Reviewer's report

Title: Does Risk for Ovarian Malignancy Algorithm excel human epididymis protein 4 and CA125 in predicting epithelial ovarian cancer: a meta-analysis

Version: 1 Date: 12 March 2012

Reviewer: Mariska M. G. Leeflang

Reviewer’s report:

On first sight, these authors seem to have done an excellent job. They followed all the steps needed for a comparative accuracy review and even assessed publication bias, heterogeneity and the effect of individual studies. However, I do miss the clinical context and a clear interpretation of all the results that they present. If I were an oncologist/gynaecologist (which I am not), what would these results mean to me and to my patients? Which are the outcomes of most interest to me and how should I interpret these? I also think the reporting of the methods should be done in a clearer and more concise way.

MAJOR COMPULSORY REVISIONS

1. The study objectives and the link between these objectives and the study methods should be better explained. The last sentence of the background indicates that the aim is twofold:
   I. assessing the accuracy of ROMA
   II. comparing the accuracy of ROMA, HE4 and CA125 against each other.

1a. For the second question, the authors limit their study base to studies comparing either HE4 against CA125 or ROMA against both HE4 and CA125. Although this makes sense, I would like to know why the authors did not include studies evaluating only CA125 or only HE4.

1b. The authors do not explain in their methods section how they will analyze the comparative question. Which studies will feed into that (only comparative studies, or also the only-ROMA studies)? Did they check whether all samples were tested for HE4 and for CA125 and whether all samples were included in the analyses of ROMA? To give an example, a study that included 100 patients of which 75 were sampled for CA125 and 85 for HE4, but only 60 patients delivered a sample for both tests, then only this subsample of 60 can be used for ROMA-analyses. I suspect that this is not an issue in this review (as it is easy to do two tests on one blood sample), but I can also imagine that sometimes not enough blood could be drawn for two test, or that CA125 was already included in the clinical routine (because approved by the FDA) and thus used more often.

1c. The authors used both QUADAS and QUADAS-2, but they did not include questions to assess the validity of this comparative question. Examples are: were the results for CA125 masked when HE4-result was read? Did all patients
undergo both HE4, CA125 and ROMA? Were the patients in these studies the same as the patients in the studies only addressing HE4/CA125/ROMA?

2. The questions above reveal another problem with this review: the clinical context. The authors state that CA125 and ROMA are approved by the FDA, but they don’t explain whether these tests are used in practice and at which stage in the diagnostic process they are used. Neither do they explain this for HE4. What will these tests be used for and in what setting? Will they be used by the general practitioner to refer patients to the hospital (when test positive) or to send them home (when test negative)? Or is the test used to decide who goes for further diagnosis by, for example, CT scanning? Or will these test be used to confirm the diagnosis of EOC?

3. A comparative question also almost asks for a null hypothesis: what do the authors expect to find? That ROMA is more sensitive than both CA125 and HE4 alone? And when would the authors decide that ROMA is ‘better’ than either CA125 or HE4?

4. Did the authors use the studies that compared HE4 with CA125 to calculate ROMA from these results and include them in the ROMA-analysis?

5. The authors state that they found 11 articles meeting the inclusion criteria. But from the text of the search results section, it appears that the three studies that compared the performance of ROMA, CA125 and HE4, were among the 6 only evaluating ROMA. Also, of the four studies comparing HE4 and CA125, three were already mentioned under evaluation of ROMA alone. So I come to 7 articles in stead of 11. Also, the numbers of patients seem to be not correct (some patients counted twice?).

Under the methods of index tests section, the numbers are different. Eight studies used EIA to measure HE4. The references 16 and 19 and 21 were not mentioned under the previous section. Three other studies used CMIA, but reference 14 was not mentioned under the previous section. This implies that the authors did search for studies only including HE4. The same seems to be true for CA125. This is all very confusing!

This is all very confusing and a clearer explanation of what analyses were included in which studies/articles is really needed.

6. The authors state that “in all studies, the spectrum of patients was considered representative”, but they do not explain what this means. Are these women coming to primary care, secondary care, etc?

7. I don’t understand the explanation about the concerns of applicability for domain 2 of QUADAS-2. Please add a few sentences of what this really means.

8. The results under the ROMA diagnostic value section seem to be analysed correctly. But I do miss an explanation of which studies feed into which analyses (although it can be extracted fro the forest plots). I also think that there are too many forest plots to be informative. I would rather restrict to only one forest plot.
and mention the results for the subgroups in the text only.

9. The results under the two performance comparison sections also seem to be correct. But I don’t understand how it is possible that only four studies compared HE4 to CA125, while all 11 studies evaluating HE4, also evaluated CA125 (see under methods of index test). Again, please explain better which studies included which analyses and how these analyses were used in the meta-analyses and why.

10. I miss a sROC plot in which all tests are included. Or at least a separate one for CA and HE.

11. Would it be possible to formally test for a difference in the accuracy of ROMA versus HE4 versus CA125? Under the Conclusions section, the authors state that ROMA and HE4 can replace CA125. But the confidence intervals of ROMA and CA125 show very much overlap and I can’t find the sensitivity of CA125 (so can’t judge how different this is from the other two).

12. Table 3 is unreadable and impossible to understand. This may be due to the format. Also, what do the numbers under the comparative section of table 3 mean? If it says 0.86 under sensitivity for EOC-ROMA in the ROMA:HE:CA section, how does this differ from the 0.90 for the only-ROMA sensitivity (should perhaps be addressed in the discussion)?

13. Please address in the discussion what these results mean and for which patients. What does it mean for example, an AUC of 0.93 for ROMA? What do all the I-squares in table 3 mean; what is the meaning of some being 0.0%, while others are 88%? What is the consequence of ROMA being better capable of detecting advanced stage EOC than detecting early stage EOC (while the specificity remains the same)? What do the numbers for the comparative analyses mean (see also table 3)?

14. Please add a short description of the methods to the abstract.

15. The authors state that they used the bivariate model to analyze sensitivity and specificity. The studies report however a wide variety of cut-off values (for HE and ROMA). What does a summary sensitivity and specificity mean then? To which cut-off value does this relate? Would it matter if someone uses one cut-off value or another? In general, when studies report varying cut-off values, the HSROC model is recommended. State reports the results for this model as well (although the results may be more difficult to interpret). Why didn't the authors use the results from this model?

MINOR ESSENTIALY REVISIONS

The methods section comes after the results section, which is very confusing and unhelpful.

Where does OC stand for and what is the meaning of the subgroup analyses for OC?
DISCRETIONARY REVISIONS

The authors used both QUADAS-1 and QUADAS-2 and I don’t see the rationale for doing so. Most questions of QUADAS-1 are incorporated in QUADAS-2 as signalling questions, so there is much overlap between the two.

Remove the abbreviations for LMP and BL and use the full words. It is only used a few times and remains confusing when reading only the abbreviations.

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

Quality of written English: Not suitable for publication unless extensively edited

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

Declaration of competing interests:

I declare that I have no competing interests