

9–12

Assessing bias: the importance of considering confounding

Authors: Andrea C Skelly, Joseph R Dettori, Erika D Brodt

Institution: Spectrum Research Inc, Tacoma, WA, USA

Overview: Why evaluate bias? Isn't statistical significance enough?

It is common to come across a study that reports that treatment A “provided significantly better pain relief” than treatment B.

When a difference in an outcome (eg, pain) between exposures (eg, treatment groups) is observed, one needs to consider whether the effect is truly because of exposure or if alternate explanations are possible. As we discussed in the previous issue of EBSJ, to evaluate the validity of a research study, factors that might distort the true association and/or influence its interpretation need to be carefully considered. This means evaluating the role of bias and considering the study's statistical precision. In the previous issue, we discussed aspects of statistical testing and precision to explore the question of whether an effect is real or due to chance. We also considered some caveats to concluding that a “statistically significant” result is clinically meaningful.

This article takes a look at the potential for bias and its impact.

Bias relates to *systematic* sources of error which need to be considered. Why? The internal validity of a study depends greatly on the extent to which biases have been accounted for and necessary steps taken to diminish their impact. In a poor-quality study, bias may be the *primary* reason the results are or are not “significant” statistically! Bias may preclude finding a true effect; it may lead to an inaccurate estimate (underestimate or overestimate) of the true association between exposure and an outcome. Significance testing in itself *does not* take into account factors which may bias study results.

Bias can be divided into three general categories: (1) selection bias; (2) information bias; and (3) confounding.

This article focuses on confounding. Recognizing it and controlling for its effects are important to a study's credibility.

What is confounding?

Confounding is often referred to as a “mixing of effects” [1, 2] 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. In a clinical trial, this can happen when the distribution of a known prognostic factor differs between groups being compared.

Confounding factors may mask an actual association or, more commonly, falsely demonstrate an apparent association between the treatment and outcome when no real association between them exists.

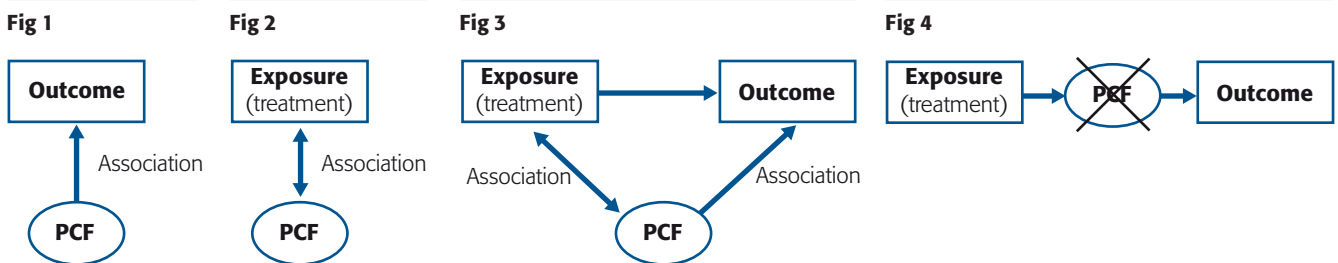
The existence of confounding variables in studies make it difficult to establish a clear causal link between treatment and outcome unless appropriate methods are used to adjust for the effect of the confounders (more on this below). Confounding variables are those that may compete with the exposure of interest (eg, treatment) in explaining the outcome of a study. The amount of association “above and beyond” that which can be explained by confounding factors provides a more appropriate estimate of the true association which is due to the exposure.

General characteristics of confounders include the following:

1. A true confounding factor is predictive of the outcome even in the absence of the exposure. Although a potential confounding factor (PCF) may be causative, it might not be. The primary requirements are that an independent relationship between the factor and the outcome exists and that the PCF not be the result of the exposure (or the outcome). In fact, many of the PCFs which often must be evaluated are proxies for variables which are complex and difficult to measure (**Fig 1**).
2. A confounding factor is also associated with the exposure being studied but is not a proxy or surrogate for the exposure. In clinical trials, confounding is often a result of unequal distribution of the potential confounding factors between treatment groups (**Fig 2**).

A situation that contains both numbers 1 and 2 sets the stage for potential confounding (**Fig 3**).

3. A confounder cannot be an intermediate between the exposure and the outcome. For example, the relationship between diet and coronary heart disease may be explained by measuring serum cholesterol level. Cholesterol is not a confounder because it may be a causal link between diet and coronary heart disease (**Fig 4**).



How does confounding influence results? An example in spine surgery research:

Let's imagine that we wanted to know if treating index osteoporotic vertebral fractures with vertebroplasty increased the risk of subsequent vertebral fractures. We evaluate 400 patients with index vertebral fractures, 200 of whom received vertebroplasty and 200 did not. After 2 years, we identified 45 subsequent fractures with the following fictitious distribution:

Vertebroplasty	Conservative care	Relative risk (95% confidence interval)
30/200 (15%)	15/200 (7.5%)	2.0 (1.1–3.6)

At first glance, it looks like those who received vertebroplasty were at a much higher risk (two times the risk) compared with those who did not (this is called a “crude” estimate of the association). However, it is important to investigate whether other reasons could account for this difference. In particular, other variables that may influence the risk of subsequent vertebral fracture should be evaluated such as age, weight, and smoking status. The data below describes these variables at the time of the incident fracture. Note that age and weight are similar between groups. But, the difference in the proportion of patients who smoke is dramatically different in that 55% of the patients in the vertebroplasty group smoke compared with only 8% in the conservative care group, as outlined below.

	Vertebroplasty N = 200	Conservative care N = 200
Age, y, mean±SD	78.2±4.1	79.0±5.2
Weight, kg, mean±SD	54.4±2.3	53.9±2.1
Smoking status, No. (%)	110 (55)	16 (8)

If we stratify the results by smoking status, we note that the risk of subsequent fractures is similar between treatment groups in each stratum (smoking and nonsmoking) such that the relative risk (RR) is closer to 1 (no effect) compared with the overall results above where RR was 2.

Smoke			No smoke		
Vertebroplasty	Conservative	RR (95% confidence interval)	Vertebroplasty	Conservative	RR (95% confidence interval)
23/110 (21%)	3/16 (19%)	1.1 (0.4, 3.3)	7/90 (8%)	12/184(7%)	1.2 (0.5, 2.9)

Thus, smoking was a confounding factor distorting the true relationship between vertebroplasty and the risk of subsequent vertebral fractures.

Confounding by indication—a special and common case of confounding

With regard to the assessment of a technology or surgical procedure, confounding may take the form of an indication for use of that technology or procedure [2, 3]. Such “confounding by indication” may be extremely important to consider in either studies of efficacy or of safety.

In a hypothetical study, let’s suppose that all patients who received treatment A had more severe disease than those who received treatment B and that there was a statistically significant difference showing that treatment B resulted in better patient function. Is it valid to conclude that treatment B is truly better than treatment A? No! Given that the severity of the condition is likely associated with the outcome and that the severity is also associated with the treatment choice, the effects of the treatment cannot be separated from those of the severity.

To compare the effectiveness of two treatments, the only way to deal with this is to ensure that the study design requires patients with the same range of condition severity are included in both treatment groups and that choice of treatment is not based on condition severity.

Dealing with confounding

The potential for confounding should be considered in the design and implementation of the study. Factors which might be associated with the outcome other than the treatment

need to be measured. To some extent, confounding can be accounted during analysis, assuming that such factors have been measured as part of the study.

Step 1: Measure and report all potential confounders

Patient characteristics are an often underreported or misreported set of measurements in spine care studies but are extremely important to quantify and report as they may be potential confounders. Diagnostic features, comorbidities, and any factor that might affect patient outcome needs to be measured and reported for each study group as well. Any and all of these characteristics, features, and factors may be potential confounders of the relationship between your “exposure of interest” (eg, a surgical treatment) and the outcome (eg, patient function). Planning for and measuring these attributes goes a long way toward dealing with the role of confounding.

Step 2: Routinely assess the role of confounding factors and adjust for them in analyses

There are a number of ways of assessing and adjusting for confounding, however a detailed discussion of this is beyond the scope of this article. Briefly, a few examples of how this could be accomplished include:

1. During study planning, inclusion could be restricted by specific confounding variables, such as age.
2. Several methods of “adjusting” the effect estimate as part of the analysis can be used. Stratification (as shown above) is one that can be relatively straightforward and involves looking at the association between the exposure and outcome for each factor category (or stratum) by calculating a stratum-specific estimate.
3. Multivariate analysis, a set of statistical methods which allows for adjustment of multiple variables simultaneously via mathematical modeling, can also be used to “control” for confounding.

Basic concepts for these methods for control of confounding during analysis are the subject of future articles.

Step 3: Report adjusted and crude estimates of association and discuss limitations of the study that may be due to confounding and the magnitude of the influence

Regardless of the method used, an *adjusted* estimate should be obtained which reflects the degree of association between the exposure and disease that remains *after* the effects of the confounder have been “removed.” In general, if the adjusted estimate is different from the crude estimate by approximately 10% or more, the factor should be considered a confounder and the adjusted estimate used as a more reliable indicator of the effect of the exposure, ie, as an estimate of the effect “above and beyond” that is due to the confounder(s).

SUMMARY

Failure to evaluate demographic and clinical factors as potential confounders can bias your study results and lead to erroneous conclusions. Study design must include the measurement and reporting of such factors. During analysis, the association between such factors and the outcome and your exposure of interest must be explored. A commonly overlooked type of confounding in the surgical literature is confounding by indication. This needs to be dealt with during study design to ensure that treatment groups include patients with the same range of condition severity and that treatment choice is not based on condition severity. In all likelihood, no matter how many variables one adjusts for, there will be residual confounding, possibly by factors that are unknown and cannot be measured.

REFERENCES

1. **Rothman KJ** (2004) A potential bias in safety evaluation during open-label extensions of randomized clinical trials. *Pharmacoepidemiol Drug Saf*; 13(5):295–298.
2. **Weiss NS** (2006) *Clinical Epidemiology: The Study of the Outcome of Illness*. 2nd ed. New York: Oxford University Press.
3. **Greenland S, Neutra R** (1980) Control of confounding in the assessment of medical technology. *Int J Epidemiol*; 9(4):361–367.