

7–10

Class of evidence (CoE): how does it relate to critical appraisal?

Author Andrea C Skelly

Institution Spectrum Research Inc, Tacoma, WA, USA



Introduction

A cornerstone of evidence-based medicine (EBM) is the critical appraisal of clinical research. EBM facilitates *informed* interpretation of the literature and allows the clinician to integrate this interpretation with clinical expertise to facilitate decision making. In previous issues of EBSJ, topics central to performing critical appraisal have included:

- Study types, uses of different designs and potentials for bias
- Random assignment
- The importance of equivalent patient care and adequate follow-up
- Appropriate statistical analysis
- Understanding if study group differences are real or due to chance, statistical significance versus clinical significance
- How confounding may affect the study results

These are important to those doing research but the basic concepts are also vital to those reading the literature. It is not uncommon for published articles to make claims that are not supported by study design, available data, and its analysis. Study limitations may not be accurately discussed. Clinicians need to be able to filter studies through the lens of evaluation of a study's limitations as well as its strengths.

Critical appraisal provides important context around the findings and potential biases of a study (or studies) so as to help one put the results of a study in perspective. The purpose is to primarily determine to what extent the findings are real, due to chance, or due to bias.

Study validity and bias

Much of the medical literature, including editorials, letters, case reports, reviews, laboratory and animal studies, is interesting and informative. However, only a small fraction of the literature reports scientifically sound advances that can and should change how clinicians care for patients or how agencies determine policy.

Understanding of the above critical appraisal topics provide insight into the internal validity and external validity of a study. In other words, these concepts provide insight into the credibility and usability of a study.

Internal validity refers to the extent to which the study methods and analysis allow for appropriate conclusions about the study group(s) evaluated. Accurate and unbiased measurement and appropriate analysis are integral to internal validity. Internal validity is present when possible alternate explanations (eg, chance, confounding, bias) for study findings have been minimized or excluded so that appropriate conclusions about the study participants can be made. You can see how previously covered topics indeed assist with determining internal validity of a study. All studies (yes, even RCTs) have some potential for bias. The key is to determine the extent to which any potential bias might influence the soundness of the results and the ability to draw meaningful conclusions. Systematic differences between study groups set the stage for compromising study internal validity and reporting of erroneous results.

Bias, broadly defined as any systematic error which may lead to an inaccurate estimate of the true association between exposure and disease, can be divided into three general categories: (1) selection bias, (2) information bias, and (3) confounding [1–3]. Failure to take steps to reduce these biases during study design, execution, and analysis compromises the internal validity of a study.

Selection bias can occur when:

- The *exposure status* (for cohort studies) or the disease status (for case-control studies) influences an individual's opportunity for inclusion in a study. (When severity of disease influences what treatments are used, this can lead to the confounding by indication described in the previous issue of EBSJ.)
- There is *differential surveillance*, diagnosis or referral of individuals into the study. For instance, patients at high risk of a specific condition maybe more likely to be referred to a specialist for evaluation. This could influence the frequency with which a particular condition is diagnosed. It may also mean that these patients are more likely to be included as cases in a study. As a result the condition may appear to occur more frequently in this population secondary to more intense surveillance and testing instead of a true difference between study groups.
- *Participants differ* with respect to important characteristics from those who refuse to participate or did not respond (eg, to a questionnaire). For example, if elderly women at risk of osteoporosis are less likely to have a telephone and the study protocol requires a telephone interview about risk factors for osteoporosis, these women are less likely to be included in the study and any observed association with the factor can be distorted compared to the true value. Participants may differ from those who chose *not* to participate with regard to health status, potential risk factors, motivations and attitudes toward health [4]. This can influence the results.

Bias can result when participants are lost to follow-up. The effect on the observed association depends on the extent to which information about the exposure and outcome are known for those who are lost or quit and how they might differ from those who remain in the study. If the probability of loss to follow-up is related to both exposure and the outcome, the observed association may be distorted. If exposed persons are more likely to leave the study than unexposed persons when they develop the condition being studied, the estimate of risk may be biased.

Although not precisely an issue of selection bias, if the choice of places to sample is not representative of the population of interest, it may limit a study's generalizability.

Information bias, which can be thought of as any error in the ascertainment or handling of information, is a potential concern in all study designs. Measurement error may be one of the major sources of bias in a study [5]. The type and characteristics of this bias determine whether the observed association is an overestimate or underestimate of the true magnitude of the association.

- It is important that all attempts be made to accurately and comparably ascertain the exposure, the outcome, and other pertinent variables including potential confounders. Blinding of investigative personnel or blinded assessment of primary outcomes is important when feasible.
- In setting up the groups to be compared, decisions about how the outcome, exposure and other variables are to be classified need to be carefully considered, since there may be opportunity for *misclassification*. Misclassification that is non-selective (also called non-differential or random) biases results to the null. Misclassification can also be differential or selective, wherein one group is treated or measured differently. In contrast to random misclassification, here the proportion of misclassified participants differs between comparison groups and the estimate of the true association will be distorted in an unpredictable way.

Confounding is often referred to as a “mixing of effects” wherein the effects of the exposure under study on a given outcome are mixed in with the effects of an additional factor (or set of factors) resulting in a distortion of the true relationship. It is possible that an observed association is due, at least in part, to factors other than the exposure of interest and that the magnitude and even the direction of the observed association can change once these factors (ie, confounders) are taken into account in the analysis. The amount of association “above and beyond” that which can be explained by confounding factors provides a more appropriate estimate of the true risk which is due to the exposure. This type of bias was extensively discussed in the first issue of EBSJ for 2012.

External validity refers to the extent to which the results of a study can be applied to a specific patient population. Selection bias may compromise external validity. Obviously the characteristics of the study group(s) and how they are treated are important considerations; however, if the study does not possess internal validity, external validity does not exist.

Class or level of evidence: overview

The relative quality of evidence provided from clinical research is based on principles of evidence-based medicine and is termed “levels of evidence” (LoE) or “class” of evidence (CoE). Class of evidence is a hierarchical rating system for classifying the overall quality of an individual study. Ideally, it should take into account factors that influence study validity and *provide a summary assessment of the potential for bias*. CoE assists in formalizing the elements of critical appraisal to arrive at an overall perspective on the credibility of a study and specific limitations that may compromise the ability to draw conclusions from it.

There are various rating systems available for rating levels of evidence, most built primarily on study design, where randomized controlled trials are considered to provide the highest level of evidence and cohort studies provide a lower level [6]. Many systems for the critical appraisal of clinical studies have been described in the literature that included factors in addition to study design [7–10]. Overall, the intent is to evaluate the potential sources of bias in studies, allowing the astute reader to assess the extent to which such biases may influence the results. In general, all schemes assess components of studies that may introduce bias into studies such as the following:

- Study design (eg, prospective cohort, randomized controlled trial)
- Patient selection and evaluation methods (including outcomes evaluation)
- Patient follow-up (length and completeness)
- Sample size and ability to detect differences beyond the role of chance
- Consideration of and controlling for potentially confounding factors

Many rating systems primarily consider study design, with randomized controlled trials (RCTs) considered the “best.” Poorly conducted RCTs may yield less valid results than well-conducted prospective cohort studies. Each of the factors above plays a role in the internal validity of a study regardless of study design. The epidemiological basis for these factors and how they relate to class/level of evidence ratings will be the focus of the next article in this series. We will discuss how the CoE for studies reported in EBSJ is determined.

References

1. **Koepsell T, Weiss N** (2003) *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press.
2. **Weiss NS, Koepsell TD** (2003) *Epidemiologic Methods: Studying the Occurrence of Illness*. New York: Oxford University Press.
3. **Weiss NS** (2006) *Clinical Epidemiology: The Study of the Outcome of Illness*. New York: Oxford University Press.
4. **Hennekens C, Buring JE** (1987) *Epidemiology in Medicine*. Boston, Toronto: Little, Brown & Co.
5. **Armstrong B, White E, Saracci R** (1992) Principles of exposure measurement in epidemiology. In: *Monographs in Epidemiology and Biostatistics, Volume 21*. New York: Oxford University Press.
6. **Wright JG, Swiontkowski MF, Heckman JD** (2003) Introducing levels of evidence to the journal. *J Bone Joint Surg Am*; 85(A):1–3.
7. **Atkins D, Best D, Briss PA, et al** (2004) Grading quality of evidence and strength of recommendations. *BMJ*: 328:1490.
8. **Phillips B, Ball C, Sackett D, et al** (2001) Levels of evidence and grades of recommendation. Available at: http://www.cebm.net/levels_of_evidence.asp.
9. **Poolman RW, Struijs PA, Krips R, et al** (2006) Does a “Level I Evidence” rating imply high quality of reporting in orthopaedic randomised controlled trials? *BMC Med Res Methodol*; 6:44.
10. **West S, King V, Carey TS, et al** (2002) Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute–University of North Carolina Evidence-based Practice Center, Contract No. 290-97-0011). Rockville, MD: US Agency for Healthcare Research and Quality.

EDITORIAL PERSPECTIVE

Spine Research Analysis 1.01

It is the entirely biased and scientifically unfounded opinion of the EBSJ editors that the AOSpine International readership takes note of this article. Of many excellent contributions in the EBSJ series titled ‘Science in Spine’ to date this article provides a most profound and insightful review of fundamentally important elements concerning how to do successful research and how to critically interpret results.

We suggest making this article required foundational reading for all trainees and other interested parties in all AOSpine locations around the world as the new academic year begins.