We thank the editors and reviewers for their positive and constructive critique of our paper. We have below summarized those questions/comments that we identified in the reviewer report for us to respond to. In the event that we missed a question/comment please let us know. Our author response is structured as follows: 1. Reviewer/editor question, 2. Author response, and 3. Author action. All page and line numbers refer to the CLEAN version of the revised manuscript.

Editor and statistical comments

Statistical comment:
The treatment of missing data is unclear and requires some amendment. Baseline observation carried forward has been shown to bias results and should not be used. The authors state that this was in addition to the imputation incorporated in the main analyses, the form of which is not specified, as they state that analytic techniques were selected to avoid missing value issues.

Author response:
We thank the statistical reviewer for this comment and agree that this could have been better explained. In the mixed linear model we analyzed the full incomplete data set (i.e. including all available data at all time points) using restricted maximum likelihood estimation. This method does not impute any data, but rather uses the data from available cases to compute (restricted) maximum likelihood estimates. Like multiple imputation, this method gives unbiased parameter estimates and standard errors and has been reported to be at least as good as multiple imputation (Ranstam et al. 2012, BMC Medical Research Methodology).
To make a conservative estimate of improvements over time in each group we used the baseline observation carried forward method (though we agree this method has limitations). In addition, we have now conducted a further best/worst case scenario sensitivity analysis. In this analysis we have imputed the 25th percentile value at each time point for missing data in one group and the 75th percentile value for missing data in the other group and vice versa to test the robustness of the results, assuming the extreme situation in opposite directions for individuals lost to follow-up. The results of these analyses largely confirm our main analysis. In the analysis assuming best case scenario in traumatic tear patients and worst case in degenerative tear patients the results of no clinically relevant difference between groups remained. In the opposite analysis (i.e. degenerative=best case and traumatic=worst case), the results indicated that there could be a clinically meaningful difference between groups as the 95% CI slightly overlapped the 10 points cut-off (supplementary table 8). However, this was in favor of the degenerative tear group and thus opposite to the general presumption of better outcome in those with traumatic tears.

Author action:
The statistics section regarding the mixed linear model approach was rephrased to clarify the analysis approach (page 9-10, lines 44-9):
"The main outcome, between-group difference in change in $KOOS_4$ from baseline to 52 weeks was analysed using a repeated measures mixed linear model (Restricted Maximum Likelihood Estimation – REML) with subject nested within surgery site as random effects, and group (traumatic vs. degenerative) and time (baseline, 12 weeks and 52 weeks) as fixed effects.[35] The main analysis was changed to the current analysis as compared with the protocolled ANalysis of COVAriance (ANCOVA) approach as the mixed model approach (including all available data at all time points) using REML is considered a valid option to create unbiased parameter estimates and standard errors and takes into account that repeated measures are non-independent.[22]"

Further, the section describing the different sensitivity analysis was also rephrased (page 10, lines 28-56):
"Sensitivity analysis by adding degree of cartilage defects as a covariate in addition to age, sex and BMI was also conducted for the main outcome (i.e. $KOOS_4$) as well as a sensitivity analysis including all participant characteristics reported in table 1 with a p-value <0.10 and a fully adjusted model. Further sensitivity analyses were conducted to assess the robustness of the results with alternative definitions of traumatic and degenerative meniscal tears. Lastly, sensitivity analyses using a non-responder imputation approach (i.e. baseline observation carried forward) and a best/worst case scenario analysis imputing 25^{th} percentile data from participants with available data at the 12 and 52 weeks follow-up for the degenerative tear group and 75^{th} percentile data for the traumatic tear group (and vice versa) to investigate if this changed the interpretation of the results."
The results section was also rephrased to include the results of the best/worst case scenario sensitivity analysis (page 12-13, lines 50-7):
"Lastly, sensitivity analysis using null responder imputation and assuming best/worst case scenario of participants lost to follow-up did not alter the interpretation of data, though the analysis assuming best case for individuals lost to follow-up in the degenerative tear group and worst case in the traumatic tear group indicated the possibility of a clinically relevant larger improvement in the degenerative tear group (supplementary table 8)"
This analysis was also commented in the discussion section (page 15, lines 19-27):
"The direction of the resulting bias due to loss to follow-up of these participants is uncertain. However, sensitivity analyses with null responder imputation or assuming best/worst case scenario for patients lost to follow-up did not change the overall interpretation of data."

Statistical comment:
Although the authors have made amendments from the protocol, they give justification and this does not appear to have substantially affected the results.

Author response:
Thank you for this comment. We acknowledge that we have made changes from what was reported in the protocol paper. We have tried to report these as transparently as possible.

Author action:
No changes made to the manuscript.

Statistical comment:
As noted by Harris, there is an inconsistency between the difference used in the power calculation (8 points) and the difference used for excluding clinically important effects in the results (10 points).

Author response:
We agree with the statistical editor that this should have been discussed in more detail. The minimal clinically relevant difference/change on the KOOS has been suggested to be around 8-10 KOOS points (Roos et al. 2003, Health and Quality of Life Outcomes). However, there is no consensus on the exact number. Thus, we powered the study to be able to detect the smallest potential clinically relevant change. Prior to analysis, we decided to use the 10 point cut-off in the interpretation of what constituted a clinically relevant change, as this has been used in previous studies comparing

the effect of surgery with exercise therapy in improving patient reported outcomes for patients with ACL tears, having knee replacement and patients with degenerative meniscal tears (Frobell et al. 2010, NEJM; Skou et al. 2015, NEJM and Kise et al. 2016, BMJ). This was described in the statistical section in methods (page 10, lines 28-36):
"The minimal clinically important difference on the KOOS is considered to be 8-10 points.[36] In the present study, a 95% CI excluding differences greater than 10 KOOS-points between groups was interpreted as indicating the absence of a clinically meaningful difference between groups as previously done in randomised trials on knee patients using the $KOOS_4$ as the primary outcome.[26-28]"

Author action:
We agree with the statistical editor about the importance of this and have added a paragraph on this in the discussion section (page 14, lines 37-51):
"The present study was powered to detect an 8-point difference in improvement in KOOS scores between groups, as 8-10 KOOS points was considered a clinically relevant difference when the study was planned.[36] However, there is no consensus on the specific value that constitutes a clinically relevant difference/change on the KOOS score. Prior to analysis, we decided to interpret a 95% CI excluding differences greater than 10 KOOS-points between groups as absence of a clinically meaningful difference between groups, as this cut-off was used in randomised trials comparing surgery to exercise therapy for patients with different knee pathologies.[26-28]"

Statistical comment:
Including in the sensitivity analyses only the participant characteristics from table 1 that are significant at 10% level may potentially miss associations. The excluded variables should be checked for association after adjustment for the other factors.

Author response:
We agree with the statistical editor that we should have performed a fully adjusted analysis including all variables reported in table 1. This sensitivity analysis has been performed and did not change the interpretation of the results (sensitivity analysis 3, supplementary table 2).

Author action:
The results section was slightly changed to include this (page 12, lines 26-31):
"In sensitivity analysis of the main outcome (i.e. $KOOS_4$) adding degree of cartilage defects as a covariate did not change the interpretation of results, which was similar for analysis including all participant characteristics with a $p<0.10$ and the fully adjusted analysis (supplementary table 2)."

Statistical comment:
Some individuals are lost to follow up. The numbers presented in table 2 suggest that the losses may be a biased subsample and this should be investigated. For example, the average change from 3 to 12 months of KOOS4 in the TT group is 61.8-57.4 = 4.4 and for the DT group this is 66.2-58.7 = 7.5, a difference of 4.4-7.5 = -3.1, which is quite different to the difference of -5.3 in those that completed both assessments.

Author response:

Thank you for this comment. To try to take into account that those who dropped out could be a selected subsample with either a more inferior or favorable outcomes we have now conducted a best/worst case scenario sensitivity analysis. In this analysis we have imputed the 25th percentile value at each time point for missing data in one group and the 75th percentile value for missing data in the other group and vice versa to test the robustness of the results assuming extreme situation in opposite directions for individuals lost to follow-up. The results of these analyses largely confirm our main analysis. In the analysis assuming best case in traumatic tear patients and worse case in degenerative tear patients the results of no difference between groups remained. In the opposite analysis (i.e. degenerative=best case and traumatic=worst case), the results indicated that there could be a clinically meaningful difference between groups as the 95% CI overlapped the 10 points cut-off slightly (supplementary table 8). However, this was in favor of the degenerative tear group and thus opposite to the general presumption of better outcome in those with traumatic tears.

Author action:
The section describing the different sensitivity analysis was rephrased to include this additional analysis (page 10, lines 37-56):
"Sensitivity analysis by adding degree of cartilage defects as a covariate in addition to age, sex and BMI was also conducted for the main outcome (i.e. $KOOS_4$) as well as a sensitivity analysis including all participant characteristics reported in table 1 with a p-value <0.10 and a fully adjusted model. Further sensitivity analyses were conducted to assess the robustness of the results with alternative definitions of traumatic and degenerative meniscal tears. Lastly, sensitivity analyses using a non-responder imputation approach (i.e. baseline observation carried forward) and a best/worst case scenario analysis imputing 25th percentile data from participants with available data at the 12 and 52 weeks follow-up for the degenerative tear group and 75th percentile data for the traumatic tear group (and vice versa) to investigate if this changed the interpretation of the results."
The results section was also rephrased to include the results of the best/worst case scenario sensitivity analysis (page 12-13, lines 50-7):
"Lastly, sensitivity analysis using null responder imputation and assuming best/worst case scenario of participants lost to follow-up did not alter the interpretation of data, though the analysis assuming best case for individuals lost to follow-up in the degenerative tear group and worst case in the traumatic tear group indicated the possibility of a clinically relevant larger improvement in the degenerative tear group (supplementary table 8)"
This analysis was also commented in the discussion section (page 15, lines 19-27):
"The direction of the resulting bias due to loss to follow-up of these participants is uncertain. However, sensitivity analyses with null responder imputation or assuming best/worst case scenario for patients lost to follow-up did not change the overall interpretation of data."

Editor comments:
- One editor felt it was difficult to get a view on how much this added to the existing literature, but felt it read easily and seemed straightforward.

- Several editors were supportive and felt the paper was clinically relevant.

- Another editor said that until there is a trial this looks like it may be the best we have and it's good to cover an orthopaedic topic that is also relevant to primary care doctors.

Author response:

We thank the editors for their positive comments on our paper and agree.
In a recent systematic review and meta-analysis of knee arthroscopy for degenerative knee disease/meniscal tears (Thorlund et al. 2015, BMJ) our search, that did not include an age criterion, identified NO randomised trials comparing the effect of meniscal surgery for younger patients or for patients with traumatic meniscal tears with non-surgical treatment such as sham surgery or exercise (as noted by one of the editors).
Many surgical procedures are based on a 'mechanical' hypothesis that removing specific pathologies will relieve symptoms (Aspenberg 2014, Acta Orthop). However, these theories have rarely been confirmed in randomised trials. Even though this study is of observational nature and cannot determine the efficacy of surgery for patients with traumatic tears *per se,* the results question the general assumption that younger patients with traumatic meniscal tears experience better symptom relief than patients with degenerative tears.
Given the large placebo effect of arthroscopic meniscal surgery for patients with degenerative tears, we believe that this study provides important equipoise to justify testing the effect of meniscal surgery for young patients with traumatic tears against non-surgical treatment in future randomized trials. Another strength is that the present observational study report the main pre-specified analysis on prospectively collected data from Knee Arthroscopy Cohort Southern Denmark (KACS), which has been pre-registered (ClinicalTrials.gov identifier: NCT01871272) and the design of the cohort and the main analysis has also been described in detail in a protocol paper (Thorlund et al. 2013).

Author action:
No changes made to the manuscript.

Patient editor comment:
Our patient editor (who didn't participate in the meeting) wrote in a note that she couldn't see in the patient involvement box a description of how you had decided which outcomes to get patients to report. She wondered that if you had no involvement, how do you know these are relevant outcomes? Finally, she added that if by contrast the outcomes were actually set by patients in any way, why have you not described that in the patient involvement box?

Author response:
Thank you for this very relevant and important question. When designing this cohort (in 2011/12) it was (unfortunately) not typical to include patients in the planning of studies. For the KACS cohort and for this specific study the main outcome was the Knee Injury and Osteoarthritis Outcome Score (KOOS). The KOOS has been translated and validated in more than 40 languages and is one of the most used patient-reported outcome measures to assess pain, symptoms, function during daily activities and sports and recreation and quality of life in patients with knee pathology. To ensure that the KOOS covers outcomes (or domains) that are important to patients with knee symptoms the questionnaire was developed and refined in an expert group consisting of orthopedic surgeons, physiotherapists and patients with knee injury (Roos et al. 1998, J Orthop Sports Phys Ther 78(2)).

Author action:
We do not feel that the involvement of patients in developing the KOOS justifies amendment of the 'Patient involvement' section in the present study. However, to clarify the involvement of patients in developing the main outcome of this study (i.e. the KOOS) we have amended the methods section describing the main outcome (page 6-7, lines 55-14).

"The KOOS was developed with involvement of patients and is intended for individuals with knee injuries that can result in post-traumatic OA such as meniscus injury, anterior cruciate ligament (ACL) injury and chondral injury.[23] $KOOS_4$ is the mean score of four of the five KOOS subscales (i.e. excluding the ADL subscale). The KOOS questionnaire has been validated in individuals undergoing APM,[23-25] and the $KOOS_4$ has been used in trials assessing the effect of knee surgery.[26-28]"

Reviewer 1:

Reviewer comment:
The study is a comparative cohort study comparing the outcomes of arthroscopic partial meniscectomy (APM) in those with traumatic meniscus tears to those with degenerative meniscus tears (TT vs DT). The question is an important one because APM surgery is common and because although many previous studies have compared APM for DT to alternative treatments, there have been no such trials for TT.
A potential problem with this study is the poor definition of the type of tear (TT vs DT). The authors address this by providing several sensitivity analyses using different definitions, however this still requires some clarification (see below). The comments below mainly concern clarity of definitions as I found this somewhat confusing at times. I have no major issues with the methodology.

Author response:
We thank the reviewer for acknowledging the importance of our study and agree with the comment regarding the lack of trials comparing the effect of arthroscopic partial meniscectomy with alternative treatments for patients with traumatic tears. We regret that we have not been clear enough about our definitions of traumatic and degenerative tears. We will provide specific answers relating to specific questions regarding this issue below.

Author action:
For actions regarding traumatic and degenerative tear definitions, see specific replies below.

Reviewer comment:
The definition of TT vs DT is unclear, and possibly varies from the published study protocol. The protocol states that all patients aged 18-34 and those aged 35-55 with 'violent' tears are TT – the rest being DT. The definition used in the study is not clear. The authors provide the definition (TT is defined as "Participants aged 18-34 years and replying that symptoms evolved as a result of a 'specific incident' or 'violent incident' AND participants aged 35-55 years replying symptoms evolving as a result of a 'violent incident'") but then state that it was changed "to include all participants aged between 18-55". All patients aged 18-55 were always included in all definitions, it is how they were divided into TT and DT that needs clarification. Was the definition changed prior to analysis (as stated)? Please provide a clear definition of the definition used in the analysis.

Author response:
We thank the reviewer for the opportunity to clarify this issue. We agree that the last statement (i.e. "to include all participants aged between 18-55") in the description of deviation from the study protocol is not very informative.

The original definition in the protocol paper was:
Traumatic:
Participants aged 18-34 years with either slowly, specific incident (i.e. semi-traumatic) or violent incident (i.e. traumatic) symptom onset AND patients aged 35-55 years with violent incident (i.e. traumatic) symptom onset.
Degenerative:
Participants 35-55 years with slowly or specific incident (i.e. semi-traumatic) symptom onset and having knee symptoms more than 6 months.
Admittedly, this definition was not optimal, as several participants in the age group of interest were left out of the analysis (i.e. those age 35-55 years with violent incident (i.e. traumatic) symptom onset was not categorized as well as to those aged 35-55 years with shorter symptom duration than 6 months).

Thus, *prior* to analysis, we decided to modify the definition to categorize all participants (including those not covered by the previous definition). We relaxed the criteria on symptom duration as we recognized this was also related to wait list times and not necessarily only to the duration of symptoms.
The modified definition used in the analysis was:
Traumatic:
Participants aged 18-34 years and replying that symptoms evolved as a result of a 'specific incident' or 'violent incident' AND participants aged 35-55 years replying symptoms evolving as a result of a 'violent incident'.
Degenerative:
Participants aged 18-34 years and replying that symptoms 'evolved slowly over time' AND participants aged 35-55 years replying symptoms evolving as a result of a 'specific incident' or 'evolved slowly over time'.

Author action:
We have changed the description of the change of definition in comparison to the protocol as we believe this sentence led to the confusion of the definition. This description now reads (page 9, lines 21-29):
"This definition was slightly changed prior to analysis as some participants between 18-55 years of age were not categorised as having either a traumatic or degenerative tear by the previous definition outlined in the study protocol.[22] Furthermore, the criterion on duration of symptoms was relaxed as it turned out that this was likely to be influenced by referral time to the orthopaedic department."

Reviewer comment:
The authors provide several sensitivity analyses based on the definition of TT vs DT, which strengthens the study. However, I would have expected one of those definitions to be the definition used in the original published protocol in order to better address bias due to selective outcome analysis.

Author response:
Thank you for this comment. As described in response to the previous question the original definition excluded a number of participants in the age group between 18-55 years as these were not categorized by the initial definition described in the protocol. Thus, we have refrained from including this analysis in the study.

Author action:
No changes made to the manuscript.

Reviewer comment:
There is inconsistency between the power analysis and the primary outcome. The power analysis is based on a difference of 8 points in KOOS but the threshold for clinical importance chosen was a difference of 10 points.

Author response:
We agree with the reviewer that this should have been discussed in more detail. The minimal clinically relevant difference/change on the KOOS has been suggested to be around 8-10 KOOS points (Roos et al. 2003, Health and Quality of Life Outcomes). However, there is no consensus on the exact number. Thus, we powered the study to be able to detect the smallest potential clinically relevant change. Prior to analysis, we decided to use the 10 point cut-off in the interpretation of what constituted a clinically relevant change, as this has been used in previous studies comparing the effect of surgery with exercise therapy in improving patient reported outcomes for patients with ACL tears, having knee replacement and patients with degenerative meniscal tears (Frobell et al. 2010, NEJM; Skou et al. 2015, NEJM and Kise et al. 2016, BMJ). This was described in the statistical section in methods (page 10, lines 28-36):
"The minimal clinically important difference on the KOOS is considered to be 8-10 points.[36] In the present study, a 95% CI excluding differences greater than 10 KOOS-points between groups was interpreted as indicating the absence of a clinically meaningful difference between groups as previously done in randomised trials on knee patients using the $KOOS_4$ as the primary outcome.[26-28]"

Author action:
We agree with the reviewer about the importance of this and have added a paragraph on this in the discussion section (page 14, lines 37-51):
"The present study was powered to detect an 8-point difference in improvement in KOOS scores between groups, as 8-10 KOOS points was considered a clinically relevant difference when the study was planned.[36] However, there is no consensus on the specific value that constitutes a clinically relevant difference/change on the KOOS score. Prior to analysis, we decided to interpret a 95% CI excluding differences greater than 10 KOOS-points between groups as absence of a clinically meaningful difference between groups, as this cut-off was used in randomised trials comparing surgery to exercise therapy for patients with different knee pathologies.[26-28]"

Reviewer comment:
For the sensitivity analysis using "null responder imputation" (which used baseline value carried forward, I cannot see how this was done "in addition to" the imputation methods used in the main analysis. Surely, it can only be one or the other. I also question the validity of using last observation carried forward using baseline observations. This may require input from a statistician.

Author response:
We agree that this could have been better explained. In the mixed linear model we analyzed the full incomplete data set (i.e. including all available data at all time points) using restricted maximum

likelihood estimation. This method does not impute any data, but rather uses the data from available cases to compute (restricted) maximum likelihood estimates. Like multiple imputation, this method gives unbiased parameter estimates and standard errors and has been reported to be at least as good as multiple imputation (Ranstam et al. 2012, BMC Medical Research Methodology).
To make a conservative estimate of improvements over time in each group we used the baseline observation carried forward method (though we agree this method has limitations). In addition, we have now conducted a further best/worst case scenario sensitivity analysis. In this analysis we have imputed the 25th percentile value at each time point for missing data in one group and the 75th percentile value for missing data in the other group and vice versa to test the robustness of the results, assuming the extreme situation in opposite directions for individuals lost to follow-up. The results of these analyses largely confirm our main analysis. In the analysis assuming best case scenario in traumatic tear patients and worst case in degenerative tear patients the results of no clinically relevant difference between groups remained. In the opposite analysis (i.e. degenerative=best case and traumatic=worst case), the results indicated that there could be a clinically meaningful difference between groups as the 95% CI slightly overlapped the 10 points cut-off (supplementary table 8). However, this was in favor of the degenerative tear group and thus opposite to the general presumption of better outcome in those with traumatic tears.

Author action:
The statistics section regarding the mixed linear model approach was rephrased to clarify the analysis approach (page 10, lines 37-56):
"Sensitivity analysis by adding degree of cartilage defects as a covariate in addition to age, sex and BMI was also conducted for the main outcome (i.e. $KOOS_4$) as well as a sensitivity analysis including all participant characteristics reported in table 1 with a p-value <0.10 and a fully adjusted model. Further sensitivity analyses were conducted to assess the robustness of the results with alternative definitions of traumatic and degenerative meniscal tears. Lastly, sensitivity analyses using a non-responder imputation approach (i.e. baseline observation carried forward) and a best/worst case scenario analysis imputing 25^{th} percentile data from participants with available data at the 12 and 52 weeks follow-up for the degenerative tear group and 75^{th} percentile data for the traumatic tear group (and vice versa) to investigate if this changed the interpretation of the results."
The results section was also rephrased to include the results of the best/worst case scenario sensitivity analysis (page 12-13, lines 50-7):
"Lastly, sensitivity analysis using null responder imputation and assuming best/worst case scenario of participants lost to follow-up did not alter the interpretation of data, though the analysis assuming best case for individuals lost to follow-up in the degenerative tear group and worst case in the traumatic tear group indicated the possibility of a clinically relevant larger improvement in the degenerative tear group (supplementary table 8)"
This analysis was also commented in the discussion section (page 15, lines 19-27):
"The direction of the resulting bias due to loss to follow-up of these participants is uncertain. However, sensitivity analyses with null responder imputation or assuming best/worst case scenario for patients lost to follow-up did not change the overall interpretation of data."

Reviewer comment:
Terminology around KOOS vs KOOS4 vs KOOS subscales vs KOOS4 subscales is confusing. If KOOS4 was used, then please define it once and use the term "KOOS4" throughout the results.

Author response:

Thank you for this comment. The different terminology was used because these actually refer to different things. The wording 'KOOS' is used when referring to the entire questionnaire, 'KOOS4' is used when referring to the main outcome (i.e. incorporating 4 of 5 of the individual KOOS subscales) and 'KOOS subscales' is used when we are referring to the 5 individual subscales of the KOOS (used as secondary outcomes in the study). We agree with the reviewer that this was not clearly explained.

Author action:
The methods section on the KOOS outcome has been rephrased into (page 6-7, lines 48-27):
"Knee Injury and Osteoarthritis Outcome Score (KOOS). The KOOS consists of 5 subscales covering pain, symptoms, function during daily activities (ADL), sport and recreational function (Sport/Rec) and quality of life (QOL). Each KOOS subscale ranges from 0-100 points with 0 representing extreme knee problems and 100 representing no knee problems.[23] The KOOS was developed with involvement of patients and is intended for individuals with knee injuries that can result in post-traumatic OA such as meniscus injury, anterior cruciate ligament (ACL) injury and chondral injury.[23] $KOOS_4$ is the mean score of four of the five KOOS subscales (i.e. excluding the ADL subscale). The KOOS questionnaire has been validated in individuals undergoing APM,[23-25] and the $KOOS_4$ has been used in trials assessing the effect of knee surgery.[26-28]
The main outcome, in the present study, was the between-group difference in change from baseline to 52 weeks in the mean score on the $KOOS_4$. In the study protocol and trial registration it was stated that change from baseline to 52 weeks on all five KOOS subscales was the main outcome.[22] However, prior to analysis we decided to use the $KOOS_4$ to simplify interpretation by having only *one* main outcome. To assist the clinical interpretation of our main outcome (i.e. $KOOS_4$) all five subscales of the KOOS were included as secondary outcomes."

Reviewer comment:
In table 3, the term "NTT" (a term used in the protocol for non traumatic tears) is used instead of "DT" (used throughout this manuscript)

Author response:
Thank you for picking up on this mistake.

Author action:
'NTT' was changed into 'DT' in table 3. Remaining tables were also checked and corrected for consistency.

Reviewer comment:
The protocol lists many other outcomes, such as SF36, global perceived effect and adverse events. Are these to be reported in separate publications? This information would be helpful.

Author response:
The reviewer is correct. As this was a prospective cohort study many other outcomes were collected in the KACS cohort. This study report the main protocolled analysis. Subsequent papers will report on other secondary outcomes. Further, as declared in the cover letter when submitting this manuscript, 3 manuscripts describing other results from the study of this cohort have already been published (Pihl et al. 2016a, Acta Orthop; Tornbjerg et al. 2016, BJSM; Pihl et al. 2016b, Acta

Orthop).

Author action:
No changes made to the manuscript.

Reviewer 2:

Reviewer comment:
The paper is very well written in all sections with respect to language and grammar. The script satisfies most of the STROBE criteria. However, I have a few comments and queries about the paper to the authors.

Author response:
Thank you for the positive comments.

Author action:
Specific queries are replied below.

Reviewer comment:
Details of structural pathology of menisci was collected in the methods and presented in the tables but not presented under results nor discussed the results in discussion with respect to the type of tear in relation to outcome. Salient findings and their relevance may be discussed for these as well as associated pathology in the knee.

Author response:
Thank you for this relevant question. The main analysis in this study was based on testing the difference in patient reported improvement over time between patients with traumatic vs. degenerative tears, based on a definition including age and patient reported symptom onset. Indeed, the structural pathology of the meniscus or the specific tear type has been suggested to impact improvement after surgery. Therefore we did several sensitivity analyses to address this. First, we performed a sensitivity analysis where we based the traumatic/degenerative definition on the meniscal tissue status (i.e. degenerative vs. non-degenerative) but this did not change the interpretation of no difference between groups (sensitivity analysis 6, supplementary table 5). We also tried to combine meniscal tissue quality with patient reported symptom onset for defining traumatic/degenerative tears, which also did not change the interpretation of the results (sensitivity analysis 7, supplementary table 6). Lastly, we also performed an extra sensitivity analysis, as suggested by the statistical editor, including all variables listed in table 1 (including mechanical symptoms, meniscal tear type, etc.), which again did not change the results (sensitivity analysis 3, supplementary table 2).

Author action:
We have added to the results and discussion that additional sensitivity analyses including meniscal and other pathology did not change the interpretation of the results.
Results have been rephrased into (page 12, lines 26-31):

"In sensitivity analysis of the main outcome (i.e. $KOOS_4$) adding degree of cartilage defects as a covariate did not change the interpretation of results, which was similar for analysis including all participant characteristics with a $p<0.10$ and the fully adjusted analysis (supplementary table 2)."
The discussion section on sensitivity analyses have been expanded (page 14, lines 21-36):
"There is no consensus on the definition of 'traumatic' and 'degenerative' tears, and there is a 'grey zone' between the two. Therefore, we conducted several sensitivity analyses testing the robustness of the results by adjusting for meniscal and other structural knee joint pathologies observed at surgery and by applying different definitions of traumatic and degenerative meniscal tears. Even though the level of statistical significance and the direction of the results varied slightly in these analyses, the overall interpretation of no clinically meaningful difference between groups remained (supplementary material)."

Reviewer comment:
What was the effect of tear type on the main outcome? Please discuss.

Author response:
There was no effect of tear type when adjusted for in the analysis (sensitivity analysis 3, supplementary table 2)

Author action:
We have added to the results and discussion that additional sensitivity analyses including meniscal and other pathology did not change the interpretation of the results.
Results have been rephrased into (page 12, lines 26-31):
"In sensitivity analysis of the main outcome (i.e. $KOOS_4$) adding degree of cartilage defects as a covariate did not change the interpretation of results, which was similar for analysis including all participant characteristics with a $p<0.10$ and the fully adjusted analysis (supplementary table 2)."
The discussion section on sensitivity analyses have been expanded (page 14, lines 21-36):
"There is no consensus on the definition of 'traumatic' and 'degenerative' tears, and there is a 'grey zone' between the two. Therefore, we conducted several sensitivity analyses testing the robustness of the results by adjusting for meniscal and other structural knee joint pathologies observed at surgery and by applying different definitions of traumatic and degenerative meniscal tears. Even though the level of statistical significance and the direction of the results varied slightly in these analyses, the overall interpretation of no clinically meaningful difference between groups remained (supplementary material)."

Reviewer comment:
Did the authors have a chance to look at the effect of duration of symptoms and outcome?

Author response:
There was no effect of symptom duration when adjusted for in the analysis (sensitivity analysis 3, supplementary table 2)

Author action:
No changes made to the manuscript.

Reviewer comment:
Sensitivity analysis was mentioned to have been performed by adding degree of cartilage defects in addition to age, sex and BMI for the main outcome under the methods section. What was the outcome? Please discuss findings in the discussion. Although cartilage loss was presented under results, other parameters were not discussed.

Author response:
Thank you for this question. The interpretation of the results did not change when further adjusting the model for degree of cartilage defects (sensitivity analysis 1, supplementary table 2). In this study several sensitivity analyses including many different variables where conducted to test the robustness of the results. In essence none of these changed the main interpretation of no clinical relevant difference in change between groups from baseline to 52 weeks. Because of the very robust results, we have refrained from commenting on specific variables/pathologies as these did not seem to change the results.

Author action:
Nevertheless, the discussion on sensitivity analyses was slightly expanded (page 14, lines 21-36):
"There is no consensus on the definition of 'traumatic' and 'degenerative' tears, and there is a 'grey zone' between the two. Therefore, we conducted several sensitivity analyses testing the robustness of the results by adjusting for meniscal and other structural knee joint pathologies observed at surgery and by applying different definitions of traumatic and degenerative meniscal tears. Even though the level of statistical significance and the direction of the results varied slightly in these analyses, the overall interpretation of no clinically meaningful difference between groups remained (supplementary material)."

Reviewer comment:
Data and its meaning is not clear in supplementary table 1. What do the authors want to convey here? The legend of table indicates baseline data but why add assessed at 52 weeks here? Is the data baseline or at 52 weeks?

Author response:
We agree with the reviewer that this may be confusing. The data in the supplementary table 1 is a loss to follow-up analysis comparing baseline characteristics of patients who were lost at the 52-week follow-up and those assessed at the 52-week follow-up. This is presented both for patients defined as having traumatic and degenerative tears.

Author action:
We have rephrased the column headlines in supplementary table 1 to make this clearer.

Reviewer comment:
Under the unanswered section, the authors imply that APM is not useful in traumatic tears. I submit that this topic was not researched in this paper and preliminary glance indicates more than 50% patients were satisfied after APM in this group at 1 year and KOOS score seems to have significantly improved from preop status to 52 weeks post op. So the conclusion is not acceptable although the recommendation to conduct further study on that topic may be accepted

Author response:
Thank you for this important comment. We completely agree with the reviewer that this paper did not assess the efficacy of APM for patients with traumatic tears, as this would require a randomised placebo-controlled trial. However, we would also argue that the changes observed in the traumatic tear group might not necessarily result from the APM per se, as implied by the reviewer. History has taught us, that such assumptions were not true for patients with degenerative meniscal tears when tested in well-designed randomised trials. We believe that we have made a very balanced statement based on results in the present study (i.e. no better improvement for patients with traumatic vs. degenerative tears), and previous studies showing that many patients with knee injury may choose not to have surgery if first given a non-operative treatment like exercise therapy (Frobell et al. 2010, NEJM; Kise et al. 2015, BMJ). Lastly, we suggest that it may be time to investigate if this could also be the case for patients with traumatic meniscal tears.

Author action:
No changes made to the manuscript.

Reviewer comment:
Authors seem to have given more weightage to the single item questions in the above recommendation. These questions were asked at 1 year and hence may suffer a retrospective memory bias and also these responses are subjective. There is definite improvement in KOOS scores assessed in these patients compared to the preoperative levels. Hence not recommending surgery to traumatic tears may not be appropriate, in my opinion, with the results presented by the authors.

Author response:
We are somewhat uncertain to which single item the reviewer is referring. However, we do not believe that we recommend against surgery for traumatic tears. Rather we argue that the effect of surgery for young patients with traumatic meniscal tears should be investigated in a comparison with either placebo (sham) surgery or non-operative treatment such as exercise therapy to document if meniscus surgery provides more benefit to these patients than sham surgery or a non-operative strategy. Such studies should also monitor harms and cost.

Author action:
No changes made to the manuscript.

Reviewer comment:
Main difference in results between TT and DT may be related to presence of degenerative meniscal changes. If degenerative menisci are included in both groups, the results would probably be equal/ not significant. Have the authors considered this point and if yes, how did they exclude this effect? This would bring accuracy of visual assessment of meniscus through scopy into question.

Author response:
Thank you for this very relevant question. We strongly agree with the reviewer that degenerative joint changes could potentially influence the outcome after surgery (hence the aim of this study). Also such changes could be present in both groups based on our definition of traumatic and degenerative tears based on age and symptom onset. To address this issue we performed sensitivity

analyses where we based the traumatic/degenerative definition on the meniscal tissue status (i.e. degenerative vs. non-degenerative) but this did not change the interpretation of no difference between groups (sensitivity analysis 6, supplementary table 5). We also tried to combine meniscal tissue quality with patient reported symptom onset for defining traumatic/degenerative tears, which also did not change the interpretation of the results (sensitivity analysis 7, supplementary table 6).

Author action:
No changes made to the manuscript.

Reviewer comment:
How accurate is assessment of degenerative meniscus through arthroscope? Has this been validated with histological appearances? The authors note that majority are degenerative in the DT group. It would be interesting to see how many patients below age 40 and 45 are included in each group. This data is not presented and may be included in the supplementary data for those who are interested.

Author response:
Thank you for this question. As all clinical assessments of pathologies this is not perfect. However, it was assessed by the surgeon at arthroscopy with full vision of the joint as compared to imaging modalities. The inter-rater reliability has been reported to be moderate to good, as described in the methods section (page 8, lines 41-47):
"The inter-rater reliability for meniscal tear type and tissue quality has been reported to be good to moderate (κ coefficients of 0.72 and 0.47, respectively)[31] and good for ICRS cartilage grading (ICC 0.83).[33]"
Also, we thank the reviewer for the suggestion to show the proportion of participants in each age group in the degenerative and traumatic tear groups. We have now added this as supplementary material. We report the distribution of the age groups applied in the study (i.e. 18-34 years and 35-55 years) and the proportion of participants above and below 40 years in each group.

Author action:
The distribution of participants in the different age groups is presented in supplementary table 9.

Reviewer 3:

We identified no specific comments/questions for us to address in the first part of this reviewers comments (thus we have removed these to improve readability), though we very much appreciate the positive feedback and the reviewers great insight into this topic. We have addressed the specific comments from the reviewer below.

Reviewer comment:
Terminology: As we can all appreciate, definition of terms lies at the heart of any study and this is particularly pertinent to the study under scrutiny, as there is no universal consensus on the definition of the terms "trauma(tic)", non-traumatic/degenerative, mechanical symptoms (locking, catching). Having said that, this also highlights how obscure the situation is in clinical orthopedics: Surgeons commonly use these definitions in clinical decision-making despite the fact that they are vague, at best. Nevertheless, from a scientific perspective, this situation introduces an obvious

source of bias, which the authors have attempted to address by having predefined (rationale) definitions for "symptom onset" and "mechanical symptoms" (page 7). As much as I appreciate the approach, I still wonder whether the authors should talk about "acute onset" vs. "gradual, non-traumatic onset", rather than "traumatic" vs. "degenerative". The downside of doing this, I must admit, is that the term "traumatic" is so commonly used (with no universal definition, I must add) that this fact almost justifies its' use here. Perhaps, the authors could still elaborate on this in their discussion? Also, although the authors have provided us with a number of different sensitivity analyses, one potentially clinically-relevant analysis would be to divide patients based on the tear type/morphology, as advocates of APM commonly argue that meniscus tears with longitudinal tear pattern, bucket handle tear or flap are "unstable" (indication for APM), whereas radial, horizontal and complex were determined as stable (less optimal for the APM).

Author response:
We thank the reviewer for bringing up this important issue about terminology. From the above we are not certain what the reviewer actually favors. We have used the following reasoning for the choice of terminology. Basing the definition purely only on symptom onset as suggested by the reviewer is in our opinion not optimal. If patients had either 'acute onset' or 'gradual/slow onset' this could have been feasible, but many patients describe symptom onset like something in between. This is also the reason for the 3 response options, when we asked patients to recall symptom onset (i.e. 'violent incident', 'minor specific incident' or 'slowly evolved over time'). Another reason is that, in our opinion, it is very unlikely for a 'minor specific incident' to cause a tear to a healthy meniscus, whereas this is much more likely to occur to an already degenerated meniscus. Since general joint degeneration is strongly linked to age, we chose to categorize tear as traumatic/degenerative based on a combination of symptom onset and age. That said, we completely agree with the reviewer that this definition is likely not perfect either, which is the reason for the substantial number of sensitivity analyses 'tweaking' the definition but with no difference in results.
We also agree with the reviewer that tear type is usually considered important and different distribution of tear types between groups could potentially also impact the effect of APM. However, conducting additional sensitivity analysis including all patient characteristics reported in table 1 (including tear type), as suggested by the statistical reviewer, did not change the interpretation of the results (sensitivity analysis 3, supplementary table 3).

Author action:
To include the results from the additional sensitivity analysis the results section and discussion were expanded slightly.
Results section now reads (page 12, lines 26-31):
"In sensitivity analysis of the main outcome (i.e. $KOOS_4$) adding degree of cartilage defects as a covariate did not change the interpretation of results, which was similar for analysis including all participant characteristics with a $p<0.10$ and the fully adjusted analysis (supplementary table 2)."
The discussion section on sensitivity analyses have been expanded (page 14, lines 21-36):
"There is no consensus on the definition of 'traumatic' and 'degenerative' tears, and there is a 'grey zone' between the two. Therefore, we conducted several sensitivity analyses testing the robustness of the results by adjusting for meniscal and other structural knee joint pathologies observed at surgery and by applying different definitions of traumatic and degenerative meniscal tears. Even though the level of statistical significance and the direction of the results varied slightly in these analyses, the overall interpretation of no clinically meaningful difference between groups remained (supplementary material)."

Reviewer comment:
Conceptual remark: As noted above, the entire practice of carrying out APM for patients with knee pain basis on nothing but an intuitively rationale: The alleged link between symptoms and "meniscus tears" has been refuted - there is no way to determine whether a "tear" seen at arthroscopy or on MRI is symptomatic (causing pain), the performance of the clinical (meniscus) tests is equivalent to a coin toss… Should the onset of symptoms (see above) be the crux of this paper?

Author response:
Thank you for this suggestion. As described in the previous answer we believe that a combination of symptom onset and age defines the groups in a better way than just symptom onset alone.

Author action:
No changes made to the manuscript.

Reviewer comment:
Observational vs. RCT data: An obvious concern related to this design is the observational nature of the study. Although the authors have provided us with a considerable amount of data suggesting that the analyses (comparison) are (is) valid, maybe a few words on the potential biases and on how they were addressed would be warranted.

Author response:
We agree that the limitations and potential biases of this study should be discussed. This has been done in the discussion section under 'Strength and weaknesses', however given the additional analyses and the reviewer comments this section has been expanded.

Author action:
Part of the 'Strength and weaknesses' section in the discussion has been expanded (page 14-15, lines 53-38):
"We decided to exclude participants 56 years or older not to enrich the data set with many participants likely to have more advanced stages of knee osteoarthritis. Excluding these participants could potentially lead to better average KOOS scores in the degenerative tear group as older age is associated with worse outcome after knee injury.[38] However, sensitivity analysis showed that the $KOOS_4$ time course did not differ between the degenerative tear group, participants aged 56 years or older or the traumatic tear group (group-by-time interaction, p=0.080).
Some participants were lost to follow-up. At 52 weeks, loss to follow-up was 18% and 11% for participants with traumatic and degenerative tears, respectively. Traumatic tear participants lost to follow-up self-reported markedly poorer on 4 of 5 KOOS subscales at the baseline assessment prior to surgery, compared to those who remained in the study. The direction of the resulting bias due to loss to follow-up of these participants is uncertain. However, sensitivity analyses with null responder imputation or assuming best/worst case scenario for patients lost to follow-up did not change the overall interpretation of data.
Participant age and sex distribution in the KACS cohort is similar to what has been reported for patients undergoing meniscal surgery in Denmark.[39] Nevertheless, participants having meniscal repair at surgery were excluded as we intended to compare patient-reported outcomes of two

distinct patient groups receiving the same type of treatment. Thus, the present results only apply to individuals having APM."

Reviewer comment:
What is already known about the subject, second bullet point: "above placebo", what is the study that has shown APM to provide a benefit (even short-term) above placebo?

Author response:
Thank you for this comment. Technically the reviewer is correct that there is no single study showing better effect of APM above placebo. However, here we refer to the results of our previous meta-analysis (Thorlund et al. 2015, BMJ) on APM for the degenerative knee. This study showed a marginal statistically significant effect when pooling studies with different comparators vs APM. Thus, the second bullet point reads:
"High quality evidence shows only marginal short-term benefit of arthroscopic partial meniscectomy above placebo or non-surgical treatment for middle-aged and older individuals with degenerative meniscal tears but we found no trials for patients with traumatic tears."
We believe this is fair given the results of our previous meta-analysis.

Author action:
No changes made to the manuscript.

Reviewer comment:
Discussion, page 13: Prior evidence on "traumatic vs. degenerative": The authors already provide a few studies on this topic, but these two should probably also be included (and their results briefly discussed) given that the evidence on the topic is indeed so sparse (and surprisingly uniform).
a. Kim et al. 1
b. Ghislain et al. 2
1. Kim JR, Kim BG, Kim JW, et al. Traumatic and non-traumatic isolated horizontal meniscal tears of the knee in patients less than 40 years of age. Eur J Orthop Surg Traumatol 2013;23(5):589-93. doi: 10.1007/s00590-012-1028-6
2. Ghislain NA, Wei JN, Li YG. Study of the Clinical Outcome between Traumatic and Degenerative (non-traumatic) Meniscal Tears after Arthroscopic Surgery: A 4-Years Follow-up Study. J Clin Diagn Res 2016;10(4):RC01-4. doi: 10.7860/JCDR/2016/16686.7569

Author response:
Thank you very much for leading our attention to these two papers. They have now been incorporated into the Introduction and Discussion.

Author action:
The references have been incorporated into the introduction, which now reads (page 4-5, lines 39-20):
"In most observational studies, meniscal tear type (i.e. traumatic or degenerative) has rarely been taken into account.[13 14] Early reports from the beginning of the 1980´ies suggest poorer results in individuals with 'degenerative' changes undergoing arthroscopic meniscectomy.[15-17] More recent studies investigating difference in outcome between individuals with traumatic or degenerative tears report conflicting results. One study, including participants less than 40 years of age with isolated

horizontal tears (a rare tear type in this population) reported similar outcome in individuals with traumatic and non-traumatic tears of this type 2 years after surgery.[18] Another study observed similar outcome at 1 year after surgery but poorer outcome in individuals with degenerative tears compared to traumatic tears 4 years after either meniscectomy or meniscal repair.[19] However, both studies were retrospective and included a limited number of participants.[18 19] Lastly, one larger study reported better outcome in individuals with traumatic compared with degenerative meniscal tears 4 years after surgery. However, this was assessed at clinical visits or by telephone interview and not validated patient-reported outcomes measures.[20] Taken together, solid evidence from larger prospective studies using validated outcomes is lacking to confirm the current presumption that individuals with traumatic tears experience larger improvements in patient-reported outcomes after APM than those with degenerative tears."

The studies have also been referenced in the discussion (page 14, lines 6-13):

"Early reports suggested better results in the absence of 'degenerative' changes when undergoing APM.[15-17] However, more recent studies show conflicting results but these are limited by poor study quality or small sample size.[18-20 37]„