Reviewer's report

Title: Consensus-Based Recommendations for Investigating Clinical Heterogeneity in Systematic Reviews

Version: 1 Date: 2 February 2013

Reviewer: David Kent

Reviewer's report:

The authors present consensus-based recommendations for investigating “clinical heterogeneity” in systematic reviews. They have gathered a considerable amount of expertise, so the recommendations should be of interest to readers. Nevertheless, I do have some 3 major concerns: 1) the investigators do not seem to have resolved major controversies and therefore the set of recommendations are somewhat vague, and largely well appreciated and accepted by practitioners, so it was not entirely clear to this reader what is “newsworthy” 2) Recommendations for exploration of HTE in MIPD (and other patient-level analyses are inadequately covered, so my recommendations would be to be explicit that these recommendations are focused on study-level analyses only; 3) the authors indicate that their investigation was motivated by concerns that systematic reviews and meta-analyses are not applicable to individual patients, yet they appear unwilling to face up to how limited systematic reviews generally are in exploring “clinical heterogeneity”, particularly at the patient level. Overall, I thought the manuscript could be considerably compressed (and perhaps combined with the “longer explanatory manuscript” in preparation.

Major compulsory comments:

1. Abstract
   The background indicates that the objective for examining HTE is to make meta-analytic results more applicable to “individual patients”. This suggests the importance of patient-level variables in determining treatment effects. One of the areas of consensus, based on several prior and cited studies is that study-level analyses are inappropriate for patient-level inferences because of ecological biases. Nevertheless, the conclusion of the abstract is that “these investigations may allow increased applicability of findings of systematic reviews to the management of individual patients.” It is unclear how the conclusions are justified by the findings of the investigators. A clear statement that patient-level inference require patient-level analyses would clarify this issue.
   The results read a bit like an expansion of the methods. There is no summary of any of the (most-important) recommendations.

2. The definition of Clinical Heterogeneity (page 8):
   The first two sentences of this paragraph offer slightly different definitions of clinical heterogeneity, and both sentences differ slightly from the definition in the abstract. These differences are not just in wording. The first sentence suggests
that clinical heterogeneity is defined as “variation in treatment effects”; the second sentence is that clinical heterogeneity is “clinical variation” that gives rise to treatment effect variation; while the abstract defines it as variability in study characteristics (without the explicit requirement that this variability give rise to heterogeneity of treatment effects [HTE]). According to the first definition “clinical heterogeneity” is a form of HTE (i.e is defined as variability in the effects on the dependent variable). According to the second definition it is variation in the independent variables, but must give rise to HTE. According to the definition in the abstract, it is merely clinical variation (variation in the independent variables that may or may not give rise to HTE). To me, it seems as if clinical heterogeneity may be present even in the absence of HTE. Please clarify, and make definitions consistent.

3. Recommendations (Table 1): Please clarify what “pre-planned” and “a priori” means in the context of systematic reviews, when presumably the data of the component studies has already been open to review and scrutiny, and would be widely known by the clinical experts participating in the review. Many of the hypotheses generated and tested in a systematic review are motivated based on observations within the component studies themselves. Indeed, the recommendation for a priori investigations may in some ways contradict the recommendations that examination of HTE be based on a strong rationale that might be derived from “evidence from previous research” – which presumably would be included in any systematic review.

4. Analyses of IPD seem to be a separate topic and should probably be removed from the table. The authors do not seem to offer a (comprehensive) framework or literature review for patient-level HTE analysis.

5. Between meta-regression and IPD, there is no discussion of meta-analysis of subgroup effects, when subgroups are reported in the original article. This seems like an important omission.

Minor essential revisions:

6. While guidelines encouraging parsimony in HTE analyses are probably needed, the recommendation for 10 studies per variable seems poorly supported. The simulation by Peduzzi et al is oft-cited, but not especially robust. Other more robust analyses have suggested that higher EPVs are often necessary to avoid over-fitting (e.g Courvoisier, DS, et al J Clin Epidemiol. 2011 Sep;64(9):993-1000). None of these studies applied to study-level analyses. It should also be noted that this (limited power) is another reason why study-level meta-analysis/meta-regression will typically have only limited utility for HTE analysis.

7. Several variables (e.g. “length of follow-up”) can be either patient-level variables or study level variables. This should probably be clarified.

8. Some of the detail provided in the methods section, which bleeds into the results, is probably unnecessary and this section can be compressed.

**Level of interest:** An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I know several of the authors and participants well; I have co-authored a paper with Doug Altman and he is a co-investigator of a project of mine; Xin Sun and I have had several conversations about various mutual interests, but have yet to collaborate. I did not think that this represented a disqualifying COI.