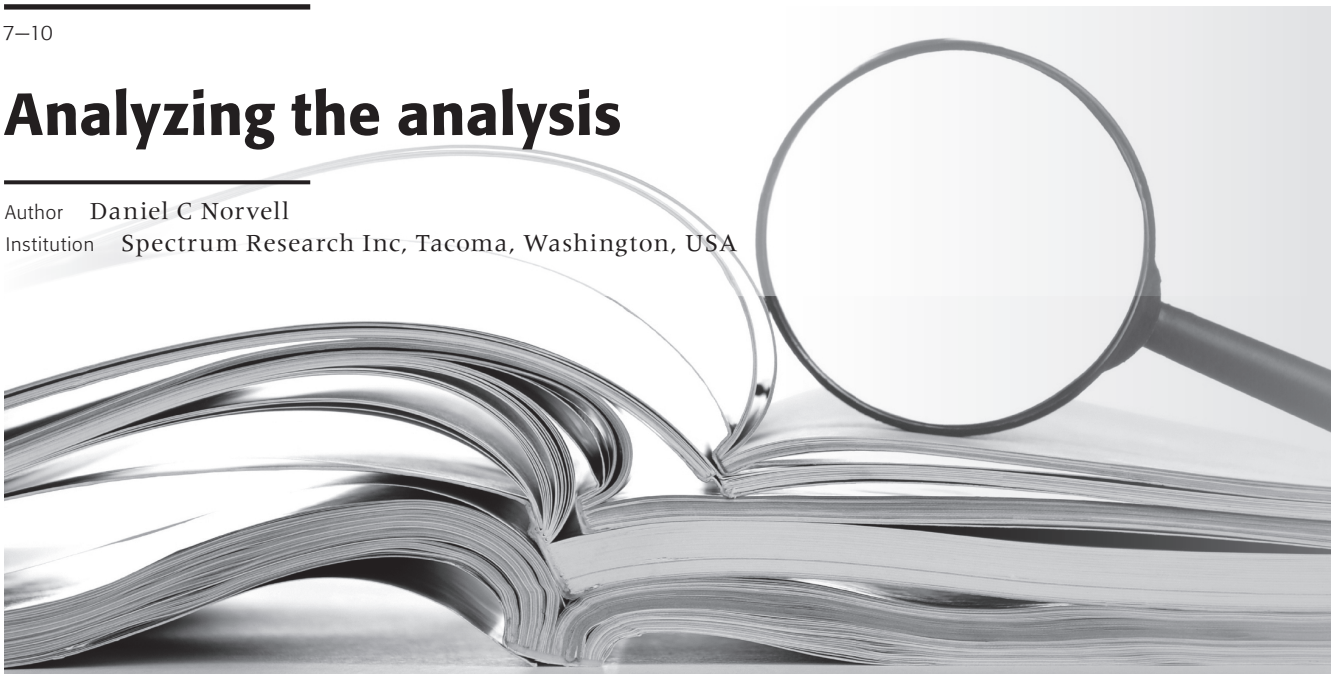
7–10

Analyzing the analysis

Author Daniel C Norvell
Institution Spectrum Research Inc, Tacoma, Washington, USA



Research is often seen as tremendously time- and resource-intensive, difficult to undertake save for the initiated few. In fact, research is actually an uncomplicated undertaking if clearly planned and executed in a straightforward method. The ultimate reward is that of discovery. Proving the expected theory can be gratifying, but nothing comes close to finding something unexpected and then searching for an explanation. It is the unraveling of a mystery which probably ranks as one of the preeminent satisfactions for an academically minded person.

Prior articles have laid the foundation of critical appraisal by discussing such topics as study types and bias, random assignment, and the importance of equivalent patient care and adequate follow-up. All these factors profoundly influence the quality and credibility of a research article. It is the analysis which forms the culmination of the research effort. If done right, it can be a major accomplishment; if flawed, it is a wasted effort or even worse—may lead to wrong conclusions with unforeseeable consequences.

Analysis is the consideration of everything from description and characterization of the study population to the analytical statistics performed for treatment comparisons. The proper statistical analysis is dependent on the study objectives and design, and the method and source of data collection. Without needing a deep understanding of biostatistics, there are a few basic tenets one can easily evaluate to determine if the appropriate analyses were used when critically appraising an article.

Most analyses are divided into *descriptive* and *analytical* statistics and should be clearly described in the methods section. Descriptive statistics are most frequently used to provide general information about patients and factors that may be related to outcomes. They set the stage for some of the analytical methods (such as control for confounding) that may be needed to ensure the most accurate estimate of a study effect. Analytical statistics allow for evaluation of treatment effects and the associations between factors. Both may involve tests of statistical significance.

The following provides an overview of what to look for when critically appraising a study's analytic methods.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Is the population described or characterized in a summary table (descriptive statistics)?
The presentation of descriptive data on the study population is important for a number of reasons, and some important elements are overlooked in published articles. For example, it enables you to determine the comparability of study groups at baseline and evaluate the likelihood of any selection bias or confounding (see definition below).

A prospective study allows one to collect all known and suspected potential confounders. For retrospective studies, the author is limited to what was already collected and often does not have access to factors that likely influence outcome (eg, smoking status).

It enables you to assess whether all important factors that may influence outcome were collected. In instances when an analysis cannot include all known or suspected confounders, the treatment effects may be biased. This is known as omitted variable or residual confounding bias and is often a problem in retrospective studies (see box). When known potential confounders cannot be included in an analysis, the author should acknowledge this as a limitation and describe the anticipated effect.
The baseline characteristics of the study population can help in determining the generalizability of the results to your own study population.

Baseline scores for pain, function, and quality-of-life measures should be presented, especially when used as an outcome or associated with the outcome of interest. The absolute scores at follow-up are often associated with the scores before treatment. Finally, the descriptive tables presented in a study report typically describe all enrolled patients. This can allow you to determine, when not explicitly stated, the extent of loss to follow-up.

Definition: Confounding is a type of bias. A *confounding factor* is something that is associated with the exposure of interest (eg, treatment) and is also prognostic factor for the outcome. Furthermore, these factors often influence which treatment the subject receives in non-randomized observational studies. As a result, studies with groups where there is an imbalance of a confounding factor between groups can lead to misleading results like overestimating or underestimating the treatment effects if these factors are not carefully identified beforehand and controlled in the analysis.

A common example is smoking. If one group has more patients who are smokers than in another, the group with more smokers may be at an unfair disadvantage when comparing outcomes. This is known as *confounding bias.* Table 1 of every manuscript should provide a detailed account of important baseline factors in each group (In fact, this is a requirement for EBSJ.)

The total number of subjects in each category should be presented, the proportion of the overall study subjects in that cohort, and a statistically significant *P* value among factors between groups should be presented. By studying the descriptive statistics in Table 1 of a paper, you can quickly discern imbalances between groups and the possibility of confounding. The potential effects of confounding will be presented in an upcoming Science in Spine article.

Are the results reported analytically (analytic statistics)?
The purpose of analytical statistics is to assess the effects of treatment and risk factors on specific outcomes and usually to determine the probability that the observed results are due to chance. This evaluation/assessment relies on the testing of statistical hypotheses. The testing of statistical hypotheses (sometimes called testing of statistical significance) is important for determining treatment safety or efficacy. Statistical tests aim to distinguish true differences (associations) from chance. As all research is performed on samples of subjects, there is always a possibility that the results observed are solely due to chance and that no true differences exist between the compared treatment groups. Statistical tests help sort out how likely it is that the observed difference is simply due to chance. Commonly, an arbitrary test threshold value (eg, alpha = .05) is used to distinguish results that are assumed to be due to chance from results that are due to other factors. If the probability

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

that the results are due to chance is less than the threshold value ($P < .05$), it is assumed the differences are because of these other factors (eg, true differences in treatment effects).

When critically appraising the appropriate analytical statistics, consider the following:

- Is there an analytical test for each stated objective?
- Was the outcome a score (eg, pain, function, or quality of life)? If so, did the authors control for the baseline score or calculate the change score?
- Was the analysis method appropriate for the outcomes used? Did they perform a stratified analysis for categorical risk factors and outcomes? Was logistic regression used when estimating treatment effects for binary outcomes when controlling for potential confounders? Was linear regression used for continuous outcomes?
- Were the data presented clearly, concisely, and transparently to include tables of regression coefficients and confidence intervals?
- Were appropriate effect measures used to measure the objectives and to support the results and claims?

The following sections briefly clarify the final three considerations.

Stratified analysis

A good start when assessing an article's analysis methods section is whether the authors conducted a stratified analysis before more sophisticated methods of regression. Such an analysis allows one to assess the distribution of individual variables and their impact on categorical outcomes that will lead to a more relevant and strategic model development. Stratified analysis also allows for the assessment of heterogeneity of treatment effects (HTE) (ie, effect modification), which was discussed in the previous Science in Spine article.

Regression

Regression is a powerful tool for evaluating treatment effects while controlling for potential confounding factors or assessing statistical effect modification. When more than a few variables (strata) are formed for stratified analysis, or when more than a few potential confounding factors need to be adjusted, multiple regression can be used. This allows for the control of multiple factors simultaneously and produces such effect estimates (often called point estimates) as regression coefficients (for continuous outcomes from a linear regression model) rate/risk/hazard ratios (RR) (for dichotomous outcomes from a binomial or cox or regression) or odds ratios (OR) (for dichotomous outcomes from a logistic regression), and their 95% confidence limits. Regressions can also be used to predict outcomes. When reading an article that uses a multiple regression technique, look for the full model (ie, description or listing of all variables included in the model) to be presented, not just the adjusted treatment effects. If the journal limits the amount of information that can be presented, the complete regression should be available to reviewers and readers in the appendix.

When considering the strength of the effect estimate (RR or OR) from a regression model, the *P value is less important than the confidence interval.* Extremely wide confidence intervals indicate wide variability and the estimate may not be stable. Results for which estimates are surrounded by wide confidence intervals should be interpreted with caution even when associations are statistically significant.

Other useful effect estimates

Effect estimates are more useful than *P* values that essentially have little to no clinical utility, especially when interpreted with a 95% confidence interval. One should consider the authors' use of effect estimates and how clinically useful they are. For example, an OR can be hard to interpret and the relative difference in treatments or factors is often

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

overestimated when the outcome of interest is high (eg, success or complication rates >10%). However, ORs are commonly presented in papers because logistic regression is so universally used. One must always ask: "How can I apply these findings clinically? And how may it change how I treat or counsel patients?"

Other effect measures with strong clinical utility that may be presented in a manuscript include the relative risk reduction, the risk difference, and the number needed to treat. Suppose an article is comparing the results of instrumented spine fusion with disc arthroplasty. The authors report that the proportion of implant failure among spine fusion is 20% and among disc arthroplasty 10%.

- *The relative risk (RR)* is simply the proportion of patients with the outcome in one treatment group divided by the proportion of patients with the outcome in another treatment group. In this case, 0.10/0.20 = 0.50.
- *The relative risk reduction (RRR)* is |1–RR| x 100, or in our case, (1–0.5) x 100 = 50%. An RRR of 50% means that the disc arthroplasty group reduced the risk of implant failure by 50% compared with instrumented spine fusion. If the treatment increases the risk of a bad event, we call that *relative risk increase (RRI).* Furthermore, when a treatment increases the probability of a good event, the term we use is *relative benefit increase (RBI).*
- The *risk difference (RD)* is the absolute difference between the proportions, 0.20 −0.10 = 0.10 or 10%.
- The number needed to treat (NNT) represents the number of patients one would need to treat to prevent a negative outcome (or allow a positive outcome, depending on which outcome is being evaluated). The formula is 1/RD. In our example, 1/.10 = 10; therefore, for every 10 patients treated with disc replacement, one implant failure can be prevented compared with spine fusion.

Table 1 A summary of how to report implant failure for fictional data comparing an instrumented spine fusion with disc replacement

Disc replacement (n = 30) No. failed (%)	Spine fusion (n = 30) No. failed (%)	RR B/N	RRR 1-(B/N)	RD N-B	NNT 1/(N-B)
3 (10)	6 (20)	0.50	.50	.10	10

Summary

When critically appraising an article, understanding the use and misuse of statistics is imperative. In summary, a cogent analysis follows a few logical steps:

- What are the study objectives?
- Was the study design appropriately for the objectives?
- Does the data analysis section report clearly the descriptive statistics used and do the analytical statistics map directly to the objectives?
- Was there a strong attempt to collect, report (**Table 1**), and account for all known and unknown potentially confounding factors?
- Was the outcome a score (eg, pain, function, quality of life)? If so, did the authors control for the baseline score or calculate the change score?
- Was the data presented clearly, concisely, and transparently to include tables of regression coefficients and confidence intervals?
- Were appropriate effect measures (and statistical analyses) used to measure the objectives and to support the results and claims?
- Answers to such questions do not affect the class of evidence which is based mostly on study design; however, it may affect how you would interpret, apply, and even trust the results of a given study you are reviewing.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.