Author's response to reviews

Title: Evaluation of using composite HPV genotyping assay results to monitor human papillomavirus infection burden through simulations

Authors:

Carol Y Lin (cvl6@cdc.gov)

Version: 5 Date: 25 January 2015

Author's response to reviews: see over
Dear Editor:

I would like to submit the attached manuscript, “Evaluation of using composite HPV genotyping assay results to monitor human papillomavirus infection burden through simulations” (MS: 5555441311471687) by Carol Lin for your reconsideration for publication in *BMC Infectious Diseases*.

I also would like to thank the three referees (Drs Xianhong Xie, Eric Chow and David Ragan) for the thoughtful and constructive comments. The point-to-point responses to the referees’ comments are given in the following pages. Attached please find one copy of the revised manuscript and. I hope that the revision and responses have adequately addressed the referees’ comments.

The submitted material has not been published and is not under consideration for publication elsewhere. Thank you for your consideration and look forward to hearing from you.

Sincerely yours,

Carol Y. Lin
Responses to referees’ comments

Referee 1: Dr. Xianhong Xie

1. Page 5, lines 97-99. Please discuss when the ratio would be greater than 1. Too many false positives? Also please mention when the ratio would be equal or less than 1. The ratio of estimated-to-true prevalence ratio is a poor measure of accuracy. In the extreme case (especially when the true prevalence is low), those who test positive could be completely different from those who are true positives, and the estimated-to-true ratio could still be equal to 1. The composite test sensitivity, specificity, ROC curve, and AUC are better accuracy measures than the estimated-to-true ratio. Some or all of the latter measures should be presented with the estimated-to-true prevalence ratios.

In the method section, on page 6, line 130, the following sentences have been added. “A ratio greater than 1 means composite prevalence estimate calculated based on a panel of genotyping assay results over-estimates the true underlying composite infection burden. The number of false positives exceeds the number of false negatives. In contrast, a ratio less than 1 means composite prevalence estimate calculated based on a panel of genotyping assay results under-estimates the true underlying composite infection burden. The number of false negatives exceeds the number of false positives. “

As the reviewer pointed out, the estimated-to-true prevalence ratio has its limitation. In the discussion, on page 13, line 284, the following sentences have been added. “Estimated-to-true prevalence ratios have been used to examine how well the prevalence estimates based on genotyping assay results measure the true underlying infection burden. One limitation is the ratio does not distinguish true and false positive rates. True and false positive rates depend on the type-specific infection burdens and genotyping assay sensitivity and specificity.”

Combined sensitivity, specificity and ROC curve have been used to measure performance of diagnostic systems with various individual tests and to select an optimal diagnostic system. A likelihood approach was proposed to derive combined sensitivity and specificity of a diagnostic system with various binary tests. The impact of number of individual tests and correlations between individual tests on the combined sensitivity, specificity and ROC curve was evaluated (Lin et al., 2012).

When genotyping assay is for the use of detecting cervical cancer, clinical sensitivity and specificity of genotyping assay are going to be different. (HPV 16 and HPV 18 are generally considered particularly high-risk genotypes and account for approximately 70% of invasive cervical cancers globally.) To select a diagnostic system with various HPV genotyping tests to detect cervical cancer, the composite sensitivity and specificity and ROC curve are important. Combined sensitivity, specificity and ROC curve will allow us to identify the subset of HPV which can better detect cervical cancer.

In this manuscript, genotyping assay is for detecting the existence of type-specific infection, analytical sensitivity and specificity of genotyping assay are considered to be the same. The objective of this manuscript is to examine how well the prevalence estimates based on genotyping assay results measure
the true underlying infection burden. The settings are different from detecting cervical cancer. The composite sensitivity, specificity and ROC curve might not be as important.

- The decision rule (cut-off point) is set. A positive outcome has been defined as having at least one positive HPV in the composite measure. Since the cut-off point is set, the composite sensitivity and composite specificity is just a point on the ROC curve.
- Composite sensitivity is a function of genotyping assay sensitivity; the composite specificity is a function of genotyping assay specificity. Since the decision rule is set, given the number of individual HPV tests in the composite measure, the composite sensitivity increases/decreases when individual genotyping assay sensitivity increases/decreases. Similarly, the composite specificity increases/decreases when individual genotyping assay specificity increases/decreases.
- The number and HPV types in the composite measure cannot be varied based on the composite sensitivity and specificity, although composite sensitivity and specificity vary with numbers of HPV included in a composite measure. The number and HPV types in the composite measures are grouped based the association with a variety of clinical conditions (i.e., cervical cancer, warts), phylogenetic position and types related to vaccines to form composite measure and may not be varied.

I agree with the reviewer that in the extreme case when prevalence, assay sensitivity and specificity are very low, those who test positives could be completely different from those who are true positives. When forming composite measure based on PCR-based tests, the extreme case is less of a concern.

- PCR-based genotyping assay is able to catch a good portion of true positives, even when the true prevalence is low. Analytical sensitivity of PCR-based assay is high. PCR-based genotyping assay can detect small amount of viral DNA.
- The composite prevalence is generally higher than type-specific prevalence. Since the outcome measure is defined as having at least one infection, the composite prevalence generally increases with increasing number of HPV types in the composite measure.
- Population prevalence is more of the interest when measuring infection burden. PCR-based genotyping assay has been used to monitor infection burden for research purpose. Subjects who test positive generally are not informed or referred for colposcopy.

2. Page 5, line 112, does the author mean that the correlation between each pair of HPV types among the 40 or so HPV types is assumed to be equal to 0.4? It seems unrealistic that all the correlations are equal to a specific value because the HPV types have genetic grouping, some HPV type pairs are more correlated than the others. It would be better that the correlations between the HPV types are estimated from the US data or Canada data, which would make the simulated data more realistic.

Thanks for pointing this out. This is a great suggestion. The simulation results have been updated using correlations of 0.05 to replace 0.4 to reflect more realistic scenarios. The correlations are now estimated using the data collected in the 2003-2006 US NHANES study. The mean of pair-wise correlations of 37 HPV types is 0.05 and the correlations vary from 0 to 0.3. The 25th, 50th, 75th, 85th and
95th percentiles are 0.007, 0.034, 0.077, 0.100 and 0.157, respectively. Eighty-five percent of the pairwise correlations are in between of 0 and 0.1. In addition, the results of the newly added sensitivity analysis suggest the impact of correlations is limited (Table 5). Since the variation between pairwise correlations is relatively small, the impact of correlations is limited, and the correlation matrix (74x74) of 37 HPV types used to generate the data is relatively complex, the mean of pairwise correlations, 0.05, is used as the correlations been types in this simulations study.

3. Page 6, lines 119-123. Please explain how the sensitivity and specificity pair (0.95, 0.95) was chosen in the simulation section instead of in the discussion section. Also, how does the pair compare to the typical setting in the routine HPV testing?

As suggested, the explanation has been added in the simulation section (page 6, line 138).

The analytical sensitivity of PCR-based assay is generally higher than HPV-clinical tests. Digene HC2 has been suggested as a more cost-effective cancer screening alternative than Pap testing in the developed countries [Kok et al, 2012] and a more plausible screening approach in the developing countries [Flore et al, 2011; Zao et al, 2010]. HC2 uses the signal amplified hybridization method and is designed to detect “high” enough (clinically relevant) viral loads. In contrast, the PCR-based test is able to detect a tiny amount of viral DNA. Analytical sensitivity of Digene HC2 is at the pictogram level of HPV DNA and Roche LA assay is at the sub-picogram level [Molijn et al., 2005; Snijders et al., 2003]. Even the commercial versions of HPV genotyping assays (e.g., Linear Array (LA) genotyping test, INNO-LiPA HPV genotyping v2) are only approved for research use not to screen subjects for clinical use in the U.S. Subjects with positive LA genotyping test results in the research studies generally are not informed or referred for colposcopy, since most of these infections can be cleared by the immune system within 2 years.

4. Page 6, line 126. Does the author mean in the reduced rows in the Tables 1a-2b, only the prevalence of these 4 HPV types: 6, 11, 16 and 18 were reduced by 50%, while the prevalence of all the other types was unchanged? This seems to contradict with the correlation of the HPV types.

Yes, for demonstration purpose, type 6, 11, 16 and 18 were reduced by 50%, while the prevalence of all the other types was unchanged. Since Quadrivalent HPV Vaccine (HPV4, Gardasil) protects against HPV 6, 11, 16 and 18, if the vaccine coverage rates are high, the infection of type 6, 11, 16 and 18 should be reduced significantly. The reduction due to sexual behaviors change is not considered.

5. Page 7, lines 143-145. Has the author considered increasing the specificity to more than 0.95 and lower the sensitivity to less than 0.95 and see the effect on the estimation of the composite prevalence? A better way of choosing the sensitivity and specificity pair is to select from the ROC curve.
Taking genotyping assay with sensitivity and specificity equal (0.87, 0.98) as an example, based on the simulations results in this manuscript, we know that the magnitude of overestimation would be lower than the scenarios when genotyping assay sensitivity and specificity equal (0.95, 0.95) is used. The simulation results suggest that increasing genotyping assay specificity reduces the false positives and reduces the magnitude of overestimation. In addition, genotyping assay sensitivity has limited impact on the composite prevalence estimates. To verify, a simulation is done based on the scenario with genotyping assay sensitivity and specificity (0.87, 0.98) and the pre-specified “true” underlying type-specific infection is distributed as in the US. As expected, the composite prevalence and the magnitude of overestimation are lower when genotyping assay sensitivity and specificity (0.95, 0.95) is used. The composite prevalence estimates are 0.761, 0.421, 0.173 and 0.093 and the estimated-to-true prevalence ratios are 1.23, 1.49, 1.44 and 1.45 for the composite measures of the 37 HPV types, 14 high-risk types and 4 vaccine types and 2 high-risk vaccine types, respectively.

6. Page 7, line 155-157. It appears that the problem is there are too many false positives. It becomes worse when the “true” prevalence is decreased. The assay sensitivity is too high, but the specificity is not high enough. If one increases the specificity and decreases the sensitivity, the problem of overestimating the composite prevalence could be alleviated.

This is a good point. There has been a tendency either to develop a new testing technique or to modify existing techniques to increase analytical sensitivity to detect HPV. One of the messages from this study is that Assay specificity is as or more important than sensitivity in monitoring HPV infections. Therefore, eliminating factors which might cause false positives (i.e., contamination introduced through reagents, laboratory disposables or equipment including carry-over contamination between tests or sample-to-sample contamination, etc.) to increase assay specificity is important.

7. Page 9, lines 203-205. Please explain how the fact that “in the NWT region, the ‘true’ type-specific infection … in the composite measure” causes that “increasing number of HPV types in the composite measures does not always help to ease the overestimating problem in the NWT, Canada” (lines 200-201). It appears that the lower HPV prevalence in NWT, Canada compared to that in US plays a dominant role.

On page 11, line 236, the following texts have been revised to clarify:
“Unlike the US scenario, increasing number of HPV types in the composite measures does not always help to ease the overestimating problem in the NWT, Canada. It is because in the US scenarios, the type-specific infection burdens of newly-added HPV types are at similar or higher levels than those already in the composite measure. In the NWT, Canada, the type-specific infection burdens of newly-added types are much lower than those already in the composite measure (Figure 1), therefore, the magnitude of false positive rates increased is much greater.”

8. Page 15, lines 346-347. Please discuss that the genotype testing on all the HPV types has the same sensitivity \( \alpha \) and specificity \( \beta \). The HPV types have different true prevalence; for a given test, one usually cannot control both the sensitivity and specificity simultaneously. But by choosing the threshold differently, one can easily achieve higher sensitivity and lower specificity or lower sensitivity and higher specificity.
In the appendix, p.17 line 392, the paragraph is revised as following:
“Suppose that n independent subjects, i=1,2,...,n, are in the study. For each subject, i, K various HPV types, j=1,2,...,k, are tested. Let δ_{ij} denote the true type-specific prevalence of the j-th HPV type and \( \delta_s=(\delta_{i1}, \delta_{i2},...\delta_{ik}) \) be the true type-specific prevalence for k various types. The true type-specific prevalence rates are considered to be different. \( \alpha \) and \( \beta \) are the analytical sensitivity and specificity of genotyping assays. Genotyping assay performance is considered to be the same for all types.”

In the simulation study, the latent vectors for type-specific infections and genotyping test results are generated from multivariate normal, \( N(0, \Sigma_k) \). The correlation between \( Y_j \) and \( Z_j \) allows us to obtain the data with desired genotyping sensitivity and specificity after dichotomized.

9. Page 15, line 353. Although the correlation matrix \( \Sigma_k \) may have been specified in the reference Lin et al. It would be better to specify the matrix for this manuscript, i.e., giving the actual correlation values.
On page 18, in the end of appendix, the following text has been added to clarify the simulation method.

“To generate the data, for each type, we first identify the threshold values of \( p_j \) and \( q_j \) based the pre-specified values of \( \delta_j, \alpha, \beta \). Second, we identify correlation matrix \( \Sigma_k \) (where correlation between each pair of \( Y_j \) and \( Z_j \) is obtained by a numerical method). Multivariate normal data \( (Y_{i1},Z_{i1},...,Y_{ij},Z_{ij},...,Y_{ik},Z_{ik}) \) are generated with mean zero and correlation matrix \( \Sigma_k \). Then, for each HPV type, we use the pre-determined threshold values of \( p_j \) and \( q_j \) to dichotomize \( Y_{ij} \) and \( Z_{ij} \) to obtain \( D_{ij} \) and \( T_{ij} \).

To illustrate, when \( \alpha=0.95, \beta=0.95 \) and \( \delta_j=(0.0283, 0.0339, 0.0469, 0.0185) \),

\[
\Sigma_4 = \\
\begin{pmatrix}
1 & 0.91 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.92 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.93 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.89 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & & & & & & & 1
\end{pmatrix}
\]

is used to generate the laten vectors.”

10. Page 19, Table 1b, the estimated-to-true ratios are missing for the reduced prevalence row under No. of HPV types = 37.

Thank you for pointing this out. It is corrected.
References:


Referee 2: Dr. Eric Chow

The method section is not clear and authors need to explain it in details.

The method section has been reorganized and revised. Simulation set up, “true” composite infection statuses, “true” composite prevalence, positive composite test result and composite prevalence are defined in the method section. Hopefully, the clarity has been improved now in the manuscript.

1. Please explain how the estimated-to-true prevalence is calculated.
   Estimated-to-true prevalence ratio is defined as estimated prevalence divide by pre-specified “true” prevalence ratio.

2. It is not clear what the ‘true’ prevalence is referred to.
   The following paragraph has been added in the method section (p. 6, line 120) to define ‘true’ prevalence.

   The prespecified “true” prevalence of each outcome measure is defined as the proportion of subjects with a positive infection status. For each subject, based on the pre-specified “true” type-specific infectious statuses, the “true” composite positive infection status of the four outcome measures is defined as having at least one HPV type-specific infection of the 37 HPV types (45 types in Canada), 14 high-risk types (22 types in Canada), 4 vaccine types, and 2 high-risk vaccine types.

3. Page 6. Please explain why the six sets of specificity and sensitivity were used.
   In the method section (p7, line 139), the text has been revised as following, “To examine the effect of genotyping assay specificity, assay sensitivity was held unchanged and specificity was reduced from 0.95 to 0.90, 0.85 and 0.80. Similarly, to examine the effect of genotyping assay sensitivity, assay specificity was held unchanged and sensitivity was reduced from 0.95 to 0.90, 0.85 and 0.80.”
1. More detail is required in the methods section on how the datasets are generated, how the simulations are run, and how correlation is applied. How the figure of 0.4 for correlation between types was arrived at also needs to be explained and, if necessary, referenced.

The method section has been reorganized and revised. Simulation setup, “true” composite infection statuses, “true” composite prevalence, positive composite test result and composite prevalence are defined in the method section. Simulation results have been updated using correlation of 0.05 which is estimated from the 2003-2006 US NHANSE data. Furthermore, the following text is added in the appendix to further clarify the simulation method.

“To generate the data, for each type, we first identify the threshold values of \( p_j \) and \( q_j \) based the prespecified values of \( \delta_j, \alpha, \beta \). Second, we identify correlation matrix \( \Sigma_k \) (where correlation between each pair of \( Y_j \) and \( Z_j \) is obtained by a numerical method). Multivariate normal data \((Y_{i1}, Z_{i1},..., Y_{ij}, Z_{ij},..., Y_{ik}, Z_{ik})\) are generated with mean zero and correlation matrix \( \Sigma_k \). Then, for each HPV type, we use the predetermined threshold values of \( p_j \) and \( q_j \) to dichotomize \( Y_{ij} \) and \( Z_{ij} \) to obtain \( D_{ij} \) and \( T_{ij} \).

To illustrate, when \( \alpha=0.95, \beta=0.95 \) and \( \delta_j=(0.0283, 0.0339, 0.0469, 0.0185) \),

\[
\Sigma_k = \begin{pmatrix}
1 & 0.91 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.92 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.93 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.93 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\
1 & 0.91 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05
\end{pmatrix}
\]

is used to generate the laten vectors.”

2. 500 data sets were generated for each scenario yet only point-values are given in the results tables. It should be specified what the point-values represent (i.e., mean, median) and a measure of spread (such as standard deviation or interquartile range) for each value should be provided.

In the method section (p), the following text is added, “For each scenario, 500 data sets were generated. Mean and standard deviation of the “true” and estimated composite prevalence of 500 data sets were calculated. Mathematical algorithms related to the simulation setup are given in the Appendix.” In addition, standard deviations for each values have been added in Table 1 (1a, 1b), Table 2 (2a, 2b).

3. The estimated-to-true ratios are missing for the 37/reduced scenario in Table 1b.

Thank you for pointed it out. It has been corrected now.

4. On lines 187-193, the importance of correlation is discussed but results are not shown. This is an important sensitivity analysis and the results should be given in the results section.
Sensitivity analysis for correlations is added in the method section. The results are given in Table 3.

5. The manuscript would benefit from a thorough review of the language and grammar. While the manuscript is generally clear, I found the numerous grammatical errors a distraction.

Thank you for the suggestion. The manuscript was reviewed and edited. Hopefully, the mistakes have been corrected.

6. It is not clear what the reductions of 8%/5%, 8%/4%, 10%/8% are in the tables. This is not described in the methods where only the 50% reduction is mentioned and is not referred to in the Results. These should be described, including what these numbers are based on, or omitted.

For demonstration purpose, type 6, 11, 16 and 18 are reduced by 50%, while the prevalence of all the other types remain unchanged. Since Quadrivalent HPV Vaccine (HPV4, Gardasil) protects against HPV 6, 11, 16 and 18, if the vaccine coverage rate is high, the infection of type 6, 11, 16 and 18 should be reduced significantly. Since only the 4 vaccine types were reduced 50%, when forming composite measures, the reductions in the composite measure with 14 high-risk types or 37 types are much lower.

7. There are numerous references to “we” in the paper (e.g., “we perform” on line 79) but only a single author is listed on the manuscript. If there were others involved in the study they should be acknowledged, otherwise different language should be used.

Thank you!! It is been corrected.

8. I’m not sure I agree with the author’s contention that individuals with weaker immune systems are more likely to get infected” (lines 66-67 and 188-189). In my opinion, the probability of becoming infected is related primarily to the individual’s level of sexual activity and the prevalence of infection in the population. The rate of clearance, on the other hand, or the probability of infection persisting is likely related to immunity. Furthermore, as far as I can tell, the model used here for the simulations makes no distinction between individuals in terms of risk factors for infection and the degree of correlation between types is applied as an average measure.

I agree with the reviewer that sexual life style and prevalence are important factors for getting infected. I was thinking when subjects are exposed to the viruses, subjects with strong immune systems are less likely to be infected than subjects with weaker immune responses. I think that the reviewer’s point that subjects with weaker immune system are less likely to clear infections is a greater point. Persistent infection of HPV high-risk types is an important factor for cervical cancer.

The confounding factors were not discussed in this manuscript. It is the limitation of this study. Hopefully, it will be address in the future study.
The text has been revised as following: “Type-specific HPV infections are considered to be correlated in this simulation study because the risk factors of getting infected by various HPV types are similar and subjects with weaker immune systems are more likely to get infected and/or stay infected.”

9. It seems strange to take US (or Canadian) reported infection burden as the true prevalence when these estimates were obtained using the very assays that are being evaluated in this study. Perhaps it would have been better to use a true prevalence that produces the observed US or Canadian estimates when an assay with baseline characteristics of 0.95/0.95 for sensitivity/specificity is applied in the simulation.

The reported type-specific prevalence in the U.S. was based on Roche Linear Array genotyping assay results. The reported prevalence in Canada was based on an in-house PCR assay which has been shown to be comparable to Roche Linear Array [Jiang et al, 2011]. If PCR process is well performed, the performance of these assays should be able to reach (0.95, 0.95).

These reported prevalence was chosen so that the simulation study can reflect some levels of realistic scenarios. More importantly, I hope that a trend can be found based on these scenarios and the results can be generalized. For example, the simulation results suggest that the magnitude of over-estimation increases when the “true” type-specific prevalence rates decrease. Therefore, we know if the “true” infection burdens are higher than what was specified in this simulation study, the magnitude of over-estimation should be lower.

10. On lines 206-221, the issue of sensitivity versus specificity is discussed. At a population level, it is infectiousness that is important for onward transmission and ultimately the burden of infection in the population. PCR, being a very sensitive methodology, is also likely to overestimate prevalence from the point of view of infectiousness. A very tiny amount of DNA detected is likely not infectious or may represent a past infection that is no longer viable or may simply be evidence of recent sexual activity. I think it is perhaps worth noting this point in your discussion.

This is a very good point. Thank you for the suggestion. The following text has been added in the discussion section (p12, line 269), “Particularly, subjects with positive HPV detected by PCR in research studies generally are not informed or referred to have colposcopy, since persistent infection of HPV high-risk types is the pivotal event in the development of cervical cancer and most of HPV detected can be cleared without treatment in about two years. In addition, on the population level, from the point of view of infectiousness, population prevalence calculations based on PCR results could overestimate true infectiousness burden, since a very tiny amount of DNA detected by PCR is likely not infectious and may just represent a past infection.”

Reference