Dear Mrs. Burch and editors of the BMJ,

We are pleased that our manuscript entitled: "Sacrospinous hysteropexy versus vaginal hysterectomy with uterosacral ligament suspension in women with uterine prolapse stage two or higher: a multicentre randomised non-inferiority trial", is being considered for publication in the BMJ.

We would like to thank the BMJ manuscript committee and reviewers for their time and valuable comments. We hope the changes in the manuscript are satisfactory for publication.

On behalf of the authors, yours sincerely,

Renée Detollenaere
Mail: r.j.detollenaere@isala.nl
Phone: +31614682330

Response to the comments of BMJ's manuscript committee:

*We appreciated the difficulties involved in randomizing women to different types of surgery, and commend the authors for this undertaking. We also felt that the clinical question is of interest.

*Reviewers raised concerns about the assessment of outcomes after only 1 year, and editors shared this concern. Was any assessment of later outcomes planned?
- As described in the SAVE U study protocol and the discussion of the manuscript women are followed annually till 60 months after surgery. We have added this information also to the Material and Methods section. As reviewer 3 suggests extending the discussion part on the follow-up duration we adjusted the text in the discussion the following:

First, our report is on a relatively short follow-up period of 12 months. However, results from registry studies suggest that the highest risk of prolapse surgery after hysterectomy is in the first year(s) after surgery. Furthermore not only long-term surgical outcome is important but also short-term secondary outcomes such as complication rate and postoperative recovery.33, 34 Women gave consent for follow-up till 60 months after surgery and these data will be further analyzed.

*Editors felt that the paper included many surgical specialty terms and may be difficult to follow for non-surgeons. Removing or clarifying some surgery-specific language may improve readability of the paper. We also thought it would be helpful to have a brief description of the procedures, and why one might be chosen over the other, for non-surgeons. Editors felt that the cover letter was very convincing, so perhaps some of the explanation from that could be included in the introduction.
- We clarified some surgery specific language: for example 1. a short description of the sacrospinous hysteropexy was added to the introduction, 2. In the material and methods section after anterior/posterior vaginal wall repair the term 'colporrhaphy' is added for those unfamiliar with this term, 3. In the methods section we explained in more detail what POP-Q stage 2 is about and 4. we added a description of both surgical procedures to the text including two links to two videos in which the sacrospinous hysteropexy procedure is being explained.
- As the individual POP-Q values might be too difficult to interpret for non-surgeons and are not really relevant for this group we removed these data from table 2 and 4. We also better explained the POP-Q system in the text as suggested by reviewer 3.
- Why one procedure might be chosen over the other is more explained in the introduction and some of the text of our cover letter was used as suggested.

*Editors felt that there might be an important patient perspective here, in that this might give women the opportunity to avoid a hysterectomy. This aspect could be emphasized more.
- We agree on this point. The opportunity for women to choose for hysterectomy was already described in the discussion section but we emphasized it more and also added this conclusion to the box ' what this study adds'.

*We would like to see all of the secondary outcomes listed in the trial registration explicitly reported in the paper. The secondary outcomes as described in the methods section seem to match what's in the trial registry, but the results section does not go on to give the outcomes in detail. Please provide these. It is acceptable to report the outcomes in a table with a descriptive summary statement in the text, but they should be easy to find.
- The secondary outcomes are now more extensively reported. We have changed the order in which the results are described as the primary and secondary outcomes were described halfway the results paragraph in the original manuscript. The primary outcome is described first (now table 3) and the secondary outcomes are presented thereafter (table 4 and 5) and we also provided an extra table with the secondary outcomes 'post-operative recovery' and 'sexual functioning' (table 6).

*Likewise, it wasn't completely clear where the original primary outcome, now the first secondary outcome, is reported. Please emphasize this outcome in the text.
- This outcome is reported in table 3 and we described the results of this outcome measure in more detail in the text.

*We had the following additional statistical concerns:
- Please provide a rationale for the choice of non-inferiority margin.
Recurrence rates after vaginal hysterectomy with respect to the apical compartment vary from 0-12%29, so that a failure rate including bothersome symptoms and repeat surgery after sacrospinous hysteropexy of 10% or more might be regarded as high. As we expected a failure rate of 3% based on outcomes of vaginal hysterectomy in a previous randomised study 16,30 the non-inferiority margin was set at 7%. This means that when the upper limit of the 95% confidence interval (CI) for the estimated difference in recurrence rate after sacrospinous hysteropexy

versus vaginal hysterectomy exceeds 7%, SH is inferior to vaginal hysterectomy.

- This is a non-inferiority study and hence the emphasis of the results should be on whether the observed difference (for the primary outcome), and its estimated 95% confidence interval, is within the non-inferiority margin. However, the emphasis throughout the results appears to be on formal comparative tests for all outcomes (primary and secondary). There are a plethora of outcomes (each with a p-value relating to a formal statistical comparison). Tables 4 and 5 contain 46 such comparisons in all. Which are the most important? I am concerned with the number of formal comparison tests for a supposedly non-inferiority study.
We agree the emphasis should be on the observed difference including 95% CI's. Therefore P-values were deleted from table 4 (table 3 in the revised manuscript). Furthermore differences and CI's were provided for the outcomes presented in table 4.

We adjusted for multiple testing and only P values <0.002 (0.05/32) were considered significant (although adjusting for multiple testing did not alter our results). P values were provided for the domain scores of the questionnaires regarding functional outcome, quality of life, post-operative recovery and sexual functioning as it was difficult to compare scores between the groups otherwise.

- Tables 4 and 5. It would seem more appropriate to present the differences here as SH-VH (rather than as VH-SH).
We changed table 4 as suggested (SH-VH).

- Table 6. The logistic regression analysis (which attempts to identify important prognostic factors) does not seem a natural adjunct to the non-inferiority study and perhaps should be omitted.
Although we believe that attempting to identify prognostic factors for recurrent prolapse after the performed surgical interventions is an interesting topic we agree that this subject might be beyond the scope of this article and therefore removed the text and table 6.

Response to the reviewer's comments
Reviewer: 1
*This is a well-written and well-designed trial on a potentially interesting topic. However, in my opinion, it presents some relevant limitations.The real question is why the uterine preservation? In 2013, Gutman and Maher stated that: "While the data on the efficacy of sacrospinous hysteropexy are conflicting, sacrospinous hysteropexy remains an alternative to vaginal hysterectomy in women who desire future fertility or uterine conservation. However, women with severe advanced prolapse desiring uterine conservation are at a high risk of recurrence and should consider alternative approaches to hysteropexy".
It is very interesting to evaluate the role of sacrospinous hysteropexy (SH) in women who require the fertility sparing, but in postmenopausal women it is unclear the role of the uterine preservation during transvaginal fascial anti-prolapse surgery. It could be useful if it is demonstrated that sacrospinous hysteropexy could be a surgical procedure more easy, more sure and more effective in comparison with vaginal hysterectomy with McCall or uterosacral ligament suspension. However, we know that at the actual state of the art, sacrospinous hysteropexy does not offer a higher cure rate and it is associated with a relevant rate of buttock pain (15-18% in the available literature). Moreover, this is true without any sexual or functional advantage. The authors proposed an interesting and well-designed non-inferiority trial but it is probably that only a superiority study could really change the actual clinical practice.
- The rationale for this study and the explanation for the question why uterus preservation is that at this moment no evidence for either vaginal hysterectomy or uterus preservation is available and this study aims to find answers on the question whether removal of the uterus is necessary or whether providing new apical support is durable and sufficient (described in the introduction). We argue that a non-inferiority trial is adequate to study this subject. There is a trend towards uterus preservation among gynaecologists and the outcomes of this trial provide evidence that SH is non-inferior to VH. Therefore women's preference can be leading after counseling that both uterus preservation and hysterectomy are save and effective.

*The 1-yr follow up is a short follow-up in case of anti-prolapse surgery. In this way, it is not possible to obtain strong evidence.
- We agree that 12 months follow-up is a relatively short follow-up period. This is however the first report on this comparison, and until now no large randomized trial have been performed comparing uterus preservation and hysterectomy. Therefore we feel that the presented information is relevant. Many women are treated with hysterectomy although performing hysterectomy is not an evidence-based practice. Furthermore secondary outcomes such as complication rate, postoperative recovery etc. add to the current knowledge on the procedures. Moreover, data from registry studies suggest that the highest risk for prolapse surgery after hysterectomy is in the first year(s) after surgery (Lykke Int Urogyn 2014; Altman AJOG 2008)

As commented above we changed the text the following:
First, our report is on a relatively short follow-up period of 12 months. However, results from registry studies suggest that the highest risk of prolapse surgery after hysterectomy is in the first year(s) after surgery. Furthermore not only long-term surgical outcome is important but also short-term secondary outcomes such as complication rate and postoperative recovery.33, 34 Women gave consent for follow-up till 60 months after surgery and these data will be further analyzed.

*SH presented a shorter operating time; however, I am not sure that 13 minutes are a difference clinically relevant. Moreover, in the VH group, the authors have performed a higher rate of posterior colporrhaphy and there was a very higher rate of procedures performed by residents. Therefore, it is obvious that the operating time was longer in VH group.
- We agree that 13 minutes difference is not a clinical relevant difference (especially not for patients). This difference can be relevant regarding cost-effectiveness although this was not a secondary endpoint of our study. We added a sentence to the discussion section that the higher rate of procedures performed by residents could have led to longer operating time in the VH group.

More procedures were performed by residents in the vaginal hysterectomy group. No statistically significant difference was found in surgical failure rate in patients who underwent surgery by either gynaecologist or resident but the higher rate of procedures performed by residents could have contributed to longer operation time in the vaginal hysterectomy group.

*The authors in the introduction stated that VH might increase the risk of stress urinary incontinence. However, the data on this topic are not at all homogeneous.
- We are aware of the fact that not all studies on this subject come to the same conclusions. Therefore we removed this sentence from the introduction and now only focus on the recurrence of prolapse after VH.

On the other hand women after vaginal hysterectomy may be at increased risk of recurrent prolapse since hysterectomy disrupts supportive structures of the pelvic floor. 13 To prevent future vaginal vault prolapse after hysterectomy additional vault suspension is recommended. 14,15

*Please report also the findings of 6-wk and 6-mo follow-up.
- As we stated that 12 months follow-up is an intermediate follow-up period to report about, we believe that the outcomes after 6 weeks and 6 months are not relevant for the readers and providing these data would make the results too extensive.

*Buttock pain rate is significantly higher in SH group. It should be mandatory to inform all the women before surgery of this side effect of the SH. It is possible that a postmenopausal woman would prefer a hysterectomy to the postoperative pain. Please debate this possibility in the discussion.
- We agree that women should be informed about the possible and most relevant complications and side effects of surgical interventions prior to surgery. Information regarding buttock pain was standard included in the pre-operative counseling in the participating hospitals. In view of the quick recovery of buttock pain and comparable VAS pain scores after surgery we believe that this difference in postoperative buttock pain should not be a reason to choose for vaginal hysterectomy.

We added the following text to the discussion part:

Two patients after uterine preservation ended up undergoing hysterectomy. In one patient this was because of persistent buttock pain. The overall rate of buttock pain in our study (9%) is in line with other studies and the majority of patients had spontaneous resolution.29 Preoperative counseling should include information about the potential risk of buttock pain postoperative.

*Please correct colporrhapy in colporrhaphy in table 3
- We adjusted text as suggested.

Reviewer: 2
This paper presents a randomised, non-inferiority study of sacrospinous hysteropexy vs vaginal hysterectomy with uterosacral ligament suspension. The study addresses an important and "live" question within urogynaecology: whether removal of the uterus is necessary in cases of uterine prolapse or whether providing new apical support is sufficient and durable. The introduction is clear and refers to the up to date and relevant literature. The study is well designed; the choice of non-inferiority design is apporpriate and the sample size calculation is valid in terms of likely effect size. The background recurrence rate (3%) is based on quoted figures from similar studies, and while perhaps a little lower than many of us would quote for standard vaginal hysterectomy, seems reasonable for the trial. The authors have dealt with the issues of "other compartment" surgery apppropriately, and the choice of outcomes is entirely appropriate. The change of definition of the primary outcome is an interesting approach, but it is well argued, and in practical terms, is essentially an unchanged outcome. I applaud them for doing this, and for their honesty. Details of surgical procedures, standardisation across the units, and all procedures are clearly mentioned and described.

*I have one question for the methodology: they state that patients with cervical elongation were included, but do not state how many of these there were, or which arm they were allocated to. It should be included, since cervical elongation alone can be considered NOT to be uterine prolapse.
- We agree that whether cervical elongation must be considered as uterine prolapse is arguable. A study by Berger et al. (IUJ 2012) showed that about 40% of women with pelvic organ prolapse have cervical elongation and the cervix is about 8.6 cm longer in women with pelvic organ prolapse. However no clear definition for cervical elongation is available at this moment and therefore it is difficult to give numbers on how many women had cervical elongation. We tried to clarify this sentence more in the manuscript:

Patients with co-existing anterior and/or posterior vaginal wall prolapse were allowed to participate and cervical elongation together with uterine prolapse was no reason to exclude women for participation.

1. The consort diagram indicates the women who received SH while randomised to VH (the crossovers) have been included in the SH group for analysis. Is this correct? If so, this is not intention to treat, since they should be retained within the VH group.
- This was not the case. The women who were randomised to VH and received SH were analysed in the VH group (ITT). We explained this in the results and clarified it more in the footnote of figure 1.

2. In line 52, the authors state "there were no significant differences in overall anatomical or surgical failure" while table 4 has a significant finding of higher failure in the posterior compartment for the VH group. This is discussed in the discussion, but should also be mentioned in the text of the results.
- The following sentence was added to the text:

No differences were found in anatomical recurrences in the different compartments except that more anatomical failures of the posterior vaginal wall were found after vaginal hysterectomy: sacrospinous hysteropexy 4.0% versus vaginal hysterectomy 14.1%, (95% CI for difference -18.2 to -1.8).

*Table 2: for "Overall POPQ stage prolapse" it should be made clear this is referring to UTERINE prolapse
- POP-Q stage uterine prolapse is reported at the top of the table. The overall POP-Q stage is equal to the POP-Q stage of the most severely prolapsed compartment. Overall POP-Q stage prolapse is therefore different from POP-Q stage uterine prolapse.

*Table 4: The significant finding of higher failure in the posterior compartment for the VH group should be mentioned in results.
- See comment above. This was added to the text.

Reviewer: 3
This is a very well written paper describing a high quality and original trial. The authors are correct to say that this trial is important since there is only very little evidence available comparing sacrospinous hysterectopexy and vaginal hysterectomy with uterosacral ligament suspension. I do have some suggestions to improve the manuscript:

* The study has a fairly short follow-up period. The authors address this shortly in the discussion section. This issue is more extensively addressed in the answer to the comment of JAMA reviewer 3. I would suggest to extend the discussion part on the follow-up duration with the arguments given in the answer to JAMA reviewer 3 (including the references given). Furthermore I wonder how the (pretty low?) recurrence rates of both interventions found in this study compare to the numbers found in other studies? I would add some information about this to the discussion as well.
-As suggested we extended the discussion part on the follow-up duration. We therefore adjusted the text in the discussion the following:

First, our report is on a relatively short follow-up period of 12 months. However, results from registry studies suggest that the highest risk of prolapse surgery after hysterectomy is in the first year(s) after surgery. Furthermore not only long-term surgical outcome is important but also short-term secondary outcomes such as complication rate and postoperative recovery.33, 34 Women gave consent for follow-up till 60 months after surgery and these data will be further analyzed.

- We compare the recurrence rates found in our study already in the discussion section (comparison with other studies). Overall our recurrence rates are comparable with previous (non-randomised) prospective studies. However the only randomized trial by Dietz et al found a higher recurrence rate after SH. It is difficult to explain the high recurrence rate after SH in that study. Possible explanations might be differences in surgical protocol or skills, precise definition of the recurrence outcome, and sample size (as the study by Dietz was underpowered and recurrence rate was a secondary endpoint).

*Another concern I have is the large amount of statistical tests that were performed. Did the authors consider a correction for multiple testing? This issue should probably be addressed in the discussion section, especially since the authors mention a 'trend towards significance' on one of the outcomes (page 11 line 54). This trend will obviously be lost when corrected for multiple testing.
- We adjusted for multiple testing and only P values <0.002 (0.05/32) were considered significant (although adjusting for multiple testing did not alter our results). This information is added to the statistical paragraph P values were provided for the domain scores of the questionnaires regarding functional outcome, quality of life, post-operative recovery and sexual functioning as it was difficult to compare scores between the groups otherwise.

- The sentence this reviewer is referring to is changed the following:

We found no difference in occurrence of anatomical anterior vaginal wall prolapse.

*The non-inferiority margin is set to 7%, but it is not completely clear to me how this margin is established. This should be clarified in the methods and/or discussion section.
- We included the following text in the methods section:

Recurrence rates after vaginal hysterectomy with respect to the apical compartment vary from 0-12%29, so that a failure rate including bothersome symptoms and repeat surgery after sacrospinous hysteropexy of 10% or more might be regarded as high. As we expected a failure rate of 3% based on outcomes of vaginal hysterectomy in a previous randomised study 16,30 the non-inferiority margin was set at 7%. This means that when the upper limit of the 95% confidence interval (CI) for the estimated difference in recurrence rate after sacrospinous hysteropexy versus vaginal hysterectomy exceeds 7%, SH is inferior to vaginal hysterectomy.

*The CONSORT statement on Reporting of noninferiority and equivalence randomized trials (Piaggio, JAMA 2012) states: "Judgement of the results in relation to the study hypothesis is based on the location of the whole CI in relation to Δ. For non-inferiority trials, the upper bound of the 2-sided (1-2α)x100% CI for the (deleterious) treatment effect or the upper bound of the 1-sided (1-α)x100% CI has to be below the margin Δ to declare that non-inferiority has been shown, with a significance level of α." (see also figure 1 of the CONSORT statement). In table 4 one can see that the upper bound of the 95% CI of the primary outcome lies above the 7% margin. This confused me, since it would mean non-inferiority could not be concluded. However from the BMC study protocol I understand that the non-inferiority margin is actually -7% and if this is the case, I think one would consider the lower bound of the 95% CI (as the authors described in the BMC study protocol). I am not a statistician, so maybe I am just wrong, but if it is confusing to me it might also be confusing to the general readers of the BMJ. Therefore I think it needs clarification in the manuscript.
- See also the comment above. We more clarified this in the manuscript. The reviewer is right concluding that when the upper bound of the 95% CI of the primary outcome lies above 7%, non-inferiority could not be concluded. As we changed the outcomes in table 3 (SH-VH) and the upper bound of the 95% for the primary outcome is 7%, SH is non-inferior to VH regarding the primary outcome difference -3.9%; 95% CI for difference -8.6 to 0.7.

*In the abstract and in the conclusions of the article the authors mention a benefit in favour of sacrospinous hysteropexy with regard to duration of surgery. However, this variable is not mentioned as a (secondary) outcome in the protocol nor in the methods section of this article and it is also not described in the results section (just in table 3). I think it should either be removed from the abstract and conclusions, or it should be more adequately described in the rest of the article.
- We agree with the reviewer that this outcome should not be described in the abstract and conclusions, as it was not a secondary outcome.

*The explanatory logistic regression analyses to identify possible risk factors for failure are not described in the study protocol or the trial registration. I wonder why the authors decided to add these analyses to the manuscript, which is already pretty extensive and informative without these extra analyses. I think they do not really add to the manuscript and I also think that they are not adequately described. Failure is defined as overall POP-Q stage 2 or greater or pessary use or repeat surgery but for which procedure? Based on the Ns it seems that both groups are put together? What would be the added value of this?
And why did the authors choose this definition of failure instead of the definition used as primary outcome for this study? From the manuscript it is also not clear whether the adjusted ORs are from a single multivariable model or from multiple models (each risk factor adjusted for centre, preoperative POP-Q stage and randomization outcome)?
- We have performed analyses for each separate risk factor (adjusted for centre, preoperative POP-Q stage and randomisation outcome) since the number of recurrences fell short to use a single multivariate model. Although we believe that attempting to identify prognostic factors for recurrent prolapse after the performed surgical interventions is an interesting topic, we agree that this subject might be beyond the scope of this article and therefore removed the text and table 6.

And some small remarks:
*The primary endpoint of this study is recurrent prolapse POP-Q stage 2 or higher in the apical compartment (uterus or vaginal vault) with bothersome prolapse complaints or repeat surgery for recurrent apical prolapse at 12 months follow-up. The definition of 'bothersome prolapse complaints' is described in the methods section on page 6 but not in the abstract and in the 'outcome measures' section on page 5. Instead of 'bothersome symptoms' (page 2 line 26) or 'bothersome prolapse complaints' (page 5 line 11) the authors might use 'bothersome bulge symptoms' to make clear which symptom(s) are considered as bothersome.
- We agree and adjusted text.

*All women with POP-Q stage two or higher uterine prolapse requiring surgery were asked for participation (page 4 line 9). This should be clarified: when do women require surgery? Is this a patients' decision? Is it decided by the surgeon? Or do they decide together?
- They decide together. This was explained in more detail in the methods section:

The decision to treat uterine prolapse surgically was a shared decision by the woman and her gynaecologist.

*According to table 2, not all women experienced bothersome bulge symptoms at baseline. The authors explained this in their answer to a comment of JAMA reviewer 2. I would advice to explain this in the article (for example in a footnote in table 2).
- We added this explanation to table 2 as suggested.

*I noticed some typos: on page 8 line 56 (differenceS), on page 13 line 25 (compaRtment), on page 20 line 42 (leadinG) and in table 4 on page 22 line 27 (, instead of .)
- We adjusted text.

*I think the description of the POP-Q measurement in the footnote of table 2 (a-c) might be unclear to general readers unfamiliar with the POP-Q system. This needs clarification, either in the methods section or in the footnote of table 2.
- As the individual POP-Q values might be too difficult to interpret for non-surgeons and are not really relevant for this group we removed these data from table 2 and 4. We also more explained the POP-Q system in the text as suggested and also changed the information on the POP-Q system in the footnote of table 2.

*How can POP-Q scores be missing at baseline (as mentioned in table 2) while a POP-Q stage 2 or greater is the main inclusion criteria of the study?
- These patients had a uterine prolapse POP-Q stage 2 or greater but unfortunately the complete POP-Q score was not recorded in the medical files. When using the web-based application to randomise patients, gynaecologists had to answer the question: uterine prolapse POP-Q stage 2 or higher requiring surgery with 'yes' and also needed to choose POP-Q stage (for stratification) before randomisation was possible.

1. Response letter.doc PDF HTML