Imperial College
London

Surgery & Cancer
Faculty of Medicine
Imperial College London

Level 2, Hammersmith House
Hammersmith Hospital
Du Cane Road
London
W12 0HS
Tel: +44 (0)20 8383 5290 Fax: +44 (0) 02 8383 8065
Mob: +44(0)7725623604

m.kygiou@imperial.ac.uk
www.imperial.ac.uk

14 May 2016

Maria Kyrgiou MSc PhD MRCOG
Senior Lecturer
Consultant Gynaecological Oncology

The Editor of The British Medical Journal

Dear Editor

Re: **Adverse obstetrical outcomes after local treatment for cervical pre-invasive and early invasive disease according to the cone depth: a systematic review and meta-analysis.**

My colleagues and I submit the enclosed revised manuscript for consideration for publication in The British Medical Journal.

The replies to the points raised by the reviewers and the editorial board are included point by point underneath.

We are aware that we had to include several new points requested by the reviewers and this has increased the word count above the recommendations in the guidance to the authors. If the editorial board feels that the manuscript needs to be shortened, we would be happy to do that. The references of the included studies could also be attached as supplementary material to shorten the manuscript.

I confirm that the material has not been published before and is not under simultaneous consideration elsewhere.

We look forward to your reply.

Yours sincerely

Maria Kyrgiou, MSc, PhD, MRCOG
Clinical Senior Lecturer – Honorary Consultant Gynaecologic Oncologist
Imperial Healthcare NHS Trust – Imperial College

Response to reviewers

Reviewer 1: Maija Jakobsson

This meta-analysis addresses effect of treatment of CIN on obstetric outcomes, especially on preterm birth. This subject has been discussed a lot. This meta-analysis is the first that compares different comparison groups and clearly shows that although women with CIN have a higher baseline risk of prematurity, cervical treatments and particularly large cone size increases this risk even more. Increased risk that is associated with multiple treatments and large cone volumes is presented at the first time in this meta-analysis. Meta-analysis is very well written, statistics is excellent. It is actually very hard to find something that could be improved.
This study is also important to general readers, because patients are also aware of this risk and often seek for advice. Authors will disseminate the results to the lay audience.

Thank you

Specific questions:
*** Scientific reliability**
Research Question - clearly defined and appropriately answered? Yes
Overall design of study - adequate ? Yes
Participants studied - adequately described and their conditions defined? Yes
Methods - adequately described? Complies with relevant reporting standard - Eg CONSORT for randomised trials ? Ethical ? Yes
Results - answer the research question? Credible? Well presented? Yes
Interpretation and conclusions - warranted by and sufficiently derived from / focused on the data? Message clear? Yes
References - up to date and relevant? Yes, very comprehensive literature search.
Abstract/summary/key messages/ What this paper adds – reflect accurately what the paper says? Yes. I suggest to add something about cone size in the key messages.
More details about the cone depth have been included in the key messages as per reviewer's recommendation.

Some minor comments:
1. Title: It could be modified a bit. Now it contains term "cervical preinvasive and early invasive disease". Current WHO classification has abandoned term "microinvasive cervical cancer" and therefore it is not recommended anymore. Authors use this term also on p 5, line 27.
We have removed the word microinvasive from p5, line 27 as per recommendation. We have included the terminology early invasive disease both in the title and throughout the text as excisional conservative local treatment is increasingly used in women with early invasive malignancy. The use of conisation for fertility-sparing surgery now goes beyond stage 1A1 disease (previously known as microinvasive) to encompass stage 1A2 and small volume low risk 1B1 tumours. Large studies are currently assessing the safety of these approaches as compared to radical trachelectomy but this is increasingly used in clinical practice. It is for these reasons that we decided to include this in the manuscript and title as the results of these meta-analyses are also relevant to women treated for early invasive disease. Having said that only one of the eligible studies by Paraskevaidis et al. includes stage 1A1 disease in the inclusion criteria. We are happy to delete this from the title if this is what the editorial board feels is more appropriate.

2. Abstract, p2, line 50. Quite often term "very preterm birth" is used instead of "severe prematurity" on gestational weeks below 34/32.
Indeed 'very preterm birth' could be an alternative terminology and we could change that throughout, although we are struggling to come up with an appropriate word to then describe 'extreme preterm

birth <28/30 weeks of gestation'. These terminologies are not standard in the literature and we feel that 'severe' and 'extreme' are more consistent. We are happy to change the terminology if the editorial board feels that the proposed term is more appropriate.

3. Absract, p 3, line 3. The increased risk with increasing cone depth could be presented in more condensed in the Abstract.
We feel that the increasing risk of preterm birth with increasing cone depth is the most novel and important finding of this paper and the absolute risks and relative risks would be very informative to clinicians. It is for this reason that this is emphasized in the abstract. If the editorial board feels this should be summarized we would be happy to do that.

4. M&M, p5, line 35-36. I do not understand term "aka" in front of procedures.
The abbreviation has been added.

5. M&M. P8, line 46. Patient involvement. This is very important topic, but is this the right place for this?
This is described in the guidance to the authors in the BMJ journal policy. The paragraph has been updated.

6. Study limitations. Cone depth measurement is not precise; therefore results regarding this measurement should be interpreted with caution. This should be discussed in the study limitations section.
A comment has been added in the strength and limitation section as per reviewers recommendations.

Reviewer 2: M E Cruickshank

There have been a number of systematic reviews on this topic. Preterm labour and adverse Obstetric outcomes remain an important issue for colposcopists and women needing treatment of CIN. Perhaps in some areas of healthcare, the argument has been accepted e.g. primary care and public health.
The methodology is appropriate and clear and adheres to STROBE.
The authors clearly set out the rational for this updated meta-analysis and the ongoing uncertainty in treating women and providing accurate information on the risk of preterm labour prior to treatment. The results are clearly presented and the sub-group analysis useful for colposcopists/clinicians. A clearer clinically relevant message on identification of high risk groups or minimising risk would be useful for the target audience.
For example is the key message that the meta-analysis results are reassuring in that RR is only slightly raised by treatment effect over and above having cervical disease and we can reassure the majority of women who will need only a single type 1 excision? Or should the reader be concerned about planning depth and volume of excision more carefully to minimise risk of future cancer and PTL? A key message on high risk groups and discussion on implications for clinical practice. I would be interested to know what advice clinicians should be giving to patients on their increased risk compared to the normal population and the size of the effect above this for single excision.
Thank you. We have included some more key clinical messages in the interpretation of findings section and in the key messages box according to the reviewer's recommendation. We have clarified that the risk is increased after any excision depth but the absolute increase in risk in Type 1 excisions as opposed to having CIN only is small, if any. The number of studies assessing this comparison is small and it is not possible to draw any firm conclusions. Further clinical messages are described in the conclusion section: 'When deciding to treat women of reproductive age, every effort should be made to perform a local treatment that will optimise the chances of a healthy pregnancy without compromising the completeness of the local treatment.' The interpretation of the level of risk introduced by treatment according to the depth can be used in light of the absolute increase in risk described in the tables for clinicians and patients' decision-making.

Minor comments:
Pg 6 line 48. Should be length of labour
This has been corrected.

Pg 17 line 20. Is there evidence to support statement that treatments have become less aggressive following increasing risk of preterm labour rather than the move to fertility preserving techniques? The Strander study referenced relates to move from hysterectomy to conserving cervix not reducing amount of cervical tissue removed/destroyed on cervical conserving techniques. The risk If residual or recurrent CIN is also important in this context since these women are the ones at high risk from repeat excisions.
Strander et al. document in the paper that 'the standardised incidence ratio for cervical or vaginal cancer was 2.05 (1.83 to 2.30) for women treated during the period 1958-1970, and 4.52 (3.47 to 5.80) for those treated in 2001-08.' Although one explanation may be indeed the reducing number of hysterectomies conducted in the more recent years, the move to less radical treatment techniques ie. the move from cold-knife conisation to LLETZ may be also a plausible explanation. Furthermore Albrechtsen et al. 2008 had documented in Norway a reduction in the risk of preterm birth with time that is possibly a reflection of the reduction in the radicality of treatment: "The time trend described could be explained by the fact that over the period studied, smaller amounts of cervical tissue were removed as new methods of conisation were introduced."
The sentence has been rephrased and a reference has been added.

Pg 18 lines 1-19. The message on risk of PTL and appropriate selection of high risk group for surveillance and intervention is confusing and I was not clear if there was any benefit or risk of disbenefits or not. It just require an additional line that the RR for women with a single type 1 treatment is not sufficiently increased to warrant antenatal monitoring of cervical length or currently offered interventions such as cervical cerclage. I do not know if the authors feel there data supports this.
We have amended the entire paragraph according to the reviewer's recommendation. We tried to convey a clear message although this is not easy, given that there is significant lack of data on the antenatal management of women after treatment for CIN. Clinical management is inconsistent and largely unit- or clinician-dependent. Some hospitals have already introduced a strategy of cervical length-indicated cerclage post conisation, even though the associated risks and benefits have not yet been fully assessed in properly designed studies. This strategy has become a 'standard of care' in some hospitals without strong evidence of benefit but it may also lead to worse outcomes. Women after cervical treatment represent a discrete group with known aetiology for preterm birth (either CIN or the effects of treatment for CIN), albeit with yet unclear pathogenic pathways. Future research should assess the value of antenatal interventions in this distinct group and devise a logical prevention strategy that can be tested and then applied. Furthermore the level of risk for preterm birth that would deem a woman suitable for intensive antenatal surveillance varies across units and therefore it is not possible to make recommendations on who should or should not attend specialist preterm birth prevention clinics as part of this manuscript.
In summary, the decisions on how to manage these women antenatally were beyond the scope of this meta-analysis and we are unable to make any recommendations based on our results. There is paucity of evidence on the value of the interventions and the population that would benefit from these and this should be the focus of future studies.

Pg 18. Line 38. It would be helpful to use a more specific term than 'young'.
This has been replaced with 'women of reproductive age'

Pg 18. Line 46. It would be helpful to be more specific than 'acceptable parameters'. Is this <10/12 mm?
We have clarified the sentence. By acceptable parameters we refer to a depth of excision or ablation of at least 8mm in order to include the crypts and any attempt to minimize reproductive morbidity

should be monitored in order to ensure that the treatment size is not compromised in the expense of high recurrence rates and worse oncological outcomes.

Reviewer 3: Björn Strander

Maria Kyrgiou and her collaborators have once again performed a large systematic review and metaanlysis in the field of associations between treatment for cervical precancer and pregnancy outcome. This is basically a needed update of their own metaanalysis in the BMJ 2008 and an important expansion and update of an Australian metaanlysis from 2011.

The methodology of the work is as far as I can judge immaculate. I find no relevant articles that are not included. The discussion is clear and concise and their conclusions are partly novel, as they state.

Thank you

I have one major concern, however. A very common, and quite difficult clinical situation for colposcopists is to assess, and inform a younger woman about the risk for future preterm birth as an effect of a standard treatment of a fully visible lesion (Transformation zone type 1). The length of an excision would probably only have to be 8 mm. This metaanalysis does not give an answer to the question if there is a risk to treat. The crucial analysis of cone length stratified for different comparison groups is unfortunately not done. The clinical relevant comparison group in this, and most cases, are untreated women with HPV infection and cervical lesions. Perhaps there are not sufficient data for this subgroup analysis, but I would expect this at least be commented upon as a limitation.
Many thanks for this important comment. We agree that this is an important point to be added and discussed. We have indeed performed the proposed subgroup analysis and we agree that this should be included in the results and discussed. We had such a plethora of data to present within the constrains of the word count and we omitted this analysis in the submitted version. This is now included in the material and methods, results and discussion of the revised draft. This subgroup analysis is also presented in Supplementary Table 4. This also includes a very informative comparison of the risk of preterm birth in treated women for different cone depth compared to women with CIN but no treatment. As the reviewer predicted the number of studies is small for many of the comparisons. The difference in the risk of preterm birth for small excisions as opposed to having CIN became insignificant but the number of studies was small and therefore definite conclusions could not be drawn. This has been added in the results, discussion and key messages.

Moreover: Data on the smallest, and most common i today's clinical practice, excisions have narrow confidence intervals, half of the studies are very small and the studies are heterogenous. This should call for specific caution in interpretation of the data on this group. Thus the first conclusion in the main findings (Page 14, line 51) and other parts of the manuscript should be rephrased as the authors have not shown that the risk for performing a small excision (<10mm) on a women with a cervical lesion should increase her risk for preterm birth and adverse sequele – as compared to abstaining from treatment.
This has been amended as per reviewers' recommendations and the limitations in this subgroup have been emphasized.

Minor points
There are no discussion of how acquisition of data on cone length can vary and the implications for the results. E.g. formalin fixed vs. fresh mesurement.
We agree that a lot of the studies do not describe the measurement technique in detail. The dimensions rely on retrospective analysis of the cone depth documented in the histopathology report in formalin fixed samples. A sentence reflecting this has been added to study limitations section together with future directions.

The double figures for cone length (e.g. ≤ 10/12 mm), volume (>3/4 cc) and pregnancy length (e.g. <32/34 w) should be explained in the text and the tables.
An explanation was added in the material and methods section and also as footnotes to the tables. Further subgroup analyses are available for the various cone depth cut-offs separately and not grouped together and these can be shown in supplementary tables if the editorial board feels that this would be useful for the readers. Currently this data is not shown as it was felt that this did not add useful information.

The terms depth and length for cone material are used alternatively in the manuscript.
We have corrected this and we use consistently the terminology 'depth' throughout the manuscript. This is based on the terminology agreed as part of a working group that met during the International HPV meeting in Lisbon in September 2015. The authors of the manuscript participated in that working group.

cc and mm (not mm3) are both used for volume.
The term 'cc' has been used throughout the manuscript

In table 3 the both the signs ≤ and ≥ are used, not explaining to what category the exact figure belong (e.g. 10 mm)
Different studies used different cut-offs for the cone depth measurements. Some included the values equal to the cut-off in the lower and others in the higher grouping. For meta-analytical pooling these had to be grouped together to allow analysis. Therefore the signs ≤ and ≥ are used correctly as some studies in both groups included values equal to these and other did not. We have included a footnote in table 3 to explain this and other reviewers' relevant comments. Analyses for the individual cone depths is also available upon request (data not shown) and this data can be added as supplementary tables if the editors feel this is necessary.

Pages 56 – 58 are forest plots. The difference between the analysis in page 56 and 57 is not clear to me.
We apologise for the mistake but it appears that we have uploaded the same figure twice by mistake. The figures should represent:
Figure 2: Meta-analysis on preterm birth (<37weeks) in treated versus untreated women
Figure 3: Meta-analysis on preterm birth (<37 weeks) in treated versus untreated women according to the cone depth a) ≤10/12mm; b) ≥10/12mm; c) ≥15/17mm d) ≥20mm
Figure 4: Meta-analysis on preterm birth (<37 weeks) in women treated with a cone depth a) ≥10/12mm versus ≤10/12mm; b) ≥15/17mm versus ≤17/15mm; c) ≥20mm versus ≤20mm
Figure 3 appears twice while Figure 2 does not appear. This will be corrected in the revised version.

This is a major work within the field. It deserves to be published and it is suitable for the BMJ audience. I suggest the authors get an opportunity to revise the manuscript
Thank you

Reviewer 4: Dr. Ioannis Biliatis

This is a very interesting meta-analysis from a team that has been studying the issue of premature birth and its relation to excision/ablation of cervix for many years. Of course there is no randomised study included in this meta-analysis and there is not going to be one in the future so the meta-analysis of observational studies are the best data that can be provided to lead clinicians in this very interesting and debatable issue.
The risk of preterm birth and reproductive morbidity after treatment for CIN has gone back and forth and then back again. I feel that many clinicians found hard to follow the literature in the last few years and find it even harder to properly counsel patients when seen in the colposcopy clinics about the associated risk.
A few examples of papers leading to this conflict come to my thoughts: Sadler (JAMA 2004), Noehr (Obs Gyn 2009), Ortoft (BJOG 2010), Albrechtsen (BMJ 2008) supporting substantial

increase in risk after treatment from population-based studies and Castanon (BMJ 2012) suggesting that this association may not be actually true (although the authors did recall these conclusions in the latest paper (BMJ 2014).
The manuscript is well presented and I must say it is the most comprehensive meta-analysis I have seen in the literature. All the important factors affecting outcomes have been included in the analysis. All the relevant obstetric outcomes are thoroughly assessed, singleton and multiple birth, single and multiple cones, nulliparous and parous women etc.
Thank you

I have the following comments to make

1) It would be interesting to see the subgroup analysis on the risk of preterm birth for different cone depth groups only for studies that have as a comparison group women with CIN but no treatment. The number of studies is likely to be small but I do think it would be useful to see the results for this subgroup comparison.
Thank you for these comments. This is a very valid point and a point also raised by one of the other reviewers, Dr Strander. We have indeed performed this analysis but the data was not shown. We have now included this in the material and methods, results and discussion. We have also included all the data as a Supplementary Table 4. The results, key messages and discussion have been amended accordingly.

2) Can the authors discuss the limitations of the previously published recent small meta-analyses.
We have added more comments about the limitations of meta-analyses in the discussion in the 'Interpretation in light of the evidence section'.

3) What does 10/12mm or 34/32weeks mean? This is not clear to me. Also if a cone biopsy had a depth of 10mm this was categorised as ≤10/12 or ≥10/12?
In the last line of second paragraph of preterm birth I assume it should be <6cc instead of <6mm
In addition, as before it is not clear for me how this categorisation works. For instance the comparison of ≥15/17 with ≤17/15 is quite confusing and it needs more clarification in the methods. Are these the groups used in the initial studies included in the analysis? In which category a 16mm deep cone belongs in?
Many thanks for the corrections. They have all been amended in the manuscript. The first and third points were also raised by Dr Stander and this has now been added in the manuscript and as a footnote to the table. Different studies used different cut-offs and these had to be grouped for the purposes of the analysis. Some studies used <10mm, 10-15mm, >15mm and others for example ≤10mm, 11-14mm, ≥15. These were grouped together in the different analysis. Separate analysis for the depths (ie. not grouped together) is available upon request and could be included as supplementary material if the editors and reviewers feel this would be informative. The second point was also raised by the previous reviewer and has been corrected.

4) Figure 2 and Figure 3 are exactly the same and depicting the same issue which is premature birth depending on the depth of the cone. I think the authors made a mistake and instead of presenting a figure of prematurity depending on the mode of treatment they submitted the same figure twice. The only thing that differs is the year. At the first figure is on descending and on figure 2 on random order. This needs to be corrected.
Also the total number of women presented in Figure 3 is different than the one presented in table 3 and the one written in the manuscript. For example the total number of women included in the first category of Figure 3 (A) should be 550929 and not 546824.
Indeed there has been a mistake during the revised uploaded version and this is now corrected in the revised version. The numbers have also been corrected and the corrected numbers included.

5) The authors used the Newcastle score to assess quality of retrospectve cohort studies but did not exclude any studies according to this score. They also included 2 studies with a score of 6. Why was this scoring system used (a lot of difficulties with this scoring system noted in the literature as different reviewers were scoring the same manuscripts differently, JCE Sep 2013) if it was not used to exclude studies? Did they have in mind a pre-defined score that would prompt them to exclude studies?
Indeed the scoring systems used have strengths and limitations. The Newcastle Ottawa score is considered to be on of the best tools to assess the quality of retrospective studies and this was recommended by the research group's statistician (MA) and also described in the Cochrane handbook. We calculate the median of the Newcastle-Ottawa scores and performed sensitivity analysis for studies above the median (=8.3) and after exclusion of studies with a score of 6 or below and 7 or below. This did not change the results of this analysis. This has been added in the results section. The results in detail are presented below in the reply to the editorial committee's comments that raised the same issue.

6) What is the difference between prematurity and spontaneous prematurity?
The term prematurity includes all pregnancies delivered before 37 weeks of gestation. This includes all spontaneous preterm births but also iatrogenic, medically-induced deliveries. The most clinically relevant outcome would be spontaneous preterm birth although this is not always reported in the registries. It is for this reason that both outcomes were included. An explanation of spontaneous prematurity has been included in the material and methods.

7) In results: you mention 4 prospective reports, but you reference 5
Many thanks. We have corrected this.

8) Page 15, Line 5: As expected….: I think I would rephrase that sentence to improve the language.
Thank you. We have rephrased this.

9) Table 3: Versus and not verses.
Suplem Table 1: the score (9) is missing from van Rooijen 1999
Methods 4t paragraph: the word spontaneous in between 2 ; doesn't make sense
Thank you pointing out these typos. They have all been corrected.

10) It will be interesting to see a comparison of treated women with a cerclage compared to untreated population in relation to prematurity?
It is not possible to extract data separately from the women that had a cerclage from the existing studies and such an analysis is not feasible.

11) How was the volume of the cone specified?
There was some heterogeneity in the way in which studies calculated the volume. Due to the retrospective nature of the studies it is generally calculated using the 'volume of a cone' equation based on measurements taken post-fixation by the reporting pathologist. Equations varied across papers. These limitations have been highlighted in the limitations of the study section. We felt that there is no need to include more detailed explanation of this in the manuscript.
Leiman et al.: the specimen were measured at a fixed state of the lateral and anteroposterior diameters and the height (length between apex and base). Cone volumes were calculated mathematically with the use of these measurements in 47 specimens.
Kitson et al.: in keeping with other studies, it was assumed that the cone of tissue removed during LLETZ approximated a hemiellipsoid shape. The volume excised was, therefore, calculated using 1/2_*4/3*π*a/2*b/2*c where a = transverse diameter, b = longitudinal diameter and c = depth.
Khalid et al.: the volume was calculated by multiplying the three dimensions as documented in the histopathology report, i.e. length*thickness*perimeter.

I think this paper is clearly needed and I think it would be very well fitted in a journal like the BMJ with wide readership. This meta-analysis 10 years after the first paper published in the Lancet by the first author will bring new insight on the existing evidence base with absolute risk that can also be used when counselling patients. I have no doubt that this is the most comprehensive meta-analysis I have seen and I think the results will be invaluable to clinicians.

In summary, I think this is a very important and well conducted meta-analysis in this field and it will be very well fitted in a journal like the BMJ with wide readership. It provides comprehensive evidence base in this field and particularly on the risk of preterm birth stratified by the cone size. I think it should be accepted for publication.

Thank you.

Editorial committee

The search is over 1 year old – December 2014, perhaps it's worth considering an update. Also, has the paper been elsewhere before it came to us? As per the ICMJE recommendations, we encourage all authors to share with us any correspondence from previous submissions to other journals.

We agree that the literature search required updating and we have run a new search until April 2016. Two further studies met the criteria for inclusion in this review and these studies have now been included in all tables and analysis. The numbers have been updated in all tables and figures and in the manuscript. The results and conclusions have not changed with the inclusion of these two studies. Furthermore we have included the PRISMA chart and references to include the new searches. The manuscript in the existing format has been submitted to the JAMA. The handling editor wanted to re-direct this to one of the JAMA Network Journals. Given the public health importance of the subject for women's health including cervical cancer screening and preterm birth, two of the major women's health issue, we declined the offer as we felt that the findings would be more widely disseminated if published in a general medical journal like the BMJ. We did not receive any comments from reviewers.

Apart from the summary of scores nothing else is done with this assessment - there are detailed tables in the appendix – it would be useful if you discussed this more.
Assessment of heterogeneity seems limited and glossed over, in the Methods you say 'If there was evidence of substantial heterogeneity, the possible reasons for this were investigated and reported' – how did you examine this? (i.e., did you carry out meta-regression); we felt this wasn't explored as well as it could have been.
We further developed the section of the summary of scores, the quality of the studies and the assessment of the heterogeneity. A sensitivity analysis was performed to assess the impact of the quality of the included studies on the outcomes, more specifically the risk of preterm birth (<37 weeks) for all the individual treatment techniques and for preterm birth in women with LLETZ when compared to women with disease but no treatment. We calculated the median for the Newcastle-Ottawa score, which was 8.3. We performed an analysis of all studies with a score above 8.3 for the outcomes above and we found no differences in the results. We further performed analysis to exclude all studies with a Newcastle-Ottawa score equal or less than 7 and equal or less than 6. We found that these analyses did not alter the findings, while there was the same direction and magnitude of effect. Furthermore, for the same outcomes and comparisons, we performed subgroup analysis to include separately studies that included a truly representative cohort versus those that included a somewhat representative cohort and also those that scored one or two marks for the comparability of the cohorts. Once again we did not find any evidence that this may have affected the effect of the analysis.

All these analyses have been conducted and are available upon request. Due to constrains with the word count, these were not included, but we would be happy to include them in supplementary tables if the editorial board feels that this is required.

An example of some of the sensitivity and subgroups analyses is included below. The point estimates for PTB < 37 weeks after any treatment or only LLETZ did not significantly change when the analyses were stratified according to the described criteria.

Outcome and population	RR, 95% CI, N of studies included
PTB, < 37 weeks	
Overall	1.78 [1.60, 1.98] (60 studies)
Only studies with score ≥9	1.76 [1.55, 2.00] (35 studies)
Only studies with score ≥8	1.73 [1.55, 1.94] (50 studies)
Only studies with score ≥7	1.78 [1.60, 1.98] (59 studies)
Studies with truly representative population	1.82 [1.55, 2.14] (15 studies)
Studies with somewhat representative population	1.76 [1.54, 2.01] (45 studies)
Studies with highly comparable groups **	1.79 [1.61, 1.99] (58 studies)
Studies with somewhat comparable groups *	1.65 [0.50, 5.49] (2 studies)
PTB, <37 weeks, LLETZ only	
Overall	1.56 [1.36, 1.79] (26 studies)
Only studies with score ≥9	1.62 [1.38, 1.89] (17 studies)
Only studies with score ≥8	1.54 [1.34, 1.78] (24 studies)
Only studies with score ≥7	1.56 [1.36, 1.79] (26 studies)
Studies with truly representative population	1.70 [1.43, 2.01] (4 studies)
Studies with somewhat representative population	1.56 [1.28, 1.90] (22 studies)
Studies with highly comparable groups **	1.56 [1.36, 1.79] (26 studies)
Studies with somewhat comparable groups *	NA
PTB, <37 weeks, LLETZ vs. colposcopy only	
Overall	1.33 [1.11, 1.60] (9 studies)
Only studies with score ≥9	1.46 [1.22, 1.74] (5 studies)
Only studies with score ≥8	1.33 [1.11, 1.60] (9 studies)
Only studies with score ≥7	1.33 [1.11, 1.60] (9 studies)
Studies with truly representative population	1.55 [1.40, 1.72] (1 study)
Studies with somewhat representative population	1.26 [1.03, 1.53] (8 studies)
Studies with highly comparable groups **	1.33 [1.11, 1.60] (9 studies)
Studies with somewhat comparable groups *	NA

Meta-regression was possible for some but not all possible confounders. When we tried to pursue a meta-regression for some chosen moderators/factors (ie. smoking, social history, previous PTB, parity etc), we found that only a proportion of the studies provided data for these. Therefore, we considered that any attempts to analyse these would further introduce bias due to a limited not representative samples of studies.

We performed meta-regression analysis to assess the impact of some factors for which we had available data on outcomes. These included the quality of the studies (based on the Newcastle-Ottawa score); year of study (1979-1989, 1990-1999, 2000-2009, 2010-2015); type of treatment (excision or ablation); type of comparator (external, internal –pretreatment pregnancies, internal – self matching, CIN but no treatment, HSIL but no treatment). The mono-variate analysis suggested that the type of treatment and comparator significantly affected the outcome but not the type of treatment and Newcastle-Ottawa score. When these factors were assessed in a multivariate regression analysis, the type of treatment (excision vs ablation) and the comparison group significantly affected outcomes while the other factors did not. When we further performed meta-regression restricting only to excisional treatments and using women with colposcopy/biopsy as

comparator, we found that all excisional techniques were associated with an increased risk of PTB. LLETZ was associated with a RR of 1.34 (1.10-1.64), CKC with RR 2.3 (1.39 – 3.85), LC with a RR 1.6 (0.91 – 2.87), NETZ with RR 4.26 (1.96-9.33). If we used women from the general population as a comparator, the RR were 1.32 (1.03-1.69) times higher. This is an estimation of how much the use of external comparator may overinflate the effect of treatment. However, when we compared effects with other comparison groups, there were no significant differences. More detailed data of the analysis can be provided upon request.

Some important points from these analyses have now been added in the manuscript and also in the limitations of the analysis section. Further details can be included in the manuscript if the editors feel this is necessary.

The title should not declare the findings. Perhaps use the term "adverse obstetrical outcomes"
The title has been amended: 'Adverse obstetrical outcomes after local treatment for cervical pre-invasive and early invasive disease according to the cone depth: a systematic review and meta-analysis.'

There are 69 studies but for many of the outcomes only a few contribute information. Furthermore the results of individual studies all seem to be in the same direction so the summary results only improve precision.
Indeed, a meta-analysis pools data from different studies in order to obtain more precise effect estimates than the in individual studies. In addition a meta-analysis documents the variation of effects among contributing studies. We agree that the majority of the studies point to the same direction. However there have been large linkage studies reporting opposing conclusions and this is one of the great strengths of meta-analytical techniques. We were able to map the entire literature, analyse according to different clinical subgroups and we were able to assess the true effect for many of the outcomes according to the comparison group and the cone length.

The first paragraph of the introduction is very UK-centric. Is this paragraph absolutely needed? If so, could it put the topic in a more international perspective?
We agree that the first paragraph is not required and indeed describes a very UK-centric summary of cervical cancer screening and prevention. This has been amended to reflect the recommendations of the editorial committee.

Excluding the 49 studies with no untreated reference population seems like throwing away information. If this were a network meta-analysis could those have been included?
In our meta-analysis, which aims to compare obstetrical harm associated with treatment of cervical cancer, it was noted that effects are influenced significantly by the choice of the comparator group and the depth of the excisional technique. We did not include studies without comparator group. This was an explicit exclusion criterion. Given the amount of studies retrieved with comparative data, sufficient information could be collected to obtain meaningful conclusions with good statistical power.

It was beyond the scope of this systematic review to assess studies that do not have an untreated comparison group and our inclusion criteria were set to include only studies with at least two compared groups. Single arm studies are at high risk of bias and would contaminate the effect of the analysis. Having been involved in network meta-analysis (NMA) previously (Kyrgiou JNCI 2006), we feel that inclusion of single arm studies would not be appropriate in this setting. The methodology of NMA allows for indirect comparisons of arms across studies that have not been directly compared in previous studies. That requires that the included studies have at least 2 comparison groups and it is not appropriate to include single arm studies due to the high risk of bias. Furthermore, the NMA methodology has been developed for analysis of randomized controlled trials (RCTs) while this has never been conducted in non-RCTs like the studies included in this review. We are not aware of any

NMA in cohort observational studies and as far as we know the methodology for this has not been developed and remains controversial.

Bottom of p 5 "We excluded studies..." Shouldn't the conjunction be OR instead of AND?
Thank you. This has been corrected.