Shoulder Surgery vs Nonoperative Treatment for Shoulder Pain and Disability in Patients With Imaging-Verified Pathology: Statistical Analysis Plan for an Emulated Target Trial from the Prospective Vejle Hospital Shoulder Cohort

Kim Gordon Ingwersen[1, 2, 3], Robin Christensen[3, 4], Tobias Haugegaard[3, 4], Tine Dimon[5], Jette Vobbe[5], Anne Kristensen[5], Sven Thingholm[1], Jean-Sébastien Roy[6, 7], David Høyrup Christiansen[8, 9]

Affiliations:

[1] Research Unit in Physiotherapy and Occupational Therapy, University Hospital of Southern Denmark – Lillebaelt Hospital – Vejle Hospital,Vejle, Denmark.
[2] Department of Regional Health Research, University of Southern Denmark, 5000 Odense, Denmark
[3] Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark
[4] Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.
[5] Orthopaedic Department, Shoulder Unit, University Hospital of Southern Denmark – Lillebaelt Hospital – Vejle Hospital, Vejle, Denmark.
[6] Center for interdisciplinary Research in Rehabilitation and Social Integration, Quebec City, Quebec, Canada
[7] School of Rehabilitation Sciences, Faculty of Medicine, Université Laval, Quebec City, Quebec, Canada
[8] University Clinic for Interdisciplinary Orthopaedic Pathways, Elective Surgery Center, Silkeborg Regional Hospital, Silkeborg, Denmark
[9] Department of Clinical Medicine, Health, Aarhus University, Aarhus, Denmark

Principal Investigator: Kim Gordon Ingwersen (KGI), Department of Physio- and Occupational therapy, University Hospital of Southern Denmark – Lillebaelt Hospital – Vejle Hospital, Denmark and Department of Regional Health Research, University of Southern Denmark, Odense, Denmark. Kim.Ingwersen@rsyd.dk

Senior Biostatistician and Analysis Support: Robin Christensen (RC), Professor of Biostatistics & Clinical Epidemiology: Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark and Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense Denmark

Statistical Analyst: Tobias Haugegaard, MSc, Biostatistics fellow.

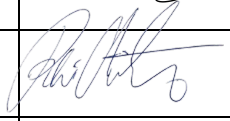
Database Manager: David Hass (DH), Open Patient data Explorative Network (OPEN), Odense University Hospital, Region of Southern Denmark

Trial Registration Number: Clinicaltrials.gov (NCT04686435).

STATISTICAL ANALYSIS PLAN (SAP) REVISION HISTORY

SAP version	Section changed	Description of and reasons for change	Date changed
1.0			**August-15th-2025**

SIGNATURES

Date	Name	Roles and Responsibility	Signature
15-08-2025	Kim Gordon Ingwersen	SAP first author and Chief Investigator	
Friday, 15 August 2025	Robin Christensen	Senior Biostatistician	
Friday August 15 2025	Tobias Haugegaard	Statistical analyst	
17-08-2025	David Høyrup Christiansen	Senior Investigator	

Background and rationale

Shoulder pain is the third most prevalent musculoskeletal diagnosis, affecting up to 55% of the population yearly.[1] Annually, it is estimated that around 2-3% of adults visit their general practitioner due to shoulder pain,[2 3] making it a frequent cause for referral to the secondary healthcare services.[4-6] Conservative treatment strategy is considered the first line treatment for most shoulder disorders, and several studies have shown that conservative treatments lead to moderate to large effects sizes.[7] However, other studies have suggested that both conservative and surgical interventions produce little benefit over placebo or wait and see strategies[8 9]or found no difference in effect between exercise therapy and surgery.[10-14]

These findings from randomised trials, suggest that surgery should not routinely be performed in patients with shoulder pain and disability. However, in daily clinic, surgeons argue that they are making a pragmatic decision whether to recommend non-surgical interventions or surgery based on patient history, clinical tests, image findings and recommendations from clinical guidelines.[15] They also argue that application of current randomised trials into clinical practice are questionable as external validity are hampered due to recruitment bias or selective inclusion.[16 17] Ideally, to determine whether surgery is superior, an well designed and properly conducted randomised trial, including all eligible participants, should be performed. However, performing an efficacy trial, randomly assigning patients to surgery or non-surgical intervention would always require the patient's willingness and therefore be affected by recruitment bias. In a well-designed randomised trial, confounding is minimised through random allocation, ensuring that both known and unknown confounders are equally distributed between groups.[18] This process establishes exchangeability, meaning that any differences in outcomes can be attributed to the intervention rather than external factors. By preventing selection bias and confounding by indication, randomisation ensures temporal precedence and balances both measured and unmeasured variables. This strengthens causal inference by allowing valid statistical testing without the need for extensive adjustments. However, small sample sizes, non-adherence, or differential loss to follow-up can introduce bias if not properly addressed.

Assessing comparative effectiveness from observational data is necessary for several reasons, particularly when randomised trials are limited in their applicability or availability.[19] First, RCTs may lack external validity, meaning their findings may not be directly transferable to real-world settings due to differences in populations, interventions, or healthcare contexts. Observational studies help address this by providing contextual or supplementary evidence that enhances the applicability of research findings. Second, long-term outcomes and rare harms may not be adequately captured in RCTs due to short follow-up periods or ethical constraints, making non-randomised studies (NRS) valuable for assessing these effects. Third, observational data can help resolve inconsistencies in trial results, inform subgroup analyses, and provide baseline risk estimates, which are crucial for determining absolute benefits and harms. Finally, when RCT evidence is insufficient or indirect, well-conducted observational studies may serve as sequential evidence or, in some cases, even replace RCTs if they provide equal or higher confidence in the estimated effects.[19] Therefore, integrating observational data into comparative effectiveness research enhances the completeness and relevance of evidence used for clinical and policy decision-making. In

order to circumvent recruitment bias and secure pragmatic sampling, so called "real-world data" from observational cohorts could potentially be used to complement the lacking evidence from randomised trials.[19] The purpose of target trial emulation is to enable causal inference from observational data by explicitly modelling a hypothetical randomised trial, enhancing validity and reliability.[20] This structured approach aligns analyses with randomised trial principles, mitigates methodological biases, and strengthens comparative effectiveness research using big data.[21] It requires data on prognostic factors that are associated with treatment decisions. However theoretical, if all such confounders are precisely gathered and adjusted for, there would be no difference between the emulated trial and corresponding randomised (target) trial recruiting all eligible patients into a pragmatic RCT.[22] Propensity score methods, including matching, stratification, and inverse probability weighting, help balance covariates between treatment groups, mimicking randomisation.[23]

Here, we therefore outline the planned structure and statistical causal methods applied for an emulated target trial, allowing us to draw causal inferences about the impact of the shared decision between patient and the surgeon's initial clinical decision to perform surgery, relative to the initial clinical decision not to perform surgery to assess efficacy. The main outcome is the Patient Acceptable Symptom State (PASS), which dichotomously captures what matters most to patients—the acceptability of their current symptom state one year after allocation of treatment.

Objectives

Primary efficacy objective:
To evaluate the effectiveness of outpatient shoulder surgery, compared with non-surgical care in increasing the proportion of patients who achieve the Patient Acceptable Symptom State (PASS) one year after allocation, among individuals presenting with shoulder discomfort and imaging-confirmed pathology at a secondary healthcare facility.

Co-primary efficacy objective:
To evaluate the effectiveness of outpatient shoulder surgery, compared to non-surgical care, in reducing disability at one-year follow up, as measured by the proportion of patients achieving a clinically meaningful improvement (greater than 41% change) in the Quick Disabilities of the Arm, Shoulder and Hand (Quick-DASH) score, among individuals presenting with shoulder discomfort and imaging-confirmed pathology at a secondary healthcare facility.

The null hypothesis for this trial is that there is no difference between patients allocated to surgery based on a surgeon's initial clinical evaluation and tacit knowledge and patients allocated to non-surgical interventions.

Secondary objectives are:

1) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on the proportion of patients achieving "Better" or "Much better" in overall function and discomfort, measured with Global Perceived Effect (GPE) score at one year from allocation of intervention, in patients diagnosed with

shoulder discomfort and medical imagine verified pathology at a secondary healthcare facility.

2) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on change in shoulder pain and discomfort, measured using the short form of Disability of the Arm, Shoulder and Hand (Quick-DASH) questionnaire, from allocation of intervention to one year follow-up, in patients diagnosed with shoulder discomfort and medical imaging verified pathology at a secondary healthcare facility.
3) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on change in maximum shoulder pain, measured using a 0-100 Visual Analog Scale (VAS), from allocation of intervention to one year follow-up, in patients diagnosed with shoulder discomfort and medical imagine verified pathology at a secondary healthcare facility.
4) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on change in shoulder pain during activity, measured using a 0-100 Visual Analog Scale (VAS), from allocation of intervention to one year follow-up, in patients diagnosed with shoulder discomfort and medical imagine verified pathology at a secondary healthcare facility.
5) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on change in shoulder pain during rest, measured using a 0-100 Visual Analog Scale (VAS), from allocation of intervention to one year follow-up, in patients diagnosed with shoulder discomfort and medical imagine verified pathology at a secondary healthcare facility.
6) To determine the effectiveness of outpatient shoulder surgery, relative to non-surgical interventions, on change in shoulder related sleep disturbance, measured using a 0-100 Visual Analog Scale (VAS), from allocation of intervention to one year follow-up, in patients diagnosed with shoulder discomfort and medical imaging verified pathology at a secondary healthcare facility.

STUDY METHODS

Trial design

In this pragmatic emulation of a stratified, randomised, open-label, superiority trial, adults referred to secondary care with various shoulder diagnoses and medical imagine verified pathology were initially allocated into either Surgery group, or the Non-Surgical Treatment group with an expected allocation ratio of 1:15. The primary endpoint in the intention-to-treat population was PASS assessed one year from allocation of treatment. Secondary endpoints was assessed also at 3 and 6 months after allocation of treatment. We will analyse observational data from the prognostic Vejle Hospital Shoulder Cohort (VHS Cohort), initiated to evaluate prognostic factors for consistent pain 1 year after diagnosis, using a "causal inference framework", in an attempt to emulate a randomised target trial, stratified per diagnostic group defined as 1) Degenerative; 2) Traumatic injury; 3) Non-traumatic instability/luxation; 4) Non-specific pain, based upon ICD-10 diagnosis and traumatic/non-traumatic anamnesis (Illustration 1), that would answer the question whether more patients referred for surgery based on a surgeon's clinical evaluation will reach a satisfactory outcome, when compared to patients not referred for surgery. Thus, causal inference from these observational data needs to be evaluated with respect to how well they emulate a particular target trial that is not feasible/ethical to perform.

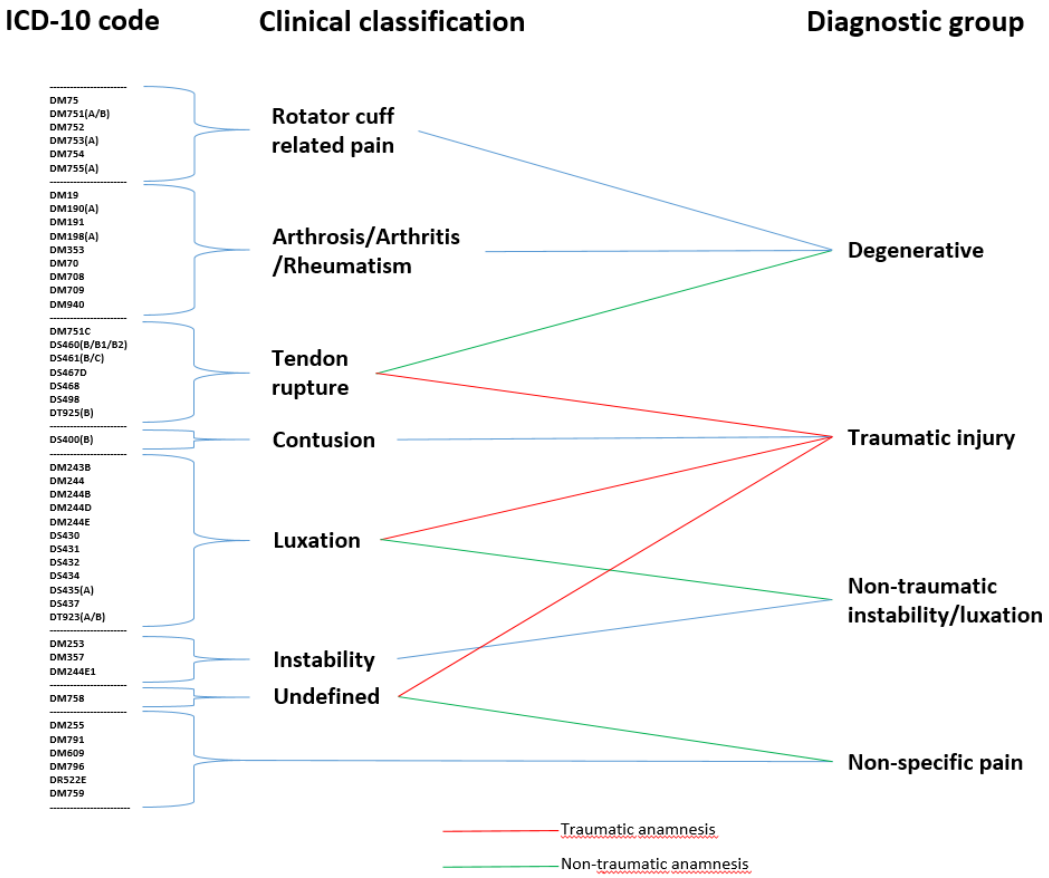


Illustration 1: Diagnostic grouping based on ICD-10 diagnoses.

The statistical analysis plan (SAP) is reported in accordance with the "Guidelines for the Content of Statistical Analysis Plans in Clinical Trials".[24] Reporting of the trial will follow the "Consolidated Standards of Reporting Trials" (CONSORT).[25 26]

The trial was registered at Clinicaltrials.gov (NCT04686435) in 12/2020. First patient was enrolled in January 2021. Last patient last follow-up was January 2025. The trial has been evaluated and approved by The Regional Committees on Health Research Ethics for Southern Denmark (Project.ID: S-20202000-16) and the Danish Data Protection Agency (journal no: 20/12316).

Group allocation

Allocation to surgery or non-surgical treatment is based on the clinician's initial overall judgment, taking into account patient history, clinical examination, imaging findings, patient expectations, relevant clinical guidelines and shared decision making. This real-world allocation process was analytically emulated to reflect a trial-like comparison.

Sample size and power considerations

The sample size calculation applies jointly to both primary efficacy objectives. A power analysis for comparing two proportions using Pearson's chi-square test indicated that, with a 1:15 allocation ratio between groups, a total sample size of 992 would provide 92% power to detect a 20 percentage point absolute difference (0.80 vs. 0.60) at a two-sided significance level of 5% ($\alpha = 0.05$). The calculation employed a normal approximation to the binomial distribution under the null hypothesis of no difference in proportions. For practical and operational considerations, the sample size was rounded up to 1000 participants to ensure sufficient power and account for potential deviations.

Framework

Both primary and secondary outcomes are analysed within a superiority framework; surgical interventions are favourable compared to non-surgical interventions.

Statistical interim analysis and stopping guidance

No statistical interim analysis is planned on the primary endpoint.

Timing of final analysis

Data extraction and analysis for the between-group comparison (Surgical vs. Non-Surgical) for the primary endpoint (1 year follow-up) is planned to be performed after making this SAP public available. The main publication of the trial will be prepared when these data have been received and cleaned (anticipated by September 2025).

Timing of outcome assessments

The overview of trial procedures and time-point of each outcome assessment is presented in the **Illustration 2** "Timeline for recruitment and follow-up" and **Box 1** "Outcomes and Covariates – Timepoint of assessment". The Treatment allocation date (Surgical or Non-surgical) is used to calculate 3 months, 6 months and 1 year follow-up time points.

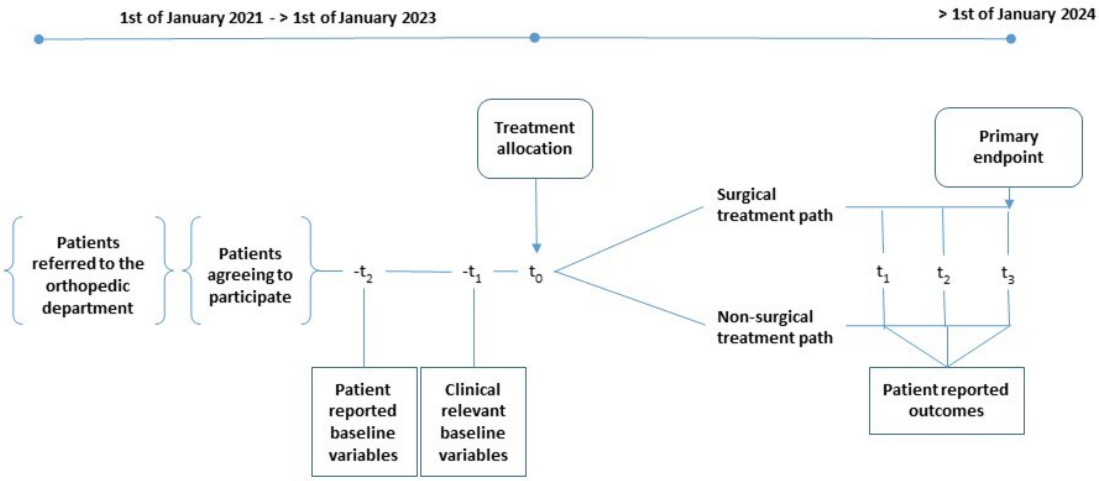


Illustration 2: Timeline for recruitment and follow-up. $-t_2$: 2-4 days prior to medical examination; $-t_1$=Day of medical examination; t_0=Allocation; t_1= 3 months; t_2= 6 months; t_3= 1 year.

STATISTICAL PRINCIPLES

Confidence intervals and p-values

All statistical tests and 95% confidence intervals (95%CIs) will be two-sided and *p-value* < 0.05 will be considered statistically significant for the primary and co-primary endpoint. Groups will be considered comparable if the two-sided 95% confidence interval for the odds ratio comparing Shoulder Surgery with Nonoperative Treatment, for both the primary and co-primary endpoint, lies entirely within the range of 0.80 to 1.25.[27] In the secondary outcomes Quick-DASH and VAS, a 95% confidence interval excluding a difference greater than 10 Quick-DASH points or 10 points on VAS between the two groups will be interpreted as indicating absence of a minimal clinically important difference (i.e. possible equivalence).

To address multiplicity, a gatekeeping procedure will be applied to both the primary objectives and the key secondary objectives to determine which will be used in the interpretation of the study results. While all outcomes will be analysed and presented, only those passing the gatekeeping procedure will be considered for inferential conclusions. The gatekeeping procedure follows a hierarchical approach in the prioritised order listed below. Statistical testing will proceed sequentially until an analysis fails to show a statistically significant difference ($p < 0.05$), or until all analyses have been completed.

If a test does not meet the specified statistical significance threshold, no further claims will be made for subsequent outcomes in the hierarchy:

1) PASS
2) Quick-DASH MIC
3) GPE
4) Quick-DASH

5) Maximum Pain
6) Pain in activity
7) Pain at rest
8) Shoulder related sleep disturbance

Adherence and protocol deviations

Adherence to the surgical intervention is assessed by the percentage of patients being referred to surgery at the initial clinical evaluation, and the actual number of these patients having a record of the surgery was performed in the REDCap database, including date and type of surgery performed: adherence% = (Number of patients with a surgical record in REDCap database/ number of patients referred for surgery at first medical examination) x 100.

Adherence to the non-surgical intervention is assessed by the percentage of patients being referred for non-surgical treatment, and the actual number of patients with a minimum of 1 record, in the REDCap database, of receiving non-surgical intervention after the initial clinical examination: Adherence% = (Number of patients with a record of receiving non-surgical intervention/number of patients referred for non-surgical intervention) x 100. Treatment adherence will be presented with descriptive statistics (N, %) by randomisation group.

Patient's crossing over to receive surgery from the non-surgical intervention group within one year from allocation and patients from the surgical group without a record in the REDCap database of surgery performed will be considered a major protocol deviation. Furthermore, patients in both groups withdrawing from the trial between baseline and one year follow-up will be considered as a major protocol deviation. Minor deviations are not completing all follow-up measurement. The number (percentage) of patients with respectively major and minor protocol deviations will be summarised by treatment group. No formal statistical testing will be conducted.

Analysis populations

The primary analysis will be performed on the Intention-to-treat (ITT) population, including all eligible patients according to the treatment they were initially allocated (initial clinical decision at first medical examination) to receive. A sensitivity analysis will be considered based on the per-protocol population, if >5% of the total number of patients in one treatment group are major protocol violators. The per-protocol analysis set will include: 1) patients who in the surgical group were allocated (initial clinical decision at first medical examination) to receive surgery, and have a record of the surgery was performed in the REDCap database, including date and type of surgery performed, 2) patient who in the non-surgical group were allocated (initial clinical decision at first medical examination) to receive non-surgical treatment and a minimum of 1 record within 4 months, in the REDCap database, of receiving non-surgical intervention after the initial clinical examination as well as no record of surgery performed.

Trial population

Screening data

The total number of patients referred for examination due to suspicion of shoulder disorder, scheduled for an examination more than one week ahead and sent an electronic invitation, 2-4 days prior to their examination, in their personal electronic mailbox (E-boks) will be presented in the CONSORT flowchart (Figure 1) to describe the percentage of participation.

Eligibility

All patients, not in need of an interpreter at their medical examination, between 18 and 75 years, referred from private practice to the orthopaedic shoulder sector at Lillebaelt Hospital (Vejle) was invited.

Eligible patients: 1) diagnosed with shoulder disorder at initial medical examination; 2) had shoulder specific pathology on medical imagine; and 3) gave written consent to participate.

Patients were excluded if: 1) patients had significant cognitive impairment; 2) were unable to read and understand Danish; or 3) was diagnosed with diagnostic codes related to fracture of shoulder or upper arm; neurological conditions, disorders of the scapulae or frozen shoulder.

Recruitment

The CONSORT flowchart[25] will comprise number of patients invited, excluded (with reasons), eligible for inclusion in the trial, allocated (initial clinical decision at first medical examination), receiving their allocated treatment, withdrawals (with reasons), and lost to follow-up (with reasons), included in ITT analysis. The CONSORT flowchart is depicted in **Figure 1.**

Withdrawal/Follow-up

The level of consent withdrawal will be tabulated, classified as: 1) withdraw from follow-up but allow data collected to date to be used; 2) withdraw from follow-up and withdraw consent for data collected to date to be used; 3) lost to contact/follow-up. Timing of withdrawal will be presented with number and reasons in the CONSORT flowchart at the 3 months, 6 months and 1 year outcome assessment by treatment group.

Baseline patient characteristics

Data presented in table 1 will be used to describe patient characteristics by randomisation group at baseline. Categorical variables will be presented with numbers and percentages. Means and standard deviations will be reported for normally distributed continuous variables; medians, interquartile ranges, and ranges will be reported for skewed data. Tests of statistical significance will not be performed for baseline characteristics. Because the groups arise from a non-randomised study design, standardised differences will be used to assess between-group imbalances, with attention to differences of potential clinical importance.

ANALYSIS

Outcome and endpoint definitions

Primary:

Outcome:

The primary outcome of this trial is the patient's perceived symptom acceptability in relation to their upper extremity condition, measured using the concept of the Patient Acceptable Symptom State (PASS). PASS was original designed for rheumatic disorders,[28] but has shown to be a useful outcome for other musculoskeletal and orthopaedic conditions.[29-32] The PASS defines the highest level of symptom at which patients find their condition acceptable.[33]

Endpoint:

The primary endpoint is the proportion of participants responding "Yes" to the PASS anchor question:

"Taking into account all your daily activities, pain and functional impairment, do you consider your current state satisfactory?"

This binary endpoint (Yes/No) will be assessed at 1 year post allocation. A higher proportion of "Yes" responses indicates a more favourable outcome in symptom acceptability.

Co-Primary:

Outcome:

The co-primary outcome of this trial is the change in upper extremity disability, assessed using the Quick Disabilities of the Arm, Shoulder and Hand (Quick-DASH) questionnaire.[34] The Quick-DASH is a validated, patient-reported measure designed to evaluate physical function and symptoms in people with musculoskeletal disorders of the upper limb.[34 35] A relative change exceeding 41% has been proposed as a clinically meaningful threshold for improvement in upper extremity conditions.[35]

Endpoint:

The co-primary endpoint is the proportion of participants achieving a reduction greater than 41%, corresponding to a minimal important change,[35] in their Quick-DASH score from baseline to 1 year post allocation (Quick-DASH MIC). This binary endpoint (achieved/not achieved) reflects whether an individual's improvement surpasses the defined clinically important change. A higher proportion of patients reaching this threshold indicates a more favourable treatment effect on physical function and symptoms.

Key Secondary:

The following key secondary outcomes will be analysed to assess the between-group differences in treatment effects over time. Unless otherwise stated, the primary analysis time point for all key secondary outcomes is 1 year after treatment allocation. Additional measurements will be taken at 3 and 6 months for descriptive and exploratory analyses of trajectories over time.

1. Global Perceived Effect (GPE)

Outcome:

Patient-reported global impression of change in "Overall shoulder problems", measured using a 15-point Likert scale ranging from "A great deal worse" (score = 1) to "A great deal better" (score = 15).

Endpoint:
The participants answer on the GPE question "All in all, how do you experience your shoulder problems now compared to before you started your treatment in the study?" will be dichotomised into "Improved" (score > 9) and "Not improved" (score ≤ 9). The key secondary endpoint is the proportion of participants classified as "Improved" at 1 year after treatment allocation. The GPE has been shown to be a valid and reliable outcome measure in shoulder trials and is recommended by OARSI.[36 37]

2. QuickDASH Score
Outcome:
Self-reported upper extremity disability and symptom severity, measured using the 11-item Quick Disabilities of the Arm, Shoulder and Hand (QuickDASH) questionnaire.[34]. Each item is rated on a 5-point Likert scale and transformed into a total score ranging from 0 (no disability/symptoms) to 100 (most severe disability/symptoms). The optional work and sports/music modules are excluded from the analysis. The QuickDASH has demonstrated high reliability, validity, and responsiveness in patients with shoulder conditions.[34 35]

Endpoint:
The key secondary endpoint is the between-group difference in change in total QuickDASH score from allocation to 1 year follow-up. The outcome will also be measured at 3 and 6 months for secondary analyses.

3. Shoulder Pain Intensity variables
Outcome:
Pain intensity in the affected shoulder over the past 7 days, assessed separately in three specific contexts:

- Maximum pain experienced
- Pain during physical activity
- Pain at rest

Endpoint:
Each pain domain is measured on a 100 mm visual analogue scale (VAS) anchored with 0 = "No pain" and 100 = "Worst pain imaginable," as recommended by the OARSI.[36] The key secondary endpoint is the between-group difference in change from allocation to 1 year follow-up. Additional measurements will be collected at 3 and 6 months.

4. Shoulder-Related Sleep Disturbance
Outcome:
Degree of sleep disturbance attributed to shoulder symptoms in the past 7 days.

Endpoint:
Measured using a 100 mm VAS, anchored with 0 = "Not disturbed" and 100 = "Extremely disturbed." The key secondary endpoint is the between-group difference in change from allocation to 1 year follow-up, with additional measurements at 3 and 6 months.

Covariates for development of propensity score

To account for potential confounding and to improve comparability between the surgical and non-surgical treatment cohorts, propensity score adjustment will be applied. The propensity

score will be estimated as the conditional probability of receiving surgical treatment, given a set of baseline covariates measured prior to treatment allocation.

Covariates are selected based on clinical relevance and their potential association with both treatment assignment and outcomes. These include patient-reported measures of general health, health-related quality of life, fear-avoidance beliefs, treatment expectations, and physical activity levels, as well as sociodemographic characteristics (e.g., age, sex, education), lifestyle factors (e.g., smoking, BMI), and clinical assessments including imaging and physical examination findings.

All covariates used in the propensity score model were collected prior to referral to either surgical or non-surgical treatment pathways. The full list of covariates and the timing of their assessment, along with the timing of primary and secondary outcome measurements, is detailed in Box 1.

Box 1: Outcome and Covariates – Timepoint of assessment.

	Timepoint of assessment			
	Baseline	3 months	6 months	1 year
Outcome measures				
Primary outcome				
PASS, no. (%)				X
Quick-DASH MIC, no. (%)				x
Key Secondary Outcomes				
GPE - Overall shoulder problems – no. (%)				X
Quick-DASH – 0-100	X	X	X	X
VAS shoulder pain 'Maximum' – 0-100	X	X	X	X
VAS shoulder pain 'Activity' – 0-100	X	X	X	X
VAS shoulder pain 'Rest' – 0-100	X	X	X	X
VAS shoulder related sleep disturbance – 0-100	X	X	X	X
Covariates (de-confounding variables)				
Demographic variables				
Age (years)	X			
Female sex, no. (%)	X			
Body Weight (kg)	X			
Height (cm)	X			
BMI (kg/m^2)	X			
Central adiposity (Waist-To-Height ratio > 0.50), no. (%)	X			

Marital status - Married or cohabitating, no. (%)	X			
Educational level - Education above high school level, no. (%)	X			
Work status - Employed for wages or Self-employed, no. (%) - Unemployed, no. (%) - Sick-leave, no. (%) - Retired, no. (%) - Other, no. (%)	X			
Smoking status - Current smoker, no. (%) - Previous smoker, no. (%) - Non-smoker, no. (%)	X			
Alcohol consumption - >7 units/week, no. (%)	X			
Multiple Pain sites - 0 pain sites (besides current shoulder pain), no. (%) - 1-2 pain sites (besides current shoulder pain), no. (%) - >2 pain sites (besides current shoulder pain), no. (%)	X			
Comorbidity, no. (%) - No comorbidities, no. (%) - 1-2 comorbidities, no. (%) - >2 comorbidities, no. (%)	X			
Preference regarding treatment - Surgery, no. (%) - Corticoid injection, no. (%) - Physiotherapy, no. (%) - Other, no. (%)	X			
Expectation of treatment effect – 0 (No effect) – 100 (Complete recovery)	X			
Treatment received prior to referral due to current shoulder pain - None, no. (%) - Physiotherapy or Chiropractor treatment, no. (%) - Corticoid injection, no. (%)	X			
Current workers compensation claim, no. (%)	X			
Current insurance compensation claim, no. (%)	x			
Sick leave due to shoulder within the last 3 months – days	X			
Clinical assessment variables				
Duration of current shoulder pain - <3 months, no. (%) - 3-6 months, no. (%) - >6-12 months, no. (%) - >12 months, no. (%)	X			
Presence of constant pain, no. (%)	X			
Injury history - Traumatic, no. (%)	X			

- Acute overuse (Short duration of hard work, without history of trauma), no. (%) - Prolonged overuse (Prolonged stress on shoulder), no. (%) - Unknown, no. (%)				
Use of analgesics due to shoulder-related pain	X			
Use of analgesics not due to shoulder-related pain	X			
Taking medication due to Hypertension, no. (%)	X			
Taking medication due to Elevated serum triglycerides or Low serum high-density lipoprotein (HDL) , no. (%)	X			
Insulin resistance, no. (%)	X			
Metabolic syndrome*, no. (%)	X			
Range of motion – Active Elevation – degrees	X			
Range of motion – Active External rotation – degrees	X			
Range of motion – Active internal rotation – degrees	X			
Range of motion – Passive Elevation – degrees	X			
Range of motion – Passive External rotation – degrees	X			
Abnormal Isometric strength in elevation, no. (%)	X			
Abnormal Isometric strength in internal rotation, no. (%)	X			
Abnormal Isometric strength in external rotation, no. (%)	X			
Major symptoms - None, no. (%) - Pain, no. (%) - Stiffness, no. (%) - Instability, no. (%) - Decreased strength, no. (%)	x			
Atrophy of one or more of shoulder muscles, no. (%)	X			
Abnormal Scapula placement, no. (%)	X			
Palpation tenderness of shoulder girdle - Acromion Clavicular joint, no. (%) - Biceps tendon, no. (%) - Other, no. (%) - Missing, no. (%)	X			
Ultrasound measures (Acromion Clavicular joint) - Normal, no. (%) - Osteoartrose (Narrowed joint space OR osteophyte), no. (%) - Hypertrophy of capsule, no. (%) - Luxation, no. (%) - Missing, no. (%)	X			
Ultrasound measures - Structural changes in Rotator cuff, no. (%)	X			
Ultrasound measures (Rotator cuff ruptures) - Normal, no. (%) - Partial, no. (%) - Complete, no. (%) - Missing, no. (%)	x			
Ultrasound measures - structural changes in m. Biceps tendon, no. (%)	x			
Ultrasound measures – Inflammation of Bursae, no. (%)	x			

Ultrasound measures - Clinically relevant Calcification, no. (%)				
Radiographic - Acromion Clavicular joint arthrosis, no. (%) - Gleno-Humeral-joint arthrosis, no. (%) - Acromion deformations, no. (%) - Other, no. (%) - Missing, no. (%)	X			
3 out of 5 positive Clinical test (Hawkins; Neer's; Full can; Empty Can; Resisted ext. rotation), no. (%)	X			
Patient reported outcome measure (PROM)				
WHO-5 – 0-100	X			
EQ VAS – 0-100	X			
PSEQ(4) – DK – 0-24	X			
TSK-13 – 0-52	X			
UCLA-as – 1-10	x			

Results will be reported as means and standard deviations unless otherwise indicated. PASS: Patient acceptable Symptom State; Quick DASH MIC: Short version of the Disability of the Arm, Shoulder and Hand Questionnaire (Quick-DASH) criteria for Minimal Important Change (41% change); GPE: Global Perveived Effect; Quick-DASH: The short version of the Disability of the Arm, Shoulder and Hand Questionnaire (Quick-DASH); VAS: Visual Analog Scale; WHO-5: The World Health Organization-Five Well-Being Index[38]; EQ-VAS: European Quality of life – Visual Analog Scale[39]; PSEQ(4): Pain Self-Efficacy Questionnaire – 4 item[40]; TSK-13: Tampa Scale of Kinesiophobia 13 item[41]; UCLA-as: University of California Los Angeles Activity Scale[42]; *Metabolic syndrome defined as 2 or more positive factors out of: 1)Medication due to hypertension; 2) Medication due to Triglycerides/HDL; 3) Insulin resistance; 4) Waist-To-Height ratio >0.5.

Analysis methods

All descriptive statistics and statistical analysis will be reported in accordance with the recommendations of the *"Enhancing the QUAlity and Transparency Of health Research"* (EQUATOR) network[43] and the CONSORT statement.[44] Visual inspection (QQ-plot, histograms, and scatterplots) of the standardised residuals from the statistical model will be used to assess the assumption of normality and homogeneity of variances. Following the flow diagram (Figure 1), which outlines the population included in the intention-to-infer analysis, baseline characteristics will be summarised by study group (Table 1).[45] This table will include demographic variables, relevant medical history, and other baseline factors potentially associated with exposure and the outcomes of interest.[27]

To analyse observational data for causal treatment effects, the most important methodological challenge is to control bias due to lack of randomisation. With the rising challenge of analysing more complex observational data, the propensity score has been the foundation for many of these approaches. Bias control methods based on propensity scores have become widely accepted. To define the propensity score, we introduce the following notation: let $X = (X_1,\ldots, X_{n\text{i}})$ represent n confounders that are measured prior to intervention initiation (referred as "baseline confounders" below), then $X_i = (X_{1i},\ldots, X_{n\text{i}})$ is a vector of the value of the n confounders for the i^{th} subject. Let T represent the available interventions, with $T = 1$ indicating the subject is in the group initially allocated to surgery and $T = 0$ meaning the subject in the control group (ie, referral to non-surgical intervention). For the i^{th} subject, the propensity score is the conditional probability of being in the treated group given their measured baseline confounders,

$$\text{Prob}\,(X_i) = (T_i = 1 \mid X_i)$$

Intuitively, the propensity score provides a way to simulate randomisation when actual randomisation is not feasible. By conditioning on the propensity score, we aim to ensure that each subject has an equal chance of receiving treatment, thereby balancing the distributions of measured baseline confounders between treatment and control groups. The stabilized inverse probability treatment weighting (sIPTW) approach will be used as the analysis. A multivariable binomial logistic regression model will be utilised to generate the propensity scores.

The primary analyses will be based on the Intention to Treat (ITT) population. Categorical endpoints (e.g., the proportion achieving PASS/Quick-DASH MIC after 1 year) will be analysed using logistic regression, with treatment group and stratification factors as covariates, based on the IPW data provided initially. The primary estimand will be the odds ratio (OR) with 95% confidence intervals (CIs) derived from the fully adjusted logistic regression model. The main analysis will use inverse probability weighting, with further adjustment for stratification variables. The OR will represent the treatment effect, with CIs providing the range within which the true treatment effect is expected to lie with 95% certainty. A p-value will be reported to assess the statistical significance of the treatment effect, testing the null hypothesis of no difference between groups. An OR >1 indicates higher odds of achieving the PASS/Quick-DASH MIC outcome with surgical treatment, while an OR <1 indicates lower odds. For secondary objectives, categorical outcomes will be analysed using the same approach as the primary endpoints, while continuous outcomes will be analysed as change from baseline using repeated-measures mixed-effects linear models. Patient identification number will be included as a random effect. Fixed effect factors will include baseline score, treatment group (Shoulder Surgery vs Nonoperative Treatment), time point (baseline, 3, 6, and 12 months), stabilised inverse probability of treatment weighting (sIPTW), and treatment-by-time interactions.

Sensitivity analysis will be performed with the purpose to test the robustness of the primary analyses, based on the per-protocol population, if >5% of the total number of patients in one treatment group are major protocol violators. The per-protocol analysis set will include: 1) patients who in the surgical group were 'allocated (initial clinical decision at first medical examination) to receive surgery, and have a record of the surgery was performed in the REDCap database, including date and type of surgery performed, 2) patient who in the non-surgical group were allocated (initial clinical decision at first medical examination) to receive non-surgical treatment and a minimum of 1 record, in the REDCap database, of receiving non-surgical intervention after the initial clinical examination as well as no record of surgery performed.

Subgroup analyses will be performed if imbalance is seen in the applied stratification sequence, only including diagnostic subgroups with balanced patient distributions between surgical treated patients and non-surgical treated patients.

Missing data

Missing data are ignorable only when they are missing completely at random (MCAR), meaning no systematic differences exist between missing and observed values. More commonly, data are assumed to be missing at random (MAR), where any systematic

differences can be explained by observed variables.[27] Under the MAR assumption, repeated measures mixed-effects models or multiple imputation is recommended as it yields valid standard errors, P values, and confidence intervals.[46] For categorical efficacy endpoints, missing data at 1 year from baseline will be addressed through non-responder imputation, which is supposedly representing a conservative approach. For continuous efficacy endpoints repeated measures mixed effects models will handle missing data; missing data will be handled indirectly with the use of the mixed-effects linear-models approach with an assumption that data were missing at random, in accordance with the working assumption underlying these models.[47]

Additional analyses

No additional analyses are planned.

Harms

No data on harm was collected.

Statistical software

All statistical analyses will be conducted using SAS (SAS Institute Inc) and R (R Foundation for Statistical Computing). SAS will be used primarily for data management and modeling, while R will be applied for supplementary analyses, data visualisation, and reproducibility checks.

REFERENCES:

1. Lucas J, van Doorn P, Hegedus E, et al. A systematic review of the global prevalence and incidence of shoulder pain. *BMC Musculoskelet Disord* 2022;23(1):1073. doi: 10.1186/s12891-022-05973-8 [published Online First: 2022/12/09]
2. Jordan KP, Kadam UT, Hayward R, et al. Annual consultation prevalence of regional musculoskeletal problems in primary care: an observational study. *BMC Musculoskelet Disord* 2010;11:144. doi: 10.1186/1471-2474-11-144 [published Online First: 2010/07/06]
3. Greving K, Dorrestijn O, Winters JC, et al. Incidence, prevalence, and consultation rates of shoulder complaints in general practice. *Scand J Rheumatol* 2012;41(2):150-5. doi: 10.3109/03009742.2011.605390 [published Online First: 2011/09/23]
4. Huisstede BM, Wijnhoven HA, Bierma-Zeinstra SM, et al. Prevalence and characteristics of complaints of the arm, neck, and/or shoulder (CANS) in the open population. *Clin J Pain* 2008;24(3):253-9. doi: 10.1097/AJP.0b013e318160a8b4

00002508-200803000-00011 [pii] [published Online First: 2008/02/22]
5. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004;33(2):73-81. [published Online First: 2004/05/28]
6. Miranda H, Viikari-Juntura E, Heistaro S, et al. A population study on differences in the determinants of a specific shoulder disorder versus nonspecific shoulder pain without clinical findings. *Am J Epidemiol* 2005;161(9):847-55. doi: 161/9/847 [pii]

10.1093/aje/kwi112 [published Online First: 2005/04/21]
7. Steuri R, Sattelmayer M, Elsig S, et al. Effectiveness of conservative interventions including exercise, manual therapy and medical management in adults with shoulder

impingement: a systematic review and meta-analysis of RCTs. *Br J Sports Med* 2017;51(18):1340-47. doi: 10.1136/bjsports-2016-096515
8. Beard DJ, Rees JL, Cook JA, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. *Lancet* 2017 doi: 10.1016/S0140-6736(17)32457-1
9. Bennell K, Wee E, Coburn S, et al. Efficacy of standardised manual therapy and home exercise programme for chronic rotator cuff disease: randomised placebo controlled trial. *BMJ* 2010;340:c2756. doi: 10.1136/bmj.c2756
10. Brox JI, Gjengedal E, Uppheim G, et al. Arthroscopic surgery versus supervised exercises in patients with rotator cuff disease (stage II impingement syndrome): a prospective, randomized, controlled study in 125 patients with a 2 1/2-year follow-up. *Journal of shoulder and elbow surgery / American Shoulder and Elbow Surgeons [et al]* 1999;8(2):102-11.
11. Brox JI, Staff PH, Ljunggren AE, et al. Arthroscopic surgery compared with supervised exercises in patients with rotator cuff disease (stage II impingement syndrome). *BMJ* 1993;307(6909):899-903. [published Online First: 1993/10/09]
12. Haahr JP, Andersen JH. Exercises may be as efficient as subacromial decompression in patients with subacromial stage II impingement: 4-8-years' follow-up in a prospective, randomized study. *Scand J Rheumatol* 2006;35(3):224-8. doi: T846N68613Q14W73 [pii]

10.1080/03009740600556167 [published Online First: 2006/06/13]
13. Haahr JP, Ostergaard S, Dalsgaard J, et al. Exercises versus arthroscopic decompression in patients with subacromial impingement: a randomised, controlled study in 90 cases with a one year follow up. *Ann Rheum Dis* 2005;64(5):760-4. doi: 64/5/760 [pii]

10.1136/ard.2004.021188 [published Online First: 2005/04/19]
14. Ketola S, Lehtinen J, Arnala I, et al. Does arthroscopic acromioplasty provide any additional value in the treatment of shoulder impingement syndrome?: a two-year randomised controlled trial. *J Bone Joint Surg Br* 2009;91(10):1326-34. doi: 10.1302/0301-620X.91B10.22094

91-B/10/1326 [pii] [published Online First: 2009/10/02]
15. Mohammed HT, Yoon S, Hupel T, et al. Unnecessary ordering of magnetic resonance imaging of the knee: A retrospective chart review of referrals to orthopedic surgeons. *PLoS One* 2020;15(11):e0241645. doi: 10.1371/journal.pone.0241645 [published Online First: 2020/11/03]
16. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37. doi: 10.1186/1745-6215-10-37 [published Online First: 2009/06/06]
17. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147. doi: 10.1136/bmj.h2147 [published Online First: 2015/05/10]
18. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869. doi: 10.1136/bmj.c869 [published Online First: 2010/03/25]
19. Schunemann HJ, Tugwell P, Reeves BC, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Res Synth Methods* 2013;4(1):49-62. doi: 10.1002/jrsm.1078 [published Online First: 2013/03/01]

20. Hernan MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med* 2021;385(15):1345-48. doi: 10.1056/NEJMp2113319 [published Online First: 2021/10/02]
21. Hernan MA, Dahabreh IJ, Dickerman BA, et al. The Target Trial Framework for Causal Inference From Observational Data: Why and When Is It Helpful? *Ann Intern Med* 2025;178(3):402-07. doi: 10.7326/ANNALS-24-01871 [published Online First: 2025/02/17]
22. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390. doi: 10.1136/bmj.a2390 [published Online First: 2008/11/13]
23. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127(8 Pt 2):757-63. doi: 10.7326/0003-4819-127-8_part_2-199710151-00064 [published Online First: 1998/02/12]
24. Gamble C, Krishan A, Stocken D, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. *JAMA* 2017;318(23):2337-43. doi: 10.1001/jama.2017.18556 [published Online First: 2017/12/21]
25. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *BMJ* 2025;389:e081123. doi: 10.1136/bmj-2024-081123 [published Online First: 2025/04/15]
26. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 explanation and elaboration: updated guideline for reporting randomised trials. *BMJ* 2025;389:e081124. doi: 10.1136/bmj-2024-081124 [published Online First: 2025/04/15]
27. Christensen R, Ranstam J, Overgaard S, et al. Guidelines for a structured manuscript: Statistical methods and reporting in biomedical research journals. *Acta Orthop* 2023;94:243-49. doi: 10.2340/17453674.2023.11656 [published Online First: 2023/05/12]
28. Tubach F, Wells GA, Ravaud P, et al. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. *J Rheumatol* 2005;32(10):2025-9. [published Online First: 2005/10/06]
29. de Vos Andersen NB, Kent P, Hjort J, et al. Clinical course and prognosis of musculoskeletal pain in patients referred for physiotherapy: does pain site matter? *BMC Musculoskelet Disord* 2017;18(1):130. doi: 10.1186/s12891-017-1487-3 [published Online First: 2017/03/31]
30. Tubach F, Dougados M, Falissard B, et al. Feeling good rather than feeling better matters more to patients. *Arthritis Rheum* 2006;55(4):526-30. doi: 10.1002/art.22110 [published Online First: 2006/07/29]
31. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. *Ann Rheum Dis* 2005;64(1):34-7. doi: 10.1136/ard.2004.023028 [published Online First: 2004/05/08]
32. Christiansen DH, Hjort J. Group-based exercise, individually supervised exercise and home-based exercise have similar clinical effects and cost-effectiveness in people with subacromial pain: a randomised trial. *J Physiother* 2021;67(2):124-31. doi: 10.1016/j.jphys.2021.02.015 [published Online First: 2021/03/22]
33. 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;66 Suppl 3(Suppl 3):iii40-1. doi: 10.1136/ard.2007.079798 [published Online First: 2007/11/21]
34. Beaton DE, Wright JG, Katz JN, et al. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am* 2005;87(5):1038-46. doi: 10.2106/JBJS.D.02060

35. Budtz CR, Andersen JH, de Vos Andersen NB, et al. Responsiveness and minimal important change for the quick-DASH in patients with shoulder disorders. *Health Qual Life Outcomes* 2018;16(1):226. doi: 10.1186/s12955-018-1052-2 [published Online First: 2018/12/12]
36. Fitzgerald GK, Hinman RS, Zeni J, Jr., et al. OARSI Clinical Trials Recommendations: Design and conduct of clinical trials of rehabilitation interventions for osteoarthritis. *Osteoarthritis Cartilage* 2015;23(5):803-14. doi: 10.1016/j.joca.2015.03.013 [published Online First: 2015/05/09]
37. Puga VO, Lopes AD, Shiwa SR, et al. Clinimetric testing supports the use of 5 questionnaires adapted into Brazilian Portuguese for patients with shoulder disorders. *J Orthop Sports Phys Ther* 2013;43(6):404-13. doi: 10.2519/jospt.2013.4441 [published Online First: 2013/03/20]
38. Bech P, Gudex C, Johansen KS. The WHO (Ten) Well-Being Index: validation in diabetes. *Psychother Psychosom* 1996;65(4):183-90. doi: 10.1159/000289073 [published Online First: 1996/01/01]
39. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33(5):337-43. doi: 10.3109/07853890109002087 [published Online First: 2001/08/09]
40. McWilliams LA, Kowal J, Wilson KG. Development and evaluation of short forms of the Pain Catastrophizing Scale and the Pain Self-efficacy Questionnaire. *Eur J Pain* 2015;19(9):1342-9. doi: 10.1002/ejp.665 [published Online First: 2015/03/15]
41. Miller RP, Kori S, Todd D. The Tampa Scale: A measure of kinesiophobia. *Clin J Pain* 1991;7(1):51-52.
42. Morup-Petersen A, Skou ST, Holm CE, et al. Measurement properties of UCLA Activity Scale for hip and knee arthroplasty patients and translation and cultural adaptation into Danish. *Acta Orthop* 2021;92(6):681-88. doi: 10.1080/17453674.2021.1977533 [published Online First: 2021/09/18]
43. Christensen R, Bliddal H, Henriksen M. Enhancing the reporting and transparency of rheumatology research: a guide to reporting guidelines. *Arthritis Res Ther* 2013;15(1):109. doi: 10.1186/ar4145 [published Online First: 2013/03/02]
44. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332. doi: 10.1136/bmj.c332 [published Online First: 2010/03/25]
45. Christensen R, Haugegaard T, Jakobsen JC. Role of prespecified analysis plans in physiological research: Encouraged or mandatory? *Exp Physiol* 2025 doi: 10.1113/EP092034 [published Online First: 2025/02/21]
46. Detry MA, Lewis RJ. The intention-to-treat principle: how to assess the true effect of choosing a medical treatment. *JAMA* 2014;312(1):85-6. doi: 10.1001/jama.2014.7523 [published Online First: 2014/07/25]
47. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. *JAMA* 2016;315(4):407-8. doi: 10.1001/jama.2015.19394 [published Online First: 2016/01/28]
48. Danish Health Authority. Sundhedsstyrelsens udmeldinger om alkohol 2021 [Available from: https://www.sst.dk/da/viden/alkohol/alkoholforebyggelse/sundhedsstyrelsens-udmeldinger-om-alkohol accessed 14-05-2021.

Tables:

Table 1. Baseline Characteristics of the Patients in the ITT population. *

	Surgical group (N=??)	Non-Surgical group (N=??)	Between-Group Differences (95% CI)	Standardised Differences
Female sex — no. (%)				
Age — yr				
Height — m				
Weight — kg				
Body-mass index — kg/m^2				
Education level beyond high school — no. (%)				
Employment				
Employed for wages — no. (%)				
Self-employed — no. (%)				
Unemployed — no. (%)				
Sick leave — no. (%)				
Retired — no. (%)				
Other — no. (%)				
Substance use				
Current smoker — no. (%)				
Alcohol consumption above recommendations† — no. (%)				
Index shoulder right — no. (%)				
Duration of shoulder symptoms — months				
Previous treatment for current shoulder disorder				
Physiotherapist/Chiropractor — no. (%)				
Corticosteroid injection — no. (%)				
Used pain medication in the past week due to shoulder-related pain				
Paracetamol — no. (%)				

Ibuprofen — no. (%)	
Morphine or opioids — no. (%)	
Other — no. (%)	
Diagnostic group	
Degenerative — no. (%)	
Non-Specific shoulder pain — no. (%)	
Non-traumatic instability/luxation — no. (%)	
Traumatic injury — no. (%)	
Quick-DASH score‡ — 0 to 100	
VAS Shoulder pain Maximum§ — 0 to 100	
VAS Shoulder pain Activity§ — 0 to 100	
VAS Shoulder pain Rest§ — 0 to 100	
VAS Shoulder related sleep Disturbance§ — 0 to 100	

* Plus–minus values are mean ±SD unless otherwise indicated.

† The Danish Health Authority guideline recommends alcohol consumption lower than 7 units per week for females and 14 units per week for males to have a low risk of developing diseases.[48]

‡ The short version of the Disability of the Arm, Shoulder and Hand Questionnaire (Quick-DASH) ranges from 0 to 100, with higher scores indicating worse disease status.

§ Visual Analog Scale (VAS) ranges from 0 to 100, with higher scores indicating worse status.

Table 2. Primary and Key Secondary Outcomes at 1 year in the Intention-to-Treat Population.*

VARIABLE	SURGICAL GROUP	NON-SURGICAL GROUP	BETWEEN GROUP		
PRIMARY OUTCOME					
			OR (95%CI)		
			Crude	**Adj.**	**p-value**
PASS – no. (%)					
QUICK-DASH MIC – no. (%)					
KEY SECONDARY OUTCOME					
GPE - OVERALL SHOULDER PROBLEMS – no. (%)					

	Mean change from baseline	SE	Mean change from baseline	SE	LSMean difference (95% CI)	
					Adj.	p-value
Quick-DASH – 0 to 100						
VAS SHOULDER PAIN MAXIMUM – 0 to 100						
VAS SHOULDER PAIN ACTIVITY – 0 to 100						
VAS SHOULDER PAIN REST – 0 to 100						
VAS SLEEP DISTURBANCE – 0 to 100						

PASS: Patient Acceptable Symptom State; GPE: Global Perceived Effect; Quick-DASH: Short form of Disability of the Arm, Shoulder and Hand questionnaire; VAS: Visual Analog Scale. * All analyses will be based on the ITT population: For dichotomous outcomes, logistic regression will be applied, with treatment group and stratification factors as covariates, based on the IPW data (missing data at 1 year from allocation will be addressed through non-responder imputation); The primary estimate will be the odds ratio (OR) with 95% confidence intervals (CIs) derived from the logistic regression model. For continuous outcomes, repeated measures linear mixed effects models (with no imputation for missing data) will be applied; Estimates are mean change from baseline and standard deviations with difference between groups reported with 95% confidence intervals.

Figures:

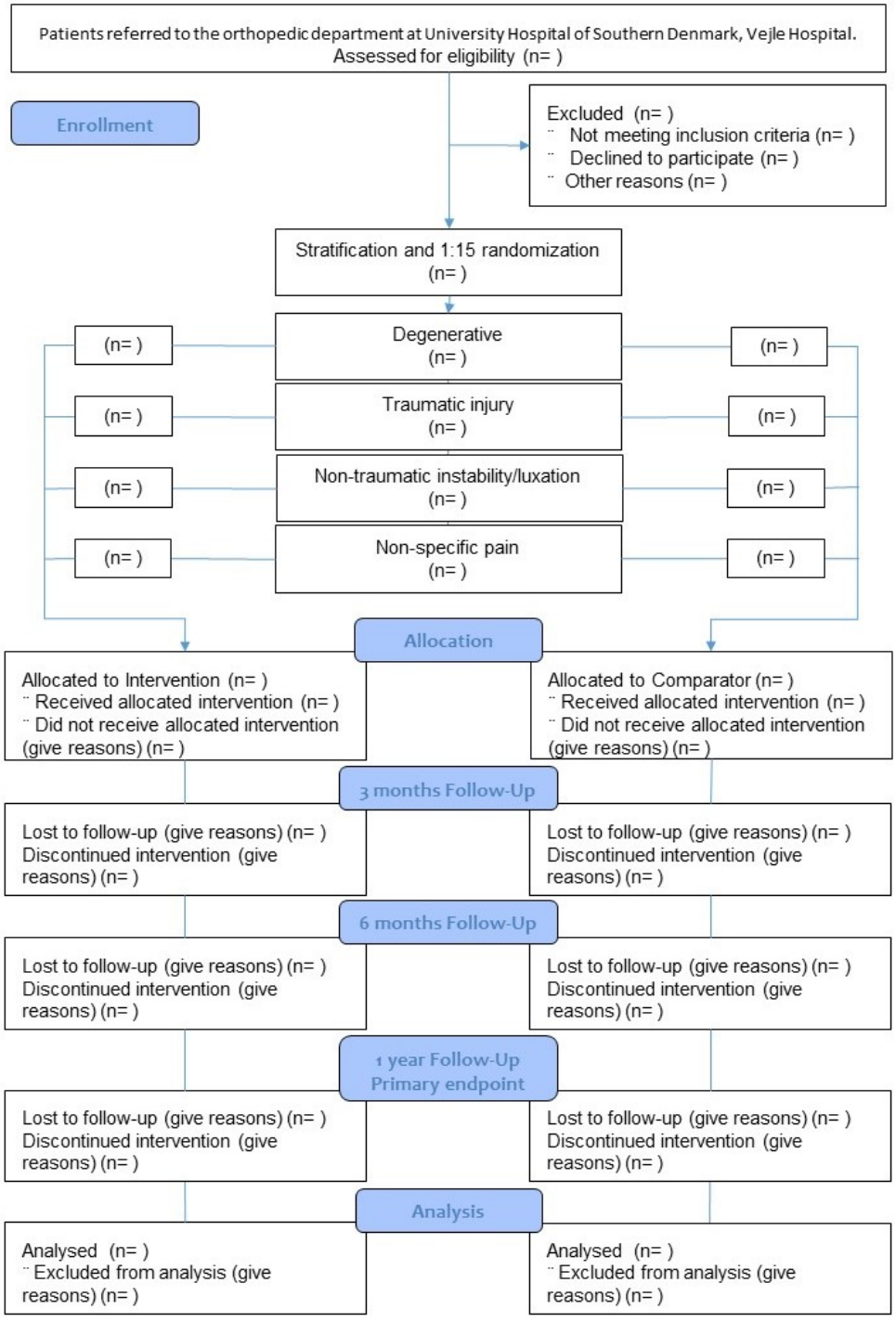


Figure 1: CONSORT flow-chart. Surgical group (Intervention), Non-Surgical group (Comparator)