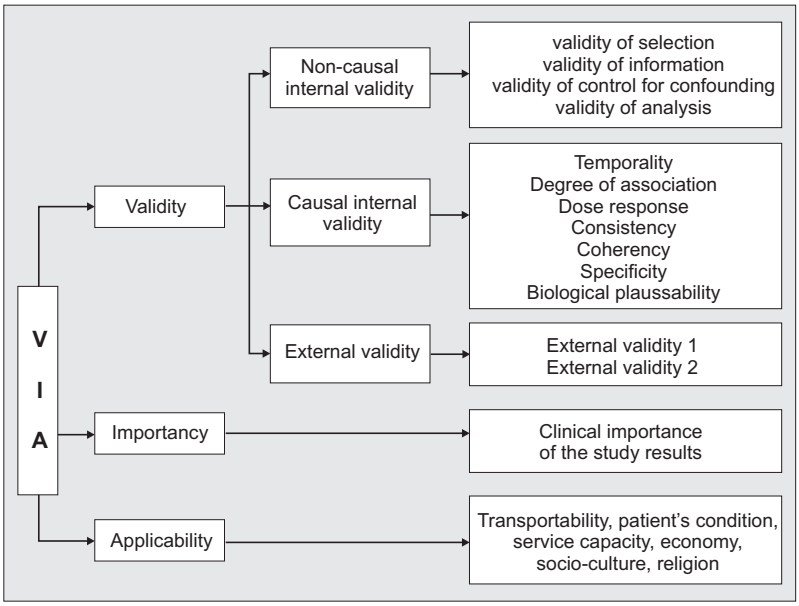

Figure 3. Components of critical appraisal[4,5]

The component of validity is aimed to proof whether the result is a fact or not; while the component of importancy and applicability are directed to evaluate the value and feasibility.

The Introduction

In the introduction component of a research article, there are usually two or three paragraph in brief statement. In this section, the researcher tries to explain about the reason of why he/she conducted the study. Generally, when we read it carefully, there are a pattern of introduction component in a research article. The pattern includes the formulation of data to determine the degrees of problem (for instance the data of prevalence, incidence, number of cases, the magnitude of effect in certain cases), elaboration of data that has been known before, formulation of current gap (problems) and the aim or hypothesis of study. The list of component associated with the introduction is shown on the following **Table 2**.

Table 2. List of questions for the introduction of clinical journal article[4]

Component	Explanation
What is the data that has been presented to show the magnitude of problems?	Prevalence, incidence, impact.
What data that has been known before (elaborated)?	Substantial and methodological elaboration
Is there still any gap?	Substantial and methodological gap and confirmation of previous study results.
What are the major aims (primary endpoint)?	The most important objective of the study. It provides the basis of sample size calculation and the study is usually designed to answer the major aims.
What are the minor aims (secondary endpoint)?	It does not serve as the basis of sample size calculation. It may be elaborative and provides baseline data for further studies.
What is the study hypothesis?	It consists of non-inferiority trial (or negative clinical trial to provide evidences that the outcomes of each intervention are equal) and inferiority trial (positive clinical trial, i.e. to proof that the outcomes of intervention are different

The Methods

The method of a research article usually provides information about the study population, sample size, inclusion and exclusion criteria, as well as randomization method (since our manuscript discusses about clinical trial, we will focus on randomization method), blinding information and the analysis plan used in the study. In brief, this part can generally divided into four sub-components, i.e. population and subjects; study design; sample size; analysis plan. **Table 3** shows a list of question that relevant to the methods section.[4,6,7]

Table 3. List of questions for the methods of clinical journal article[4]

Components	Explanation
Population and subjects	- Who are the target and accessible population in the study?(target population refers to the population that has become the objective of generalizing the study, which has clinical and demographical components; while accessible population is part of target population wh ich has clinical, demographical, temporal and spatial components)
	- What are the inclusion and exclusion criteria?
Study design	- What are the methods used for randomization? (for example, simple, block, or stratified randimisation etc).
	- What tools that has been used for randomization? (table or computer)?
	- Who perform the randomization? Is there any concealment? What kind of concealment that has been performed?
	- Is there any blinding and how the blinding was performed?
Sample size	- What is the formula of sample size that (should be) used? (Is there any comparison between unpaired and paired proportion, or between unpaired and paired mean value?).
	- What are the rationales for determining sample size? (for example, does the study determine the degree of type I and II error and the expected power?)
Analysis plan	- What kind of statistical tests that have been used? Are they consistent with the hypothesis and the aims of the study?
	- What are the methods used for analysis? (analysis per protocol or intention to treat?)
	- What computer program used for statistical analysis?
	- Is there any interim analysis that has been planned for the study?

We should recognize some basic terms that are strongly associated with the methods section, which are randomization, blinding and concealment. Randomization refers to random allocation, i.e. a process of randomization to determine to which group a subject will be allocated. Randomization is useful to eliminate counfounding factors. Blinding is a procedure to mask information about the treatment given; while concealment is about how to conceal the randomization table.[4,6]

Analysis plan usually will provide information about what kind of statistical tests that will be used to analyse data and how the level of significance is used. Usually, the statistical analysis being used is highly associated with hypothesis of the study; for example, Chi-square test is used when the study aims for comparing the proportion of unpaired group. Moreover, we should also find out about what kind of analysis method that has been used in the study, either it is an analysis per protocol or intention to treat (ITT) analysis. When the analysed subjects were those who had participated and completed the study, then the analysis per protocol would be appropriate. In contrast, ITT analysis is used for those who had been randomized although they had not complete the study.[4,6,8] Another analysis is the interim analysis, i.e. any assessment of data performed before the number of all study subjects has been completed. The analysis is usually used when there is a suspicion that one of treatment is more superior than others and if the analysis has proven the suspicion, the study should be stopped.[4]

The Results

The Result section of a clinical trial research article usually consists of the flow of the study and the study data including primary and secondary outcomes. **Table 4** shows a list of overall questions relevant to the results section.[4,7]

The flow of the study could be occassionally found in the methods section. It is common to present the flow of the study in diagram or flow chart to provide easy reading about the process involved in the study. Through the flow, we could obtain some information about how many subjects have been recruited, how many subjects excluded from the study and how many eligible subjects. Other information that could be found is about how large is the percentage of eligible subjects and what is the percentage of subjects that were lost-to follow up.

Baseline data is a series of initial data prior to the intervention. Our task here is to evaluate whether the groups are comparable in regards of demographical, socio-economic, and history of illness. If the baseline data are not comparable, there are potentially confounding variables. Therefore, researchers usually use stratification or multivariate methods to control confounding variables in the analysis.

Table 4. List of questions for the results of clinical journal article[4]

Components	Explanation
Flow and Baseline data	- How many subjects are randomized for each group? - How many and what is the percentage of subjects that completed the study? - How many and what is the percentage of subjects that did not complete the study? - What are the reasons for subjects being excluded from the study? - Are the baseline data comparable? - What had been done by the researcher when the baseline data are not comparable?
Primary outcomes	- What are the primary outcomes? - Are they statistically or clinicaly significant?
Secondary outcomes	- What are the secondary outcomes? (for example, side effect)

Primary outcomes of the study provide answers to primary research question; while secondary outcomes answer the additional research question. Principally, when reading the results of each study, we should perform clinical and statistical interpretation. The statistical interpretation could have been done by evaluating the p value and confidence interval of each hypothesis test that has been performed (usually if $p < 0.05$, the relative risk does not lied in the range of 1; proportional or mean value does not located in the range of 0). Moreover, clinical interpretation refers to comparison between the study outcome and effect size. Outcomes may appear as the value of relative risk and proportional difference (if the outcomes are categorical scale); mean difference (when the outcomes are numerical scale); it may also appear as hazard-ratio and incidence difference (for outcomes of survival).[4] Please notice that statistical significance, which is usually demonstrated as p value <0.05, should always be considered with its clinical significance.

In clinical trials, cohort and case-control studies, there are usually two comparable groups and the results are presented as a relative risk, i.e. a comparison of outcomes in the intervention and control groups. If the outcomes are measured as comparison of the occuring event (for example the number of recovered subjects compared to those who does not in a group), then such relative risk is known as odd ratio (OR). If the outcomes are calculated as a frequency, i.e. by comparing the study subjects who experienced the event with the total number of subjects in a group, then the relative risk is defined as risk ratio (RR). When there is no difference found between groups, the value of OR and RR is 1. Relative risk (either OR or RR) greater than 1 means that the number of events are greater in the intervention group compared to the control group. In contrast, when the relative risk is lesser than 1, the events in the intervention group is lesser.

Results of the study would be more useful when being presented as the risk differences. In this setting, we devide the proportion of events in the control group with those in the intervention group. Risk difference is also usually demonstrated as the number needed to treat (NNT) expressing how many subjects that should be treated during a period of time to achieve an outcome (treatment) or to prevent an event (prophylaxis/protection from disease). Epidemiologically, NNT characterizes of how many patients that would need health care in the form of treatment to reduce the number of cases, which is presented as a final endpoint. The following **Table 5** illustrates those parameters.

The Discussion

In the discussion section, we will find information about the researcher's steatement regarding validity, clinical significance and applicability that related to the study results. However, the researchers are usually only discuss

Table 5. How to calculate odd-ratio, risk ratio, and number needed to treat

	Number of patients	Number of events (Ex. recovery)	Odds of cure	Odd ratio	Risk of cure (frequency)	Risk ratio	Risk difference	Number needed to treat
Intervention	1,000	150	150/850	$\frac{150/850}{100/900} = 1.59$	150/1000	$\frac{150/1000}{100/1000} = 1.5$	$\frac{150}{1000} - \frac{100}{1000} = 0.05$ (=0.05%)	1 / 0.05 = 20
Control	1,000	100	100/900		100/1000			

validity limited to three ascpects, which are the power and limitation of study; comparing the study with other studies; and its explanation (biological plausabiity). A list of questions relevant to the discussion section is presented in **Table 6** as follows.

Table 6. List of questions for the discussions of clinical journal article[4]

Components
What are the power and limitation of study according to the researcher's point of view?
How is the comparison between the study results and other studies?
How is the biological plausibility of the study results?
Can the results of study be generalized?
How is the clinical significance of the study results?
How is the applicability of study results according to the researcher?

CONCLUSION

Critical appraisal is extremely needed as an assessment of information found in scientific journal articles before we are convinced to apply the information as guidelines in clinical practice. The level of reliability of a study tremendously depends on the study design, which places clinical trial as the highest in rank. Critical appraisal evaluates all components in a research article including the introduction, method, results and discussion. A good knowledge about how to evaluate each component is very necessary to perform a good critical appraisal. However, it should be considered that statistical significance should be compared with clinical significance.

REFERENCES

1. Oxford Centre for Evidence-Based Medicine. Levels of evidence. 2012. [cited on June 10, 2012]. Available online on http://www. cebm.net.
2. Belsey J. What is evidence-based medicine? London: Hayward Medical Communications; 2009.
3. Pwee KH. What is this thing called EBM? Singapore Med J. 2004;45:413-7.
4. Dahlan MS. Membaca dan menelaah jurnal uji klinis. Jakarta: Salemba Medika; 2010.
5. Makela M, Witt K. How to read a paper: critical appraisal of studies for application in healthcare. Singapore Med J. 2005;46(3):108-15.
6. Moore A, McQuay H. Clinical trials. Bandolier's little book of making sense of the medical evidence. Oxford: Oxford University Press, 2006.
7. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ. 2001;323:42–6.
8. Burls A. What is critical appraisal? What is series? London: Hayward Medical Communications; 2009..