Reviewer's report

Title: The relative timing of VMO and VL in the aetiology of anterior knee pain: a systematic review and meta-analysis.

Version: 2 Date: 30 December 2007

Reviewer: Chris Sutton

Reviewer's report:

Major Compulsory Revisions

1) Further reporting of between- and within-study heterogeneity should be included in the Results section. The assumption is made early that it is appropriate to provide a pooled estimate across studies; I feel that a pooled estimate should be provided (using random effects meta-analysis) only if the studies can be deemed sufficiently homogeneous for this to make sense. Whilst the "Discussion" covers potential reasons for the substantial heterogeneity observed, further salient points should be included in the Results (and possibly the "Methods", in terms of the possibility of not combining the results from the studies if highly heterogeneous). A particularly important statistical issue which is not addressed in the paper (and which relates to between-study and within-study heterogeneity) is the magnitude of the within-group standard deviation in each group in the individual studies. In particular, Boling (2006) has a far larger SD (in both groups for stair descent and for the AKP group for stair ascent) than any of the other studies whilst the SD for the control group in Crossley (2004) appears implausibly small (for both stair ascent and stair descent); these differences imply substantial differences may have applied to the entry criteria or measurement within the different studies.

2) The authors do not make a clear distinction between matching in the study design (to achieve balance on important confounding factors) and balancing on such factors (which may happen through matching or by chance). Most of the studies are not matched even on age and there appear to be moderate differences in the mean ages between AKP and control groups in most studies, with the AKP participants being typically a few years older than the controls. Whilst they argue that the age of the participants did not account for the heterogeneity between the studies, their argument is not fully convincing, only confirming that there are other sources of heterogeneity.

3) The Conclusions start with a point relating to the pooled results. Given the degree of heterogeneity, recognised elsewhere, the emphasis should be less on the pooled results and more on the findings from the systematic review and failure to explain the considerable heterogeneity within and between studies in most respects.

4) When assessing potential publication bias, the emphasis on the p-value is
inappropriate, given the small number of studies (although the interpretation of the p-value is sound).

5) The sentence “Some primary studies did not provide sufficient data for meta-analysis” is rather misleading and should be rewritten. The point the authors are making is that the necessary statistics were not presented explicitly in the paper.

Minor Essential Revisions

1) In the penultimate paragraph of the “Results”:
   - the I²-statistic is written as I² (rather than I²);
   - the word “neither” in brackets following Voight [6] appears inappropriate should this read “did not report”?

Discretionary Revisions

1) When reporting on the significance and degree of the heterogeneity in effect sizes between studies, the degree of heterogeneity (as measured using the I²-statistic) should be prioritised over the statistical significance (or non-significance) of the heterogeneity (as assessed using the p-value).

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests