PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Physiotherapist-led treatment for Femoroacetabular Impingement Syndrome (The PhysioFIRST study): A protocol for a participant and assessor-blinded randomised controlled trial.
AUTHORS	Kemp, Joanne; Johnston, Richard; Coburn, Sally; Jones, Denise; Schache, Anthony; Mentiplay, Benjamin; King, Matthew; Scholes, Mark; De Oliveira Silva, Danilo; Smith, Anne; McPhail, Steven; Crossley, Kay

VERSION 1 – REVIEW

REVIEWER	Nadine Foster Keele University UK
REVIEW RETURNED	03-Aug-2020

GENERAL COMMENTS	This paper is a trial protocol paper for a trial that also have trial registry information. There are key strengths of the trial including 3rd party randomisation, attempt to limit disclosure of patient participants (they are unaware of the specific differences between the two intervention packages), the protocol follows SPIRIT guidance, the intervention description follows TIDier guidance, the way it builds on a previous pilot and feasibilty RCT and the plans for long-term follow-up. However, there are important areas / issues that need to be improved, before any publication: 1. the details on the trial registry are different in important ways that some of the details in this protocol paper, and this is worrisome. The key examples are a) the description of the trial's primary outcomes are different - in the registry there are 2 primary outcomes - the first is listed as the perceived global rating of change (Likert scale) and the second listed is the iHOT-33, yet this paper states the iHOT-33 is the primary and the perceived change is a secondary outcome. This is important to sort out, as the trial's key results and conclusions will be determined by the primary outcome(s). Also its important for sample size, trials with two co-primary outcomes need to power their trial to detect differences on both primary outcomes, whereas the sample size calculation included in this paper mentions only the iHOT-33. Its also important for the statistical analysis plan (sometimes trials change the significance level set when using more than one primary outcome), b) the descriptions of the two intervention groups are different in important ways between the trial registry information and this paper - again this is worrisome (although I do understand

	these complex multicomponent complex interventions are always complex to convey in words within word-counts of journals) - key examples are in the registry information there is clearly tailoring in the control arm eg. manual therapy tailored to ROM deficits, standardised stretching with some modifications, the same individualised health education sessions (as the intervention arm), yet the description in the paper is that this arm is a standardised control (standardised health education, non-specific standardised stretching, standardised manual therapy, and a physical activity program). I find the differences in description of the intervention components worrisome. Has the plan changed since the trial was registered? 2. The justification for having stretching in the control arm package of care is based on their pilot RCT, the paper references it following the statement that it has previously not been shown to be effective - yet the pilot RCT paper (pg 311 and pg 312) showed that participants in that arm had moderate to large gains in hip adductor (ES 0.54) and extensor (ES 0.66) strength, and that ' the improvement seen in the control group did exceed the MIC'. Could the authors clarify what they mean and what the justification is for the components within the control arm? 3. The overall aim of this trial is described throughout various sections of the paper as 'to test the efficacy of a physiotherapy-led intervention for FAI' but actually that is not what this trial is designed to do. This trial compares a complex physiotherapy-led intervention of type A, B and C with another complex physiotherapy-led intervention of type A, B and D - many components of the two packages are the same (this is clear is the table); hence actually this trial is really testing the comparative effectiveness of two physiotherapy-led multicomponent packages of care and I strongly believe the title, abstract, aim, and conclusion need to use this language. Whilst I realise there are different understandings of terms like efficacy and effectiveness - usually efficacy is about determining if a treatment 'works' (in this case, if physio-led treatment 'works' - but this would then need to be compared with a treatment that is not physio-led, which is not the case with this trial). The key difference between arms is not whether the treatment is physio-led but the inclusion of muscle strengthening and functional task retraining activities? 4. Its not clear if physiotherapists are providing treatment in both arms of this trial, or whether different therapist provide treatment to those in arm A than the physiotherapists providing care to patients in arm B. This is important given a) readers need to understand the level of care taken to avoid intervention-deliverer bias (in the case that the physios delivering the treatment are more strongly in favour themselves of say the targeted intervention, which is very possible) and to avoid contamination bias (where the same therapists are trained to deliver both interventions and given one cannot be 'untrained' once trained, elements of the targeted arm leak into the standardised arm). If different therapists are delivering each intervention, then clustering effects might need to be at least considered in the statistical analysis. If the same therapists are delivering each intervention, then some description of how the research team has identifed, engaged and trained therapists to be in collective equipoise about the comparative effectiveness of the two interventions is needed in the paper. How many therapists and where are they from? Do they reflect usual

	physiotherapists that (outside of this trial) people with FAI might be seeking care from? This research team have previously published strong statements about the need for FAI care to include treatments that address muscle weakness, low trunk strength, poor balance and low functional task performance (ie. they have argued in the peer reviewed literature that rehab must include muscle strengthening and functional task training) - so I think it is possible that therapists involved in this trial may well be very aware of these views and as such not really be in collective equipose (thus potentially biasing the between arm comparison if they are treating patients in both arms). It would be useful to know how the team identified, engaged, and trained the physiotherapists involved, and if not too late to do, identify and be able to describe the physiotherapists views about equipoise in this trial. If possible it would be also useful to describe the previous experience of these physiotherapists in treating FAI, and in particular whether they believe muscle strengthening and functional task retraining are essential components of their care. How many physiotherapists are trained/delivering care overall and to each arm? 5. The paper, erroneously, states this is the first full-scale study to test the efficacy of a physio-led intervention for FAI - this is clearly not the case given at least 3 RCTs that have previously done just this (comparing physio-led interventions with surgery). Also it makes erroneous statements about some of those previous trials (one of which I was involved in, so I know it particularly well). Page 4 states that those previous physio-led interventions consisted of non-targeted, non-progressive exercises. This is misleading. UK FASHIoN PHT comprised the key characteristics of individualisation, supervision and progression of an exercise program (prescribed and supervised by physiotherapists in the UK NHS and supplemented by a home programme of exercise that was also reviewed and progressed, supported by exercise diaries). Its important to ensure what is written in this paper is accurate. Minor points: 1. when referring to previous guidelines, the references are rather old (10 years old for example for back pain), yet there are more recently developed/published guidelines in countries such as the USA, the UK and Denmark 2. page 6 says people with other MSK conditions were excluded, yet FAI patients often present with pain/problems in other body regions often in the back and indeed in the appendices the intervention clearly includes treatment of lumbar dysfunction. Can this be clarified? 3. Great that an MRI scan is included at 12 months, and there are plans for long-term follow-up on various outcome measures, but its not clear at what point (if any) patients will be unblinded in this trial. It would be good to specify whether patients will be told which group they were in, and when any results from the primary timepoint of 6 months will be shared with participants (as this may affect their longer-term outcomes). 4. The trial is powered to detect at least a moderate effect size (0.5) between the intervention arms, which given previous RCTs of exercise A versus exercise B for other MSK conditions (which mostly show no between group difference) is ambitious. Most RCTs that have previously demonstrated a between-arm difference of moderate or higher ES have been comparing exercise with a no-exercise comparison. I would have powered

	this trial for a small ES, as that would be more realistic (both arms of this trial are getting intensive interaction and a package with multiple components, with 9 physio sessions, a further 12 supervised gym sessions etc). I did wonder that given the team are collecting iHOT-33 and global rating of change at 6 and 4 time-points respectively that they might like to consider using all of these time-points in their primary analyses to increase the power they will have to detect between arm differences. If their statistical analysis plan is not already finalised this may be possible. 5. Keele STarT MSK tool is written Keele STartT MSK tool in error, its also not clear which version the team are using (the validated tool has 12 items, so they may be using the earlier development version?). I hope my review of this paper is helpful. Its great to see this team doing a trial to work out the comparative effectiveness of two physio-led treatment packages for FAI. I also like that they are including cost-effectiveness analyses, although did wonder whether they might really be able to see a difference in treatment costs given they have matched the two arms for the same number of sessions/treatment and gym time. This will mean there will need to be important differences in general quality of life, or further healthcare costs, between arms, for there to be a clearly better intervention in terms of cost-effectiveness.

REVIEWER	Karen Barker Professor of Physiotherapy Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences University of Oxford Oxford UK
REVIEW RETURNED	09-Sep-2020

GENERAL COMMENTS	Introduction - the synthesis of the existing surgery versus physio trials does conflates the design of several trials and incorrectly states that the comparison was non-targeted, non progressive exercises which is not accurate for all of the trials you have described. Please separate out the interventions used for these trials and re-write to accurately reflect content (albeit that does not alter rationale for this study. Justify decision to compare 2 physiotherapy interventions as opposed to best practice physio to control or waiting list intervention. Given that the intervention is based on 'best practice from consensus - describe how will stop physiotherapists in 'standard treatment providing best practice treatment - in effect are artificially constraining 'standard' so better to describe as comparison of two physiotherapy regimes and describe more clearly differences between them. Clear description in methods of procedures and intervention described clearly in accordance with TIDieR checklist. Standardised arm less clearly described - care with wording this is not reflecting standard care but a carefully constrained alternative intervention. More description of steps taken to avoid contamination between arms and therapists would be helpful.

	Sample size - remove sentence 'likely to be clinically meaningful in this population - reference 62 does not give the data to define the MCID of the iHOT-33.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Nadine Foster
Keele University
UK

Please state any competing interests or state ‘None declared’:
No competing interests

Comments to the Author
This paper is a trial protocol paper for a trial that also have trial registry information.

There are key strengths of the trial including 3rd party randomisation, attempt to limit disclosure of patient participants (they are unaware of the specific differences between the two intervention packages), the protocol follows SPIRIT guidance, the intervention description follows TIDier guidance, the way it builds on a previous pilot and feasibilty RCT and the plans for long-term follow-up.
Author response: Thank you for highlighting the strengths of the trial.
Author action: None

However, there are important areas / issues that need to be improved, before any publication:

1. the details on the trial registry are different in important ways that some of the details in this protocol paper, and this is worrisome. The key examples are:
a) the description of the trial's primary outcomes are different - in the registry there are 2 primary outcomes - the first is listed as the perceived global rating of change (Likert scale) and the second listed is the iHOT-33, yet this paper states the iHOT-33 is the primary and the perceived change is a secondary outcome. This is important to sort out, as the trial's key results and conclusions will be determined by the primary outcome(s). Also its important for sample size, trials with two co-primary outcomes need to power their trial to detect differences on both primary outcomes, whereas the sample size calculation included in this paper mentions only the iHOT-33. Its also important for the statistical analysis plan (sometimes trials change the significance level set when using more than one primary outcome).
Author response: Thank you for identifying this inconsistency in reporting of trial primary outcomes. We agree this was an oversight and should match the clinical trial registry. We have amended the manuscript to reflect the multiple primary outcomes1 as indicated in the clinical trial registry. We have also amended the sample size justification and statistical analysis section to reflect this correction,
Author action: This has been amended in the study aims, outcome measures, and outcomes tables. The aims now read “Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching in 164 participants with FAI syndrome on hip-related QOL (International Hip Outcome Tool 33 (iHOT-33)) and patient-perceived global improvement at 6-months. We hypothesise that, compared to the standardised stretching physiotherapist-led intervention, the targeted strengthening physiotherapist-led intervention will result in greater improvement in: (i) hip-related QOL and (ii) patient-perceived global improvement. Secondary aims are to measure: (i) the cost-effectiveness of the targeted strengthening physiotherapist-led intervention compared to the standardised stretching physiotherapist-led intervention; (ii) the effects of targeted strengthening

physiotherapist-led intervention on physical activity levels; (iii) the effects of targeted strengthening physiotherapist-led intervention on hip strength; and explore (iv) the effects of targeted strengthening physiotherapist-led intervention on hip biomechanics; and (v) the effects of targeted strengthening physiotherapist-led intervention on hip joint structure."
The outcomes section now reads
"Primary Outcomes
We will collect multiple (two) primary endpoints.29
Hip-related QOL will be measured using the iHOT-33. The iHOT-33 questionnaire consists of 33 individual questions scored on a visual analogue scale from zero (worst possible score) to 100 (best possible score). The iHOT-33 has acceptable psychometric properties and is recommended for use in active adults with hip-related pain.40 41 It has a low standard error of measurement (6 points),42 is responsive,43 with reported minimal clinically important differences ranging from 6 to 10 points 43 and minimal detectable change (groups) of 2 points.42
Patient-perceived global improvement will be measured on a 7-point Likert scale ('much improved', 'improved', 'a little improved', 'no change', 'a little worse', 'worse', 'much worse'). This is a clinically relevant tool for evaluating an individual patient's perspective on meaningful improvement.44"
Patient-perceived global improvement has now been listed as a primary outcome in Table 2.
The sample size calculation now reads "…Therefore, the proposed SMD was reduced to 0.50. This is consistent with previously reported SMD for the second primary outcome (patient-perceived global improvement) of 0.50. Estimated sample sizes for a two-sample means test t test assuming 80% power, α=0.025 (accounting for multiple primary outcomes), results in a sample size estimate of 156 participants."
The statistical analyses section now includes "The two primary endpoints chosen will be evaluated separately, such that a significant treatment effect against either of the endpoints will be taken as evidence of efficacy.29"

b) the descriptions of the two intervention groups are different in important ways between the trial registry information and this paper - again this is worrisome (although I do understand these complex multicomponent complex interventions are always complex to convey in words within word-counts of journals) - key examples are in the registry information there is clearly tailoring in the control arm eg. manual therapy tailored to ROM deficits, standardised stretching with some modifications, the same individualised health education sessions (as the intervention arm), yet the description in the paper is that this arm is a standardised control (standardised health education, non-specific standardised stretching, standardised manual therapy, and a physical activity program). I find the differences in description of the intervention components worrisome. Has the plan changed since the trial was registered?
Author response: Many thanks for identifying the inconsistencies in reporting of the trial intervention in the manuscript compared to the clinical trial registry. The treatment plan has not changed since the trial was registered. This is reflected in the treatment manuals attached to the original manuscript submission as supplementary files 1 and 2. We have modified the manuscript to reflect the individualised nature of the education and manual therapy received by the second treatment group. This now reflects the trial registration document which stated "In Phase I (week 0-12), the control participants' face to face treatment will also include manual therapy tailored to any assessed range of motion deficits in hip or spinal joints. However, they will receive a standardised cardiovascular training program with one progression based on achievement of training goals within parameters of perceived effort and pain. The control participants will receive the same individualised health education sessions covering topics such as exercise, diet, weight loss and appropriate stretching. Control participants will not do any hip or trunk muscle retraining or strengthening nor any functional proprioceptive or sports/activity specific retraining. Control participants will instead do a standardised stretching program which may be minimally modified by exercise selected if they cause discomfort."
Author action: We have updated the description of the standardised stretching group in the text and the table to accurately reflect the interventions.

“Standardised physiotherapist-led intervention
The standardised physiotherapist-led intervention consists of tailored health education, non-specific, standardised stretching, a standardised physical activity program and manual therapy individualised to participants’ needs. In order to control for the psychosocial effects of therapist contact inherent with physiotherapy intervention, this program will provide a credible alternative to physiotherapy exercises to reduce the possibility of resentful demoralisation. Stretching was chosen as our pilot work showed a smaller effect than a targeted strengthening intervention on hip-related quality of life and muscle strength.6 (Supplementary File 2).”

2. The justification for having stretching in the control arm package of care is based on their pilot RCT, the paper references it following the statement that it has previously not been shown to be effective - yet the pilot RCT paper (pg 311 and pg 312) showed that participants in that arm had moderate to large gains in hip adductor (ES 0.54) and extensor (ES 0.66) strength, and that ' the improvement seen in the control group did exceed the MIC'. Could the authors clarify what they mean and what the justification is for the components within the control arm?
Author response: We have clarified our rationale for using a standardised stretching program as our comparator intervention in the methods section
Author action: “Stretching was chosen as our pilot work showed a smaller effect than a targeted strengthening intervention on hip-related quality of life and muscle strength.”

3. The overall aim of this trial is described throughout various sections of the paper as 'to test the efficacy of a physiotherapy-led intervention for FAI' but actually that is not what this trial is designed to do. This trial compares a complex physiotherapy-led intervention of type A, B and C with another complex physiotherapy-led intervention of type A, B and D - many components of the two packages are the same (this is clear is the table); hence actually this trial is really testing the comparative effectiveness of two physiotherapy-led multicomponent packages of care and I strongly believe the title, abstract, aim, and conclusion need to use this language. Whilst I realise there are different understandings of terms like efficacy and effectiveness - usually efficacy is about determining if a treatment 'works' (in this case, if physio-led treatment 'works' - but this would then need to be compared with a treatment that is not physio-led, which is not the case with this trial). The key difference between arms is not whether the treatment is physio-led but the inclusion of muscle strengthening and functional task retraining activities?
Author response: Thank you for raising this point which has highlighted the challenge of designing robust comparator arms in physiotherapist-led, exercise-based clinical trials. We have adjusted the aims of the trial throughout the manuscript to more accurately reflect the two treatment arms.
Author action: The aim of the study now reads “Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching in 164 participants with FAI syndrome on hip-related QOL (International Hip Outcome Tool 33 (iHOT-33)) and patient-perceived global improvement at 6-months. We hypothesise that, compared to the standardised stretching physiotherapist-led intervention, the targeted strengthening physiotherapist-led intervention will result in greater improvement in: (i) hip-related QOL and (ii) perceived improvement. Secondary aims are to measure: (i) the cost-effectiveness of the targeted strengthening physiotherapist-led intervention compared to the standardised stretching physiotherapist-led intervention; (ii) the effects of targeted strengthening physiotherapist-led intervention on physical activity levels; (iii) the effects of targeted strengthening physiotherapist-led intervention on hip strength; and explore (iv) the effects of targeted strengthening physiotherapist-led intervention on hip biomechanics; and (v) the effects of targeted strengthening physiotherapist-led intervention on hip joint structure.”

4. Its not clear if physiotherapists are providing treatment in both arms of this trial, or whether different therapist provide treatment to those in arm A than the physiotherapists providing care to patients in arm B. This is important given a) readers need to understand the level of care taken to avoid

intervention-deliverer bias (in the case that the physios delivering the treatment are more strongly in favour themselves of say the targeted intervention, which is very possible) and to avoid contamination bias (where the same therapists are trained to deliver both interventions and given one cannot be 'untrained' once trained, elements of the targeted arm leak into the standardised arm). If different therapists are delivering each intervention, then clustering effects might need to be at least considered in the statistical analysis. If the same therapists are delivering each intervention, then some description of how the research team has identifed, engaged and trained therapists to be in collective equipoise about the comparative effectiveness of the two interventions is needed in the paper. How many therapists and where are they from? Do they reflect usual physiotherapists that (outside of this trial) people with FAI might be seeking care from? This research team have previously published strong statements about the need for FAI care to include treatments that address muscle weakness, low trunk strength, poor balance and low functional task performance (ie. they have argued in the peer reviewed literature that rehab must include muscle strengthening and functional task training) - so I think it is possible that therapists involved in this trial may well be very aware of these views and as such not really be in collective equipose (thus potentially biasing the between arm comparison if they are treating patients in both arms). It would be useful to know how the team identified, engaged, and trained the physiotherapists involved, and if not too late to do, identify and be able to describe the physiotherapists views about equipoise in this trial. If possible it would be also useful to describe the previous experience of these physiotherapists in treating FAI, and in particular whether they believe muscle strengthening and functional task retraining are essential components of their care. How many physiotherapists are trained/delivering care overall and to each arm?
Author response: Thank you for highlighting your concerns about training, equipoise and contamination of treating therapists. We have now provided more detail of these aspects in the Interventions section of the Methods. We did not specifically gain equipoise of the treating therapists for this study, as we have published this previously in our pilot RCT, where treating therapists felt that both interventions were credible (Kemp JOSPT 2018).
Author action: “Study participants will receive one of two physiotherapist-led interventions (targeted strengthening physiotherapist-led intervention or standardised stretching physiotherapist-led treatment) across four clinical sites within Victoria (Australia). Registered physiotherapists will lead the two-phase intervention (Table 2) that will be delivered over a 6-month period and has been described using the Template for Intervention Description and Replication (TIDieR) guidelines.32 Physiotherapists will be trained to deliver the intervention to both groups. Training of the physiotherapists will occur at the commencement of the study and annually thereafter. Treating physiotherapists will also be provided with written treatment manuals and training materials to refer to. In order to limit the likelihood of contamination between treatment groups, treating physiotherapists will be instructed to not have participants from different treatment groups attend the clinic at the same time. We have previously reported treating therapists’ beliefs that both interventions are credible. 6 In order to maintain participant blinding, treating physiotherapists will be trained to deliver both interventions with equal enthusiasm. Each of the four clinical sites will have between three and five therapists trained, depending on clinic requirements. The treating physiotherapists were recruited from four large private physiotherapy clinics in Australia and represent a typical therapist in an Australian private practice where people with FAI syndrome might seek care.”

5. The paper, erroneously, states this is the first full-scale study to test the efficacy of a physio-led intervention for FAI - this is clearly not the case given at least 3 RCTs that have previously done just this (comparing physio-led interventions with surgery). Also it makes erroneous statements about some of those previous trials (one of which I was involved in, so I know it particularly well). Page 4 states that those previous physio-led interventions consisted of non-targeted, non-progressive exercises. This is misleading. UK FASHIoN PHT comprised the key characteristics of individualisation, supervision and progression of an exercise program (prescribed and supervised by physiotherapists in the UK NHS and supplemented by a home programme of exercise that was also reviewed and progressed, supported by exercise diaries). Its important to ensure what is written in

this paper is accurate.
Author response: We have modified this section of the introduction to (1) provide more detail of the individual RCTs that compared physiotherapy to hip arthroscopy; and (2) highlighted the absence of full scale RCTs that offer a head-to-head comparison of two exercise-based physiotherapist-led interventions.
Author action:
“The physiotherapist-led interventions used for comparison to hip arthroscopy were varied in the degree of detail reported and content of the exercise interventions. The RCT by Griffin et. al. compared hip arthroscopy to personalised hip therapy, which included an exercise programme featuring individualisation, progression, and supervision.20 Palmer et. al. described a tailored programme to improve core stability and movement control, but little detail was provided on how this was achieved.22 Mansell et. al. described in detail their programme of stretching and motor control exercises.21 However it is unclear whether the exercises described in these studies were developed based on contemporary knowledge of impairments in FAI syndrome23, or be of sufficient stimulus and dosage12 24 to address the deficits in strength and functional performance that exist in these patients25. Thus, a physiotherapist-led intervention that compares exercise interventions should be developed and tested.
“However, absence of a full-scale RCT comparing the head-to-head effectiveness of two exercise-based, physiotherapist-led interventions for FAI syndrome25-27 limited the strength of such recommendations.”

Minor points:
1. when referring to previous guidelines, the references are rather old (10 years old for example for back pain), yet there are more recently developed/published guidelines in countries such as the USA, the UK and Denmark
Author response: Thank you for highlighting the age of the references cited, these have been replaced with more recent references.
Author action:
“…evident from clinical guidelines for osteoarthritis,13 low back pain,14 and chronic whiplash - associated disorders15”
“13. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation Guideline for the Management of Osteoarthritis of the Hand, Hip, and Knee. Arthritis Care Res (Hoboken) 2020;72(2):149-62. doi: 10.1002/acr.24131 [published Online First: 2020/01/08]
14. Oliveira CB, Maher CG, Pinto RZ, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society 2018;27(11):2791-803. doi: 10.1007/s00586-018-5673-2 [published Online First: 2018/07/05]
15. Bussières AE, Stewart G, Al-Zoubi F, et al. The Treatment of Neck Pain–Associated Disorders and Whiplash-Associated Disorders: A Clinical Practice Guideline. Journal of Manipulative & Physiological Therapeutics 2016;39(8):523-64.e27. doi: 10.1016/j.jmpt.2016.08.007”

2. page 6 says people with other MSK conditions were excluded, yet FAI patients often present with pain/problems in other body regions often in the back and indeed in the appendices the intervention clearly includes treatment of lumbar dysfunction. Can this be clarified?
Author response: Thank you for highlighting the complexity of clinical diagnoses of causes of pain in the hip and groin region. We believe our inclusion criteria, which reflect the Warwick agreement’s requirement for the diagnosis of FAI syndrome are as robust as is possible in a real-life clinical setting. In order to ensure our criteria are as clear as possible, we have added extra detail to the exclusion criteria, that a condition where FAI syndrome is not considered to be the primary cause of pain will be excluded.
Author action: “(vi) neurological, other MSK, or systemic arthritis conditions including other significant

musculoskeletal conditions where FAI syndrome was not considered to be the primary cause of hip pain"

3. Great that an MRI scan is included at 12 months, and there are plans for long-term follow-up on various outcome measures, but its not clear at what point (if any) patients will be unblinded in this trial. It would be good to specify whether patients will be told which group they were in, and when any results from the primary timepoint of 6 months will be shared with participants (as this may affect their longer-term outcomes).
Author response: We have now clarified that participants will become unblinded once the data analyses are complete. We do not expect that emergency unblinding will be required due to the very low incidence of adverse events seen in our pilot study of the same trial interventions.
Author action: "Participants will become unblinded once the data analyses are complete. We do not expect that emergency unblinding will be required due to the very low incidence of adverse events seen in our pilot study of the same trial interventions."

4. The trial is powered to detect at least a moderate effect size (0.5) between the intervention arms, which given previous RCTs of exercise A versus exercise B for other MSK conditions (which mostly show no between group difference) is ambitious. Most RCTs that have previously demonstrated a between-arm difference of moderate or higher ES have been comparing exercise with a no-exercise comparison. I would have powered this trial for a small ES, as that would be more realistic (both arms of this trial are getting intensive interaction and a package with multiple components, with 9 physio sessions, a further 12 supervised gym sessions etc). I did wonder that given the team are collecting iHOT-33 and global rating of change at 6 and 4 time-points respectively that they might like to consider using all of these time-points in their primary analyses to increase the power they will have to detect between arm differences. If their statistical analysis plan is not already finalised this may be possible.
Author response: Thank you for this suggestion, we have modified our primary outcomes to accurately reflect the published clinical trial registration and feel that deviating from this plan would cause conflict with the project ethics and clinical trial registration.
Author action: none.

5. Keele STarT MSK tool is written Keele STartT MSK tool in error, its also not clear which version the team are using (the validated tool has 12 items, so they may be using the earlier development version?).
Author response: We have corrected the typo identified. We have highlighted that we are using the 10-item clinical version of the tool.
Author action: "Keele STarT MSK Tool© Clinical version, contains 10 items that ask the participant about their function and disability, pain and coping, comorbidity, and the impact of pain. Once scored, it places the patient into three categories based on their risk of a poor outcome (low, medium, high). This tool has moderate-to-good level predictive ability in the identification of patients who develop persistent disabling pain."

I hope my review of this paper is helpful. Its great to see this team doing a trial to work out the comparative effectiveness of two physio-led treatment packages for FAI. I also like that they are including cost-effectiveness analyses, although did wonder whether they might really be able to see a difference in treatment costs given they have matched the two arms for the same number of sessions/treatment and gym time. This will mean there will need to be important differences in general quality of life, or further healthcare costs, between arms, for there to be a clearly better intervention in terms of cost-effectiveness.
Author response: Thank you Prof Foster, your detailed review was very helpful. We appreciate your comments regarding cost-effectiveness analysis and agree it may be difficult to see differences given the similar amount of physiotherapist-led treatment received.

Author action: none.

Reviewer: 2

Karen Barker
Professor of Physiotherapy
Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences
University of Oxford
Oxford
UK

Please state any competing interests or state ‘None declared’:
None Declared
Comments to the Author
Introduction - the synthesis of the existing surgery versus physio trials does conflates the design of several trials and incorrectly states that the comparison was non-targeted, non progressive exercises which is not accurate for all of the trials you have described. Please separate out the interventions used for these trials and re-write to accurately reflect content (albeit that does not alter rationale for this study.
Author response: This concern was also highlighted by reviewer 1. We have rewritten the paragraph in question in the introduction to more explicitly describe the body of evidence to date, especially in relation to previously reported physiotherapist-led interventions.
Author action:
“The physiotherapist-led interventions used for comparison to hip arthroscopy were varied in the degree of detail reported and content of the exercise interventions. The RCT by Griffin et. al. compared hip arthroscopy to personalised hip therapy, which included an exercise programme featuring individualisation, progression, and supervision.20 Palmer et. al. described a tailored programme to improve core stability and movement control, but little detail was provided on how this was achieved.22 Mansell et. al. described in detail their programme of stretching and motor control exercises.21 However it is unclear whether the exercises described in these studies were developed based on contemporary knowledge of impairments in FAI syndrome23, or be of sufficient stimulus and dosage12 24 to address the deficits in strength and functional performance that exist in these patients25. Thus, a physiotherapist-led intervention that compares exercise interventions should be developed and tested.
“However, absence of a full-scale RCT comparing the head-to-head effectiveness of two exercise-based, physiotherapist-led interventions for FAI syndrome25-27 limited the strength of such recommendations.”

Justify decision to compare 2 physiotherapy interventions as opposed to best practice physio to control or waiting list intervention. Given that the intervention is based on 'best practice from consensus - describe how will stop physiotherapists in 'standard treatment providing best practice treatment - in effect are artificially constraining 'standard' so better to describe as comparison of two physiotherapy regimes and describe more clearly differences between them.
Author response: This concern was also highlighted by reviewer 1. We have now modified the descriptions of the 2 intervention arms to more accurately describe them as two physiotherapist-led treatment arms rather than describing the second treatment arm as a control intervention. The details are provided above in the response to reviewer 1.
Author action: The description of the 2 treatment arms has been changed throughout the manuscript.

Clear description in methods of procedures and intervention described clearly in accordance with TIDieR checklist.

Author response: None required.

Standardised arm less clearly described - care with wording this is not reflecting standard care but a carefully constrained alternative intervention.
Author response: Thank you for this suggestion – we have changed the wording of the second treatment arm as suggested by reviewer 1, with changes described above in our response to reviewer 1.
Author action: The description of the second treatment arms has been changed throughout the manuscript to more accurately reflect the components of care received.

More description of steps taken to avoid contamination between arms and therapists would be helpful.
Author response: Reviewer 1 had similar concerns. All treating physiotherapists were trained to treat both trial arms. We have added more detail on how the therapists were trained and other steps taken to ensure minimal contamination between treatment arms. These are described in detail above in response to reviewer 1's comments.
Author action: "Study participants will receive one of two physiotherapist-led interventions (targeted strengthening physiotherapist-led intervention or standardised stretching physiotherapist-led treatment) across four clinical sites within Victoria (Australia). Registered physiotherapists will lead the two-phase intervention (Table 2) that will be delivered over a 6-month period and has been described using the Template for Intervention Description and Replication (TIDieR) guidelines.32 Physiotherapists will be trained to deliver the intervention to both groups. Training of the physiotherapists will occur at the commencement of the study and annually thereafter. Treating physiotherapists will also be provided with written treatment manuals and training materials to refer to. In order to limit the likelihood of contamination between treatment groups, treating physiotherapists will be instructed to not have participants from different treatment groups attend the clinic at the same time. We have previously reported treating therapists' beliefs that both interventions are credible. 6 In order to maintain participant blinding, treating physiotherapists will be trained to deliver both interventions with equal enthusiasm. Each of the four clinical sites will have between three and five therapists trained, depending on clinic requirements. The treating physiotherapists were recruited from four large private physiotherapy clinics in Australia, and represent a typical therapist in an Australian private practice where people with FAI syndrome might seek care."

Sample size - remove sentence 'likely to be clinically meaningful in this population - reference 62 does not give the data to define the MCID of the iHOT-33.
Author response and action: We have removed this sentence as suggested by reviewer 2.

1. McLeod C, Norman R, Litton E, et al. Choosing primary endpoints for clinical trials of health care interventions. Contemp Clin Trials Commun 2019;16:100486-86. doi: 10.1016/j.conctc.2019.100486

VERSION 2 – REVIEW

REVIEWER	Nadine Foster Keele University, UK
REVIEW RETURNED	28-Dec-2020

GENERAL COMMENTS	This is a revised protocol paper, summarising a RCT that is ongoing. The requested revisions have in large part been made, mainly to make the description of previous RCTs' exercise interventions more accurate, to improve the detail of the RCT

	methods section, and to address the differences between the detailed protocol paper and the previous trial registry information. However, the key areas that still need improvement are: 1. clarification for the rationale for this trial - the background section offers two rationales. The first focuses on the limitations of previous RCTs (that all compared physiotherapy-led exercise based treatment versus surgery) and the second seems to be about evidencing recommendations that a recent expert consensus group made that were not based on RCT evidence. I have tried to follow these two rationales and have struggled to see how EITHER lead to the current trial. I therefore believe the rationale for THIS trial needs clearer articulation in the background section. This point is similar to the point made before about being really clear what the aims are of this trial (as a reminder I previously commented "The overall aim of this trial is described throughout various sections of the paper as 'to test the efficacy of a physiotherapy-led intervention for FAI' but actually that is not what this trial is designed to do"). - in the first rationale for the current trial in this revised paper - the text says - "However it is unclear whether the exercises described in these (previous) studies were developed based on contemporary knowledge of impairments in FAI syndrome, or be of sufficient stimulus and dosage to address the deficits in strength and functional performance that exist in these patients. Thus, a physiotherapist-led intervention that compares exercise interventions should be developed and tested". All 3 previous RCTs that are referenced compared physiotherapy-led exercise based treatments (that offered mixed programmes comprising strengthening and stretching as far as I can tell) versus arthroscopic surgery - which is of little relevance in terms of providing a rationale for the current trial (focused as it is on comparing two packages of physiotherapy-led care with much similar content and some different features of the exercise content). I also don't follow the point about sufficient dosage in the previous RCTs exercise interventions, since the current trial is not actually comparing different dosages in each arm of the trial (in fact participants will have the same number of sessions over same timeframe etc). The key differences in the two interventions in the current trial centre on i) individualisation versus standardisation and ii) strengthening exercise versus stretching exercise. The team are comparing two packages of physio-led care that have largely similar content except for one offers individualised (they use the term targeted) strengthening and the other offers standardised stretching. Therefore I believe the rationale for the current trial needs to be more clearly focused on providing the reader with an explanation of why these TWO key features of exercise based interventions are the ones these authors feel are important to focus on. And why mix the two in the one trial? (other options for comparisons would have included individualised strengthening v individualised stretching - to tease out the difference in type of exercise (should we focus on strengthening or stretching?) assuming individualisation is important; and there were several other options for the comparisons). Its not clear why the team are focusing on this particular comparison. - the second rationale given relates to the recent consensus recommendations - "A recent consensus meeting recommended exercise-based intervention as the first-line treatment for young adults with hip-related pain. However, absence of a full-scale RCT comparing the head-to-head effectiveness of two exercise-based,

	physiotherapist-led interventions for FAI syndrome limited the strength of such recommendations. Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching". Again I didn't follow this rationale. If there is uncertainty if the first-line treatment for FAI should involve exercise-based treatment then one would need a RCT comparing exercise-based treatment with NO exercise-based treatment to provide evidence . Yet the authors have designed a RCT comparing two packages of exercise based treatment (with some different features of the exercise - focusing on individualising or standardising, and on strengthening or stretching). Would it make more sense to provide a rationale along the lines of - first line care has been recommended to be exercise-based but we don't know whether some features or characteristics of exercise for FAI are superior to others, in particular whether exercise programme should focus on strengthening or stretching, or whether exercise should be individualised or standardised? To me, this suggested rationale underpins this particular trial more appropriately. I believe that the rationale for this RCT needs revised, so that it is much clearer why this trial is comparing the interventions that it is. I think the team can make this further minor revision - focusing in on key features of exercise programmes that they are manipulating in the trial - the hypothesis that individualisation will yield better outcomes than standardisation of exercise, and the hypothesis that strengthening exercise will yield better outcomes than stretching for FAI. This would require reworking of a couple of paragraphs in the background section. 2. The authors have now clarified there are two primary outcomes (iHOT-33 and global rating of change) and this matches the trial registry information. However, they then also clarify that significant between arm difference in EITHER of these two primary outcomes will result in a conclusion of efficacy (I assume they mean comparative superiority), and so it would be helpful to ensure the language throughout the text matches this. For example, the text states "Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching in 164 participants with FAI syndrome on hip-related QOL (International Hip Outcome Tool 33 (iHOT-33)) and patient-perceived global improvement at 6-months." This should read 2..." or patient-perceived global improvement..." not "and". It would also be helpful to explain how the 7 point global rating of change will be grouped for analysis - will it be dichotomised into better/not better? Minor points: - the trial flow chart needs to use the same language for the two arms of the trial as the rest of the manuscript, it currently still refers to physiotherapy or control arms (which are not reflective of the actual content of each of the trial interventions). - the sample size section has been amended to now ensure both primary outcomes are included in the rationale for the sample size required. This is important to do when a trial has two specified primary outcome measures. I have to admit however that I do not completely following the explanation. Its not clear how the global rating of change responses will be analysed/dichotomised for analysis. And if the iHOT-33's MCID ranges between 6 and 10, with a SD of 25, then ending up powering the trial for a between

	group effect size of 0.5 means the trial will be underpowered to detect a between arm difference of even the largest of the MCID range that is published (ie. to detect a between arm difference of 10 in the iHOT-33, with a SD of 25, this trial would need to be powered for an ES of 0.4 not 0.5; if the actual between arm difference is lower, say as low as 6 (with SD of 25), then the trial would need to be powered to detect an ES of 0.24). As I described in my previous review of this protocol and sample size, the trial interventions are very similar with some features of exercise the key differences between the interventions, and thus this trial risks being underpowered to detect between arm differences that are smaller than 'moderate in size based on Cohen's ES' but are nevertheless still clinically important. I appreciate the team cannot change their sample size, but the explanation of it could be made clearer given the two primary outcomes. I hope the above points help to ensure clarity in the protocol paper underpinning this important RCT and congratulate the team on conducting this trial to help inform the content of physiotherapy-led exercise based interventions for FAI.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Nadine Foster, Keele University

Comments to the Author:

This is a revised protocol paper, summarising a RCT that is ongoing. The requested revisions have in large part been made, mainly to make the description of previous RCTs' exercise interventions more accurate, to improve the detail of the RCT methods section, and to address the differences between the detailed protocol paper and the previous trial registry information.

However, the key areas that still need improvement are:

Reviewer comment 1. Clarification for the rationale for this trial - the background section offers two rationales. The first focuses on the limitations of previous RCTs (that all compared physiotherapy-led exercise based treatment versus surgery) and the second seems to be about evidencing recommendations that a recent expert consensus group made that were not based on RCT evidence. I have tried to follow these two rationales and have struggled to see how EITHER lead to the current trial. I therefore believe the rationale for THIS trial needs clearer articulation in the background section. This point is similar to the point made before about being really clear what the aims are of this trial (as a reminder I previously commented "The overall aim of this trial is described throughout various sections of the paper as 'to test the efficacy of a physiotherapy-led intervention for FAI' but actually that is not what this trial is designed to do").

Author response: The aim has been reworded to address Professor Foster's concerns, and reads "This double-blind, randomised controlled trial (RCT) aims to estimate the effect of a physiotherapist-led intervention with targeted strengthening compared to a physiotherapist-led intervention with standardised stretching, on hip-related quality of life (QOL) or perceived improvement at 6-months in people with femoroacetabular impingement (FAI) syndrome." This reworded aim provides clarity about the comparison between the 2 different exercise programmes.

Reviewer comment 2. In the first rationale for the current trial in this revised paper - the text says - "However it is unclear whether the exercises described in these (previous) studies were developed based on contemporary knowledge of impairments in FAI syndrome, or be of sufficient stimulus and dosage to address the deficits in strength and functional performance that exist in these patients. Thus, a physiotherapist-led intervention that compares exercise interventions should be developed and tested". All 3 previous RCTs that are referenced compared physiotherapy-led exercise based treatments (that offered mixed programmes comprising strengthening and stretching as far as I can tell) versus arthroscopic surgery - which is of little relevance in terms of providing a rationale for the current trial (focused as it is on comparing two packages of physiotherapy-led care with much similar content and some different features of the exercise content). I also don't follow the point about sufficient dosage in the previous RCTs exercise interventions, since the current trial is not actually comparing different dosages in each arm of the trial (in fact participants will have the same number of sessions over same timeframe etc). The key differences in the two interventions in the current trial centre on i) individualisation versus standardisation and ii) strengthening exercise versus stretching exercise. The team are comparing two packages of physio-led care that have largely similar content except for one offers individualised (they use the term targeted) strengthening and the other offers standardised stretching. Therefore I believe the rationale for the current trial needs to be more clearly focused on providing the reader with an explanation of why these TWO key features of exercise based interventions are the ones these authors feel are important to focus on. And why mix the two in the one trial? (other options for comparisons would have included individualised strengthening v individualised stretching - to tease out the difference in type of exercise (should we focus on strengthening or stretching?) assuming individualisation is important; and there were several other options for the comparisons). Its not clear why the team are focusing on this particular comparison.

- the second rationale given relates to the recent consensus recommendations - "A recent consensus meeting recommended exercise-based intervention as the first-line treatment for young adults with hip-related pain. However, absence of a full-scale RCT comparing the head-to-head effectiveness of two exercise-based, physiotherapist-led interventions for FAI syndrome limited the strength of such recommendations. Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching". Again I didn't follow this rationale. If there is uncertainty if the first-line treatment for FAI should involve exercise-based treatment then one would need a RCT comparing exercise-based treatment with NO exercise-based treatment to provide evidence. Yet the authors have designed a RCT comparing two packages of exercise based treatment (with some different features of the exercise - focusing on individualising or standardising, and on strengthening or stretching). Would it make more sense to provide a rationale along the lines of - first line care has been recommended to be exercise-based but we don't know whether some features or characteristics of exercise for FAI are superior to others, in particular whether exercise programme should focus on strengthening or stretching, or whether exercise should be individualised or standardised? To me, this suggested rationale underpins this particular trial more appropriately.

I believe that the rationale for this RCT needs revised, so that it is much clearer why this trial is comparing the interventions that it is. I think the team can make this further minor revision - focusing in on key features of exercise programmes that they are manipulating in the trial - the hypothesis that individualisation will yield better outcomes than standardisation of exercise, and the hypothesis that strengthening exercise will yield better outcomes than stretching for FAI. This would require reworking of a couple of paragraphs in the background section.

Author response: Thank you for highlighting your concerns around the rationale of the study. As suggested, we have reworked the second half of the introduction to more appropriately justify the interventions tested.

We have also highlighted that when we designed our interventions, we deliberately sought to compare what we considered “best practice” based on our understanding of impairments (reduced hip muscle strength) against a standardised comparator that would seem credible to participants, to allow for participant blinding and same level of patient-clinician contact between groups. However, we acknowledge that this does not allow us to test whether any between group differences are due to the strengthening components of the programme alone, or the individualised nature of the intervention. We also acknowledge that the differences between the groups may be smaller. We have now described these limitations in the manuscript.

Manuscript changes: The second half of the introduction now reads:

“Treatment options for FAI syndrome can be surgical or non-surgical.[12] Non-surgical approaches are recommended as the first line options for other musculoskeletal pain conditions (evident from clinical guidelines for osteoarthritis,[13] low back pain,[14] and chronic whiplash associated disorders[15]), due to the higher costs and risks associated with surgery. Recently published RCTs comparing hip arthroscopic surgery to physiotherapist-led interventions for FAI syndrome found small[20 21] to moderate[22] between-group differences favouring hip arthroscopy, with a greater cost and risk of adverse events associated with surgery.[20-22] The physiotherapist-led interventions used for comparison to hip arthroscopy involved a diversity of exercise interventions including stretching, motor control, core stability and strengthening, and provided varied detail regarding the individualisation and the content of the exercise interventions. Hence, the specific components of exercise programmes that are effective are not known. A recent consensus meeting recommended individualised, exercise-based interventions as the first-line treatment for young adults with hip-related pain, however no recommendation was made regarding one type of exercise over another.[12] Such a recommendation could not be provided because of the absence of a full-scale RCT comparing the head-to-head effectiveness of different exercise-based, physiotherapist-led interventions for FAI syndrome[23-25] . Thus, a physiotherapist-led intervention that compares exercise interventions needs to be developed and tested.”

We have also added the following section to the limitations section of the manuscript.

“When we developed the two intervention groups, we deliberately sought to compare what we considered “best practice” based on our understanding of impairments (reduced strength) against a standardised comparator that would seem credible to participants, to allow for participant blinding and same level of patient-clinician contact between groups. However, this does not allow us to test whether any between group differences are due to the different exercise components of the programme (strength v stretch), or to the nature of the interventions (individualised v standard), and this would need to be explored in future studies.”

Reviewer comment 3. The authors have now clarified there are two primary outcomes (iHOT-33 and global rating of change) and this matches the trial registry information. However, they then also clarify that significant between arm difference in EITHER of these two primary outcomes will result in a conclusion of efficacy (I assume they mean comparative superiority), and so it would be helpful to ensure the language throughout the text matches this. For example, the text states "Therefore, the primary aim of this RCT is to compare effectiveness of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching in 164 participants with FAI syndrome on hip-related QOL (International Hip Outcome Tool 33 (iHOT-33)) and patient-perceived global improvement at 6-months." This should read 2..." or patient-perceived global improvement..." not "and".

Author response: Thank you for highlighting this inconsistency. We have amended the text throughout the manuscript to accurately reflect the comparative superiority of the primary outcomes.

Manuscript changes: In abstract and at the end of the background section “Therefore, the primary aim of this RCT is to estimate the effect of a physiotherapist-led intervention with targeted strengthening to a physiotherapist-led intervention with standardised stretching in 164 participants with FAI syndrome on hip-related QOL (International Hip Outcome Tool 33 (iHOT-33)) or patient-perceived global improvement at 6-months28. We hypothesise that, compared to the standardised stretching physiotherapist-led intervention, the targeted strengthening physiotherapist-led intervention will result in greater improvement in: (i) hip-related QOL or (ii) perceived improvement.”

Reviewer comment 4. It would also be helpful to explain how the 7-point global rating of change will be grouped for analysis - will it be dichotomised into better/not better?

Author response: Thank you, we realise that we did not explain this as clearly as we should have done. We have provided additional information in the methods section of the manuscript to clarify that the 7-point global rating of change will not be dichotomised, but used as a continuous scale.

Manuscript changes: In the outcomes section, “For the analysis, patient-perceived global improvement will be used as a continuous scale.”

In the statistical analysis section, we have provided further information.

“For the primary analysis, patient-perceived global improvement will be assessed as a 7-point scale, with bootstrapped standard errors to account for non-normality of residuals. A secondary analysis will assess the between-group difference in the proportion of participants reporting being ‘much improved’ or ‘improved’, as an indicator of successful treatment outcome.”

Minor points:

Reviewer comment 5. The trial flow chart needs to use the same language for the two arms of the trial as the rest of the manuscript, it currently still refers to physiotherapy or control arms (which are not reflective of the actual content of each of the trial interventions).

Author response: Thank you for highlighting this inconsistency, we have modified the flow chart for the two arms of the trial to reflect the manuscript.

Reviewer comment 6. The sample size section has been amended to now ensure both primary outcomes are included in the rationale for the sample size required. This is important to do when a trial has two specified primary outcome measures. I have to admit however that I do not completely following the explanation. It is not clear how the global rating of change responses will be analysed/dichotomised for analysis. And if the iHOT-33's MCID ranges between 6 and 10, with a SD of 25, then ending up powering the trial for a between group effect size of 0.5 means the trial will be underpowered to detect a between arm difference of even the largest of the MCID range that is published (ie. to detect a between arm difference of 10 in the iHOT-33, with a SD of 25, this trial would need to be powered for an ES of 0.4 not 0.5; if the actual between arm difference is lower, say as low as 6 (with SD of 25), then the trial would need to be powered to detect an ES of 0.24). As I described in my previous review of this protocol and sample size, the trial interventions are very similar with some features of exercise the key differences between the interventions, and thus this trial risks being underpowered to detect between arm differences that are smaller than 'moderate in size based on Cohen's ES' but are nevertheless still clinically important. I appreciate the team cannot change their sample size, but the explanation of it could be made clearer given the two primary outcomes.

Author response: We appreciate Professor Foster's concerns about our sample size, which reflect our own discussions about the uncertainty surrounding the clinically important difference for the iHOT-33 in a non-surgical patient group. The MCID of the iHOT-33 has only been estimated in hip arthroscopy cohorts and in the context of within group change and not between group difference, anchored to a global rating of change scale by 2 studies. We acknowledge that our target effect size (SMD=0.5) would represent a larger between group difference than the lower bound of the previously reported MCID for iHOT-33 of 6 points. We believe that it is important to aim for an effect size that would represent the higher end of the previously reported MCID for iHOT-33 of 10 points. While our two interventions do contain some similar elements, our pilot trial indicated we could potentially expect larger differences than 6 points between treatment groups. Therefore, we powered the study for an effect size of SMD=0.50, because a moderate effect would be considered clinically meaningful. To achieve this effect size, with a SD of 25 (estimated from our pilot study), we would need to be able to detect a between group MD of 12.5. Our pilot data indicates that a MD of 12.5 is possible, with MD of 15 observed.

Manuscript changes: We have provided more detail in the sample size section of the manuscript explain/clarify this point and to reflect the uncertainty around the iHOT-33 MCID and the similarities between the interventions. We also added an additional limitations section in discussion to acknowledge that this means we may miss a smaller, but potentially clinically significant difference between groups.

"A power calculation was conducted for this RCT, informed by data from our previous pilot study that utilised and compared a similar tailored strengthening intervention to a standardised stretching intervention.[6] The MCID of the iHOT-33 is still uncertain in non-surgical patients with FAI syndrome and has only been estimated in hip arthroscopy cohorts.[39 40] Therefore, the power calculation was based on the observed baseline standard deviation (SD) and the between-group differences in the scores of our first primary outcome measure (hip-related QOL (iHOT-33)) from our pilot study (baseline SD = 25 points; mean difference 15 points out of 100)[6] , which exceeded the previously reported MCID of 6-10 points[40]. Our pilot trial[6] observed a standardised mean difference (SMD) of 0.68 for the iHOT-33. We reduced the proposed SMD to 0.50 for this study to account for the small sample (n=24) in the pilot study, the similarities between the interventions and the difference in the expertise of treating physiotherapists in a full-scale study. This is consistent with previously reported between-group SMD for the second primary outcome (patient-perceived global improvement) of 0.50.[60] Estimated sample sizes for a two-sample means test t test assuming 80% power, α=0.025 (accounting for both primary outcomes), results in a sample size estimate of 156 participants. To account for an estimated 5% drop-out due to the study duration, a recommended sample size of 164 participants (82 in each group) will be recruited in this RCT."

"**Limitations**

We acknowledge that our target effect size (SMD=0.5) might represent a larger between group difference than the lower bound of the previously reported between group difference (for example the lower end of the previously reported MCID for iHOT-33 of 6 points)[36]. Therefore, we powered the study for an effect size of SMD=0.50, because a moderate effect would be considered clinically meaningful. While our two interventions do contain some similar elements, our pilot trial indicated we could potentially expect larger differences than 6 points between treatment groups[6]."

I hope the above points help to ensure clarity in the protocol paper underpinning this important RCT and congratulate the team on conducting this trial to help inform the content of physiotherapy-led exercise-based interventions for FAI.

Author response: We wish to thank Professor Foster once again for her very constructive and insightful feedback.

VERSION 3 – REVIEW

REVIEWER	Nadine Foster University of Queensland Australia None, I have reviewed this paper several times already and provided detailed feedback previously - which the authors have responded to
REVIEW RETURNED	15-Mar-2021

GENERAL COMMENTS	I have already reviewed previous versions of this paper, and the authors have responded to those and amended the paper accordingly.