Trial title: Effect of ChatGPT and DeepSeek-Guided Rehabilitation on Clinical Outcomes in Individuals With Subacromial Pain Syndrome

ClinicalTrials.gov: NCT07148687

Version date: 24 February 2026

Design: Triple-masked (participant-, diagnosing clinician-, and outcome assessor–blinded), 3-arm, parallel-group randomized controlled trial

Setting: Outpatient Physical Medicine and Rehabilitation clinic, Kırşehir Ahi Evran University Training and Research Hospital (Kırşehir, Türkiye)

Ethics and Registration

The trial was prospectively registered at ClinicalTrials.gov (NCT07148687). Ethical approval was obtained from Selçuk 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. Recruitment occurred from 25 Sep 2025 to 1 Dec 2025, and final follow-up assessments were completed on 20 Dec 2025.

Participants

Adults with shoulder pain were screened and referred when subacromial pain syndrome (SAPS) was diagnosed based on standardized history and physical examination consistent with rotator cuff–related shoulder pain. SAPS classification was supported using a test-cluster approach (painful arc, Hawkins–Kennedy, and resisted external rotation/infraspinatus strength test); SAPS was considered clinically supported when ≥2 tests were positive.

Inclusion criteria: age 18–65 years; clinical SAPS ≥4 weeks; sufficient active shoulder ROM to participate; able to attend supervised sessions; written consent.

Exclusion criteria: suspected/confirmed full-thickness rotator cuff tear, adhesive capsulitis, glenohumeral instability, or labral tear; cervical radiculopathy/neurological disorders/inflammatory rheumatic disease/fibromyalgia; shoulder surgery in past 6 months; systemic or neuromuscular disease; pregnancy; withdrawal/refusal of consent.

Randomization, Allocation Concealment, and Masking

After baseline assessment, participants were randomized (1:1:1) to:

Conventional physiotherapy + conventional exercise program

Conventional physiotherapy + ChatGPT-guided exercise program

Conventional physiotherapy + DeepSeek-guided exercise program

An independent administrative staff member (not involved in recruitment, treatment delivery, or outcome assessment) generated and implemented allocation using Randomizer.org with simple randomization. The

allocation list was stored as an encrypted file and remained inaccessible to recruiters, treating physiotherapists, and outcome assessors until baseline assessments were completed. Group assignment was revealed to the treating physiotherapist after baseline measurements.

Masking: Participants were informed that they would receive one of several supervised exercise-based rehabilitation programs for SAPS, but were not informed of their specific group assignment or the use of large language models. The diagnosing clinician and the outcome assessor were blinded to allocation. Treating physiotherapists were aware of group assignment but were not involved in outcome assessments or data analysis. Data management and statistical analysis were performed using coded group labels until primary analyses were completed. At week 6, the outcome assessor recorded a forced guess of group allocation to descriptively evaluate assessor blinding.

Interventions (6 weeks; 3 sessions/week; 18 sessions)

All participants completed supervised outpatient sessions (60 min; duration recorded). Attendance (0–18) was recorded. Pain response was monitored with a prespecified rule: acceptable pain during/after exercise ≤4/10 (NPRS) without meaningful increase in night pain or next-day baseline pain; if exceeded, exercises were regressed (reduced load/ROM/lever arm/sets/reps, added support).

Common physiotherapy modalities (all groups; identical co-interventions):

Moist hot pack 20 min

Conventional TENS 100 Hz, 60 μs; 20 min

Therapeutic ultrasound 1 MHz, 1.5 W/cm²; 5 min

Delivered 3 sessions/week for 6 weeks.

Exercise programs: All programs were structured into early/middle/late phases and progressed using the same pain-monitoring rule and performance criteria (technique, tolerance, symptom stability).

Conventional program: developed a priori by experienced musculoskeletal physiotherapists; mobility/active-assisted elevation, scapular stabilization, progressive rotator cuff strengthening, upper-quadrant stretching/strengthening.

LLM-guided programs:

ChatGPT group: ChatGPT-4o (OpenAI; web interface)

DeepSeek group: DeepSeek-R1 (DeepSeek; web interface)

A standardized English prompt requested a 6-week (18-session) supervised outpatient, exercise-only program for SAPS with basic clinic equipment, 3 phases, with exercise-specific dosage, symptom-based progression/regression, and the pain-monitoring rule. The identical prompt was submitted in 3 independent sessions per model. A musculoskeletal physiotherapist synthesized a single master program per model, retaining safe/feasible items aligned with SAPS rehabilitation principles and removing potentially unsafe/disproportionate

items (examples documented in Supplementary Table S2). Final session-by-session programs are provided in Supplementary Table S1; prompt text is provided in Supplementary Appendix S1.

Outcomes and Assessment Schedule

Outcomes were assessed at baseline (week 0) and post-intervention (week 6) by the same blinded assessor.

Primary outcome:

SPADI (0–100; higher = worse pain/disability)

Key secondary outcomes:

NPRS activity pain (0–10)

QuickDASH (0–100)

SF-12 Physical Component Summary (PCS)

Exploratory outcomes: NPRS rest/night; SF-12 MCS; active shoulder ROM (goniometer); isometric strength (Kinvent K-Force; Newtons); TSK; ASES; PCS; HADS; WORC. Internal rotation strength was not analyzed due to measurement standardization concerns in the handheld dynamometry setup.

Safety Monitoring

Adverse events were monitored at each supervised session and at week 6 using standardized questioning and spontaneous reporting. No intervention-related adverse events were reported.

Sample Size

A priori sample size estimation (G*Power 3.1) targeted detection of a clinically meaningful difference in SPADI, assuming medium effect size ($f=0.25$), $\alpha=0.05$, 80% power for a 3-group, 2-time-point repeated-measures framework, yielding a minimum of 54 participants (18/group). To allow for attrition and equal group sizes, the target was 75 participants (25/group).

Statistical Analysis Plan (SAP)

Analyses were conducted in IBM SPSS Statistics v25.0. Continuous variables are summarized as mean (SD) or median (IQR); categorical variables as n (%). Baseline comparability was assessed with 1-way ANOVA (or Kruskal–Wallis as appropriate) and chi-square tests.

Primary inference: For each outcome, between-group comparisons were based on change scores (week 6 minus baseline). Change scores were compared across 3 groups using 1-way ANOVA; Kruskal–Wallis was used when assumptions were not met. If an omnibus test was significant, pairwise post hoc comparisons of change scores were conducted with Bonferroni adjustment.

Multiplicity control: SPADI primary hypothesis tested at two-tailed α=0.05. Holm–Bonferroni adjustment was applied across prespecified key secondary outcomes (NPRS-activity, QuickDASH, SF-12 PCS). Exploratory outcomes were analyzed similarly but interpreted as hypothesis-generating without multiplicity adjustment.

Effect sizes: η^2 for omnibus between-group comparisons; Hedges g (or Cohen's d) for pairwise differences.

Missing data: Complete-case approach (participants with both baseline and week-6 data); missing outcomes not imputed. Participant flow and attrition were documented; discontinuations were due to scheduling conflicts (2 per group).

IPD/Data Sharing

Individual participant data (IPD) will not be shared due to ethical restrictions and institutional policy.