Study Protocol

Trial Title: ChatGPT and DeepSeek-Assisted Rehabilitation in Subacromial Pain (LLM-RehabSP)

NCT Number: NCT07148687

Organization ID: 2025/874

Date: December 2025

Principal Investigator: Dr. Ömer Alperen Gürses

Affiliation: Kırşehir Ahi Evran University, Türkiye

Methods

Study Design and Ethics

This assessor-blinded, three-arm, parallel-group randomized controlled trial was conducted in the outpatient Physical Medicine and Rehabilitation clinic of Kırşehir Ahi Evran University Training and Research Hospital (Kırşehir, Türkiye). The trial was prospectively registered at ClinicalTrials.gov (NCT07148687). Ethical approval was obtained from the Selcuk University Faculty of Health Sciences Ethics Committee for Non-Interventional Clinical Investigations (approval no. 2025/874). All participants provided written informed consent prior to enrollment. The study was conducted in accordance with the Declaration of Helsinki and is reported following CONSORT recommendations.

Participants

Adults presenting with shoulder pain were screened by specialist physicians and referred to the study team when subacromial pain syndrome (SAPS) was diagnosed based on history and physical examination (eg, painful arc during elevation and positive impingement/resisted tests consistent with rotator cuff–related shoulder pain). When clinically indicated, ultrasound or MRI was used to exclude full-thickness rotator cuff tears and other major structural pathology.

Inclusion criteria were: age 18–65 years; clinical SAPS for ≥4 weeks; sufficient active shoulder range of motion to participate in exercise; ability to understand instructions and attend supervised sessions; and written informed consent. Exclusion criteria were: suspected or confirmed full-thickness rotator cuff tear, adhesive capsulitis, glenohumeral instability, or labral tear; differential diagnoses (eg, cervical radiculopathy, neurological disorders, inflammatory rheumatic disease, fibromyalgia); shoulder surgery within the previous 6 months; systemic or neuromuscular disease; pregnancy; or withdrawal/refusal of consent.

Randomization and Blinding

After baseline assessment, participants were randomized (1:1:1) to (1) conventional physiotherapy plus a conventional exercise program, (2) conventional physiotherapy plus a ChatGPT-guided exercise program, or (3) conventional physiotherapy plus a DeepSeek-guided exercise program.

An independent administrative staff member generated a computer-based block randomization sequence. Allocation concealment was ensured using sequentially numbered, opaque, sealed envelopes opened only after baseline measurements were completed.

Participants were informed that the trial compared different supervised exercise-based rehabilitation programs for SAPS but were not informed of the specific exercise-generation approach. The outcome assessor remained blinded to group allocation. Treating physiotherapists delivered the assigned programs but did not perform outcome assessments or data analysis. Data were coded prior to analysis.

Common Physiotherapy Modalities (All Groups)

All groups received the same conventional physiotherapy modalities at each session: moist hot pack (20 min), conventional TENS (100 Hz, 60 μs; 20 min), and therapeutic ultrasound (1 MHz, 1.5 W/cm²; 5 min). These modalities were delivered 3 sessions/week for 6 weeks (18 sessions). Because modalities were identical across groups, between-group differences were attributed to differences in exercise prescription.

Exercise Interventions

All participants completed a supervised 6-week exercise program (3 sessions/week; 45–60 min/session; 18 sessions total). Pain response was monitored using a standardized rule: acceptable pain during/after exercise was ≤4/10 on the NPRS without a meaningful increase in night pain or next-day baseline pain. If exceeded, exercises were regressed (eg, reduced range/load or increased support). Attendance (completed sessions out of 18) was recorded.

All three programs were structured into early, middle, and late phases and progressed according to the same pain-monitoring rule. Session-by-session programs are provided in Supplementary Table S1.

Conventional Exercise Program

The conventional program was developed a priori by two experienced musculoskeletal physiotherapists based on contemporary guideline-informed practice and published supervised exercise approaches for SAPS. It emphasized mobility/active-assisted elevation, scapular stabilization, progressive rotator cuff strengthening, and upper-quadrant stretching/strengthening.

LLM-Guided Exercise Programs (ChatGPT and DeepSeek)

The ChatGPT group used ChatGPT-4o (OpenAI; web interface; accessed June 2025) and the DeepSeek group used DeepSeek-R1 (DeepSeek; web interface; accessed June 2025). A standardized English prompt requested a 6-week, 18-session supervised outpatient exercise program using basic clinic equipment and explicitly restricted outputs to exercise and brief education only (no medications, injections, surgery, manual therapy, or electrotherapy). The prompt required inclusion of scapular stabilization, rotator cuff strengthening, thoracic mobility/posture exercises, phased progression, and the same pain-monitoring rule. The full prompt is provided in Supplementary Appendix S1.

For each model, the prompt was submitted in three independent sessions. An experienced musculoskeletal physiotherapist (>5 years' experience) synthesized a single master program by prioritizing exercises consistently proposed across runs and retaining only exercises meeting predefined criteria for safety, feasibility, and alignment with SAPS rehabilitation principles. A small number of unsafe or infeasible suggestions (eg, advanced plyometric drills) were removed or simplified (examples in Supplementary Table S2). Treating physiotherapists received protocol training, and session duration/structure were standardized across groups.

Outcome Measures

Outcomes were assessed at baseline (week 0) and post-intervention (week 6) by the same blinded assessor. Co-primary outcomes were: pain intensity (NPRS; 0–10) at rest, during activity, and at night; shoulder pain and disability (SPADI; 0–100); and health-related quality of life (SF-12 physical and mental component summary scores). Secondary outcomes included QuickDASH, active shoulder range of motion (goniometer), isometric strength (Kinvent K-Force; Newtons), kinesiophobia (TSK), shoulder function (ASES), pain catastrophizing (PCS), emotional status (HADS), and disease-specific quality of life (WORC).

Sample Size

A priori sample size estimation (G*Power 3.1) for a 3 (group) × 2 (time) mixed ANOVA assumed a medium effect size ($f = 0.25$), $\alpha = 0.05$, and 80% power, requiring 54 participants (18 per group). To account for potential attrition of approximately 10–15%, the planned target was 60 participants (20 per group).

Statistical Analysis

Analyses were performed using IBM SPSS Statistics (v25.0). Continuous variables are presented as mean (SD) or median (IQR), and categorical variables as n (%). Baseline comparability was examined using one-way ANOVA (or Kruskal–Wallis, as appropriate) and chi-square tests.

Primary inference was based on 3 (group) × 2 (time) mixed ANOVA models for each outcome to test the group × time interaction. Holm–Bonferroni adjustment was applied across the co-primary outcomes. When interaction effects were significant, post hoc comparisons were performed using pairwise tests on change scores with appropriate multiplicity control. Effect sizes are reported as partial η^2 for ANOVA effects and Hedges g (or Cohen's d) for pairwise differences. All tests were two-tailed with $\alpha = 0.05$. Analyses followed an available-case approach; attrition and missingness are reported.