Reviewer's report

Title: Diagnostic performance of alpha-fetoprotein, Lens culinaris agglutinin-reactive alpha-fetoprotein, des-gamma carboxyprothrombin, and glypican-3 for the detection of hepatocellular carcinoma: a systematic review and meta-analysis

Version: 1 Date: 6 April 2013

Reviewer: Mariska M. G. Leeflang

Reviewer's report:

It was a pleasure to review this excellent protocol for a systematic review of diagnostic test accuracy. Most of my comments are therefore minor comments, or questions for clarification. The only major comments are about the role of these tests in practice and their role compared to each other.

Major compulsory revisions

1. Could the authors please elaborate a little bit on what they expect the role of these tests to be in practice. I mean, what do the authors expect to happen to people who test positive? Will they be referred for transplantation or resection immediately (as to me, the add-on role could imply)? Or will first another test be done before resection? Or is the resection the final test as well?

2. Only in the last paragraphs it becomes clear that the authors have thought about comparing the biomarkers against each other as well. However, from the background it is not clear if there could be a choice made for one of these tests. In the background, the author do explain that these tests may have value if they are used together, in parallel or serial. Do they plan to assess this hypothesis further, if the data are available? Could the authors be clear about these two questions (simple comparisons and the combination testing) from the beginning? Would it, for example, be part of the second objective, about assessing the potential role of these tests?

3. On one hand, these biomarkers are described as screening tests; on the other hand they are described as add-on tests. These are for me two different roles. Or is this the question that the second objective refers to (what is the role of the tests)? In that case, it may be made a bit more explicit.

Discretionary Revisions

4. Applicability in QUADAS-2 is phrased in terms of low or high concerns regarding applicability, while the authors state it here as low applicability or high applicability. I found this a bit confusing in first instance.

5. Will only AFP be potentially used as part of the reference standard? Or DCP or glypican as well? This should then be addressed in the relevant QUADAS-2 item.

6. The section about displaying sensitivity and specificity in the forest plots, but not if AFP was used in the reference standard, was a bit confusing. It may
already help to add a white line between these sentences and the previous one.

7. Would it be possible to be consistent about AFP-L3 or AFP-L3/AFP ratio? I have the feeling that these terms are used inconsistently (but I am not an expert in this field, so perhaps there is a reason for this mixed usage).

**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