SPINE Volume 28, Number 15, pp 1631–1634

A Critical Guide to Case Series Reports

Timothy S. Carey, MD, MPH,* and Scott D. Boden, MD†

Objective. Provide guidance to investigators and authors regarding appropriate conduct and reporting of case-series studies.

Summary of Background Data. Evidence-based practice has provided a substantial contribution to advancing clinical science. Many study designs have been critically examined, and the quality of the research literature has improved. A common study design in musculoskeletal medicine is the case series: a description of the course of patients over time. Case series can provide valuable information as to: case definition, trend analyses regarding outcomes, and clues as to causation. Case series cannot be used to draw inferences regarding treatment effect.

Methods. Examination of previous work on identification of characteristics of high quality study designs such as cohort studies; extending this work to case series.

Results. We identified draft characteristics that good case series studies should address: clearly defined study question; well- described study population; well-described intervention; use of validated outcome measures; appropriate statistical analyses; well-described results; discussion/conclusions supported by the data presented; funding sources acknowledged.

Conclusions. We propose these measures to authors and journal editors as one mechanism to improve the quality of the case series study. [Key words: case series study, evidence-based medicine, clinical epidemiology, study design] **Spine 2003;28:1631–1634**

Evidence-based practice has provided a substantial contribution to advancing clinical science. The emphasis on properly conducted randomized controlled trials (RCT) and critical evaluation of the medical literature has substantially helped medical specialties, including orthopedic surgery and other musculoskeletal specialties. Many articles and texts discussing evidence-based practice have emphasized the importance of the randomized, placebo controlled trial in advancing clinical knowledge. Although many clinical questions are optimally addressed by the RCT, and this study design is still underused in addressing the efficacy and effectiveness of musculoskeletal interventions, the RCT is not appropriate for all clinical questions. The RCT design is particularly challenging when a surgical procedure is being studied. Yet, just as a surgical RCT can be conducted according to rigorous standards, other, more descriptive studies should be conducted with equal care.

The case series is one of a group of descriptive studies that by their very nature do not test the hypothesis of treatment efficacy.[1] That is, a case series is not the appropriate design to determine whether a treatment works or not. The case series belongs to a group of descriptive studies that includes: case reports (single case), case series, cross-sectional (prevalence) studies, surveillance studies, and ecological correlational studies. Descriptive studies have multiple uses. These include improved case definition, trend analyses regarding outcomes, health care planning including financial analyses, registry data regarding outcomes and complications used in "benchmarking" analyses, and clues about cause. When one detects a clue regarding cause from a descriptive study, it is called a hypothesis-generating study. That is, a follow-up study is then needed in order to test the hypothesis generated from the descriptive study. These descriptive study designs represent the basis on which other, hypothesis-testing analyses occur. Although the case series is often maligned as a mechanism for drawing inferences regarding the effectiveness of a given treatment, some medical and surgical treatments have been accepted as standard treatment only on the basis of the clinical experience of the case series: antibiotics for meningitis and pneumococcal pneumonia, surgery for appendicitis. However, such inferences can generally only be drawn when there is an enormous effect size of treatment— unfortunately, this is rare in medicine.

We define the case series as: a group of patients with similar diagnoses or undergoing the same procedure followed over time. The case series may be used in several settings:

- Initial reports of a new diagnosis or innovative treatment.
- Single physician or hospital reports of outcomes.
- Multi-institutional registry.

The size of a case series can range from two or three cases to hundreds or even thousands. An important distinction lies between a case series and a cohort study. Cohort studies are characterized by assembling patients with similar characteristics at a common point in their disease course and following them over time, using predefined measures of their outcomes: pain, functional status, employment, satisfaction with care, *etc.*

A major characteristic of a case series, and its major drawback, is the lack of a comparison group. If patients feel better (or return to work) after a procedure or other

From *The Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, and †Orthopaedics, Emory Clinic, Emory University School of Medicine, Atlanta, Georgia.
Acknowledgment date: October 2, 2002. First revision date: January 21, 2003. Acceptance date: January 24, 2003.
The manuscript submitted does not contain information about medical device(s)/drug(s).
No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address correspondence to Timothy S. Carey, MD, MPH, The Cecil G. Sheps Center for Health Services Research, 725 Airport Road, CB# 7590, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7590, USA; E-mail: carey@mail.schsr.unc.edu

treatment, how do we know whether this is the result of the efficacy of the procedure or the natural history of the disease under study? Of course, if the patients do not appear to benefit, or benefit minimally from the procedure, then the case series may provide a valuable service in that no randomized controlled trial or other more sophisticated study design is necessary; the procedure under study should not be widely disseminated.

Case series can also be valuable in describing the natural history of a condition or the recovery and complication rates after a treatment or procedure. Examination of consecutive cases can be useful as a "benchmark" so that other practices can compare their own complication rates (or length of stay, cost, *etc.*) with published results. Of course, a problem with the use of the case series for such a "benchmarking" function is that authors may submit for publication only those series of cases that have the best outcomes. Such publication bias is a common problem for all types of clinical research.

■ Characteristics of Well Done Case Series Articles

The Evidence-Based Practice Centers sponsored by the Agency for Healthcare Research and Quality (AHRQ) recently published a summary examining the quality characteristics of various study designs.[2] This comprehensive review examined observational studies as well as randomized trials. However, they did not examine the case series study design. The authors identified nine methods domains across which there was agreement among expert study design methodologists. For most of these methods domains, there was no empirical evidence that incorporating the quality measure into the study design led to more valid outcomes—these quality measures are generally formed from expert opinion. The two areas in which there does appear to be empiric evidence are comparability of subjects between intervention and control groups at the outset of a study and the lack of potential conflict of interest. Because case series studies by definition do not have a control group, comparability of subjects at the outset is problematic. However, below we adapted the observational study quality characteristics identified by West *et al* to the case series design. Our draft characteristics of a good case series are:

- Clearly defined question.
- Well-described study population.
- Well-described intervention.
- Use of validated outcome measures.
- Appropriate statistical analyses.
- Well-described results.
- Discussion/conclusions supported by data.
- Funding source acknowledged.

Below we will address each of these characteristics. We propose that authors, journal reviewers, and editors consider using the quality indicators in writing and evaluating case series studies.

■ Clearly Defined Question

All research, of course, should have a clearly defined question. Simply collecting data for the sake of collecting data rarely provides interesting results. For case series studies, the question should be appropriate to the study design. That is, the study question should not be couched in terms of determining whether Treatment A is better than Treatment B, or even whether Treatment A is an effective treatment for a particular disease. As noted above, there have been rare cases in which case series have been useful in identifying effective treatments. However, these have been few and far between and usually are diseases that have had extraordinarily high case mortality. This is rarely the case with musculoskeletal illness. Similarly, clear study questions are best when they are focused. This is true for all research, and especially true for descriptive studies. What is it about the natural history of the disease that the author wants to address? Health care utilization? Return to work? Functional status? All scientific writing is best done when addressing a focused question.

■ Well-Described Study Population

The case series should begin with a case definition. For example, if an investigator is examining a series of patients with spinal stenosis, what is the definition of spinal stenosis used? Specific clinical and/or radiologic criteria should be used so that readers can compare their patients with those described in the case series. Remarkably, many medical and surgical problems do not have standardized case definitions. When the definition is not standardized, making the definition explicit will allow the reader to compare the cases described in the article with the cases in his or her own practice. Are the criteria used clinical? Radiographic? Have they been used before? Is a reference included? Similar to other studies, the case definition should involve explicit inclusion and exclusion criteria. The standard descriptive information should include age, gender, socioeconomic status (income, insurance status, *etc.*), and for musculoskeletal problems, whether the patients are covered by worker's compensation or not. Important descriptive information should include the stage of disease or illness, especially its duration. Case series derived from tertiary care centers tend to have patients with a greater severity, more advanced illness, and greater comorbidity. This descriptive information can allow physicians in community practice to determine whether the patients evaluated are similar to the patients they see in their own practice. The number of patients evaluated should be described. How many years did it take to arrive at the sample size? Usually case series include all the patients the investigator could find, but if patients were excluded, the reader needs to know how many and for what reason. Otherwise, the researcher might select only the patients with the most favorable prognosis or the best outcomes. Follow-up of the patients should be consistent, with description of why pa-

tients were lost to follow-up, refused the study intervention, *etc.* Patients may be lost to follow-up if they die, are dissatisfied with care, or if they are feeling better and do not perceive a need to return to a physician. This is an important distinction between a case series and a cohort study in that the case series is often not identified in advance as a study population. This type of loss to follow-up information is just as important in a case series as with cohort studies or randomized trials.

■ Well-Described Intervention

When a medication intervention or surgical procedure is performed, there should be a clear description of the procedure or treatment. Ideally, this description should be sufficiently clear that another center could replicate the study. If the procedure is not identified in detail, references to other articles should be provided. An issue frequently omitted from case series is description of cointerventions. Too often, postoperative patients are described as having received "standard postoperative rehabilitative care." Such care may vary substantially from one area of the country to another (let alone between the U.S. and Europe). Therefore, cointerventions such as physical therapy, medication therapy, work rehabilitation, *etc.* should be described in reasonable detail. Such interventions may have a significant specific effect in improving patient outcome, which may substantially augment the effect of a surgical procedure. An example would be a lumbar fusion for degenerative disc disease with the cointervention of a brace or physical therapy. Either of these cointerventions could in and of itself result in an improved patient outcome.

■ Use of Validated Outcome Measures

How was improvement detected in the case series? The field of musculoskeletal medicine is fortunate in that there are numerous valid outcome measures available. However, the number of available outcome measures greatly exceeds the number of valid outcome measures. Previous use of the measure is not the same as appropriate validation against generally accepted gold standard measures. Some sort of reference to previous validation should be made in the article. Fortunately, in the care of patients with spinal problems, multiple validated instruments are available, such as the Oswestry scale or the Roland-Morris disability scale. The individual assessing the patient's outcome should ideally be masked to the specific intervention. That is, the assessor should not know whether the patient has received the treatment or not. In the case of some surgical interventions, this may be impossible. However, the use of a research assistant who is not in the direct employ of the clinical office may be a reasonable second best. The length of observation and the intervals between clinical observations should be standardized and of sufficient duration to be clinically meaningful. For example, patients may have a functional decline immediately after an operative procedure and then improve. After injection treatment, patients may feel better following a treatment but the duration of improvement is sometimes relatively brief. Some justification as to the duration of follow-up should be provided.

■ Appropriate Statistical Analyses

Case series are quite different from most study designs in that statistical tests yielding *P* values or confidence intervals are not needed and in most cases are inappropriate. However, some types of analyses may be useful. For example, a lack of improvement over time (or an improvement followed by return to baseline) may be quite important. Statistical tests and power calculations may be helpful when addressing a lack of improvement. Case series sometimes use historical controls, comparing the outcomes of patients to those from previous case series. Because cointerventions such as early mobilization, active physical therapy, *etc.* have become more common, functional status tends to improve much more rapidly following procedures than it did 10 to 20 years ago. Therefore, use of historical controls is often misleading.

Another type of control is the prepost analysis, with the patient serving as his or her own control. When statistical tests are used, the paired nature of the data (the patient's predata are compared with the same patient's postdata) should be taken into account. Though these analyses may sometimes be useful, investigators and readers should recognize that there exists a powerful nonspecific effect of procedures.[3] Patients who have been followed for a prolonged period of time without benefit are often looking for a new treatment and may derive a significant perceived benefit from a novel procedure. In addition, many chronic painful illnesses tend to wax and wane. Surgical procedures tend to be offered when symptoms are at a crescendo, and a case series cannot differentiate spontaneous improvement from a specific benefit from the procedure. Finally, authors sometimes assess many clinical and biologic variables and seize on one or two that demonstrate an improvement. Any statistical testing, recognizing that such testing is generally inappropriate in case series, should address the fact that multiple comparisons are made.

■ Well-Described Results

As noted above, the case series should utilize only validated outcome measures. In addition, adequacy of follow-up should be described. This includes: number of patients who are lost to follow-up, number of patients who switch to another provider or pursue other treatment choices, and number of patients who die from other causes. These outcomes should be consistently measured. As an example, a case series of patients treated for spinal stenosis should indicate the number of patients evaluated with spinal stenosis in the practice, the proportion operated on, the number of patients lost to follow-up, and the reasons for loss to follow-up (death, move out of state, transfer to another provider, unable to contact, *etc.*).

Table 1. Case Series Characteristics

Characteristic	Davis and Onik[6]	Saal and Saal[8]
Study question	+	+
Study population	−	+
Comparability of subjects	−	−
Well-described intervention	+	+
Valid outcome measure	−	+
Appropriate statistical analysis	?	+
Well-described results	−	+
Appropriate discussion	−	+
Funding issues	?	?

+ = characteristic present; − = characteristic absent;
? = unclear whether present or absent.

■ Discussion/Conclusions Supported by the Data

The issues here are similar to those in any research study. The conclusion should be supported by the data in the article. Where other information is used to buttress these conclusions, this should be explicitly stated and referenced. Similarly, limitations (and there are always limitations) should be made explicit. Often most helpful in the discussion is a description of next research or clinical steps in furthering this field. If the investigator thinks that the procedure or treatment is ready for a randomized trial, he or she should say so and possibly describe what that trial might look like. Bland assertions that "more research is needed" add little to any article; the more specific the recommendations for the next steps in the research, the better.

■ Funding/Sponsorship

Empiric evidence suggests that positive studies demonstrating improvement are more likely with certain types of funding of the study. That is, studies that have received private or industry funding are more likely to demonstrate a positive result than studies with nonprofit or government funding.[4] Reasons for this phenomenon are likely multiple. They may include multiple publications and especially publication bias. Publication bias in no way implies scientific misconduct, but rather is the phenomenon of authors submitting for publication studies that appear to support the efficacy of a treatment and not submitting studies that do not support the efficacy of a treatment.[5] Even when the case series itself is not externally funded, if the author has a consulting or board relationship with a device or pharmaceutical manufacturer that might be perceived as a potential conflict of interest, it is best to fully disclose it at the time of submission. Journals should be strongly encouraged in requiring authors to explicitly state sources of funding for all articles.

An Example.

For illustrative purposes, we compared two case series articles that examined the outcomes of patients undergoing percutaneous procedures for low back pain (Table 1). We critically read each study and evaluated its adherence to the above quality measures. The study by Davis and Onik examined percutaneous discectomy, a procedure that a later randomized clinical trial by Chatterjee *et al* demonstrated to have little utility.[6,7] Several parts of the study description would have benefited from further elaboration. The study population was tersely described as "patients who had failed the usual conservative care. . ." The outcome measure was described as "moderate to complete pain relief" with little information on how these data were gathered or who did the measuring. Although the reader is told that "longer follow-up studies are needed. . .," the discussion implies efficacy of the treatment. In contrast, the recent case series by Saal and Saal examines a case series of intradiscal electrothermal therapy.[8] The study population appears to be highly selected (62 out of 1116 patients presenting), and we are provided with limited baseline information on SF-36. Acknowledgment of the need for comparative studies is made. In neither study were sponsorship or conflict of interest issues clarified; the Saal and Saal study indicates that it belongs to "conflict of interest category 12."

■ Conclusions

Although often maligned as a study design, the case series can provide an important initial indication of treatment efficacy or the description of a novel clinical problem or complication. The impact of a case series report can be strengthened through adhering to standard design and reporting guidelines as illustrated above. Case series may be small or large, sometimes even in the hundreds or thousands of patients. All research, including case series, is time and labor intensive. The case series is worth doing in some circumstances, and therefore worth doing well.

■ Key Points

- The case series articles design is commonly used.
- Case series should not be used to assess treatment efficacy.
- Draft quality characteristics of quality case series reports are presented.

References

1. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet* 2002;359:145–9.
2. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. File Inventory, Evidence Report/Technology Assessment Number 47. AHRQ Publication No. 02-E0106, April 2002. Agency for Healthcare Research and Quality, Rockville, MD. Available at http://www.ahrq.gov/clinic/strevinv.htm.
3. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81–8.
4. Djulbegovic B, Lacevic M, Cantor A, et al. The uncertainty principle and industry-sponsored research. *Lancet* 2000;356:635–8.
5. Montori V, Guyatt G. Summarizing the evidence: publication bias. In: User's Guide to the Medical Literature: A Manual for Evidence-based Clinical Practice. AMA Press, Chicago, IL, 2002:530–538.
6. Davis GW, Onik G. Clinical experience with automated percutaneous lumbar discectomy. *Clin Orthop* 1989;238:98–103.
7. Chatterjee, Fory PM, Findlay GF. Report of a controlled clinical trial comparing automated percutaneous lumbar discectomy and microdiscectomy in the treatment of contained lumbar disc herniation. *Spine* 1995;20:739–42.
8. Saal JA, Saal JS. Intradiscal electrothermal treatment for chronic discogenic low back pain. *Spine* 2000;25:2622–7.