Author's response to reviews

Title: Is HPV-16 seropositivity a correlate of immunity? Insights from compartmental models

Authors:

Igor A Korostil (ikorostil@kirby.unsw.edu.au)
Suzanne M Garland (suzanne.garland@thewomens.org.au)
Matthew G Law (mlaw@kirby.unsw.edu.au)
David G Regan (d.regan@unsw.edu.au)

Version: 4 Date: 30 November 2012

Author's response to reviews: see over
Response to Reviewers

My co-authors and I thank the editor and reviewers for their efforts in providing this review and feedback to improve the paper

**MS: 1757886043755892**

Is HPV-16 seropositivity a correlate of immunity? Insights from compartmental models.

We have made substantial changes to our paper. In particular, we selected prior distributions for model parameters more carefully, introduced separate rates of seroreversion for males and females, where applicable, and in general, made an effort to present our (updated) results in a clearer and more detailed way.

As it is stated in our paper, our approach was to use a number of simple models of HPV-16 transmission to test which implementation of seroreactivity would produce the closest fit to the available data using the minimal number of parameters. We intentionally chose to compare parsimonious models to potentially increase the explanatory power of parameters and consequently, increase likelihood of being able to arrive at meaningful conclusions. Given that the current data describing HPV-16 transmission can not be seen as particularly accurate (for example, even the reported durations of infection are rather contradictory, “short” durations around 1 year, or “long” ones, closer to 2 years, not to mention durations of immunity we know almost nothing about), having fewer parameters in the models is crucial. Hence, we tried to parameterise the models so that they implement simpler explanations suggested by the literature, even if the competing more complex explanations are supported by more published studies. For example, we do not make model parameters age dependent. As one of the reviewers pointed out, the advantages we gain by such an approach may be somewhat offset by a degree of plausibility achieved, which may not be completely up to what one would expect from a typical HPV transmission model used these days in various modeling studies. While we agree with this, we believe that our approach is sufficient to provide insights into the relation between seropositivity and natural immunity based on the presently available Australian data.

Please allow us to address each of the suggestions and to illustrate the work we have done to address these comments in the following response letter.

We have included below the comments from the reviews and the changes made to address these comments.

Reviewers’ comments (preserved in their original form) are in blue while responses are in black standard font.

**Reviewer: Zoe R Edelstein**
Major Compulsory Revisions

1. The authors use 8 disease transmission models in total. Clearly a lot of thought went into the construction of these models and the authors describe their components in the Background and Methods section, as well as in Table 1, Figures 1-3 and Appendix. However, I still had some trouble following these explanations and how the models differed from one another. Some suggested revisions and clarifications are listed below. Though some are minor, taken together they appear to be a major revision:

a. The definition of “+” as seropositive should be added to the methods section text

As suggested by the reviewer, we have clarified this in “Natural history of HPV-16”.

b. I found Table 1 to be the most effective guide for understanding the differences between models once the number of model types went from three (SIS, SIR and SIRS) to eight. I suggest referencing Table 1 earlier in the text and describing the differences between models with reference to it.

As suggested by the reviewer, we rearranged the text to first reference Table 1 and then proceed with explanations of the differences between models. We have also improved the table to make it more informative.

c. ... However, with regard to Table 1, “immunity” against infection, as it appears to be defined (more this below) does not seem to apply to SIRS3 and SIRS4, unless I misunderstanding the loops R+ to S+ to I+.

The reviewer is right, we corrected this error in Table 1.

d. For the list of parameters in Table 2,

i. My understanding is that betas come be represented as lambda in the technical appendix and in Figures 1-3. This needs to be clearer either by footnote or other edit, as do other differences in definition between the list of coefficients and Table 2 (e.g. risk reduction versus degree of protection; probability of transmission versus force of infection)

Betas are used to calculate lambda (the force of infection). We have added a section with definition of the force of infection to the technical appendix (“Calculation of the force of infection”).

ii. Are the durations in years?

Yes, they are in years. This has been clarified in Table 2 caption.

2. It should be stated if the DIC value is based on calibration to the serology data, the DNA detection data or both, as well as either or both genders. It appears there is only data for DNA prevalence for women aged 15-39 (based on the results and the cited article). This should be clarified in the methods and if used in the DIC value please state how the data limitations for DNA prevalence limit the DIC estimation.

We acknowledge data limitations mentioned in the comment, namely, (1) we have no DNA prevalence data for men; (2) the available DNA prevalence data is only for women aged 15-39;
clarifications have been added to “Model comparison and calibration”. We have also addressed this in “Discussion”.

3. The authors compartmentalize age and gender. There are some limitations to their models regarding age (regarding serology parameters), but these are acknowledged. For gender they state that they assume the natural history of infection to be the same in men and women, though rates of transitions may differ. This may be a strong assumption especially when comparing SIS to SIRS models, based on the little that is known about HPV natural history regarding serology and also simulated serology results in Figure 4. Unless I am mistaken, the authors have the ability to use their data to test this assumption (e.g. a best fit by DIC for each gender).

The same natural history was not assumed – we have clarified this in “Natural history of HPV-16”.

DIC was calculated for calibration to all available data at once (i.e. to seroprevalence for both males and females and DNA prevalence for females 15-39 y.o. only). We have mentioned this in “Model comparison and calibration”.

It is possible that this is what is being referred to in the sentence on page 14 that begins “Unfortunately...” (Either way, the results from which this statement is derived should probably be made more clear.)

This sentence is a part of discussion of the best fitting SIS models which incorporated degrees of immunity (Table 2). We initially assumed that females are protected to some extent (risk reduction coefficients for them was at least 0.4). Regarding males, we let that coefficient take any values from 0 (no protection at all) to 1 (full protection). Hence the differences between females and males in natural history would be apparent from the results of the calibration procedure.

However, it turned out that given the data available for calibration (see above), protection for males could not be determined: the posterior distribution for risk reduction coefficient for males was effectively as its prior.

This shows that we didn’t have enough observational data and/or more specific HPV natural history data to be able to detect the level of protection for males via the calibration procedure.

Note that in the revised version of the paper we simply mention a flat posterior distribution for the degree of immunity for males (in “Results”), which represents the same conclusion.

4. I am not entirely sure that the results support the conclusion that seropositivity and immunity are completely “decoupled,” especially for women. In the model with the best fit (SIS2), the degree of protection inferred is between 0.7 and 0.9 for women, suggesting a high probability of immunity. This high degree of protection is explained in the discussion, but not the abstract or conclusions. It appears that the authors define immunity as it exists in the SIR model: 100% (at a population level) and as lifetime immunity. Without some tempering statements regarding the lack of relationship between seropositivity and immunity (for example mentioning the degree of protection or using a term such as “lifetime immunity”), their conclusions could be misinterpreted.

We have thoroughly revised the paper to make sure such conclusions prone to misunderstandings are no longer present in the text.

5. The authors should discuss the possibility of re-activation of infection as opposed to re-infection. They should also discuss any limitation of using prevalence of infection (by DNA) versus a measure
of incidence, as well as only measuring DNA prevalence at the cervix.

While we agree that possible re-activation as an alternative to reinfection may be the key to better understanding of natural history of HPV, discussion of this subject is beyond the scope of our paper, since we do not include any models incorporating reactivation in the comparison. The reason for this is that at the moment, reactivation mechanism is not well understood, and the very existence of reactivation as such is arguable. To implement reactivation we would have to consider a number of possibilities, such as at what age it should occur (women over 40?), what would be prerequisites for reactivation (age alone, or sexual history, or something else?). Then there is the question of reactivation in males. Should we look at sexual history, the compartmental models we use in this study would not be well suited for the task, and also we don’t have temporal sexual behavior data. In view of these numerous difficulties, implementation of reactivation would require a different modeling setup and many uncertain parameters to be introduced, which is not a feasible task from the modeling perspective we adopt in this paper. Clearly, reactivation will be a subject of future research. We acknowledge that we do not aim to perform a comprehensive analysis of plausible HPV natural history interpretations in this study and discuss only selected scenarios. We think that the suggested discussion on DNA prevalence is very important, but regardless of the implications of such discussion, we still have to use whatever data available, which is only seroprevalence and DNA prevalence for females 15-39 y.o. Effectively, our results only valid under assumption that we trust these data. However, we agree that a reader should to be explicitly informed about possible limitations related to the data.

We mentioned the data limitations in the text (“Model comparison and calibration” and “Discussion”).

Minor Essential Revisions

1. On page 4, the sentence that starts “However, in view...” is quite confusing. Please either restructure the sentence or split it up.

We followed the reviewer’s suggestion.

2. On page 8, the sentence that begins “Differences...” is also confusing. Consider cutting the text from “in” to “models”

We followed the reviewer’s suggestion.

Discretionary Revisions

1. In the abstract, the authors should consider mentioning that they included both men and women, as many readers will want to know this

We have added the suggested clarification to the abstract.

2. The text on the 60+ category seems non-essential unless the authors want to discuss Burchell et al’s review (Vaccine, 2006) that showed HPV DNA prevalence goes up in the older women w/o adjustment for age (but not with adjustment)

The text on the 60+ could be placed in the appendix. However, we have it in the main paper in order to clarify why our models do not include the 60+, which may not be obvious to a reader.

We have reworded and abridged the relevant text.
Reviewer: Mark Jit

Major compulsory revisions

1. p. 6: It's not clear to me why excluding over 60s reduces the dimensionality of the model since the age-dependent sexual matrix is already data defined, and it hence would not require any additional variables to incorporate this age group. On the other hand, excluding this group removes an additional data point to which the model could be fitted to so would appear to increase rather than reduce overspecification.

In our paper we used the sexual behavior data derived in (Regan, Philp, Hocking, & Law, 2007) from the results of the Australian Sexual Health and Relationships (ASHR) study. This study covered only 16-59 year olds. Therefore, the sexual mixing matrix with 60+ is not data defined. We could assume that 60+ as not more sexually active than the 45-59 year olds (the least active age group according to (Regan et al., 2007)). This, however, would have to be implemented either as a fixed partner change rate selected to be lower than that for the 45-59 y.o., or a variable partner change rate. The latter would certainly be preferable given the uncertainty associated with the sexual behavior in 60+.

Indeed, by including the 60+ in the model we would be able to fit the models to an additional data point, namely, seropositivity for the 60-69 reported in (Newall et al., 2008). No new data points would be available for HPV DNA prevalence in females because the WHINURS study covered only women aged up to 40.

To summarize, adding the 60+ would result in 1 extra variable and 1 extra data point to calibrate the models to. While this appears a valid alternative to having only individuals under 60 in the models, we took into account the fact that acceptability of assuming no mortality in the models (which is a helpful simplification) would seem debatable with the 60-69 year olds incorporated.

2. p. 7: Is the Australian population structure between 12 and 60 also uniform (in addition to age-specific mortality)? If not, then the birth and death process described would not adequately capture it.
The figure presented above is available on the web site of Australian Bureau of Statistics ([www.abs.gov.au](http://www.abs.gov.au)) along with other information relevant to this question. Considering that we model the entire Australian population, the percentages shown in the figure suggest, in our opinion, that our implementation of the birth and death process is acceptable given that the data we calibrate our models to; namely, we consider a model fitting the data well if the model produced outcomes fall within the confidence intervals which are quite wide.

3. p. 7: If sexual activity commences at age 15, then why is there a need to model 12-14 year olds at all? Although to be honest I am sceptical that 15 really is the minimum age of sexual debut for the whole population; I suspect there is a small but still important proportion of children who commence sexual activity before 15.

We agree with the reviewer. There is not need to include the 12-14 year olds in the model, we have them only to potentially extend the model should we ever have some representative data on their sexual behavior. Their presence, however, does not bring in any complications as they are assumed to be sexually inactive. The age of sexual debut is 15 since we did not have any relatively credible sexual behavior data for individuals under 15 to be able to add them to the models without introducing a number of parameters such as the percentage of sexually active individuals under 15 (in males and females), their sexual partner change rates. Hence our approach should clearly be viewed as a limitation aimed at keeping the models simple (please, see our explanation of why we had to aim for model simplicity at the beginning of this letter).

4. p. 8: The age-independence of infection clearance is a problem because the detailed natural history of pre-cancerous neoplasias is not modelled. Older women are less likely to clear infections because their lesions are further progressed and hence less likely to clear. Models that capture lesion progression do not need age-dependent clearance rates but a simplified model like this would appear to require this age dependence to be built in explicitly.

We agree with the reviewer. While it may seem a more appropriate choice to make the clearance rates age-dependent, this choice would be not the only one supported by literature. Since our
priority was to ensure parsimony of the models (please, see the explanations provided on page 1), we opted for the same durations of infection for all ages since it has been reported by some studies (for example, (Trottier et al., 2008)).

5. p. 12: The posterior probability of seroconversion for females and males is 10-40% and 3-20% respectively; this is much lower than previous empirical estimates. For example, Dillner et al. Semin Cancer Biol 1999; 9:423 suggests >60% of females seroconvert, and Desai et al. Sex Transm Dis 2011; 38:5 suggests the figures for males might be about 1/3 of that.

Previously, the prior distribution assigned to the probability of seroconversion was too wide, allowing for values notably lower than reported in literature. We corrected this in the revised version of the paper: the new prior distributions (and hence the posterior distributions) are in agreement with the reported values.

6. p. 15: While the point about the need to consider SIS models may be valid, it should be noted that the "SIS models" in this paper are not true SIS models. In the second susceptible state individuals still have a degree of protection which they would not have in a classic SIS structure.

We have clarified this in the text ("Natural history of HPV-16"). The models we refer to as SIS (see Figure 1) are not classical SIS models where individuals become fully susceptible to reinfection following clearance. In our SIS models, individuals enter a model in the S state (susceptibles) being seronegative (hence the state is marked as S-). Then they can acquire a sexual partner who is infected and move to state I (infected) themselves. This is followed by clearance of infection, which can result either in a move back to the state S- (no seroconversion occurred), or to state S+. Individuals in S+ are seropositive and susceptible to reinfection with HPV-16. However, we let them have a degree of immunity varying from none to full (parameters changing from 0 to 1). This degree is to be inferred by the calibration process.

7. Table 3: I assume that the betas are the probabilities per partnership rather than per sex act? And if so, shouldn't they be dependent on risk group? (High risk individuals with many sexual partners may have a smaller probability of transmission per partnership since the majority of these partnerships will be casual.)

Our implementation of sexual mixing in the model is based on an implementation suggested in (Garnett & Anderson, 1994). In this formulation, the transmission probability (beta) is a per-partnership transmission probability and does not vary by sexual activity. The point made by the reviewer is valid, however our formulation does not account for this but is a very widely used formulation. While this is a flaw in the formulation, it will not make very much difference because the values we assume are very high (~0.5-1) such that the probability of transmission approaches 1 after only a few partnerships. For example, if the transmission probability per partnership is 0.5, then after 3 partnerships the probability of transmission is 0.75, and after 5 partnerships it is 0.97; these probabilities are 0.98 and 0.999, respectively, if the transmission probability per partnership is 0.75.

8. It would be helpful to show the posterior distributions of the fitted parameters (at least the marginals; the pairwise joint distributions would be even better). It would also be good to compare results to other model-based estimates of natural history parameters (eg. Van de Velde Am J Epidemiol 2007; 165:762, Bogaards et al. Am J Epidemiol 2010; 171:817, Jit et al. Med Decis Making 2010; 30:84, Insinga et al. BMC Inf Dis 2009; 9:119 to name just a few).

We have presented comprehensive descriptions of all posteriors for all model parameters in the
updated Technical Appendix. We have compared some of the estimates in “Results”. This is only a nominal comparison, because the other studies used models different from ours fitted to data from different countries, so we are not in a position to provide a more meaningful discussion going beyond simply quoting the values obtained by other modelers.

**Minor essential revisions**

1. p. 10: It would be good to show the preliminary analyses that indicate that having seroreversion in state R+ only produces an inferior fit.

   Following editor’s suggestions we have removed the line about preliminary results. Initially, we did not present the preliminary analyses for the following reason: the typical problems SIR models have with maintaining a required level of infection inevitably translated in them failing to produce appropriate seroprevalence levels; any of our SIR models would produce even lower seroprevalence with added seroreversion, which would be done with the help of additional parameters (rates of seroreversion for males and females), and DIC would heavily penalize such models.

2. The actual data (serology and DNA) do not seem to be described anywhere. It would be useful to know how it was collected, what the source population is, what the sample sizes are etc.

   We have added a brief description of the data to the Technical Appendix. The data are discussed in detail in the papers we provide references to both in the main text and Technical Appendix.

3. p. 29: In the model equations, shouldn't lambda be a function of [I-) and [I+] as well?

   Yes, it depends on all infected, no matter seropositive or seronegative. We have clarified this in the Technical Appendix (“Calculation of the force of infection”).

4. p. 9: “we associate seroreversion with state S+”. This is confusing since it is not clear whether people serorevert FROM state S+ or TO state S+. Better would be "seroreversion moves individuals from state S+ to state S-".

   We have followed the reviewer’s suggestion and carefully reworded model descriptions (“Natural history of HPV-16”).

5. Figure 4: The actual data points (serology and DNA) are not described in the legend; I assume they are the points with error bars around them.

   We have removed Figure 4 from the revised version. Please, see calibration figures for all models in the Technical Appendix.

**Reviewer**: Johannes A Bogaards

**Major Compulsory Revisions**
The paper is clearly written and the mathematics appears to be well understood. The possibility of inferring natural immunity from cross-sectional age-specific prevalence data of HPV infection (DNA prevalence) and seropositivity is an interesting notion that deserves more investigation.

The simplicity of the models under investigation in this paper is attractive as well as a drawback. For instance: all models perform poorly in predicting HPV DNA prevalence among women, and it seems that the selection of SIS-type models is mainly driven by their ability to at least provide predictions within the 95% confidence intervals of the data. My guess is that predictions would be substantially improved if viral persistence is included in the models, which could give a completely different ordering of models from "best" to "worst" according Fig 5. As far as HPV-16 goes: the proportion of DNA-positive women with a persisting infection rapidly increases as a function of age, which lessens the requirement for reinfection as a means to reproduce a certain level of HPV infection in the population.

We understand that the simplicity of our models may be easily viewed as a drawback, though it also is a clear advantage. However, our approach was to ensure parsimony of the models as well as relative plausibility, which means that they had to be supported at least by some literature. This has now been mentioned in "Background". Please, also see page 1 for justification of our approach.

We understand that the potentially improved predictions the reviewer mentioned are those for SIR(S) models. We agree that the main “problem” of these models is that they can’t reproduce a certain level of HPV infection in the population. So, any assumptions aimed at increasing the level of infection would immediately improve their fit. For example, it could be a higher probability of transmission and/or longer duration of infection. Introducing additional complexity in each of the presented model would lead to an improved fit, but that would not be in line with our selected approach outlined above (please, see page 1).

Another major concern is the construction of priors and the lack of information given about posteriors. It seems that the authors have used highly informative priors, even if there is no substantial (or clear-cut) empirical evidence (a case in point: duration of natural immunity; Tsc, Tsr and Tcl are also not clear from Table 3). In the few instances that information is given about posteriors, it seems that they do not diverge strongly from the priors. This suggests that MCMC does not contribute much to DIC and raises the possibility that model selection is performed on a rather static (and far from optimal) set of models. More information on model fitting procedures is definitely needed to judge the validity of model estimates, and of model selection.

We have explained how we choose prior distributions in the Technical Appendix. Note that these distributions are now different from those in the previous edition of the paper. We have also presented a very extensive description of all posteriors for all model parameters in the Technical Appendix. As it is clear from the description, the posteriors are usually quite different from priors (and we use uniform priors for all parameters). The table with parameters has been substantially improved.

The authors should contrast their findings and interpretations with those of others. Although it is claimed that "these [i.e. serological] data have generally not been considered in the development and calibration of transmission models", some of the cited references already calibrated models to HPV-16 seroprevalence (e.g. Barnabas, PLoS Med 2006) with explicit assumptions about the relationship between seropositivity and immunity. Others (e.g. Bogaards, Am J Epidemiol 2010) have also estimated rates of natural immunity from prevalence data over the range from SIR, SIRS, to SIS models; and concluded that SIRS provided the description of data. It is of interest to know if (and if so; why?) the authors arrive at different conclusions as compared to previous modellers.
We mentioned the relevant findings of other modelers in the main text. It was not possible to directly compare them with ours since their models were not quite the same as ours, and a relatively fair comparison would necessarily involve running the models of interest on the same parameter set as well as analyzing differences in model implementations.

There was indeed one paper (Barnabas2006) where they used HPV-16 prevalence for calibration. Since we are not aware of any other papers employing such data for similar purposes, we claimed that the data have generally not been considered. To clarify this point, we now mention the Barnabas2006 paper in the main text (“Background”).

As for the SIRS model describing the data better than SIS and SIR (as in Bogaards2010), (1) our SIS models are not ‘classical’ SIS models (where one is susceptible right after clearance to the same extent as before becoming infected), hence the fact that they outscored SIRS models does not mean that we place the ‘classical’ SIS over SIRS; (2) our SIRS models turned out clearly superior to SIR models, which is in line with the mentioned result from Bogaards2010; (3) ‘classical’ SIS models are essentially present in our study implicitly, as limit cases of our SIS models (when the degrees of immunity are zero); (4) in the updated results, SIRS4 is the second best model, and its score is slightly worse than that of SIS2 most probably because it has more parameters than SIS2. Please, also see the calibration plots for the models in the Technical Appendix – SIS2 and SIRS4 are indeed very similar.

Finally, the take-home message (that SIS-type models provide superior fit of "real data" and that HPV-16 seropositivity is not a correlate of natural immunity) might be slightly misleading. Regarding the former: the authors do not investigate SIS models, because the risk of reinfection is always reduced after resolution of infection. In effect, the authors model partial immunity on a population level, which translates into a certain probability to become immune upon viral clearance on the individual level. In this sense, I cannot imagine that a SIRS model with initially fast decay of natural immunity could not describe the data as well as a "SIS-type" (=partial SIR) model does. Regarding the latter: the authors might as well conclude that HPV-16 seropositivity is a correlate of immunity (albeit a poor one) since there is a correlation between seropositivity and being protected from reinfection in their models. The point is that this correlation is far from perfect, as seropositive individuals may still be at risk for reinfection. The authors must soften their conclusions and the epidemiological literature on this topic could also be discussed to greater extent.

The main text has been thoroughly rewritten to take this comment into account. We agree that as a ‘classical’ SIS is essentially SIRS with a very short duration of immunity, SIRS should potentially describe data just as well. However, SIRS with a very short immunity would have an extra parameter (average duration of natural immunity) for which it would be penalized by DIC. So, in general SIRS would outscore ‘classical’ SIS if they described the data considerably better than SIS regardless the penalization for extra parameters.

In conclusion, I think that the authors hint at an interesting interpretation of seropositivity in HPV infection dynamics but must make a more convincing case that their conclusions are valid and robust to alternative assumptions.

**Minor Essential Revisions**

Table 2 = Table 1 and Table 3 = Table 2. Fig 5 is not at all informative; this information should be in a Table (Table 3).

We have followed the reviewer’s suggestion and placed the DIC scores into Table 1, which has
been improved.

**Editor's comment:**

All referees agree that this paper looks at an interesting scientific question. However, there are major issues that authors need to address before a decision can be made on the paper. Most importantly - the key observation that "models in which seropositivity is assumed to be a correlate of immunity behave poorly" is not supported by the analysis. Indeed, in the best fitting model, seropositive females have a reduction in the risk of infection of 70-90%. So I would strongly argue that seropositivity appears to be a pretty good proxy for immunity. Overall, I think there is a lot of confusion here: the lack of clarity in the description of the models and the absence of any description of parameter estimates for the different models prevent readers from an in-depth understanding of what is done, what the results are and how they should be interpreted.

It is therefore essential that authors act on these 3 aspects:

- Improve the description of the models. Description in pages 9-10 and Table 1 in particular on how seroconversion/seroreversion is modelled isn't clear enough. The text should rely on the flow diagrams more. Description on how the rates of seroconversion/seroreversion depend on the infection status need to be clarified. Some notations (e.g. S+, S- etc) need to be explained.

We have improved the description of our models (“Natural history of HPV-16”).

- Description of results is largely incomplete. Surprisingly, authors present the priors of their parameters; but they do not present the posteriors. Posterior means (95% CI) and priors should be presented in a Table. The trace of the MCMC output should be shown in the SI. Posterior values of parameters and their interpretation should be discussed.

We provided detailed description of posteriors in the Technical Appendix, along with MCMC trace plots for all parameters and all models.

- Interpretation of the results. Again, I think that there is a lot of confusion here. Authors seem to have an extremely restrictive interpretation of the word "correlate". At the top of page 9, the 2 options they mention are: seropositive individual = immune individual (i.e. correlation= 1 or identity); or seropositivity is not a correlate of immunity (i.e. correlation=0). But clearly, there's a wide range of possibilities between these 2 extreme scenarios, for any correlation between 0 and 1. "Correlate" shouldn't be restricted to "identity". The little authors say about their parameter estimates seem to indicate that their best fitting model is closer to "identity" (correlation=1) than to the scenario correlation=0.

Without all these points being addressed, it is not possible to properly assess the paper. In particular, authors must be careful when interpreting the relatively poor fit of the SIRS model compared to the SIS model. The difference could also be due to certain assumptions made in the models. For example, authors assume that the risk to change status is constant over time (that is, the duration of stay in a compartment is Exponentially distribution). But such assumption has often little biological basis. Consider seroreversion - it's likely that the risk to serorevert will be low just after an individual has seroconverted, but will increase over time. So, authors should investigate whether fit is improved when the duration individuals stay in compartments is taken from a non-Exponential distribution (Gamma for example - by partitioning compartments in sub-compartments; authors can control the variance of the distribution- see appropriate literature).
We agree that the abovementioned improvements to the models would certainly be necessary should we aim at identifying the most plausible description of seroreactivity in the context of HPV transmission. However, the aim of this paper is to consider intentionally simplified yet plausible models. The level of plausibility which has been achieved is a matter of highly subjective judgment, given the present lack of understanding of the issue.

Looking at the raw seroprevalence data - there is a strong difference between patterns for females and males (a decay at older age groups in females that is not observed for males) that is not currently captured by the models. This is worrying and may explain why overall, the models have a relatively poor fit to the data. I suggest authors explore if this difference could be due to sex-specific differences in natural history.

We have made a very substantial revision of our paper. Please, see the new calibration plots for all models in the Technical Appendix.

Other comments:

- Bottom of page 8 "While it is assumed that the natural history of HPV-16 infection is the same for females and males, the rates at which transitions between infection states occur may differ between genders." If it's for compartments other than infection rate, it means that natural history is different between genders.

We have edited the relevant part of the text to avoid misunderstandings.

- End of page 9-beginning of page 10: Section on how seroconversion/seroreversion is modelled is too long and poorly written. In the method section, authors should briefly describe how seroconversion/seroreversion is modelled in the different model variants and the assumptions they make. A short paragraph should be enough. Authors should clarify vague and obscure sentences of the type "for SIS models we associate seroreversion with state S+ (SIS2)" - does it mean that "in the SIS model a Susceptible individual is either seropositive or seronegative; if seropositive, the individual seroreverts at a constant rate"?

We have improved the description (“Natural history of HPV-16”).

The assumptions made should be discussed/justified in the discussion section, not in the methods section. Authors should also remove results of the type "However, preliminary analysis..." at the top of page 10 from the methods section and move them either in the results section or in the discussion (if it's to justify some choices they have made). At the end of the day, the description of the models page 9-10 and Table 1 remains very obscure and doesn't make it possible to a clear understanding of the different model variants. Readers shouldn't have to look at equations in SI to understand nuances between the different models. It is essential that this description is clarified in subsequent revisions.

We have followed editor’s suggestions and improved the text accordingly, as well as Table 1.

- The fact that the best model doesn't manage to capture the age pattern of seropositivity in female is particularly worrying.

This is no longer the case: the revised version of the paper is using the updated priors and includes separate seroreversion rates for males and females. Please, see the calibration plots in
the Technical Appendix.

- Page 12: "While we do not present or discuss the posterior parameter estimations (as compared to their priors, Table 2)" again this is not acceptable. All parameters should be presented and their interpretation discussed.

All parameters are now presented, and the most relevant ones are discussed.

References


