STATISTICAL ANALYSIS PLAN (SAP) FOR:

Comparing Standard Care Approaches for Greater Trochanteric Pain Syndrome in General Practice: A Randomised Clinical Trial

Version: 1.0

Date: 15-05-2026

Protocol version: Protocol version 2, 14-04-2026

Trial registration ID: H-26000025

Contents

1. Study Synopsis 3
2. Study Objectives and Outcomes 3
2.1. Primary Objective and Outcome 5
2.2. Secondary Objectives and Outcomes 5
2.3. Economic Evaluation 6
2.4. Descriptive Outcomes 6
2.5. Specification of Endpoints 6
2.5.1. Primary Endpoint 6
2.5.2. Secondary Endpoints 6
3. Study Design 7
3.1. Sample Size 7
3.2. Randomization and Blinding 8
4. Study Population 8
4.1. Subject Disposition 8
5. Statistical Analysis 8
5.1. Primary Endpoint 8
5.2. Secondary Endpoints 9
5.3. Sensitivity Analyses 9
5.4. Descriptive Outcomes 9
6. Implementation of Analysis Plan 10
7. References 11

1. Study Synopsis

Lateral hip pain or greater trochanteric pain syndrome (GTPS) is a common pain condition presenting on the outside of the hip. GTPS most commonly affects middle-aged women, with a prevalence of 15% in women aged 50-70 (1). The cause of this condition is tendinopathy of the gluteal tendons with or without involvement of the trochanteric bursa, resulting in pain while walking, sitting, stair climbing and lying on the side (2–4). Hence, GTPS interferes with many daily activities and affects quality of life to a degree similar to that of late-stage hip osteoarthritis (4). A quarter of GTPS patients will experience sick leave due to pain, and the prognosis is poor with 63% of patients in primary care still reporting pain at 5-year follow-up after various treatments (5). At present, there is no protocol for the treatment of GTPS in primary care, and there is a lack of evidence on which treatment approach is the most effective.

This statistical analysis plan (SAP) describes the planned analyses for a pragmatic, multicentre, three-armed parallel group randomised superiority trial comparing three treatment strategies for GTPS in primary care. Participants are randomised in a 1:1:1 ratio to one of three groups: (1) patient advice, consisting of a patient information leaflet; (2) patient advice plus a home exercise programme; or (3) patient advice plus a home exercise programme and a corticosteroid injection. Randomisation will be stratified by site to ensure a balanced distribution of site-specific factors across treatment arms. The primary outcome is global perceived change since baseline measured by the Global Rating of Change Scale (GROC) at 3 months. Follow-up is conducted at 1 month, 6 weeks, 2 months, 3 months, 6 months, and 12 months. A full description of the study methods is provided in the trial protocol.

2. Study Objectives and Outcomes

The purpose of this trial is to investigate the effectiveness of patient advice, versus patient advice plus a home exercise programme, versus patient advice plus a home exercise programme and a corticosteroid injection on global improvement at 3 months in patients with GTPS. The main hypotheses are:

1. Patient advice plus exercises plus CSI will be superior to patient advice alone.
2. Patient advice plus exercises will be superior to patient advice alone.

All outcomes will be obtained from all participants at baseline and at follow-ups at 1 month, 6 weeks, 2 months, 3 months, 6 months, and 12 months, as specified in Table 1. Baseline demographic and clinical characteristics are described in Table 2 (section 2.4).

Table 1. Overview of outcomes, data types, and planned analysis methods.					
Outcome	**Data type**	**Scale**	**Notes**	**Timepoints**	**Analysis method**
Primary					
GROC	Dichotomous	Categories 1-7	Dichotomisation: Categories 6-7 = success; categories 1-5 = no success	3 months (primary); 1, 2, 6, 12 months (secondary)	Binomial regression, log-link, adjusted for baseline NRS and sex
Secondary					
NRS average pain	Interval	0-10; higher = more pain; MCID: -1.5	-	1, 2, 3, 6, 12 months	Linear mixed-effects model, adjusted for baseline NRS and sex
NRS night pain	Interval	0-10; higher = more pain	-	1, 2, 3, 6, 12 months	
NRS % time in pain	Ratio	0-100%; higher = more pain	-	1, 2, 3, 6, 12 months	
VISA-G	Interval	0-100; higher = better function	-	1, 2, 3, 6, 12 months	
PSFS	Interval	0-10; higher = better function	Mean of scores for 1-5 self-selected activities	1, 2, 3, 6, 12 months	
PSEQ	Interval	0-60; higher = better self-efficacy	-	1, 2, 3, 6, 12 months	
EQ-5D-5L VAS	Interval	0-100; higher = better health	-	1, 2, 3, 6, 12 months	
EQ-5D-5L utility score	Interval	0,00-1,00; higher = better health	Calculated using Danish population preference weights	1, 2, 3, 6, 12 months	
EQ-5D-5L items	Ordinal	Level 1-5 per domain	-	1, 2, 3, 6, 12 months	Descriptive
IPAQ	Ratio	MET-min/week	MET-min/week = MET level × minutes × days/week; categorical score: low (<600), moderate (600-3000), high (>3000) MET-min/week	1, 2, 3, 6, 12 months	Median and IQR; linear mixed-effects model if normally distributed
PASS status	Dichotomous	Yes/no	-	1, 2, 3, 6, 12 months	Binomial regression, log-link and identity link
PASS time to event	Ratio	Days	Time from randomisation to first PASS achieved	-	Kaplan-Meier and Cox regression
Adverse events	Ordinal	Grade 1-5 (CTCAE v5.0)	Incidence by grade and causality	-	Descriptive

2.1. Primary Objective and Outcome

The primary outcome measure is Global Rating of Change Score (GROC), which is a 7-point numerical rating scale anchored by ‘very much worse’ and ‘very much better’, on which participants rate their perceived overall change in their hip condition from baseline to the respective follow-up (6). Participants will be asked: “How is your lateral hip pain now compared with when you first came in for treatment in this project?” The scale includes seven response categories (‘very much worse’, ‘much worse’, ‘a little worse’, ‘no change’, ‘a little better’, ‘much better’, and ‘very much better’). Participants will be dichotomised as improved if they report being ‘much better’ or ‘ very much better’, and as not improved if they select any of the remaining categories. The GROC scale has previously been shown to be clinically relevant for interpreting meaningful improvements from the patient perspective (7).

2.2. Secondary Objectives and Outcomes

The key secondary outcome is self-reported average hip pain over the previous week, measured using an 11-point Pain Numeric Rating Scale (NRS, 0–10; 0 = 'no pain', 10 = 'worst pain imaginable'). The minimal clinically important difference (MCID) for the NRS is 1.5 points for musculoskeletal disorders (8).

Other secondary outcomes include:

Pain:

- Self-reported percentage of time pain has been present during the last week (0-100%; 0% = no pain, 100% = constant pain).
- Self-reported worst experienced night pain during the last week (0-10 NRS).

Function and disability:

- Self-reported disability measured by the Victorian Institute of Sports Assessment for gluteal tendons (VISA-G), a validated tool for evaluating disability in gluteal tendinopathy (scores 0–100; higher = better function). A Danish validated version will be used (9).
- Functional level measured by the Patient Specific Functional Scale (PSFS) (scores 0–10; higher = better function) (10).

Psychological measures:

- Pain self-efficacy measured by the Pain Self-Efficacy Questionnaire (PSEQ) (scores 0–60; higher = better self-efficacy) (11). A Danish validated version will be used (12).

Quality of life:

- Health-related quality of life measured by the EQ-5D-5L (13,14). The Danish translation of the EQ-5D-5L will be used, with utility scores calculated using Danish population preference weights (15).

Physical activity:

- Physical activity measured by the International Physical Activity Questionnaire (IPAQ) (16).

Symptom satisfaction:

- Patient Acceptable Symptom State (PASS), assessed by asking participants: "Is your lateral hip pain at an acceptable level that does not require any further treatment?" (yes/no) (17). Time to first PASS will be recorded, along with any subsequent loss of PASS status.

Adverse events:

- Adverse events will be graded according to the Common Terminology Criteria for Adverse Events v5.0 (CTCAE) (18).

2.3. Economic Evaluation

A health economic evaluation will be conducted following international guidelines (19,20). Health benefits will be measured using the EQ-5D-5L and expressed as quality-adjusted life years (QALYs) calculated using Danish population preference weights (15,21). Healthcare utilization, productivity losses, and out-of-pocket expenses will be captured using a modified version of the Oslo Cost Questionnaire (OCC-Q) at each follow-up. Costs will be summarized at both the societal and healthcare system level. A detailed statistical analysis plan for the health economic evaluation will be developed separately in collaboration with a health economist.

2.4. Descriptive Outcomes

Baseline characteristics will be presented in a table. A full overview can be seen in table 2.

2.5. Specification of Endpoints

2.5.1. Primary Endpoint

The primary endpoint is global perceived improvement at 3 months as measured by the GROC.

2.5.2. Secondary Endpoints

The secondary endpoints are the changes from baseline to 1 month, 2 months, 6 months, and 12 months. Baseline measures are collected at Day 0 prior to randomisation and serve as covariates in all analyses.

Table 2. Specification of demographics and clinical characteristics to be collected at baseline				
Category	**Variable**	**Data type**	**Notes and specifications**	**Presentation**
Demographics				
	Age	Ratio	Years	Mean (SD) or median (IQR)
	Sex	Nominal	Female/male/other	n (%)
	BMI	Ratio	kg/m², calculated from self-reported height and weight	Mean (SD) or median (IQR)
	Employment status	Nominal		n (%)
Clinical characteristics				
	Pain duration	Ratio	Months	Median (IQR)
	Recruitment site	Nominal	Study site 1/study site 2	n (%)
	Pain location,	Dichotomous	Unilateral/bilateral	
	Previous healthcare contact	Nominal	Multiple responses possible	
	Previous treatment type	Nominal	Multiple responses possible	
Baseline outcome scores				
	NRS, average pain	Interval	0-10; higher = more pain	Mean (SD) or median (IQR)
	NRS, night pain	Interval	0-10; higher = more pain	
	NRS % time in pain	Ratio	0-100%; higher = more pain	
	VISA-G	Interval	0-100; higher = better function	
	PSFS	Continuous	0-10; higher = better function	
	PSEQ	Interval	0-60; higher = better self-efficacy	
	EQ-5D-5L VAS	Interval	0-100; higher = better health	
	EQ-5D-5L utility score	Interval	0-1; higher = better health	
	IPAQ	Ratio	MET-min/week	
	PASS	Dichotomous	Yes/no	n (%)

3. Study Design

3.1. Sample Size

The sample size primary was calculated for the primary analysis of the Global Rating of Change (GROC) at 3 months, using results from the LEAP trial (22). The calculation was based on Dunnett's test to adjust for the two planned comparisons against group 1 (patient advice). A success rate of 35% was assumed in group 1, 65% in group 2 (patient advice plus exercises), and 75% in group 3 (patient advice plus exercises and CSI). The sample size is determined by the comparison yielding the smallest expected difference (control vs. education plus exercise). Allowing for a familywise type I error rate of 0.05 and a power of 85%, this comparison requires 54 analysable participants per group. After accounting for an anticipated 15% loss to follow up, 64 participants per group must be enrolled, resulting in a total sample of 192 participants.

3.2. Randomization and Blinding

Participants will be randomly allocated to one of three arms after inclusion using REDCap's randomization module. Randomization will be performed in blocks to ensure balanced group sizes throughout the recruitment period. The randomization procedure, including preparation of the allocation sequence, will be performed by a colleague not otherwise involved in the project to ensure allocation concealment. The randomization will be stratified by site to ensure a balanced distribution across the treatment arms. The allocation will remain concealed from those responsible for including participants to minimize risk of bias.

The person responsible for performing the statistical analyses will remain blinded to the coding until after the analyses have been performed and only the person delivering the interventions will know which intervention is linked to which group number (group 1, 2 or 3).

4. Study Population

4.1. Subject Disposition

Participants will be recruited from Bispebjerg sports medicine department, general practice clinics, private rheumatology/orthopaedic clinics, social media, and advertising at social events. Any patient who presents with self-reported lateral hip pain will be considered for inclusion. A full description of inclusion and exclusion criteria is provided in the trial protocol.

5. Statistical Analysis

All statistical analyses will be performed by a blinded member of the project group according to a pre-established analysis plan using the intention-to-treat principle. Reporting will adhere to CONSORT guidelines (23). This plan is written in consultation with a statistician and will be published before inclusion of the first participant. SPSS software (IBM Corporation, Armonk, NY, USA) and STATA will be used as statistical software. Missing data will be handled using multiple imputation combined using Rubin's rules (24). The number of imputed datasets will be determined by the percentage of incomplete cases, following the rule of thumb that the number of imputations should be at least equal to the percentage of incomplete cases (25). For example, if 20% of cases are incomplete, at least 20 datasets will be imputed. Data are assumed to be missing at random. Imputation will be performed separately within each treatment group.

5.1. Primary Endpoint

The primary analysis will investigate the between-group difference in the proportion of participants reporting improvement on the GROC at 3 months. Participants will be dichotomised as improved if they report 'much better' or 'very much better' (GROC categories 6–7), and as not improved if they select any of the remaining categories (GROC categories 1–5. Two primary comparisons will be made: group 1 (patient advice) versus group 2 (patient

advice plus exercises), and group 1 versus group 3 (patient advice plus exercises and CSI). To control the familywise type I error rate at 0.05 across these two primary comparisons, Dunnett-adjusted contrasts will be used. Between-group differences will be estimated using binomial regression models with a log link and with an identity link to calculate absolute risk differences, both with 95% confidence intervals, adjusting for baseline NRS and sex. Comparisons between group 2 and 3 are considered exploratory and will be presented descriptively.

5.2. Secondary Endpoints

Continuous secondary outcomes, including NRS pain, VISA-G, PSFS, PSEQ, EQ-5D-5L, and IPAQ, will be analysed using linear mixed-effects models with random intercepts and slopes for each participant. Fixed effects will include time (1, 2, 3, 6, and 12 months), group allocation, baseline value of the respective outcome, the interaction between time and group, and sex. Results will be presented as mean (SD) or median (IQR) based on assessment of normality using Q-Q plots and histograms. IPAQ will be summarised using median and IQR due to expected skewness in physical activity data (26) and will not be included in the linear mixed-effects model.

Dichotomous secondary outcomes, including Patient Acceptable Symptom State (PASS) status at each follow-up and loss of PASS (defined as change from PASS to no longer achieving PASS at a subsequent follow-up), will be analysed using binomial regression models with a log link to estimate risk ratios and with an identity link to estimate risk differences, adjusting for baseline NRS and study site. Numbers needed to treat (NNT) will be derived from these regression-adjusted risk differences. Time-to-event analyses for achieving PASS will be conducted using Kaplan-Meier survival curves and Cox regression, considering only the time to first PASS achieved.

5.3. Sensitivity Analyses

Sensitivity analyses will include per-protocol analyses and best-/worst-case scenarios to assess the robustness of findings under different assumptions for missing data. No subgroup analyses are planned.

5.4. Descriptive Outcomes

Baseline demographic and clinical characteristics will be summarised descriptively for all randomised participants by treatment group. Continuous variables will be presented as mean (SD) or median (IQR) based on assessment of normality using Q-Q plots and histograms. Categorical variables will be presented as n (%). No formal inferential analysis will be performed on baseline characteristics.

6. Implementation of Analysis Plan

All analyses will be performed by a blinded member of the project group, and none of the investigators involved in delivering the interventions will perform any of the statistical analyses. The implementation of the SAP will be as follows:

1. The database manager will code each treatment arm into 'group 1', 'group 2' and 'group 3', leaving the group member blinded to treatment allocation during analyses.
2. Blinded data will be delivered to the group member for analysis.
3. Primary and secondary endpoint analyses will be performed blinded from treatment allocation.
4. Results will be presented to the project group where any uncertainties will be clarified prior to unblinding of data.

7. References

1. Segal NA, Felson DT, Torner JC, Zhu Y, Curtis JR, Niu J, et al. Greater trochanteric pain syndrome: epidemiology and associated factors. Arch Phys Med Rehabil. 2007 Aug;88(8):988–92. doi:10.1016/j.apmr.2007.04.014 PubMed PMID: 17678660; PubMed Central PMCID: PMC2907104.

2. Bird PA, Oakley SP, Shnier R, Kirkham BW. Prospective evaluation of magnetic resonance imaging and physical examination findings in patients with greater trochanteric pain syndrome. Arthritis Rheum. 2001 Sep;44(9):2138–45. doi:10.1002/1529-0131(200109)44:9<2138::AID-ART367>3.0.CO;2-M PubMed PMID: 11592379.

3. Kong A, Van der Vliet A, Zadow S. MRI and US of gluteal tendinopathy in greater trochanteric pain syndrome. Eur Radiol. 2007 Jul;17(7):1772–83. doi:10.1007/s00330-006-0485-x PubMed PMID: 17149624.

4. Fearon AM, Cook JL, Scarvell JM, Neeman T, Cormick W, Smith PN. Greater trochanteric pain syndrome negatively affects work, physical activity and quality of life: a case control study. J Arthroplasty. 2014 Feb;29(2):383–6. doi:10.1016/j.arth.2012.10.016 PubMed PMID: 24210307.

5. Lievense A, Bierma-Zeinstra S, Schouten B, Bohnen A, Verhaar J, Koes B. Prognosis of trochanteric pain in primary care. Br J Gen Pract J R Coll Gen Pract. 2005 Mar;55(512):199–204. PubMed PMID: 15808035; PubMed Central PMCID: PMC1463090.

6. Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. J Man Manip Ther. 2009;17(3):163–70. doi:10.1179/jmt.2009.17.3.163 PubMed PMID: 20046623; PubMed Central PMCID: PMC2762832.

7. ten Klooster PM, Drossaers-Bakker KW, Taal E, van de Laar MAFJ. Patient-perceived satisfactory improvement (PPSI): interpreting meaningful change in pain from the patient's perspective. Pain. 2006 Mar;121(1–2):151–7. doi:10.1016/j.pain.2005.12.021 PubMed PMID: 16472915.

8. Abbott JH, Schmitt J. Minimum important differences for the patient-specific functional scale, 4 region-specific outcome measures, and the numeric pain rating scale. J Orthop Sports Phys Ther. 2014 Aug;44(8):560–4. doi:10.2519/jospt.2014.5248 PubMed PMID: 24828475.

9. Jorgensen JE, Fearon AM, Mølgaard CM, Kristinsson J, Andreasen J. Translation, validation and test–retest reliability of the VISA-G patient-reported outcome tool into Danish (VISA-G.DK). PeerJ. 2020 Mar 4;8:e8724. doi:10.7717/peerj.8724

10. Horn KK, Jennings S, Richardson G, Vliet DV, Hefford C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. J Orthop Sports Phys Ther. 2012 Jan;42(1):30–42. doi:10.2519/jospt.2012.3727 PubMed PMID: 22031594.

11. Nicholas MK. The pain self-efficacy questionnaire: Taking pain into account. Eur J Pain Lond Engl. 2007 Feb;11(2):153–63. doi:10.1016/j.ejpain.2005.12.008 PubMed PMID: 16446108.

12. Rasmussen MU, Rydahl-Hansen S, Amris K, Samsøe BD, Mortensen EL. The adaptation of a Danish version of the Pain Self-Efficacy Questionnaire: reliability and construct validity in a population of patients with fibromyalgia in Denmark. Scand J Caring Sci. 2016 Mar;30(1):202–10. doi:10.1111/scs.12232 PubMed PMID: 25924021.

13. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2011 Dec;20(10):1727–36. doi:10.1007/s11136-011-9903-x PubMed PMID: 21479777; PubMed Central PMCID: PMC3220807.

14. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021 Mar 1;30(3):647–73. doi:10.1007/s11136-020-02688-y

15. Sørensen J, Davidsen M, Gudex C, Pedersen KM, Brønnum-Hansen H. Danish EQ-5D population norms. Scand J Public Health. 2009 Jul;37(5):467–74. doi:10.1177/1403494809105286 PubMed PMID: 19535407.

16. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003 Aug;35(8):1381–95. doi:10.1249/01.MSS.0000078924.61453.FB PubMed PMID: 12900694.

17. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis. 2007 Nov;66 Suppl 3(Suppl 3):iii40-41. doi:10.1136/ard.2007.079798 PubMed PMID: 17934093; PubMed Central PMCID: PMC2095292.

18. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0 [Internet]. Bethesda, MD: National Cancer Institute; 2017. Available from: https://dctd.cancer.gov/research/ctep-trials/trial-development

19. Drummond M. Methods for the economic evaluation of health care programmes. Fourth edition. Oxford, United Kingdom ; New York, NY, USA: Oxford University Press; 2015. 445 p. (Oxford medical publications).

20. Ehlers LH, Vestergaard AS. Costing in health economic evaluation: theory and practice. 1. edition. Aalborg: Aalborg University Press; 2019. 86 p.

21. Jensen CE, Riis A, Petersen KD, Jensen MB, Pedersen KM. Economic evaluation of an implementation strategy for the management of low back pain in general practice. Pain. 2017 May;158(5):891–9. doi:10.1097/j.pain.0000000000000851 PubMed PMID: 28114182.

22. Mellor R, Bennell K, Grimaldi A, Nicolson P, Kasza J, Hodges P, et al. Education plus exercise versus corticosteroid injection use versus a wait and see approach on global outcome and pain from gluteal tendinopathy: prospective, single blinded, randomised

clinical trial. Br J Sports Med. 2018 Nov;52(22):1464–72. doi:10.1136/bjsports-2018-k1662rep PubMed PMID: 30385462.

23. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010 Mar 23;340:c332. doi:10.1136/bmj.c332 PubMed PMID: 20332509; PubMed Central PMCID: PMC2844940.

24. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987. 1 p. (Wiley series in probability and mathematical statistics Applied probability and statistics). doi:10.1002/9780470316696

25. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice [Internet]. doi:10.1002/sim.4067

26. Sjostrom M, Ainsworth B, Bauman A, Bull F, Hamilton-Craig C, Sallis J. Guidelines for data processing analysis of the International Physical Activity Questionnaire (IPAQ) - Short and long forms [Internet]. 2005 [cited 2026 Apr 24]. Available from: https://www.semanticscholar.org/paper/Guidelines-for-data-processing-analysis-of-the-and-Sjostrom-Ainsworth/efb9575f5c957b73c640f00950982e618e31a7be