## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | EXPANDING THE UPPER AGE LIMIT FOR CERVICAL CANCER SCREENING: A PROTOCOL FOR A NATIONWIDE NON-RANDOMIZED INTERVENTION STUDY |
| **AUTHORS** | tranberg, mette; Petersen, Lone; Elfström, Klara; Hammer, Anne; Blaakær, Jan; Bennetsen, Mary; Jensen, Jørgen; Andersen, Berit |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | FJ van Kemenade<br>Dep of pathology & clin bioinformatics<br>Erasmus MC Univ Med Center, Rotterdam |
| **REVIEW RETURNED** | 11-May-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | Expanding the upper age limit for cervical cancer screening: a protocol for a nationwide non-randomized intervention study.<br><br>Given that in DK screening ends at 65 (with an exit test), the author's premise is that a lack of appropriate exposure to the programme may explain the current rise of carcinoma in 75-80 years old. Their approach is a quasi-experimental design in which one-region offers either GP based of self-sampled hrHPV testing. The four other regions serve as 'population control'.<br>It is a non-randomized approach: women aged 65-69 are targeted. They can decide for they if the opt for the GP visit of want to have the self-sampling device.<br>The in the style of the Nordic country to have such a careful approach. Yet, I have some question that merit attention from the authors while conducting their experiment.<br><br>1. Authors should give numbers on the current perceived 'rise of carcinoma's' in the group of 75-80. What is the size of the problem? One would expect that since 1960 (start of the programme) even in this age group protection would have set in.<br>2. How do authors explain the observation that there is no rise in hrHPV prevalence (study ref 17)? If that is the case, how certain can the authors be that offering hrHPV detection can mitigate cancer in older women<br>3. Authors settle, understandably, for CIN2+ as primary end-point, while admitting this is a surrogate marker. However, what evidence authors propose that this end point (i.e.CIN2) is a valid surrogate marker in this age group? Isn't CIN3+ a better end point, even if this end point may limit the current power calculation<br>4. Have the authors modelled how long an hrHPV positive test result at the age of 65 would 'translate into a lower incidence of cervical cancer? Isn't such an endpoint competed out by other causes of death, such a cardiac disease, more frequent cancer<br>5. I cannot find any mention of excluded women that have had an |

| | |
|---|---|
| | exit test with hrHPV. Isn't more useful to concentrate on women that did not receive such a HPV exit test?<br>6. For risk factors, I found smoking missing. This probably needs to be asked<br>7. Authors should consider to evaluate patient satisfaction by the offered test.<br>8. I wonder if the COBAS test is validated with use of surepath?<br>9. How can authors excluded bias by intention? Will women that do opt for the self test be similar to physician obtained test. Authors do clearly realized the presence of confounders.<br>10. Who funds this study? This may not be very cheap…<br>11. As for the figure: they are clear, but authors should add when they measure their end points. Why did they choose for 6 months? Isn't that too early<br>12. A more concise formula should be given for the harms measured. Authors seem to propose the number of colposcopy / conization performed. I would suggest that colposcopy ending with CIN1 or less is clearly harmfull (no conization here) in comparison to colposcopy leading to treatment. Treatment of CIN2 is a moot point. How do authors distinguish this with end point CIN3? Do they include p16 or some objective marker?<br>13. There are subtle difference in the triage schemes with the self-test and the GP cytology. This may introduce ascertainment/ verification bias. In such a well thought of trial, that surprises me a bit. |

| | |
|---|---|
| **REVIEWER** | Paolo Giorgi Rossi<br>Azienda USL - IRCCS di Reggio Emilia, Italy |
| **REVIEW RETURNED** | 19-May-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | In general the study is interesting and well described. I have only few comments to improve the reporting. In particular the abstract is not very informative and should be reframed.<br><br>Abstract:<br>Methods<br>The description about allocation to self-sampling or GP sample is not clear: "The intervention group is invited for HPV-based screening by attending routine screening at the GP or by requesting a self-sampling kit." Can they choose? If they cannot, they are not randomized but how are they allocated?<br>Analyses<br>Please describe how the different interventions are compared: extending age and self-sampling. From the abstract it is not clear if there are two (GP lower age vs. self-sampling higher age), three (GP lower, GP higher age, self-sampling higher age) of four groups (full factorial design).<br><br>Strengths and limitations<br>Please try to explicit the hypotheses that are tested or at least the comparisons: the increase in age is not mentioned at all.<br><br>Background<br>Very clear and well written. The hypotheses about the peak after screening stop-age are well presented. I suggest to give some numbers about the magnitude of this peak.<br><br>Methods<br>Outcomes: the choice of the main outcome should be better justified: |

| | |
|---|---|
| | in this context the outcome cannot be used as a comparison in sensitivity since the women are not undergoing to the same rounds of screening, the outcome should represent how many possible future cancers we are preventing. In this prospective CIN3, AIS and microinvasive Ca are probably a success, CIN2 we do not know, and FIGO 2 or more severe cancers are not a success. So FIGO 2 cancers are rare probably so they do not affect the comparison, but the inclusion of CIN2 should be justified. In any case, the evaluation of a public health intervention, particularly screening, should be based on the balance of many outcomes, desirable and undesirable, so probably the primary endpoint will not differ too much from secondary endpoints in the weight of the evaluation. I suggest to emphasize this point in the discussion: how all the information collected will be summarized to inform decision making; it is absolutely not the case that if the study has a $p<0.05$ on the CIN2+ outcome automatically all the intervention will be implemented as it is…<br><br>Statistical analyses<br>I think comparisons could be better represented also graphically (I miss the study flowchart as I mentioned below).<br><br>Discussion<br>Pagl 18/19: is it true that many CIN will not have time to progress, but it is also true that the probability that a missed CIN3 became a cancer will be a cancer in the next 6 years is much higher in women over 50 (if you compare the data by Ronco at al lancet 2014 and the relative detection rate in the original trials at baseline, this is clear, Ronco 2010, Rijkaart 2012, Naucler 2007).<br><br>Figures<br>I miss a flowchart of the study, where it is clear which decisions are made by the investigator (by geographic allocation) and which by the woman (using self-sampling or attending GP clinic); which samples will be used for biomarkers (only GP in the intervention area or also in control areas). |

| | |
|---|---|
| **REVIEWER** | Andrew Coldman<br>BC Cancer Research Centre |
| **REVIEW RETURNED** | 02-Jun-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is a clearly written protocol which describes the study planned by the investigators. I have only minor suggestions:<br>1. I wonder if it would be helpful to use the same terminology between figures 2 and 3 so that HPV test results are called GPHPV and SSHPV (or something similar) to distinguish them in a consistent way.<br>2. Some states in Figure 3 (states indicating 3-month retest) do not lead anywhere and also indicate repeat HPV testing following a valid HPV test but inadequate cytology – is this correct?<br>3. The quasi-experimental design used in the study appears to be based on a post-intervention comparison in the intervention and control regions. Should pre-intervention outcome statistics in the intervention and control regions be considered to potential adjust for any regional differences? It may be helpful to indicate in the strengths and weaknesses discussion why this is not considered in the study design.<br>4. In the abstract introduction it indicates that the study will seek to determine if “cervico-vaginal self-sampling for HPV is superior to GP based screening in reaching long-term unscreened women”. I |

| | |
|---|---|
| | wonder if it would be more accurate to state that the study attempts to determine whether an invitation providing the option for self-sampling or GP based screening is superior to existing practice where women are not invited to screening? |

**VERSION 1 – AUTHOR RESPONSE**

**REVIEWER 1:**
Given that in DK screening ends at 65 (with an exit test), the author's premise is that a lack of appropriate exposure to the programme may explain the current rise of carcinoma in 75-80 years old. Their approach is a quasi-experimental design in which one-region offers either GP based of self-sampled hrHPV testing. The four other regions serve as 'population control'. It is a non-randomized approach: women aged 65-69 are targeted. They can decide for they if the opt for the GP visit of want to have the self-sampling device. The in the style of the Nordic country to have such a careful approach. Yet, I have some questions that merit attention from the authors while conducting their experiment.

**RESPONSE:**
Thank you very much for the kind words regarding our study and your excellent comments to your manuscript.

**1.** Authors should give numbers on the current perceived 'rise of carcinoma's' in the group of 75-80. What is the size of the problem? One would expect that since 1960 (start of the programme) even in this age group protection would have set in.

**RESPONSE:**
We thank the reviewer for the suggestion. We have included detailed information, including a reference, in the revised manuscript, p. 6 lines: 99-100.

**2**. How do authors explain the observation that there is no rise in hrHPV prevalence (study ref 17)? If that is the case, how certain can the authors be that offering hrHPV detection can mitigate cancer in older women

**RESPONSE:**
We thank the reviewer for raising an important point. We cannot tell if offering HPV-screening can mitigate cervical cancer in older women. However, several studies have reported HPV-screening to be more sensitive for detecting cervical precancer and cancer compared to screening with cytology, also in older women. Based on this, the high cervical cancer incidence and mortality, and an ongoing debate with respect to postponing screening exit, we designed the present study. Hopefully, our study results will be valuable for future decision making.

**3.** Authors settle, understandably, for CIN2+ as primary end-point, while admitting this is a surrogate marker. However, what evidence authors propose that this end point (i.e.CIN2) is a valid surrogate marker in this age group? Isn't CIN3+ a better end point, even if this end point may limit the current power calculation?

**RESPONSE:**
Thanks for this important comment. To our knowledge there is no evidence supporting that CIN2+ is a valid surrogate marker in this age group. We acknowledge that a CIN2 diagnosis is associated with low reproducibility and high interobserver variation. Furthermore, it is estimated that CIN2 lesions are more likely to regress. However, studies also find that the likelihood of CIN2 regression declines significantly with increasing age. Based on this, surgical treatment is recommended for older women diagnosed with CIN2 according to Danish guidelines, and is routine practice in developed countries. Thus, for screening purposes, including CIN2+ and CIN3+ cases as the primary and secondary outcomes, respectively, may be justified by them being treatable endpoints in older non-reproductive women (conization), while still recognizing that the detection and treatment of CIN3 and especially

CIN2 may be considered as overtreatment. We have revised the manuscript to make this clearer (page 19, lines: 394-397).

**4.** Have the authors modelled how long an hrHPV positive test result at the age of 65 would 'translate into a lower incidence of cervical cancer? Isn't such an endpoint competed out by other causes of death, such a cardiac disease, more frequent cancer

**RESPONSE:**
Thanks for this important comment. Not much is known about the HPV natural history in older women. For example, studies find that new HPV detection at older age may be a result of recent acquisition, auto-inoculation, viral reactivation of a latent infection, etc. However, it remains unclear if viral reactivation confers similar risk of progression compared to recent acquisition. As it was not the scope of the present study, no modelling regarding this matter was performed. Given the lack of randomization and the targeted age group, we agree that comorbidity is an important confounder to account for when comparing outcomes between the intervention and reference groups. Fortunately, we will be able to assess whether the distribution of comorbidity (Charlston index) is well-balanced between the groups using individual-level registry data. This information is included in the revised manuscript, page: 18, lines: 377-381.

**5.** I cannot find any mention of excluded women that have had an exit test with hrHPV. Isn't more useful to concentrate on women that did not receive such a HPV exit test?

**RESPONSE:**
If the National Health Authorities were to decide to expand the upper screening age, the invitation algorithm would follow today's routine procedure, meaning that women with a negative or positive hrHPV exit test would be invited for screening when 5 years had passed since their latest invitation or cervical cytology sample (whichever came last). To gain real-life experiences, the routine invitation algorithm is used in our study, which is the rationale for not excluding women with a hrHPV positive exit-test from receiving a screening invitation.

**6.** For risk factors, I found smoking missing. This probably needs to be asked

**RESPONSE:**
Indeed, smoking but also sexual behavior are relevant risk factors to account for. However, unfortunately, we were unable to retrieve information on these potential confounders due to the study design. Thus, when reporting the results of our study, we will describe that there could be residual confounding effects from these factors. This information is included in the revised manuscript. p. 18, lines: 377-381.

**7.** Authors should consider to evaluate patient satisfaction by the offered test.

**RESPONSE:**
Thanks for this excellent suggestion. The authors are currently planning on evaluating the patient satisfaction of the two offered screening methods through questionnaires. These results will be published in a separate paper. Furthermore, qualitative interviews with women who have ordered the self-sampling kit, but did not return it to the laboratory will be considered.

**8**. I wonder if the COBAS test is validated with use of surepath?

**RESPONSE:**
Thanks for this question. Yes, the Cobas 4800 assay is clinically validated with use of Surepath collected samples. This information together with a reference is included in the revised manuscript. See page 11, line: 215.

**9.** How can authors excluded bias by intention? Will women that do opt for the self test be similar to physician obtained test. Authors do clearly realized the presence of confounders. .

**RESPONSE:**
We thank the reviewer for raising an important point. Women who chose self-sampling may be

different that those opting for GP-based sampling. The aims of the present study were not only to assess the potential impact on CIN2+/CIN3+, but also to explore whether self-sampling would result in a higher participation rate. Details regarding the included potential confounders are provided in the revised manuscript. See p. 18, lines: 377-381.

**10.** Who funds this study? This may not be very cheap…

**RESPONSE:**
The initiative and the study was partly funded by the Department of Public Health Programmes, Randers Regional Hospital, which is located in the Central Denmark Region. Some public funding had been provided by the Health Foundation and other fundraising is on-going. The Health foundation had no role in the design of the study and collection, analysis, and interpretation of the data, and in writing the manuscript. This has been clarify in the revised manuscript. See p. 20, lines: 411-414.

**11**. As for the figure: they are clear, but authors should add when they measure their end points. Why did they choose for 6 months? Isn't that too early

**RESPONSE:**
Thanks for your suggestion regarding including information about when the different outcomes are measured in Figure 2 and 3. Please see the revised Figures 2 and 3. Your concern regarding the measurement of CIN2+ at 6 months is highly justified, and the authors have decided to expand this to 18 months. This correction is included in the figures and revised manuscript.

**12.** A more concise formula should be given for the harms measured. Authors seem to propose the number of colposcopy / conization performed. I would suggest that colposcopy ending with CIN1 or less is clearly harmfull (no conization here) in comparison to colposcopy leading to treatment. Treatment of CIN2 is a moot point. How do authors distinguish this with end point CIN3? Do they include p16 or some objective marker?

**RESPONSE:**
Thanks for this comment. We agree that colposcopy ending with ≤CIN1 is clearly harmful, although no conization is performed. Thus, this outcome is included in the revised manuscript. Please see page: 14, line: 294.
We do not use p16 to distinguish between CIN2 and CIN3, but it is used to distinguish between precancer (CIN2-CIN3) and mimics of precancer.

**13.** There are subtle difference in the triage schemes with the self-test and the GP cytology. This may introduce ascertainment/ verification bias. In such a well thought of trial, that surprises me a bit.

**RESPONSE:**
It is correct that there are minor differences in the triage schemes with the self-sample and GP cytology. We assume that you are referring to the fact that women attending GP-screening and having a hrHPV negative result do not have reflex cytology testing performed, whereas HPV positive self-samplers with a HPV negative GP-triage sample will undergo cytology co-testing. The rationale behind this difference is that women with a hrHPV positive self-sample/HPV negative GP-triage cytology have just recently been tested HPV positive. Thus, cytology co-testing among these women is performed as an extra safety net.

Ideally, we should have performed HPV and cytology co- testing on women attending GP-screening as well, but according to available funding and resources in the laboratory, this approach was not possible. However, the likelihood of overlooking CIN2+ cases due to the lack of co-testing is considered low, as the sensitivity of cytology testing among postmenopausal women is modest due to low cellularity and the presence of atrophic epithelial cells. Furthermore, papers conducted in the USA have shown that primary HPV testing is very comparable to co-testing with respect to the ability to detect CIN3+. Yet, we acknowledge this limitation of our study which will be taken into account when interpreting the results.

**REVIEWER 2**
In general the study is interesting and well described. I have only few comments to improve the reporting. In particular the abstract is not very informative and should be reframed.

**RESPONSE:**
Thank you very much for the kind words regarding our paper and your excellent comments to your manuscript.

**Abstract:**
**Methods**
The description about allocation to self-sampling or GP sample is not clear: “The intervention group is invited for HPV-based screening by attending routine screening at the GP or by requesting a self-sampling kit.” Can they choose? If they cannot, they are not randomized but how are they allocated?

**RESPONSE:**
The aim in the abstract has been reworded according to comments from reviewer 3, page. 3, lines: 53-57. It now states that invited women can choose to attend screening either at their GP or by self-sampling. Because the aim has been clarified, we have not reworded this exact sentence in the method. We hope this is acceptable to the reviewer. If not, we would be happy to reword the sentence.

**Analyses**
Please describe how the different interventions are compared: extending age and self-sampling. From the abstract it is not clear if there are two (GP lower age vs. self-sampling higher age), three (GP lower, GP higher age, self-sampling higher age) of four groups (full factorial design).

**RESPONSE:**
Thanks for this comment. In the revised manuscript, we have included information on the comparisons made to explore the second purpose of our study p. 3 lines: 68-70.

**Strengths and limitations**
Please try to explicit the hypotheses that are tested or at least the comparisons: the increase in age is not mentioned at all.

**RESPONSE:**
In the revised manuscript, we have reworded the first sentence listed in the short bullet points. p. 5 lines: 80-83.

**Background**
Very clear and well written. The hypotheses about the peak after screening stop-age are well presented. I suggest to give some numbers about the magnitude of this peak.

**RESPONSE:**
The magnitude of the problem is included together with a reference in the revised manuscript, p. 6 lines: 99-100.

**Methods**
Outcomes: the choice of the main outcome should be better justified: in this context the outcome cannot be used as a comparison in sensitivity since the women are not undergoing to the same rounds of screening, the outcome should represent how many possible future cancers we are preventing. In this prospective CIN3, AIS and microinvasive Ca are probably a success, CIN2 we do not know, and FIGO 2 or more severe cancers are not a success. So FIGO 2 cancers are rare probably so they do not affect the comparison, but the inclusion of CIN2 should be justified. In any case, the evaluation of a public health intervention, particularly screening, should be based on the balance of many outcomes, desirable and undesirable, so probably the primary endpoint will not differ too much from secondary endpoints in the weight of the evaluation. I suggest to emphasize this point in the discussion: how all the information collected will be summarized to inform decision making; it is absolutely not the case that if the study has a $p<0.05$ on the CIN2+ outcome automatically all the intervention will be implemented as it is…

**RESPONSE:**
Thanks for your considerations regarding our choice of outcome. We have tried to justify our choice of primary and secondary outcomes, p. 19, lines: 394-397.
By stating the following in the discussion, p.18, lines: 363-366: "*Importantly, this study will evaluate whether the potential harms of screening in older women are outweighed by the potential benefits of decreasing the incidence of cervical (pre)-cancer (2, 49). Overall, this knowledge will address important research gaps and may help guide future screening recommendations",* we believe that we have already emphasized that our study will be used to guide future decision making and screening guidelines.

**Statistical analyses**
I think comparisons could be better represented also graphically (I miss the study flowchart as I mentioned below).

**RESPONSE:**
To clarify the planned comparisons between the groups and within the intervention group, we have included Table 1.

**Discussion**
Page 18/19: is it true that many CIN will not have time to progress, but it is also true that the probability that a missed CIN3 became a cancer will be a cancer in the next 6 years is much higher in women over 50 (if you compare the data by Ronco at al lancet 2014 and the relative detection rate in the original trials at baseline, this is clear, Ronco 2010, Rijkaart 2012, Naucler 2007).

**RESPONSE:**
Thanks for drawing our attention to these interesting and important publications. When reporting the results of our study, we will certainly look closer into the mentioned results.

**Figures**
I miss a flowchart of the study, where it is clear which decisions are made by the investigator (by geographic allocation) and which by the woman (using self-sampling or attending GP clinic); which samples will be used for biomarkers (only GP in the intervention area or also in control areas).

**RESPONSE:**
The need of a flowchart may be limited now that the outcome measures are stated in the Table 1, and the time of outcome measurements are provided in Figure 2 and 3, as recommend by reviewer 1. Thus, we decided to not include a flowchart in the revised manuscript. We hope this is acceptable to the reviewer. Of note, only in the intervention region, p16/ki67 dual stain cytology will be performed on HPV positive GP-samples with sufficient material for cytology testing. This is clarified in the revised manuscript, p.12 lines: 233-234.

**REVIEWER 3**
This is a clearly written protocol which describes the study planned by the investigators. I have only minor suggestions:

**RESPONSE:**
Thank you very much for the kind words regarding our paper and your excellent comments to your manuscript.

1. I wonder if it would be helpful to use the same terminology between figures 2 and 3 so that HPV test results are called GPHPV and SSHPV (or something similar) to distinguish them in a consistent way.

**RESPONSE:**
Due to the comments from reviewer 1, we have divided Figure 2 and 3 according to outcome measures which should ease the interpretation of the figures. For that reason, we have not changed the terminology in Figures 2 and 3. We hope this is acceptable to the reviewer. If not, we would be happy to include these abbreviations.

2. Some states in Figure 3 (states indicating 3-month retest) do not lead anywhere and also indicate repeat HPV testing following a valid HPV test but inadequate cytology – is this correct?

**RESPONSE:**
Thanks for this comment. In Figure 3, the algorithm for women having HPV testing repeated after 3 months is now clarified in the figure notes. Yes, it is correct that HPV testing is performed after 12 months in women positive for other hrHPV types than HPV16/18 with insufficient cytology. This follow-up matches the routine screening guidelines for women aged 60-64 years in Denmark.

3. The quasi-experimental design used in the study appears to be based on a post-intervention comparison in the intervention and control regions. Should pre-intervention outcome statistics in the intervention and control regions be considered to potential adjust for any regional differences? It may be helpful to indicate in the strengths and weaknesses discussion why this is not considered in the study design.

**RESPONSE:**
Thanks for this important comment. We have mentioned this issue in the revised manuscript, p. 18-19, lines: 383-387.

4. In the abstract introduction it indicates that the study will seek to determine if “cervico-vaginal self-sampling for HPV is superior to GP based screening in reaching long-term unscreened women”. I wonder if it would be more accurate to state that the study attempts to determine whether an invitation providing the option for self-sampling or GP based screening is superior to existing practice where women are not invited to screening?

**RESPONSE:**
Thanks for this important comment. We agree that the objectives in both the abstract and manuscript may benefit from being clarified. Please see p. 3, lines: 53-57 and p. 7, lines: 122-127.

**Other revisions:**
The annual proportion of women who are eligible to receive a screening invitation in the intervention region is lower than first calculated due to small birth cohorts. Thus, the recruitment is expected to take about 4.5 years instead of 2.5 years. This correction is added in the abstract and manuscript.

HPV self-sampling has been changed to self-sampling through the manuscript to avoid confusion regarding that all women in the intervention group will undergo HPV-based cervical cancer screening.

## VERSION 2 – REVIEW

| | |
|---|---|
| **REVIEWER** | Folkert J van Kemenade<br>Erasmus MC Univ Med Center, Rotterdam, NL |
| **REVIEW RETURNED** | 23-Jul-2020 |

| | |
|---|---|
| **GENERAL COMMENTS** | I had some more questions and not all are adressed. I understand that the study already started. I'm still puzzled as to why this study is deemed appropriate while at the same time authors state that the hrHPV prevalence has not increased. That would means that the ' problem' will be solved sooner or later as more screened cohorts ' move into the olders ages'. Authors should clarify that point.<br>Most of my questions have been adressed (included funding)<br>Yet, I still miss a paragraph on the harms of this trial. Some women will be treated for lesions that do no harm. This should be adressed. Certainly if no modelling has been done prior to the study. |

## VERSION 2 – AUTHOR RESPONSE

**REVIEWER 1:**
**1)** I understand that the study already started. I'm still puzzled as to why this study is deemed appropriate while at the same time authors state that the hrHPV prevalence has not increased. That would mean that the ' problem' will be solved sooner or later as more screened cohorts ' move into the older ages'. Authors should clarify that point.

**Response:**
Please note that the age group of the population in reference 18 is older (69 and above) than the population we target in our intervention (65-69 years). Extension of the upper screening age as we elaborate in our study may thus prevent the cancers developed in the older age-groups. Further, as we already write in the introduction, the increasing female life expectancy has raised the question if the uper age limit for screening should be extended to 69 or 70 years, and it is therefore relevant to study the effect of such an extension.

To make it clearer we have revised in the relevant paragraph in the introduction (p. 6, lines: 97-105), and it is now stated that:

*It has been hypothesized that the incidence peak* at older ages *could be a result of a mid-life change of sexual partners or reactivation of a latent HPV infection as the immune system weakens with age [14-17]. However, a recent Danish study of HPV DNA prevalence in women aged 69 and above showed no increase in prevalence that could explain the cervical cancer peak at older ages [18]. The authors stated that this result may be explained by the fact that an HPV infection may become undetectable at a late stage in the oncogenic process [18 19]. It has also been hypothesized that the current peak in older ages could be attributed to an insufficient screening history in older birth cohorts [20]. Whatever the reason, the increasing female life expectancy (at age 65 years it is about 20 additional years) has raised the question if the upper age limit for screening should be extended to 69 or 70 years[10 21,22].*

**2)** Most of my questions have been addressed (included funding). Yet, I still miss a paragraph on the harms of this trial. Some women will be treated for lesions that do no harm. This should be addressed. Certainly if no modelling has been done prior to the study.

**Response:**
The risk for harms is already mentioned in the introduction, p. 6, lines: 92-93 and in the strengths and limitations section, p. 19, lines: 391-392. However, to clarify the subject is elaborated in the revised manuscript under the strengths and limitations section (p. 19, lines: 393-396) and it is now stated:

*while still recognizing that the detection and treatment of CIN3, and especially CIN2, may be considered as overtreatment, because an unknown proportion of these lesions would never have progressed to cancer in the woman's lifetime[55]. Specifically, it is important to take into account that conization is associated with an increased risk of bleeding and stenosis, which may hinder or challenge sampling from the cervix post-conization[52], and that false-positive screening results may place some women in a surveillance cycle of unclear end, which may cause distress[55].*