

RESEARCH

Open Access



A cross-sectional study on ChatGPT's alignment with clinical practice guidelines in musculoskeletal rehabilitation

Ertuğrul Safran^{1*} and Sefa Yildirim^{1,2}

Abstract

Background AI models like ChatGPT have the potential to support musculoskeletal rehabilitation by providing clinical insights. However, their alignment with evidence-based guidelines needs evaluation before integration into physiotherapy practice.

Objective To evaluate the performance of ChatGPT (GPT-4 model) in generating responses to musculoskeletal rehabilitation queries by comparing its recommendations with evidence-based clinical practice guidelines (CPGs).

Design This study was designed as a cross-sectional observational study.

Methods Twenty questions covering disease information, assessment, and rehabilitation were developed by two experienced physiotherapists specializing in musculoskeletal disorders. The questions were distributed across three anatomical regions: upper extremity (7 questions), lower extremity (9 questions), and spine (4 questions). ChatGPT's responses were obtained and evaluated independently by two raters using a 5-point Likert scale assessing relevance, accuracy, clarity, completeness, and consistency. Weighted kappa values were calculated to assess inter-rater agreement and consistency within each category.

Results ChatGPT's responses received the highest average score for clarity (4.85), followed by accuracy (4.62), relevance (4.50), and completeness (4.20). Consistency received the lowest score (3.85). The highest agreement (weighted kappa = 0.90) was observed in the disease information category, whereas rehabilitation displayed relatively lower agreement (weighted kappa = 0.56). Variability in consistency and moderate weighted kappa values in relevance and clarity highlighted areas requiring improvement.

Conclusions This study demonstrates ChatGPT's potential in providing guideline-aligned information in musculoskeletal rehabilitation. However, due to observed limitations in consistency, completeness, and the ability to replicate nuanced clinical reasoning, its use should remain supplementary rather than as a primary decision-making tool. While it performed better in disease information, as evidenced by higher inter-rater agreement and scores, its performance in the rehabilitation category was comparatively lower, highlighting challenges in addressing complex, nuanced therapeutic interventions. This variability in consistency and domain-specific reasoning underscores the need for further refinement to ensure reliability in complex clinical scenarios.

*Correspondence:
Ertuğrul Safran
esafran@bezmialem.edu.tr

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Clinical trial number Not applicable.

Keywords Artificial intelligence, Guideline adherence, Musculoskeletal rehabilitation, Natural Language processing

Background

Artificial intelligence (AI) has become an integral component of modern healthcare, offering innovative tools to enhance diagnostics, clinical decision-making, and patient care efficiency [1, 2]. Among AI applications, natural language processing (NLP) models such as ChatGPT have garnered significant attention due to their ability to generate human-like responses to complex medical inquiries [3, 4]. These models are trained on vast datasets derived from diverse information sources, enabling them to produce contextually relevant and coherent outputs. However, ensuring the alignment of AI-generated content with evidence-based clinical practice guidelines (CPGs) is paramount to establishing their reliability in medical practice [5].

Recent studies have demonstrated the potential of AI tools in various healthcare domains, including disease diagnosis, treatment planning, and patient education. In musculoskeletal physical therapy, ChatGPT has shown notable promise, achieving up to 80% compliance with CPGs [6–8]. Additionally, it has been proposed as a means to optimize workflows in resource-limited settings where rapid access to accurate clinical information is crucial [9]. Despite these advantages, concerns persist regarding the consistency and reliability of AI-generated responses, particularly in complex clinical scenarios that demand nuanced clinical reasoning [10, 11]. While AI models like ChatGPT can generate well-structured and plausible content, they may also produce inaccurate or misleading information, underscoring the necessity for critical evaluation by healthcare professionals [12].

Musculoskeletal rehabilitation, a cornerstone of physical therapy, plays a crucial role in restoring function, alleviating pain, and enhancing the quality of life for individuals with musculoskeletal disorders [13]. For physiotherapists, adherence to evidence-based CPGs ensures the effectiveness and standardization of treatments, thereby optimizing patient outcomes and maintaining professional accountability [14, 15]. Integrating AI systems such as ChatGPT presents an opportunity to support physiotherapists by providing rapid, guideline-consistent recommendations, particularly in time-constrained or resource-limited settings [9]. Previous research has predominantly examined ChatGPT's responses in diagnosing specific musculoskeletal pathologies or surgical conditions, such as lumbar radicular pain [11], degenerative spondylolisthesis [7], as well as surgical procedures like anterior cruciate ligament (ACL) reconstruction [10], and rotator cuff repairs [16]. Additionally, its application in clinical decision support

has been explored [8]. However, physiotherapy practice encompasses not only diagnostics but also detailed patient assessment and evidence-based rehabilitation interventions, requiring comprehensive clinical reasoning beyond mere diagnostic capabilities. Yet, the current literature lacks an in-depth evaluation of ChatGPT's capability to provide physiotherapy-specific recommendations, particularly regarding assessment strategies and rehabilitation interventions aligned with CPGs. Unlike previous studies that focused primarily on medical diagnoses or surgical scenarios, our study specifically investigates how ChatGPT performs in physiotherapy-specific domains such as assessment and rehabilitation—core components of physiotherapeutic management in musculoskeletal care. Therefore, the novelty of this study lies in its focused evaluation of ChatGPT's performance in three physiotherapy-specific domains: disease information, patient assessment, and evidence-based rehabilitation practices—areas that have been less explored in previous research. By identifying areas where AI aligns with or diverges from established guidelines, this research seeks to inform the potential role of ChatGPT in clinical education and decision support within musculoskeletal rehabilitation.

Methods

Study design

This study employed a cross-sectional observational design to compare recommendations generated by evidence-based CPGs with those provided by ChatGPT's GPT-4 model for musculoskeletal conditions. A total of twenty questions were systematically developed by two physiotherapists, each with over eight years of clinical experience in musculoskeletal rehabilitation. The development process began by identifying the most commonly encountered musculoskeletal conditions and clinical decision-making challenges in physiotherapy. Only conditions with established and accessible CPGs were included to ensure objective benchmarking. The questions were then categorized into three key domains: disease information, patient assessment, and rehabilitation. Additionally, efforts were made to ensure a balanced anatomical distribution—seven questions for the upper extremity, nine for the lower extremity, and four for the spine. A purposive sampling approach was adopted, and the number of questions was determined based on the need to cover a diverse range of physiotherapy-relevant scenarios while ensuring feasibility for expert rating and statistical analysis.

Each question was submitted in a new session to avoid memory retention or contextual influence between prompts. While ChatGPT does not learn from individual prompts in real time, periodic updates released by OpenAI may influence model performance; therefore, specifying the data collection period enhances reproducibility. The model was explicitly instructed to respond from the perspective of a physiotherapist to simulate clinical reasoning and decision-making comparable to that of an experienced practitioner. All responses were recorded verbatim to ensure accuracy and consistency during subsequent analysis.

Outcome measurement

The responses generated by ChatGPT were independently assessed by two musculoskeletal physiotherapists, each with over eight years of clinical experience. The two physiotherapists who evaluated the responses were also involved in developing the questions. To mitigate potential bias, each question was discussed and evaluated independently by both raters using a predefined Likert scale, and inter-rater agreement was statistically analyzed using weighted kappa coefficients.

A 5-point Likert scale (ranging from 1 = Strongly Disagree to 5 = Strongly Agree) was used to evaluate the following predefined response characteristics:

Relevance Does the response directly address the question posed?

Accuracy Is the information provided accurate and consistent with current clinical practice guidelines?

Clarity Is the response well-organized and easy to comprehend?

Completeness Does the response comprehensively address all aspects of the question?

Consistency How consistent are the responses generated by ChatGPT when the same question is posed multiple times?

To facilitate comparative analysis, Table 1 presents the questions, CPG-derived answers, and ChatGPT's verbatim responses, offering a comprehensive overview of the dataset.

Statistical analysis

The average scores from the two raters' Likert scale evaluations were computed. To assess inter-rater agreement, weighted kappa coefficients were calculated for each evaluation criterion: relevance, accuracy, clarity, completeness, and consistency. Furthermore, the dataset was stratified into three primary categories—disease

information, assessment, and rehabilitation—to examine inter-rater agreement within each category. Weighted kappa values were computed both for the overall dataset and within these specific domains.

All statistical analyses, including weighted kappa computations, were conducted using IBM SPSS Statistics version 25.

Results

A comparative analysis was conducted on the 20 predefined questions and their corresponding ChatGPT-generated responses, with the original questions, ChatGPT responses, and guideline-based answers presented in Table 1.

Section-wise average scores revealed that clarity received the highest mean score of 4.85, suggesting that ChatGPT's responses were well-structured and easy to comprehend. In contrast, consistency obtained the lowest mean score of 3.85, indicating notable variability in responses when the same question was posed multiple times. Additional scores included accuracy at 4.62, relevance at 4.50, and completeness at 4.20.

Inter-rater agreement, evaluated using weighted kappa values, varied across the assessed criteria. Consistency exhibited the highest inter-rater agreement, with a weighted kappa of 0.88, followed by completeness ($\kappa=0.71$) and accuracy ($\kappa=0.57$). Moderate agreement was observed for relevance ($\kappa=0.45$), whereas clarity exhibited the lowest agreement, with a weighted kappa of 0.27. The question-wise average scores and weighted kappa values for each criterion are detailed in Table 2.

Further subgroup analysis based on question categories (disease information, assessment, and rehabilitation) revealed notable differences. The disease information category exhibited the highest agreement, with a weighted kappa of 0.90, suggesting ChatGPT provided consistent and accurate responses in this domain. In contrast, the rehabilitation category showed lower agreement, with a weighted kappa of 0.56, indicating greater variability in ChatGPT's performance when addressing rehabilitation-related queries. The average scores and weighted kappa values for each category are summarized in Table 3.

Discussion

The findings of this study provide valuable insights into both the potential and limitations of ChatGPT's GPT-4 model in addressing queries related to musculoskeletal rehabilitation. By systematically comparing ChatGPT's responses with evidence-based CPGs, this study highlights the model's ability to generate clinically relevant and accurate information while also identifying areas requiring improvement. ChatGPT demonstrated notable strengths in clarity and accuracy, as reflected in its high ratings. The highest mean score for clarity underscores

Table 1 Questions with corresponding ChatGPT and guideline-based responses

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
What are the intrinsic risk factors for carpal tunnel syndrome?	The intrinsic risk factors with the strongest link to carpal tunnel syndrome are obesity, age, and female sex. The risk increases linearly with body mass index and age. The risk doubles in individuals with a body mass index greater than 30 kg/m ² and in those over the age of 50. Female sex increases the risk by 1.5 to 4 times. Intrinsic risk factors linked to carpal tunnel syndrome, but to a lesser extent include diabetes mellitus, osteoarthritis previous musculoskeletal disorders, estrogen replacement therapy, cardiovascular disease risk factors, hypothyroidism, family history of carpal tunnel syndrome, lack of physical activity, wrist ratio greater than 0.70, wrist-palm ratio greater than 0.39, a short, wide hand, and short stature [21].	Intrinsic risk factors for carpal tunnel syndrome (CTS) include anatomical variations (e.g., a smaller carpal tunnel), female sex (due to generally narrower carpal tunnels), obesity, hormonal changes (such as during pregnancy or menopause), and certain medical conditions like diabetes, hypothyroidism, and rheumatoid arthritis. These factors may increase pressure within the carpal tunnel or make the median nerve more susceptible to compression.
How is neck pain classified based on clinical presentation?	Neck pain is classified as neck pain with limited mobility, neck pain with movement coordination disorders, neck pain with headache, and neck pain with radiating pain [22].	Neck pain is commonly classified into four categories based on clinical presentation: 1. Neck pain with mobility deficits – Characterized by limited range of motion and localized neck pain, often linked to mechanical dysfunction or poor posture. 2. Neck pain with movement coordination impairments – Includes whiplash-associated disorders, often with instability, motor control deficits, and hypersensitivity. 3. Neck pain with headaches (cervicogenic) – Associated with headaches triggered by neck movements or sustained postures, often linked to upper cervical dysfunction. 4. Neck pain with radiating pain - Involves radicular symptoms such as pain, numbness, or weakness radiating into the arm, often due to nerve root compression.
What are the common characteristics, symptoms, and recurrence rates of a hamstring strain injury (HSI)?	An hamstring strain injury can occur anywhere along the length of the muscle, but occurs most frequently in the proximal biceps femoris at the musculotendinous junction. At the time of injury, an individual experiences a sudden, sharp pain in the posterior thigh. Additionally, an audible or palpable popping sensation often occurs during an activity that overloads and/or overstretches the hamstring muscle. The individual may stop the event or activity due to the pain and limited function. The recurrence rate of HSI ranges between 13.9% and 63.3% when followed over the same and subsequent seasons. Also, injuries with more extensive myofascial damage extending into the tendon are more prone to reinjury and delayed return to play [23].	Hamstring strain injuries are characterized by sudden onset pain in the posterior thigh, often occurring during activities involving rapid acceleration or deceleration, such as sprinting or kicking. Symptoms include localized tenderness, swelling, bruising, and reduced strength or flexibility. There are two main types: sprint-related strains (more severe, affecting the long head of the biceps femoris) and stretch-related strains (associated with slower movements like high kicks). Recurrence rates are high, ranging from 12–63%, often due to inadequate rehabilitation, poor neuromuscular control, or returning to activity prematurely. Strength imbalances and previous injury are key risk factors.
What are the risk factors for hip fractures in older individuals?	Individual risk factors for hip fracture in individuals 65 years of age and older have been extensively explored. In addition to reduced bone mineral density, a range of other factors have been identified, including older age, female sex, low body mass index, ethnicity, being postmenopausal without estrogen replacement therapy, femoral geometry, previous hip fracture, smoking, vitamin D deficiency, low dietary calcium intake, and hypervitaminosis A, among many others [24].	Risk factors for hip fractures in older individuals include osteoporosis (weakened bone density), advanced age (due to decreased bone mass and muscle strength), and female sex (postmenopausal bone loss). Other factors include poor balance and reduced mobility, which increase fall risk, as well as chronic conditions like arthritis, diabetes, or neurological disorders. Medications that cause dizziness or sedation, nutritional deficiencies (e.g., calcium and vitamin D), and lifestyle factors like sedentary behavior or smoking also contribute. Environmental hazards, such as loose rugs or poor lighting, further increase the risk of falls leading to hip fractures.

Table 1 (continued)

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
What are the risk factors for achilles tendon disorder?	Risk factors are commonly classified as those that are intrinsic or extrinsic to the individual. Intrinsic risk factors that have been associated with Achilles tendon disorders include abnormal ankle dorsiflexion range of motion, abnormal subtalar joint range of motion, decreased ankle plantar flexion strength, increased foot pronation, and associated diseases. Extrinsic risk factors that have been associated with Achilles tendinopathy include training errors, environmental factors, and faulty equipment [25].	Risk factors for Achilles tendon disorders include intrinsic factors such as reduced flexibility, poor ankle dorsiflexion, decreased calf muscle strength, and abnormal foot biomechanics (e.g., overpronation). Extrinsic factors include sudden increases in activity level, repetitive overuse from high-impact sports like running or jumping, and wearing inappropriate footwear. Other contributing factors include age-related tendon degeneration, male sex, obesity, and systemic conditions like diabetes or inflammatory arthritis. Certain medications, such as fluoroquinolones or corticosteroids, are also associated with increased risk of tendon disorders.
Assessment		
What provocative tests are used to evaluate people with carpal tunnel syndrome?	Provocative tests used to evaluate carpal tunnel syndrome are the Phalen Test, Tinel Sign, Carpal Compression Test, Reverse Phalen Test, Upper Extremity Neurodynamic Test and Scratch-Collapse test [21].	Common provocative tests used to evaluate carpal tunnel syndrome (CTS) include: 1. Phalen's Test: The patient flexes their wrists to 90 degrees and holds the position for 60 s. A positive test reproduces symptoms like tingling or numbness in the median nerve distribution. 2. Tinel's Sign: Gentle tapping over the carpal tunnel elicits tingling or "electric shock" sensations in the median nerve distribution. 3. Durkan's Compression Test: Direct pressure is applied over the carpal tunnel for up to 30 s. Symptom reproduction indicates a positive result. 4. Hand Elevation Test: Raising both hands above the head for one to two minutes may provoke symptoms. These tests help confirm median nerve compression but should be combined with history and diagnostic imaging for accuracy.
What key historical and physical examination findings are most indicative of lumbar disc herniation with associated radiculopathy?	Manual muscle testing, sensory testing, supine straight leg raise, Lasegue sign, and crossed Lasegue sign are suggested for use with recommendation grade A in the diagnosis of lumbar disc herniation with radiculopathy. The supine straight leg raise is suggested for use with recommendation grade B in diagnosing of lumbar disc herniation with radiculopathy. There is an insufficient evidence to make a recommendation for or against the use of the cough impulse test, Bell test, hyperextension test, femoral nerve stretch test, slump test, lumbar range of motion, or absence of reflexes in diagnosing lumbar disc herniation with radiculopathy [26].	In lumbar disc herniation with radiculopathy, the history typically includes low back pain radiating into the leg along a dermatomal pattern, often described as sharp, burning, or electric-like. Pain is often exacerbated by activities that increase intradiscal pressure, such as sitting, bending, coughing, or sneezing. Patients may also report numbness, tingling, or weakness in the affected leg. On physical examination, findings may include: • Positive straight leg raise (SLR) test: Pain radiates down the leg when the affected leg is raised. • Decreased reflexes: Commonly in the Achilles (S1) or patellar (L4) reflexes. • Sensory deficits: Numbness in dermatomal patterns (e.g., L5 or S1). • Motor weakness: Weakness in ankle dorsiflexion (L5) or plantarflexion (S1). • Antalgic posture or gait: To reduce nerve root tension. These findings correlate with the affected nerve root and guide diagnosis.
Which conditions should physiotherapists consider in the differential diagnosis of lateral elbow pain?	The conditions to consider with a differential diagnosis of lateral elbow pain, although not all-inclusive, include the following: • Cervical radiculopathy • Radial tunnel syndrome • Posterior interosseous syndrome • Plica syndrome • Radio-capitellar chondromalacia • Posterolateral rotatory instability • Myofascial trigger points in the wrist extensors [27]	In the differential diagnosis of lateral elbow pain, physical therapists should consider lateral epicondylalgia (tennis elbow) as the most common cause, but also evaluate for radial tunnel syndrome (nerve entrapment causing pain without significant weakness), posterior interosseous nerve syndrome (with motor deficits), cervical radiculopathy (referred pain from C5-C6), osteoarthritis of the elbow, ligamentous injuries (e.g., lateral collateral ligament sprain), and intra-articular pathology (e.g., synovitis or loose bodies). Additionally, systemic conditions like rheumatoid arthritis or referred pain from the shoulder should be considered if the presentation is atypical.

Table 1 (continued)

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
What diagnostic criteria should clinicians use to identify patellofemoral pain?	Clinicians should use criteria such as the presence of retropatellar or peripatellar pain, recurrence of retropatellar or peripatellar pain with squatting, climbing stairs, sitting for prolonged periods, or other functional activities that load the patellofemoral joint in a flexed position, and exclusion of all other conditions that may cause anterior knee pain, including tibiofemoral pathologies. Additionally, clinicians may use patellar tilt testing in the presence of hypomobility to support the diagnosis of patellofemoral pain [28].	Clinicians should diagnose patellofemoral pain based on the following criteria: 1. Anterior knee pain: Pain localized to the peripatellar or retropatellar region, often described as dull or aching. 2. Activity-related symptoms: Pain is exacerbated by activities that load the patellofemoral joint, such as squatting, running, climbing stairs, prolonged sitting (the “theater sign”), or jumping. 3. Absence of other specific knee pathology: Ruling out conditions like ligament tears, meniscal injuries, or patellar instability through clinical evaluation and imaging if necessary. 4. Physical examination findings: o Pain during patellar compression or palpation of the patellar facets. o Positive Clark’s test (patellar grind test) or pain during resisted quadriceps contraction. o Possible maltracking or poor patellar alignment during movement. These criteria emphasize pain location, activity provocation, and exclusion of differential diagnoses.
How should physical therapists diagnose the International Classification of Disease (ICD) category of plantar fasciitis and the associated International Classification of Functioning, Disability and Health (ICF) impairment-based category of heel pain using the patient’s history and physical examination findings?	<ul style="list-style-type: none"> Plantar medial heel pain: most noticeable with initial steps after a period of inactivity, but also worse following prolonged weight-bearing Heel pain precipitated by a recent increase in weight-bearing activity Pain with palpation of the proximal insertion of the plantar fascia Positive windlass test Negative tarsal tunnel tests Limited active and passive talocrural joint dorsiflexion range of motion (ROM) Abnormal Foot Posture-6 (FPI-6) score High body mass index in nonathletic individuals [29] 	<p>Patient History:</p> <ol style="list-style-type: none"> Location of pain: Pain is usually at the medial plantar heel, often at the insertion of the plantar fascia. Onset and progression: Symptoms often develop gradually and are worse with the first steps in the morning or after prolonged inactivity (post-static dyskinesia). Aggravating factors: Pain increases with prolonged standing, walking, or running and improves with rest. Risk factors: History of recent changes in activity levels, inappropriate footwear, or prolonged weight-bearing activities. <p>Physical Examination:</p> <ol style="list-style-type: none"> Palpation: Localized tenderness at the medial calcaneal tubercle or along the plantar fascia. Windlass test: Pain reproduced during dorsiflexion of the great toe, which stretches the plantar fascia. Passive dorsiflexion: Limited dorsiflexion of the ankle, often associated with tightness in the gastrocnemius or soleus muscles. <p>Exclusion of other causes: No signs of systemic conditions (e.g., rheumatoid arthritis or infection) or neural involvement (e.g., tarsal tunnel syndrome).</p>
Rehabilitation		
What are the strength and motor control training recommendations in the rehabilitation program after anterior cruciate ligament (ACL) reconstruction?	Strength and motor control training recommendations in the rehabilitation program are plyometric and agility training, motor control training, isotonic and isokinetic training, concentric and eccentric training, open and closed kinetic chain exercises, core stability and aquatic therapy [30].	ACL reconstruction, strength and motor control training should focus on progressive, structured exercises tailored to recovery stages. In the early phase (0–6 weeks), emphasize quadriceps activation (e.g., isometric exercises) and restoring normal gait. In the intermediate phase (6–12 weeks), incorporate progressive resistance training targeting quadriceps, hamstrings, and hip musculature, along with neuromuscular control exercises (e.g., balance and proprioceptive training). In the late phase (12+ weeks), focus on sport-specific drills, dynamic stability, and plyometric exercises to improve strength, motor control, and functional performance. Progressions should consider pain, swelling, and achieving symmetry between limbs to reduce re-injury risk.
Which interventions can be used in the treatment of patients with hip osteoarthritis?	Flexibility, strengthening, endurance exercises, and manual therapy can be used in the treatment of patients with hip osteoarthritis with evidence grade A. Additionally, modalities and patient education combined with exercises can be used at level B evidence, while functional, gait and balance training and weight loss can be used at level C evidence [31].	Treatment for hip osteoarthritis includes a combination of exercise therapy, manual therapy, and patient education. Strengthening exercises targeting the hip abductors, extensors, and flexors, along with aerobic and flexibility training, are essential to improve joint stability and mobility. Manual therapy techniques, such as joint mobilizations and soft tissue release, can reduce pain and improve range of motion. Patient education should focus on weight management, activity modification, and joint protection strategies. Additionally, assistive devices (e.g., canes) and modalities like heat or ultrasound may provide symptomatic relief. Advanced cases may require referral for pharmacological or surgical interventions.

Table 1 (continued)

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
Which interventions are performed in patients with neck pain and limitation of movement?	The timing of the introduction of shoulder ROM exercises by physical therapists may be delayed up to 4 weeks without negatively impacting patient-reported outcomes in patients with glenohumeral osteoarthritis who have undergone total shoulder arthroplasty (Moderate quality of evidence) [32].	In the acute phase, clinicians should provide thoracic manipulation, a program of neck ROM exercises, and scapulothoracic and upper extremity strengthening to increase compliance with the program. Additionally, clinicians may provide cervical manipulation and/or mobilization. In the subacute phase, clinicians should provide neck and shoulder girdle endurance exercises and thoracic manipulation and cervical manipulation and/or mobilization. In the chronic phase, clinicians should provide a multimodal approach of the following: thoracic manipulation and cervical manipulation or mobilization, mixed exercise for cervical/scapulothoracic regions: neuromuscular exercise (eg, coordination, proprioception, and postural training), stretching, strengthening, endurance training, aerobic conditioning, and cognitive affective elements. Additionally, dry needling, laser, or intermittent mechanical/manual traction may be used [22].
When should range of motion exercises begin for patients with glenohumeral osteoarthritis following total shoulder arthroplasty?	The timing of the introduction of shoulder ROM exercises by physical therapists may be delayed up to 4 weeks without negatively impacting patient-reported outcomes in patients with glenohumeral osteoarthritis who have undergone total shoulder arthroplasty (Moderate quality of evidence) [32].	ROM exercises for patients with glenohumeral osteoarthritis following total shoulder arthroplasty typically begin within the first week post-surgery, as long as the surgical approach and healing allow. In the early phase (0–4 weeks), passive ROM exercises are prioritized to protect the joint and soft tissues while preventing stiffness. Progression to active-assisted and active ROM exercises usually occurs in the intermediate phase (4–6 weeks), depending on patient tolerance and surgeon guidelines. Careful monitoring is essential to avoid overloading the joint or compromising the surgical repair.
Which interventions can be used for Heel Pain – Plantar Fasciitis?	Manual therapy (A), stretching (A), taping (A), foot orthoses (B-C), night splints (A), ultrasound (A), low level laser therapy (B), phonophoresis (C), electrotherapy (D), education and counseling for weight loss (E), therapeutic exercise and neuromuscular re-education (B), dry needling (B) [29].	Interventions for heel pain from plantar fasciitis focus on reducing pain, improving flexibility, and restoring function. Stretching exercises targeting the plantar fascia and calf muscles are highly effective. Strengthening exercises, particularly for the intrinsic foot muscles, can improve arch support. Manual therapy, including myofascial release and joint mobilizations, helps reduce stiffness and pain. Additional interventions include night splints, orthotics to offload the plantar fascia, and taping for short-term symptom relief. Patients should also be educated on activity modification, appropriate footwear, and gradual return to activity. For persistent cases, shockwave therapy or corticosteroid injections may be considered.
What should be the patient education strategies of physiotherapists in the management of patients with adhesive capsulitis?	Clinicians may utilize patient education that describes natural course of the disease, promotes activity modification to encourage functional, pain-free ROM, and matches the intensity of stretching to the patient's current level or irritability [33].	Patient education for adhesive capsulitis should focus on explaining the condition's phases (freezing, frozen, and thawing) to set realistic expectations for recovery, which can take months to years. Reassure patients that the condition is self-limiting and improves over time, but emphasize the importance of staying active within pain-free limits to maintain joint mobility. Educate on gentle stretching and avoiding aggressive movements that may worsen pain. Discuss the role of pain management strategies, such as heat or cold therapy, and the potential benefits of corticosteroid injections if recommended by their physician. Encourage adherence to home exercise programs and emphasize patience and consistency in rehabilitation.

Table 1 (continued)

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
What are the joint mobilization/manipulation interventions used for Lateral Elbow Tendinopathy (LET)?	<p>A preponderance of level 2 evidence demonstrates that lateral glide MWM technique to the elbow, Mill's manipulation technique, or regional mobilization techniques all demonstrate a positive effect compared with a placebo or control group on pain, pain-free grip strength, and function in the short term. Clinicians should use local elbow joint manipulation or mobilization techniques to reduce pain and increase pain-free grip strength in individuals with LET, as a stand-alone or adjunctive treatment in improving short-term outcomes for those who can tolerate the specific technique (B). Clinicians may use manipulation or mobilization techniques directed at the cervical spine, thoracic spine, and/or wrist as an adjunct to local treatment for short-term pain relief in individuals with LET when impairments in those regions are identified (C).</p> <p>Gaps in Knowledge</p> <p>The current literature does not address which type of joint mobilization technique is superior to others. The midterm and long-term outcomes of joint mobilization on outcomes in LET are unknown. Joint mobilizations/manipulations may contribute to diminishing pain and improving motor function via neurophysiologic mechanisms. Although these mechanisms are not completely understood, joint mobilizations/manipulations may involve reflex inhibition of pain mediated through joint mechanoreceptors [27].</p>	<p>Joint mobilization and manipulation for lateral elbow tendinopathy aim to reduce pain, improve joint mobility, and promote tissue healing. Common interventions include:</p> <ol style="list-style-type: none"> 1. Lateral glide mobilization: A sustained or oscillatory lateral glide of the humeroulnar joint, often combined with gripping exercises to enhance pain-free grip strength. 2. Radial head mobilization: Anterior or posterior mobilizations of the radial head to address stiffness and improve forearm movement. 3. Cervical and thoracic mobilizations or manipulations: Techniques targeting the cervical spine (e.g., C5-C6) and thoracic spine to address referred pain and improve overall upper extremity mechanics. 4. Soft tissue mobilization: Targeting the extensor muscle group and common extensor tendon to reduce tension and pain. <p>These techniques are often integrated into a broader rehabilitation program including strengthening and neuromuscular retraining.</p>
When should rehabilitation begin for patients with distal radius fractures (DRF)?	<p>Clinicians should initiate early therapy that consists of hand, wrist, and shoulder active range of motion (AROM) exercises along with light daily activity within the first 3 weeks after a surgically repaired DRF to improve short-term (up to 3 months) outcomes for pain, wrist AROM, grip strength, and functional, and long-term (≥ 6 months) outcomes for wrist AROM and grip strength (A). Clinicians should initiate submaximal progressive strengthening, such as towel and putty squeezing and light-load gripping exercises at 2 weeks following a surgically repaired DRF or during the second week of cast immobilization (only the uncomplicated individuals with stable DRF, satisfactory radius-ulna articular alignment, and no ulnar-sided pain) to improve short-term (up to 6 months) outcomes for pain, wrist AROM, grip strength, and functional capacity with negligible risk of compromising proper fracture healing (B) [34].</p>	<p>Rehabilitation for patients with distal radius fractures should typically begin as soon as the acute inflammatory phase subsides, usually within 1–2 weeks after injury or surgical fixation, provided the fracture is stable. Early rehabilitation focuses on reducing pain and swelling through gentle range-of-motion exercises for the fingers, wrist, and forearm to prevent stiffness. For immobilized fractures, therapy starts after cast or splint removal (4–6 weeks), emphasizing gradual wrist and forearm mobility, strength, and functional use. Early intervention is critical for optimal outcomes but must respect fracture healing timelines and surgeon recommendations.</p>
What should be the patient education strategies of physiotherapists in the management of patients with low back pain (LBP)?	<p>Physical therapists may use active education strategies rather than passive strategies (i.e., providing access to educational materials only). Active education strategies include one-on-one education on the biopsychosocial contributors to pain and self-management techniques, such as remaining active, pacing strategies, and back-protection techniques. Physical therapists may also incorporate counseling on the favorable natural history of acute LBP as part of the education strategy (B) [35].</p>	<p>Patient education for managing low back pain should focus on reassurance, promoting active recovery, and addressing misconceptions. Physiotherapists should educate patients about the benign and often self-limiting nature of most low back pain, emphasizing the importance of staying active and avoiding prolonged bed rest. They should explain the role of posture, ergonomics, and regular exercise in prevention and management while encouraging gradual return to normal activities. Addressing fears and beliefs about pain and movement (e.g., fear of harm) is crucial to reduce kinesiophobia. Providing strategies for managing acute flare-ups, such as pacing and relaxation techniques, is also beneficial.</p>

Table 1 (continued)

Questions	Guidelines Recommendation	ChatGPT Response
Disease Informations		
Which interventions can be performed for an ankle ligament sprain?	For acute/protected motion phase; Early weight bearing with support (A), manual therapy (A), cryotherapy (A), diathermy (C), electrotherapy (D), low- level laser therapy (D), and therapeutic exercises (A). For progressive loading/sensorimotor training phase; Manual therapy (A), therapeutic exercise and activities (C), and sport - related activity training (C) [36].	Interventions for an ankle ligament sprain focus on reducing pain and swelling, restoring mobility and strength, and preventing recurrence. In the acute phase, use the RICE protocol (Rest, Ice, Compression, Elevation) alongside protected weight-bearing with braces or taping if needed. Gentle range-of-motion exercises, such as ankle circles or dorsiflexion-plantarflexion movements, should be introduced early. In the subacute phase, progress to strengthening exercises (e.g., resistance bands) and proprioceptive training, such as balance exercises on unstable surfaces. The rehabilitation phase includes functional activities like hopping, agility drills, and sport-specific movements to restore full performance. Education on proper footwear and prevention strategies is essential.

Table 2 Question-wise average scores with weighted kappa values

Question	Relevance	Accuracy	Clarity	Completeness	Consistency
Q1	4.00	4.00	5.00	3.50	4.00
Q2	5.00	5.00	5.00	5.00	5.00
Q3	5.00	3.50	5.00	4.00	3.00
Q4	5.00	5.00	5.00	5.00	5.00
Q5	5.00	5.00	5.00	5.00	5.00
Q6	4.50	5.00	5.00	3.00	3.00
Q7	4.00	5.00	4.50	3.50	3.00
Q8	4.50	4.00	5.00	3.50	3.00
Q9	5.00	5.00	5.00	5.00	4.00
Q10	5.00	4.50	4.50	4.00	3.00
Q11	4.00	4.00	5.00	4.00	4.00
Q12	4.50	5.00	5.00	4.00	4.00
Q13	4.00	4.00	4.00	3.50	4.00
Q14	4.00	4.50	5.00	5.00	4.00
Q15	4.00	5.00	4.50	4.00	4.00
Q16	5.00	5.00	5.00	5.00	4.00
Q17	4.50	5.00	5.00	4.00	3.50
Q18	4.50	4.50	5.00	4.50	4.00
Q19	4.50	5.00	5.00	5.00	4.00
Q20	4.00	4.50	4.50	3.50	3.50
Average	4.50	4.62	4.85	4.20	3.85
Weighted kappa	0.45	0.57	0.27	0.71	0.88

the AI model's ability to present well-structured and easily comprehensible information. Another noteworthy strength was the model's performance in relevance and accuracy (4.62), reflecting its ability to provide contextually appropriate and largely evidence-aligned responses. Furthermore, the high weighted kappa values for consistency and completeness in the disease information category suggest that ChatGPT's responses are not only comprehensive but also exhibit minimal variability when addressing well-defined queries. Despite these strengths, the lowest scores for completeness and consistency highlight areas in need of improvement. Variability in responses to repeated queries raises concerns about reliability, particularly in scenarios requiring precise and consistent recommendations. The rehabilitation category exhibited the lowest weighted kappa value, suggesting challenges in addressing the nuanced and complex aspects of therapeutic interventions.

These results align with prior research emphasizing the strengths and challenges of AI-driven decision-support tools in healthcare [17, 18]. For instance, Bilika et al. explored ChatGPT's application in physiotherapy decision-making, underscoring the importance of its cautious use and the need for informed judgment in clinical practice [19]. Such clarity enhances the usability of ChatGPT as an educational tool for physiotherapists and other healthcare professionals, particularly in environments where rapid access to clear and concise information is critical. These results are consistent with earlier studies demonstrating ChatGPT's adherence to clinical

Table 3 Category-wise average scores and weighted kappa values

Category	Relevance	Accuracy	Clarity	Completeness	Consistency	Weighted Kappa
Disease Information (Question 1–5)	4.80	4.50	5.00	4.50	4.40	0.90
Assessment (Question 6–13)	4.60	4.70	4.80	3.80	3.20	0.79
Rehabilitation (Question 14–20)	4.30	4.65	4.80	4.25	3.90	0.56

guidelines in musculoskeletal care, with reported compliance rates reaching up to 80% in similar contexts [8]. This finding underscores ChatGPT's potential as a decision-support tool, particularly for routine clinical queries. This may be attributed to limitations in ChatGPT's training data, particularly concerning specialized and context-sensitive rehabilitation scenarios. For example, Gianola et al. found inconsistencies in ChatGPT's recommendations for lumbosacral radicular pain compared to CPGs, raising concerns about accuracy and internal consistency [11]. Similarly, Sawamura et al. [12] concluded that although ChatGPT can generate accurate responses, its reference reliability and selection remain notable limitations. Consequently, they emphasized the necessity of cautious use, as ChatGPT is not entirely dependable for clinical decision-making. Additionally, the moderate inter-rater agreement for relevance and the low agreement for clarity emphasize the need for refining ChatGPT's training algorithms to enhance alignment with evidence-based guidelines. Such variability in content quality has also been documented in studies examining AI-generated responses to complex medical queries [20].

Limitations

This study has several limitations that should be acknowledged. First, only two assessors were involved in the evaluation, which may limit the generalizability of inter-rater agreement results. Second, the assessors also developed the questions, introducing potential bias despite independent scoring and statistical analysis of agreement. Third, while the assessment criteria were informed by existing literature and expert consensus, the rating tool itself has not been psychometrically validated. Additionally, the study was limited to a fixed set of 20 questions, which, although diverse, may not fully represent the range of real-world clinical scenarios. Finally, the findings reflect ChatGPT's performance during a specific time window and may not apply to future updates of the model.

Implications for clinical practice

The findings of this study carry significant implications for the integration of AI tools like ChatGPT into musculoskeletal rehabilitation. While the model demonstrates promise in delivering clear and accurate information, its limitations necessitate careful implementation. ChatGPT should be utilized as a supplementary resource rather than a primary decision-making tool, ensuring that AI-generated responses are cross-checked against established clinical guidelines to maintain accuracy. Furthermore, the results underscore ChatGPT's potential role in clinical education. By providing readily accessible, guideline-consistent information, ChatGPT can serve as a valuable resource in the training of physiotherapists and other healthcare professionals. However,

further refinement of the model is necessary to enhance its ability to address the complexities of rehabilitation interventions and ensure adherence to domain-specific practices. Particularly, improving the reliability of ChatGPT's recommendations in rehabilitation contexts could significantly expand its clinical utility. At the same time, educators should be mindful that reliance on AI tools does not replace the need for students to engage critically with original sources, learn how to interpret clinical guidelines, and develop independent reasoning skills.

In summary, ChatGPT exhibits high potential in musculoskeletal rehabilitation education and information retrieval but requires cautious integration into practice due to its variability in complex clinical reasoning tasks. Ongoing evaluation and refinement of AI tools are essential to enhance their reliability, particularly in context-specific applications such as physiotherapy.

Future directions

To address the identified limitations, future research should focus on fine-tuning ChatGPT for specific healthcare domains, particularly musculoskeletal rehabilitation. Incorporating domain-specific datasets and enhancing the model's ability to interpret complex clinical scenarios could substantially improve its practical applicability. Additionally, longitudinal studies assessing the long-term impact of ChatGPT's integration into clinical practice are warranted.

Furthermore, the role of contextual and placebo effects in shaping user perceptions of AI-generated responses requires further exploration. As highlighted in this study, non-specific factors—such as response presentation and perceived authority—may influence both the acceptability and perceived reliability of AI tools. Investigating these factors could inform strategies to optimize the design and deployment of AI systems in healthcare, ensuring their effective and responsible use. Future research should include controlled trials assessing the impact of ChatGPT use in clinical education and patient care decision-making, particularly in real-time clinical scenarios.

Conclusion

This study underscores ChatGPT's dual role as both a promising and a challenging tool in musculoskeletal rehabilitation. While it excels in clarity, relevance, and accuracy, its limitations in consistency and domain-specific reasoning necessitate careful oversight and ongoing refinement. Addressing these challenges could enable ChatGPT to evolve into a reliable decision-support tool, ultimately enhancing clinical practice, education, and patient care in musculoskeletal rehabilitation.

Key Points

Findings

- ChatGPT's responses demonstrated high clarity (average score of 4.85) and relevance (4.50) when compared to clinical guidelines for musculoskeletal rehabilitation.
- The lowest score was observed in consistency (3.85), highlighting variability in repeated responses.

Implications

- These findings suggest that ChatGPT has potential as a supplementary tool for physiotherapists, offering guideline-aligned recommendations in musculoskeletal care.
- Enhancing the model's consistency could improve its reliability for routine clinical use and decision support.

Caution

- The study relied on a predefined set of questions, which may not represent the full complexity of clinical practice.
- The evaluation was based on subjective scoring, which, despite using experienced raters, introduces potential variability. Furthermore, although the assessment criteria were informed by existing literature and expert consensus, the rating tool itself has not been formally validated.

Abbreviations

ACL	Anterior Cruciate Ligament
AI	Artificial Intelligence
CPG	Clinical Practice Guideline
GPT-4	Generative Pre-trained Transformer 4
NLP	Natural Language Processing
SPSS	Statistical Package for the Social Sciences

Acknowledgements

The authors have no acknowledgments.

Author contributions

E.S. was responsible for the conception and design of the study, data collection, and interpretation of the data. S.Y. contributed to the study design, data collection, and interpretation of the data. Both E.S. and S.Y. were involved in the drafting of the manuscript, and all authors contributed to the interpretation of the data for the work and revising it critically for important intellectual content. All authors have finally approved the manuscript.

Funding

No.

Data availability

All data supporting the findings of this study, including ChatGPT-generated responses and Likert-scale ratings, are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Bezmialem Vakıf University, Istanbul, Turkey

²Institute of Health Sciences, Department of Physiotherapy and Rehabilitation, Istanbul Medipol University, Istanbul, Turkey

Received: 30 January 2025 / Accepted: 11 April 2025

Published online: 24 April 2025

References

1. Sarella PNK, Mangam VT. AI-driven natural Language processing in health-care: transforming patient-provider communication. *Indian J Pharm Pract*. 2024;17(1).
2. Kalra N, Verma P, Verma S. Advancements in AI based healthcare techniques with FOCUS ON diagnostic techniques. *Comput Biol Med*. 2024;179:108917.
3. Jain K, Prajapati V. NLP/deep learning techniques in healthcare for decision making. *Prim Health Care: Open Access*. 2021;11(3):373–80.
4. Biswas SS. Role of chat Gpt in public health. *Ann Biomed Eng*. 2023;51(5):868–9.
5. Van Dis EA, Bollen J, Zuidema W, Van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023;614(7947):224–6.
6. Rajjoub R, Arroyave JS, Zaidat B, Ahmed W, Mejia MR, Tang J, et al. ChatGPT and its role in the decision-making for the diagnosis and treatment of lumbar spinal stenosis: a comparative analysis and narrative review. *Global Spine J*. 2024;14(3):998–1017.
7. Ahmed W, Saturno M, Rajjoub R, Duey AH, Zaidat B, Hoang T et al. ChatGPT versus NASS clinical guidelines for degenerative spondylolisthesis: a comparative analysis. *Eur Spine J*. 2024:1–22.
8. Hao J, Yao Z, Tang Y, Remis A, Wu K, Yu X. Artificial Intelligence in Physical Therapy: Evaluating ChatGPT's Role in Clinical Decision Support for Musculoskeletal Care. *Ann Biomed Eng*. 2025:1–5.
9. Ismail AMA. ChatGPT: an expected excellent future technology in enhancing patient care education and physiotherapists' continuous training. *Eur J Physiotherapy*. 2024;26(1):62–3.
10. Johns WL, Martinazzi BJ, Miltenberg B, Nam HH, Hammoud S. ChatGPT provides unsatisfactory responses to frequently asked questions regarding anterior cruciate ligament reconstruction. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*; 2024.
11. Gianola S, Barger S, Castellini G, Cook C, Palese A, Pillastrini P, et al. Performance of ChatGPT compared to clinical practice guidelines in making informed decisions for lumbosacral radicular pain: a cross-sectional study. *J Orthop Sports Phys Therapy*. 2024;54(3):222–8.
12. Sawamura S, Bito T, Ando T, Masuda K, Kameyama S, Ishida H. Evaluation of the accuracy of ChatGPT's responses to and references for clinical questions in physical therapy. *J Phys Therapy Sci*. 2024;36(5):234–9.
13. Finucane LM, Stokes E, Briggs AM. Its everyone's responsibility: responding to the global burden of musculoskeletal health impairment. *Musculoskeletal science & practice*; 2023. p. 102743.
14. Lin I, Wiles L, Waller R, Goucke R, Nagree Y, Gibberd M, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. *Br J Sports Med*. 2020;54(2):79–86.
15. Hoffmann TC, Lewis J, Maher CG. Shared decision making should be an integral part of physiotherapy practice. *Physiotherapy*. 2020;107:43–9.
16. Kolac UC, Karademir OM, Ayik G, Kaymakoglu M, Familiari F, Huri G. Can popular AI large Language models provide reliable answers to frequently asked questions. About Rotator Cuff Tears? *JSES International*; 2024.
17. Neravetla AR, Nomula VK, Mohammed AS, Dhanasekaran S, editors. Implementing AI-driven Diagnostic Decision Support Systems for Smart

- Healthcare. 2024 15th International Conference on Computing Communication and Technologies N. (ICCCNT); 2024: IEEE.
18. Dlugatch R, Georgieva A, Kerasidou A. AI-driven decision support systems and epistemic reliance: a qualitative study on obstetricians' and midwives' perspectives on integrating AI-driven CTG into clinical decision making. *BMC Med Ethics*. 2024;25(1):6.
 19. Bilika P, Stefanouli V, Strimpakos N, Kapreli EV. Clinical reasoning using ChatGPT: is it beyond credibility for physiotherapists use? *Physiother Theory Pract*. 2024;40(12):2943–62.
 20. AlShehri Y, McConkey M, Lodhia P. ChatGPT provides satisfactory but occasionally inaccurate answers to common patient hip arthroscopy questions. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*; 2024.
 21. Erickson M, Lawrence M, Jansen CWS, Coker D, Amadio P, Cleary C, et al. Hand pain and sensory deficits: carpal tunnel syndrome: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of hand and upper extremity physical therapy and the academy of orthopaedic physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2019;49(5):CPG1–85.
 22. Blanpied PR, Gross AR, Elliott JM, Devaney LL, Clewley D, Walton DM, et al. Neck pain: revision 2017: clinical practice guidelines linked to the international classification of functioning, disability and health from the orthopaedic section of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2017;47(7):A1–83.
 23. Martin RL, Cibulka MT, Bolgia LA, Koc TA Jr, Loudon JK, Manske RC, et al. Hamstring strain injury in athletes: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy and the American academy of sports physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2022;52(3):CPG1–44.
 24. McDonough CM, Harris-Hayes M, Kristensen MT, Overgaard JA, Herring TB, Kenny AM, et al. Physical therapy management of older adults with hip fracture: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy and the academy of geriatric physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2021;51(2):CPG1–81.
 25. Carcia CR, Martin RL, Houck J, Wukich DK, Altman RD, Curwin S, et al. Achilles pain, stiffness, and muscle power deficits: Achilles tendinitis: clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopaedic section of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2010;40(9):A1–26.
 26. Kreiner DS, Hwang SW, Easa JE, Resnick DK, Baisden JL, Bess S, et al. An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *Spine J*. 2014;14(1):180–91.
 27. Lucado AM, Day JM, Vincent JJ, MacDermid JC, Fedorczyk J, Grewal R, et al. Lateral elbow pain and muscle function impairments: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of hand and upper extremity physical therapy and the academy of orthopaedic physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2022;52(12):CPG1–111.
 28. Willy RW, Hoglund LT, Barton CJ, Bolgia LA, Scalzitti DA, Logerstedt DS, et al. Patellofemoral pain: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2019;49(9):CPG1–95.
 29. Koc TA Jr, Bise CG, Neville C, Carreira D, Martin RL, McDonough CM. Heel pain–plantar fasciitis: revision 2023: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy and American academy of sports physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2023;53(12):CPG1–39.
 30. Kotsifaki R, Korakakis V, King E, Barbosa O, Maree D, Pantouveris M, et al. Aspetar clinical practice guideline on rehabilitation after anterior cruciate ligament reconstruction. *Br J Sports Med*. 2023;57(9):500–14.
 31. Ensek KR, Bloom NJ, Harris-Hayes M, Cibulka MT, Disantis A, Di Stasi S, et al. Hip pain and movement dysfunction associated with nonarthritic hip joint pain: A revision: clinical practice guidelines linked to the international classification of functioning, disability, and health from the academy of orthopaedic physical therapy and American academy of sports physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2023;53(7):CPG1–70.
 32. Michener LA, Heitzman J, Abbruzzese LD, Bondoc SL, Bowne K, Henning PT, et al. Physical therapist management of glenohumeral joint osteoarthritis: a clinical practice guideline from the American physical therapy association. *Phys Ther*. 2023;103(6):pzad041.
 33. Kelley MJ, Shaffer MA, Kuhn JE, Michener LA, Seitz AL, Uhl TL, et al. Shoulder pain and mobility deficits: adhesive capsulitis: clinical practice guidelines linked to the international classification of functioning, disability, and health from the orthopaedic section of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2013;43(5):A1–31.
 34. Mehta SP, Karagiannopoulos C, Pepin M-E, Ballantyne BT, Michlovitz S, MacDermid JC, et al. Distal radius fracture rehabilitation: clinical practice guidelines linked to the international classification of functioning, disability, and health from the academy of orthopaedic physical therapy and academy of hand and upper extremity physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2024;54(9):CPG1–78.
 35. George SZ, Fritz JM, Silfies SP, Schneider MJ, Beneciuk JM, Lentz TA, et al. Interventions for the management of acute and chronic low back pain: revision 2021: clinical practice guidelines linked to the international classification of functioning, disability and health from the academy of orthopaedic physical therapy of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2021;51(11):CPG1–60.
 36. Martin RL, Davenport TE, Paulseth S, Wukich DK, Godges JJ, Altman RD, et al. Ankle stability and movement coordination impairments: ankle ligament sprains: clinical practice guidelines linked to the international classification of functioning, disability and health from the orthopaedic section of the American physical therapy association. *J Orthop Sports Phys Therapy*. 2013;43(9):A1–40.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.