Assessing the Validity of Published Randomized Controlled Trials in Podiatric Medical Journals

Michael A. Turlik, DPM*
Donald Kushner, DPM*†
Dina Stock, DPM†

The purposes of this study were to develop an instrument to assess the validity of randomized controlled trials and to report on the differences in the validity of randomized controlled trials between two podiatric medical journals and a mainstream medical journal. The study demonstrated that after adequate training, there can be agreement among reviewers evaluating the quality of published randomized controlled trials using an established instrument and guidelines. The results of the study indicate that randomized controlled trials published in podiatric medical journals are less credible than those published in a mainstream medical journal. (J Am Podiat Med Assoc 93(5): 392-398, 2003)

The amount and sources of information regarding therapeutic options received by podiatric physicians on a regular basis is staggering. Practicing podiatric physicians attend continuing medical education seminars; read publications; review textbooks; speak with colleagues, students, and patients; and encounter drug and equipment sales personnel regularly. Attempting to determine which source of information is best relied on in formulating a personal treatment paradigm is often confusing and frustrating.

Evidence-based medicine attempts to make sense of the wide array of information available by categorizing the sources of information into a hierarchy from most to least useful (Fig. 1). Although many types of publications contain useful information depending on the type of question to be answered (prognosis, diagnosis, harm, etc), practitioners of evidence-based medicine advise that the most useful sources of information on which to base therapeutic decisions are definitive randomized controlled trials (RCTs), with the least useful being expert opinion.1, 2 Properly planned and executed RCTs minimize bias. Bias can be defined as systematic deviation from the truth. Bias is most often unintentional, but it can be intentional.

A recent review of the podiatric medical literature3 revealed that most published articles in podiatric medical journals are case studies or case series. Although a few RCTs were found, no meta-analyses were documented. That study sought to quantify the types of publications, but no attempts were made to assess the quality of the publications.

The quality of published RCTs varies from well-performed, definitive trials to trivial, underpowered studies with serious methodologic flaws. In an attempt to improve the quality of published RCTs, a group of journal editors, clinical epidemiologists, and statisticians met in 1995 to formulate reporting guidelines.4 The result of their efforts was the CONSORT (Consolidated Standards of Reporting Trials) statement, a checklist and flow diagram that could be used by journal editors and clinical researchers to standardize and improve the reporting of RCTs. A 2001 issue of the Journal of the American Podiatric Medical Association (JAPMA) featured an article on the CONSORT statement for the podiatric audience.5

*Ohio College of Podiatric Medicine, Cleveland.
†Ohio College of Podiatric Medicine, Cleveland.
Corresponding author: Michael A. Turlik, DPM, Beachwood Foot and Ankle Center, 3690 Orange Pl Dr, Ste 415, Beachwood, OH 44122.
It is important for the practicing podiatric physician to evaluate the quality of an RCT before considering implementing its conclusions. The two most important questions for the reader to answer about an RCT are 1) Is the article relevant to my practice? and 2) Are the results of the study valid?

Relevance

Assessing the relevance of an article varies with the reader’s training, experience, location, and type of practice. An important issue to resolve when considering relevance is whether the results of the study can be generalized to one’s own patient population. If not, then the article may not be worth reading. Another important issue is whether the therapeutic interventions described are common and feasible in one’s practice. If this seems unlikely after reading the title and abstract, it is not necessary to read the article further to assess its validity.

Validity

If after reading the title and abstract of the article the reader believes that the article is relevant to his or her own practice setting, the next step is to determine whether the conclusions of the article are valid. How close to the truth are the results?

A major problem in assessing trial quality is that the reader must rely on written information on the study. A well-designed but poorly reported trial may be viewed as lacking validity and therefore not credible.

The purpose of this study was to develop and evaluate the reliability of an instrument used to assess the validity of RCTs. In addition, using the new instrument, this study sought to evaluate the validity of RCTs published in podiatric medical journals versus a mainstream medical journal that has adopted the CONSORT statement.

Materials and Methods

There are several examples of instruments used to evaluate the validity of published RCTs. After reviewing the literature, the senior author (M.A.T.) developed such an instrument, which was used at the Ohio College of Podiatric Medicine, Cleveland, in the instruction of senior podiatric medical students during the 1999–2000 academic year.

The authors modified the instrument used at the Ohio College of Podiatric Medicine (Fig. 2) and developed guidelines for its use (Appendix). After several training sessions, interrater agreement reached an acceptable level. A kappa (κ) reliability coefficient with 95% confidence intervals (CIs) was calculated as a measure of agreement, with statistical significance set at $P < .05$.

A κ coefficient of at least 0.7 was accepted as agreement between reviewers. Beginning with the December 2000 issue and working backward, two of the authors (D.K. and D.S.) reviewed 10 years of JAPMA and the Journal of Foot and Ankle Surgery (JFAS). After reading the title and abstract of each article, the reviewers selected all of the articles that the authors described as RCTs. Only parallel-group design studies were included. Crossover studies and matched design studies were not included. The number of RCTs found in these journals during this period limited the number of RCTs used in the study. The same procedure was used to select an equal number of RCTs from the Journal of the American Medical Association (JAMA). The corresponding number of articles was found in the issues published in the first 4 months of 2001.

The clinicians then met again as a group and reviewed all of the articles using the worksheet and the guidelines. A scoring consensus was reached for each article using a modified nominal group process. The completed worksheets were entered into a database for analysis, and a $\chi^2$ test was performed; statistical significance was set at $P < .05$ between the podiatric medical journals and JAMA using SPSS version 9.0 (SPSS Science, Chicago, Illinois).

Results

Comparison of the κ values for the independently reviewed questions regarding validity (N = 162) reveals that point estimates for κ values ranged from 0.80 to
0.86. All of the \( \kappa \) values were statistically significant \((P < .001)\), and the lowest 95% CI value was 0.70 (Turlik versus Stock). All of the reported values were greater than 0.70, the predetermined value for agreement between reviewers decided on by the authors at the start of the study (Table 1).

Nine articles from JAMA and nine articles from JFAS (two) and JAPMA (seven) were reviewed. Each article was judged on nine separate categories of validity, for a total of 81 separate measurements on each set of articles (Table 2).

Neither group of articles consistently explained the concept of concealment allocation (question 2) effectively. None of the podiatric medical journal articles demonstrated any acceptable explanation of concealment allocation, and only three of the nine articles from JAMA adequately described concealment allocation.

The podiatric medical journal articles scored well on only two questions: on question 3 (comparability of groups at baseline), five of the nine articles had satisfactory explanations, and on question 8 (appropriateness of statistical methods), seven of the nine articles had satisfactory explanations.

The articles from JAMA scored well on all of the questions except question 2 (concealment allocation). On four questions, the articles from the mainstream medical journal received perfect scores by the reviewers: questions 3 (comparability of groups), 4 (all patients accounted for at end of study), 6 (validated outcome instrument), and 8 (appropriateness of statistical methods).

Podiatric medical journal articles provided satisfactory explanations for 26 of the 81 potential questions, whereas the articles from JAMA had adequate explanations for 67 of the 81 possible questions. Comparison of the questions from the worksheets reviewed by the group \((N = 162)\) revealed that the podiatric medical journals \((n = 81)\) differed significantly from JAMA \((n = 81)\) \((df = 1; \chi^2 = 42.4; P < .001)\).

<table>
<thead>
<tr>
<th>Validation Topic</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the randomization explained?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Was the concealment allocation explained?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Were the groups comparable at baseline?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Were all of the patients accounted for at the end of the study?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Were the assessors of outcomes blinded?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Was the outcome instrument used to collect the data validated?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Was statistical analysis of sample size performed before the experiment?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Was the statistical analysis of the main outcome variable compatible with the type of data collected and their distribution?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Was the 95% confidence interval calculated for the point estimate of the main outcome?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.** Worksheet used to evaluate the validity of published randomized controlled trials.

<table>
<thead>
<tr>
<th>Table 1. Independent Measures of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewers No.</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Kushner versus Stock</td>
</tr>
<tr>
<td>Kushner versus Turlik</td>
</tr>
<tr>
<td>Turlik versus Stock</td>
</tr>
</tbody>
</table>
Discussion

The results of this study indicate that RCTs published in podiatric medical journals are less credible than those published in a mainstream medical journal.

Only two of the podiatric medical journal articles\textsuperscript{13, 14} contained a description of randomization consistent with the published study guidelines. Nonrandomized trials have been shown to generate a treatment bias\textsuperscript{28}; therefore, caution should be used when applying the results of nonrandomized trials to practice.

Neither group of articles consistently contained explanations of concealment allocation that were judged to be acceptable by the authors. A lack of concealment allocation may lead to intentional or, more likely, unintentional selection of patients by the investigators for inclusion in certain treatment arms, thereby subverting the randomization sequence and introducing selection bias into the study.\textsuperscript{29}

Four of the podiatric medical journal articles\textsuperscript{13, 15-17} had less than 10% of the randomized patients drop out of their studies. None of the podiatric medical journal articles had an adequate explanation as to the handling of missing patients. It is inevitable in most clinical trials, as in real life, that some participants may not finish the entire course of therapy.

How and whether the author accounts for the patients lost to follow-up is an important matter when considering the validity of a pragmatic study. Participants who withdraw or are noncompliant with the study often have a different prognosis than those who remain to the end.\textsuperscript{30} Simply ignoring dropouts and missing responses is unsatisfactory because it increases ascertainment bias, destroys the randomization sequence, and undermines the validity of the clinical research.

Only two of the podiatric medical journal articles\textsuperscript{13, 14} had adequate explanations of blinding of assessors of outcomes. On average, randomized trials that have not used appropriate levels of blinding show larger treatment effects than blinded studies (11% to 17% bias).\textsuperscript{31} These studies have shown that assessments under masked conditions were more likely to yield lower and more consistent scores than assessments under open conditions. When blinding of participants and investigators is impractical, blinding those who assess the clinical outcomes is usually necessary to minimize bias. Blinding measurement of outcomes may be more critical than blinding the administration of the therapeutic intervention to prevent assessment bias, especially when the outcome measures involve subjective measurements.

Four of the podiatric medical journal articles\textsuperscript{10-12, 18} were judged to have an adequate description of the validity of the instrument used to measure a primary outcome. Two of the studies measured hard outcomes of care, and the remaining two studies focused on soft outcomes. When measuring subjective outcomes of care, a validated instrument is preferable to limit assessment bias and provide meaningful information on which to base therapeutic decisions.\textsuperscript{32}

Only one podiatric medical journal article\textsuperscript{16} was judged to have an adequate explanation of how the sample size was calculated before the experiment. Small sample sizes may produce a type II error or, in studies that are statistically significant, may yield CIs that are so large as to be meaningless therapeutically. Enrolling more patients in the study than is necessary is costly and unproductive. The authors of an RCT should enroll enough patients to have a high probability of detecting a statistically significant difference and a clinically important treatment effect, if one exists.

Only one podiatric medical journal article\textsuperscript{16} presented the results of the study in terms of 95% CIs. The present authors believed that the range of the CIs in this study was extremely large. In addition, the present authors did not know how to interpret the end points because no clinically significant estimate was given despite a reference to a power calculation. Even if the podiatric medical journal articles were not methodologically flawed, none of the trials could be judged as definitive.

The present study was consistent with an earlier study\textsuperscript{3} on the frequency of RCTs authored by podiatric physicians in podiatric medical journals. A total of nine RCTs were found in two podiatric medical journals for a 10-year period, and nine RCTs were found in the first four issues of 2001 of JAMA. The authors did not search other publications in which

\begin{table}
\centering
\caption{Composite Evaluation of Validation Worksheets}
\label{tab:composite_evaluation}
\begin{tabular}{|c|c|c|c|c|}
\hline
Validation Question & Podiatric Medical Journal Articles & Mainstream Medical Journal Articles & & \\
& Yes & No & Yes & No & \\
\hline
1 & 2 & 7 & 5 & 4 & \\
2 & 0 & 9 & 3 & 6 & \\
3 & 5 & 4 & 9 & 0 & \\
4 & 4 & 5 & 9 & 0 & \\
5 & 2 & 7 & 8 & 1 & \\
6 & 4 & 5 & 9 & 0 & \\
7 & 1 & 8 & 8 & 1 & \\
8 & 7 & 2 & 9 & 0 & \\
9 & 1 & 8 & 7 & 2 & \\
\hline
Total & 26 & 55 & 67 & 14 & \\
\hline
\end{tabular}
\end{table}
The quantity of RCTs published by podiatric physicians in podiatric medical journals is limited, and the quality is lower than in mainstream medicine. It is important for podiatric physicians to produce pragmatic, valid, and definitive RCTs to demonstrate the effectiveness and safety of therapeutic interventions unique to podiatric medicine. The authors believe that if this recommendation is followed, the standing of podiatry in mainstream medicine will improve.

Conclusion

The instrument developed by the authors to measure the validity of RCTs reflects their personal views on the subject. They tended to limit the numbered items on the form to those they believed were the most important for assessing the validity of published reports. Although there was agreement between reviewers, inclusion of other aspects of validity in an evaluation worksheet may have altered the results of the study.

The RCTs were evaluated on the basis of information supplied by the authors in their articles. A well-planned and well-executed but poorly reported study would be judged as being not valid. This is an additional limitation of the present study.

Appendix

Guidelines Used to Evaluate the Validity of Published Randomized Controlled Trials

Investigators should not only minimize or eliminate bias but should also communicate those efforts to the readers. Readers should not have to assume or guess at the methods used to produce high-quality clinical studies. The reviewers of the articles will answer "yes" on the validation worksheet if the author describes in enough detail, in the Materials and Methods section or the Results section, convincing evidence that adequate measures were taken to ensure that the validation topic was addressed. The reviewers of the articles will answer "no" on the validation worksheet if no description of the validation topic is given or if the description given is not credible.

1. Randomization. It is acceptable if the authors state the method of randomization used to generate the allocation sequence. The most likely method involves a random number list generated by a computer or use of a table of random numbers. It is unacceptable not to state the method of randomization used or to describe a method that is consistent with haphazard allocation.

2. Concealment Allocation. To be considered appropriate, the authors must describe the method used to separate the generator of the sequence from the executor of the experiment. This usually involves the use of numbered, opaque, sealed envelopes used by the generator to communicate with the clinician how the participants are to be randomized at entry. In addition, it should be noted where the allocation code was kept and when the sequence was broken. It is inappropriate not to mention the method by which the executors of the experiment were separated from the generator of the sequence.

3. Baseline Differences Between Treatment Arms. It should be considered appropriate if the author describes, in a table or adequately in the narrative, the demographic characteristics of the treatment groups. The authors should comment on the comparability between the two groups and, if different, describe the adjustments made before evaluating the data. It should be considered inappropriate and unsatisfactory if the author does not present enough material for the reader to evaluate the demographic differences between the treatment groups.

4. Dropout Rate. It is acceptable provided the authors state that the data were analyzed on an intention-to-treat basis. In addition, if dropout rates are greater than 10%, the author must attempt to analyze the data with and without corrections for the missing participants and present the results. Many methods are acceptable (worst-case scenario, etc). Dropout rates greater than 10% that are not adjusted are unacceptable.

5. Blinding of Outcome Assessors. If the study is not a double-blind study or it lacks convincing evidence that the method of blinding was appropriate, the authors should clearly state the method by which the assessors of the outcome were blinded to the treatment of the randomized groups. It is unsatisfactory if the people assessing the main outcome are aware of the type of treatment the patients received.

6. Validation of Outcome Instruments. The instrument used to measure and collect the main outcome data should provide valid and reliable measures. For hard outcomes such as blood pressure, body weight, and serum glucose levels, it would be desirable for the author to use a standardized instrument. Measurements of variables that involve evaluation of an articulation, palpation of pedal pulses, etc) should detail the method used and information that would convince the reader that there was agreement.
between evaluators. For measurement of softer outcomes such as pain and function, a validated instrument is preferred. The authors should clearly state that the instruments used to collect the data for soft outcomes have been validated, and they should provide appropriate references. It is inappropriate if the instruments used to collect the data have not been validated and references have not been supplied.

7. Sample Size Calculation. The authors should state the method and results of the analysis used to calculate the sample size. Statistical significance should be set at less than .05, and power should be fixed at 80% or more. The author should then enroll an appropriate number of subjects in the study, as determined by the power calculation. It is inappropriate merely to present calculated sample size. Statistical significance needs to clearly state why these tests have been selected. It is inappropriate for the authors not to discuss the statistical test used and the rationale for its use.

8. Analysis of Data. The authors should state the statistical tests used to analyze the data and whether underlying assumptions for the tests have been met. If obscure statistical tests have been used, the author needs to clearly state why these tests have been selected. It is inappropriate for the authors not to discuss the statistical test used and the rationale for its use.

9. 95% Confidence Interval. The authors should calculate 95% confidence intervals for point estimates for the main outcome measure. It is inappropriate merely to present P values and point estimates.

References


