Reviewer's report

Title: Osteopathic Manipulative Treatment for Low Back Pain: A Systematic Review and Meta-Analysis of Randomized Controlled Trials

Version: 2 Date: 17 May 2005

Reviewer: Elmer Villanueva

Reviewer's report:

General
My first advice (one I'm afraid I will classify as a major compulsory revision) is the adoption of the QOROM recommendations in the reporting of meta-analyses. I worked this out with Fiona Godlee when she was at BMC. This resulted in the very clear statement of authors that the BMC supports QOROM recommendations. At the very least, use of these recommendations will make most of the following redundant:

- The Authors will need to show the exact search strategy.
- The Authors will need to describe the study flow.
- The Authors will need to define the population of interest and interventions clearly.
- The Authors will need to specify the outcomes of interest. Descriptions of "contrasts" result in the difficult situation in which the Authors find themselves -- that of several hundred possible combinations from a thin evidence base (six studies). In effect, this places all outcomes on a single footing; in a sense, death is the same as debility. The Authors will need to be bold and say "We are most interested in return to function (for instance) at six weeks as the primary outcome of interest". By assiduously examining all contrasts, any results are difficult to separate from the "this pill only works for those born under the sign of Libra" saga.

My second advice is for the Authors to be very careful in the excessive use of stratified and sensitivity analysis in the presence of sparse data. For a particular PICO (patient, intervention, control, outcome) combination the Authors have, at most, a sample size of six. In most cases, it is difficult to extract information from such a small sample size. Running several exploratory analyses on such a small sample size is even less likely to be justifiably prudent.

My third advice is for the Authors to ensure that, even in the absence of statistical heterogeneity, the more substantive problem of clinical heterogeneity is considered. Statistical heterogeneity only measures the departure of each study's "average" effect from the theoretical "middle" effect (however each of these concepts is defined). Ignoring heterogeneity arising from variations in study designs, one common (and perhaps the most important) source is clinical variation.

My fourth advice is to temper findings if such findings arise from a small number of studies, in spite of the performance of a meta-analysis. The worst-case scenario is one in which a finding is drawn from the results of a single study. Since no meta-analysis actually occurs, it is misleading that such a finding is the result of the synthesis of six studies.

My fifth advice is to make "effect sizes" more clinically relevant by back-transforming Cohen's d
scores to their original scales. It is particularly difficult for a clinician to interpret an effect size of 2 (say) in decrease in medication use at six weeks (say). What this means is that medication use decreased by 2 standard deviation units at six weeks. Let me quote a common response to the latter statement: "Huh?". The difficulty is one in which the scale is incomprehensible to the common reader.

My sixth advice is for the Authors to cease comparing effect sizes of OMT to other treatments like NSAIDs because of the simple fact that no head-to-head comparisons were performed and the cross-comparisons result in a Clayton's comparison (for those not up with Australian advertising folklore, Clayton's Kola-based mixer at one stage had a campaign where, it being alcohol-free, was promoted to adults as "the drink you're having when not having a drink"). In other words, this result is the one you report when you don't have an actual result to report.

-------------------------------------------------------------------------------
Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
-------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)