

# Evaluating the results of a Systematic Review/Meta-Analysis

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

The Foot and Ankle Online Journal 2 (7): 5

*This is the second of two articles discussing the evaluation of systematic reviews for podiatric physicians. This article will focus on publication bias, heterogeneity, meta-analysis analytic and sensitivity analysis. A recent article related to plantar foot pain will be critically evaluated using the principles discussed in the paper.*

**Key words:** Evidence-based medicine, review article, meta-analysis

Accepted: June, 2009

Published: July, 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))

In the event that the primary studies selected for the systematic review are so dissimilar (heterogeneity) that it is ill-suited to combine the treatment effects, the systematic review will end with a table describing all of the articles abstracted. The table should contain each individual reference with the abstracted information to include: the results of the study as well as, the quality evaluation of the article done by the authors of the systematic review. The results of a systematic review are qualitative rather than quantitative (meta-analysis). The evaluation of individual randomized controlled trials has been covered earlier in this series.<sup>1,2,3</sup> The authors in the narrative results section should explain why the studies were unable to be combined into a pooled estimate of effect (meta-analysis).

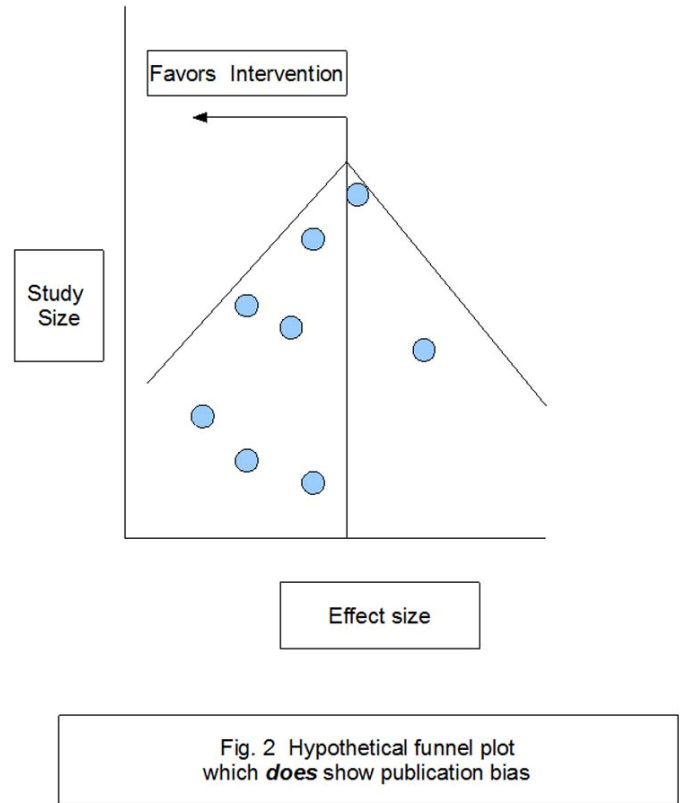
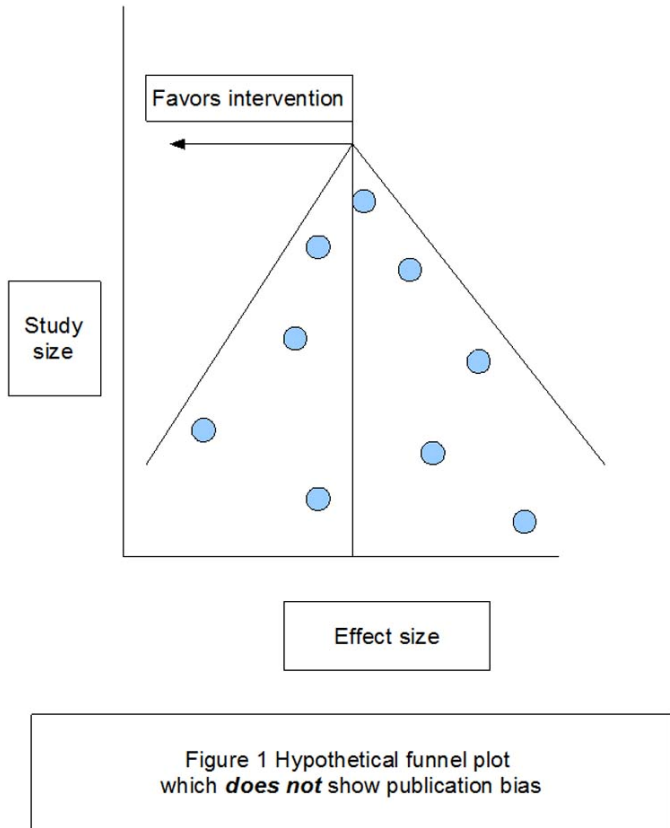
## Meta-Analysis

The results of a systematic review are a function of the quantity and quality of studies found during the review. The conclusion of a systematic review may be that after reviewing the published studies the clinical question cannot be answered and that there is a need for a larger, or a more rigorous study design to answer the clinical question.<sup>4,5</sup>

This article is the second and final article explaining systematic reviews/meta-analysis. The first article evaluated the internal validity of a systematic review.<sup>6</sup> The purpose of this article is to explain the results section of a meta-analysis using a recent meta-analysis of extracorporeal shockwave therapy (ESWT) for mechanically induced heel pain<sup>7</sup> as a guide.

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

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



A meta-analysis uses statistical techniques to combine data from various studies into a **weighted** pooled estimate of effect. Meta-analysis overcomes small sample sizes of primary studies to achieve a more precise treatment effect. In addition, meta-analysis is thought to increase power and settle controversies from primary studies. When **not** to perform a meta-analysis: the studies are of poor quality, serious publication bias is detected or the study results are diverse.

**Publication Bias**

Reporting bias can be defined as the author's inclination not to publish either an entire study or portions of the study based upon the magnitude, direction or statistical significance of the results.<sup>8</sup> A type of reporting bias is publication bias, which refers to the fact that the entire study has not been published.

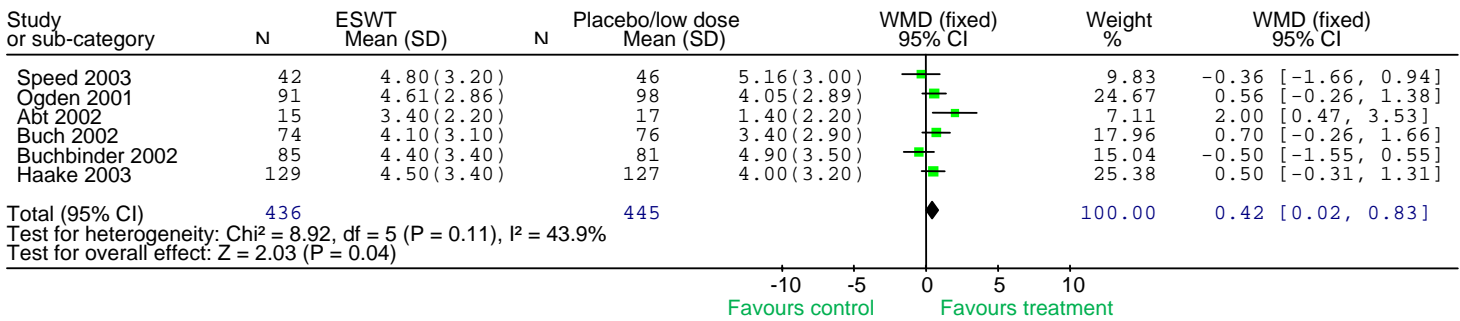
Systematic reviews which fail to search and find unpublished studies which report negative results may lead to an over estimation of the treatment effect.

Small trials with negative results are unlikely to be published and if they are may be published in less prominent journals.

Large studies which report positive results may receive a disproportionate amount of attention. They may be actually published more than once. This is the opposite of publication bias. Therefore, it is important for the authors performing a meta-analysis to eliminate duplicate publications otherwise the treatment effect will be overestimated.

A common method to search for publication bias is to construct a funnel plot (Fig 1, 2). A funnel plot for evaluation of publication bias is a scatter diagram of randomized controlled trials found as a result of the systematic review in which the treatment effect of the intervention appears along the X axis while the trial size appears along the Y axis.

Review: Interventions for treating plantar heel pain  
 Comparison: 22 morning pain - eswt  
 Outcome: 01 VAS 10cm



**Figure 3** Results from ESWT study<sup>7</sup> (presented in forest plot).

The precision of estimating a treatment effect from a clinical trial increases with increasing sample size and event rate. Smaller studies show a large variation in treatment effect at the bottom of the funnel plot. When no publication bias is present the graphical representation reveals an inverted funnel (Fig 1).

When publication bias is present typically it will be noticed that smaller studies are missing which do not favor the intervention typically the lower right-hand side of the plot resulting in an asymmetrical presentation (Fig 2). It is difficult to evaluate publication bias in the meta-analysis using a funnel plot if the study is composed of a small number of trials with small sample sizes.<sup>9</sup> The reader is referred to the following references for a more complete explanation of the subject matter.<sup>10,11</sup>

Returning to our article evaluating ESWT for mechanically induced heel pain,<sup>7</sup> in the methods section the authors state that they will use a funnel plot to evaluate for publication bias. A funnel plot could not be found when reviewing the figures in the results section. At the end of the article the authors discuss in the narrative of the study their findings regarding publication bias. The authors were unable to recognize the existence of small, unpublished studies showing no statistically significant benefits. As a result it is likely that the treatment effect found many overestimate the actual treatment effect.

### Heterogeneity

It is common to expect some variability between studies. However, if the variability between studies is significant the inference of the meta-analysis is decreased, and it may no longer make sense to pool the results from the various studies into a single effect size.

There are two types of heterogeneity, clinical and statistical.<sup>12</sup> Are the patient populations, interventions, outcome instruments and methods similar from study to study (clinical heterogeneity)?

Are the results similar from study to study (statistical heterogeneity)? Large differences in clinical heterogeneity improves generalizability however, may produce large differences in results which weakens any inference drawn from the study.

Clinical heterogeneity is best evaluated qualitatively. It is a clinical judgment based upon the reader's understanding of the disease process. The reader needs to ask the following question; is it likely based upon the patient populations, the outcomes used, interventions evaluated and methodology of the study that the results would be similar between studies? If the answer to this question is no then a meta-analysis does not make sense. If the answer to this question is yes the authors should proceed to evaluate statistical heterogeneity.

Statistical heterogeneity can be evaluated both qualitatively and quantitatively. Qualitative evaluation involves developing a forest plot of the point estimates and corresponding 95% confidence intervals of the various primary studies selected for pooling (Fig 3). Are the point estimates from the various primary trials similar from study to study and do the 95% confidence intervals about the point estimates overlap? If the answer is yes, there is not significant heterogeneity and a pooled treatment estimate makes sense. For example, in the forest plot from the ESWT study<sup>7</sup> (Fig 3) although the point estimates do not all favor the intervention they are fairly close to each other. In addition, there appears to be overlap of the 95% confidence intervals for all of the studies. The conclusion one should reach is that there is not significant heterogeneity in this systematic review and therefore one should proceed to pool the data. In contrast, when the point estimates are not grouped together and the 95% confidence levels do not overlap then significant heterogeneity exists and that data should not be pooled.

Statistical heterogeneity can also be evaluated by statistical tests.<sup>15</sup> The two common tests are Cochran's Q and the  $I^2$  statistic. Cochran's Q is the traditional test for heterogeneity. It begins with the null hypothesis that the magnitude of the effect is the same across the entire study population. It generates a probability based upon the Chi squared distribution. The test is underpowered therefore;  $p > 0.1$  indicates lack of heterogeneity.  $I^2$  is a more recent statistical test to evaluate for heterogeneity.<sup>14</sup> The closer to zero  $I^2$  is the more likely any difference in variability is due to chance. Less than 0.25 is considered mild, between 0.25 and 0.5 is considered moderate greater than 0.5 is considered a large degree of heterogeneity.

The options for systematic reviews which demonstrate significant heterogeneity are the following: do not perform a meta-analysis, perform a meta-analysis using a random effects model, explore and explain heterogeneity of the study<sup>15</sup> using sensitivity analysis / meta-regression.

The authors of the ESWT study<sup>7</sup> present in the results section in narrative and table format clinical characteristics of the primary studies.

In addition, they presented the point estimates and 95% confidence intervals of the primary studies in a forest plot with results from Cochran's Q, as well as,  $I^2$  (Fig 3). Their conclusion is that there was not significant heterogeneity present and therefore pooling of the data was appropriate.

### Meta-Analytic Models

The two different models used to combine data in a meta-analysis are random effect and fixed effect.<sup>8</sup> Both involve calculating a weighted average from the results of the primary studies. The larger the study the more impact it will have on the combined treatment effect. The fixed effect model assumes data between studies are roughly the same and any differences are due to random error. There are different fixed effect tests which can be used depending upon the type of data and the precision of the studies included. The random effects model is used when heterogeneity is encountered in the primary studies and offers a more conservative estimate. The main method is the DerSimonian Laird test. The random effects model provides less weight to larger studies and has larger confidence intervals generated about the effect size. The estimates of effect should be similar between fixed effect and random effect models if the studies do not show heterogeneity. If there is significant heterogeneity the results will differ sometimes greatly. If the meta-analysis combines different types of outcomes the results may be reported as an effect size. An effect size less than 0.2 indicates no effect greater than 0.2 indicates a small effect, greater than 0.5 indicates a moderate effect greater than 0.8 indicates a large effect.

The results of the meta-analysis should be presented as a summary point estimate with 95% confidence intervals. The authors of the meta-analysis should place the results in a clinical perspective and determine if the results are clinically significant.

The authors of the ESWT study<sup>7</sup> chose to use a fixed effect model to pool the data from the primary studies. The authors presented their findings in the results section using figures (Fig 3) and text. The pooled estimate of a 10 cm VAS scores for morning pain at 12 weeks with 95% confidence intervals is reported. The authors conclude that the pooled estimate although statistically significant in favor of ESWT is not clinically significant.

### Sensitivity Analysis

Sensitivity analysis is often carried out in meta-analyses to evaluate potential sources of bias. For example, do the results of the meta-analysis vary with trial quality, trial size, type of intervention, patient characteristics, outcome or any other variable usually determined a priori. As with any other type of subgroup analysis precautions should be undertaken when interpreting their results.<sup>8</sup>

The authors of the ESWT study<sup>7</sup> performed a sensitivity analysis comparing the results as a function of study quality. When only the trials which were judged to be a higher quality were used in the meta-analysis the results failed to reveal a statistically significant result. This is consistent with the concept that trials which lack methodological rigor overestimate the treatment effect of interventions. The authors conclude that the meta-analysis performed does **not** support the use of ESWT in the treatment of mechanically induced heel pain.

### References

1. Turlik M: Evaluating the internal validity of a randomized controlled trial. *Foot and Ankle Online Journal*. 2 (3): 5, 2009.
2. Turlik M: How to interpret the results of a randomized controlled trial. *Foot and Ankle Online Journal*. 2 (4): 4, 2009.
3. Turlik M: How to evaluate the external validity of a randomized controlled trial. *Foot and Ankle Journal* 2 (5): 5, 2009.
4. Edwards J: Debridement of diabetic foot ulcers. *Cochrane Reviews*, <http://www.cochrane.org/reviews/en/ab003556.html>. Accessed 2/23/09.
5. Valk G, Kriegsman DMW, Assendelft WJJ: Patient education for preventing diabetic foot ulceration. *Cochrane Reviews*. <http://www.cochrane.org/reviews/en/ab001488.html>. Accessed 2/23/09.
6. Turlik, M. Evaluation of a Review Article. *Foot and Ankle Journal* 2:, 2009.
7. Thomson CE, Crawford F, Murray GD: The effectiveness of extra corporeal shock wave therapy for plantar pain: a systematic review and meta-analysis. *Musculoskeletal Disorders* 6:19, 2005.
8. Guyatt G, Drummond R, Meade M, Cook D: *Users' guides to the medical literature*. New York, McGraw-Hill Medical, 2008.
9. Egger M, Davey Smith G: Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315: 629 – 634, 1997.
10. Sterne JAC, Egger M, Davey Smith G: Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 323: 101 – 105, 2001.
11. John PA, Ioannidis J, Trikalinos T: The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *176* (8): 1091 – 1096, 2007.
12. Hatala R, Keitz S, Wyer P, Guyatt G: Tips for teachers of evidence-based medicine: 4. Assessing heterogeneity of primary studies in systematic reviews and whether to combinetheir results. *CMAJ* 172: 661 – 665, 2005.
13. Fletcher J: What is heterogeneity and is it important? *BMJ*: 334: 94 – 96, 2007.
14. Higgins JPT, Thompson SG, Deeks JJ, Altman DG: Measuring inconsistency in meta-analyses. *BMJ* 327:557 – 560, 2003.
15. Ioannidis J, Patsopoulos NA, Rothstein HR: Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 336: 1413 – 1415, 2008.