PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Finnish Subacromial Impingement Arthroscopy Controlled Trial (FIMPACT): A protocol for a randomized trial comparing arthroscopic subacromial decompression and diagnostic arthroscopy (placebo control), with an exercise therapy control, in the treatment of shoulder impingement syndrome
AUTHORS	Paavola, Mika; Malmivaara, Antti; Taimela, Simo; Kanto, Kari; Jarvinen, Teppo

VERSION 1 - REVIEW

REVIEWER	Brox, Jens Ivar Dep Phys Med & Rehabil, Oslo University Hospital; University of Oslo, Norway.
REVIEW RETURNED	23-Sep-2016

GENERAL COMMENTS	Shoulder pain is next to low back pain and neck pain the most common musculoskeletal condition. Subacromial pain syndrome, rotator cuff disease or tendinosis often referred to as the impingement syndrome is the major diagnosis. Paradoxically, surgical treatment has exploded after a trial in BMJ in 1993 found that it was not more effective than exercises. Later studies have replicated this finding, but current overuse of surgery has not decreased. The present protocol describes a sham controlled trial evaluating the effectiveness of surgery. The secondary aim is to compare surgical results to physiotherapy guided exercises. This trial is highly warranted and the protocol describes a gold standard design even using blinded interpretation of results and writing of the manuscript. Patients are already included according to the estimated sample size and follow-ups are continued until the last 2-year follow-up is completed. The protocol is very interesting and of excellent quality in my opinion. I have only some minor comments: Abstract: A visual analogue scale (0-100) means that a scale is used as an analogue for pain intensity. To express this in mm is not the point, in my opinion it is better to delete mm. Introduction, page 5: subacromial pain syndrome was used by K Engebretsen et al in BMJ in 2009, long before the referred guideline was published. The introduction suggests that surgery may be either therapeutical, placebo (or both) while neither the role or evidence of exercises is clearly reported. Also the similar ongoing UK trial lack the exercise arm. Methods, page 8 and 9, table 1, sick leave is not listed. Page 11, postoperative care is briefly described, how many physiotherapy

	sessions are expected as compared to the exercise group? Page 12, table 2, sick-leave is not listed, and how is it estimated? According to a previous study (Grovle L, Haugen A et al.) patients reliably report their current sickness benefit status but not the number of days on sick-leave. The trial has three related primary outcomes: arm activity pain, rest pain, and acceptable symptom intensity at 24 months. This is acceptable in my opinion, while others had preferred a self-reported pain disability index. Y Roe et al. (2016) published a comprehensive review of the most frequent outcomes using ICF taxonomy. My most important objection is the choice of 15 points as the minimally clinically important difference (MCID). MCID is not an exact estimate because it is calculated by using the most subjective post-treatment global evaluation as a gold standard. For neck and back pain MCID of VAS-pain is always higher than for an index made from a set of questions. Surprisingly MCID for shoulder pain is estimated to 15, lower than for relevant shoulder scores (see Ekeberg et al). Another way of evaluating the minimally clinically important difference is to ask the patient on beforehand, in this case: would you prefer surgery with a benefit of 15 point reduction of pain on a 0-100 scale? Also the choice of 30 as an acceptable state may represent a problem. How many of the included patients have a preoperative score of 30 to 45? Others have used patients reporting excellent or good as an acceptable state (by example Brox et al 1999) which is probably not much different from < 30, although possibly a stronger criterium. The study includes three pre-specified subgroup hypotheses (page 19/20). Analyses this hypotheses are just exploring and results are rarely important (see Pocock, NEJM 2016). Nevertheless, when subgroups are included I miss a fourth hypothesis: patients on sick leave versus patients not on sick leave (Koljonen et al 2009, Brox et al 1996 and 1999). Under limitations (page 25) it is reported that 14 patients were excluded because of A-C arthrosis or intra-articular findings. It is normally wrong to exclude patients after randomisation because it reduces the effect of randomisation, Fortunately exclusion includes only a small percentage of patients. Surgeons often believe that they can observe the painful lesion during arthroscopy, but this is not possible, pain and morphology are weakly correlated. This is a pragmatic trial and it is difficult to handle various surgeons who argue that this patient has to be excluded. Still only the clinical criteria should have been applied in order to reduce bias. I hope my minor comments may contribute in the process of evaluating and understanding the future results of this trial.

REVIEWER	Lars Adolfsson Department of Orthopaedics Linköping University Hospital Sweden
REVIEW RETURNED	28-Sep-2016

GENERAL COMMENTS	the study protocol is clear and well described. A number of questions and concerns however need further attention.

	* Clinical signs of full thickness rotator cuff ruptur would render exclusion. In my experience it would be impossible to ascertain a non-existing full thickness supraspinatus rupture in a patient with subacromial pain. I suggest that this is changed as I assume that the authors intend patients with Clinical signs of a major rupture resulting in marked weakness in any of the examined muscles? Later on apparently all patients underwent MRI with contrast and this was the final decision for eligibility? * Was fulfillment of all inclusion criteria mandatory? In that case, what is the reliability of item 5) isometric tests? Are they at all necessary? * Timing of steroid injections in relation to inclusion? * Calcific deposits allowed? Disregarded? * During the surgery apparently some patients were subjected to a partial bursetomy which reportedly may be a treatment for subacromial pain in itself. How is that managed in the analysis? * Was sucessful blinding controleld and assured? * "adequate relief of symptoms" at 6 months would allow crossover and change of treatment. If for instance a patients that underwent ASD was unsatisfied with the outcome, in what way was then treatment changed? Was the patient excluded or unblinded and regarded as a failure? How could blinding of the assessor be continued in such a situation? * How was inadequate relief of symptoms determined?

REVIEWER	Harvinder pal Singh University of Leicester United Kingdom
REVIEW RETURNED	08-Oct-2016

GENERAL COMMENTS	Protocol appears to have been published after the significant portion of the project has been completed as some data appears to have been analysed already, Like on page 25 first paragraph, "Despite thorough preoperative screening..... (n=13)). Does it need to be included. Page 17: Recruitment rate: Does this need to be published in the protocol of the study? Also the use of was and were in the paragraph gives the impression, study analysis is already complete. The authors have not discussed about night pain that these patients very commonly suffer and can be the presenting complaint. Authors have covered it by asking patients to comment about the shoulder pain in last 24 hours but night pain could specifically be discussed as the presenting problem in most patients. Page 2 Paragraph 2, The biggest difference between this study and the CSAW trial seems to be the Primary outcome measure. That could be mentioned in this paragraph. Page 12 Para 1: 'Deliberating', does it need to be debilitating? Page 16, Paragraph 2: Authors jump between present tense and past tense in this article. One example is use of was " each follow-up was documented" and page 16 para 1: 99% followup was achieved. Is this a publication of protocol or early results.

	Why was the allocation for surgery versus physiotherapy kept as 2:1, Its limitations could be discussed in the limitations section. Authors have not addressed how the partial tears would be addressed in the protocol, would they be excluded or included when assessing patients during surgery. Would patients be included if they had bilateral disease or would only one shoulder be included?

VERSION 1 – AUTHOR RESPONSE

Comments from the Reviewer Number 1:

** Shoulder pain is next to low back pain and neck pain the most common musculoskeletal condition. Subacromial pain syndrome, rotator cuff disease or tendinosis often referred to as the impingement syndrome is the major diagnosis. Paradoxically, surgical treatment has exploded after a trial in BMJ in 1993 found that it was not more effective than exercises. Later studies have replicated this finding, but current overuse of surgery has not decreased.

The present protocol describes a sham controlled trial evaluating the effectiveness of surgery. The secondary aim is to compare surgical results to physiotherapy guided exercises. This trial is highly warranted and the protocol describes a gold standard design even using blinded interpretation of results and writing of the manuscript. Patients are already included according to the estimated sample size and follow-ups are continued until the last 2-year follow-up is completed. The protocol is very interesting and of excellent quality in my opinion.

Authors' response: We thank the reviewer for the careful review of our paper and an attentive summary of the prevailing status on the treatment of SIS. We appreciate the positive general remarks on our paper and particularly the insightful suggestions for improvement.

**Abstract: A visual analogue scale (0-100) means that a scale is used as an analogue for pain intensity. To express this in mm is not the point, in my opinion it is better to delete mm.

Authors' response: Thank you for this attentive suggestion. We agree with the reviewer's suggestion (to delete "mm").

Authors' action: We have deleted "mm" from the VAS scale.

In addition, this suggestion by the reviewer also made us realize that we have not defined the word descriptors of the VAS scale properly. Accordingly, we have revised our Materials and Methods as follows (page 13, under the section on VAS): "Shoulder pain was assessed on a 100 mm scale ranging from 0 (no pain) to 100 (extreme pain)". Later on, format 0-100, is used.

** Introduction, page 5: Subacromial pain syndrome was used by K Engebretsen et al in BMJ in 2009, long before the referred guideline was published.

Authors' response: We thank the reviewer for reminding us of the history of the term "subacromial pain syndrome".

Authors' action: We have added the reference suggested by the reviewer (Engebretsen et al. Radial extracorporeal shockwave treatment compared with supervised exercises in patients with subacromial

pain syndrome - single blind randomised study. BMJ 2009 Sep 15; 339:b3360) as a reference #3 in the revised manuscript.

** The introduction suggests that surgery may be either therapeutical, placebo (or both) while neither the role or evidence of exercises is clearly reported. Also the similar ongoing UK trial lack the exercise arm.

Authors' response: We thank the reviewer for pointing out the limited discussion on the role/evidence of exercise therapy on SIS and also for highlighting that the inclusion of an exercise therapy group is an obvious difference/distinct feature between our trial and that of the CSAW.

Authors' action: Prompted by the reviewer's remarks, we have rephrased the Introduction as follows (page 4, 2nd para.): "Remarkably, there is a stark absence of evidence from high-quality controlled trials to support the existing practice of performing subacromial decompression for patients with SIS. Two recent systematic reviews concluded that subacromial decompression provides no superior benefits in terms of pain relief, function, or quality of life compared to non-surgical treatment 12,13. There is even a placebo controlled trial to show the beneficial effect of exercise therapy over placebo physiotherapy.14"

** Methods, page 8 and 9, table 1, sick leave is not listed.

Authors' response: Our baseline assessment included the following question: "Have the shoulder symptoms limited your ability to work normally?", with response options: "yes" or "no". However, unfortunately we did not specifically inquire about sick leave in our baseline questionnaire.

Authors' action: No action.

** Page 11, postoperative care is briefly described, how many physiotherapy sessions are expected as compared to the exercise group?

Authors' response: Participants in the exercise therapy (ET) group had 15 visits to an independent physiotherapist for guidance and monitoring of the progress. In the surgical groups (both ASD and DA groups), in turn, the postoperative care was pragmatic in nature: In essence, all surgically treated participants received one visit to an independent physiotherapist for guidance and instructions for home exercises. They were also given a referral to the physiotherapy unit of their own participating centre, but adherence (the number of physiotherapy visits) was not monitored in the ASD and DA arms.

Authors' action: The above noted has now been added to the text as follows: "In both ASD and DA groups, the postoperative rehabilitation was identical. All surgically treated participants received one visit to an independent physiotherapist for guidance and instructions for home exercises. Subsequent rehabilitation was carried out according to the standardized rehabilitation protocols of the participant centres." (page 11, under the title: "Postoperative care").

** Page 12, table 2, sick-leave is not listed, and how is it estimated? According to a previous study (Grovle L, Haugen A et al.) patients reliably report their current sickness benefit status but not the number of days on sick-leave.

Authors' response: Return to work was monitored at each follow-up visit (3, 6, 12 and 24 months) using the following questions:
• Have you been able to return to your normal employment? (yes/no)
• If not, is this due to the shoulder related problems? (yes/no)
We did not inquire about the number of days on sick-leave.

Authors' action: Information about return to work is added to the table 2 (Page 12).

** The trial has three related primary outcomes: arm activity pain, rest pain, and acceptable symptom intensity at 24 months. This is acceptable in my opinion, while others had preferred a self-reported pain disability index. Y Roe et al. (2016) published a comprehensive review of the most frequent outcomes using ICF taxonomy. My most important objection is the choice of 15 points as the minimally clinically important difference (MCID). MCID is not an exact estimate because it is calculated by using the most subjective post-treatment global evaluation as a gold standard. For neck and back pain MCID of VAS-pain is always higher than for an index made from a set of questions. Surprisingly MCID for shoulder pain is estimated to 15, lower than for relevant shoulder scores (see Ekeberg et al). Another way of evaluating the minimally clinically important difference is to ask the patient on beforehand, in this case: would you prefer surgery with a benefit of 15 point reduction of pain on a 0-100 scale?

Authors' response: We thank the reviewer for attentive remarks on the MCID. We share the reviewer's concerns about the appropriateness of the straightforward 15 points as the pre-determined cut-off point for the MCID. However, given that a criterion had to be chosen for the MCID and Tashjian et al. had carried out this particular analysis for rotator cuff disease (SIS) (and this value was recently used in another RCT on SIS, Ketola et al. 2013), we decided to stick with the value chosen. In retrospect, it is easy to agree with the reviewer that anchoring the MCID to a question/set of questions similar to those suggested by the reviewer would have been more optimal. However, at this point, such change unfortunately cannot be made, as we did not ask the required question(s).

Authors' action: No action.

** Also the choice of 30 as an acceptable state may represent a problem. How many of the included patients have a preoperative score of 30 to 45? Others have used patients reporting excellent or good as an acceptable state (by example Brox et al 1999) which is probably not much different from < 30, although possibly a stronger criterium.

Authors' response: We thank the reviewer for another insightful remark. Again, similar to the MCII/-D (previous comment), this issue poses a formidable intellectual challenge.

Regarding the reviewer's query on how many of the included patients have a preoperative score of 30 to 45: Although we have made a deliberate decision not to analyze any of the trial data beforehand, we decided to address this specific query. Our preliminary exploration demonstrated an evident inconsistency between the patient responses and the chosen PASS threshold of 30. As a result, we decided to re-evaluate the whole issue according to the reviewer's remarks. Based on this reassessment, we decided to use "patient's global satisfaction to the treatment" as the criterion/anchoring question for PASS (and PDSS). This change will have an effect on the entire responder analysis (see below).

Authors' action: Paragraph on "responder analysis" revised as follows (page 23-24): "As noted above, instead of focusing only on the statistical significance of the mean differences between treatment

groups in the VAS (i.e., the mean improvement from baseline to 24 months), we will also carry out “a responder analysis”. In principle, this analysis allows physicians to inform a patient of his or her chance of experiencing a clinically meaningful improvement from the treatment, both in absolute terms and in comparison, to a control group. The difference between responders and non-responders can be considered the net-benefit of the treatment. One proposed means to carry out a responder analysis relies on the assessment of the proportion of patients reaching the patient-acceptable symptom state (PASS) and the patient-disappointing symptoms state (PDSS). As no universal consensus exists on either the PASS or the PDSS in the context of SIS, we chose to anchor our responder analysis to the patient’s assessment of satisfaction with the shoulder treatment outcome: Patients reporting very satisfied or satisfied will be categorized as “Responders” and those reporting very dissatisfied or dissatisfied as “Non-responders”. Given the obvious coarseness of this approach, we plan to evaluate the appropriate criteria for PASS and PDSS in more detail in the future, exploring the potential contribution of, e.g., arm pain at rest and at activity, shoulder function, and night pain.”

This notion also prompts a following change to our assessment of the patient satisfaction/secondary outcome measures (page 14, under the title “Patient satisfaction…”): ”We elicited patients’ global assessment of satisfaction to the treatment with this question: “Are you satisfied with the treatment you have received?” We used a VAS scale ranging from 0 (completely disappointed) to 100 (completely satisfied).

Additionally, we elicited patient satisfaction to the treatment outcome with the following question at each follow-up time point (Table 2): “How satisfied are you with the outcome of your treatment?” on a 5-item scale. Participants who reported very satisfied or satisfied will be categorized as “Responders” and patients who responded very dissatisfied or dissatisfied as “Non-responders”.

** The study includes three pre-specified subgroup hypotheses (page 19/20). Analyses this hypotheses are just exploring and results are rarely important (see Pocock, NEJM 2016). Nevertheless, when subgroups are included I miss a fourth hypothesis: patients on sick leave versus patients not on sick leave (Koljonen et al 2009, Brox et al 1996 and 1999).

Authors’ response: We share the reviewer’s concerns regarding the credibility (confidence in the estimates) of subgroup analyses, as also highlighted in our manuscript (page 19, under the section entitled: “Secondary analyses…”): “These secondary analyses will be supportive, explanatory and/or hypothesis generating”. As for the suggestion to include a fourth hypothesis (based on sick leave), see our response to question #4. As noted, we unfortunately do not have preoperative data on sick leave status.

Authors’ action: No action.

** Under limitations (page 25) it is reported that 14 patients were excluded because of A-C arthrosis or intra-articular findings. It is normally wrong to exclude patients after randomisation because it reduces the effect of randomisation, fortunately exclusion includes only a small percentage of patients. Surgeons often believe that they can observe the painful lesion during arthroscopy, but this is not possible, pain and morphology are weakly correlated. This is a pragmatic trial and it is difficult to handle various surgeons who argue that this patient has to be excluded. Still only the clinical criteria should have been applied in order to reduce bias.

Authors’ response: We agree with the reviewer about the complexities of carrying out clinical (surgical) RCTs. The exclusion of patients after the first randomization (Phase I) is indeed an obvious limitation that we have discussed (page 24, under the section on “Limitations…”). However, be it also

noted that although this issue potentially hampers our secondary comparison (ET vs. ASD), the patients were excluded before the second randomization (Phase II), and thus, it has no effect on our primary comparison (ASD vs. DA).

Authors' action: No action.

Comments from the Reviewer Number 2:

** Clinical signs of full thickness rotator cuff rupture would render exclusion. In my experience it would be impossible to ascertain a non-existing full thickness supraspinatus rupture in a patient with subacromial pain. I suggest that this is changed as I assume that the authors intend patients with Clinical signs of a major rupture resulting in marked weakness in any of the examined muscles?

Authors' response: We agree with the reviewer's remark that it is difficult/impossible to clinically ascertain a non-existing full thickness supraspinatus tear in a patient with SIS. If a patient with subacromial pain has a marked weakness of rotator cuff muscles, this is probably caused by major tendon rupture or nerve impairment. However, a seemingly normal/only slightly weak muscle strength does not rule out a (minor) cuff tear. And this is why a comprehensive clinical examination including history and physical examination was always complemented by MRA before inclusion of the patient in the trial. MRA has been shown to yield 94% sensitivity and 92% specificity for detecting a major (clinically relevant) rotator cuff tear (Lenza et al. Cochrane Database of Systematic Reviews 2013, Issue 9. Art. No.: CD009020. DOI: 10.1002/14651858.CD009020.pub2). We thank the reviewer for pointing out that adding "marked weakness in any of the examined muscles" clearly specifies the criterion.

Authors' action: The first Exclusion criterion revised as follows: "1. Full thickness tear of the rotator cuff tendons diagnosed on clinical examination (marked weakness in any of the examined muscles) or magnetic…"(page 7, under the title "Exclusion criteria").

** Later on apparently all patients underwent MRI with contrast and this was the final decision for eligibility?

Authors' response: Yes, all potentially eligible patients did indeed undergo magnetic resonance image with intra-articular contrast
medium (MRA) and if any pathology (other than those "suggestive of subacromial pain syndrome") was found in MRA, patient was excluded from the study.

Authors' action: No action.

** Was fulfillment of all inclusion criteria mandatory? In that case, what is the reliability of item 5) isometric tests? Are they at all necessary?

Authors' response: Yes, fulfillment of all inclusion criteria was mandatory. Similar to any individual clinical tests for SIS, isometric (strength) tests are indeed quite unreliable (Abduction: sensitivity 55%, specificity 75%; External rotation: sensitivity 70%, specificity 50%)(e.g., Kelly et al. Clin Rehabil 2010). In general, clinical tests with high sensitivity tend to have low specificity, and vice versa. Having said that, it has been argued that "combination of shoulder physical tests provide better accuracy, but marginally so" (Hegedus et al. BJSM 2012). Because FIMPACT was designed as an efficacy trial, we

chose our eligibility criteria so that the participants recruited would represent patients with as “pure” an SIS as possible. As we used these criterion during the eligibility screening, we prefer leaving the eligibility criteria as is.

Authors’ action: No action

** Timing of steroid injections in relation to inclusion?

Authors’ response: Timing of steroid injections was not recorded.

Authors’ action: No action.

** Calcific deposits allowed? Disregarded?

Authors’ response: We thank the reviewer for pointing out this pertinent issue. Patients with substantial calcific deposits found in preoperative imaging (large enough to be considered requiring removal) were excluded from the study.

Authors’ action: We have now added this as an exclusion criterion (page 7): ”Substantial calcific deposits in the rotator cuff tendons found in the preoperative imaging”

** During the surgery apparently some patients were subjected to a partial bursetomy which reportedly may be a treatment for subacromial pain in itself. How is that managed in the analysis?

Authors’ response: We thank the reviewer for pointing out this important issue regarding our study. It is indeed true that there is evidence to suggest that bursectomy alone provides similar outcomes to bursectomy with acromioplasty in patients with SIS (e.g., Henkus et al. JBJS-Br 2009, Donnigan The Iowa Orth J. 2011). In our FIMPACT trial, bursa tissue was addressed (either bluntly stretched with troachar or resected) only if adequate visualisation of the tendons could not be achieved without it. Resection was kept minimal and only on the tendon side of the bursa. We believe that this kind of minimal resection of bursa has no true clinical effect. Furthermore, we felt that reliable visualisation of the rotator tendon insertions was more important than possible effect of a small bursa resection.

Authors’ action: No action.

** Was successful blinding controlled and assured?

Authors’ response: Yes, in the two operative groups, patients were asked to guess whether they had undergone ASD or DA (page 15, para. entitled “Patients’ perception of operative treatment-group assignment”) at the three month follow-up. Regarding “assurance” (of successful blinding), all patients still remain blind to their treatment group (except of those who have requested unblinding due to persisting symptoms).

Authors’ action: No action.

** "adequate relief of symptoms" at 6 months would allow crossover and change of treatment. If for instance a patients that underwent ASD was unsatisfied with the outcome, in what way was then

treatment changed? Was the patient excluded or unblinded and regarded as a failure? How could blinding of the assessor be continued in such a situation?

Authors' response: We addressed the clinical scenario outlined above very pragmatically: If a patient was not satisfied with the outcome of surgery, he/she was referred to physiotherapy until he/she requested unblinding. If no ASD was carried out (DA group), then ASD was offered. If, in turn, the unblinding proved that the patient had undergone ASD, we offered extended physiotherapy. In both of these occasions, the unblinding was documented.

As for the blinding of the outcome assessors, be it noted here that all our primary outcomes are patient-reported outcome measures. In addition, the designated physiotherapists carrying out the follow-up clinical examinations were different from those treating the participants and were also instructed not to inquire about group assignment/unblinding. Finally, all follow-up examinations were carried out with participants wearing a t-shirt.

Authors' action: Prompted by these attentive comments, we have added a sentence explaining our protocol for dealing with participants who had undergone ASD as follows (page 12, 1st para.): "If the participant was allocated to DA group, ASD was then offered. If the participant had undergone ASD, he/she was offered extended physiotherapy."

Similarly, the section on follow-up assessments was revised as follows (page 15-16, last three to 1st three lines): "…in addition to which they were also assessed clinically at 6 and 24 months (and 5 and 10 years) post randomisation by a study physiotherapist unaware of treatment allocation, treatment given or possible unblinding. Outcome assessors were instructed not to inquire anything about prior treatment. Further, all follow-up examinations were carried out with participants wearing a t-shirt."

** How was inadequate relief of symptoms determined?

Authors' response: Thank you for this attentive remark. We had no pre-specified criteria for "inadequate relief of symptoms" (debilitating symptoms). Rather, we left it to the participants and the study physicians' clinical judgment to make this call.

Authors' action: We have now revised this section as follows (page 12, 1st para.): "No pre-specified criteria were used for determining "inadequate relief of symptoms/debilitating symptoms", rather it was left to the participants and the study physicians to make the clinical judgment together."

Comments from the Reviewer Number 3:

** Protocol appears to have been published after the significant portion of the project has been completed as some data appears to have been analysed already, Like on page 25 first paragraph, "Despite thorough preoperative screening..... (n=13)). Does it need to be included.

Authors' response: The reviewer is indeed correct in pointing out that significant portion of the project is already completed (actually, we are getting ready to start the actual data analysis). As for the section pointed out by the reviewer, we agree that this detail could well be omitted from the protocol (and published in the actual study report). However, as we feel that providing this data (number of preoperative exclusions) is crucially important (an obvious limitation of our trial), we felt that entering into this discussion here (in the protocol) will save some precious space to be used for other purposes in the actual study report soon to be written.

Authors' action: No action.

** Page 17: Recruitment rate: Does this need to be published in the protocol of the study? Also the use of was and were in the paragraph gives the impression, study analysis is already complete.

Authors' response: We agree with the reviewer's suggestion to remove any mention on the recruitment rate.

Authors' action: Paragraph on recruitment rate omitted.

** The authors have not discussed about night pain that these patients very commonly suffer and can be the presenting complaint. Authors have covered it by asking patients to comment about the shoulder pain in last 24 hours but night pain could specifically be discussed as the presenting problem in most patients.

Authors' response: We fully agree with the reviewer that night pain can be considered one of the hallmark symptoms in patients with SIS. As also pointed out by the reviewer, our primary outcome measures, patient's perceived pain intensity at rest and at arm activity in the last 24 hours, do not specifically address this issue. However, the Constant-Murley score covers this particular issue as follows:
Unaffected sleep? Yes (2 points) No (0 points)

Authors' action: We have now added sleep disturbance as another secondary outcome measure and revised the text as follows (page 13, 2nd para. under the section on "Constant-Murley score"): "In addition, as night pain is considered one of the hallmark symptoms in patients with SIS and our two primary outcome measures (patient's perceived pain intensity at rest and at arm activity in the last 24 hours) do not specifically address this issue, a specific question from the Constant-Murley score (unaffected sleep: "Yes" or "No") will be analysed separately.

** Page 2 Paragraph 2, The biggest difference between this study and the CSAW trial seems to be the Primary outcome measure. That could be mentioned in this paragraph.

Authors' response: Thank you for this suggestion. Prompted by this comment (along with a suggestion by reviewer #1), we have added a sentence delineating the minor differences between our trial and that of the CSAW.

Authors' action: The following sentence is now added in our Introduction (page 5, 2nd para.): "As readily apparent, the two trials (FIMPACT vs. CSAW) are very similar in design with the only notable differences being the primary outcome measure (Pain at rest and after activity vs. Oxford Shoulder Score, a score that assesses both pain and ADL impairment), the primary outcome assessment point (24 months vs. 6 months), and the intervention delivered for the third group (exercise therapy vs. active monitoring with specialist reassessment), respectively.

** Page 12 Para 1: 'Deliberating', does it need to be debilitating?

Authors' response: Apologies for the typo.

Authors' action: Corrected.

** Page 16, Paragraph 2: Authors jump between present tense and past tense in this article. One example is use of was " each follow-up was documented" and page 16 para 1: 99% follow-up was achieved. Is this a publication of protocol or early results.

Authors' response: We thank the reviewer for this remark. As this protocol is being written while the trial is already/still ongoing, we decided to use a tense that is chronologically accurate for each individual phase of the study. Accordingly, phases that we had already completed were written in past tense, whereas phases/action to be done or currently undergoing were described in present tense. We hope that this clarifies our intentions. As for the 99% follow-up, this refers to our previous placebo-surgery controlled trial.

Authors' action: No action.

** Why was the allocation for surgery versus physiotherapy kept as 2:1, Its limitations could be discussed in the limitations section.

Authors' response: Actually, to obtain three balanced study groups (of 70 participants), we had to use a two-phase sequential randomization, in which participants were first randomized into non-operative or operative group with 1:2 ratio immediately after the baseline appointment (Phase I). Then, in Phase II, those allocated to operative treatment were further randomized to ASD or DA with 1:1 ratio, thus yielding a 1:1:1 allocation ratio for the three study arms, the ET, DA, and ASD.

Authors' action: We have rephrased this section, under the heading "Overview of study design", as follows: "To obtain three balanced study groups (of similar group size), we performed a two-fold, sequential randomization as follows: First, we randomized patients to surgical or conservative treatment in 2:1 ratio and then randomized those allocated to surgery to ASD or DA in 1:1 ratio" (page 6, 1st para.).

** Authors have not addressed how the partial tears would be addressed in the protocol, would they be excluded or included when assessing patients during surgery.

Authors' response: Thank you very much for pointing out this important issue. Patients with partial tears were included in the study, while those with full-thickness tears were excluded and cuff repair was carried out.

Authors' action: Prompted by the reviewer's attentive remarks, we have added the following notion on this issue (page 11, 1st full sentence): "Patients with partial tears were included in the study, while patients with a full-thickness tear were excluded and rotator cuff repair was carried out."

** Would patients be included if they had bilateral disease or would only one shoulder be included?

Authors' response: We thank the reviewer for another attentive remark. Patients with bilateral symptoms were not excluded from the study. If patient had bilateral symptoms, only one shoulder was included in the study.

Authors' action: We have now added this information to the manuscript as follows (page 6, last sentence): "If patient had bilateral symptoms, only one shoulder was included in the study."

VERSION 2 – REVIEW

REVIEWER	Jens Ivar Brox Dep of Physical Medicine and Rehabilitation Oslo University Hospital University of Oslo Norway
REVIEW RETURNED	09-Dec-2016

GENERAL COMMENTS	All my questions are satisfactory answered. I have read the revised version carefully and I recommend it for publication.

REVIEWER	Lars Adolfsson Department of Orthopaedics Linköping University Hospital Sweden
REVIEW RETURNED	02-Jan-2017

GENERAL COMMENTS	The authors have adequately addressed most of the important concerns and questions raised in the first Review. The protocol is relevant and although a difference between the treatment Groups will be difficult to demonstrate with the study design it will add to present knowledge. Two minor comments remain: 1. since one of the primary outcomes is pain at rest it seems appropriate to have this aspect clearly listed as one of the inclusion criteria 2. although mentioned briefly it would be nice if it was clearly stated that it is the change in VAS from baseline to 24 months that will be compared between the Groups and not median or mean values which one might assume when Reading the abstract and synopsis.

REVIEWER	Harvinder Pal Singh University of Leicester United Kingdom
REVIEW RETURNED	29-Jan-2017

GENERAL COMMENTS	Authors have addressed my comments, thank you

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Jens Ivar Brox
Institution and Country: Dep of Physical Medicine and Rehabilitation, Oslo University Hospital, University of Oslo, Norway

All my questions are satisfactory answered. I have read the revised version carefully and I recommend it for publication.

Authors' response: We thank the reviewer for the very constructive and insightful suggestions for edits that clearly improved our paper.

Reviewer: 2
Reviewer Name: Lars Adolfsson
Institution and Country: Department of Orthopaedics, Linköping University Hospital, Sweden

The authors have adequately addressed most of the important concerns and questions raised in the first Review. The protocol is relevant and although a difference between the treatment Groups will be difficult to demonstrate with the study design it will add to present knowledge.

Authors' response: We thank the reviewer for the careful review of our paper and we appreciate the positive general remarks and suggestions for improvement.

Two minor comments remain:

1. Since one of the primary outcomes is pain at rest it seems appropriate to have this aspect clearly listed as one of the inclusion criteria

Authors' response: We fully agree with the reviewer that pain at rest is a hallmark symptom of SIS. However, given that we did not explicitly define pain at rest as an inclusion criterion for our trial, we feel reluctant to change the eligibility criteria at this phase. It be noted though that our key inclusion criterion (#2) was "Subacromial pain for greater than 3 months with no relief from non-operative means", which can be considered to contain the requested information.

Authors' action: No action.

2. Although mentioned briefly it would be nice if it was clearly stated that it is the change in VAS from baseline to 24 months that will be compared between the Groups and not median or mean values which one might assume when Reading the abstract and synopsis.

Authors' response: We thank the reviewer for pointing out this important issue. This information is indeed provided in the detailed statistical analysis plan, but it is lacking from the abstract.

Authors' action: Prompted by the reviewer's remark, we have rephrased the abstract as follows (Page 2, Paragraph 2): "Our two primary outcomes are pain at rest and at arm activity, assessed using visual analog scale (VAS). We will quantify the treatment effect as the difference between the groups in the change in the VAS scales with the associated 95% confidence interval (CI) at 24 months."

Please note that this addition pushed us well over the word limit for the abstract (300 words), and accordingly, we had to shorten other parts of the abstract.

Reviewer: 3
Reviewer Name: Harvinder Pal Singh
Institution and Country: University of Leicester United Kingdom

Authors have addressed my comments, thank you

Authors' response: We thank the reviewer for the careful review of our paper and valuable suggestions for improvement.

VERSION 3 – REVIEW

REVIEWER	Lars Adolfsson Department of Orthopaedics Linköping University Hospital Linköping Sweden
REVIEW RETURNED	22-Feb-2017

GENERAL COMMENTS	The raised concerns have been addressed by the authors and following the revision I believe that the manuscript is fully acceptable