9—12

Class or level of evidence: epidemiologic basis

Author Joseph Dettori
Institution Spectrum Research Inc, Tacoma, WA, USA

"I wouldn't have seen it if I hadn't believed it." Yogi Berra

Introduction

Class of evidence (CoE) is a hierarchical rating system used by EBSJ and most major scientific publications for classifying the overall quality of an individual study. It is a shortcut to identifying what is likely the best (or worst) evidence on a given topic. The "classes" range from I to IV with "CoE I" representing the highest level of evidence, and "CoE IV" representing the lowest level. Assigning a CoE to an individual article is an attempt to provide the reader with a relative assessment of the research study's risk of bias; that is, the likelihood that the results of the study are influenced by various biases rather than the intervention. This article intends to open the eyes of its readership to the many potential confounders and to look behind the claims of CoE 1.

Common sources of bias EBSJ considers when critically appraising a study include:

- Patient selection and allocation of treatment
- Intention-to-treat analysis
- Blind or independent assessment for important outcomes
- Co-interventions applied equally to study groups
- Patient follow-up rate of less than 85%
- Adequate sample size
- Controlling for possible confounding

Patient selection and allocation of treatment

How patients are selected and allocated for treatment in a clinical study of efficacy and safety is paramount. Ideally, patients are selected based on chance to protect against selection bias and confounding [1]. That is why a randomized controlled trial (RCT) is considered the best study design in reducing the risk of bias and achieving a high CoE. It is possible, however, when one conducts an RCT, to still introduce bias into the allocation

process. How? Bias can be introduced by allowing those who enroll patients into a study to have access to upcoming assignments. Having access gives the enroller knowledge of the next assignment that could then influence whether a patient is included or excluded based on perceived prognosis. Therefore, care must be taken to ensure that the allocation of the patient to a particular treatment group is concealed; in other words, that the implementation of the random allocation sequence occurs without prior knowledge of treatment assignment [2]. Some argue that RCTs that do not provide for proper allocation concealment overestimate the effect of a treatment as much as 30%–40% [3]. In the critical appraisal process, one should evaluate whether the allocation was concealed. If it is not reported, be suspicious of potential bias.

Intention-to-treat analysis

Investigators can undermine random assignment in another way—systematically excluding from the results those patients who do not receive the assigned treatment. The reason that patients do not receive the treatment they are assigned often relates to prognosis [4]. For example, some patients who are randomized to a surgical arm of a study may not undergo surgery due to other comorbidities. If these patients who are likely to have a poor outcome are excluded from the surgical arm of the trial because they did not receive treatment, and are instead included in the control arm of the study, bias in favor of the surgery will be erroneously reported. Therefore, it is important to evaluate whether investigators analyzed all patients in the groups to which they were randomized, the so-called intention-to-treat analysis. Having a comprehensive denominator with accounting for all patients who received treatment for a certain condition is essential to allow outside reviewers to screen for bias.

Blind or independent assessment for important outcomes

Personnel who measure or assess the outcomes of interest often have a belief or suspicion of which treatment offers the best outcome. If they are privy to the treatment administered, they may interpret marginal results in a way that favors their presupposition. That is why studies, when possible, should have those who are evaluating the results blinded to the treatment. Another way that bias can enter into the unblinded assessment process is through differential encouragement during a performance test. In some cases, the effect of differential encouragement can be as large as the effect of a beneficial therapy[5]. Some outcomes are not measured by a third party but rather are reported directly by the patient (patient-reported outcomes), such as with the Scoliosis Quality of Life Index (SQLI) or the Neck Disability Index (NDI). In these cases, it is best if the patient is blinded to the treatment. Often in surgical trials, neither the patient nor the evaluator can be blinded, particularly when surgery is compared with nonoperative care. In these situations, certain measurements can be obtained by independent individuals not part of the research study. A measurement of radiographs from a radiologist not associated with the study is an example of independent assessment. Blinding is most often done in prospective studies. However, retrospective studies can also qualify for blinding in cases when outcomes are absolute and reliable, such as in death or reoperation. These outcomes need no interpretation and are not subject to differential encouragement.

Co-interventions applied equally to study groups

Co-interventions (additional treatments or therapies) should be applied equally between study groups. Co-interventions are not applied equally when patients in one treatment arm receive additional interventions not given to the comparison group, or when one treatment arm is followed-up more intensely than the other.

Patient follow-up rate of less than 85%

At the end of a clinical study, the investigators should know the status of each patient with respect to final evaluation. Patients who do not provide outcomes at the evaluation time (those lost to follow-up for any reason) often have a different prognosis from those that do. For example, some patients may have done so well following treatment that they decided there was no need for follow-up, or they may have experienced adverse events that prevented them from returning or induced them to seek care elsewhere. The larger the proportion of patients who do not return for follow-up, the greater the likelihood the validity of the study is compromised.

Adequate sample size

Many spine studies have relatively few patients, particularly for those conditions that are not so prevalent. Compounding the problem of few study subjects is that some outcomes, such as complications or adverse events, are rare. Too few patients and rare outcomes both contribute to the problem of inadequate sample size. The result of an inadequate sample size is that the investigator may not have the necessary statistical power to detect important differences in outcomes between treatments. As a result, the conclusions that there are no differences between groups may be wrong. When this occurs (when investigators claim there is no difference when there really is a difference) it is called a type II error. A type II error is most often caused by an inadequate sample size. The validity of the results from a study that demonstrates no statistical difference when an important clinical difference is present should be suspect.

Controlling for possible confounding

The purpose of random assignment is to create two or more treatment groups that are similar at baseline with respect to prognosis. Studies with small sample sizes are more prone to have unbalanced prognostic factors between groups. Furthermore, non-random (cohort) studies, no matter how large, are likely to have differences in characteristics between groups that could influence prognosis. In either situation, investigators should evaluate the distribution of all known baseline prognostic factors in the treatment and control groups. If differences are substantial, look for an analysis that adjusts for these differences using regression or stratified analysis. These analyses control for possible confounding due to unequal distribution of baseline prognostic factors.

Putting it together

The highest class of evidence (the lowest risk of bias) for EBSJ is Level I, defined as a good quality RCT. A good quality RCT demonstrates all principles discussed above. A Level II study is either a RCT that violates any of the above criteria, or a good quality cohort study that includes blind or independent assessment, follow-up rate of ≥ 85%, adequate sample size, and controlling for possible confounding. A violation of any of the principles establishing a good quality cohort study reduces it to a Level III study. Likewise, all case-control studies are considered Level III studies when assessing therapeutic effectiveness and safety. Finally, all case series are considered Level IV studies since there are no controls with which comparisons can be made.

With these principles clearly stated, it must be recognized that randomized controlled surgical trials present problems not seen in pharmaceutical trials. Barriers that make RCTs involving surgery difficult to design and perform are well documented [6, 7]. It is clear that the scientific spine community cannot perform RCTs to answer every clinical inquiry. This begs a bigger question: how does the scientific spine community prioritize

potential trials weighing the resources required to conduct the trial and the value of the information likely gained? Some have suggested that at least three lines of inquiry are required: (1) evaluation of the information that would be gained if the trial was executed successfully, (2) feasibility of the study, and (3) the resource cost of conducting the study [8]. In addition to prioritizing RCT topics, spine surgeons need to improve the quality of non-randomized comparative studies. In doing so, not only will the spine community get closer to the truth of the effectiveness of treatment but it will also be able to apply the results to a wider, real-world population.

References

1. **Norvell DC** (2010) Study types and bias: don't judge a study by the abstract's conclusion alone. *EBSJ;* 1(2):7–10.
2. **Dettori J** (2010) The random allocation process: two things you need to know. *EBSJ;* 1(3):7–10.
3. **Schulz KF, Chalmers I, Hayes RJ, et al** (1995) Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA;* 273(5):408–412.
4. **Bhandari M, Guyatt GH, Swiontkowski MF** (2001) User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am;* 83-A(6):916–926.
5. **Guyatt GH, Pugsley SO, Sullivan MJ, et al** (1984) Effect of encouragement on walking test performance. *Thorax;* 39(11):818–822.
6. **Fung EK, Loré JM Jr** (2002) Randomized controlled trials for evaluating surgical questions. *Arch Otolaryngol Head Neck Surg;* 128(6):631–634.
7. **Wright JG, Katz JN, Losina E** (2011) Clinical trials in orthopaedics research. Part I. Cultural and practical barriers to randomized trials in orthopaedics. *J Bone Joint Surg Am;* 93(5):e15.
8. **Katz JN, Wright JG, Losina E** (2011) Clinical trials in orthopaedics research. Part II. Prioritization for randomized controlled clinical trials. *J Bone Joint Surg Am;* 93(7):e30.