Author's response to reviews

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

Authors:

Gerd Boehmer (gboehmer@gmx.net)
Lisa Wang (Lisa.wang@med.uni-tuebingen.de)
Angelika Iftner (Angelika.iftner@med.uni-tuebingen.de)
Barbara Holz (Barbara.holz@med.uni-tuebingen.de)
Juliane Haedicke (Juliane.haedicke@med.uni-tuebingen.de)
Reinhard von Wasielewski (Reinhard.vonWasielewski@krh.eu)
Peter Martus (Peter.martus@med.uni-tuebingen.de)
Thomas Iftner (Thomas.iftner@med.uni-tuebingen.de)

Version: 3
Date: 25 September 2014

Author's response to reviews: see over
Revision

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

Dear Editor,

We would like to submit a revised version of our manuscript “A population-based observational study comparing Cervista and Hybrid Capture 2 methods: Improved relative specificity of the Cervista assay by increasing its cut-off” (MS: 1248854987129785).

We greatly appreciated the comments of our reviewers and have made modifications to our manuscript in accordance to their suggestions. We hope that this new version of our paper will be considered suitable for publication in BMC Infectious Diseases.

Yours faithfully,
Thomas Iftner

Our responses to the reviewer’s comments are as follows:

**General Corrections**
We have made minor editorial changes throughout the manuscript.

**Editorial Requirement**
Please clarify within the manuscript the regulations which make this study exempt from needing formal ethics approval.
Please refer to Reviewer 1 comment 1.

**Reviewer: Jesper Bonde**

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.
We thank the reviewer for reviewing our manuscript and for providing us with a fresh view onto our study.

Please accept our sincere apology for the confusion our ethical approval statement has caused. It had been added after an internal communication problem between Bad Münder and Tübingen. We would like to strongly emphasize that ethical approval was indeed not required due to the purely observational character of this study and hence written informed consent was
No ethical approval was required for this purely observational study, because residual patient samples were used that were completely anonymized and study results had no influence on the patients’ follow-up strategy. Written informed consent was obtained from each patient prior to study recruitment. A total of ...”

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. We agree with the reviewer and have made the requested insertion.

3) Finally, please note how many pathology adjudicators samples were seen by. After initial diagnosis, samples were seen by at least one reviewer. In case of discrepant results another review was performed. We can see how the description of the histology reviews might have been misleading and we have therefore changed that section as follows:

“Histology Review. All samples with a primary histology result of CIN2+ were reviewed by an independent external expert. In case of a discrepant review readings, a second/third histology review was performed. If two out of three diagnoses were identical, the result was considered final.”

4) 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. We thank the reviewer for his points of concern regarding the genotyping results.

- We have tried to reassess the data based on the reviewer’s suggestion. However, we were not able to reach the same conclusion. According to table 3B of the 39 HC2+ve/CER-ve samples 16 were low risk non-target types (41.0%) while the same number for CER+ve/HC2-ve is indeed 5 of 14 (35.7%). Based on this comparison, we cannot conclude that CER is superior in detecting HR HPV especially considering the fact that CER actually missed 23 target-type HR infections including 3 CIN2+ cases, while HC2 only missed 7. Also, despite HC2 having detected 16 LR infections, one might argue that CER detected 56 cases where no HPV DNA was present whatsoever.

- N=16 in the text is correct. It refers to the number of CER non-target types including 2 HPVX samples in the CER-ve, HC2+ve column. However, in order to address the high cross-reactivity of HC2 with low risk HPV types, we have
added the following sentence to the discussion:
“...[28]. In line with previous studies (summarized by [29]) our discrepancy analysis demonstrated substantial cross-reactivity of HC2 with non-target HPV types (n=16 versus n=3 for HC2 and Cervista, respectively). Furthermore, discrepancy analysis revealed ...

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

The INNO-LiPA genotyping assay has previously been used as an adjudicating assay in test comparison studies (e.g. Hwang, Ann Lab Med, 2012). The LiPA assay is one of the most sensitive commercial genotyping tests currently available, as it is based on the detection of a very short HPV L1 amplicon of 65bp in length. The analytical sensitivity of the LiPA assay is excellent at 20 to 70 copies per assay (Torres, Open Virol J, 2012).
We have added the following text to the discussion:
“... group (p<0.0001; table 2B). In order to resolve discordant results genotyping by INNO-LiPA Extra was performed. LiPA genotyping has previously been used as an adjudicating assay in test comparison studies [26, 27] because of its excellent analytical sensitivity of 20 to 70 copies per assay [28]. Discrepancy analysis revealed ...

6) Why were 21 samples inadequate for Lipa testing?
Invalid LiPA results occasionally occur and are based on the fact that samples return a negative result on the hDNA control line targeting an amplicon of the human HLA-DPB1 gene. Reasons for sample failure may include inadequate specimen collection, processing or presence of inhibitors in the DNA extract.

7) 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.
We thank the reviewer for this statement.

8) 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.
We have added the following statements to the discussion:
“... screened by LBC. Overall HR HPV prevalence was 27.8% and 25.2% for Cervista HPV HR and HC2, respectively. Due to the inclusion of 60 additional patients with CIN3+, the overall prevalence is higher than HR HPV prevalence rates of ~9% previously reported for a comparable screening population with participants aged ≥30 years [14]. The overall agreement ...”
“... NILM cytology [12-17]. This disagreement with previous reports might, however, be reflect differences in study design [12, 13, 14, 17], age range of participants [12, 13, 15-17] or overall HR HPV prevalence rates [14]. Nonetheless, ...”

9) 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.

We thank the reviewer for detailing his concern. In order to emphasize the difference between detection of disease and HPV prevalence we replaced the section title as follows: “Different Cervista cut-off values reduce HPV-positivity without affecting detection of CIN3+.”

We also added the following statement: “Therefore, we also calculated HPV positivity as detected by the Cervista HPV HR test using two different modified cut-off criteria (Table 5), which were selected for the detection of CIN3+ lesions. The manufacturer-recommended ...”

10) 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”.

We here showed a reduction of the false positive Cervista results by changing the company cut-off. False positive results really are true negative results as determined by LiPA genotyping and HC2 co-testing. This means that genotypes for the false positive Cervista results do not exist. Furthermore, genotyping was only performed on the 133 discrepant samples; all other samples were not genotyped, which makes a comparison of the general distribution of genotypes within the dataset impossible. However, most false positive results were indeed distributed close to the company set cut-off, which is why we decided to adjust the Cervista cut-off values.

11) 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 randomized 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 cut 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.

We thank the reviewer for his suggestions. We have added the following statements and the recommended references to our discussion:

“... further population-based studies are necessary to confirm our finding. Studies evaluating the company set cut-off values for the detection of disease have previously been performed for other HPV tests including HC2 where adjustment led to an increased specificity of disease detection and fewer false-positive test results [31]. It is worth ...”

“... to HC2 in ASC-US specimens [13, 14, 18, 19, 21]. Finally, split sample test comparison results of multiple HPV tests other than Cervista have recently been published from the Horizon study where the authors found substantial disagreement between HPV tests especially in women aged 30-65 years. These disagreements were attributed to screening false positives and were probably based on the different assay designs [32].”

Reviewer: Ming Guo

1. The Pap classification was converted into TBS (Table 1). However, the converted data is not completely consistent with TBS. For example, 372 cases were classified as LSIL and HSIL (Pap IIID). It would be helpful to separate LSIL cases from HSIL cases. The classification can also be consolidated to put HISIL, CIS (2 cases) and microinvation (3 cases) into a single category.

We thank the reviewer for reviewing our manuscript and for providing us with his points of concern.

Previous publications by us (e.g. Petry, BJC, 2003) and others (e.g. Griesser, Am J Clin Pathol, 2009) have already shown that the conversion of the Munich classification to The Bethesda system is not completely congruent (especially PapIIID). A retrospective analysis of the PapIIID results in order to separate HSIL from LSIL samples is unfortunately not possible, because accessing the original LBC readings would mean to revoke sample anonymisation, which is out of scope of this purely observational study. Furthermore, we agree that the separation of the Pap groups as shown in table 1 has only limited value for readers using TBS, however, for readers using the Munich classification this information is by all means useful. To satisfy readers of both systems we have analysed the data in a way that shows both the separate Pap groups as well as combinations of these (AGUS+ and HSIL+; tables 2B and 4).
2. To compare follow up biopsy in 139 cases, it would be helpful to compare between CIN2+ vs. CIN1/benign lesions. In addition, the cases without follow up, such as cases with normal Pap (1,208) or ASC-US (65) should not be included in the comparison. Our study was purely observational therefore biopsies were only available from patients who were sent to colposcopy as part of the usual follow-up procedure. As no biopsy results from patients with Pap normal and ASC-US cytology are available, women with ≤ASC-US were excluded from the analysis. Table 4 therefore only shows results from patients with AGUS+ cytology. Our analysis is based on CIN3+ vs ≤CIN2, because only a small percentage of women in this study had histology results and CIN2 results are frequently downgraded.

3. It appears that concordance between HC2 and Cervista is low in cases with negative Pap results. High concordances observed in cases of ASC-US, ASC-H, LSIL, HSIL may reflect higher HPV load in these cases than that in cases with negative Pap. It would be interesting to have follow up data for women with Pap-/HPV + results since these women should have colposcopic evaluation and/or biopsy. Without follow up data, it would be difficult to evaluate clinical efficacy of HC2 vs. Cervista assays in the women with Pap-/HPV+ results. The same issue is also implicated in women with ASC-US.

We agree that a follow-up study of Pap-/HPV+ cases would be interesting. In Germany HPV testing is recommended for women with abnormal cytology results (PapIIw, PapIII and PapIIID), hence a Pap normal woman would not be send to HPV testing. The observational design of our study meant that women and GPs were not informed about the HPV test results and therefore these women were not invited for follow-up visits. The reviewer is right by stating that comparison of clinical efficacies of HC2 vs Cervista is difficult with our dataset and we would like to point out that a clinical comparison of both tests was out of the scope of our manuscript, which rather focusses on the comparison of relative performances within the same residual LBC sample.

4. It is very interesting to see the data of HPV genotyping in cases with discrepancy between the 2 assays. The data (Table 3B) showed high negative results in cases with Cervista HPV+ results although it is not clear what cause the false positive results in Pap specimens with negative Pap results. Clinically, HPV testing is used as a triage tool in women with ASC-US Pap results or contesting with Pap in women aged 30 years and older. Since the follow up data in the above categories is not available for comparison, the presented data can be considered an analytical comparison. The author may consider to focus on this issue since false positive HPV testing in women age 30 years and older may result in unnecessary colposcopic evaluation and biopsy. We agree with the reviewer. The high false positive rate of Cervista has been addressed by changing its cut-off values (table 5) thereby reducing the HPV prevalence rate while still detecting all CIN3+ cases.

5. In Table 2B, case number should be listed in addition to the percentage. There is a repeated
data presentation in Table 1 and Table 2. We agree that data were repeatedly presented in Table 1 and 2B. We have replaced the HPV prevalence percentages with the actual numbers in Table 1.

6. In Materials and Methods (page 6), it is not clear how the HPV testing was performed, i.e. the volume of the sample for each test. HPV testing has been performed according to the protocols provided by the manufacturers. These also include details about the exact sample volumes to be used for each test. We have now included a more detailed description in the materials and methods section as follows:

“Residual LBC samples were sent to the central molecular testing laboratory (UKT, Tübingen) within 2 weeks of collection for HPV testing. For testing by CervistaTM HPV HR assay, DNA extraction was performed on all 2ml of each ThinPrep sample. It was performed using the GenFind® DNA Extraction Kit in combination with the Cervista HPV HR assay using the ThinPrep 5000 STS (Sample Transfer System, HTA system) (from Hologic). To test for a possible risk of cross-contamination during sample transfer by the STS, we ran a pilot experiment with alternating blank samples and samples containing HPV16-positive SiHa cells and found no evidence of cross-contamination. DNA extraction was carried out according to the manufacturer’s instructions. DNA integrity was measured by PCR using primers targeting the human β-globin gene. Using the Cervista high throughput automation (HTA) system DNA samples (10µl per reaction) were then tested by CervistaTM HPV HR assay, which was performed …”

“... considered positive for HPV. 4ml of each ThinPrep sample were processed for HR HC2 testing, which was performed using the Rapid Capture System 1 (RCS-1) …”

7. In Table 4, the clinical efficacy (sensitivity, specificity...) comparison between the 2 assays may not be clinically relevant and can be deleted since women with abnormal Pap results (ASC-H+) need colposcopic evaluation and biopsy. HPV testing has not role in these women for clinical management. Clinical efficacy of HPV testing is meaningful in women with ASC-US or co-testing. However, the data is not available for review.

As earlier pointed out in our response to comment 3, a clinical evaluation of both tests was not within the scope of this manuscript. The study was intended to compare test performances within the identical sample used for cytology testing. Table 4 shows the relative test diagnostics for the detection of CIN3+ by Pap cytology group, as there was no colposcopy performed in women with ≤ASC-US (observational study), relative sensitivity/relative specificity/PPV/NPV for detection of CIN3+ could not be calculated. This lack of a gold standard was the reason for us to introduce a novel method of comparing test performances. In addition, we would like to point out that current guidelines for HPV testing greatly differ by location. In Germany guidelines recommend HPV testing in women with repeated PapIIw, PapIII and PapIIId.