Reviewer's report

**Title:** Constraint-induced movement therapy: trial sequential analysis applied to Cochrane Collaboration systematic review results.

**Version:** 1  
**Date:** 14 July 2014

**Reviewer:** Gordon Murray

**Reviewer's report:**

I have a number of major reservations with this paper.

1. The most serious is the way the authors raise the issue of multiple testing in cumulative meta-analysis, likening this issue to, for example, repeated interim analysis in a clinical trial. The crucial difference is that with interim analysis the trial continues as planned unless an efficacy or safety or futility boundary is crossed, at which point the trial stops and is reported. This process can clearly ring false warning bells, and some correction is needed to allow for multiple testing. However, in meta-analysis one wants to use the totality of available (randomised) evidence, and the most up-to-date meta-analysis (if robust, etc) will always trump a previous one. In particular, one is not 'chasing a p-value'.

2. I accept that cumulative meta-analysis is interesting if one wants to look back and ask at what point the evidence became strong, or to see evidence of early bias, or potentially as a means to power a new trial (but I have yet to encounter a funding body which is prepared to fund a trial on the basis that it could tip the balance of a meta-analysis, rather than being powered to stand alone). However, I am not convinced that one needs to take account of history to make a valid inference on the strength of the currently available totality of evidence.

3. The authors seem to be saying that early, small trials are typically biased, and that as more evidence accumulates then the treatment effect estimated from the meta-analysis tends to head towards the null. This is certainly true, but one should not attempt to solve a problem which relates to bias by using a conservative inference technique which is based on adjusting for multiplicity.

4. The authors seem to get mixed up with the MID, and how it should impact on the estimated treatment effect. These two points are unrelated, although clearly the MID is crucial for a power calculation, or for interpreting an inference. In terms of their case study, this observation seems to circumvent their problem. The estimated effect size for the conventional meta-analysis for the primary endpoint is 2.88 units with a 95% CI of 0.08 to 5.68 and a p-value of 0.04. Although conventionally 'significant' it is obvious that the p-value is so borderline and the 95% CI comes so close to including a null effect, that it would be ridiculous to take this as evidence of clinical efficacy. It does not need a spurious injection of a multiplicity correction to add a degree of conservatism, just common sense!
5. Not that it is that relevant (see point 4), I am concerned to see that authors advocating 0.5 SD as a way to establish the MID. The MID depends totally on the clinical context, with, for example, a cheap, safe intervention needing only a modest MID to justify its use in clinical practice whereas an expensive intervention with a worrying safety profile would need a much larger MID before its use could be considered.

6. A minor point, but given the authors’ focus on multiplicity, it would be good to see some discussion of how to handle the two outcome measures which they are exploring in their case study.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

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

**Declaration of competing interests:**

I declare that I have no competing interests.