

# How to Interpret the Results of a Randomized Controlled Trial

by Michael Turlik, DPM<sup>1</sup> 

The Foot and Ankle Online Journal 2 (4): 4

*This is the third in a series of articles explaining basic concepts of evidence-based medicine to podiatric physicians. This article will explain how to evaluate the results section of a randomized controlled trial. This will be illustrated using actual examples from the foot and ankle literature to include an article which was referenced in the first article of this series.*

**Key words:** Evidence-based medicine, randomized controlled trial

Accepted: March, 2009

Published: April, 2009

This is an Open Access article distributed under the terms of the Creative Commons Attribution License. It permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©The Foot and Ankle Online Journal ([www.faoj.org](http://www.faoj.org))

The preceding article in the series dealt with the evaluation of internal validity of a randomized controlled trial (RCT).<sup>1</sup> The first step in critically analyzing a randomized controlled trial is to assess its internal validity. Once you feel comfortable that the authors have taken sufficient methods to minimize bias you will need to determine what the results of the study are. The authors should present the results of the study so that the reader can *easily* determine whether the intervention is effective, independent of the author's interpretations. This is the third in a series of articles introducing podiatric physicians to evidence-based medicine (EBM).

## Primary Outcome

The authors should specify a *single* primary outcome which is patient oriented and was chosen *a priori* to determine effectiveness/efficacy of the intervention. It has been reported that trial outcomes may be selectively reported post hoc<sup>2</sup>, this is considered outcome reporting bias. One of the purposes of a trial registration is to prevent post hoc changes in the primary outcome<sup>3</sup> and minimize outcome reporting bias.

Hard outcomes such as ulcer closure or amputation are less common in the podiatric literature than soft outcomes such as functional improvement or pain relief. Hard outcomes usually occur or they do not occur and are considered binary or dichotomous data. Soft outcomes usually are treated as continuous measures even though they may be categorical. When measuring soft outcomes the authors should provide a reference which validates the instrument used and provide a clinically important difference.

---

**Address correspondence to:** Michael Turlik, DPM  
Email: [mat@evidencebasedpodiatricmedicine.com](mailto:mat@evidencebasedpodiatricmedicine.com)

<sup>1</sup> Private practice, Macedonia, Ohio.

Outcome reported as mean with (95% CI)	Prefabricated  vs.  sham	Custom  vs.  sham	Prefabricated  vs.  custom
<b>Pain at three months</b>	8.7 (-0.1-17.6)	7.4 (-1.4-16.2)	1.3 (-7.6-10.2)

**Table 1** Comparison of foot orthotics.

In one review of orthopedic publications, the authors found that 31% of the publications modified a validated outcome instrument and then used the instrument without re-validating it.<sup>4</sup> Patient reported outcomes which are not validated may produce data which is biased (measurement bias). The primary outcome should be used in the calculation of the sample size to insure adequate power to detect a clinically important difference. Articles which report multiple primary outcomes should be viewed with caution because, of the increased incidence of type I errors due to multiple hypothesis testing<sup>5</sup> and lack of power. Surrogate outcomes<sup>6</sup>, physician centered outcomes<sup>7</sup> and composite outcomes<sup>8</sup> should be considered with care. The practice of EBM stresses patient centered outcomes.

Secondary outcomes should be reported in the same manner as the primary outcome. They help to explain in a positive trial the intervention used. If the primary outcome is not statistically and clinically significant then the secondary outcomes are best utilized to develop further research questions. Subgroup analysis should be viewed with caution due to type I errors and should be considered hypothesis generating rather than hypothesis confirming.<sup>9</sup>

### Magnitude and precision of treatment effect

The purpose of reviewing the results section of a RCT is to determine if the difference between the intervention and comparator is due to chance and if statistically significant is the result clinically important.

Large studies may produce statistically significant results which are so small as to be clinically insignificant. The authors should report for *any* outcome a magnitude of effect with 95% confidence limits.<sup>10</sup> In the case of binary outcomes results should be presented as a 2 X 2 table, risks or (less desirable) odds. Using the information the reader can calculate the number needed to treat (NNT) with 95% CI if not reported. For continuous measures the author should report the difference between the intervention and comparator as a point estimate with 95% confidence intervals. This data is best presented in a table rather than text. When the authors present data other than described above it will be difficult to understand and place in a clinical context. Commonly a chi-squared test will be used to assess statistical significance for binary outcomes. For continuous measures common significance tests are *t*-test, ANOVA or ANCOVA.<sup>11</sup> If unusual statistical tests are employed the author should provide appropriate references for review.

### Evaluation of soft outcome

Landorf<sup>12</sup> in assessing the effectiveness of three different types of foot orthotic devices in the treatment of mechanically induced heel pain chose The Foot Health Status Questionnaire for the primary outcome. The Foot Health Status Questionnaire is a validated instrument which is patient oriented and measures both pain and function. The outcome was selected a priori and analyzed using ANCOVA. The results of the primary outcome were reported in a table of the article, a portion of which are summarized. (Table 1)

Intervention	Resolved	Not resolved	Event Rate
Control	36	60	62.50%
Graftskin	63	49	43.75%

**Table 2** Comparison of ulcer resolution.

Control Event Rate (CER) = 62.5%,  
 Experimental Event Rate (EER) = 43.75%,  
 CER-EER = Absolute Risk Reduction (ARR) = 18.75,  
 Number Needed to Treat (NNT) =  $1/ARR = 6$  (3-19)

The largest difference in pain at three months (Table 1) was shown between the prefabricated and sham device, the smallest difference was shown between the prefabricated and custom device. During the calculation for the sample size a 15 point improvement in pain was used. If we accept this as being the clinically significant improvement the authors were looking for none of the point estimates were clinically significant. In addition, since the lower end of the 95% confidence interval was below zero for all of the comparisons none of the results were statistically significant either. A question which should always be asked in a negative trial is, was there potentially a clinically significant outcome possible if more participants were enrolled in the study? This is evaluated by looking at the upper end of the 95% confidence interval to see if it is at or above a clinically significant level, in this case 15 points. If it is it indicates that if more participants were enrolled in the study this difference may become apparent. Since the upper end of the prefabricated versus custom foot orthotic device is below 15 points we can say with confidence that there is no difference between prefabricated and custom foot orthotic devices in the treatment of mechanically induced heel pain based upon the results of this study.

### Evaluation of hard outcome

Veves<sup>13</sup> evaluated the use of Graftskin in the healing of diabetic non-infected neuropathic foot ulcers. The primary outcome was complete wound closure at 12 weeks.

The authors reported in the text that 63 of the 112 patients receiving graft skin healed by 12 weeks whereas, 36 of the 96 control patients healed by 12 weeks. They related that this was a statistically significant result ( $p = 0.0042$ ). Using the data provided in the article a 2 x 2 table could be constructed. (Table 2)

Using an online calculator<sup>14</sup> and the information from the article the Number Needed to Treat (NNT) with 95% confidence intervals can be calculated. NNT is defined as the number of patients you need to treat to prevent one bad outcome.<sup>15</sup> NNT allows the results of the study to be placed in a clinician friendly metric which can be evaluated easily. The lower the number is the more effective the treatment. The number needed to treat is the inverse of absolute risk reduction (ARR). The absolute risk reduction equals control event rate (CER) minus experimental event rate (EER). Based upon the results (Table 2) of the study 6 people will need to be treated with Graftskin for 12 weeks in order to prevent one additional ulcer from failing to heal. The reader will need to decide if the costs and adverse effects justify the use of this product when compared to the control (external validity). Understand that this is only an estimate and that the number needed to treat may be as high as 19.

### Magnets in the Treatment of Diabetic Neuropathy

In the first article of this series<sup>16</sup> a RCT<sup>17</sup> was found which evaluated the usefulness of static magnets for painful diabetic neuropathy. The second article<sup>1</sup> in the series evaluated the internal validity of this RCT.

Using the information previously presented in this article the results of the article evaluating static magnets for the relief of painful diabetic neuropathy are summarized:

### Primary outcome measure

The authors described *four* different primary outcome measures using *two* different outcome instruments: foot pain, sleep (Quality of Life [QOL]), burning pain, numbness and tingling (11 point Visual Analogue Scale [VAS]) which were analyzed *once per month for four* months. It wasn't clear if they provided references related to validation of the outcome instruments. The authors describe a fairly complex method of statistical analysis of the extensive amount of data collected. The sample size calculation utilized a 17% difference between treatment arms for the calculation. It was unclear which of the primary outcomes were used for the 17% difference. Therefore, it is difficult to tell which of the outcome measures were adequately powered. It does not appear the trial was registered.

### Magnitude and precision of treatment effect

The design of the study was a randomized controlled trial using static magnetic insoles compared with a placebo. The objective of the study as stated by the author's was "to determine if constant wearing of multipolar, static magnetic (450G) shoe insoles can reduce neuropathic pain and quality of life (QOL) scores in symptomatic diabetic peripheral neuropathy (DPN)".<sup>17</sup>

The author's conclusion is "*the present study provides convincing data confirming that the constant wearing of static, permanent, magnetic insoles produces statistically significant reduction of neuropathic pain*".

The authors presented their results for the primary outcomes in a table format. For each month and for each outcome measured they presented a mean and a standard deviation.

In the accompanying text the authors related that changes from baseline to various months were statistically significant for both magnets and control with all of the four primary outcomes at some point in time.

The authors did *not* present point estimates with 95% confidence intervals during any time period *between magnets and placebo*. It isn't clear nor do they clearly state in the article that static magnetic insoles are more effective than placebo for the reduction of neuropathic pain seen in diabetic neuropathy. Are magnets more effective than placebo in the treatment of diabetic neuropathy?

Typically the authors at the conclusion of the trial using the original data would have used a statistical test to generate a *p* value and a point estimate with 95% confidence interval for a comparison between groups for the primary outcome. Since this data is not available to the reader the information in the article can be used to determine an effect size. Using the number of participants, means and standard deviations reported by the authors for the various outcome measures at four months an effect size between magnets and placebo can be calculated. Calculation of effect sizes is commonly seen in meta-analysis to determine treatment efficacy. There are different methods available to calculate an effect size, a common method to determine significance is to calculate an effect size as described by Cohen.<sup>18</sup> An effect size less than 0.2 is not consistent with a clinical/statistical significant effect. Greater than 0.2 is a small effect, greater than 0.5 is a moderate effect, greater than 0.8 is a large effect.

Although all of the measurements at four months (Table 3) for the various primary outcomes favor magnets, the only outcome at four months which shows a small clinically/statistically significant effect favoring magnetic insoles is numbness and tingling. The other three outcome measures are all less than 0.2 therefore, neither clinically nor statistically significant.

Outcome	Effect Size Cohen's $d^*$ (CI)
Burning pain	0.15 (+/-0.22)
Numbness and tingling	0.23 (+/- 0.23)
Foot pain	0.07 (+/- 0.24)
Sleep	0.1 (+/- 0.26)

**Table 3** Comparison using effect size.

\*Effect Size (Cohen) = experimental group mean- control group mean / standard deviation of control group.

The upper limit of all of the primary outcomes is compatible with a small significant effect. Would this be apparent if more patients were enrolled? Was this study underpowered for any or all of the primary outcomes? Would the results of the study be different if the data were analyzed in an intention to treat analysis rather than using a per protocol analysis? In conclusion, based upon the data presented in the article it is unlikely that magnetic insoles have any effect in relieving symptomatic diabetic peripheral neuropathology.

## References

1. Turlik M. How to Evaluate the Internal Validity of a Randomized Controlled Trial. *The Foot and Ankle Journal*, 2 (3) 5, 2009.
2. Chan A., et. al. Evaluating the Internal Validity of a Randomized Controlled Trials. *JAMA* 291: 2457 – 2465, 2004.
3. Trial Registration. <http://clinicaltrials.gov/him> accessed 1/23/09.
4. Poolman RW, Struijs, Krips R, Sierevelt IN, Marti RK, Farrokhyar F, Zlowodzki M, Bhandari M. Reporting of Outcomes in Orthopaedic Randomized Trials: Does Blinding of Outcome Assessors Matter? *J Bone Joint Surg* 89A: 550, 2007.
5. Austin P. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *Journal of Clinical Epidemiology* 59: 964, 2006
6. D'Agostino R. Debate: The slippery slope of surrogate outcomes. *Curr Control Trials Cardiovasc Med* 1: 76 – 78, 2000.
7. Martin R. Current Concepts Review: Foot and Ankle Outcome Instruments. *Foot & Ankle International* 27: 383 - 390, 2006
8. Montoti V. How should clinicians interpret results reflecting the effect of an intervention on composite end points: should I dump this lump? *EBM* 10: 162, 2005
9. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *NEJM* 357: 2189 – 2194, 2007.
10. Consort Statement. <http://www.consort-statement.org/> accessed 1/23/09
11. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. *BMJ* 323: 1123 – 1124, 2001.
12. Landorf KB, Keenan A-M., Herbert A. Effectiveness of Foot Orthoses to Treat Plantar Fasciitis. *Archives internal medicine* 166: 1305 – 1310, 2006.
13. Veves A, Armstrong DG, Sabolinski ML, and the Apligraf Diabetic Foot Ulcer Study. Graftskin, a Human Skin Equivalent, Is Effective in the Management of Noninfected Neuropathic Diabetic Foot Ulcers. *Diabetes Care* 24: 290 – 295, 2001.
14. Number needed to treat calculator <http://www.graphpad.com/quickcalcs/NNT2.cfm> accessed 1/23/09
15. McAlister F. The "number needed to treat" turns 20 — and continues to be used and misused. *CMAJ* 179: 549 – 553, 2008.
16. Turlik M. Introduction to Evidence-based Medicine. *The Foot and Ankle Journal*, 2 (2), 4, 2009.
17. Wientraub M, Wolfe GI, Barohn RA. Static Magnetic Field Therapy for Symptomatic Diabetic Neuropathy: A Randomized, Double-Blind, Placebo-Controlled Trial. *Arch Phys Med Rehabil* 84: 727, 2003.
18. Effect Size calculator [http://davidmlane.com/hyperstat/effect\\_size.html](http://davidmlane.com/hyperstat/effect_size.html) accessed 1/23/09.