PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effect of preoperative duloxetine treatment on postoperative chronic residual pain after total hip or knee arthroplasty: A Randomised Controlled Trial
AUTHORS	Rienstra, Wietske; Blikman, Tim; Dijkstra, Baukje; Stewart, Roy; Zijlstra, Wierd; van Raaij, Tom; ten Hagen, Anita; Bulstra, Sjoerd; Stevens, Martin; van den Akker-Scheek, Inge

VERSION 1 – REVIEW

REVIEWER	Dykukha, Igor Almirall Hermal GmbH, Medical Affairs
REVIEW RETURNED	28-Jun-2021

GENERAL COMMENTS	The study tackles an important real/world clinical issue and is plausibly designed to address the scientific question. It is sufficiently powered and was conducted according to the ICH GCP principles. Comparison of the interventions in the treatment arms made, however, an impression that in fact study had an add-on design, because there were no important mutually exclusive background interventions. Hence, a double-blind design with masking using placebo capsules might have been reasonable and possible, in my opinion. Crusial for the interpretation of the results, might be a choice of the baseline mPDQ score thresholds. There is a real-world evidence (Überall et al. J Pain Res 2019) that chronic pain may respond very differently to pharmacological interventions depending on the phenotype, e.g. mixed pain may have different sensitivity to treatment comparing to neuropathic pain. Such powerful confounder should be probably considered during the stratified randomisation (mPDQ <19 vs. mPDQ >=19). Authors plausibly explained in the manuscript, why they decided to include population with mix pain into the study, but it might be still very interesting to look at the subgroups with baseline mPDQ <19 / >=19 (post-hoc). There are just a few editorial comments and suggestions. [Methods, participants] - Please transfer exclusion criteria, at least the most important of them, from the Supplement to the main manuscript, it is important information. The full list of inclusion/exclusion criteria is available in the Study Protocol. [Methods, Measurement instrument, secondary endpoints]

	- Primary endpoint: “The primary endpoint is the difference in hip- or knee specific postoperative pain…” Is it “mean difference”? Please clarify in the text. - Secondary endpoint – please add time-points of assessment - PainDETECT was developed as a screening tool, not as a measure of the pain intensity (Freynhagen R et al, Curr Med Res Opin 2006), so might be not quite meaningful outcome measure. On the other hand, it was interesting to see, how the structure of pain phenotype changed after arthroplastic [Methods, Measurement instrument, statistical procedure] - Statistical procedure: How missing data was handled? Was there any imputation of the missing data? Please clarify in the manuscript. - “In case of a participants’ discontinuation of the study, all data obtained up to discontinuation was analysed according to the intention to treat principle. All participants with at least one measurement after baseline were included in the study analyses.” That sounds as As Observed analysis. Was that so? Please clarify. As observed alone without sensitivity analysis with imputation, e.g. NRI / BOCF, LOCF or MM, may be misleading. - Please describe how primary endpoint was assessed (timepoint, handling of missing data, which statistical test). [Results, general comment] - Not quite clear which dataset Table 2 or Table 3 corresponds to the inferential test for the primary endpoint. If it is Table 2, the results of the statistical tests should be added - Proposal concerning possible post-hocs: o Subgroup analysis baseline painDETECT <19 / >=19 o Try responder analysis, e.g. improvement of VAS-R / VAS-M pain >=30% - “Mixed model for repeated measures”. Do you mean that the mean and mean difference was adjusted for the strata “heap arthroplasty” and “knee arthroplasty”? (Same comment concerns also “Limitations” part) [Results, Table 1] - Herader: “Duloxetin”, “standard of care”, for better readability. - The groups are well-balanced, except of VAS-R (more sever in Control group), Pain sub-score of KOOS/HOOS (more severe in control group) - Important: patients with baseline painDETECT 12-18 / >=19. Please add % to the demographic data. [Results, Table 2] - Consider reporting graphically the course of the outcomes over the time instead of table [Results, Table 3] - Probably, it should be “mean difference”? (instead of “est. difference”) - Is KOOS/HOOS-p 6 wks after arthroplasty is the predefined primary endpoint? Please clarify in the table and/or text. The primary endpoint has to be reported explicitly. [Results] - Important: Please ad information about adverse events in both groups. Safety parameters is an implicit part of a ICH GCP study

	and is described in the study protocol in the Supplement. It should be reported despite of failed primary endpoint. [Discussion] - Reference to additional analysis not reported in manuscript but available in the Supplement. In my opinion, it would be useful to shortly report main results of the additional analysis in the main manuscript. - Pain phenotype according to painDETECT changed over the time, in the direction of nociceptive type (Table 2), even in the late post-operative period when acute post-operative trauma should not play important role any longer. Any comments about that? [Limitations] - I think, in case of a study of pharmacological treatment mainly aimed at neuropathic pathogenetic pathways, an important and relevant study limitation is combining patients with neuropathic and mixed pain phenotypes and/or lack of stratification for this confounder. - Lack of blinding. Particularly taking into consideration high drop-out rate for safety reasons in duloxetine group, blinding with placebo could address uncertainty about whether or not nocebo effect biased the study results. [Applicability] - I am not sure, whether I understand the using of the term “applicability”. What exactly do you mean? Consider replacing with more common term. - I strongly suggest transferring all safety data to section “Results”. [Abstract] - Please add the time of the primary endpoint measure.

REVIEWER	Plancher, Kevin Plancher Orthopaedics & Sports Medicine
REVIEW RETURNED	13-Jul-2021

GENERAL COMMENTS	Title • Sufficient. Abstract • Page 2, Line 23 – Please clarify the operational definition of ‘sensitization’. • Page 2, Lines 23-25 – Please revise sentence to improve clarity. • Page 2, Line 28 – Please specify ‘signs of sensitization’. Based on the Methods section, it appears to be based solely on a single questionnaire, PainDETECT. What other signs and symptoms were used to classify patients with chronic pain preoperatively? • Page 2, Line 30 – Please specify ‘usual care’. • Page 2, Lines 30-31 – This sentence is awkwardly worded. If this was a prospective randomized clinical trial, then are you suggesting that 17.1% were lost to follow-up? • Page 2, Lines 38-39 – Please clarify the meaning of this sentence. • Page 2, Line 39 – Please outline the time points. • Page 2, Lines 40-41 – Please clarify the timepoint for which this mean improvement is referring to. Is this the preoperative intervention period or at some point postoperatively? • Page 2, Line 42 – Why did you use pain score at 6 months as your primary outcome measure if your aim was to evaluate through 1 year?

	• Page 2, Lines 43 – Please clarify what specific adverse events were captured. • Introduction • Page 4, Lines 68-69 – Is there any newer data on chronic pain after arthroplasty? The data cited here is approximately 10 years old. • Page 4, Line 81 – The authors should define 'chronic residual pain'. • Page 5, Line 99 – Please specify your hypothesis. Methods • Page 6, Lines 102-103 – This is an incomplete sentence, please revise. • Page 6, Line 120 – Awkward sentence; please consider removing. • Page 6, Lines 122-124 – Please explain the modified PainDETECT and cutoffs further. • Page 7, Line 125 and Line 130 – Please spell out the word inclusion. • Page 7, Lines 132 – It would be important for the authors to spell out the inclusion and exclusion criteria for the study within the manuscript itself rather than including it as supplementary materials. • Page 7, Line 134-137 – Please specify the type of randomization. • Page 7, Line 139 – Please clarify what is meant by 'patient characteristics' as well as the demographic and medical history information collected. • Page 7, Line 143 – Please outline what was considered an adverse event. Specifically, how was pain considered or not an adverse event. • Page 7, Lines 146-147 – Please clarify. Why would patients be in the hospital 1 day prior to knee arthroplasty surgery. • Page 7, Lines 147-148 – Please clarify what is meant by 'discontinuation of symptoms'. Presumably you are refereeing to pain? How was this evaluated? • Page 7, Line 149 – Please clarify the 'local protocol' specifically with respect to preoperative, intraoperative, and postoperative pain control measures. • Page 8, Lines 154-155 – Please clarify what 'usual care' consisted of for 10 weeks preoperatively. Additionally, please clarify if these patients were on a waiting list for surgery. • Page 8, Line 161 – Was compliance with the duloxetine regime evaluated? If so, how? Was a daily diary completed? • Page 8, Line 162 – Usual Care must be outlined. • Page 8, Lines 164-166 – Please clarify if usage of breakthrough pain medication was captured. These data would be important to be presented in the Results section. Were there any protocol deviations with patients using gabapentinoids postoperatively? • Page 8, Line 168 – Why was 6 months chosen as the primary endpoint? Additionally, please clarify the 'difference in hip- or knee specific postoperative pain'. Was the baseline pain used or pain evaluated after the 7-week treatment period? • Page 9, Line 174 – Please clarify 'the effect of treatment on general pain relief'. Are the authors using absolute scores here or change scores? • Page 9, Line 176 – Again please clarify 'relief of neuropathic-like pain'. It is also unclear why HOOS/KOOS would be considered the primary outcome measure as opposed to PainDETECT. • Page 9, Lines 180-181 – Please clarify how secondary study endpoint #3 is different than #1 and #2.

	• Page 9, Line 188 – Please spell out the word respectively throughout the manuscript. • Page 9, Line 189 – Please clarify which measure was used for your sample size calculation. Was an improvement by 10 points on the HOOS/KOOS from the preoperative values extracted from the literature? • Page 9, Line 191 – Based on your calculations presented, your sample size would be 56-57 patients per group or a total of 112-114; however, the manuscript states 118. • Page 9, Lines 196-198 – Were independent samples t-test and Mann-Whitney U performed at every time point? What was the p-value used? Was p-value corrected for multiple comparisons? • Page 10, Line 201 – Please clarify how data were handled in your analysis. • Page 10, Lines 202-208 – No supplementary files are available. Please provide more detail on your analysis. Additionally, please clarify the phrase 'development of pain over time'. Is this a subset of patients whose pain increased instead of improved? • Have the authors considered a per protocol analysis versus an intention to treat analysis? How many patients completed the study and adhered o the treatment regime? Results • Page 10, Line 212 – No figures were available. • Page 10, Line 221-222 – Why would patients have to relinquish the option of another TJA within 1 year if they had a complication? Page 11, Line 223 – Please see previous comment regarding sample size calculation. It appears your sample size is too small based on your calculation. • Page 11, Lines 226-227 – Based on the Methods section 1:1 randomization was performed. It is unclear how there could be unequal groups. Please clarify. • Page 11, Lines 228-229 – HOOS/KOOS pain suggests one group had worse pain; however, VAS pain at rest suggests the other group had worse pain. Please comment on this. Additionally, please specify which score is associated with which group. • Page 11, Lines 234-235 – Please clarify if only 92 cases were included in the analysis. How were data handled? Were only 92 cases included in the analysis or all 111 cases? Additionally, of the 92 cases, how many were in the treatment group and how many were in the control group? How many were patients with THA vs. TKA? • Page 13, Line 244 – Please clarify. Wouldn't one expected there to be significant and substantial improvement in pain by 6 months postoperatively given pain is primary indication for surgery? • Page 13, Lines 246-250 – Please provide more detail. Presumably baseline is being referred to as the first assessment rather than pain following 7 week treatment or day before surgery. Please clarify why this datapoint was used. Data from the independent samples t-test do not appear to have been presented. Please clarify. Did the authors consider an analysis of co-variance considering there was a difference at baseline? • Page 13, Line 251-254 – Please provide a descriptive summary of the study findings. Figure 2 was not included. • Page 14, Line 259 – Please clarify what is meant by 'moderate chronic residual pain'. • Page 14, Lines 262-265 – Please present the data by arthroplasty type as well as treatment group. • Page 14, Line 265 – Did the authors consider looking at the percentage of patients that had chronic residual pain as the primary outcome measure (responders vs. non responders)? Also,

	why was HOOS/KOOS pain scores used to look at chronic residual pain versus PainDETECT? • Page 14 – Please present adverse events and complications? Was compliance evaluated? Was breakthrough pain medication utilization captured? Discussion • Page 16, Lines 273-274 – Please clarify that you are referring to chronic residual pain postoperatively. • Page 16, Lines 275-277 – Please clarify. What is meant by the study population being enriched? Perhaps, the authors should explore other analyses such as the severity of neuropathic-like pain symptoms especially considering the categories outlined of 'likely sensitization' and 'possible sensitization'. • Page 16, Lines 282-283 – Please provide reference after 'knee OA patients' and 'with low back pain'. • Page 16, Lines 290-291 – Did you evaluate the effect of treatment preoperatively? What was the change in symptoms from baseline to end of treatment and then end of treatment to day before surgery? Is this the work that was previously published? • Page 17, Lines 293-295 – Please revise sentence to improve clarity. • Page 17, Lines 308-309 – Please provide supplementary files for review. It appears that your analysis would be significantly underpowered to make definitive statements within these groups. • Page 18, Line 338-345 – Please revise to improve clarity. What are the authors trying to say here? References • Please provide full reference for reference 58. • Many of the references (~70%) are over 5-10 years old. Please provide more up-to-date references. Tables & Figures • Table 1 – Please define all abbreviations in the Table footer. • Table 1 – Was hospital length of stay and discharge status captured? • Table 2 – Did pain scores significantly improve from baseline to 7 weeks after targeted treatment? • Table 2 – Please include pain scores obtained 1 day prior to surgery. Was there a change? • Table 2 – Please include scores at the time point of 48 hours after surgery. • Table 2 – Please include appropriate p-values and associated statistical test.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Dr. Igor Dykukha, Almirall Hermal GmbH	
- Please transfer exclusion criteria, at least the most important of them, from the Supplement to the main manuscript, it is important information. The full list of inclusion/exclusion criteria is available in the Study Protocol.	Exclusion criteria transferred, see page 7, section 2
- Primary endpoint: "The primary endpoint is the difference in hip- or knee specific	Yes, the mean difference is meant between the intervention and the control group. See page 9, section 3

postoperative pain…" Is it "mean difference"? Please clarify in the text.	
- Secondary endpoint – please add time-points of assessment	Description of different timepoints added, see page 9 and 10. Also a supplementary figure was added (suppl 1.) for clarification of the different timepoints.
- PainDETECT was developed as a screening tool, not as a measure of the pain intensity (Freynhagen R et al, Curr Med Res Opin 2006), so might be not quite meaningful outcome measure. On the other hand, it was interesting to see, how the structure of pain phenotype changed after arthroplastic	Indeed. It was primarily used as a screening tool in this study, to screen patients for inclusion who showed possible signs of sensitisation. It was not primarily included as a measurement tool for change in this study, as the questionnaire was not designed for that purpose. However, we did include it as a secondary outcome in order to see if the pain phenotype changed at the different timepoints during the study.
- Statistical procedure: How missing data was handled? Was there any imputation of the missing data? Please clarify in the manuscript.	See page 11, section 1 The data was not imputed. We decided to use a Full Information Maximum Likelyhood technique using multilevel mixed model analysis for repeated measures. Multilevel models have the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimates even in the presence of missing data. Missing items in the primary outcome scores, the pain subscales of the KOOS/HOOS questionnaires, were handled according to the KOOS/HOOS manual (www.koos.nu).
- "In case of a participants' discontinuation of the study, all data obtained up to discontinuation was analysed according to the intention to treat principle. All participants with at least one measurement after baseline were included in the study analyses." That sounds as As Observed analysis. Was that so? Please clarify. As observed alone without sensitivity analysis with imputation, e.g. NRI / BOCF, LOCF or MM, may be misleading.	That is correct, this is an As Observed Analysis. This is because we have used a Full Information Maximum Likelyhood technique, in which all available data can be used in the analases, including the incomplete cases (this is a multi-level technique in which the dataset was transferred from wide to long). Multilevel models have the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimates even in the presence of missing data.

- Please describe how primary endpoint was assessed (timepoint, handling of missing data, which statistical test).	A description of the primary endpoint is given on page 9, section 3. Primary endpoint is the mean difference between the intervention and control groups in hip- or knee-specific pain six months postoperatively, assessed with the pain subscale of the *Knee injury and Osteoarthritis Outcome Score* (KOOS) or the *Hip disability and Osteoarthritis Outcome Score* (HOOS). handling of missing data in the primary outcome measure and statistical analysis is described on page 11: For normally distributed data, differences between treatment groups were assessed using an independent samples student T-test. For non-normally distributed data a Mann-Whitney U-test was. In case of discontinued participation in the study, all data obtained up to the participant's discontinuation was analysed according to the intention-to-treat principle. All participants with at least one measurement after baseline were included in the study analyses. The data was not imputed. We decided to use a Full Information Maximum Likelyhood technique using multilevel mixed model analysis for repeated measures. Multilevel models have the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimateseven in the presence of missing data. Missing items in the primary outcome scores, the pain subscales of the KOOS/HOOS questionnaires, were handled according to the KOOS/HOOS manual
- Not quite clear which dataset Table 2 or Table 3 corresponds to the inferential test for the primary endpoint. If it is Table 2, the results of the statistical tests should be added	Indeed, we can imagine that this is confusing. In the original manuscript, both Tables 2 and 3 display data concerning the primary endpoint. However, table 2 only consisted of a display of the mean/median outcomes at the different timepoints, no inferential tests were done. Therefore, it did not provide any additional information. After consideration with our statistician we decided to slightly alter the presentation of the data. The original table 2 was replaced by figure 2 in order to give a visual report of the course of the primary outcome over time.

	Our planned analysis was the inferential test between the change scores of the intervention and the control group at 6 months after surgery compared to baseline. The p-values of the cross-sectional Mann-Whitney U is provided in the Results section (page 15, section 1). The original table 3 (table 2 in the revised manuscript) provides the results of the multilevel mixed model for repeated measures, which uses all available data, and therefore gives a longitudinal data analysis. This table provides p-values for all different timepoints, based on the model.
- Proposal concerning possible post-hocs: o Subgroup analysis baseline painDETECT <19 / >=19 o Try responder analysis, e.g. improvement of VAS-R / VAS-M pain >=30%	We have indeed considered these post-hoc analyses, but due to the (unforeseen) differences in response between knee and hip patients and the (foreseen) relatively small number of patients with baseline painDETECT score >19, both suggested subgroup/post-hoc analyses would consist of very small subgroups, rendering them (in our eyes) not justified. We chose to present the subgroup analyses of the knee and hip patients separately. In order to keep the article readable and focus on the primary outcome, we withdrew from further subgroup/post hoc analyses
- “Mixed model for repeated measures”. Do you mean that the mean and mean difference was adjusted for the strata “heap arthroplasty” and “knee arthroplasty”? (Same comment concerns also “Limitations” part)	Two types of mixed models for repeated measures were conducted, one as a primary analysis (see two comments above). I have transferred the description of this model from the supplementary file to the manuscript for clarification (pages 11 and 12). The second model concerns a sub-analysis adjusting for differences between knee and hip patients. In order to keep the article readable, I have left the information regarding the sub-analysis in the supplementary files. We did not particularly stratify for hip and knee, instead we added a fixed variable for joint to the multilevel mixed model. In this way, the difference explained by whether the hip or knee was the affected joint could be taken in consideration whilst still including all cases in the analysis.
- Herader: “Duloxetin”, “standard of care”, for better readability.	See table 1.
- The groups are well-balanced, except of VAS-R (more sever in Control group), Pain	Actually, the VAS pain and the KOOS/HOOS pain subscales have reversed scoring systems.

sub-score of KOOS/HOOS (more severe in control group)	For the KOOS/HOOS scales, 0 represents extreme symptoms and 100 represents no symptoms. For the VAS 0 represents no pain and 100 represents worst pain imaginable (as explained on pages 9 and 10. We have dealt with the issue of unbalanced groups at baseline as described on page 15.
- Consider reporting graphically the course of the outcomes over the time instead of table	Thank you for this suggestion. Indeed, after consideration we decided to replace table 2 with figure 2 in order to give a visual report of the course of the primary outcome over time. For readability we decided to restrict this graph to the primary outcome measure. Information regarding the secondary outcome measures on the other timepoints are presented in table 2 (this was table 3 in the original article),
- Probably, it should be “mean difference”? (instead of “est. difference”)	No: because Table 2 concerns the estimated differences based on the multilevel mixed model for repeated measures, the correct header is estimated difference. See also the description of the multilevel mixed model for repeated measures in the comments above. For clarification we removed the term est. from the column in the table, but it is still in the correct header for the table.
- Is KOOS/HOOS-p 6 wks after arthroplasty is the predefined primary endpoint? Please clarify in the table and/or text. The primary endpoint has to be reported explicitly.	The primary endpoint is reported explicitly on page 9 (see earlier comments from reviewer). For further clarification I have made the primary endpoint bold in Table 2.
- Important: Please ad information about adverse events in both groups. Safety parameters is an implicit part of a ICH GCP study and is described in the study protocol in the Supplement. It should be reported despite of failed primary endpoint.	I transferred the information about adverse events and loss to follow-up from the supplement to the article itself. See page 17.
- Reference to additional analysis not reported in manuscript but available in the Supplement. In my opinion, it would be useful to shortly report main results of the additional analysis in the main manuscript.	Indeed, information on the sub-analysis was added to the Methods section (see page 12) and the Results section (see page 16)
- Pain phenotype according to painDETECT changed over the time, in the direction of nociceptive type (Table 2), even in the late post-operative period when acute post-operative trauma should not play important role any longer. Any comments about that?	Great care needs to be taken in interpreting these scores, considering the lack of evidence surrounding the evaluative rather than discriminative measurement properties of the painDETECT Questionnaire (it was developed as a screening tool). Furthermore, the painDETECT Questionnaire is tightly correlated to the overall pain score (Rienstra et al. Plos One 2015 and Rienstra et al. Disability and Rehabilitation 2019), so the fact that the

	painDETECT scores are so low at six months and twelve months post-arthroplasty are most likely a reflection of the low overall pain scores during those timepoints (see the VAS scores at those timepoints).
- I think, in case of a study of pharmacological treatment mainly aimed at neuropathic pathogenetic pathways, an important and relevant study limitation is combining patients with neuropathic and mixed pain phenotypes and/or lack of stratification for this confounder.	This study was not aimed at neuropathic pathogenetic pathways, but used neuropathic-like symptoms as described in the painDETECT Questionnaire as an indication that sensitisation may play a role in a certain subgroup of OA patients (nociplastic mechanisms), in order to identify these patients for inclusion. It is important to note that there is no gold standard to measure sensitisation. This is not a problem specific to OA patients, but to all diseases and syndromes in which sensitisation is believed to play a role. One way to measure is by measuring neuropathic-like pain symptoms. When considering the identification of a dominant role of sensitisation in a knee or hip OA patient, it is important for clinicians to realise that sensitisation is not a single entity that is either present or absent, but rather that it forms a continuum, as stated by Lluch et al. 2018. Pain in OA patients is likely not purely due to sensitisation, and rarely neuropathic: as nociceptive processes due to the OA process likely play a role, it is probably a more mixed-pain profile considering all the different mechanisms that are involved in OA (including sensitisation). Only patients with a painDETECT score >12 were considered for inclusion, and it can be seen in Table 1 that only a small proportion had a painDETECT score >19. We were primarily interested in patients in whom sensitisation likely plays a substantial role. Therefore, I do not consider it a confounder that patients with painDETECT scores >12 were combined with a small number of patients with scores >19, as this is unlikely to be an expression of neuropathic pathways – it would rather point to severe sensitisation in combination with severe overall pain scores due to advanced OA. The application of the painDETECT questionnaire as an indication for sensitisation was addressed in the introduction and discussion section (pages 7 and 18).
- Lack of blinding. Particularly taking into consideration high drop-out rate for safety	This is a valid point. Considering the set-up and results of this trial, there is a possible

reasons in duloxetine group, blinding with placebo could address uncertainty about whether or not nocebo effect biased the study results.	nocebo effect in the duloxetine group, especially during and shortly after the intervention period. However, due to the extensive time period between the actual study intervention and the surgery that took place in-between, this effect is not very likely to have influenced the primary endpoint of this study at six months post-arthroplasty. I have added this point to the limitations section of the Discussion (page 21, last section)…
- I am not sure, whether I understand the using of the term "applicability". What exactly do you mean? Consider replacing with more common term.	The term applicability was changed to suitability and safety (end of page 21, start of page 22). Maybe these are more common terms? Considering this is a pragmatic trial I think it is important to elaborate in the Discussion what the consequences are of the frequent adverse effects of duloxetine, in combination with its numerous medical contraindications (see for more detail the description in our design paper), for clinical practice suitability/applicability.
- I strongly suggest transferring all safety data to section "Results".	Information about adverse events was added to the article, see page 17
- Please add the time of the primary endpoint measure.	Added, see page 2
Reviewer: 2 Dr. Kevin Plancher,	
• Page 2, Line 23 – Please clarify the operational definition of 'sensitization'.	Added, see page 2
• Page 2, Lines 23-25 – Please revise sentence to improve clarity.	I tried to make it clearer, and it was revised by our professional editor.
• Page 2, Line 28 – Please specify 'signs of sensitization'. Based on the Methods section, it appears to be based solely on a single questionnaire, PainDETECT. What other signs and symptoms were used to classify patients with chronic pain preoperatively?	Indeed, as explained in the Methods section, the screening on symptoms of sensitisation took place using the painDETECT Questionnaire. This was added to the line. See page 2
• Page 2, Line 30 – Please specify 'usual care'.	Clarified, see page 2
• Page 2, Lines 30-31 – This sentence is awkwardly worded. If this was a prospective randomized clinical trial, then are you suggesting that 17.1% were lost to follow-up?	This sentence is indeed awkward, I deleted it from the abstract. Information on missing data/loss to follow-up is described extensively in the Results section and figure 1 the Consort flow-chart. See page 9.
• Page 2, Lines 38-39 – Please clarify the meaning of this sentence.	I mean that the primary outcome measures were presented as originally described in the design paper; no changes were made based on lack of results, so no publication bias in this sense. This is important information for the CONSORT checklist. I have removed this line from the

	abstract as it is also included in the (enclosed) CONSORT checklist.
• Page 2, Line 39 – Please outline the time points.	Added, see page 2.
• Page 2, Lines 40-41 – Please clarify the timepoint for which this mean improvement is referring to. Is this the preoperative intervention period or at some point postoperatively?	This is the primary endpoint, so six months after surgery. I added this for clarification, see page 3.
• Page 2, Line 42 – Why did you use pain score at 6 months as your primary outcome measure if your aim was to evaluate through 1 year?	This primary outcome measure was chosen at 6 months, as in practice this was considered the first possible timepoint to evaluate chronic residual pain after arthroplasty. Because it is known from practice that the amount of chronic residual pain is not likely to change after one year postop, we aimed to follow up to one year postop in order to be as thorough as possible. For clarification I changed the aim in the line mentioned by the reviewer (page 2).
• Page 2, Lines 43 – Please clarify what specific adverse events were captured.	This is described extensively in the article itself, and considering it is quite a list of possible adverse events, we chose not to include this information in the abstract.
• Page 4, Lines 68-69 – Is there any newer data on chronic pain after arthroplasty? The data cited here is approximately 10 years old.	The reference list was updated
• Page 4, Line 81 – The authors should define 'chronic residual pain'.	Added, see page 4.
• Page 5, Line 99 – Please specify your hypothesis.	I specified the description of the aim of the study in order to formulate our hypothesis more explicitly, see page 5.
• Page 6, Lines 102-103 – This is an incomplete sentence, please revise.	Completed
• Page 6, Line 120 – Awkward sentence; please consider removing.	Thank you for this remark. I agree that it is an awkward sentence, so it was removed…
Page 6, Lines 122-124 – Please explain the modified PainDETECT and cutoffs further.	I transferred this information from the supplementary file to the article for clarification, see pages 6 and 7.
• Page 7, Line 125 and Line 130 – Please spell out the word inclusion.	Done, thank you for this remark
• Page 7, Lines 132 – It would be important for the authors to spell out the inclusion and exclusion criteria for the study within the manuscript itself rather than including it as supplementary materials.	As also suggested by reviewer 1, I added the exclusion criteria to the article, see page 7.

• Page 7, Line 134-137 – Please specify the type of randomization.	Added, see end of page 7. Randomisation was stratified for type of arthroplasty with block randomisation, block sizes of 4 and 6.
• Page 7, Line 139 – Please clarify what is meant by 'patient characteristics' as well as the demographic and medical history information collected.	As Table 1 gives a clear overview of the characteristics that were assessed in the patients, I decided to refer to this table instead of clarifying this in the text (start of page 8). I consider patient characteristics and demographic information as broadly accepted terms in scientific literature, regularly described in the first tables of an article in order to provide information on generalisability of the study population.
• Page 7, Line 143 – Please outline what was considered an adverse event. Specifically, how was pain considered or not an adverse event.	I can see now that this was not clear, thank you. Side effects and discontinuation effects of duloxetine were meant by this. I changed this section for more clarification. See page 8. Pain is not a part of these side effects or discontinuation effects.
• Page 7, Lines 146-147 – Please clarify. Why would patients be in the hospital 1 day prior to knee arthroplasty surgery.	I understand your question but we did not include an explanation for this in the paper as this is not really relevant in the context of the study. In the Netherlands it is common practice for arthroplasties to be performed at the beginning of the operative programs for orthopaedic surgeons, given the lowest possible risks of perioperative infections. This means the arthroplasties take place early in the morning, so patients are admitted to the hospital the day before surgery for logistical reasons to get them prepared in the preop hours. We visited them in the hospital to physically pick up the questionnaires as a service to them, and so they did not have to bother with mailing the questionnaires on the day of hospital admission.
• Page 7, Lines 147-148 – Please clarify what is meant by 'discontinuation of symptoms'. Presumably you are refereeing to pain? How was this evaluated?	What is meant is discontinuation symptoms of duloxetine. The specific possible discontinuation symptoms and side effects of duloxetine are described in the design paper and in the paper of T. Blikman et al., which focuses on the results of the preoperative study period of this trial. Because this topic is described extensively in these prior publications, we chose not to elaborate any further in the present paper, in order to focus on primary endpoints and keep the article focused.
• Page 7, Line 149 – Please clarify the 'local protocol' specifically with respect to preoperative, intraoperative, and postoperative pain control measures.	Our study is designed as a pragmatic trial, so no restrictions were imposed on usage of escape (pain) medication or other medication. However, usage of agents to specifically address neuropathic pain symptoms,

	like gabapentinoids, should be avoided since this could potentially interfere with study outcomes. To this end, local care-as-usual will be slightly modified for study patients at Martini Hospital and MC Leeuwarden, since these two hospitals use gabapentinoids in the perioperative and early postoperative period in a subset of patients. Upon request I could provide you with the complete protocols from the different hospitals regarding perioperative pain control measures. But this varies per individual patient and is conducted as habitual custom/standard medical care in Dutch Hospitals. No adaptations were made for the study except for the avoidance of gabapentinoids/neuropathic pain medication perioperatively. I think it is beyond the scope of this article to elaborate further on this topic in order to keep the article readable and within word limits as requested by the journal.
• Page 8, Lines 154-155 – Please clarify what 'usual care' consisted of for 10 weeks preoperatively. Additionally, please clarify if these patients were on a waiting list for surgery. • Page 8, Line 162 – Usual Care must be outlined.	As described in the article, all patients were on the waiting list for surgery in order to be eligible for the study. I added additional information on usual care. See page 9.
• Page 8, Line 161 – Was compliance with the duloxetine regime evaluated? If so, how? Was a daily diary completed?	Yes it was, at timepoints 1, 2 and 3 daily diaries were assessed for compliance and any side effects or discontinuation symptoms. This topic is described in the design paper and in the paper of T. Blikman et al., which focuses on the results of the preop study period of this trial. A description of protocol violation can be found in Figure 1 (Consort flow chart) and page 17.
• Page 8, Line 168 – Why was 6 months chosen as the primary endpoint? Additionally, please clarify the 'difference in hip- or knee specific postoperative pain'. Was the baseline pain used or pain evaluated after the 7-week treatment period?	This primary outcome measure was chosen at 6 months, as in practice this was considered as the first possible timepoint to evaluate chronic residual pain after arthroplasty. Because it is known from practice that the amount of chronic residual pain is not likely to change after one year postoperatively, we aimed to follow up to one year after surgery in order to be as thorough as possible. See page 9, last section for further clarification. All timepoints in the study were always compared to the baseline pain, We had to use the change scores (difference between baseline and 6 months postop) instead

	of the absolute scores at 6 months postop because there was a significant difference between the intervention and control groups in baseline KOOS/HOOS pain score.
• Page 9, Line 174 – Please clarify 'the effect of treatment on general pain relief'. Are the authors using absolute scores here or change scores?	Originally, we intended to perform inferential tests on the difference in absolute scores for intervention and control group at 6 months postop, but just as for the KOOS/HOOS pain subscale the VAS pain scales were also significantly different at baseline. Therefore, the inferential tests were performed on the mean difference in change scores at 6 months postop compared to baseline, rather than absolute scores (as explained in page 15, first section) For clarification I changed relief to change.
• Page 9, Line 176 – Again please clarify 'relief of neuropathic-like pain'. It is also unclear why HOOS/KOOS would be considered the primary outcome measure as opposed to PainDETECT.	I changed relief to amount because we looked at absolute scores here instead of change scores. Note that this is a secondary outcome measure. The primary endpoint was chronic residual pain and that is not a construct that you can measure with the painDETECT − instead the KOOS/HOOS questionnaires are suitable for this construct. The painDETECT was primarily used as a screening tool in this study, to screen patients for inclusion who showed possible signs of sensitisation. It was not primarily included as a measurement tool for change in this study, as the questionnaire was not designed for that purpose. However, we did include it as a secondary outcome in order to see if the pain phenotype changed at the different timepoints during the study (see table 2).
• Page 9, Lines 180-181 – Please clarify how secondary study endpoint #3 is different than #1 and #2.	See addition on page 10. 1. the effect of treatment on change in general pain six months postoperatively, measured using a Visual Analogue Scale (VAS) (100-mm horizontal line representing pain from 0 (no pain) to 100 (worst pain imaginable)); 2. amount of neuropathic-like pain measured using the modified PainDETECT-Questionnaire (m-PDQ) six months postoperatively; 3. course of the above-mentioned pain scores at different timepoints. A detailed description of all timepoints and the measurements performed during those timepoints is provided in the design paper.[41] Timepoints 1, 2 and 3 cover the

	preoperative intervention phase, timepoints 4, 5, 6 and 7 cover the postoperative period, ranging from 48 hours (the primary outcome measure was not assessed at this timepoint) to six weeks, six months and twelve months, respectively. See also figure 1 of the supplementary files for a visual overview of the study timepoints. As timepoints 1, 3 and 4 were appointed primarily for the evaluation of side effects, discontinuation effects, or peri-operative complications and not for the evaluation of the primary outcome measures, these timepoints were omitted from analyses in the present paper.
• Page 9, Line 188 – Please spell out the word respectively throughout the manuscript.	Done, thank you for pointing this out.
• Page 9, Line 189 – Please clarify which measure was used for your sample size calculation. Was an improvement by 10 points on the HOOS/KOOS from the preoperative values extracted from the literature?	G-power was used for the power analysis. The power analysis was based on an independent samples T-test, 0.05 two-sided significance, with a power of 80%, with a minimal mean difference of 10 points on the KOOS/HOOS pain subscale (based on literature) between the intervention and control group at 6 months after surgery with a standard deviation between 17.2-17.7 (based on literature) A decription of this calculation is described in the Sample Size Calculation section, see page 10. Also, the reference to literature on which the standard deviations and minimally important improvement values were based are given directly after that sentence. I attempted to clarify this section further. For further information concerning G-power see Reference: Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.
• Page 9, Line 191 – Based on your calculations presented, your sample size would be 56-57 patients per group or a total of 112-114; however, the manuscript states 118.	As also described by the comments of the editorial office and reviewer 1. See page 10 Based on the sample size calculation, a total sample size of 47 participants per group was needed, i.e. a total sample

	size of 94 participants. However, we anticipated a discontinuation rate of 20% (this was an educated guess) leading to an aimed sample size of 59 subjects per group (118 participants in total). In the end, 57 patients were randomised into the intervention group and 54 into the control group, coming to a total sample size of 111 participants, which surpassed the minimum sample size of 47 participants per group (94 participants in total). All randomised patients were included in the multilevel lmixed model data analyses for repeated measures, based on our pragmatic intention-to-treat protocol (see Figure 1. Consort flow-chart). Therefore/span>, the study was not underpowered. We changed the wording in the methods section to make this clearer.
• Page 9, Lines 196-198 – Were independent samples t-test and Mann-Whitney U performed at every time point? What was the p-value used? Was p-value corrected for multiple comparisons?	No, we did not perform cross-sectional independent samples T-tests and Mann-Whitney U tests at every timepoint. Before data collection had started, we had decided that our primary outcome measure would be at 6 months after surgery, as described in the design paper. The independent Mann Whitney U test for this timepoint is provided on page 15, section 1. Thus, there was a planned analysis. This is in contrast to an unplanned analysis or data fishing. Due to the planned analysis, no multiple statistical comparisons were made. The Multilevel model that was used for the longitudinal data analyses has the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, no correction for multiple comparisons is necessary in these models.
• Page 10, Line 201 – Please clarify how data were handled in your analysis.	As far as we understand the reviewer's question, we can answer it as follows: The section on statistical analysis has been extended based on the comments from reviewer 1. Also, a section was added considering loss to follow-up and missing data to the article. A description of the primary endpoint is given on page 9, handling of missing data in the primary outcome measure and statistical analysis is described on pages 10 and 11. The data was not imputed. We decided to use a Full Information Maximum Likelyhood technique using multilevel mixed model analysis for

	repeated measures. Multilevel models have the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimates even in the presence of missing data. Missing items in the primary outcome scores, the pain subscales of the KOOS/HOOS questionnaires, were handled according to the KOOS/HOOS manual (www.koos.nu). Multilevel models have the ability to handle models by using all available data, which is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimates even in the presence of missing data.
• Page 10, Lines 202-208 – No supplementary files are available. Please provide more detail on your analysis. Additionally, please clarify the phrase 'development of pain over time'. Is this a subset of patients whose pain increased instead of improved?	This was added as mentioned in the previous comment. I don't know why the supplementary files were not available to reviewer 2. By 'a change in the development of pain over time' I did not mean that in some patients the pain increased. I changed 'development' to the more neutral 'modification'.
• Have the authors considered a per protocol analysis versus an intention to treat analysis? How many patients completed the study and adhered o the treatment regime?	This would have been problematic to perform as most patients who discontinued the intervention treatment did use the intervention treatment for varying time periods. So we would have had to perform analyses between three groups: no intervention treatment, partial intervention treatment and full intervention treatment. An intention-to-treat analysis was considered more appropriate in a pragmatic study than a per-protocol analysis. Information on protocol deviations were added on page 17.
• Page 10, Line 212 – No figures were available.	I don't know why the figures were not available to reviewer 2.
• Page 10, Line 221-222 – Why would patients have to relinquish the option of another TJA within 1 year if they had a complication?	They wouldn't. As described in the exclusion criteria, if patients had OA in multiple joints, and the intention was to receive an elective arthroplasty in more than 1 joint within the year, we anticipated that this would interfere with the 1-year follow-up of these

	patients in the study. This does not concern necessary TJA due to complications.
Page 11, Line 223 – Please see previous comment regarding sample size calculation. It appears your sample size is too small based on your calculation.	See previous comments.
• Page 11, Lines 226-227 – Based on the Methods section 1:1 randomization was performed. It is unclear how there could be unequal groups. Please clarify.	This is due to the block randomisation for hip and knee patients. I added information regarding this issue to the section on randomisation (end of page 7).
• Page 11, Lines 228-229 – HOOS/KOOS pain suggests one group had worse pain; however, VAS pain at rest suggests the other group had worse pain. Please comment on this. Additionally, please specify which score is associated with which group.	Actually, the VAS pain and the KOOS/HOOS pain subscales have reversed scoring systems. For the KOOS/HOOS scales, 0 represents extreme symptoms and 100 represents no symptoms. For the VAS, 0 represents no pain and 100 represents worst pain imaginable (see end of page 9 and start of page 10). I added the groups to the scores (based on Table 1).
• Page 11, Lines 234-235 – Please clarify if only 92 cases were included in the analysis. How were data handled? Were only 92 cases included in the analysis or all 111 cases? Additionally, of the 92 cases, how many were in the treatment group and how many were in the control group? How many were patients with THA vs. TKA?	See previous comments and pages 11 and 12 of the manuscript Complete follow-up of all timepoints 1-7 was retrieved in 92 cases (82.9%). 17.1% of participants had one or more timepoints in which the pain subscale of the KOOS/HOOS was missing. These can be considered incomplete, or interrupted cases. Those cases were not all lost to follow-up, see figure 1 Consort Flow Chart and page 17. For complete information regarding the moment of interruption, whether it was a knee or hip patient and whether control or intervention group. For 92 cases, the data were available for all timepoints. This doesn't mean that only those (complete) cases were included in the analysis. All patients were included in the longitudinal data analysis, so 111 in total. Because we used a Full Information Maximum Likelyhood technique using multilevel mixed model analysis for repeated measures, we had the ability to use all available data. this is an advance over traditional repeated-measures analysis, where the usual treatment is to remove the entire patient if one of the outcomes is missing. With the multilevel model, we use as estimated strategy Full information maximum likelihood, where we get parameter estimates even in the presence of missing data.

• Page 13, Line 244 – Please clarify. Wouldn't one expected there to be significant and substantial improvement in pain by 6 months postoperatively given pain is primary indication for surgery?	Yes, one would expect that. It was expected that the postoperative pain scores would be skewed and that most patients would only experience a moderate amount of postop pain.
• Page 13, Lines 246-250 – Please provide more detail. Presumably baseline is being referred to as the first assessment rather than pain following 7 week treatment or day before surgery. Please clarify why this datapoint was used. Data from the independent samples t-test do not appear to have been presented. Please clarify. Did the authors consider an analysis of co-variance considering there was a difference at baseline?	Indeed, baseline is baseline and not after 7 weeks of duloxetine treatment, but the results of 7 weeks of duloxetine treatment are included in the longitudinal analysis (table 2). I added a section regarding the build-up of the mixed-model analysis to make this clearer (pages 11 and 12). A mixed model for repeated measures was constructed including, time, treatment allocation, and baseline KOOS/HOOS pain scale (in order to correct for the differences between groups at baseline). A variable was added differentiating between preoperative and postoperative timepoints (coded 0 or 1 for preop and postop timepoints, respectively), thereby creating a piece-wise analysis. This way the postoperative effect of duloxetine treatment could be distinguished while including the data from all timepoints. Apart from baseline KOOS/HOOS pain subscale, interaction terms between this piece-wise variable and all other separate variables were added, as well as a three-way interaction term between time, treatment and the 'piece-wise' variable. A random intercept was added for individual subjects. Analysis of co-variance is not necessary in a multilevel mixed model for repeated measures.
• Page 13, Line 251-254 – Please provide a descriptive summary of the study findings. Figure 2 was not included.	I don't know why Figure 2 and Table 3 were not available to reviewer 2, but the findings are presented there.
• Page 14, Line 259 – Please clarify what is meant by 'moderate chronic residual pain'.	Moderate pain was defined as a KOOS/HOOS pain subscale score <70. See end of page 16.
• Page 14, Lines 262-265 – Please present the data by arthroplasty type as well as treatment group.	This can be found in the sub-analysis, I transferred the main outcomes of the sub-analysis to the article based on comments from reviewer 1. Full description can be found in the supplementary files, in order to keep the article readable and within word limits.
• Page 14, Line 265 – Did the authors consider looking at the percentage of patients that had chronic residual pain as the primary outcome measure (responders vs. non responders)? Also, why was HOOS/KOOS pain scores used to look at chronic residual pain versus PainDETECT?	We have indeed considered this, but due to the (unforeseen) differences in response between knee and hip patients this suggested subgroup/post-hoc analysis would consist of very small groups, rendering them (in our eyes) not justified. Also, there is a (foreseen) substantial effect of the arthroplasty on pain in

	the entire study population, rendering it very difficult to find a additional difference in pain relief between intervention groups in such small subgroups of patients, this reached beyond the scope of the present article. The primary endpoint was chronic residual pain and that is not a construct that you can measure with the painDETECT − instead, the KOOS/HOOS questionnaires are suitable for measuring this construct.
• Page 14 – Please present adverse events and complications? Was compliance evaluated? Was breakthrough pain medication utilization captured?	A description of protocol violation can be found in Figure 1 and page 17 (added section). Moreover, this topic is described in the design paper and in the paper of T. Blikman et al., which focuses on the results of the preoperative study period of this trial.
• Page 16, Lines 273-274 – Please clarify that you are referring to chronic residual pain postoperatively.	Clarified, see page 18, first section.
• Page 16, Lines 275-277 – Please clarify. What is meant by the study population being enriched? Perhaps, the authors should explore other analyses such as the severity of neuropathic-like pain symptoms especially considering the categories outlined of 'likely sensitization' and 'possible sensitization'.	Aim was to identify the sensitized subpopulation of OA patients for inclusion in the study by screening with the painDETECT Questionnaire. For clarification I removed the term 'enriched' and changed this sentence (page 18, section 2). Here it is explained why we chose the 'possible sensitisation' cut-off point instead of the 'likely sensitisation' cut-off point. Because of the (foreseen) relatively small number of patients with baseline painDETECT score >19, both suggested subgroup/post-hoc analyses would consist of very small and unbalanced groups, rendering them (in our eyes) not jusitfied.
• Page 16, Lines 282-283 – Please provide reference after 'knee OA patients' and 'with low back pain'.	Done, the reference was given in the earlier sentence, it was transferred.
• Page 16, Lines 290-291 – Did you evaluate the effect of treatment preoperatively? What was the change in symptoms from baseline to end of treatment and then end of treatment to day before surgery? Is this the work that was previously published?	Yes, the preoperative effects were described in a previous manuscript. This paper is currently accepted, pending revision by BMC. The information is available in the PhD thesis of one of the co-authors (T. Blikman), access to this thesis is provided in the references. This is the reason why I do not elaborate extensively on the preoperative outcomes in the current paper, also, in order to keep the focus of the present article on the primary outcome measure and time point. The p-values of the different timepoints based on the multilevel mixed model for repeated measures are presented in Table 2, including the perioperative effect. As explained in the relevant section in the Methods, the preoperative effects can be distinguished

	from the postoperative effects in the mixed-model longitudinal analysis (Table 2)
• Page 17, Lines 293-295 – Please revise sentence to improve clarity.	Clarified. See page 19: Second, if we weren't successful in adequately desensitising patients prior to surgery this could explain the lack of effect on chronic residual pain after TJA. A statistically significant treatment effect of 8.3 points (CI 1.3-15.3) was found in the preoperative treatment phase, yet this difference does not seem clinically relevant compared to reported minimally important changes (MIC) of 10 points in literature.[50,56,57] It should be noted that these MIC values are mostly reported after operative treatments and therefore cannot automatically be extrapolated to relevant changes following conservative treatment.
• Page 17, Lines 308-309 – Please provide supplementary files for review. It appears that your analysis would be significantly underpowered to make definitive statements within these groups.	As our original power analysis was never designed to divide these groups for separate analyses, the term 'underpowered' does not seem appropriate. Indeed, these sub analyses consist of small subgroups, which is why we do not make definitive statements within these groups. The results give incentive for future studies. It is not clear to me why reviewer 2 did not have access to the supplementary files.
• Page 18, Line 338-345 – Please revise to improve clarity. What are the authors trying to say here?	Several authors and our professional native-English translator and editor looked at this section and we cannot determine what changes are necessary to make it any clearer. Could the reviewer please be more specific about what is not clear?
• Please provide full reference for reference 58.	The reference list was updated
• Many of the references (~70%) are over 5-10 years old. Please provide more up-to-date references.	The reference list was updated
• Table 1 – Please define all abbreviations in the Table footer.	Thank you for pointing this out, abbreviations added.
• Table 1 – Was hospital length of stay and discharge status captured?	No, this was not captured. In general, hospital stay is around 3 days and patients go to their homes afterwards and can manage with some help of a partner/spouse.
• Table 2 – Did pain scores significantly improve from baseline to 7 weeks after targeted treatment?	Table 2 in the original manuscript was replaced by a figure as described in earlier comments. P-values regarding the different timepoints can be found in the new table 2 (originally table 3) based on the multilevel mixed model for repeated measures.
• Table 2 – Please include pain scores obtained 1 day prior to surgery. Was there a change?	No longer applicable as table 2 was removed. See page 10, point 3.

• Table 2 – Please include scores at the time point of 48 hours after surgery.	At 48 hours after surgery, the primary endpoint, the pain subscale of the KOOS and HOOS were not assessed, as this subscale is not valid so shortly after surgery. This is described in the design paper, for clarification I also added this information to the Methods section (page 10, point 3) Therefore this timepoint was not included in the table.
• Table 2 – Please include appropriate p-values and associated statistical test.	Table 2 in the original manuscript was replaced by a figure as described in earlier comments. P-values regarding the different timepoints can be found in the new table 2 (originally table 3) based on the multilevel mixed model for repeated measures.

VERSION 2 – REVIEW

REVIEWER	Dykukha, Igor Almirall Hermal GmbH, Medical Affairs
REVIEW RETURNED	14-Sep-2021

GENERAL COMMENTS	Thank you for comprehensively and transparenly adressing all my questions and comments.