

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The natural history, dynamics, and ecology of Human papillomaviruses in genital infections of young women: protocol of the PAPCLEAR cohort study
AUTHORS	Murall, Carmen Lia; Rahmoun, Massilva; Selinger, Christian; Baldellou, Monique; Bernat, Claire; Bonneau, Marine; Boué, Vanina; Buisson, Mathilde; Christophe, Guillaume; D'Auria, Giuseppe; De Taroni, Florence; Foulongne, Vincent; Froissart, Rémy; Graf, Christelle; Grasset, Sophie; Groc, Soraya; Hirtz, Christophe; Jaussent, Audrey; Lajoie, Julie; Lorcy, Frédérique; Picot, Eric; PICOT, Marie-Christine; Ravel, Jacques; Reynes, Jacques; Rousset, Thérèse; Seddiki, Aziza; Teirlinck, Martine; Tribout, Vincent; Tuailon, Edouard; Waterboer, Tim; Jacobs, Nathalie; Bravo, Ignacio; Segondy, Michel; Boulle, Nathalie; Alizon, Samuel

VERSION 1 – REVIEW

REVIEWER	DR. DORCAS OBIRI-YEBOAH SCHOOL OF MEDICAL SCIENCES, COLLEGE OF HEALTH AND ALLIED SCIENCES, UNIVERSITY OF CAPE COAST, CAPE COAST, GHANA
REVIEW RETURNED	18-Sep-2018

GENERAL COMMENTS	<p>This is a very important protocol. I have very few comments and edits and they have been inserted as comments on the attached PDF file. In simple terms:</p> <ol style="list-style-type: none">1. please age 'years' when stating the age of the women2. few typos to correct3. kindly offer an explanation for the follow up intervals of 2 and 4 months for the HPV positive and negative women respectively4. some sentences are incomplete before the references5. were the investigating physicians who collected cervical and vaginal samples of a specific gender? Clarify6. kindly revise the colour scheme for figure one to make the text more legible <p>- The reviewer also provided a marked copy with additional comments. Please contact the publisher for full details.</p>
-------------------------	---

REVIEWER	Andrew Brouwer University of Michigan, USA
REVIEW RETURNED	11-Jan-2019

GENERAL COMMENTS	<p>Overall assessment: The authors describe an exciting study that has the potential to expand our understanding of the dynamics of early HPV infections, with an emphasis on the viral and immunological kinetics. Although the sample size is not large, which may result in a lack of clear associations at the end of the trial, the study does collect a large array of data on each participant, including HPV prevalence, HPV antibody prevalence, cytokine profiles, and assessment of the vaginal microbiome. One limitation of the study that the authors fail to highlight in a meaningful way is that, because of the study recruitment from an STI clinic and as inclusion criteria further constrain enrollment, the population is not representative and should not be used to assess prevalence. However, the study is well positioned to instead add to our understanding of the underlying biology of HPV infections in young women.</p> <p>The protocol, as written, needs additional work. In many places the language is vague or could be interpreted in different ways. The authors should work to improve the specificity of their language throughout. Many of my specific comments below highlight places that I found the language or explanation insufficient or unclear.</p> <p>Specific comments: Title: I'm not sure what is meant by the "kinetics" of an HPV infection. Do you mean viral kinetics, host immunological kinetics, both, or something else? I'm sure I will understand better after reading the protocol, but please consider being more specific in your title.</p> <p>Page 5, line 8. The word "however" here is odd, since this sentence does not really modify the previous one. Could you please rephrase?</p> <p>Page 5, line 36-45. What you have written here is primarily a "conclusion" section. It describes neither the ethics nor the proposed dissemination. At a minimum, this section should include the ethical review body that approved the study.</p> <p>Page 6, line 3. None of the bullet points outline limitations of this study. One limitation is that all participants are recruited from an STI detection center and thus are not representative of the general population. Another limitation is that the sample size is not particularly large (but this limitation is likely balanced by the expected high prevalence of HPV).</p> <p>Page 6, line 6. This bullet point is neither a strength nor limitation but a statement of fact.</p> <p>Page 6, line 9. The 2-month follow-up time is indeed an important strength of this study. I would not describe 24 months as being a</p>
-------------------------	--

	<p>particularly “long” follow-up in comparison to other HPV studies, though, but it is adequate.</p> <p>Page 6, line 12. This sentence should be reframed as a strength, as having all of these varied measure is an important strength of this protocol. I would particularly emphasize that you will have both measures of infection and antibodies.</p> <p>Page 6, line 16. This bullet point is neither an obvious strength nor limitation. In fact, I’m not sure what to make of this sentence without additional information.</p> <p>Page 6, line 19. You could probably reframe this point as a strength, as you are using finite resources to maximize your inferences.</p> <p>Page 7, line 13. Since you are contrasting the 16% prevalence in women with prevalence in women under 25, you should say the age range for the first estimate.</p> <p>Page 7, line 18. Given that this study addresses HPV infection in young people, your opening paragraph is quite light on information about prevalence in that population. Please briefly summarize what is known about prevalence in young people elsewhere in Europe, in the US, and elsewhere. (I know that there is a wealth of information for the US, at least). You should also spend more time justifying why your study is only in 18-25 year olds. Why not all women or a different age group? What is specifically of interest for this population?</p> <p>Page 7, line 21-27. Each of these sources is over a decade old. Do they still represent the best estimates of the literature?</p> <p>Page 7, line 41. You cite 1, 4, 6, and 7 above as the basis for our knowledge about HPV persistence, but then use 9 as the example of long-follow-up.</p> <p>Page 7, line 44. Why the word “kinetics”, which, to me, has connotations of molecular reactions, as opposed to something like “dynamics?” Is there a particular reason behind the choice?</p> <p>Page 7, line 53. “but we do not know how these play out...” I do not quite understand what is meant here. Do you mean that we don’t know how the aforementioned risk factors translate into infection and host biology?</p> <p>Page 8, lines 20-22. “long-term efficacy of the vaccine at the population level will largely depend on the virus within-host dynamics.” Please expand and explain what you mean by this.</p> <p>Page 8, line 35. It is strange to say that the standard of care of HPV infection is to avoid overtreatment when there is no treatment for HPV. I think you mean that no action is taken if a young woman tests positive for a high-risk HPV type. Please be clear that any treatment is for cervical lesions rather than HPV.</p> <p>Page 8, line 59. “the most oncogenic and the most prevalent HPVs.” I don’t believe that this statement is correct as written (at</p>
--	---

	<p>least in regards to prevalence). I think you mean that HPV 16 and 18 are implicated in the majority of cervical cancers (together accounting for XX% of cervical cancers).</p> <p>Page 9, line 55. You cannot actually say much about prevalence given the study design, and this objective should be removed or modified. You are recruiting from an STI center and including only sexually active participants. While this is understandable from the point of viewing of studying infection dynamics, it means that your study will not be representative of the general population.</p> <p>Page 10, line 40. "in terms of STI exposure, the centre is equally visited by people with high-risk and low-risk behaviours." I don't know what this means exactly.</p> <p>Page 10, line 57. For readers outside of the study country, please explain the implications of a person being on or nor being on a social security scheme (and what exactly it is).</p> <p>Page 11, line 2. Genital warts do not constitute a history of cervical pathology.</p> <p>Page 11, line 4. "heavy treatment" is neither technical nor particularly evocative. Do you mean something like "intensive medical treatment or pharmaceutical therapy"? In Figure 1, you describe it as "chronic medical treatment."</p> <p>Page 11, line 13. Could you please explain what is meant by "vulnerable group (guardianship)." Are either of these technical terms? Is guardianship the only vulnerable group excluded, or is it just an example?</p> <p>Page 11, line 28. You have not described the participant consent process.</p> <p>Page 11, line 32. "a number of samples are collected." Please be very specific about what samples are collected and for what purpose. Or, as it appears you explain these later, let the reader know that the sample will be enumerated later in the protocol.</p> <p>Page 11, line 53. Throughout, you have implicitly used HPV to mean an alpha HPV. Why are you changing terminology now?</p> <p>Page 12, line 6. Why are you dropping HPV infected women who clear, as opposed to switching them to a 4-month follow-up? Please explain the reasoning behind this design choice.</p> <p>Page 12, line 16. What do participants do with the self-swabs? Mail them? Store them? Are the instructions about refrigeration, or not?</p> <p>Page 12, line 42. "Results of the study will be disseminated to study participants via email." Do you mean individual participant results will be emailed to them (which seems like a violation of confidentiality)? Or that you will write a newsletter with the overall study results?</p> <p>Page 13, line 27. Is there any reason to think that taking cervicovaginal samples first might affect the later readings</p>
--	---

	<p>(physical or mental discomfort could potentially affect blood pressure, body temperature)?</p> <p>Page 13, line 43. What sort of self-sample: vaginal or urine? Does the sentence about syphilis mean that only some people are tested depending on whether they meet the guidelines or that everyone is tested in accordance with the guidelines?</p> <p>Page 14, line 41. Are you saying that the participants switches to the other arm? Please be more clear about if and how participants in each arm can switch to the other (and, later, how that impacts the analysis).</p> <p>Page 15, line 18. I would have expected presence of HPV to be included on this list.</p> <p>Page 15, line 26. “the bacterial, fungal and viral communities” is vague. At what taxonomic level will you be assessing these? Are you interested in absolute abundance, or relative abundance, etc?</p> <p>Page 16, line 7. Is there an extra “not” in this sentence? It doesn’t quite make sense to me as written.</p> <p>Page 16, line 36. It would be helpful to the reader to explain why and how and which cytokines are likely to be of interest.</p> <p>Page 17, line 2. Please provide some citations for these methods.</p> <p>Page 17, line 21. In general, “population dynamics” evokes populations of individuals. If you mean, “within-host” or “cell population” dynamics, please specify.</p> <p>Page 17, line 25. Please explain why you expect 2-month intervals to be sufficient for informing these models.</p> <p>Page 17, line 55. What do OTU and NGS stand for?</p> <p>Page 18, line 12. Please justify that your sample size will be sufficient for conducting a GWAS.</p> <p>Page 19, line 7. This sentence is difficult to parse. Please rephrase.</p> <p>Page 19, line 13. I do not understand. Are you enrolling 150 women, or are you only enrolling enough to achieve 75 HPV+? Indeed, I find this whole sample size section to be confusing. Could you please revisit and rewrite the section?</p> <p>Page 20, line 7. Please elaborate on your choice of a sample size of 300 for these analyses.</p> <p>Page 20, line 15. You do not specifically say what ethics review board approved this study.</p> <p>Page 22, line 27. Contamination of what samples, from what sources, at what stages? Please explain why entering the Proficiency Panel will help you control this potential issue.</p>
--	--

	<p>Page 22, line 51. Compensation of participants should have been indicated in the study enrollment section.</p> <p>Overall, the writing is unpolished. There are numerous minor typos throughout. Please do a thorough, close read. Some problems that I noticed (non-exhaustive): Page 8, line 2. Page 8, line 12. Page 8, line 16. Page 8, line 22. Page 10, line 40. Page 10, line 51. Page 10, line 55. Page 16, line 16. Page 19, line 17. Page 21, line 43. Page 21, line 58.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Response to reviewers

Reviewer #1

1. please age 'years' when stating the age of the women

Done

2. few typos to correct

We reread the texte carefully.

3. kindly offer an explanation for the follow up intervals of 2 and 4 months for the HPV positive and negative women respectively

This is a good point. This general intervals were based on earlier studies. First, evidence suggests that incident infections last from 9 to 18 months depending on the HPV type [Insinga et al. 2007]. Furthermore, another study with a close follow-up (twice per week for 4 months) showing that a longer follow-up was necessary [Brotman et al. 2014]. The difference between HPV-negative and HPV-positive is motivated by the estimated incidence rate of HPV, which is 30% per year [Winer et al. 2006]. Having a more regular follow-up would lead to less than 10% chances to detect new infections.

4. some sentences are incomplete before the references

We edited the text.

5. were the investigating physicians who collected cervical and vaginal samples of a specific gender?

Clarify

Physicians performing the gynaecological exam are women. The nurse team is composed of two women and one man. The field team is only composed of women.

6. kindly revise the colour scheme for figure one to make the text more legible

Thank you for the feedback. The colour was due to BMJ Open's conversion. We nevertheless redid Figure 1 to clarify several point.

Reviewer #2

Overall assessment:

The authors describe an exciting study that has the potential to expand our understanding of the dynamics of early HPV infections, with an emphasis on the viral and immunological kinetics. Although

the sample size is not large, which may result in a lack of clear associations at the end of the trial, the study does collect a large array of data on each participant, including HPV prevalence, HPV antibody prevalence, cytokine profiles, and assessment of the vaginal microbiome. One limitation of the study that the authors fail to highlight in a meaningful way is that, because of the study recruitment from an STI clinic and as inclusion criteria further constrain enrollment, the population is not representative and should not be used to assess prevalence. However, the study is well positioned to instead add to our understanding of the underlying biology of HPV infections in young women.

Thank you for the careful reading and positive feedback. We fully agree with the limitation raised by the reviewer about the epidemiological representativeness of our sample. This should not affect our longitudinal follow-up but it can be a constraint if we want to draw epidemiological conclusions. We tried to clarify this point in the revised version.

The protocol, as written, needs additional work. In many places the language is vague or could be interpreted in different ways. The authors should work to improve the specificity of their language throughout. Many of my specific comments below highlight places that I found the language or explanation insufficient or unclear.

We have gone carefully through the protocol. Thank you also for the numerous suggestions!

Specific comments:

Title: I'm not sure what is meant by the "kinetics" of an HPV infection. Do you mean viral kinetics, host immunological kinetics, both, or something else? I'm sure I will understand better after reading the protocol, but please consider being more specific in your title.

Kinetics is a term classically used to refer to within-host dynamics of viruses or drugs. However, we agree that for a general audience, using less jargon might be better. We now refer to "Natural history" instead of kinetics.

Page 5, line 8. The word "however" here is odd, since this sentence does not really modify the previous one. Could you please rephrase?

Indeed. We removed however.

Page 5, line 36-45. What you have written here is primarily a "conclusion" section. It describes neither the ethics nor the proposed dissemination. At a minimum, this section should include the ethical review body that approved the study.

We edited the abstract to fit BMJ Open's criteria better.

Page 6, line 3. None of the bullet points outline limitations of this study. One limitation is that all participants are recruited from an STI detection center and thus are not representative of the general population. Another limitation is that the sample size is not particularly large (but this limitation is likely balanced by the expected high prevalence of HPV).

We now list the limitation on representativeness of the general population. The limitation about the sample size is related to the previous one because HPV prevalence is indeed high (as confirmed by our first enrolments).

Page 6, line 6. This bullet point is neither a strength nor limitation but a statement of fact.

We removed this point.

Page 6, line 9. The 2-month follow-up time is indeed an important strength of this study. I would not describe 24 months as being a particularly "long" follow-up in comparison to other HPV studies, though, but it is adequate.

We removed the 24 months to focus on the density of the follow-up.

Page 6, line 12. This sentence should be reframed as a strength, as having all of these varied measure is an important strength of this protocol. I would particularly emphasize that you will have both measures of infection and antibodies.

We rephrased to insist on the diversity of data.

Page 6, line 16. This bullet point is neither an obvious strength nor limitation. In fact, I'm not sure what to make of this sentence without additional information.

This is a complex issue because it is more about the analysis of the data but, at the same time, having mathematical models included from the beginning is an asset. We rephrased to make this clearer.

Page 6, line 19. You could probably reframe this point as a strength, as you are using finite resources to maximize your inferences.

Done.

Page 7, line 13. Since you are contrasting the 16% prevalence in women with prevalence in women under 25, you should say the age range for the first estimate.

We now give the prevalence for women less than 25 years old.

Page 7, line 18. Given that this study addresses HPV infection in young people, your opening paragraph is quite light on information about prevalence in that population. Please briefly summarize what is known about prevalence in young people elsewhere in Europe, in the US, and elsewhere. (I know that there is a wealth of information for the US, at least). You should also spend more time justifying why your study is only in 18-25 year olds. Why not all women or a different age group? What is specifically of interest for this population?

We indicate international estimates [Bruni et al. 2010] in the Introduction and better justify our age class when presenting the participants. In short, the age was chosen to maximise HPV prevalence and minimise technical difficulties (i.e. avoid the need for parental consent).

Page 7, line 21-27. Each of these sources is over a decade old. Do they still represent the best estimates of the literature?

There are some new studies but in our opinion they do not bring any new details to the ones from 2007 and 2008. We do cite now the study by Ramanakumar et al. (2016, BMJ Open). Incidentally, the bibliography of this study proves our point since they too cannot find a reference to cite on HPV infection duration after 2009... We also cite a study on a smaller cohort of girls that shows shorter infection duration for primo-infection (Houlihan et al. 2016 Int. J. Epidemiol).

Page 7, line 41. You cite 1, 4, 6, and 7 above as the basis for our knowledge about HPV persistence, but then use 9 as the example of long-follow-up.*

We now clarify that the data originates from the control arm of vaccine studies, which explains the reference. We also added references to the studies by Ramanakumar et al., Rodríguez et al. and Insinga et al. since they support this statement.

Page 7, line 44. Why the word "kinetics", which, to me, has connotations of molecular reactions, as opposed to something like "dynamics"? Is there a particular reason behind the choice?

The term does originate from chemistry but it is often applied to virus dynamics models. We rephrased and now cite a recent review on the topic by Canini & Perelson (2014).

Page 7, line 53. "but we do not know how these play out..." I do not quite understand what is meant here. Do you mean that we don't know how the aforementioned risk factors translate into infection and host biology?

We rephrased to be more specific. Our point was to determine how external factors affect viral kinetics.

Page 8, lines 20-22. “long-term efficacy of the vaccine at the population level will largely depend on the virus within-host dynamics.” Please expand and explain what you mean by this.

Our point was that selection occurs within hosts (largely due to the immune response). The reason why vaccination acts as a selective pressure on HPVs is because it modifies this within-host environment. We cite a recent essay on the topic to further illustrate this point.

Page 8, line 35. It is strange to say that the standard of care of HPV infection is to avoid overtreatment when there is no treatment for HPV. I think you mean that no action is taken if a young woman tests positive for a high-risk HPV type. Please be clear that any treatment is for cervical lesions rather than HPV.

Thank for for spotting this. We made the change.

Page 8, line 59. “the most oncogenic and the most prevalent HPVs.” I don’t believe that this statement is correct as written (at least in regards to prevalence). I think you mean that HPV 16 and 18 are implicated in the majority of cervical cancers (together accounting for XX% of cervical cancers). We made the change.

Page 9, line 55. You cannot actually say much about prevalence given the study design, and this objective should be removed or modified. You are recruiting from an STI center and including only sexually active participants. While this is understandable from the point of viewing of studying infection dynamics, it means that your study will not be representative of the general population. We clarified our text because many of our participants are also recruited directly from the university. We also rephrased our sentence to insist on the HPV type diversity rather than the prevalence.

Page 10, line 40. “in terms of STI exposure, the centre is equally visited by people with high-risk and low-risk behaviours.” I don’t know what this means exactly.

Approximately half of the people who get tested for STIs have a risky behaviour (e.g. no use of condoms), whereas the other half are long-lasting couples who wish to stop using condoms. We anyway deleted the sentence since the previous one seemed clear enough.

Page 10, line 57. For readers outside of the study country, please explain the implications of a person being on or nor being on a social security scheme (and what exactly it is).

Benefiting from French social security means that any health expense is covered by the collectivity. This is mandatory to enter a clinical study in France.

Page 11, line 2. Genital warts do not constitute a history of cervical pathology. We changed it to HPV-associated pathology.

Page 11, line 4. “heavy treatment” is neither technical nor particularly evocative. Do you mean something like “intensive medical treatment or pharmaceutical therapy”? In Figure 1, you describe it as “chronic medical treatment.”

Thanks for the suggestion!

Page 11, line 13. Could you please explain what is meant be “vulnerable group (guardianship).” Are either of these technical terms? Is guardianship the only vulnerable group excluded, or is it just an example?

This is indeed a technical term (see article 8 of the Helsinki declaration) and guardianship is an example. We replaced it by children to be more specific.

Page 11, line 28. You have not described the participant consent process.
We added the details about informing participants and obtaining the signed consent.

Page 11, line 32. "a number of samples are collected." Please be very specific about what samples are collected and for what purpose. Or, as it appears you explain these later, let the reader know that the sample will be enumerated later in the protocol.
We added "see below" since the samples are indeed detailed later.

Page 11, line 53. Throughout, you have implicitly used HPV to mean an alpha HPV. Why are you changing terminology now?
This is because our test (DEIA) only detects alpha-papillomaviruses, which are the clinically-relevant ones. We rephrased to be clearer.

Page 12, line 6. Why are you dropping HPV infected women who clear, as opposed to switching them to a 4-month follow-up? Please explain the reasoning behind this design choice.
This is a very clever suggestion, which we had not thought about when building the study. However, it would have led to a different study structure. Here, we decided to focus on infections to maximise our chances to obtain full longitudinal follow-ups. Setting a time window during which we would have followed a given number of women (independently of their status) seemed to increase the risk for incomplete time series.

Page 12, line 16. What do participants do with the self-swabs? Mail them? Store them? Are the instructions about refrigeration, or not?
Participants are asked to keep the samples in their freezers and bring them back at each visit. The nurses then examine whether the samples are still frozen and, if not, they are processed immediately.

Page 12, line 42. "Results of the study will be disseminated to study participants via email." Do you mean individual participant results will be emailed to them (which seems like a violation of confidentiality)? Or that you will write a newsletter with the overall study results?
Sorry, we meant a newsletter. This will only be sent to participants who have exited the study to avoid interference with behaviour.

Page 13, line 27. Is there any reason to think that taking cervicovaginal samples first might affect the later readings (physical or mental discomfort could potentially affect blood pressure, body temperature)?
No, we have no reason to expect this.

Page 13, line 43. What sort of self-sample: vaginal or urine? Does the sentence about syphilis mean that only some people are tested depending on whether they meet the guidelines or that everyone is tested in accordance with the guidelines?
Self-sample are vaginal swabs. Syphilis was tested for any participant who met the CeGIDD's guideline since the test is not routinely prescribed. We clarified the formulation.

Page 14, line 41. Are you saying that the participants switches to the other arm? Please be more clear about if and how participants in each arm can switch to the other (and, later, how that impacts the analysis).
The participant indeed switches to the HPV+ arm upon HPV infection detection. Participants from the HPV+ arm cannot switch anymore (they remain in the study until the infection is cleared or has

reached chronicity, defined here after 24 months of HPV positivity). Analyses will be run at the within-host level by fitting kinetic models so we do not expect the switch in arm to affect them.

Page 15, line 18. I would have expected presence of HPV to be included on this list.
We rephrased the main endpoint for clarity.

Page 15, line 26. “the bacterial, fungal and viral communities” is vague. At what taxonomic level will you be assessing these? Are you interested in absolute abundance, or relative abundance, etc?
We will investigate total abundance. As explained below, bacterial diversity is assessed through 16S RNA and fungal diversity through ITS-1 loci.

Page 16, line 7. Is there an extra “not” in this sentence? It doesn’t quite make sense to me as written. Indeed, thanks!

Page 16, line 36. It would be helpful to the reader to explain why and how and which cytokines are likely to be of interest.
We rephrased the paragraph and now list the cytokines considered.

Page 17, line 2. Please provide some citations for these methods.
We added the reference to the primers used.

Page 17, line 21. In general, “population dynamics” evokes populations of individuals. If you mean, “within-host” or “cell population” dynamics, please specify.
We changed the formulation to viral kinetics and added a reference to Canini & Perelson.

Page 17, line 25. Please explain why you expect 2-month intervals to be sufficient for informing these models.
This is based on the estimates for infection duration (on average 12 months) and from the fact that a 4 month duration seemed too short to distinguish these patterns (Brotman et al. 2014). Furthermore, we will also have access to self-samples if we need to increase the resolution.

Page 17, line 55. What do OTU and NGS stand for?
Done. We also added NGS to the list of Abbreviations (OTU was already there).

Page 18, line 12. Please justify that your sample size will be sufficient for conducting a GWAS.
As indicated in the text, we are at the limit for GWAS analyses. However, strong effects have been described with similar sample sizes. Furthermore, another possibility is to use our dataset to look for already-described SNPs or to focus on particularly rare clinical events.

Page 19, line 7. This sentence is difficult to parse. Please rephrase.
Page 19, line 13. I do not understand. Are you enrolling 150 women, or are you only enrolling enough to achieve 75 HPV+? Indeed, I find this whole sample size section to be confusing. Could you please revisit and rewrite the section?
We apologize for the particularly unclear writing. We have simplified the formulation to focus on the number of women enrolled.

Page 20, line 7. Please elaborate on your choice of a sample size of 300 for these analyses.
The size of the cross-sectional study was the minimal one to match the size of earlier GWAS studies.

Page 20, line 15. You do not specifically say what ethics review board approved this study.

This is due to BMJ Open's format, which imposes the Ethics section to be at the end of the manuscript.

Page 22, line 27. Contamination of what samples, from what sources, at what stages? Please explain why entering the Proficiency Panel will help you control this potential issue.

The contamination can occur at any stage prior to HPV typing. The proficiency panel will help us assess the presence of such contaminations in our laboratory routine, which is a necessary step towards finding means to reduce them.

Page 22, line 51. Compensation of participants should have been indicated in the study enrollment section.

We have moved this paragraph earlier in the text.

Overall, the writing is unpolished. There are numerous minor typos throughout. Please do a thorough, close read. Some problems that I noticed (non-exhaustive): Page 8, line 2. Page 8, line 12. Page 8, line

16. Page 8, line 22. Page 10, line 40. Page 10, line 51. Page 10, line 55. Page 16, line 16. Page 19, line 17.

Page 21, line 43. Page 21, line 58.

We apologize for the numerous mistakes and re-read the text carefully.

VERSION 2 – REVIEW

REVIEWER	Andrew Brouwer University of Michigan
REVIEW RETURNED	27-Feb-2019
GENERAL COMMENTS	The manuscript is much improved by the revisions, and I have no further comments.