

Appraisal of Journal Articles: Asking the Right Questions

By Michael Haneline, DC, MPH, and Robert Cooperstein, MA, DC

There are no strict rules that can be applied to evaluating research articles. Indeed, there are quite a few different methods and checklists that reflect varying opinions on how an appraisal should be conducted¹.

These different methods, however, are intended to assist research publication readers in determining the extent to which a study is valid and if its findings are applicable to a given clinical situation. The purpose of this article is to develop questions suitable for the appraisal of chiropractic-related research.

The assessment of any form of clinical evidence requires familiarity with the methods utilized in the specific type of research design being evaluated. In order of decreasing rigor, randomized controlled trials (RCTs), cohort studies, case-control studies, literature reviews, and case reports are the main research designs encountered by doctors of chiropractic. In addition to accomplishing primary studies, investigators may perform a study evaluating the evidence produced by other studies (systematic and narrative reviews²) or may even combine the data from several small studies to increase the sample size, and thus the power of the research conclusions (meta-analysis). No particular research design is always better or worse than another, although one may be more appropriate for the phenomena under investigation. The primary concern is that the study design be capable of answering the question posed by the investigators. Once the reader becomes acquainted with

the basic methods involved in these designs, the various components utilized in a given study can be examined to judge their degree of compliance with acceptable procedures.

Primary Issues in Appraising Research Articles

Three issues should be considered when reviewing any journal article.^{3,4}

1. Are the results of the study valid? This issue has to do with whether or not the methodology employed in the study was carried out correctly, whether the statements made by the authors were logical and supported by the actual findings of the study, and whether severe confounding factors were present—at minimum discussed by the authors as limitations of the study.

2. What are the results of the study? If a treatment effect was demonstrated, was it large enough to be considered clinically significant? If a cause-and-effect relationship was suggested, was the research design capable of supporting such a conclusion?

3. Will the results of the study help my patients? The study's findings must be applicable to the doctor's practice setting and patients to be helpful. Studies that involve populations that differ from the doctor's target patients (such as emergency room patients) may not be relevant.

Each of these 3 issues is comprised of several steps that ultimately generate an appraisal checklist. Although not required for the general

reader, appraisal checklists are commonly employed as aids during the process of evaluating articles (please refer to the appendix of this article for a general checklist). We would not expect a doctor to customarily use such a checklist, completing one for each article read, but it would be worthwhile to read it from time to time to sharpen one's ability to assess the quality of a research publication.

After appraising evidence, a clinician has to make a decision about whether to apply its findings to practice. Some studies may be thought so flawed following the appraisal process that they should be considered worthless, others may be thought worthy of changing clinical practice to some degree, while the best studies may be judged to warrant an immediate and major impact upon the practice. Since evidence is rarely black and white, and since some study conclusions may seem more founded than others, we must all develop means of making judgment calls in evaluating research literature. The tools of appraisal can be very helpful in making the right decision.

General Appraisal Questions

Since explicit protocols are utilized in carrying out research, the appraisal of research articles is essentially a matter of determining whether each of the required steps was actually completed and if so, whether each was implemented properly. Points to consider when are listed below, along with explanation as necessary. Because of the general nature of these questions,

Philosophy & Technique

some items will not apply when reviewing certain types of research designs. Simply ignore these inapplicable items and go on to the next question.

- *Was the purpose of the study conveyed plainly and rationally?*

Without a clearly stated purpose in the article's introduction, the reader can never be certain that the authors conducted the research with a purpose in mind or whether they were in reality conducting a "fishing expedition." The statement of purpose is the one of the most important elements of a scholarly paper. If missing or incomplete, the reader may want to consider not reading the remainder of the article. Sometimes it becomes painfully obvious that the investigators gathered a lot of data without a primary question in mind, and then mined (or even "laundered") the data looking for correlations. This increases the likelihood of chance associations coming to light. Although there are statistical methods for dealing with such fishing expeditions, it would have been better to have never embarked upon one.

- *Was the research design clearly described and was it consistent with the study's purpose?* A number of different research designs are available. Each one has different strengths and weaknesses. For instance, randomized controlled trials (RCTs) are used to investigate the effects of treatment, cohort studies are used to examine relationships between risk factors and health status, and cross-sectional studies are used to define the incidence or prevalence of a condition. Thus, it would be inappropriate to use a cohort study design to demonstrate a treatment effect. Although it is common to extol the rigor of the RCT, often at the expense of more modest practice-based research

(PBR: a design in which doctors simply describe what they did and what happened in typical clinical settings), each has its legitimate role to play in adding to the evidentiary basis of clinical practice.

What PBR lacks in rigor, it may make up in relevance, especially in situations such as those encountered in the practice of chiropractic where it is difficult to effectively blind the various study participants as to "real" vs. "sham" or "placebo" treatments.

- *Was the study population adequately described?* Participants selected for a study should be appropriate to the type of information sought. The population should adequately reflect such characteristics as the condition under study, a particular age group, or the general population. The makeup of the study population is also a major factor regarding the applicability of the findings to clinical practice. Essentially, study participants should be similar enough to commonly encountered patient populations for its findings to be generalizable.

- *Were the study's inclusion and exclusion criteria described and were they reasonable?* Patient selection in some studies is so restrictive that the results of the study may not be applicable to your patient base. For instance, many studies dealing with whiplash have excluded patients with Grade III injuries (those having neurologic signs)⁵, even though these types of patients are commonly seen by DCs. One would have to be cautious about extrapolating findings from such a study.

- *If random assignment to treatment/control groups was carried out, was it done correctly?* The method used to randomize patients

in a study should be described and should seem likely to produce equivalent groups. If done by coin-toss, sealed envelopes, random numbers table, or by computer, then randomization was probably appropriate. If it looks as if the sequence of assignment to groups could be guessed by either participants or study personnel, it is probably inappropriate. Methods used to balance the randomization, like stratification or blocking, should be described and justified.

- *If the study design featured randomization, was it carried out without bias and was the process well-described in the methods section of the article?* Sometimes, randomization results in treatment and control groups that are so different that the study goals cannot be met. When that occurs, there are ways of addressing the problem, following which the treatment and control groups are well-balanced, especially regarding the severity of the outcomes under investigation. The groups should also be equivalent with respect to medical histories and demographic information. Any differences noted between groups at the beginning of the trial should be considered as possible confounders, and must be acknowledged in the "limitations of the study" section that is included in virtually any well-written research publication. Another factor to consider is that RCTs in which the randomization was not well-described tend to show greater treatment benefits, suggesting that bias itself is far from random in clinical studies; in other words, lesser-quality studies are more likely biased in favor of the treatment under study⁶.

- *Were study personnel as well as participants blinded as to group assignment?* Although blinding

Philosophy & Technique

(some prefer the newer term “masking”) is difficult with chiropractic interventions, the researchers should make an effort and describe their methodology. Determine what efforts were made by the researchers to achieve blinding and also consider how much it matters in that particular case. Consider, however, that Schulz⁷ estimated a 17% exaggeration of treatment effect in studies that were not double-blinded. Frequently researchers evaluate how successful blinding was after the study has concluded by asking both clinicians and patients to what group they thought the patient was assigned. If too many patients or clinicians are able to guess the assignment correctly, then blinding was not really effective⁸. Moreover, Kunz and Oxman⁹ indicated that the correctness of a study’s blinding procedures may be a more sensitive means of appraising bias than checklists used to evaluate their quality.

- *Was the sample size large enough to find a treatment effect if one was present?* Higher-quality articles provide calculations of the number of participants needed in the study to have enough power to discern a treatment effect when actually present. Small samples increase the risk of a Type II error, in which a false hypothesis is not rejected. Accordingly, it is important for researchers to make sure they have included enough participants in the study. Unfortunately, this calculation is reported less than 25% of the time in chiropractic journals¹⁰ and less than 20% of the time in the spine surgical literature¹¹.

- *Were the groups treated identically, except for the intervention, throughout the study?* It is important that examinations, treatments, and follow-ups be carried out at the same time intervals; also, study

participants should receive the same amount of attention from the clinical personnel. Differences in the way participants are handled is termed “performance bias,” and such differences in care provided beyond the intervention under investigation may have altered the results.

- *Were all of the study participants accounted for at the end of the study?* The number of participants completing a study should equal the number starting the study, minus dropouts and those eliminated for some reason. When some participants are identified as “lost-to-follow-up,” their outcomes should still be analyzed based on the groups to which they were originally assigned. Similarly, *intention to treat analysis* specifies how to handle noncompliant patients in a randomized control trial. Excluding non-compliant patients from analysis may produce a study population that is healthier than those commonly seen in practice. As if all this were not enough, Kruse et al.¹² reviewed 100 articles where authors claimed to have performed an intention-to-treat analysis and found that 47% of the time they had not.

- *Are the results of the study statistically significant?* A reader needs to know that the appropriate statistical tests were utilized and were calculated correctly. Mistakes are quite common and very few people have the specialized knowledge to recognize them. Higher-quality journals are more likely to ensure quality peer review, so that a reader may have more confidence in the correctness of the statistical analysis. Readers must be aware that not all “peer-reviewed” journals warrant that confidence; the term “peer-reviewed” is easily abused, since the process of peer-review is

only as good as the peer-reviewers themselves.

- *Are the results of the study clinically significant?* Statistical significance is almost always addressed, but clinical significance is not. We need to know not only the amount of a treatment effect, but whether it is clinically significant. Competent researchers will define what differences would be considered clinically important at the outset of a study. If the minimum differences to attain clinical importance are not met, the results would not be considered important to clinicians, even if statistical significance was attained. For example, a few years ago, the coffee world was rocked by the revelation that heavy drinking of non-filtered coffee beverages such as espresso and French press, which contain more of a compound called cafestol, could raise serum cholesterol by about 5%¹³. Thus, 5 cups per day might raise cholesterol from 200 to 210. This sensational report was of virtually no clinical importance, but many people in the media and in the public had trouble understanding this.

- *Were the outcomes measured clinically important and relevant?* Outcomes that are clinically important are those that are of interest to patients. Essentially, patients want to feel less pain and function better. In contrast, they are not nearly as concerned about improvements in their orthopedic tests or the degree of palpation tenderness, nor the exact amount of their cervical lordosis. There can be no evidence-based chiropractic¹⁴ without an ethic of “patient-centered chiropractic” that is alive and well in this profession.

- *Who sponsored and/or funded the research?* Approximately 70% of the funding for clinical drug trials

Philosophy & Technique

is from commercial sources¹⁵ and consequential biased findings are often reported¹⁶. Chiropractic research is not immune to this type of bias, although different commercial interests are involved. For instance, a study that was sponsored or authored by someone closely connected to a chiropractic technique that is commercially promoted to the profession may have a tendency to be biased¹⁷. The same could be said about research sponsored by vendors of nutritional supplements or therapeutic/diagnostic devices.

- *Were the conclusions appropriate?* All conclusions presented in

articles should be consistent with and supported by the study's findings. Be cautious about authors who "put a spin" on their findings in an attempt to support their point of view¹⁸. Conclusions should also be in agreement with the study's predetermined purpose.

Conclusions

The previously mentioned appraisal questions can be used to evaluate just about any type of journal article. Alternatively, there are a number of checklists available that were designed to appraise specific types of research designs¹. For readers who would like to delve further

into this subject, the Centre for Health Evidence (CHE) at the University of Alberta (<http://www.cche.net>) provides a wealth of information to assist with the appraisal of virtually any type of health-related journal article. The CHE website presents the complete series of "Users' Guides to Evidence-Based Practice," a renowned series of articles published in the *Journal of the American Medical Association*. We would also recommend that readers visit the website of the Institute of Evidence-Based Chiropractic (www.chiroevidence.com). ■

References

1. Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care*, 2004. 16(1):9-18.
2. Feise RJ & Cooperstein R. From narrative to systematic reviews. *Jour Amer Chiropr Assoc*, 2006. 43(3):7-9.
3. Guyatt GH, Sackett DC & Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*, 1993. 270(21):2598-601.
4. Guyatt, G.H., Sackett DL & Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA*, 1994. 271(1):59-63.
5. Spitzer WO, et al. Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management. *Spine*, 1995. 20(8 Suppl): p. 1S-73S.
6. Cooperstein R. "Once a treatment has become respectable, it may be impossible to determine whether it actually works." *Dynamic Chiropractic*, 1996. 14(15):24-26.
7. Schulz KF, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*, 1995. 273(5):408-12.
8. Morschhauser E. Does the typical procedure for evaluating the success of patient masking assess expectation or outcome of treatment? A literature review and suggested new practice. *Journal of Chiropractic Education*, 2004. 18(1):21-22.
9. Kunz R. & Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *Brit Med J*, 1998. 317(7167):1185-1190.
10. Long DC, Nick TG, & Kao C. Original research published in the chiropractic literature: evaluation of the research report. *J Manipulative Physiol Ther*, 2004. 27(4):223-8.
11. Bailey CS, Fisher CG & Dvorak MF. Type II error in the spine surgical literature. *Spine*, 2004. 29(10):1146-9.
12. Kruse RL, et al. Intention-to-treat analysis: who is in? Who is out? *J Fam Pract*, 2002. 51(11):969-71.
13. De Roos B, et al. Consumption of French-press coffee raises cholesteryl ester transfer protein activity levels before LDL cholesterol in normolipidaemic subjects. *J Intern Med*, 2000. 248(3):211-6.
14. Feise R. The Evidence-Based Approach. *J Amer Chiropr Assoc*, 2002. 39(8):30-3.
15. Bodenheimer T. Uneasy Alliance -- Clinical Investigators and the Pharmaceutical Industry. *N Engl J Med*, 2000. 342(20):1539-1544.
16. Djulbegovic B, et al. The uncertainty principle and industry-sponsored research. *Lancet*, 2000. 356(9230):635-8.
17. Cooperstein R, et al. Flawed trials, flawed analysis: Why CBP should avoid rating itself. *J Amer Chiropr Assoc*, in press.
18. Cooperstein R. Ten theses on the tyranny of unfettered factoids. *Dynamic Chiropractic*, 1997. 15(20):6-7, 14.

Philosophy & Technique

Appendix

Checklist for the Appraisal of Journal Articles

- The purpose of the study was conveyed plainly and rationally. Yes NO
- A suitable literature review was presented that adequately covered the topic. Yes NO
- The research design was clearly described. Yes NO
- The research design was in concordance with the study question. Yes NO
- The research design was adequately implemented. Yes NO
- The target population was identified. Yes NO NA*
- A determination of the needed sample size was carried out. Yes NO NA
- The sample size was adequate. Yes NO NA
- Subjects were randomly assigned to groups. Yes NO NA
- Groups were equivalent. Yes NO NA
- Data collection methods were adequately described and were appropriate. Yes NO NA
- Demographic characteristics of the groups were presented. Yes NO NA
- The data were reported with sufficient detail. Yes NO NA
- Statistical tests were described and were appropriate for the type of data involved. Yes NO NA
- The discussion corresponded to and was supported by the data. Yes NO
- Discussion statements were reasonable and logical. Yes NO
- Conclusions were in agreement with the study's predetermined purpose. Yes NO
- Study limitations were listed. Yes NO
- Was there any evidence of a "fishing expedition"? Yes NO
- References were accurate and appropriately used. Yes NO
- References were from respected evidence sources. Yes NO
- Key points: _____
-

Overall impression of this article: Good Fair Poor
Can I apply the study's findings to my patient or practice? Yes Partly No

* Not Applicable