Reviewer's report

Title: Glutathione S-Transferase L1 Multiplex Serology as a Measure of Cumulative Infection with Human Papillomavirus

Version: 2 Date: 23 December 2013

Reviewer: Mirte Scherpenisse

Reviewer's report:

Major Compulsory Revisions

The manuscript 'Glutathione S-Transferase L1 multiplex serology as a measure of cumulative infection with Human Papillomavirus describes the comparison between several HPV serological assays. The comparison of the GST-L1 assay with the cLIA, VLP-MIA and PBNA is very interesting. The GST-L1 assay, cLIA and VLP-ELISA are well described in literature and commonly used in seroprevalence and vaccine studies. However, all these serological assays have different characteristics with their own advantages and disadvantages which make it very difficult to choose which assay is best for a certain study population. It is interesting whether there are differences in risk factors and seropositivity determined by different serological assays. However, it is not clear to me why it would be interesting to investigate if the GST-L1 assay correlates with immune protection. It is already known that this assay detects both neutralizing and non-neutralizing antibodies just as the VLP-MIA. In contrast, the cLIA and PBNA are designed to detect neutralizing antibodies and therefore provide information about immune protection. More important is the difference between the serological assays which are all used frequently in seroprevalence studies. This study should not only focus on the comparison between the GST-L1 assay and the VLP-MIA but should also focus on the comparison with the other assays tested. There are population-based studies, vaccination studies, studies within high risk groups, but which serological assay in most appropriate to use? This is not mentioned in this manuscript.

Major point:

The comparison of the GST-L1 assay, cLIA and VLP-ELISA has not been described previously. However, a comparison of the GST-L1 assay, cLIA and VLP-MIA has been published recently (Scherpenisse et al, CVI, aug 2013). This paper is not cited in the manuscript but is definitely interesting for this manuscript and provides information on the a-specific background in the GST-L1 assay which is also observed in this manuscript in figure 2AB.

A lower cut-off in the GST-L1 assay showed a higher percentage of positive agreement with the VLP-ELISA. However, by decreasing the cut-off value in the GST-L1 assay, a higher percentage of false-positivity could be expected. Although the percentage of positive agreement increases with lower cut-off values, the percentage of false positive samples in the GST-L1 assay also
increases. This observation is very important but not mentioned in the manuscript. In figure 2 it is clear that the percentage of false-positivity in the GST-L1 assay is high and increases by lowering the cut-off values. What is the cause of this false-positivity and how are you going to distinguish between true positive results (VLP-ELISA results) and false positive results measured by GST-L1 assay?

What is the influence of false-positive results on the measurement of cumulative infection?

It is doubtful whether the GST-L1 assay is appropriate to use in studies in which naturally HPV infection is studies as it is difficult in this assay to distinguish between HPV seropositive and seronegative.

Abstract:
- In the method section it is not mentioned that the sera are also tested with the cLIA and for a selection of samples with the PBNA.
- The conclusion is not strong enough. See comment on conclusion section.

Introduction:
- The introduction is too long. To shorten the introduction you can move, for example, the paragraph on line numbers 103-110 to the discussion section.
- On line number 89-90 is stated that there is no gold standard for HPV serological assays. However, the PBNA is currently the ‘gold standard’, as this is the only assay that detects all neutralizing antibodies derived after HPV infection or vaccination.
- On line numbers 103-104 is stated that there are a few comparisons between the HPV serological assays. However, no references of these studies are included. Please, add references about the available publications that describe HPV assay comparisons.
- Line numbers 116-119: It is logical that the GST-L1 assay measures cumulative HPV infection rather than immune protection as this assay does not distinguish between antibodies directed against neutralizing and non-neutralizing epitopes (which is the same for the VLP ELISA). More important is the comparison of the GST-L1 assay with other assays.
- Line numbers 117-120: It’s true that the GST-L1 assay has not been compared directly with the VLP-ELISA. However, the correlation between the GST-L1 MIA and the VLP-MIA is published (Scherpenisse et al). The VLP-MIA is based upon the VLP-ELISA.
- Line number 123: Please, define the established correlates of natural HPV exposure.
- Line number 124: You could change ‘evaluating’ in ‘investigating’ its relationship. We evaluated the assay by evaluating….does not sound correct.
Methods:

Sample selection

It is not clear how the samples tested are collected. Stage 1 is for HPV16 seropositivity and stage 2 for HPV18 seropositivity. All samples were DNA negative at baseline. For HPV16 seropositivity 388/2814 samples were selected, but at which time point were the serum samples selected? How many samples were HPV seropositive and HPV seronegative based on the VLP-ELISA? Maybe a sample collection schedule will clarify the sample selection. In table 1, 2, and 3 you showed data of a sample collection of more 2000 sera? So you did not test 388 samples? Are this follow-up samples?

- Line numbers 151-153: How many samples were tested in the PBNA for HPV16 also 500? This is not clear described in the method section. To strengthen this manuscript I would advise to test a panel of HPV18 antibody positive sera in the PBNA if possible, as the comparison of serological assays with the PBNA makes this manuscript interesting. Already several papers are published on comparisons of HPV serological assays but comparisons with the PBNA are limited. However, I understand that testing sera for HPV18 in the PBNA is extremely labour intensive.

- Line numbers 158-166: this paragraph describes a statistical method and belongs in the paragraph ‘Statistical methods’. As the paper of Li et al is not published, more information is necessary about the method to calculate sampling weights.

- Line numbers 172-174: Luminex technology is well known information, please delete or change the sentence.

- The parts about the CV and ICC of the VLP-ELISA, GST-L1 assay, cLIA and PBNA do not belong in the method section. You could write a paragraph in the results section in which you describe the ICC and CV of all the assays tested. Why is there no ICC calculated for the VLP-ELISA? You did calculate a CV for the VLP-ELISA but this is not mentioned on line numbers 153-156.

Statistical methods:

- Line number 222: Do you mean ‘number of partners’ or ‘number of lifetime sexual partners’? Please be consistent in the manuscript.

- Line numbers 228-229: This is a result and should be placed somewhere in the result section or discussion section in stead of the method section. You have now already given away the answer on the research question you would like to answer in this manuscript.

- Line numbers 240-242: This is an explanation why your study is interesting. Don’t place this sentence in the method section but in the last paragraph of the introduction.

Results:
- Line numbers 249-257: It would be nice to see in Figure 1AB also the comparison of the cLIA and PBNA for at least HPV16. It will make the comparison with the GST-L1 assay and the other assays more clear.

- Line numbers 257-258: How much higher is the seroprevalence in the VLP-MIA compared with the GST-L1 assay? Please add this to the sentence.

- Line number 252: Is this a significant difference (medians of 40 and 30)? Please provide p-value.

- Line numbers 260-261: Seropositivity by the GST-L1 assay at baseline does not indicate a lower risk of incident HPV16 infection over time with adjustment for number of lifetime sexual partners. Have you observed immune protection against subsequent HPV infections by determining the risk of incident HPV infection for the other assays? This would be interesting for the comparison between the GST-L1 assay with the other serological assays. Did you observe immune protection with the cLIA and the PBNA and did you observe no immune protection with the VLP-MIA and the GST-L1 assay?

- Line number 262-263: Have you observed immune protection in the higher quintiles of the VLP-MIA?

- Line numbers: 264-274: Start this paragraph with the total percentage of agreement between both assays (GST-L1 assay and VLP-ELISA) and than the total percentage of discordant results. Than you can describe the different cut-off values you tested and the influence of the different cut-off values on the percentages of agreement and discordance.

- Line numbers 293-301: Start this paragraph with the total percentage of agreement between both assays for HPV18 (GST-L1 assay and VLP-ELISA) and than the total percentage of discordant results.

Discussion:

General comment: In the discussion section it is not necessary to add the figure and table numbers as the tables and figures are already described in the results section. In the discussion section I miss a discussion about your data. You mostly compare you’re data with other data available but the impact of your data is not enough described.

Line number 304-308: I do not agree with these statements. What is the added value of the GST-L1 assay as a marker of cumulative HPV16/18 infection? If we take into account the high a-specific background in the assay when using a lower cut-off value or the detection of low seroprevalences in a study due to a too high cut-off value, which is necessary because of the a-specific background, can we use this assay for the detection of HPV antibodies after natural infection in which antibody concentrations are generally low?

Although the positive agreement percentages increases with lower cut-off values, the percentage of discordant results, especially samples seropositive in the
GST-L1 assay and seronegative in the VLP-MIA negative results. This is very important for interpretation seroprevalence data, but these problems are not mentioned in the manuscript. The VLP-MIA can better distinguish between seronegative and low positive results.

Line number 323: Are there other papers describing the relation between antibody concentration and immune protection? If there are, add some sentences in which you describe other research about the association between the level and the neutralizing capacity of naturally induced antibodies among naturally infected persons. Did you observe immune protection with the VLP-MIA?

Line number 337: is there a difference between neutralizing and protective antibodies? Please change the sentence in which you use ‘neutralizing’ or ‘protective’ antibodies.

Line number 338-339: I don’t think that it is correct to say that a higher seroprevalence is expected by the VLP ELISA due to a lower cut-off value between the two assays. First, the antibody concentrations in the VLP ELISA are expressed in EU/ml and in the GST-L1 assay in MFI, which means that you can’t compare both cut-off values. The calculated cut-offs are specific for both assays. Second, with the GST-L1 assay seroprevalences are low because a high percentage of seropositive samples can’t be separated from the background in the assay. I think that between different assays with comparable assay characteristics the calculated seroprevalence should be comparable independent of the cut-off value.

Line numbers 346-350: Next to the serological assay used, study design could also have a great influence on seroprevalence data.

Line numbers 351-356: Currently, the lack of a universal reference standard is more important than universal cut-off values. The cLIA express antibody concentrations in mMU, the VLP-ELISA in EU/ml, and the GST-L1 assay doesn’t use a reference standard. This hampers comparisons between seroprevalence studies.

Line number 357-362: Do not focus on the cut-offs so much. The problem goes beyond the cut-off values. The cut-off values of the assays are based upon the best distinguishing between seronegative and seropositive. As all the assays have different characteristics also the cut-off values will differ between the assays. Important is, as also mentioned in the previous point, to use a reference standard with a known antibody concentration preferably for more HPV types, that can help with the comparisons between assays. What can we do to make better comparisons between assays? For example, we could use quality control panels and a universal reference standard.

Conclusions:
This is not a conclusion but a summary of your results.
Please, give also advice about which assay can be used for what kind of studies. What are the advantages and disadvantages of each assay? For example: a
disadvantage of the GST-L1 assay is the cross-specific background in the assay which could hamper the measurement of low antibody concentrations. However, the advantage of this assay is that antibodies against many HPV types can be measured rather easily. A disadvantage of the cLIA is that it can only measure antibodies directed against one epitope. An advantage of this assay is that we can measure neutralizing antibodies very easily.

Tables/Figures:
-Figure 1AB: Show also the results of the cLIA and PBNA. This is a good point in study, as, as far as I know, nobody has compared risk factors within one study between several serological assays.

-Figure 2AB: This is an important figure for the comparison between GST-L1 assay and the VLP-MIA. You can see clearly that with the lower cut-off value a high percentage is seropositive in the GST-L1 assay, but seronegative in the VLP-ELISA.

Although a lower cut-off value in the GST-L1 assay improves the correlation of positive agreement, but you forgot to describe (in the result section) what happens with the total percentage of discordant results.

- Table 1: This table is not clear. In the title is written that you tested 488 women, but in the table you show data of 2786 women. The values under Min, Max, Mean are those in MFI? This is not clear. What does 25%, 50% and 75% mean? I could not find these data in the results section of the manuscript.

Mean and Geomean are two different things. Why haven’t you used ‘GM’ in stead of ‘Mean’? You can describe in the subtext ‘GM = Geomean’.

- Table 2: This is an important table in which you show whether there is a correlation between incident infection and GST-L1 seropositivity. However the table is not designed properly and is not so neat.

-Combine table 3 and 4. In table 4 the seroprevalence columns can be deleted and can be described in the results section. Also for table 3 it is not clear how many samples were tested.

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 declare that I have no competing interests