Reviewer's report

Title: A population-based observational study comparing Cervista and Hybrid Capture 2 methods: Improved relative specificity of the Cervista assay by increasing its cut-off

Version: 2
Date: 27 July 2014
Reviewer: Jesper Bonde

Reviewer's report:

“A population-based observational study comparing Cervista and Hybrid Capture 2 methods: Improved relative specificity of the Cervista assay by increasing its cut-off”

Gerd Boehmer, Lisa Wang, Angelika Iftner, Barbara Holz, Juliane Haedicke, Reinhard von Wasielewski, Peter Martus and Thomas Iftner

BMC Infectious Diseases (Submitted: 2014-05-15)

The manuscript from Boehmer et al was a pleasure to review, and I have only minor revisions to be entrusted to the Authors.

The manuscript puts forth a very interesting question as to the precision of the company set cut off of a new generation HPV test, the Cervista HPV assay. Thereby the data reported represents a data driven evaluation of whether the FDA approved and CE-IVD marked Cervista assay in fact has the right distinction between true disease and more or less transient infection, in a non-US population. Moreover, the manuscript offers comparative data in a true split sample fashion between the former marked leading assay, HC2, and Cervista, mostly I assume as HC2 is outlined in the European Meijer Criteria as the comparator for any new HPV test with respect to clinical sensitivity and specificity of detection of CIN2+. However, the value of true split sample studies can in this respect not be emphasised enough.

The study design was based upon available residual samples. I fail to understand why written consent was obtained when the study was not required for ethical approval under German law. However, this is probably more this reviewer’s lack of insight into German rules and regulations, and should not negatively influence the evaluation of the manuscript.

1) A brief comment however as to why informed consent was obtained when no approval was required would be appreciated in the Authors reply, if not in the paper itself.

2) Moreover, in Methods, I would appreciate an insertion at line 112 stipulating “…according to routine guidelines” to ensure that the reader understands that the follow up is dictated by regional/national guidelines, not individual laboratory set follow up definitions.

3) Finally, please note how many pathology adjudicators samples were seen by.
The results are easily presented in a shot Epidemiology intro and four well written sections, and a discussion. A few things however, spring to mind when reading the result section, the Tables and Discussion:

1) In the discrepancy analysis and the concluding discussion no real discussion of the genotypes detected by the two assays discrepantly are presented, and either the implications are discussed, or Table 3B can be reduced to a simpler HR/Intermediate/LR table. For instance, what is 21 or 39 discrepant HC2+ results are intermediate or LR (44- 3 “No DNA”- 2 “No result”), whereas the same number for Cervista is 5 of 14, by my calculations (text says 16?). This indicates to me that the Cervista design at least is superior in specifically detecting HR HPV over HC2´more than 50% cross reactivities, and a comment to this end could be beneficial to the readers in order to facilitate a full appreciation of this story.

2) The second point is that I lack a discussion of the InnoLipa as an adjudicating genotyping assay. Strength, weakness and a short reference to why this assay is a good adjudicator of molecular HPV screening assays.

3) Why were 21 samples inadequate for Lipa testing?

4) I like the way of calculating a relative measure between two tests when full scale sending women for colpo is not an option. Let time show if this way as comparison will gain broad acceptance.

5) Furthermore, in the discussion of the sensitivity and specificity data I would like a comment in the discussion comparing the German prevalence against the other referenced Cervista studies, just to ensure that any differences in background prevalence would not be at play causing, as noted on pp 13, results not in line with previous reports.

6) Analysis of HPV prevalence by Cervista HPV HR assay using a different cut off: Well written, the key point of the paper. However, from a philosophical infectious medicine point of view, does the prevalence change by altering the cut off? I would argue that the prevalence of infection remains the same, however for a screening assay tasked with detecting imminent disease, cut off matters. The Authors fail to make this distinction which is the brilliant point the manuscript makes and I suggest re-phrasing the subtitle of this result section, not to confuse prevalence and detection of disease. The Authors actually make this distinction in lines 298-301 already.

7) The data and following discussion clearly shows that a reduction of false negative Cervista results can be obtained by changing the company set cut off. Without loss of detection of CIN3. Could I suggest utilizing the genotype information at hand to elucidate the genotype distribution of the clinically false positive Cervista results, and make a comment on comparing these information’s to the general distribution of genotypes in the dataset as a whole? The value would be to point to individual genotypes potentially overdiagnosed clinically with respect to disease (not infection) by the designers of Cervista. Alternatively, to show that the distribution equals that of the general population as a whole, however, the false positives being “close to cut off samples”.
Finally, and this is not a “mandatory request by reviewer”, more a point of discussion, as I will not request the Authors to quote our papers, I will however point out that our group previously published a systematic review of randomised control trials using HC2, where we compared the detection ability of HC2 at different cut off points (Rebolj, BMJ, 2011) and a description of assay discordance between HC2, APTIMA, cobas, and CLART HPV assays (Rebolj, PLOS ONE, 2014).

Firstly, we point out in BMJ that false negative rate in screening using HC2 can be reduced by increasing the cut off of HC2. So evaluation of cut off’s is not a foreign concept, and even HC2 has been subjected to this line of analysis. We find by the way that you could increase the HC2 cutt off to 10 without substantial loss of detection of high grade disease. That’s a massive reduction in false positive screening samples, typically NILMs.

Secondly, in the PLOS ONE publication of this year we do a split sample comparison between APTIMA, HC2, cobas, and CLART on SurePath taken samples. The conclusion is that especially in NILMs, a high rate of discordance is observed and that this discordance, assay by assay, represents screening false positives.

So there is ample published evidence supporting the line of thinking of the Authors, AND showing that this is not a Cervista localised issue, though this has only now been thoroughly evaluated for the Cervista HR HPV assay by this manuscript.

Jesper Bonde

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

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

Declaration of competing interests:

I have served as a paid advisor in the past to Roche and Genomica, and received honoraria in the past from Hologic/Gen-Probe, Roche, Qiagen, BD Diagnostic and Genomica for lectures.

I am not on retainer from any biotechnology company as advisor or lecturer, and I declare no competing interest.