Western Graduate&PostdoctoralStudies

Western University
Scholarship@Western

Electronic Thesis and Dissertation Repository

7-31-2024 10:30 AM

Psychometric Properties of Patient-Reported Outcome Measures in Chronic Pain Conditions with Central Sensitization:- A Systematic Review

Mst Farjana Akhter, *The University of Western Ontario*

Supervisor: Gagnier, Joel J. Department of Epidemiology and Biostatistics and Department of Surgery, Schulich School of Medicine & Dentistry, Western University, *The University of Western Ontario*
: Karp, Igor Department of Epidemiology and Biostatistics, Western University, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Epidemiology and Biostatistics

© Mst Farjana Akhter 2024

Follow this and additional works at: https://ir.lib.uwo.ca/etd

 Part of the Epidemiology Commons

Recommended Citation

Akhter, Mst Farjana, "Psychometric Properties of Patient-Reported Outcome Measures in Chronic Pain Conditions with Central Sensitization:- A Systematic Review" (2024). *Electronic Thesis and Dissertation Repository*. 10235.
https://ir.lib.uwo.ca/etd/10235

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlswadmin@uwo.ca.

Abstract

Objective: To identify available patient-reported outcome measures (PROMs) used to evaluate central sensitization (CS) manifestations in chronic pain conditions and evaluate the quality of psychometric properties of those instruments.

Methods: A comprehensive search was conducted across multiple databases. Methodological quality and psychometric properties were assessed and summarized using the COSMIN checklist and scoring manual.

Results: A total of fifty-eight studies addressing eight PROMs were identified. The Central Sensitization Inventory (CSI) received the highest overall ratings for most measurement properties among all the instruments, followed by the Pain Sensitivity Questionnaire and Fibromyalgia Survey Questionnaire. Based on pooled data, the test-retest reliability of the CSI was found to be excellent, with an intra-class correlation coefficient of 0.93 (95% CI: 0.91- 0.95) for overall chronic pain conditions.

Conclusion: CSI could be a reliable PROM for chronic pain with CS. More studies should be performed to comprehensively evaluate all measurement properties of the PROMs.

Keywords

Central sensitization; Patient-reported outcome measures; Chronic pain; Measurement properties; Systematic review; COSMIN

Summary for Lay Audience

Background: Chronic pain is a prevalent public health issue. An underlying pathophysiological mechanism of many chronic pain conditions is central sensitization (CS) in which the central nervous system becomes overly sensitive, leading to persistent pain even without obvious tissue injury. To effectively evaluate treatment outcomes in managing chronic pain with CS, healthcare providers need reliable and valid tools to measure the severity and impact of pain. These tools, known as patient-reported outcome measures (PROMs), are questionnaires that patients fill out to describe their pain experience and outcome of care. However, it's crucial to assess the quality of a PROM to ensure that it can accurately measure what it intends to measure i.e., validity and produce consistent results i.e., reliability. This study focused on identifying various PROMs used to assess CS manifestations in chronic pain conditions and evaluated those PROMs' quality of measurement properties.

Methods: A comprehensive search was conducted across multiple databases for relevant articles. The methodological quality of studies and psychometric properties of PROMs were assessed and summarized using the COSMIN checklist and scoring manual which is a consensus-based and well-accepted guideline.

Results: A total of fifty-eight studies with eight instruments were identified. Most identified PROMs have limited evidence regarding their psychometric properties. The Central Sensitization Inventory (CSI) received the highest overall ratings for most measurement properties among all the instruments, followed by Pain Sensitivity Questionnaire and Fibromyalgia Survey Questionnaire. Based on pooled data from available studies, the test-retest reliability of the CSI was found to be excellent, with an intra-class correlation coefficient (ICC) of 0.93 (95% CI: 0.91- 0.95) for overall chronic pain, ICC of 0.90 (95% CI: 0.87- 0.93) for chronic musculoskeletal pain and ICC of 0.93 (95% CI: 0.88- 0.99) for chronic neck pain. PSQ also demonstrated excellent test-retest reliability, showing an ICC of 0.86 (95% CI: 0.72- 0.99) for chronic pain conditions.

Conclusion: Although not all properties have been studied, the CSI, which received the highest overall ratings, could serve as a reliable PROM for chronic pain associated with CS. More studies should be performed to comprehensively evaluate all measurement properties of all included instruments.

Acknowledgements

I express my sincere gratitude and deep regards to my supervisors, Dr. Joel Gagnier and Dr. Igor Karp for their supervision and advice to carry out the study successfully. Their consistent academic support throughout my MSc journey distinguishes them as exceptional mentors. I am deeply grateful for their generosity in sharing their time and expertise. This project could not have been accomplished without their persistent help.

I also wish to extend my heartfelt gratitude to Dr. Pavlos Bobos for his invaluable guidance and support, which significantly enhanced the quality of this project.

I am also thankful to my colleagues and friends, who assisted in completing this project. I particularly extend my appreciation to Nana Ama Tiwaa-Boateng and Edith Ginika Otalike for their kind help and contribution to this project.

I thank Dr. Andrew Firth for his valuable insights during the review process.

Finally, I wish to acknowledge the loving support extended to me by my husband, Shibli Shams, and daughter, Yamina Shams. Their presence was crucial in keeping me motivated and offering me emotional strength.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

CS = Central Sensitization

PROM = Patient-Reported Outcome Measure

COSMIN = COnsensus-based standards for the selection of health Status Measurement INstruments

COPCs = Chronic Overlapping Pain Conditions

GRADE = Grading of Recommendations Assessment, Development and Evaluation

QST = Quantitative Sensory Testing

CPM = Conditioned Pain Modulation

HRQoL = Health-Related Quality of Life

QoL = Quality of Life

ICC = Intraclass Correlation Coefficient

CI = Confidence Interval

SEM = Standard Error of Measurement

LoA = Limits of Agreement

SDC = Smallest Detectable Change

MIC = Minimal Important change

CFA = Confirmatory Factor Analysis

EFA = Exploratory Factor Analysis

PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-analyses

SD= Standard Deviation

CSI=Central sensitization Inventory

PSQ= Pain Sensitivity Questionnaire

FSQ= Fibromyalgia Survey Questionnaire

NFF= Nociplastic-based Fibromyalgia Feature

GPQ= Generalized Pain Questionnaire

SHS= Sensory Hypersensitivity Scale

L-VISS and VDS= Leiden Visual Sensitivity Scale and Visual Discomfort Scale

IMMPACT = Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials

Chapter 1

1 Introduction

Chronic pain is a prevalent public health issue, with clear evidence of its detrimental effects on both patient well-being and productivity [1,2]. According to some studies, chronic pain causes a significant socioeconomic burden impacting over 30% of the global population [3]. Chronic pain is categorized as cancer pain and non-cancer pain which can include fibromyalgia, migraine, irritable bowel syndrome, musculoskeletal pain, pelvic pain, temporomandibular joint disorder and others [4,5]. Chronic pain adversely affects an individual's overall quality of life including physical and emotional well-being, sleep quality, and functional status [6,7] and can lead to significant psychosocial consequences [8].

Chronic pain often lacks specificity. Many chronic pain patients may not have identifiable pathology for persistent pain or may experience pain even after the healing of tissue damage [9]. Central sensitization (CS) is a proposed underlying pathophysiological mechanism of many chronic pain conditions offering insights into the persistent pain experience and the overall clinical presentation [10]. CS refers to a complex phenomenon marked by the dysregulation of the central nervous system, triggering both neuronal imbalance and hyperexcitability [11] which leads to increased sensitivity to pain or persistent pain despite the resolution of injury [12]. The presentation of clinical features associated with CS is indicative of unfavourable treatment outcomes in managing pain for individuals with various chronic pain conditions [13,14,15,16,17]. Evaluation of treatment effectiveness aimed at improving chronic pain is challenging due to the complex and subjective nature of pain [18,19]. Additionally, pain associated with CS often responds more to centrally acting drugs with analgesic efficacy (e.g. anticonvulsants, antidepressants, neuromodulators) rather than by peripheral pain-relieving agents such as non-steroidal anti-inflammatory drugs (NSAIDs) [12]. Moreover, in addition to pharmacological treatment, nonpharmacological approaches might be helpful for better management of patients with central sensitization such as cognitive-behavioural therapy, physical and occupational therapy, and graded exercise therapy [20,21]. Therefore, a standardized approach to assess CS and outcomes of care in chronic pain patients is important to implement suitable treatment strategies for better management.

Patient-reported outcome questionnaires are used to assess pain associated with CS [22]. As pain is innately subjective and a multifactorial construct which cannot be accurately assessed by objective measures or clinician assessment, PROMs remain widely accepted as the standard for the assessment of an outcome of care associated with pain [23]. PROMs aim to directly capture the patient's perspectives and experiences regarding treatment outcomes [24] which is essential for providing high-quality clinical care.

Psychometric quality (e.g. reliability, validity and responsiveness) of PROMs is vital [25], because it is important to ensure that a PROM accurately measures what it intends to measure (validity), produces consistent results in repeated administration under similar conditions (reliability) and is sensitive enough to detect meaningful changes in the patient's condition (responsiveness) [25,26,27]. Highlighting this quality is crucial to confirm that the clinical utility of those PROMs is appropriate [28].

Therefore, assessment of the quality of measurement properties of PROMs targeting central sensitization in chronic pain conditions is fundamental in clinical practice to derive meaningful benefit from their application, to assist healthcare providers in comprehending patients' pain experiences, assessing the effectiveness of treatment, and subsequently customizing treatment plans in improving pain management. There is a scarcity of PROMs explicitly designed to assess central sensitization in chronic pain conditions [29]. There hasn't been done a systematic review that comprehensively covers the range of existing PROMs utilized to evaluate manifestations of central sensitization; whether these instruments are multidimensional to adequately capture the various aspects of central sensitization; or whether these instruments have robust psychometric properties in the population of interest. To address this research gap, this systematic review aims to identify available PROMs used in assessing pain associated with central sensitization and related manifestations and to evaluate and summarize the overall quality of measurement properties of those outcome instruments using a well-established critical appraisal tool.

Chapter 2

2 Overview of Chronic Pain, Central Sensitization, Health-Related Quality of Life, Patient-Reported Outcome Measures and Research Objectives

2.1. Understanding of Chronic Pain and Central Sensitization

Chronic pain is typically characterized as any form of persistent pain lasting beyond a period of three months [30]. A review of chronic pain prevalence studies has indicated that the prevalence of chronic pain in the adult population ranges from 2% to 40% globally [31], although this broad range highlights the complexity of measuring chronic pain prevalence. Whereas chronic pain is highly prevalent, it is still poorly understood and notably challenging to manage effectively. Chronic pain can persist in regions where damaged tissue is no longer evident. Often, despite thorough examination, there may be no detectable signs of tissue damage, inflammation, or peripheral sensitization, or if these are present, they may not fully account for the reported severity of pain and associated symptoms [32,33,34]. Traditionally, chronic pain was thought to originate from nociception or neuropathy although later centralized pain or nociplastic pain was introduced [35]. Aetiologies of chronic pain disorders are not entirely clear, as these disorders exhibit common clinical features such as functional impairment, disturbances in sleep, cognitive challenges, and emotional distress.

2.1.1. Association Between Central Sensitization and Chronic Pain

Central sensitization (CS) can serve as an explanation for the experience of chronic pain with or without significant tissue damage [11,36]. The relationship between CS and chronic pain appears to involve various mechanisms that affect the spinal cord [37] and ascending or descending modulatory systems [38]. Clinical and experimental features linked with central sensitization frequently manifest across a range of chronic pain conditions [12,39,40]. These include fibromyalgia [41], chronic traumatic neck pain (i.e., whiplash) [42], osteoarthritis [43], knee osteoarthritis [44], shoulder pain [45], migraine [46], irritable bowel syndrome ([47], chronic fatigue syndrome [48], low back pain [49], rheumatoid arthritis [50] and cancer treatment-induced pain [51]. Patients with rheumatoid arthritis typically experience inflammatory and nociceptive pain linked to joint inflammation [52] but there are also features of central sensitization contributing to the rise of pain [53]. People with osteoarthritis demonstrate elevated pain intensity, prolonged discomfort, and a wider distribution of pain when subjected to a standardized injection of hypertonic saline compared

with pain-free controls. They also manifest extensive sensitivity to pressure-induced pain, increased temporal summation of pain, and impaired endogenous analgesia [54].

Multiple common chronic pain conditions like fibromyalgia, chronic migraine headaches, chronic low back pain, temporomandibular dysfunction, irritable bowel syndrome, chronic fatigue syndrome, interstitial cystitis, endometriosis, and chronic tension-type headaches often coexist. Numerous research findings suggest that these conditions share overlapping clinical and pathophysiological features, in which central sensitization plays a key role [55,56,57]. Although various terms have been used to describe this overlap, such as central sensitivity syndromes [58], the National Institutes of Health recognized these co-occurring pain conditions as Chronic Overlapping Pain Conditions (COPCs). People with COPCs often experience similar symptoms beyond pain, including insomnia, headaches, difficulty in concentrating, physical and cognitive limitations, emotional challenges like anxiety and depression, and fatigue [55].

2.1.2. Peripheral Sensitization

Peripheral sensitization manifests on tissue sites refers to the reduced threshold and augmented response of sensory nerve endings when they are exposed to noxious stimuli. Typically, pain commences with the activation of peripheral sensory neurons, which subsequently interpret and transmit nociceptive messages to the higher brain centre.

Nociception is the sensory mechanism through which various types of stimuli, such as mechanical, chemical, and thermal stimuli, are identified by the peripheral nerve fibres called nociceptors. These nociceptors play a critical role in converting noxious stimuli into electrical signals as action potentials. These signals are then relayed to the spinal cord and eventually to the brain for further processing [59]. After an injury and the resulting inflammation, nociceptors can undergo sensitization due to pro-nociceptive mediators like bradykinin, substance P, prostaglandins, or extracellular ATP [60].

Pain signals are typically initiated through the activation of first-order afferent nociceptors situated at the endpoints of primary afferent nerves. These nociceptors can be categorized based on the type of afferent fibres involved, specifically into Aδ fibres (myelinated) and C fibres. C fibres, which are both small and unmyelinated, have broader receptive fields and are linked to secondary nociceptive sensations. The majority of C fibre afferent nerves react to

various painful stimuli, resulting in the activation of nociceptive afferent neurons [61,62]. This increased sensitivity is referred to as peripheral sensitization.

Peripheral sensitization commonly arises in the aftermath of peripheral nerve injury or tissue damage. The endogenous chemicals released at the site of tissue injury or inflammation can stimulate and heighten the sensitivity of peripheral sensory neurons, thereby giving rise to peripheral sensitization [63,64]. Although this phenomenon effectively explains the occurrence of heightened pain sensitivity after an injury, it doesn't provide a satisfactory explanation for the spread of pain sensitivity beyond the specific region of tissue damage.

2.1.3. Endogenous Pain Modulation

Pain modulation involves the neural mechanism that either enhances or suppresses the transmission of pain signals. The adjustment of neural impulses within the central nervous system significantly shapes the experience of pain. The modulation of nociceptive signals by central pain inhibitory and facilitatory processes is a well-established phenomenon [65]. These modulatory mechanisms which influence pain act at different levels of the central nervous system. They are believed to be crucial in both the development of chronic pain and the variability observed among individuals in the course of persistent pain conditions [66,67,68].

It is widely recognized that endogenous pain mechanisms exist which reduce pain through overall "inhibition" [9]. There is likely a descending pain modulatory circuit in place that modulates pain perception. Over the past fifty years, numerous studies have solidified the understanding that the periaqueductal gray-rostral ventromedial medulla system plays a crucial role in the complicated process of descending pain modulation, encompassing both inhibition and facilitation [69,70]. The periaqueductal gray receives inputs from higher brain centers, including the hypothalamus, the amygdala, and the rostral anterior cingulate cortex. The periaqueductal gray transmits pain-altering signals to dorsal horn neurons in the spinal cord, either through direct pathways or indirectly via the rostral ventromedial medulla.

Within the rostral ventromedial medulla, neurons have been observed to project to the dorsal horns of the spinal cord or medulla, influencing nociceptive inputs either by amplifying or reducing them. This process serves as a pivotal relay in the regulation of descending pain facilitation, ultimately shaping the perception and experience of pain [71]. Neurons within the rostral ventromedial medulla exhibit categorizations as "ON-cells," "OFF-cells," and

"NEUTRAL-cells." ON-cells are implicated in promoting descending facilitation, which is pro-nociceptive, while OFF-cells function to inhibit nociception, demonstrating anti-nociceptive properties [72,73]. Consequently, heightened pain processing is linked to increased ON-cell activity and the suppression of OFF-cell firing. In patients with Central Sensitization, an imbalance in descending pain modulatory systems contributes to sustaining hyperalgesia and/or an increase in endogenous pain facilitation [74].

2.1.4. Pathophysiology and Symptoms of Central Sensitization

Woolf and King introduced the concept of central sensitization in 1989 following their research on rats demonstrating increased excitability of spinal cord neurons following injury. Further investigations revealed that central sensitization could persist even in the absence of ongoing peripheral input [75,76]. Healthcare professionals and clinical researchers worldwide unite in the consensus that central sensitization should extend beyond laboratory confines. Understanding central sensitization may promote enhancing the characterization and diagnosis of pain patients and the development of novel treatment strategies based on underlying mechanisms. This has led to an extensive body of research, with hundreds of studies investigating characteristics associated with assumed central sensitization in various chronic pain conditions. Consequently, central sensitization is now firmly recognized as a prevalent underlying mechanism among many patients experiencing chronic pain [77].

Several mechanisms have been suggested as contributors to persistent central sensitization, including cortical reorganization and adverse neural adaptability, changes in neurochemical composition, reduction of inhibitory neuron function, disruption of internal pain modulation mechanisms, and alterations in the structural integrity and connectivity of white matter [11]. Such alterations within the central nervous system can lead to disruptions in sensory processing of peripheral inputs, causing a discrepancy between the individual's experience and the information received from the periphery [29,78,79], altering anti-nociceptive mechanisms [74,80], and increased temporal summation of second pain or wind-up [81,82]. This results in general hypersensitivity of the somatosensory system, showing symptoms such as widespread pain or pain in various body regions, allodynia, hyperalgesia or secondary hyperalgesia [12]. Patients with CS often experience sensory hyperresponsiveness which is excessive sensitivity to various stimuli, physical pressure, and widespread intolerance to other external stimuli (light, sounds, electrical impulses, smells, etc.) [10]. Furthermore, studies have indicated that patients identified as having central sensitization represent more

diffuse pain, increased pain-related functional impairment, psychological distress, fatigue, and higher levels of depression and anxiety [83,84].

2.1.5. Concept of Nociplastic Pain or Centralized Pain

Pain may be mechanistically classified as nociceptive, neuropathic, and CS pain (non-nociceptive pain). In 2017, the International Association on the Study of Pain categorized pain stemming from altered nociceptive function due to central sensitization as 'nociplastic pain' [85,86] though ‘CS-pain’ or ‘centralized pain’ and nociplastic pain have often been viewed as synonymous [87]. The importance of the term lies in recognizing that not all chronic pain conditions can be linked to structural or tissue damage. Increased occurrences of nociplastic pain states, as illustrated by conditions such as fibromyalgia, are observed in individuals with osteoarthritis, rheumatoid arthritis, and other nociceptive pain disorders [88]. Diagnosing and managing nociplastic conditions poses a significant challenge due to the absence of distinct objective markers and their association with complicated underlying mechanisms. Addressing this difficulty, the International Association for the Study of Pain introduced, in 2021, the initial set of clinical criteria and a grading system for nociplastic pain [85].

2.1.6. Differentiation of Nociplastic/Centralized Pain from Nociceptive and Neuropathic Pain

Nociceptive pain is usually acute and develops in response to a particular circumstance or an external stimulus. Nociceptive pain arises from the damage of nonneural tissue and activation of nociceptors [89]. This type of pain is identifiable by the intensity of the pain sensation corresponding to the extent of tissue damage [39]. Neuropathic pain, as per the International Classification of Diseases 11th Revision (ICD-11), is caused by a lesion or disease of the somatosensory nervous system which can lead to the loss of normal function, heightened pain sensitivity, and the occurrence of spontaneous pain [90]. Centralized pain is characterized by the central nervous system's improper processing of pain signals due to central sensitization. Those suffering from centralized pain may interpret non-noxious stimuli as painful and experience enhanced pain intensity from normally painful stimuli. Unlike nociceptive or neuropathic pain, centralized pain is difficult to localize precisely, and patients experience more pain linked with other different symptoms [91].

2.1.7. Clinical Importance of Central Sensitization

Chronic pain patients with predominant central sensitization exhibit higher pain severity and lower quality of life compared to patients without CS [90,92]. As per a World Health Organization study, individuals enduring chronic pain face a fourfold higher risk of experiencing depression or anxiety compared to those without pain, and they are more than twice as likely to encounter challenges in their ability to work [92]. Therefore, there is a growing global recognition that addressing central sensitization should be a key focus in chronic pain treatment. Recognizing central sensitization in chronic pain patients is crucial, as it correlates with poorer treatment outcomes, particularly when therapies target local tissue or presumed nociceptive sources, and reflects a paradigm shift in understanding and managing chronic pain beyond traditional muscle and joint-centric approaches [13,14,15,16,17,51]. Patients' pain control and quality of life can be significantly improved by addressing co-morbid symptoms and/or co-morbid conditions associated with central sensitization [55,93]. Notably, the concept of quality of life is progressively gaining recognition as one of the paramount factors to be assessed when evaluating the efficacy of any therapeutic or health-related intervention [94].

2.2. Assessment of Central Sensitization

2.2.1 Neuroimaging

Neuroimaging methods provide a precise and quantitative means to evaluate the pathology of pain by assessing the extent of brain activation associated with pain sensations [95]. It is widely acknowledged that the central nervous system plays a pivotal role in pain processing, and numerous characteristics of chronic pain can be attributed to alterations within the nervous system [96]. Neuroimaging methods offer a tool for gaining insights into the neurophysiological mechanisms that underlie the initiation and persistence of chronic pain and thus contribute to the understanding of the perception and modulation of the pain experience. In chronic pain patients with central sensitization, various brain imaging techniques like functional magnetic resonance imaging, positron emission tomography, and single photon emission computed tomography are being utilized to examine the responsiveness of specific brain regions and how they respond to specific pain modalities [97,98,99,100].

Though brain imaging methods can provide valuable information regarding neural mechanisms that contribute to pain and offer the most objective means to assess central

sensitization, there has been a recommendation to consider psychophysical evaluations for measuring pain modulation processes in healthy individuals and the disrupted pain modulation processes in chronic pain patients.

2.2.2 Quantitative Sensory Testing (QST)

Quantitative sensory testing (QST) is a psychophysical method in which different experimental pain stimuli (such as thermal, mechanical, electrical, and chemical stimuli) are applied to various tissues at regulated intensities and then responses are evaluated by thresholds and stimulus-response functions [101,102].

QST comprises both static and dynamic assessments. In static tests, pain intensity ratings or pain thresholds are determined by comparing the actual stimulus intensity to the level of pain perceived by the participant in response to a specific stimulus [103]. Dynamic QST is utilized to study how the central nervous system processes painful sensations [104]. Typically conducted dynamic QSTs include temporal summation (increase in pain in response to repeated noxious stimuli) and conditioned pain modulation (CPM) which evaluates endogenous descending inhibitory modulation. CPM is a method used to determine if the sensation of pain by a painful test stimulus at a specific location is reduced when a second painful (conditioning) stimulus is administered to a distant area of the body [103]. Typically, in individuals with a functioning pain regulation system, the pain felt from the test stimulus diminishes while or after the conditioning stimulus is administered, indicating the effectiveness of the body's endogenous pain-inhibiting mechanism.

QST can provide indications that CS might be present [101] by showing decreased pain thresholds, heightened pain in response to stimuli exceeding the standard threshold, and the occurrence of temporal summation of pain. A comprehensive review indicates that the clear manifestation of reduced efficiency in conditioned pain modulation is primarily observed in individuals suffering from pain conditions with unknown etiology like fibromyalgia, temporomandibular disorders, irritable bowel syndrome, and tension-type headaches [74]. Although there is substantial evidence indicating impairments in pain modulation in various chronic pain conditions [105], the question of whether CPM can serve as a reliable predictor of pain symptoms and be deemed a valid clinical pain biomarker remains unclear. This matter holds considerable significance, given the persistent challenge of chronic pain, which continues to lack effective treatment options [106]. QST offers a thorough examination of pain sensitivity patterns but it requires a standardized set of instructions, in order to achieve

trustworthy results. QST frequently demands specialized training, costly laboratory apparatus, and substantial patient time within the laboratory, posing challenges to clinical implementation [102]. QST and neuroimaging techniques are employed in research settings but prove challenging to easily and validly implement in clinical practice.

2.2.3. Questionnaires

Diagnostic approaches alternative to objective measures in clinical settings often involve patient-reported questionnaires [107,108] to detect symptoms associated with central sensitization and outcome of care. Patient-reported measures offer the most straightforward, prompt, and convenient assessment of pain, psychological well-being, pain interference with physical functionality and overall quality of life in patients with central sensitization at minimal expense.

2.3. Quality of Life and Patient-Reported Outcome Measures (PROMs)

2.3.1. Health-Related Quality of Life (HRQoL)

In the last three decades, the assessment of quality of life has gained growing significance within the realm of healthcare. The terms Health-Related Quality of Life (HRQoL) and Quality of Life (QoL) are frequently employed interchangeably. Nevertheless, many authors state that QoL is a broader concept covering all aspects of human life. The term HRQoL was introduced with the specific aim of refining the focus to consider the effects of illness, and the impact of treatment on overall quality of life [109]. The foundation of HRQoL lies in the interplay of health and quality of life and provides a comprehensive evaluation, shaped by an individual's experiences, beliefs, expectations, and insights [110]. HRQoL is characterized as a multidimensional concept that involves the subjective assessment of how the health status influences various domains, including physical, mental, and social functioning, psychological well-being, and potentially other domains that hold relevance to the specific disease under investigation [111,112,113].

Health-related quality of life is evaluated through self-reports. The significance of measuring HRQoL is growing not only in healthcare but also in clinical trials. The adoption of a patient-centred approach in both clinical research and care settings has increased the acceptance of PROMs as valuable and trustworthy tools for assessing HRQoL [114]. Both nationally and internationally, the evaluation of HRQoL through PROMs is progressively supplementing

conventional clinical methods for assessing health and the impact of treatment on patients [115,116].

2.3.2. PROMs

The term "patient-reported outcome" broadly refers to any health-related data provided by the patient [117]. PROs offer valuable insights into how patients perceive not just the clinical advantages but also the influence of treatment on their daily lives, encompassing potential side effects and the challenges associated with treatment administration. PROMs are assessments of direct reports from patients and are progressively employed as quality improvement tools which provide clinically important information [20]. These are generally in the form of self-administered questionnaires.

PRO questionnaires are distinctive in that they produce scores for informing decisions at the individual patient level. Most PRO instruments are designed according to psychometric principles and methods [118] to ensure their scientific rigour and effectiveness and apply the summated rating scale principle [119]. Patients report their symptoms, behaviours, abilities, or perceptions usually by Likert, numeric or categorical response scale for each item. The responses to individual items assessing the same construct (e.g., fatigue) are combined to derive a score.

PROMs can be classified as generic and disease or condition-specific. Sometimes a combination of these two types of PROMs is used. The generic PROM measures general aspects of health relevant to diverse patient groups and the general population, enabling a holistic assessment of care, quality of life, and the cost-effectiveness of interventions [25]. In comparison to generic measures, disease-specific instruments are more likely to be clinically relevant and responsive to changes in health status and can capture unique information not obtained by generic instruments [120]. PROMs are utilized in diverse settings, such as research, clinical practice, and policy development. They can function as screening tools or evaluative tools, depending on their application. When used as screening tools, PROMs facilitate the initial detection of risk or potential issues within a population. In their evaluative role, PROMs are instrumental in assessing the effectiveness of treatments or interventions and in monitoring a patient's subjective outcomes over time [121,122,123].

The effective utilization of PROMs in routine clinical practice may encounter some barriers. A systematic review highlighted several patient-level barriers to utilizing PROMs, including

concerns regarding time limitations, achieving optimal patient engagement, and difficulty in completing PROMs due to physical limitations or challenges in recalling information [124]. Patients might also worry about how their responses could negatively influence the care provided by healthcare providers, leading them to adjust their answers accordingly. However, these barriers can be addressed by adopting appropriate ways such as selecting suitable tools, enhancing patient education, and improving accessibility [124].

Despite some barriers, PROMs are important in clinical research. They are vital tools in enhancing the comprehensiveness, relevance, and patient-centeredness of study findings. A study analyzing the ClinicalTrials.gov databases from 2007 to 2013 found that the adoption of PROMs in clinical research has increased, particularly within oncology trials [125]. Upon reviewing evidence from randomized controlled trials (RCTs), multiple reviews have subsequently assessed the impact of incorporating PROMs into clinical practice. These reviews have concluded that PROMs could potentially impact patient care. These patient-centred measures provide researchers with a comprehensive understanding of treatment outcomes, ensuring that interventions align with patients' needs which further promotes patient-centred care and shared decision-making between healthcare providers and patients [126]. However, for all of this, an accurate selection of PROMs for clinical use or research is essential.

2.4. Importance of Systematic Reviews and Methodological Standards in Evaluating PROMs

Systematic reviews of publications on measurement properties of PROMs could help to effectively clarify and summarize the psychometric properties of tools used to assess specific constructs within target populations. They have a pivotal role in offering a comprehensive perspective and facilitating evidence-based ideas for the optimal selection of a PROM designed to specific needs, whether for research or clinical use, and evaluative or predictive applications [127]. In conducting a systematic review of PROMs, it is essential to evaluate the methodological quality of the included studies and separately assess the quality of the PROMs, focusing on their measurement properties. The evaluation of methodological quality is crucial as it directly influences the trustworthiness of study results [127]. Studies with poor methodological quality are more likely to introduce biases in the measurement and analysis of PROMs, leading to less reliable or inaccurate findings. Again, PROMs of poor quality can bias the treatment effects [128]. Therefore, data collected from these measures may be

unreliable or inaccurate as they may underestimate or overestimate the true effects of interventions or treatments, leading to flawed assessments. Using poor-quality PROMs in clinical research may lead to the generation of misleading or inconclusive findings, which further mislead decision-making. The quality of the journal in which an article is published is also important. Journals with high-impact factors and rigorous peer-review processes are more likely to publish high-quality research [129]. This means that the research has undergone thorough evaluation and critique, enhancing its credibility and reliability. Systematic reviews are not without limitations that can undermine their applicability. The heterogeneity of included studies, with variations in methodologies and outcome measures, can make it difficult to draw consistent conclusions.

Studies evaluating the psychometric properties of health-related patient-reported outcome instruments need to follow established methodological standards to produce reliable and valid conclusions. To clarify concepts, taxonomy, terminology, and definitions related to measurement properties and their meanings, a Delphi procedure could be a useful approach for experts to reach a consensus [25,130]. The COnsensus-based standards for the selection of health Status Measurement INstruments (COSMIN) guidelines help to critically evaluate the methodological quality of studies detailing the development and measurement properties of health-related measures [127,131]. It provides a structured framework for assessing the overall quality of outcome measurement instruments used in both research and clinical practice. This guideline evolved in concordance with current recommendations for reviews, along with the Cochrane Handbook for systematic reviews of interventions [132] and for diagnostic test accuracy reviews [133], the Preferred Reporting Items for Systematic Reviews and meta-analyses (PRISMA) Statement [134], the Institute of Medicine (IOM) standards for systematic reviews of comparative effectiveness research [135] and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) principles [136]. The COSMIN steering committee, comprising a diverse team of researchers specialized in different fields has formulated a comprehensive methodological guideline for systematic reviews of PROMs. Prinsen and colleagues (2018) outlined a consecutive ten-step procedure for systematically reviewing PROMs which includes steps such as conducting the literature search and selecting relevant studies (steps: 1-4), evaluating the quality of the eligible studies and the measurement properties(steps: 5-7), Evaluating interpretability and feasibility, formulating recommendations, and reporting the systematic review (steps: 8-10) [127]. The present review has been carried out using these methods.

2.5. Justification of the Current Review

A preliminary search in PUBMED, Cochrane Database of Systematic Reviews, and COSMIN Database has been conducted to identify existing and in-progress systematic reviews that have evaluated the measurement properties of pain assessment tools focused on centralized pain or central sensitization for chronic pain conditions. From this search, three systematic reviews were identified. Among the existing reviews, one systematic review revealed the prevalence of pain sensitivity and neuropathic-like pain in Inflammatory Arthritis by using questionnaire data [137]. Another review was limited to evaluating The CSI to assess symptoms of CS [138]. A review by Middlebrook and colleagues (2019) on measurement properties of CS measures in the musculoskeletal trauma population mainly focused on QST [139]. The three mentioned reviews mainly focused on one or two most commonly used PROMs and objective measures. There has yet to be conducted a systematic review to evaluate the quality of available self-reported outcome tools used in chronic pain with central sensitization by systematic approach with established criteria. This review thus aims to thoroughly assess the quality of these tools by applying the updated COSMIN criteria for evaluating both study quality and psychometric properties. Accurate assessment tools are essential in clinical practice and research to evaluate treatment outcomes which could contribute to the optimization of therapy. This comprehensive review will provide valuable insights for clinicians and researchers in selecting appropriate tools for chronic pain with CS and guidance for future research in this area.

2.6. Research Objectives.

The objectives of this study are to (1) identify all PROMs available to assess centralized pain or central sensitization manifestations in chronic pain conditions; (2) appraise the methodological quality of studies on measurement properties of those PROMs; (3) evaluate and summarize the quality of measurement properties of those PROMs based on their psychometric evidence and grading of the level of evidence for each measurement property.

Chapter 3

3 Materials and Methods

3.1. Study Protocol

The review protocol was registered in PROSPERO, the International Prospective Register of Systematic Reviews, with an assigned registration number CRD42023460050.

3.2. Eligibility Criteria

Inclusion criteria

The following inclusion criteria were applied:

1) Full-text articles published in English covering the period from the earliest available records of up to September 2023, the period during which the review was conducted; 2) Studies that evaluated CS manifestations using PROMs in adult patients with chronic pain conditions; 3) Studies that evaluated measurement properties (one or more) of either an original or translated version of PROM outlined by the COSMIN steering committee within three primary domains of reliability, validity, and responsiveness; 4) Additionally, this review considered studies that reported the development and/or assessment of the interpretability of the PROMs (e.g., evaluating the distribution of scores in the study population, floor and ceiling effects, the availability of scores and change scores for appropriate groups or subgroups, and the minimal important change). Note that criterion 4 was not a mandatory requirement for eligibility.

Exclusion criteria

1) Studies not aimed at evaluating the psychometric properties of a PROM; 2) Studies employing the PROM in the validation of another instrument; 3) Studies limited to opinion pieces, review articles, editorials, letters, case studies, conference abstracts, or technical notes.

3.3. Search Strategy

The following six electronic databases were searched: *MEDLINE, EMBASE, SCOPUS, PubMed, Web of Science, and CINAHL*, for studies evaluating measurement properties of PROMs assessing aspects of central sensitization in chronic pain conditions. A

comprehensive search of databases was conducted to locate the most relevant and highest-quality evidence. While PsycINFO is a valuable resource for psychology-specific studies, the chosen databases (specifically Scopus) sufficiently encompass relevant publications for current review's objectives. The main terms used for the database search were: conditions of interest (chronic pain, central sensitivity syndrome, central pain syndrome, Functional somatic syndrome, chronic overlapping pain condition, chronic musculoskeletal pain), central sensitization (centralized pain, central pain, nociplastic pain, central hyperexcitability, pain sensitization, pain hypersensitivity), psychometric properties (reliability, validity, responsiveness), and outcome measures (assessment, scale, instrument, patient-reported). Due to significant diversity in terminology related to measurement properties, this review applied search filters recommended by the COSMIN initiative to each database [140]. Additional searches were conducted by combining the names of identified instruments and applying the associated terms with the help of a librarian scientist. The details of these database searches are provided in Appendix A. The reference lists of identified articles were also manually screened to find more relevant studies.

3.4. Study Selection

Covidence which is a web-based collaboration software platform, was used to facilitate the organization of records retrieved from searches. This process enabled the detection and elimination of any instances of duplication. Two reviewers, one of whom was the author and the other a PhD student trained in conducting systematic reviews, independently assessed studies for eligibility. Conflicts arising during abstract screening were resolved by a third reviewer, who was a post-doctoral researcher. Initially, the titles and abstracts were screened based on the eligibility criteria for inclusion. The full-text screening was then performed on citations believed to be potentially eligible.

3.5. Data Extraction

Two reviewers extracted data from the eligible studies using a standardized data extraction form. Data extraction forms were first pilot-tested on a selection of studies and then revised by two reviewers to ensure all relevant information was collected from the included studies and updated if needed. The following information from the included studies and PROMs was extracted: (1) characteristics of included studies including author/year, country, name of PROMs, objectives of the study, sample size, mean age, sex, and chronic pain characteristics (type); (2) characteristics of PROMs including the name of the instrument, response option,

scoring system, and the original language of the instruments. Additionally, the author(s) responsible for the development of the PROM was referenced. If any study simultaneously evaluated the measurement properties of two or more instruments, the results of measurement properties for each instrument were separately collected and assessed.

3.6. Quality Assessment

3.6.1. Methodological Quality Assessment

The full-text articles that meet the criteria for eligible studies were evaluated for their quality using the COSMIN checklist [141]. In assessing the quality of studies and quality of outcome instruments, the COSMIN checklist's nine measurement properties are considered relevant for health-related PROMs. These measurement properties are categorized into three broad domains ‘Reliability’ (internal consistency, reliability, and measurement error), ‘Validity’ (content validity, structural validity, hypothesis testing for construct validity, cross-cultural validity/ measurement invariance and criterion validity) and ‘Responsiveness’.

Reliability

Reliability

“Reliability refers to the proportion of the total variance in the measurements which is due to ‘true’ differences between patients” [141]. Reliability can be assessed through test-retest reliability, intra-rater reliability or inter-rater reliability [142, 143]. Test-retest reliability measures the stability of test results over time by administering the same test to the same group of people on different occasions. Intra-rater reliability assesses the consistency of a single rater's measurements or ratings over different occasions. When different observers evaluate the same population at the same time, the reliability is assessed through inter-rater reliability [143,144].

The Intraclass Correlation Coefficient (ICC), Cohen’s kappa or weighted Kappa are used to assess different aspects of reliability, depending on the context. The preferred statistical method for continuous scores is the ICC [145] while for dichotomous scores or nominal scores is Cohen’s kappa [145]. Partial chance agreement should be considered when dealing with ordinal scales, making a weighted kappa a preferred choice for assessing reliability [145,146]. McGraw and Wong outlined 10 forms of ICC, classified by the “Model” (one-way random effects, two-way random effects, or two-way fixed effects), the “Definition” of the

key relationship (consistency or absolute agreement) and the “Type” (single rater/measurement or the average of multiple raters/measurements) [147]. The benchmarks for interpreting ICC values can vary depending on the variable or context. However commonly used thresholds are as follows: < 0.4= poor reliability, 0.4 to 0.59= fair reliability, 0.6 to 0.74= good reliability, and 0.75 to 1.0= excellent reliability [148]. The application of Pearson's and Spearman's correlation coefficients is considered inappropriate when the presence of systematic differences is uncertain, as these correlation measures do not consider systematic errors [141]. These correlation coefficients are useful for assessing linear relationships but ICC is specifically designed to evaluate the reliability and consistency of measurements, including systematic differences.

Internal consistency

“Internal consistency refers to the degree of interrelatedness among the items” [141]. Various metrics exist to quantify internal consistency, such as Split-half Reliability and McDonald's Omega, but the most commonly used measure is Cronbach's Alpha [149]. Internal consistency assesses the cohesiveness among items and the extent to which they estimate the same trait leading to reliable outcomes [143,150]. Cronbach’s alpha above 0.7 is often considered an indication of adequate internal consistency for a Scale [151]. However, this threshold can vary depending on the context and purpose of the scale. To ensure a meaningful interpretation of the internal consistency parameter, it is essential that the items collectively comprise a unidimensional scale or subscale [149,151]. Unidimentionality can be investigated for example by factor analysis [152] or item response theory methods.

Measurement error

“Measurement errors can occur systematically or randomly. Measurement error refers to the systematic and random error of an individual patient’s score that is not attributed to true changes in the construct to be measured” [141]. In studies that rely on Classical Test Theory, the recommended metric for quantifying measurement error is the Standard Error of Measurement (SEM) derived from a test-retest design. Alternative valid statistical approaches for evaluating measurement error include the Limits of Agreement (LoA) and the Smallest Detectable Change (SDC) [153]. Both of these parameters have a direct relationship with the SEM. Calculating SEM using Cronbach's alpha is generally considered inappropriate because it doesn't consider the variance that occurs between different time points [154].

Validity

Content validity

“Content validity is the degree to which the content of an instrument is an adequate reflection of the construct to be measured” [141]. Content validity is of paramount importance in measurement properties since it needs to be ensured that the items in a PROM are relevant, comprehensive, and comprehensible in relation to the specific construct under investigation and the target study population because lack of content validity may affect all other measurement properties. Assessing content validity involves the reviewers making subjective judgments, which consider the PROM development study, the quality, and findings of any supplementary content validity studies on the PROMs (if applicable), and the reviewers' subjective assessment of the PROMs' content [113].

Structural validity

“Structural validity refers to the degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured” [141] and is usually assessed by factor analysis or IRT/Rasch analysis. The concept of structural validity applies specifically to instruments that are designed based on a reflective model. In a reflective model, all the items within the instrument are reflections of a common underlying construct. To evaluate the structural validity of a PROM, factor analysis is conducted on every item in the PROM. This analysis aims to determine the number of expected subscales in the PROM and how well items group together within those subscales (i.e., structural validity studies). Multiple factor analyses are carried out on the items within each respective subscale to evaluate each subscale's unidimensionality. The purpose of these analyses is to ascertain whether each subscale independently captures a single construct (i.e., unidimensionality studies) [113]. While confirmatory factor analysis is the more favoured approach, both exploratory factor analysis and confirmatory factor analysis can be beneficial in assessing structural validity [155]. Confirmatory factor analysis reports model fit indices to assess how well the hypothesized factor structure fits the observed data. Exploratory factor analysis helps in identifying the number of factors (subscales) and the pattern of relationships between items and factors.

Cross-cultural validity

“Cross-cultural validity\measurement invariance refers to the degree to which the performance of the items on a translated or culturally adapted instrument is an adequate reflection of the performance of the items of the original version of the instrument” [141].

When a PROM is or will be used in a ‘culturally different population’ i.e., populations with different ethnicity, language, gender, age groups, or different patient populations, Cross-cultural validity\ measurement invariance needs to be evaluated. In Classical Test Theory, when evaluating cross-cultural validity, appropriate techniques include employing regression analyses or conducting multigroup confirmatory factor analysis. Conversely, within the framework of Item Response Theory, the focus shifts to examining Differential Item Functioning as the primary method for assessing cross-cultural validity [141,156]. Measurement invariant and non-Differential Item Functioning apply to examining whether individuals from distinct groups, who share the same underlying trait level while acknowledging group variations respond similarly to a particular item [157].

Criterion validity

“Criterion validity refers to the degree to which the scores of a PROM are an adequate reflection of a gold standard” [141]. The recommended statistical approach for Criterion validity is correlation analysis when both the PROM and the gold standard yield continuous scores. The area under the receiver operating characteristic (ROC) is the preferred method if the instrument scores are continuous but scores on the gold standard are dichotomous. However, if both the instrument scores and the gold standard generate dichotomous results, the preferred methods to employ are sensitivity and specificity analysis [141].

Hypotheses testing for construct validity

“Hypotheses testing for construct validity refers to the degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured” [141].

Responsiveness

“Responsiveness refers to the ability of a PROM to detect change over time in the construct to be measured” [141]. While responsiveness and validity are distinct measurement

properties, the primary distinction between cross-sectional (construct and criterion) validity and responsiveness lies in the fact that validity assesses the accuracy of a single score, whereas responsiveness evaluates the accuracy of a change score [150].

In this review, the quality assessment of the included PROMs involved three sequential stages. Two reviewers independently conducted the quality assessments. In cases where a consensus couldn't be reached, reviewer conflicts were referred to an expert- a professor and researcher with specialized knowledge and experience in this area, who helped in reaching a consensus.

3.6.2. Stage 1. COSMIN Risk of Bias Checklist

Evaluation of the methodological quality of each study was performed using the COSMIN risk of bias checklist [113,116]. There are 4 to 13 items in each box of the checklist for each measurement property which are evaluated using a four-point rating system: 'V' for very good, 'A' for adequate, 'D' for doubtful, and 'I' for inadequate. The risk of bias for each study was rated using this four-point scale, and the overall rating relied on the lowest favourable rating (which is called "worst score counts") given to any of the items within each measurement property [141].

3.6.3. Stage 2. Employing updated Criteria for Good Measurement Properties

Stage 2a: Content validity

Each study's findings on PROM development and content validity were evaluated using ten criteria for content validity [158]. The evaluations of all available studies were then qualitatively summarized to determine whether the relevance, comprehensiveness, comprehensibility, and overall content validity were sufficient (+), insufficient (-), indeterminate (?), or inconsistent (±) considering all evidence [158]. If the content validity assessment of the PROM receives an inadequate rating, it should not be suggested for use and will consequently be omitted from subsequent evaluations of the other measurement properties [141].

Stage 2b: Remaining measurement properties

For instruments with sufficient content validity, each study's results on the remaining measurement properties were evaluated using updated criteria for good measurement properties and rated as sufficient (+), insufficient (-), or indeterminate (?) [113]. These

updated criteria, drawn from the COSMIN manual for systematic reviews of PROMs, are detailed in Table 1 [141].

Table 1: Updated criteria for good measurement properties

Measurement properties	Rating	Quality Criteria*
Structural validity	+	CTT: CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 IRT/Rasch: No violation of unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 AND no violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 OR Q3's<0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: IRT: $\chi 2 > 0.01$ Rasch: infit and outfit mean squares ≥0.5 and ≤1.5 OR Z-standardized values >−2 and <2
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence for sufficient structural validity AND Cronbach's alpha(s) ≥0.70 for each unidimensional scale or subscale
	?	Criteria for "At least low evidence for sufficient structural validity" not met
	-	At least low evidence for sufficient structural validity AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale
Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA < MIC
	?	MIC not defined
	-	SDC or LoA > MIC

Measurement properties	Rating	Quality Criteria*
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis
Cross-cultural validity\measurement invariance	+	No important differences were found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
	?	No multiple group factor analysis OR DIF analysis performed
	-	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis OR AUC < 0.70

* Quality criteria drawn from COSMIN manual [141], "+" = sufficient, "–" = insufficient, "?" = indeterminate, AUC = area under the curve, CFA = confirmatory factor analysis, CFI = comparative fit index, CTT = classical test theory, DIF = differential item functioning, ICC = intraclass correlation coefficient, IRT = item response theory, LoA = limits of agreement, MIC = minimal important change, RMSEA: root mean square error of approximation, SEM = standard error of measurement, SDC = smallest detectable change, SRMR: standardized root mean residuals, TLI = Tucker–Lewis index.

Although the range of fit indices for confirmatory factor analysis has been determined in the updated COSMIN criteria, the cutoff value for acceptable factor loading in exploratory factor analysis (EFA) was not specified. In this review, a cutoff value of 0.4 for factor loading was applied for EFA based on existing publications [159]. As the target population was chronic pain patients, a consensus was made to consider 5 days to 2 weeks as the appropriate time interval for two administrations of the same questionnaire for test-retest. According to the updated guideline from COSMIN, the methodological quality of included studies is now independent of whether or not previous hypotheses were formulated within those studies. Because the reported results, such as correlations or mean differences between groups, may still provide valuable information and not necessarily biased just because the study lacked pre-defined hypotheses [141]. To evaluate convergent validity (construct validity), generic

hypotheses recommended in the COSMIN guideline were followed to determine the expected direction (positive or negative) and magnitude (absolute or relative) of the correlations between the PROM under study and the comparator instruments [141] for all included studies with or without priori hypothesis. Furthermore, according to guidelines, for assessing discriminative validity (construct validity), a predetermined expectation for the difference in total PROM scores between patient subgroups in comparison to a pain-free control group or between different patient subgroups was set. The threshold was set at 10 points or more.

3.6.4. Stage 3. Summary of Evidence

Stage 3a. Content validity

Each PROM's overall ratings of content validity were determined, followed by grading the level of evidence as high, moderate, low, or very low using a modified GRADE approach. The quality of evidence according to the GRADE approach is usually determined by considering five factors: risk of bias (quality of the studies), inconsistency (variability in the outcomes of the studies), imprecision (total sample size of the available studies), indirectness (evidence derived from different populations, interventions, or outcomes that differ from those of interest in the review) and publication bias [122]. However, only four factors (risk of bias, inconsistency, imprecision, and indirectness) were applicable for evaluating measurement properties in systematic reviews of PROMs [141]. Assessment of publication bias in studies on measurement properties represents a significant challenge due to the lack of registries of these types of studies [113, 141]. Therefore, as per the COSMIN user manual [141], this factor was not taken into consideration during the grading of the level of evidence.

Stage 3b. Remaining measurement properties

The results of available studies for each measurement property need to be consistent to come to an overall conclusion on the quality of a PROM. If the results show inconsistencies (e.g., indicating both sufficient and insufficient reliability), it is essential to investigate the underlying reasons for these discrepancies. The findings of available studies were summarized qualitatively and rated again using the updated COSMIN criteria to conclude whether, overall, the measurement properties of each PROM are sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?). As a general guideline, a minimum of 75% of the results should align with the criteria to rate the qualitatively summarized results as sufficient (or insufficient).

Afterward, the quality of the evidence of each measurement property was graded using the GRADE approach. Table 2 presents the Modified GRADE approach for grading the quality of evidence for each measurement property per PROM and Table 3 presents Guidance on reducing grade for the risk of bias adapted from COSMIN guidelines [141].

Table 2: Modified GRADE approach*

Quality of evidence	Lower if
High	Risk of bias -1 Serious -2 Very serious -3 Extremely serious Inconsistency -1 Serious -2 Very serious Imprecision -1 total n=50-100 -2 total n<50 Indirectness -1 Serious -2 Very serious
Moderate	
Low	
Very low	

*Modified GRADE approach was drawn from the COSMIN manual for systematic reviews of PROMs [141]

Table 3: Guidance on reducing grade for risk of bias

Risk of bias	Downgrading for Risk of Bias*
No	There are multiple studies of at least adequate quality, or there is one study of very good quality available
Serious	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
Very serious	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
Extremely serious	There is only one study of inadequate quality available

* Downgrading guidance for Risk of Bias drawn from the COSMIN manual for systematic reviews of PROMs [141]

3.7. Statistical Analysis

All the results of available studies on measurement properties were first summarized qualitatively and overall ratings were given against updated criteria. However, where appropriate and feasible, the results were also statistically pooled in a meta-analysis. The decision to pool results was based on the availability of data and the consistency of results. In this review, data related to the reliability coefficient (intraclass correlation coefficient) were quantitatively pooled specifically when studies reported the required parameter estimates including standard errors (and/or 95% confidence interval) for meta-analysis as the best measure of reliability for continuous data is the intraclass correlation coefficient (ICC). Then a meta-analysis of the reliability coefficient was conducted by fitting a random-effects model specified with the restricted maximum likelihood method using Stata (version 16.4). To facilitate the interpretation of the meta-analysis results, the average reliability coefficients and their confidence limits, obtained through Fisher's Z transformations, were converted back to intraclass correlation metrics. Heterogeneity was evaluated using the Q statistic and the I^2 index and visualized through a forest plot. Heterogeneity is defined in accordance with Cochrane guidelines [158]. Again, for construct validity, all correlations of PROMs under review with comparator PROMs measuring similar constructs need to be considered for pooling [141]. In this review majority of comparator PROMs were with related but dissimilar constructs. Therefore, quantitative pooling was not appropriate for construct validity. A ‘Do file’ outlining the codes for meta-analysis is included in Appendix B.

Chapter 4

4 Results

4.1. Search Results

The search results revealed 4415 studies. Following the removal of duplicates, 2739 studies were retained. After screening the titles and abstracts, 2574 studies were excluded as they were irrelevant to current review objectives. Subsequently, 165 articles were included in the full-text screening. In the full-text review, 109 articles were excluded for not meeting the inclusion criteria. Finally, 58 studies were found to be eligible for this review. The PRISMA flow chart presented the details of the study selection process (Figure 1).

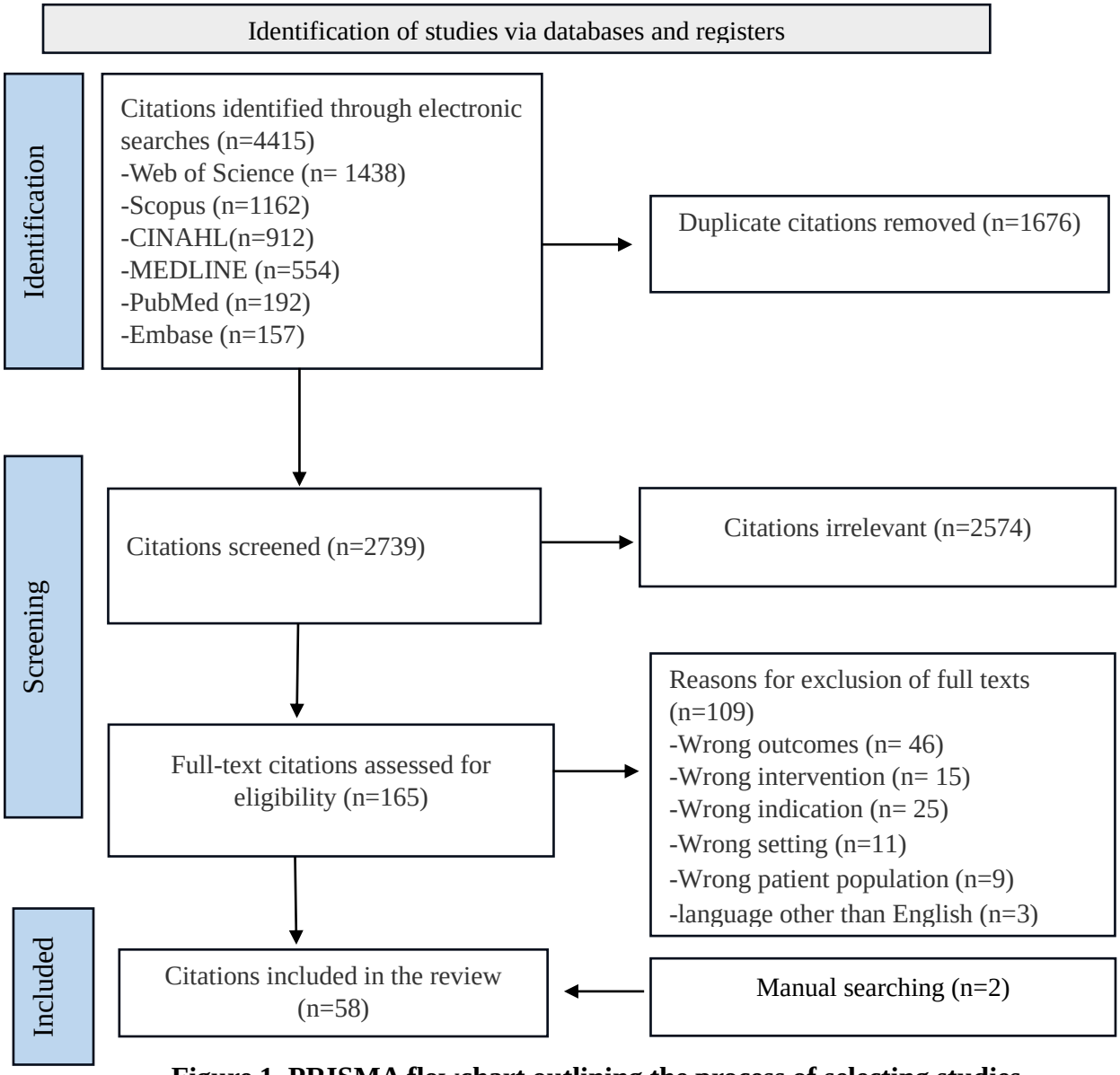

Figure 1. PRISMA flowchart outlining the process of selecting studies.

4.2. Characteristics of Included Studies

The characteristics of the 58 included studies which addressed eight instruments [159-216] are presented in Table 4. Sixteen studies [159,163,172,175,180,187,193,194,196,208,210,212-216] investigated the measurement properties of the original version of PROMs. Fourty-two studies [160-162,164-171,173,174,176-179,181,182-186,188-192,195,197-207,209] assessed the measurement properties of the translated version of the original questionnaire. The number of participants in the studies varied and ranged from 31 [160] to 1651 [204]. The mean age of patients was 48 years (SD = 11.8) with ages ranging from 22 years (SD = 4.7) [159] to 71 years (SD = 7.7) [184]. The distribution of male and female participants in the studies varied, with some studies including only female participants [187,201,209,211,212] and the remaining studies having a disproportionate ratio of females to males. A wide range of various chronic pain conditions was included across the studies including fibromyalgia (22 studies), chronic widespread pain (5), chronic low back pain (14), chronic neck pain (4), osteoarthritis (5), temporomandibular joint disorder (4), migraine/headaches (4), knee osteoarthritis (2), rheumatoid arthritis (9), chronic spinal pain (5), myofascial pain syndrome (2), persistent pelvic pain (1), mixed chronic pain (6), musculoskeletal pain (Hip pain, leg pain, ankle pain, shoulder pain) (9), central sensitization syndrome (CSS) (3), complex regional pain syndrome (6). However, 21 studies were conducted only on a single chronic pain condition, (e.g. chronic low back pain [160], chronic neck pain [165]) while 37 studies took a broader approach, incorporating multiple chronic pain disorders into a single study. Across the studies, patients with fibromyalgia, chronic low back pain, and musculoskeletal pain were most commonly included among the various chronic pain conditions.

Table 4. Characteristics of Included Studies

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Mayer et al., 2012 [159]	USA	CSI	To develop the Central Sensitization Inventory (CSI) and to assess the psychometric validity and clinical utility of the CSI to differentiate among different subject groups of chronic pain patients.	Study 1: 149 participants for Reliability and 359 participants including both chronic pain patients (210 patients) and healthy participants for factor analysis Study 2: 105 chronic pain patients and 40 healthy participants	Study 1: 22 ± 4.7 Study 2: FM= 47 ± 7.6, CWP= 46 ± 11.5, CLBP= 43 ± 10.0, Control group= 21 ± 13.6	Study1: 70 Study 2: FM= 73, CWP=26, CLBP= 25, Control group= 77	Mixed chronic pain	Fibromyalgia (FM n=30), Chronic widespread pain without FM (CWP n=31), Chronic low back pain (CLBP n=44)
Bid et al., 2016 [160]	India	CSI	To translate and cross-culturally adapt the CSI into Gujarati, and subsequently check its psychometric properties among patients with CLBP.	20 patients for content validity, 31 patients for other psychometric evaluation	53 ± 13.2	23 (74.2)	Chronic Musculoskeletal pain	Chronic low back pain (CLBP)
Noorollahzadeh et al., 2021 [161]	Iran	CSI	To perform translation, cultural adaptation of CSI into Persian and to assess its psychometric properties.	20 patients for pretesting, and 256 patients for validation studies	patient group= 42 ± 11.8, Control group = = 36 ± 9.9	98 (38.3)	Mixed chronic pain	Chronic regional pain syndrome, restless leg syndrome, hip, knee and spinal (cervical, thoracic and lumbar), osteoarthritis, Spondylolisthesis, frozen shoulder, Coccydynia, trigeminal neuralgia
Liang et al., 2022 [162]	China	CSI	To adapt CSI-25 to Chinese and test its validity and reliability	6 patients for pilot testing, 237 patients for validation studies	44 ± 12.7	223 (76.4)	Mixed chronic pain	Fibromyalgia (n=114), musculoskeletal pain (n=123: Hip pain, knee pain, ankle pain, shoulder pain, elbow pain, hand and wrist pain, lateral epicondylitis and temporomandibular joint pain, lumbago, back pain, cervicodynia)
Roby et al., 2022 [163]	Canada	CSI	To evaluate the validity of the CSI through Rasch analysis in knee osteoarthritis patients	293 patients for validation studies	64 ± 9.5	172 (58.7)	Chronic Musculoskeletal pain	Knee osteoarthritis
Bakhtadze et al., 2022 [164]	Russia	CSI	To investigate the psychometric properties of the Russian version of the Central Sensitization Inventory in patients with non-specific neck pain associated with migraine and tension-type headache.	204 patients	38 ±10.5	173 (84.8)	Mixed chronic pain	non-specific neck pain(n=37), non-specific neck + episodic headache (n=30), non-specific neck + chronic headache (n=137)
Wiangkham et al., 2022 [165]	Thailand	CSI	To translate and cross-culturally adapt the CSI into Thai, and subsequently evaluate its psychometric properties among patients with chronic non-specific neck pain.	30 patients for the cross-cultural adaptation process, and 340 patients for psychometric evaluation	34 ± 14 (20–68)	226 (66.50)	Chronic musculoskeletal pain	chronic non-specific neck pain

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Hendriks et al., 2020 [166]	Netherland	CSI	To evaluate the convergent validity of the dutch version of the Central Sensitization Inventory in chronic whiplash-associated patients.	125 patients	40 ± 11.3 (18-65)	71 (56.8)	Chronic musculoskeletal pain	chronic whiplash-associated disorder
Bakhtadze et al., 2021 [167]	Russia	CSI	To evaluate the validity and reliability of the Russian version of the Central sensitization inventory in patients with chronic non-specific neck and back pain.	195 patients for validation studies	41 ±11.4 (18-65)	142 (72.8)	Chronic musculoskeletal pain	chronic non-specific neck pain (n=106) and /or back pain (n=75)
Sharma et al., 2020 [168]	Nepal	CSI	To translate and culturally adapt the CSI into Nepali and to test its internal consistency, test-retest reliability, measurement error and construct validity among patients with subacute and chronic musculoskeletal pain.	20 patients for pretesting and 115 patients for psychometric evaluation	42 ± 14.6	67 (58)	Chronic musculoskeletal pain	
Düzce Keleş et al., 2021 [169]	Turkey	CSI	To translate the CSI into Turkish and perform a psychometric validation among patients with chronic spinal pain with an organic origin (CSPO), and fibromyalgia patients.	200 patients and 100 controls for psychometric evaluation	Fibromyalgia= 45 ±8.4 (25 - 60), CSPO = 44 ±9.7 (21 - 60), Healthy control = 36 ±10.1 (25 - 55)	276 (92)	Mixed chronic pain	Fibromyalgia patients with widespread pain(n=100), chronic spinal pain with an organic origin (CSPO n=100)
Salaffi et al., 2022 [170]	Italy	CSI	To evaluate convergent and discriminant validity of the Italian version of CSI in patients with fibromyalgia (FM).	562 FM patients	53 ± 9.6	511 (90.1)	Mixed chronic pain	Fibromyalgia
Tanaka et al., 2017 [171]	Japan	CSI	To test criterion validity and construct validity of the Japanese version of the CSI and to investigate prevalence rates of CS severity levels in patients with musculoskeletal disorders.	6 patients for pretesting, 290 patients for validation studies	51 ± 15.6	188 (64.83)	Mixed chronic pain	Musculoskeletal pain disorders (Neck injury (including whiplash) shoulder, low back, hip, knee, or ankle pain) Restless leg syndrome, Chronic fatigue syndrome, Irritable bowel syndrome, Temporomandibular joint disorder, Migraine, or tension headaches [No CSS (n=209), 1 CSS (n=63), 2 and above CSSs (n= 18)]
Neblett et al., 2013 [172]	USA	CSI	To investigate the original version of CSI scores in chronic pain patients with Central sensitization syndrome, and nonclinical samples to determine a clinically relevant cutoff value.	121 patients and 129 nonpatient comparison sample	CSS patients = 45 ± 13.3, non-CSS patients = 46 ± 12.2, non-patient comparison sample = 21 ± 3.6	167 (66.8)	Mixed chronic pain	Central sensitization syndrome (CSS n=89: tension headaches/migraines, myofascial pain syndrome, fibromyalgia, IBS, TMD and PTSD), non-CSS (n=32)
Caumo et al., 2017 [173]	Brazil	CSI	To examine the psychometric characteristics of the Brazilian-adapted Central Sensitization Inventory (CSI-BP), including internal consistency, construct validity, reproducibility, and factor structure among patients with chronic musculoskeletal pain.	20 patients for pretesting, 222 patients and 63 nonpatients for comparison sample	OA= 67 ±8.2 MPS= 43 ± 11.5 CTTH= 36 ± 12.2 FM= 50 ±11 Control group= 38 ± 14.34	248 (87)	Mixed chronic pain	Osteoarthritis (OA n=31), myofascial pain syndrome (MPS n=65), chronic tension-type headache (CTTH n=53), fibromyalgia (FM n=73),

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Kregel et al., 2016 [174]	Netherland	CSI	To perform Psychometric properties assessment of Dutch CSI among patients with chronic pain.	368 patients for validation studies.	Chronic pain patients= 43 ± 13.2 Healthy control = 37 ±14.8	269 (64.5)	Chronic Musculoskeletal pain	
Coronado and George, 2018 [175]	USA	CSI	To assess construct validity and concurrent validity of the original version of CSI and PSQ among patients with shoulder pain.	78 patients	Patients group= 39 ± 14.5 Control group= 35 ± 11.1	36 (46.2)	Chronic Musculoskeletal pain	Shoulder pain
Feng et al., 2022 [176]	China	CSI	To create the Chinese Cultural Adaptation of the CSI and assess its psychometric properties in chronic pain patients.	30 patients for pretesting, and 235 patients for validation studies	64 ± 9.5	196 (83.40)	Chronic Musculoskeletal pain	
Cuesta-Vargas et al., 2016 [177]	Spain	CSI	To translate the CSI into Spanish and subsequently conduct its psychometric validation among chronic pain patients.	395 patients	56 ± 12.7	176 (44.4)	Chronic Musculoskeletal pain	
Knezevic et al., 2018 [178]	Serbia	CSI	To translate the original CSI into Serbian and to examine its psychometric properties in chronic pain patients.	8 patients for cognitive debriefing and 355 patients for validation studies	52 ± 12.9	233 (65.6)	Mixed chronic pain	complex regional pain syndrome(n=48), fibromyalgia (n=23), Low back pain, cervical pain, knee pain, hip pain, ankle pain, shoulder pain, lateral epicondylitis, temporomandibular joint pain
Mikkonen et al., 2021 [179]	Finland	CSI	To translate and cross-culturally adapt the CSI into Finnish (CSI-FI) and subsequently test its reliability and validity.	20 patients for face validity, 187 patients and 42 controls for validation studies	Chronic pain group: 40 ± 10.6, Control group: 46 ± 11.8	162 (70.7)	Mixed chronic pain	Chronic low back pain, other chronic musculoskeletal pain, chronic headache
Neblett et al., 2017 [180]	USA	CSI	To establish a gradient of clinically relevant symptom severity levels for the original version of CSI.	287 patients (Study-2)	Average age= 50	201 (70)	Mixed chronic pain	CSS- Fibromyalgia, Tension/Migraine Headaches, Post-traumatic Stress Disorder, Restless Leg Syndrome, Temporomandibular Joint Disorder, Chronic Fatigue Syndrome, Irritable Bowel Syndrome [No CSS (n=120), 1 CSS (n=109), 2 CSSs (n=40), 3 CSSs (n=6), 4+ CSSs(n=12)]
Kosińska et al., 2021 [181]	Poland	CSI	To validate the Polish CSI in patients with chronic spinal pain.	151 patients and 30 healthy controls	patient group= 56 ± 14.1 controls= 42 ± 12.6	145 (80.1)	Chronic Musculoskeletal pain	chronic neck pain (CNP n=24), Chronic low back pain (CLBP n=73), both CNP and CLBP(n=54)

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Bilika et al., 2020 [182]	Greece	CSI	To translate and cross-culturally adapt the CSI into Greek and to perform psychometric properties assessment among patients with chronic pain.	20 patients for pilot testing and 200 patients for validation studies	patient group= 49 ± 14.7, controls= 28 ± 8.7	149 (59.6)	Mixed chronic pain	Low back pain, cervical pain, knee pain, hip pain, ankle pain, shoulder pain, epicondylitis, temporomandibular joint pain (multiple pain complaint n=55; single pain complaint n=115), and fibromyalgia (n=30)
Kregel et al., 2018 [183]	Belgium	CSI	To assess the convergent validity of the Dutch version of the CSI, and to explore its relationship with psychophysical pain measures and self-reported assessments in patients experiencing chronic spinal pain.	116 patients	40 ± 12.5	72 (62.1)	Chronic musculoskeletal pain	Non-specific chronic spinal pain (including CLBP n=54; chronic idiopathic neck pain n=62)
Kim et al., 2020 [184]	Korea	CSI	To translate the CSI into the Korean version and investigate its psychometric properties in patients with knee OA.	20 patients for pilot testing and 269 patients for validation studies	71 ± 7.7	236 (87.7)	Chronic musculoskeletal pain	Knee osteoarthritis
van der Noord et al., 2018 [185]	Netherland	CSI	To investigate the convergent validity and to present clinically relevant categories for the Dutch CSI in chronic pain patients.	198 patients	47 ± 15.5	115 (58.1)	Chronic pain	
Chiarotto et al., 2018 [186]	Italy	CSI	To cross-culturally adapt the CSI into Italian, and subsequently evaluate its structural and construct validity in chronic pain patients.	20 patients for pilot testing and 220 patients for validation studies	55 ± 15.5	172 (78.8)	Mixed chronic pain	Low back pain (LBP n=73), temporomandibular disorder (TMD n=37), hand osteoarthritis (HOA n=43), fibromyalgia (FM n=20), or rheumatoid arthritis (RA n=44).
Valera-Calero et al., 2022 [187]	Spain	CSI	To analyze the convergent validity of CSI with psychological and psychophysical outcomes in patients with Fibromyalgia.	126 patients	53 ±11.0	126 (100)	Mixed chronic pain	Fibromyalgia
Madi et al., 2021 [188]	Jordan	CSI	To adapt CSI into Arabic and to investigate its psychometric properties in chronic pain patients.	15 patients for pre-testing and 171 patients for validation studies	37 ± 12.9	128 (74.9)	Mixed chronic pain	
Knezevic et al., 2020 [189]	Serbia	CSI	To explore evidence of convergent and discriminant validity of the serbian version of the CSI in chronic pain patients.	399 patients and 146 control subjects	49 ± 14.5	361 (67.5)	Mixed chronic pain	Low back pain (n=155, cervical pain(n=26), knee pain, hip pain, ankle pain, shoulder pain, lateral epicondylitis (Pain in 2 or more locations n= 95), temporomandibular joint pain (TMJ n=20), complex regional pain syndrome (n=46), and fibromyalgia (FM n=47)

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Van Wilgen et al., 2018 [190]	Netherland	CSI	To examine the convergent validity of the Dutch version of the CSI in chronic pain patients	114 patients	47 ± 15.9	64 (56.1)	Chronic pain	
Klute et al., 2021 [191]	Germany	CSI	To cross-culturally adapt the CSI into German version, and subsequently evaluate its psychometric properties in chronic pain patients.	247 patients and 63 controls	55 ± 13.1	217 (70)	Mixed chronic pain	Fibromyalgia syndrome (FMS n=37), Multisite chronic pain (MCP n=63), Regional chronic pain (RCP n=17), Chronic back and/or neck pain (CBNP n=83), Rheumatoid arthritis in remission (RAR n=47)
Holm et al., 2021 [192]	Denmark	CSI	To examine the convergent validity of the Danish version of the CSI with quantitative sensory testing and self-reported psycho-social questionnaires in patients with LBP.	168 patients	56 ± 14.9	65 (39.1)	Chronic musculoskeletal pain	Low back pain
Neblett et al., 2015 [193]	USA	CSI	To determine the ability of the CSI to distinguish between chronic pain patients, with and without central sensitivity syndromes (CSSs).	161 patients	44 ± 11.6	92 (57.14)	Chronic pain	
Neblett et al., 2017 [194]	USA	CSI	To assess CSI scores, and their associations with other clinically relevant psychosocial variables, in patients with chronic spinal pain disorder.	763 patients	47± 10.6	270 (35%)	Chronic musculoskeletal pain	Chronic spinal pain
Bid et al., 2017 [195]	India	CSI	To compare whether the McKenzie exercise program (MEP) reduces CS better than the Conventional physiotherapy program (CPP) in patients with chronic non-specific low back pain by using Gujarati version of CSI.	128 patients	Patients group: 41 ± 7.3 Control group: 41 ± 7.76	Experimental: 36 (46.2) Control: 42 (53.8)	Chronic musculoskeletal pain	Chronic non-specific low back pain.
Ruscheweyh et al., 2012 [196]	Germany	PSQ	To assess the validity of the PSQ in chronic pain patients.	134 patients and 185 healthy control subjects	control= 45 ± 21; CTTH= 42 ± 17 CLBP= 49 ± 15; TMD= 48 ± 14; Mixed= 50 ± 13	195 (61.1)	Mixed chronic pain	chronic tension-type headache, chronic low back pain (CLBP), chronic temporomandibular disorder (TMD), mixed chronic pain
Sellers et al., 2013 [197]	USA	PSQ	To validate the English version of PSQ in the chronic pain population.	136 patients	54 ± 14	83 (61.0)	Chronic Musculoskeletal pain	Chronic low back pain (CLBP)
Ibancos-Losada et al., 2021 [198]	Spain	PSQ	To cross-culturally adapt the PSQ into Spanish and subsequently analyze its psychometric properties in fibromyalgia syndrome patients.	15 participants for pilot testing 58 patients and 296 controls for validation studies	37 ± 11.9	235 (66.4)	Mixed chronic pain	Fibromyalgia syndrome (FMS)
Inal et al., 2021 [199]	Turkey	PSQ	To translate the PSQ into Turkey and investigate its validity in chronic pain patients.	10 participants for pilot testing and 73 patients for validation studies	57 ± 15.4 (18-90)	48 (65.8)	Chronic Musculoskeletal pain	Chronic back pain
Latka et al., 2019 [200]	Poland	PSQ	To translate and cross-culturally adapt the CSI into Polish and to perform psychometric properties assessment among patients with low back pain.	12 patients for pretesting and 144 patients for validation studies	53 (19–80)	64 (44.4)	Chronic Musculoskeletal pain	Low back pain

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Grundström et al., 2019 [201]	Sweden	PSQ	To assess associations between the Sweedish version of PSQ scores and QST and to determine the extent of psychological distress influenced PSQ scores.	37 patients	Patients group= 26 ± 5.9 (18-40) Control= 30 ± 5.6 (18-40)	37 (100)	Mixed chronic pain	persistent pelvic pain
Kim et al., 2014 [202]	Korea	PSQ	To translate and cross-culturally adapt the CSI into Korean and subsequently conduct its psychometric validation among chronic pain patients.	30 patients for pretesting and 72 patients for validation studies	66 ± 8.1	45 (62.5)	Chronic Musculoskeletal pain	chronic leg pain and/or back pain caused by degenerative spinal disease
Coronado and George, 2018 [175]	USA	PSQ	To assess construct validity and concurrent validity of CSI and PSQ among patients with shoulder pain.	78 patients	Patients group= 39 ± 14.5, Control group= 35 ± 11.1	36 (46.2)	Chronic Musculoskeletal pain	Shoulder pain
Carrillo-de-la-Peña et al., 2015 [203]	Spain	FSQ	To evaluate the extent of agreement between the 1990 and 2010 diagnostic criteria of the American College of Rheumatology (ACR) and to validate the Spanish adaptation of the Fibromyalgia Survey Questionnaire (FSQ).	65 patients	50 (22-65)	64 (98.5)	Mixed chronic pain	Fibromyalgia
Häuser et al., 2012 [204]	German	FSQ	To validate the German version of FSQ in fibromyalgia patients.	1651 patients	54 ± 9.8 (19–84)	1562 (94.8)	Mixed chronic pain	Fibromyalgia syndrome (FMS), Chronic widespread pain
Fors et al., 2020 [205]	Norway	FSQ	To validate the Norwegian version of FSQ against the ACR1990 criteria.	120 patients and 62 controls	FM= 53 ± 10.9 Control= 48 ± 13	FM= 119 (99.2) Control= 59 (95.2)	Mixed chronic pain	Fibromyalgia
Jiao et al., 2023 [206]	China	FSQ	To assess the reliability and validity of the ACR 2011 and 2016 survey diagnostic criteria for fibromyalgia in China.	2 patients for pretesting, 200 FM patients and 200 RA patients (as control) for validation studies	FM group= 49 ± 13.4, RA group (Control) = 49 ± 13.2	FM= 174 (87.0) Control= 174 (87.0)	Mixed chronic pain	Fibromyalgia (FM n=200), Rheumatoid arthritis (RA n=200)
Kang et al., 2019 [207]	Korea	FSQ	To validate the Korean version of 2016 revised version of the 2010/2011 Fibromyalgia Survey Questionnaire (FSQ)	30 patients for pretesting, 86 FM and 89 other rheumatological disorders patients for validation studies	51 ± 11.8	135 (71.1)	Mixed chronic pain	Fibromyalgia (FM n=86), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), osteoarthritis (OA), and myofascial pain syndrome (MPS) [all other disorders n=89]
Moore et al., 2022 [208]	USA	FSQ	To examine the correlation between FSQ and QST as measures of pain centralization among patients with Rheumatoid arthritis.	285 patients	55 ± 13.7	234 (82.1)	Chronic musculoskeletal pain	Rheumatoid arthritis
Aguirre Cárdenas et al., 2021 [209]	Chile	FSQ	To examine the psychometric properties of the Chilean version of the Fibromyalgia Survey Questionnaire (FSQ).	290 Patients and 117 participants without chronic pain	49 ± 14.3	407(100)	Mixed chronic pain	Fibromyalgia (FM n=194), Rheumatoid arthritis (RA n=96)

Table 4. Characteristics of Included Studies *cont.*

Author (ref)	Country	PROM	Objective of the study	Sample size (N)	Age mean ± SD (range) year	Sex (% Female)	Chronic pain	
							Type	Conditions
Neville et al., 2018 [210]	USA	FSQ	To assess the convergent validity of FSQ in knee osteoarthritis patients.	129 patients	Female: 64 ± 8.6 Male: 65 ± 8.7	68 (52.7)	Chronic musculoskeletal pain	Rheumatoid arthritis
Bidari et al., 2015 [211]	Iran	FSQ	To evaluate the validity of the Persian version of the FSQ and Polysymptomatic Distress Scale (PSD) in chronic pain patients.	263 patients	FM group= 42 ± 11 non-FM group= 48 ± 11	263(100)	Mixed chronic pain	Fibromyalgia (FM n=169), osteoarthritis, periarthritis, regional pain syndromes [non-FM n=94]
Ghavidel-Parsa et al., 2022 [212]	Iran	NFF	To develop and validate the preliminary Nociplastic-based Fibromyalgia Features (NFF) in patients with chronic pain.	185 patients	FM= 45 ± 9.9 NON-FM= 48 ± 11.5 (18-65)	185 (100)	Mixed chronic pain	Fibromyalgia, non-FM non-inflammatory chronic pain (osteoarthritis, tendonitis or painful periarticular conditions, mechanical low back pain, mechanical neck pain)
van Bemmel et al., 2019 [213]	Dutch	GPQ	To develop and evaluate the psychometric performance of this generalized pain questionnaire (GPQ) in patients with FM and rheumatoid arthritis (RA).	212 patients	FM= 45 ± 11.6 RA= 60 ± 12.1	164 (77.4)	Mixed chronic pain	Fibromyalgia (FM n=98), Rheumatoid arthritis (RA n=114)
Dixon et al., 2016 [214]	USA	SHS	To develop and validate the Sensory Hypersensitive Scale (SHS).	Study 4: 124 patients and 66 healthy controls Study 5: 5 patients and 44 healthy controls	Study 4: FM= 44 ± 11.3; Osteoarthritis= 59 ± 8.0, Osteoarthritis with FM= 56 ± 8.2, Healthy control= 35 ± 13.2 Study 5: Patient group=41 ± 11.3 Control group= 40 ± 11.6	Study 4: 190 (100) Study 5: 51 (49.5)	Mixed chronic pain	Fibromyalgia, Rheumatoid arthritis, chronic low back pain
Austin et al., 2020 [215]	Australia	novel self-report instrument	To develop and assess psychometric properties of a novel self-report questionnaire to assess symptoms associated with altered central nervous system pain processing in people with and without chronic pain.	99 patients and 84 healthy controls	Case= 57 ± 18.3 (19-89), Control= 41 ± 15.2	153 (83.6)	Mixed chronic pain	Osteoarthritis, Widespread pain, Neck and back pain, headaches/migraine, Abdominal/pelvic pain, Peripheral neuropathies, Radiculopathy
Ten Brink et al., 2021 [216]		L-VISS and VDS	To validate L-VISS and VDS scale in chronic pain patients.	185 patients and 125 pain-free controls		251 (81)	Mixed chronic pain	Complex Regional Pain Syndrome (CRPS) fibromyalgia, general chronic pain

The mean age was rounded to the nearest whole number, rather than being displayed in decimal format. SD= Standard deviation, CSI=Central sensitization Inventory, PSQ= Pain Sensitivity Questionnaire, FSQ= Fibromyalgia Survey Questionnaire, NFF= Nociplastic-based Fibromyalgia Feature, GPQ= Generalized Pain Questionnaire, SHS= Sensory Hypersensitivity Scale, L-VISS and VDS= Leiden Visual Sensitivity Scale and Visual Discomfort Scale, PROM= Patient-reported outcome measure.

4.3. Summary of Included PROMs

The included studies evaluated eight different instruments. The number of items/ questions across the PROMs ranged from seven to twenty-five, with the total number of subscales or domains varying between two and nine. The most frequently assessed instruments for the pain associated with central sensitization were Central Sensitization Inventory (CSI) (total of 37 studies), Fibromyalgia Survey Questionnaire (FSQ) (total of 9 studies) and Pain Sensitivity Questionnaire (PSQ) (total of 8 studies). Five studies were identified for other instruments, including Nociplastic-based Fibromyalgia Features (NFF), Generalized Pain Questionnaire (GPQ), Sensory Hypersensitivity Scale (SHS), Novel self-report Questionnaire, Leiden Visual Sensitivity Scale & Visual Discomfort Scale and each of these instrument's psychometric properties were assessed in a single study. The characteristics of eight identified PROMs are presented in Table 5. However, in a study conducted by Coronado and George (2018) [175], they concurrently evaluated two PROMs, the CSI and PSQ. Therefore, the results on the measurement properties of each instrument were reported and assessed separately in this review. This breakdown led to 37 studies being attributed to the CSI and 8 studies to the PSQ, even though the total number of studies included in the review was 58.

Table 5. Summary of PROM Identified

PROM	Developer	Subscales / Outcome domains	Number of items	Response options	Description	Range of scores/scoring	Original language	Available translations	Number of studies evaluating the instrument
Central Sensitization Inventory (CSI)	Mayer et al., 2012 [159]	4 factors: Physical functioning, Emotional Distress, Headache/Jaw Symptoms, Urological Symptoms	Two part- Part A: 25 items Part B: 10 items (this part is for information only and is not scored)	(0-4)	Each of these items is measured on a 5-point temporal Likert scale, with the following numeric rating scale: never =0, rarely = 1, sometimes =2, often =3, and always= 4.	Score range: 0 - 100.	English	19	37
Pain Sensitivity Questionnaire (PSQ)	Ruscheweyh et al., 2009 [217]	Domain: Pain Sensitivity	17-items.	0-10	11-point scale with 0 meaning “not painful at all” 1 meaning 'only just noticeable pain' and 10 meaning “worst pain imaginable.” PSQ-total score = summed score of PSQ moderate (items: 1, 2, 4, 8, 15, 16, and 17) and PSQ minor (items: 3, 6, 7, 10, 11, 12, and 14) subscale or factor.	0-10.	German	5	8
Fibromyalgia Survey Questionnaire (FSQ) / Fibromyalgia Survey Diagnostic Criteria (FSDC)	Wolfe et al., 2010 [218]	2 subscales: Widespread pain distribution in designated body locations, severity of cognitive and presence of somatic symptoms	Symptom Severity Score (SS): 6 items -3 items on Severity of symptoms -3 items on somatic symptoms	0-3	3 items on Severity of symptoms each of which is scored by a Likert format from 0 (no problem) to 3 (severe: continuous, life-disturbing problems). 3 questions with a positive or negative response to somatic symptoms with a maximum score of 3.	Score range: 0–12.	English	5	9
			Widespread Pain Index (WPI): 19 number of painful body regions	0-1	Identifying 19 body areas with pain or tenderness.	Score range: 0–19. A cumulative score range is 0–31 {(sum of the (0–19) WPI and the 6-item (0–12) SS scale)}. Fibromyalgia diagnostic criteria: (WPI ≥7 AND SS ≥5) OR (WPI 4– 6 AND SS ≥9).			

Table 5. Summary of PROM Identified *Cont.*

PROM	Developer	Subscales / Outcome domains	Number of items	Response options	Description	Range of scores/scoring	Original language	Available translations	Number of studies evaluating the instrument
Nociplastic-based Fibromyalgia Features (NFF)	Ghavidel-Parsa et al., 2022 [212]	Domain: pain extent, migratory pain, pain aggravation with emotional or physical stress, affective component of pain perception, morning fatigue and pain hypersensitivity.	7 items	0-1	binary items with 1= Yes and 0= No responses	Cut-off score is 4	Persian		1
Generalized Pain Questionnaire (GPQ)	van Bemmel et al., 2019 [213]	Domain: pain sensitivity, after sensation, spreading of pain.	7-items	0-4	Each of these items is measured on a 5-point Likert-type rating scale with the following numeric rating scale:(0= never, 1= hardly noticed, 2= moderately, 3= strongly, 4= very strongly	A cutoff score >10 is suggested for identifying possible generalized pain hypersensitivity	Dutch		1
Sensory Hypersensitivity Scale (SHS)	Dixon et al., 2016 [214]	subscales: touch, taste, smell, hearing, light, pain, allergies, heat and cold.	25-item	0-5	5-point Likert-type rating scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral / Not Sure, 4 = Agree, 5 = Strongly agree		English		1
Novel self-report Questionnaire	Austin et al., 2020 [215]	2 clusters: pain symptoms, emotional and fatigue symptoms.	18-item	0-3	4-point Likert-type rating scale 0 = never, 1= rarely, 2= sometimes, 3= always	Cumulative total score ranged from 0 - 54 (factor 1= 0-33, factor 2 = 0-21)	English		1
Leiden Visual Sensitivity Scale and Visual Discomfort Scale	Perenboom et al., 2018 [219]	Domains: Visual sensitivity to light and patterns	Leiden Visual Sensitivity Scale: 9-items scale	0-4	Each of these items is measured on a 5-point Likert scale to measure the degree of visual sensitivity, with the following numeric rating scale: not at all = 0, slightly=1, moderately =2, severely =3, and very severe =4.	Total score range 0-36	Dutch		1
	Conlon et al., 1999 [220]	Domains: somatic (e.g., being bothered, sore eyes), perceptual (e.g., afterimages, flickering, shimmering), and performance difficulties (e.g., worse eyesight, blurring) to different light sources or patterns.	Visual Discomfort Scale (VDS): 23-items	0-3	Each of these items is measured on a scale from 0 = (Event never occurs), 1= (Occasionally. A couple of times a year), 2= (Often. Every few weeks), to 3= (Almost always)	Total score ranging from 0–69	English		

PROM = Patient-reported outcome measure

4.4. Quality Assessment

5.4.1. Methodological Quality Ratings of Each Study

The risk of bias in 58 studies which evaluated the psychometric properties of PROMs was assessed.

5.4.2. Rating of Results of Each Included Studies using Good Measurement Properties Criteria

The results of each study for each measurement property underwent rating based on updated criteria for good measurement properties.

5.4.3. Overall Rating and Grading of the Quality of Evidence

The results of available studies were summarized for each measurement property of each PROM and overall ratings were given again based on criteria for good measurement properties to examine each PROM's quality. It is important to note that, to rate the qualitatively summarized results of structural validity and internal consistency of CSI as sufficient (or insufficient), the higher percentage of results which met the criteria was taken into account. Finally, using a modified GRADE approach, the quality of the evidence was graded.

Ratings on methodology quality, results of each measurement property and overall ratings are provided in Table 6. A summary of overall ratings is presented in Table 7 for a concise overview of the results. The grading of the quality of evidence for each measurement property of each PROM is provided in Table 8. Two raters experienced some disagreements, with a few being resolved through discussion among them and the remaining conflicts being referred to an expert to reach a consensus. Agreements/disagreements regarding the rating of measurement properties are listed in Table 9.

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Content validity	Bid 2016 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	**Content validity: (+)**	Ibancos-Losada 2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	**Content validity: (+)**	Carrillo-de-la-Peña 2015 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	**Content validity: (+)**
	Noorollahzadeh2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)		Latka 2019 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)		Aguirre Cárdenas 2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	
	Liang 2022 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)		Kim 2014 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)				
	Wiangkham 2022 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Sharma 2020 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Caumo 2017 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Knezevic 2018 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Mikkonen 2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Chiarotto 2018 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Madi 2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Klute 2021 (A)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							
	Holm 2021 (D)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)							

** V = very good, A = adequate, D = doubtful, I = inadequate; ** + = sufficient, - = insufficient, ± = inconsistent, ?= indeterminate*

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****
Structural validity	Mayer 2012 (D)	EFA → factor 1 =0.48–0.91; factor 2 =0.47–0.82; factor 3= 0.55–0.73; factor 4= 0.41- 0.71 (+)	**Acceptable factor loading (+)**	Ibancos-Losada 2021 (I)	EFA → The factorial analysis explained 69% of the variance. PSQ-moderate =0.627-0.783; PSQ-minor= 0.609-0.847 (+)	**2 factors with acceptable factor loading (+)**	NA	NA	NA
	Noorollahzadeh 2021 (V)	CFA → Root Mean Square Error of Approximation (RMSEA)=0.049 (+)		Latka 2019 (A)	EFA → The factorial analysis revealed 70.69% of the total variance. (+)				
	Liang 2022 (V)	CFA → factor 1=0.45–0.84; factor 2 =0.43–0.80; factor 3= 0.42–0.73; factor 4= 0.44- 0.64. Fit Indices were not reported. (+)		Kim 2014 (I)	PSQ-moderate (factor 1)= 0.510-0.782, PSQ-minor (factor 2)= 0.529-0.871 (+)				
	Roby 2022 (V)	IRT → No violation of unidimensionality, local independence (Following iterative creation of subtests), monotonicity and the chi-square value for item-trait interaction (X2 (80)=99.1, P=0.071) was non-significant which indicates fit to the Rasch model. (+)							
	Bakhtadze 2022 (A)	EFA → one-factors model: 0.263-0.715 (8 out of 25 items had factor loading < 0.40). (-)							
	Wiangkham 2022 (V)	CFA → One-factor model: CFI=0.89, TLI=0.88, RMSEA=0.09; Correlated 4-factor model: CFI= 0.90, TLI= 0.89, RMSEA= 0.08 ; Bifactor model: CFI= 0.93, TLI= 0.92, RMSEA= 0.070 (-)							
	Bakhtadze 2021 (A)	EFA→ 6 factor model: factor 1=0.48-0.75; factor 2= 0.49-0.68; factor 3= 0.55-.66; factor 4= 0.46-0.79; factor 5= 0.49-0.68; factor 6= 0.62-0.64 (+)							
	Düzce Keleş 2021 (A)	EFA→ The first 7, which was greater than an eigenvalue of the scale of 7 described 61.1% of the variance. There was a clear levelling off in the scree plot after the first factor. (+)							
	Tanaka 2017 (A)	EFA→ 5 factors model factor 1= 0.40-1.03; factor 2= 0.41-0.75; factor 3= 0.49-0.75; factor 4= 0.45-0.57; factor 5= 0.42-0.54. The factor loading of 7 items: <0.40. (-)							
	Caumo 2017 (V)	CFA → factor 1 = 0.42-0.88; factor 2 = 0.5-.78; factor 3= 0.5-0.64; factor 4= 0.47-0.64 (+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property cont.

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)
	V/A/D/I*	+/–/? **	+/–/±/?**	V/A/D/I*	+/–/? **	+/–/±/?**	V/A/D/I*	+/–/? **	+/–/±/?**
Structural validity	Kregel 2016 (V)	EFA→ factor 1 = 0.41-0.71; factor 2 = 0.41-.77; factor 3= 0.44-0.72; factor 4= 0.45-0.83 CFA→ CFI: 0.97; NFI: 0.95 NNFI:0.97; RMSEA: 0.065 (+)							
	Feng 2022 (A)	EFA→ factor 1= 0.44-0.78; factor 2=0.495-0.78; factor 3=0.77-0.80; factor 4 =0.46-0.56; factor 5= 0.48-0.59 (10 items not loaded on the factor) (-)							
	Cuesta-Vargas 2016 (A)	EFA→ The percentage of total variance explained by the one factor = 25.9% (-)							
	Knezevic 2018 (V)	CFA→ factor 1= 0.41-0.78; factor 2=0.56-0.81; factor 3= 0.51-0.64; factor 4=0.63-0.77; RMSEA = 0.07; CFI = 0.93; TLI = 0.93. (-)							
	Mikkonen 2021 (A)	EFA→ One factor model: 0.30-0.64; The percentage of total variance explained by factor 1= 28.1% (Eigenvalue 7.026) and factor 2= 34.9% (1.706) (-)							
	Kosińska 2021 (A)	EFA→ Factor 1= 0.406-0.787; factor 2= 0.517-0.801; factor 3= 0.465-0.716; factor 4= 0.349-0.758; CFA→ RMSEA=0.07 (+)							
	Kim et al., 2020 (A)	EFA→ factor 1= 0.462-0.836; factor 2= 0.403-0.768; factor 3= 0.652-0.755; factor 4= 0.411-0.863; factor 5= 0.436-0.883; factor 6= 0.512-0.872. (+)							
	Chiarotto 2018 (A)	EFA→ EFA revealed a first eigenvalue accounted for 26% of the total variance, and the ratio of the first to the second eigenvalue was 3.8. Factor loading = -0.09-0.61. (-)							
	Madi 2021 (A)	EFA→ factor 1= 0.503-0.754; factor 2= 0.504-0.747; factor 3= 0.467-0.621; factor 4=0.400- 0.724 (+)							
	Klute 2021 (V)	CFA→ 4-factor model; TLI = 0.99; RMSEA = 0.06; x2 (269) = 553.09, p < 0.001. 1-factor model; TLI = 0.98; RMSEA = 0.08; $\chi 2(275) = 756.39$, p < 0.001. Bifactor model; TLI = 0.99; RMSEA = 0.05; x2 (250) = 430.6, p < 0.001. (+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****
	Mayer 2012 (V)	Cronbach's alpha (total): 0.88 (+)		Ibancos-Losada 2021 (V)	Cronbach's α PSQ-total = 0.95; PSQ-moderate = 0.91; and PSQ-minor = 0.92 (+)	**Cronbach's α = 0.87–0.96 (+)**	Carrillo-de-la-Peña 2015 (V)	Cronbach's α: FSQ total (PSD) = 0.85 (?)	**Cronbach's α = 0.71–0.94 (?)**
	Bid 2016 (I)	Cronbach's alpha (total): 0.91 (?)		Latka 2019 (V)	Cronbach's α PSQ-total = 0.96 (+)		Häuser 2012 (V)	Cronbach's α: FS (WPI+SSS) = 0.71 (?)	
	Noorollahzadeh 2021 (V)	Cronbach's α = 0.87 (+)		Kim 2014 (V)	Cronbach's α PSQ-total = 0.93; PSQ-moderate = 0.88; and PSQ-minor = 0.87 (+)		Fors et al 2020 (V)	Cronbach's α: FS (WPI+SSS) = 0.90 (?)	
	Liang 2022 (V)	Cronbach's α(total)= 0.88 (+)					Jiao 2023 (V)	Cronbach's α: FS (WPI+SSS) = 0.82 (?)	
	Roby 2022 (V)	Cronbach's alpha (total)= 0.89 and Person Separation Index (PSI)= 0.91.after rescoring the thresholds. (+)					Kang 2019 (I)	Cronbach's α (total) = 0.94 (?)	
	Bakhtadze 2022 (V)	Cronbach's α (total) = 0.88 (?)					Aguirre Cárdenas 2021 (V)	Cronbach's α: FSQ (total) at T1= 0.91 and T2= 0.78 (?)	
	Wiangkham 2022 (V)	Cronbach's α (total) = 0.91 (?)					Bidari 2015 (V)	Cronbach's α: FSQ total (WPI+SSS=PSD) = 0.81 (?)	
Internal consistency	Bakhtadze 2021 (V)	Cronbach's α (total) =0.89 (+)	**Cronbach's α = 0.87–0.99 (+)**						
	Sharma 2020 (I)	Cronbach's α (total) = 0.87 (?)							
	Düzce Keleş 2021 (V)	Cronbach's α (total) = 0.92 (+)							
	Tanaka 2017 (V)	Cronbach's α (total) =0.89 (?)							
	Caumo 2017 (V)	Cronbach's α (total) = 0.91 (+)							
	Kregel 2016 (V)	Cronbach's α (total) = 0.91 (+)							
	Feng 2022 (V)	Cronbach's α (total) = 0.89 (?)							
	Cuesta-Vargas 2016 (V)	Cronbach's α (total) = 0.87 (?)							
	Knezevic 2018 (V)	Cronbach's α (total) = 0.91 (?)							
	Mikkonen 2021 (V)	Cronbach's α (total) = 0.88 (?)							
	Kosińska 2021 (V)	Cronbach's α (total) = 0.93 (+)							
	Bilika 2020 (V)	Cronbach's α (total) = 0.99 (?)							
	Kim 2020 (V)	Cronbach's α (total) = 0.94 (+)							
	Chiarotto 2018 (V)	Cronbach's α (total) = 0.87 (?)							
	Madi 2021 (V)	Cronbach's α (total) = 0.88 (+)							
	Klute 2021 (V)	Cronbach's α (total) = 0.93 (+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****
Cross cultural validity	NA	NA	NA	NA	NA	NA	NA	NA	NA
Reliability	Mayer 2012 (D)	ICC or weighted kappa not reported. Pearson correlation (test–retest correlation) for the total score = 0.817 (?)	**Test-retest ICC = 0.79–0.86 (+)**	Ruscheweyh 2012 (I)	Test-retest: ICC (PSQ-total) = 0.72; ICC (PSQ-minor)= 0.71. (+)	**Test-retest ICC = 0.71–0.93(+)**	Fors et al 2020 (A)	Test-retest: ICC (FS =WPI+SSS) = 0.86 (+)	
	Bid 2016 (A)	Test-retest: ICC (total) = 0.971 (+)		Latka 2019 (I)	Test-retest: ICC (PSQ-total) = 0.92; ICC (PSQ-moderate)= 0.87; ICC (PSQ-minor)= 0.91. (+)		Jiao 2023 (D)	ICC or weighted kappa not reported. Spearman's correlation analysis for the FS scale, its subscales r = 0.53 to 0.82 (?)	
	Noorollahzadeh 2021 (A)	Test-retest: ICC (total) =0.934 (+)		Kim 2014 (I)	Test-retest: ICC (PSQ-total) = 0.78; ICC (PSQ-moderate)= 0.79; ICC (PSQ-minor)= 0.75. (+)		Kang 2019 (D)	ICC or weighted kappa not reported. Spearman's correlation analysis ranged from 0.616 to 0.910 (?)	
	Liang 2022 (A)	Test-retest: ICC (total) =0.934 (+)					Aguirre Cárdenas 2021 (I)	Test-retest: ICC (FSQ total) = 0.79 (+)	
	Bakhtadze 2022 (I)	Test-retest: ICC (total) =0.91 (+)	**Test-retest ICC = 0.85–0.99 (+)**						
	Wiangkham 2022 (I)	Test-retest: ICC (total) = 0.90 (+)							
	Bakhtadze 2021 (I)	Test-retest: ICC (total) = 0.89 (+)							
	Sharma 2020 (I)	Test-retest: ICC (total) = 0.98 (+)							
	Düzce Keleş 2021 (A)	Test-retest: ICC (total) = 0.93 (+)							
	Tanaka 2017 (A)	Test-retest: ICC (total) = 0.85 (+)							
	Caumo 2017 (A)	Test-retest: ICC (total) = 0.91 (+)							
	Kregel 2016 (I)	Test-retest: ICC (total = 0.88 (+)							
	Feng 2022 (I)	Test-retest: ICC (total) = 0.932 (+)							
	Cuesta-Vargas 2016 (A)	Test-retest: ICC (total) = 0.91 (+)							
	Knezevic 2018 (V)	Test-retest: ICC (total) = 0.947 (+)							
	Mikkonen 2021 (A)	Test-retest: ICC (total) = 0.933 (+)							
	Kosińska 2021 (A)	Test-retest: ICC (total) = 0.96 (+)							
	Bilika 2020 (A)	Test-retest: ICC (total) = 0.991 (+)							
	Kim 2020 (A)	Test-retest: ICC (total) = 0.888 (+)							
	Madi 2021 (A)	Test-retest: ICC (total) = 0.94 (+)							
	Klute 2021 (A)	Test-retest: ICC (total) = 0.917(+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)
	V/A/D/I*	+/−/? **	+/−/±/?**	V/A/D/I*	+/−/? **	+/−/±/?**	V/A/D/I*	+/−/? **	+/−/±/?**
Measurement error	Bid 2016 (A)	Variability → SEM= 1.84; Smallest detectable change (SDC) / MDC = 5.09 (?)	**SEM = 0.31– 4.14; SDC(MDC) = 0.86 –11.5 (MIC not determined) (?)**	NA	NA	NA	NA	NA	NA
	Bakhtadze 2022 (I)	Minimal detectable changes (MDC)= 10 (?)							
	Wiangkham 2022 (I)	Variability → SEM=2.33; Minimal detectable changes (MDC)= 6.47 (?)							
	Bakhtadze 2021 (I)	Minimal detectable changes (MDC)= 10 (?)							
	Sharma 2020 (I)	Variability → SEM= 0.31; Smallest detectable change (SDC)= 0.86 (?)							
	Cuesta-Vargas 2016 (A)	Variability → SEM= 2.52 ; minimum detectable change (MDC90)= 7.83% (?)							
	Knezevic 2018 (V)	Variability → SEM= 3.16% ; minimum detectable change (MDC90)= 8.12% (?)							
	Mikkonen 2021 (A)	Variability → SEM= 0.43 (?)							
	Kosińska 2021 (A)	Variability → SEM= 0.99; minimum detectable change (MDC90)= 2.31% (?)							
	Bilika 2020 (A)	Variability → SEM= 2.1 (?)							
	Madi 2021 (A)	Variability → SEM= 3.45; minimum detectable change (MDC)= 9.657 (?)							
	Klute 2021 (A)	Variability → SEM= 4.14; Smallest detectable change (SDC)= ± 11.49 (?)							
Criterion validity	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Hypothesis testing (construct validity)	Mayer 2012 (V)	Discriminative validity → There was a significant difference in the total score of CSI (≥ 10 points higher) between patient subgroups compared to the Control group, p<0.001. The fibromyalgia group scored significantly the highest than other patient subgroups. (1+)	**Results in line with 106 hypotheses but not with 22 hypotheses (+)**	Ruscheweyh 2012 (A)	Convergent validity→ PSQ (total) and QST r = -0.47, P < .001; PSQ (minor) and QST r = -0.52, P < .001; PSQ(total) and QST r = 0.76, P < .001; PSQ (minor) and QST r = 0.71, P < .001; PSQ-minor and BDI r= 0.09, P> 0.05; PSQ-minor and PCS r= 0.25, P < 0.01; PSQ-minor and STAI r=0.19, P < 0.05. Discriminative validity → PSQ-total scores and PSQ-minor scores were significantly higher in chronic pain patients (but < 10 points higher) than the control group, P < .001. (3+, 5-)	**Results in line with 38 hypotheses but not with 37 hypotheses (-)**	Carrillo-de-la-Peña 2015 (A)	Convergent validity→ FSQ total (PSD) and PHQ-15 r= 0.76, p < .001; FSQ (subscale) and PHQ-15 r= 0.73- 0.67, p < .001; FSQ total (PSD) and PSQI r= 0.45, p < .001; FSQ (subscale) and PSQI r= 0.32-0.58, p < .001; FSQ total (PSD) and PHQ-9 r= 0.62, p < .001; FSQ (subscale) and FIQ (symptom) r= 0.49-0.72, p < .001; FSQ total (PSD) and BDI r= 0.51, p < .001; FSQ (subscale) and BDI r= 0.41- 0.56, p < .001; FSQ total (PSD) and FIQ r= 0.72, p < .001; FSQ (subscale) and FIQ r=0.57-0.82, p < .001; FSQ total (PSD) and SF-12 r= -0.57, p < .001; FSQ (subscale) and SF-12 r= -0.42- to -0.71, p < .001. (12+)	**Results in line with 51 hypotheses but not with 15 hypotheses (+)**
	Noorollahzadeh 2021 (V)	Discriminative validity → The patient subgroups showed the highest CSI score (≥ 10 points) than the control group as expected, p<0.001. (1+)		Sellers 2013 (A)	Convergent validity→ PSQ (total) and PCS r=0.32, P < 0.001; PSQ-minor and PCS r= 0.33 P < 0.001; PSQ- moderate and PCS r= 0.33 P < 0.001; PSQ(total) and HADS (depression, anxiety) r= 0.05-0.14, P>0.05; PSQ(subscales) and HADS (depression, anxiety) r= 0.03-0.13, P>0.05; PSQ-total and BPI (Pain, interference) r= 0.15-0.25, P>0.05; PSQ (subscales) and BPI (Pain, interference) r= 0.11-0.24, P>0.05; PSQ-total and RMQ r= 0.21, P>0.05; PSQ (subscales) and RMQ r=0.17- 0.22, P>0.05; PSQ(total) and VAS1 r= 0.23, P < 0.05; PSQ (subscale) and VAS1 r= 0.19-0.26, P < 0.05; PSQ(total) and VAS2 r= 0.33, P < 0.001;PSQ (subscale) and VAS2 r= 0.29-0.34, P < 0.001. (4+, 10-)		Häuser 2012 (A)	Convergent validity→ SSS and PHQ-4 r= 0.56, p< 0.0001; FS (WPI+SSS) and PHQ-4 r= 0.48, p< 0.001. (2+)	

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Hypothesis testing (construct validity)	Liang 2022 (A)	Convergent validity → CSI and PCS (total) r=0.709; CSI and PCS subscales (rumination, magnification and helplessness) r=0.630–0.695; CSI and BPI (total) r=0.773; CSI and BPI (mean item score) r=0.773. Discriminant validity → The patient subgroups showed the highest CSI score (≥ 10 points) than control group as expected, p<0.001. (5+)		Ibancos-Losada 2021 (V)	Convergent validity→ PSQ (total) and HADS (anxiety, depression) r= 0.45-0.49, p < 0.01; PSQ (subscales) and HADS (anxiety, depression) r=0.37-0.52, p < 0.01; PSQ (total) and PCS r= 0.58, p < 0.01; PSQ (subscales) and PCS r=0.50-0.60, p < 0.01; PSQ (total) and FIQ r=0.26, p < 0.05; PSQ-minor and FIQ r= 0.31,p < 0.01; PSQ-moderate and FIQ r= 0.19; PSQ (total) and CSI r=0.33, p < 0.01; PSQ-minor and CSI r= 0.36, p < 0.01; PSQ-moderate and CSI r=0.29, p < 0.05; PSQ (total) and CPT (pain intensity) r= 0.65, p = 0.01; PSQ (subscales) and CPT (pain intensity) r= 0.57- 0.60, P= 0.01; PSQ (total) and CPT (Tolerance) r= -0.56, p = 0.01; PSQ (subscales) and CPT (Tolerance) r= - 0.52 to -0.57, p = 0.01; PSQ (total) and PPT r= -0.59 p = 0.01; PSQ(subscales) and PPT r= -0.50 to -0.60, p = 0.01. (13+, 3-)		Fors et al 2020 (V)	Convergent validity→ FS (WPI+SSS) and FIQ rh= 0.74; WPI and FIQ rh= 0.59; SSS and FIQ rh= 0.85; FS (WPI+SSS) and TPC rh= 0.63; WPI and TPC rh= 0.55 SSS and TPC rh= 0.61 (6+)	
	Bakhtadze 2022 (I)	Convergent validity → CSI and NDI-RU rS=0.57, p<0.05; CSI and HADS (anxiety, depression) rS=0.57- 0.56, p <0.05; CSI and McGill Pain Questionnaire (SF-MPQ-2) rS=0.46, p<0.05).(3+)		Inal 2021 (A)	Convergent validity→ PSQ (total) and BPI (pain, interference) r= 0.28-0.31, p < 0.05; PSQ (subscale)and BPI (pain) r= 0.24-0.31, p < 0.05; PSQ (subscale)and BPI (interference) r= 0.22-0.34, PSQ (total) and BDI r= 0.14, P>0.05; PSQ (subscale) and BDI r= 0.10-0.19, P>0.05; PSQ (total) and BAI r= -0.01, P>0.05; PSQ (subscale) and BAI r= -0.44 to 0.09, P>0.05; PSQ (total) and PCS r= 0.02, P>0.05; PSQ (subscale) and PCS r= -0.02 to 0.12, P>0.05; PSQ (total) and VAS1 r= 0.70, P<0.001; PSQ (subscale) and VAS1 r= 0.63-0.68, P<0.001; PSQ (total) and VAS2 r= 0.82, P<0.001; PSQ (subscale) and VAS2 r= 0.76-0.79, P<0.001; (7+, 9-)		Jiao 2023 (V)	Convergent validity→ FS (WPI+SSS) and FIQR rh= 0.487; WPI and FIQR rh= 0.292; SSS and FIQR rh= 0.589. (2+,1-)	

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****
Hypothesis testing (construct validity)	Wiangkham 2022 (V)	Convergent validity → CSI and VAS (pain intensity) rs= 0.36, P<0.001; CSI and NDI rs= 0.59 <0.001; CSI and FABQ (total score) rs= 0.42, <0.001; CSI and FABQ (Physical Activity and work) rs= 0.37-0.44, <0.001; CSI and PCS (total score) rs= 0.10, P= 0.07; CSI and PCS (subscale) rs= 0.10-0.11, P= 0.06; CSI and SF-36 (total score) rs= –0.58, P<0.001; SF-36 (mental and physical) rs= –0.48 to- 0.54, P<0.001; (6+, 2-)		Latka 2019 (I)	Convergent validity→ PSQ and CSQ- total rho= 0.27, P<0.01. (1-)		Kang 2019 (A)	Convergent validity→ WPI and FIQR rh= 0.815; SSS and FIQR rh= 0.854;m WPI and EQ-5D rh= 0.717; SSS and EQ-5D rh= 0.724; WPI and MD-HAQ rh= 0.712; SSS and MD-HAQ rh=0.743. (6+)	
	Hendriks 2020 (A)	Convergent validity → CSI and SCL-90-R r_s =0.648, P <0.001, CSI and 4DSQ (somatization) r_p= 0.681, P<0.001, CSI and 4DSQ (anxiety, depression, distress) r_s= 0.54-0.66, P <0.001; CSI and IES r_s= 0.450, P <0.001, CSI and TSK r_p= 0.219, P=0.020; CSI and CIS20R r_s = 0.436, P<0.001; CSI and VAS r_s = 0.187, P=0.048; CSI and PPT (quadriceps, trapezius) rs= - 0.20 to –0.11; CSI and TS (quadriceps, trapezius) rp= –0.008 to 0.028, P >0.05. (5+, 4-)		Grundström 2019 (V)	Convergent validity→ PSQ (total) and HPTm r= - 0.63, P<0.001; PSQ (total) and CPTm r= 0.56, P<0.001; PSQ (total) and PPTm r= - 0.43, P<0.001; PSQ (total) and HADS r = 0.19-0.27, P<0.001; PSQ (subscale) and HPTm r= -0.53 to -0.65, P<0.001; PSQ (moderate) and CPTm r= 0.44, p=0.006; PSQ (minor) and CPTm r= 0.60, P<0.001; PSQ (subscale) and PPTm r= -0.34 to – 0.46, P<0.001; PSQ-moderate and HADS r = -0.03 to 0.23, P>0.05; PSQ-minor and HADS r = 0.31-0.38. (5+, 5-)		Moore 2022 (V)	Convergent validity→ FSQ and PPT rh = -0.13 to -0.20; FSQ and TS rh =0.12-0.13; No significant correlation was found between FSQ and CPM. The degree of correlation between QST and individual components of FSQ did not exhibit any substantial differences. (3-)	
	Bakhtadze 2021 (I)	Convergent validity → CSI and NDI-RU rs=0.56, p<0.05; CSI and ODI-RU (rS = 0.36, p <0.05). (2+)		Kim 2014 (A)	Convergent validity→ PSQ (total) and PCS r = 0.38, P = 0.002; PSQ (subscale) and PCS r = 0.36-0.37, P = 0.003. (2+)		Aguirre Cárdenas 2021 (V)	Convergent validity→FSQ and PHQ-15(r = 0.62, p < 0.0001; FSQ and FIQ-R r = 0.60, p < 0.0001; FSQ and NPRS r = 0.51, p < 0.0001; FSQ and PHQ-9 r = 0.49, p < 0.0001; FSQ and BPI-PI r = 0.47, p < 0.0001; FSQ and PCS r = 0.31, p < 0.0001; FSQ and SF-12 r = −0.46, p < 0.0001; FSQ and PVAQ r = 0.22, p = 0.002. Discriminative validity → The FM group scored significantly higher on the FSQ total score than the RA group and G group, p < 0.0001. (8+, 1-)	

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Hypothesis testing (construct validity)	Sharma 2020 (A)	Convergent validity → CSI and PCS r = 0.50, P < 0.001; CSI and NRS pain intensity r = 0.25, P = 0.013; Discriminative validity → Mean CSI scores were significantly higher in women compared to men, P = 0.005. (2+, 1-)		Coronado and George, 2018 (A)	PSQ (total scores) and QST (pressure pain threshold, heat pain thresholds, suprathreshold heat pain rating) rho =−0.27 to 0.17; PSQ (minor) and QST (pressure pain threshold, heat pain thresholds, suprathreshold heat pain rating) rho = -0.18 to 0.12, p > 0.05; PSQ (total) and BRS rho = −0.39, p < 0.05; PSQ (minor) and BRS rho = −0.37, p < 0.05; PSQ (total) and DASS (anxiety) rho = 0.31, p < 0.05; PSQ(minor) and DASS (anxiety) rho = 0.25, p < 0.05; PSQ (total) and PANAS (negative) rho = 0.31, p < 0.05; PSQ(minor) and PANAS (negative) rho = 0.27, p <0.05. (4+, 4-)		Neville 2018 (V)	Convergent validity→ Female: FSQ and QST(PPT) = all |r| ≥ 0.27, all p ≤ 0.021 except thumbnail PPT; FSQ and TM r = 0.22, p = 0.078; FSQ and CPM r = 0.01, p= 0.945. No QST outcomes correlated with FM score in males. (6-)
	Düzce Keleş 2021 (V)	Discriminative validity → The fibromyalgia subgroup scored higher on the CSI score (≥ 10 points)than the CSP subgroup and the healthy control group scored lowest, p <0.001. (1+)					Bidari 2015 (A)	Convergent validity→ FSQ (total) and FIQ (total) r= 0.45, P < 0.05; FSQ (subscale) and FIQ (total) r= 0.33- 0.49, P < 0.05; FSQ (total) and FIQ (function) r= 0.06, P > 0.05; FSQ (subscale) and FIQ (function) r= 0.05-0.10, P > 0.05; FSQ (total) and FIQ (symptom) r= 0.24, P < 0.05; FSQ (subscale) and FIQ (symptom) r= 0.21-0.24, P < 0.05; FSQ (total) and SF-12 (PCS, MCS) r= -0.38 to -0.45, P < 0.05; FSQ (subscale) and SF-12 (PCS, MCS) r= -0.30 to -0.48, P < 0.05. Discriminative validity → The mean score of FSQ (PSD) and its components in the FM group were significantly higher than in non- FM chronic pain group, p=0.01. (5+, 4-)

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Hypothesis testing (construct validity)	Salaffi 2022 (V)	Convergent validity → CSI and modFAS ρ = 0.58 p<0.0001; CSI and FIQR ρ = 0.542, p<0.0001; CSI and PDS ρ = 0.518, p<0.0001 Discriminative validity → The CSI score among the FM very severe subgroup was higher than other severity subgroups and the remission subgroup scored lowest, p<0.0001. (4+)							
	Tanaka 2017 (A)	Convergent validity → CSI and EQ-5D r = −0.44, p<0.01; CSI and BPI (pain intensity) r = 0.42, p < 0.01; CSI and BPI (pain interference) r = 0.48, p < 0.01. Discriminative validity → one CSS group and 2 or more CSS groups scored higher on the CSI (≥10 points) than those with no CSS diagnosis group, p < 0.01. (4+)							
	Neblett 2013 (A)	Discriminative validity → Patients with diagnoses of CSS scored higher on the CSI (≥10 points) than the non-CSS patient sample and nonpatient control group, P =0 .05; with the AUC for the CSI=0 .86. (1+)							
	Caumo 2017 (A)	Convergent validity → CSI and PCS (subscale) r=0.62-0.68; CSI and PCS (total score) r=0.68. Discriminative validity → Patients with diagnoses of FM were higher on the CSI (≥10 points) than other patient subgroups and the nonpatient control group, p<0.001. (3+)							
	Kregel 2016 (V)	Discriminative validity → The patient group scored higher on CSI than the control group. (1+)							
	Coronado and George 2018 (A)	Convergent validity → CSI and QST (pressure pain threshold, heat pain thresholds, suprathreshold heat pain rating) rho = −0.13 to 0.13, (p > 0.05); CSI and BRS rho = −0.29, p < 0.05; CSI and DASS-21 (depression, anxiety, and stress) rho = 0.64 - 0.67, p < 0.05) CSI and PANAS (negative) rho = 0.67, p < 0.05. (2+, 2-)							
	Feng 2022 (I)	Convergent validity → CSI and EQ-5D index r= −0.375, P <0.001; CSI and HADS (anxiety and depression) r= 0.467-0.525, P <0.001. Discriminative validity → The CSI scores in chronic pain patients with 2 or above CSSs were significantly higher (≥10 points), compared with those without CSS, p < 0.001. (3+)							
	Knezevic 2018 (V)	Discriminative validity → The patient subgroups scored significantly higher on the CSI total Scores (≥10 points) than the pain-free group, p < 0.001. (1+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)
	V/A/D/I*	+/–/? **	+/–/±/?**	V/A/D/I*	+/–/? **	+/–/±/?**	V/A/D/I*	+/–/? **	+/–/±/?**
	Mikkonen 2021 (A)	Convergent validity → CSI and TSK (kinesiophobia) r = 0.463; CSI and RMDQ (disability) r = 0.387; CSI and the DEPS (depression) r = 0.615; CSI and PSQ-3 (impact of pain on sleep) r = 0.505; CSI and EQ-5D-5L (quality of life) r = −0.554. Discriminative validity →The CSI scores in patient subgroups were significantly higher (≥10 points) compared with those pain-free controls, < 0.001 (6+)							
	Neblett 2017 (V)	Discriminative validity→ The mean CSI scores moved up into higher severity ranges from the non-CSS patients to the patients with multiple CSSs. (1+)							
	Kosińska 2021 (V)	Convergent validity → CSI and NDI r=0.593; CSI and ODI r=0.422. Discriminative validity → CSI values were statistically higher (≥10 points) in the patient subgroup with both pain locations (CNP and CLBP) compared with only one location (CNP or CLBP), p<0.03. (3+)							
	Bilika 2020 (V)	Convergent validity → CSI and PCS r = 0.68. Discriminative validity → The FM group scored the highest on the total CSI scores (≥10 points) than other subgroups whereas the control group scored the lowest, p=0.000 (2+)							
Hypothesis testing (construct validity)	Kregel 2018 (V)	Convergent validity → CSI and pain intensity r = 0.320, P < 0.001; CSI and SF-36 (emotional, physical) r = -0.635 to -0.617, P < 0.001; CSI and PCS r = 0.464, P < 0.001; CSI and PDI r = 0.472, P < 0.001; CSI and PPT r = -0.276 to -0.237, P ≤ .01; CSI and CPM r = 0.017, P = 0.858. (4+, 2-)							
	Kim 2020 (V)	Convergent validity → CSI and WOMAC(pain scores) r = 0.524, p < 0.001; CSI and VAS r = 0.496, p < 0.001; CSI and WOMAC (function) r = 0.408, p < 0.001; CSI and EQ-5D r = 0.437, p < 0.001. (4+)							
	van der Noord 2018 (V)	Convergent validity → CSI and SCL-90 (anxiety, depressive) rs = 0.65- 0.67; p < .01; CSI and WPI rs = 0.43; p < .01; CSI and NRS rs = 0.36; p < .01; CSI and PCS rs = 0.39; p < .01. (5+)							
	Chiarotto 2018 (A)	Convergent validity → CSI-I score and NRS r= 0.427, P < 0.01; CSI-I score and SF36-PF score r= -0.479, P < 0.01; CSI-I score and HADS (Anxiety) r= 0.706, P < 0.01; CSI-I score and HADS (depression) r= 0.551, P < 0.01; CSI-I score and PSEQ r= −0.618 P < 0.01; CSI-I score and ODI r= 0.356, P < 0.01; CSI-I score and RMDQ r= 0.450, P < 0.01. Discriminative validity → FM patients was ≥ 10 points higher than in the other subgroups, p < 0.001. (6+)							
	Valera-Calero 2022 (A)	Convergent validity → CSI and NRS r= 0.305, p<0.001; CSI and PPT (mastoid) r= −0.372, p<0.001; CSI and HADS (Depression, Anxiety) r= 0.415-0.541, p<0.001. (3+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property cont.

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Hypothesis testing (construct validity)	Madi 2021 (A)	Convergent validity → CSI and VRS rs= 0.38, p < 0.0001; CSI and PCS rs= 0.43, p < 0.0001; CSI and PCS (subscales) rs= 0.30- 0.44, p < 0.0001; CSI and EQ-VAS rs= -0.61, p < 0.0001; CSI and EQ-5D-3L rs= -0.48, p < 0.0001. Discriminative validity → Patients with more than one chronic pain complaint had significantly higher scores than patients with only one complaint, p= 0.002 and patients with a confirmed diagnosis of CS or CSS than those with no confirmed diagnosis, p = 0.04 (7+)								
	Knezevic 2020 (V)	Convergent validity → CSI and pain intensity τb = 0.271, P< 0.001; CSI and FACS (total) τb= 0.381, P< 0.001; CSI and FACS (Factor 1) τb = 0.410, P< 0.001; CSI and FACS (Factor 2) τb = 0.241 P< 0.001; CSI and PCS (total) τb= 0.369, P< 0.001; CSI and PCS (magnification, helplessness) τb= 0.343-0.400, P< 0.001; CSI and PCS (Rumination) τb= 0.290 P< 0.001; CSI and ODI τb= 0.381, P< 0.001; CSI and MOS (cognitive functioning scale, sleep scale) τb= -0.409 to -0.504, < 0.001; CSI and SF-36v2-(PCS) τb= -0.292, P< 0.001; CSI and SF-36v2-(MCS) τb= -0.402, P< 0.001; CSI and MSPSS (total) τb= -0.186, P= 0.043; CSI and MSPSS (subscales) τb= - 0.177 to -0.138. Discriminative validity → The CSI scores of the fibromyalgia group were significantly higher (≥10 points) than those of all other subgroups, while the control group exhibited lower scores than all patient groups, p < 0.001. (8+, 6-)								
	Van Wilgen 2018 (V)	Convergent validity → CSI and SCL-90 rs= 0.75, P < 0.001; CSI and WPI rs= 0.58, P < 0.001); CSI and VAS rs = 0.29, P < 0.01; CSI and PCS r = 0.27, P < 0.01. (2+, 2-)							
	Klute 2021 (A)	Convergent validity → CSI and PSQminor τ = 0.23; p < 0.001; CSI and PHQ15 τ = 0.57; p < 0.001; CSI and FSQ (SSS) τ = 0.56; p < 0.001; CSI and FSQ (WPI)τ = 0.47; p < 0.001; CSI and PainDETECT τ = 0.43; p < 0.001; CSI and MPSS τ = 0.32; p < 0.001; CSI and PCS τ = 0.28; p < 0.001 Discriminative validity → The CSI scores of the FMS group were significantly higher (≥10 points) than those of all other subgroups, while the control group exhibited lower scores than all patient groups, p < 0.001. (6+, 2-)								
	Holm 2021 (A)	Convergent validity → CSI and ODI rs = 0.52, p= <0.001; CSI and KEDS rs = 0.74, p= <0.001; CSI and WAI1 rs = -0.42, p= <0.001; CSI and QST (total) r = 0.22, p=0.008. (3+, 1-)							
	Neblett 2015 (A)	Discriminative validity → The CSS patient group had significantly higher total CSI scores (≥10 points) than the non-CSS patient group, < 0.001. (1+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	CSI [159-195]			PSQ [175,196-202]			FSQ [203-211]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****	**V/A/D/I***	**+/−/? ****	**+/−/±/?****
Responsiveness	Neblett 2017 (V)	There was a significant change in CSI scores (change score 10.2) between admission (without FRP-receiving patients) and discharge (with FRP-receiving patients) ($p < 0.001$). (+)	**Results in line with two hypothesis (+)**	NA	NA	NA	NA	NA	NA
	Bid 2017 (V)	Significant changes were observed in CSI scores across three assessments for two groups of patients with chronic non-specific low back pain who underwent either a conventional physiotherapy program or a McKenzie exercise program ($p < 0.001$). The McKenzie program yielded significantly better results. (+)							

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	NFF [212]			GPQ [213]			SHS [214]		
	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)	Studies (Meth Qual Rating)	Results (Rating)	Summary of Results (Overall Rating)
	V/A/D/I*	+/−/? **	+/−/±/?**	V/A/D/I*	+/−/? **	+/−/±/?**	V/A/D/I*	+/−/? **	+/−/±/?**
Content validity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Structural validity	N/A	N/A	N/A	van Bemmel 2019 (I)	Mokken analysis: no secondary dimensions emerged, item scalability > 0.30, (non-significant violations of monotonicity) and Adequate model fit. (+)	(+)	N/A	N/A	N/A
Internal consistency	Ghavidel-Parsa 2022 (V)	Cronbach's α: s 0.72 but lack of redundancy between the items. (?)	(?)	van Bemmel 2019 (V)	Reliability coefficient r=0.90 (+)	(+)	Dixon 2016 (V)	Cronbach's alpha: SHS (total)= 0.86, SHS (factors)= 0.62-0.88 (?)	(?)
Cross-cultural validity	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Reliability	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Measurement error	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Criterion validity	Ghavidel-Parsa 2022 (V)	AUC= 0.87, sensitivity 82.5% and specificity 91.5% (+)	(+)	N/A	N/A	N/A	N/A	N/A	N/A
Construct validity	Ghavidel-Parsa 2022 (V)	Discriminative validity → The ROC AUC = 0.87 indicated a good ability of NFF to discriminate between FM and non-FM with a score of 4 as the best cut-off.(1+)	(+)	van Bemmel 2019 (A)	Convergent validity→ GPQ (total) and FSQ r= 0.72, <0.001; GPQ (total) and PDQ r= 0.87, <0.001; GPQ (total) and Pain intensity r= 0.81, <0.001; GPQ (total) and SF-36 (PCS) r= -0.62, <0.001; GPQ (total) and SF-36 (MCS) r= -0.75, <0.001; GPQ (total) and HAQ-DI r= 0.72, <0.001. Discriminative validity → Patients with FM scored significantly higher on the GPQ compared to patients with RA, P<0.001). The GPQ had excellent accuracy in predicting FM, with an AUC of 0.89. (6+)	(+)	Dixon 2016 (A)	Convergent validity→ SHS (total) and heat pain threshold r = -0.40, p = 0.019; SHS (total) and cold pressor duration r = -0.50, p = 0.002; SHS (total) and heat pain tolerance r= -0.009, p >0.05. Discriminative validity → study 4: SHS total scores of the fibromyalgia with osteoarthritis group were significantly higher (but <10 points) than those of the healthy control group, p< 0.001 Study 5: SHS total scores of the CLB patients did not significantly differ from the control group. (2+,3-)	(-)
Responsiveness	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 6. COSMIN ratings on methodology quality, results, and overall rating per measurement property *cont.*

COSMIN measurement properties	Novel self-report instrument [215]			L-VISS and VDS [216]		
	Studies (Meth Qual Rating)	**Results (Rating)**	**Summary of Results (Overall Rating)**	**Studies (Meth Qual Rating)**	**Results (Rating)**	**Summary of Results (Overall Rating)**
	V/A/D/I*	**+/–/? ****	**+/–/±/?****	**V/A/D/I***	**+/–/? ****	**+/–/±/?****
Content validity	N/A	N/A	N/A	N/A	N/A	N/A
Structural validity	Austin 2020 (D)	EFA → factor 1 = 0.59- 0.91, factor 2= 0.60-0.91 (+)	**(+)**	NA	NA	NA
Internal consistency	Austin 2020 (V)	Cronbach's α: factor 1 = 0.94, factor 2 = 0.90 (+)	**(+)**	Ten Brink 2021 (V)	Cronbach's α: L-VISS = 0.85, VDS = 0.94 (?)	**(?)**
Cross-cultural validity	N/A	N/A	N/A	N/A	N/A	N/A
Reliability	N/A	N/A	N/A	N/A	N/A	N/A
Measurement error	N/A	N/A	N/A	N/A	N/A	N/A
Criterion validity	N/A	N/A	N/A	N/A	N/A	N/A
Hypothesis testing (Construct validity)	Austin 2020 (D)	Convergent validity→ Novel self-reported questionnaire (factor1, factor2) and PVAQ r= 0.54 – 0.58, p<0.0001; Novel self-reported questionnaire (factor1, factor2) and PASS-20 r= 0.65– 0.72, p<0.0001; Novel self-reported questionnaire (factor1, factor2) and PSEQ = -0.64 to – 0.71, p<0.0001; Novel self-reported questionnaire (factor1, factor2) and PCS r= 0.63 – 0.66, p<0.0001; Novel self-reported questionnaire (factor1, factor2) and DASS-21 Total r= 0.59 – 0.66, p<0.0001; Novel self-reported questionnaire (factor1, factor2) and CPM r= -0.41– 0.46, p=0.0001. (6+)	**(+)**	Ten Brink 2021 (D)	Convergent validity→ L-VISS and Visual Distortion Scores of the Pattern Glare Test (pattern 2) r= 0.33-0.42, p <.05; VDS and Visual Distortion Scores of the Pattern Glare Test (pattern 2) r=0.35-0.39, p <.05 for all patients except FM patient. Discriminative validity → Patients with fibromyalgia; Complex Regional Pain Syndrome (CRPS), and other types of pain exhibited elevated L-VISS and VDS scores compared to those without pain, p < .001. (3+ 1-)	**(+)**
Responsiveness	N/A	N/A	N/A	N/A	N/A	N/A

* V = very good, A = adequate, D = doubtful, I = inadequate; ** + = sufficient, - = insufficient, ± = inconsistent, ?= indeterminate; meth qual = methodological quality; CSI=Central sensitization Inventory; PSQ= Pain Sensitivity Questionnaire; FSQ= Fibromyalgia Survey Questionnaire; NFF= Nociplastic-based Fibromyalgia Feature; GPQ= Generalized Pain Questionnaire; SHS= Sensory Hypersensitivity Scale; L-VISS and VDS= Leiden Visual Sensitivity Scale and Visual Discomfort Scale; PROM= Patient-reported outcome measure; CFA = confirmatory factor analysis, CFI = comparative fit index, CTT = classical test theory, ICC = intraclass correlation coefficient, IRT = item response theory, MIC = minimal important change, RMSEA: root mean square error of approximation, SEM = standard error of measurement, SDC = smallest detectable change, SRMR: standardized root mean residuals, TLI = Tucker–Lewis index, QST=Quantitative sensory testing; BDI= Beck Depression Inventory; PCS= Pain Catastrophizing Scale; STAI= State-Trait Anxiety Inventory; PSD= polysymptomatic distress; PHQ-15 = Patient Health Questionnaire-15; FIQ= Fibromyalgia Impact Questionnaire; SF-12= Short-Form-12; HADS= Hospital Anxiety and Depression Scale; BPI= Brief Pain Inventory; RMQ= Roland-Morris Back Pain Questionnaire; VAS= Visual analogue scale; PHQ-4= Patient Health Questionnaire-4; CPT= Cold Pressor Test; PPT=Pressure pain threshold; TPC= tender points counts; NDI= Neck Disability Index; SF-MPQ-2= Short form McGill Pain Questionnaire-2; BAI= Beck Anxiety Inventory; FIQR= revised fibromyalgia impact questionnaire; FABQ= Fear- Avoidance Beliefs Questionnaire; SF-36= Short-Form-36; CSQ= Coping Strategy Questionnaire; EQ-5D= EuroQol five-dimensional questionnaire; MD-HAQ= Multidimensional Health Assessment Questionnaire; SCL-90-R= Symptom Checklist 90-R; 4DSQ= Four-Dimensional Symptom Questionnaire ; IES= Impact of Event Scale; TSK= Tampa Scale for Kinesiophobia; CIS20R= Checklist Individual Strength; TS= Temporal summation; HPT= Heat Pressor Test; ODI= Oswestry disability index; NPRS= Numeric Pain Rating Scale; PHQ-9= Patient Health Questionnaire-9; BPI-PI= Brief Pain Inventory–Interference Subscale; PVAQ= Pain Vigilance and Awareness Questionnaire; NRS=Numeric Rating Scale; BRS= Brief Resilience Scale; DASS= Depression, anxiety, and Stress Scale; PANAS= Positive and Negative Affect Schedule; CPM= conditioned pain modulation; modFAS= Modified Fibromyalgia Assessment Status; PDS= Polysymptomatic Distress Scale; RMDQ= Roland Morris Disability Questionnaire; PSQ-3= Pain and Sleep Questionnaire Three-Item Index; EQ-5D-5L= EuroQol five-dimensional five-level questionnaire; PDI= Pain Disability Index; WOMAC= Western Ontario and McMaster Universities Osteoarthritis Index; PSEQ = Pain Self-Efficacy Questionnaire; EQ-5D-3L= EuroQol five-dimensional three-level questionnaire; FACS= Fear-Avoidance Components Scale; MOS= Medical Outcomes Study; MSPSS= Multidimensional Scale of Perceived Social Support; MPSS= Mainz pain staging system; KEDS= Karolinska exhaustion disorder scale; WAI1= Work ability index item 1. AUC= Area under the curve.

Table 7. Summary table of COSMIN overall ratings of each measurement property of each PROM

PROMs	Content validity		Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Hypothesis testing for construct validity	Responsiveness
	+/−/±/?**		+/−/±/?**	+/−/±/?**	+/−/±/?**	+/−/±/?**	+/−/±/?**	+/−/±/?**	+/−/±/?**	+/−/±/?**
CSI (159-195)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	Overall (+)	Acceptable factor loading (+)	Cronbach's α = 0.87–0.99 (+)	N/A	Test-retest ICC = 0.85–0.99 (+)	SEM = 0.31–4.14; SDC (MDC) = 0.86 –11.5 (MIC not determined) (?)	N/A	>75% of the results aligned with the hypotheses (+)	Results in line with two hypothesis (+)
PSQ (175, 196-202)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	Overall (+)	Acceptable factor loading (+)	Cronbach's α = 0.87–0.96 (+)	N/A	Test-retest ICC = 0.71–0.93 (+)	N/A	N/A	>25% of the results were not aligned with the hypotheses (-)	N/A
FSQ (203-211)	Relevance: (+) Comprehensiveness: (+) Comprehensibility: (+)	Overall (+)	N/A	Cronbach's α = 0.71–0.94 (?)	N/A	Test-retest ICC = 0.79–0.86 (+)	N/A	N/A	>75% of the results aligned with the hypotheses (+)	N/A
NFF (212)	N/A		N/A	Cronbach's α= 0.72 (?)	N/A		N/A	AUC= 0.87, sensitivity 82.5% and specificity 91.5% (+)	100% of the result aligned with the hypothesis (+)	N/A
GPQ (213)	N/A		Mokken analysis: Adequate model fit. (+)	Reliability coefficient r=0.90 (+)	N/A	N/A	N/A	N/A	100% of the result aligned with the hypothesis (+)	N/A
SHS (214)	N/A		N/A	Cronbach's alpha: SHS (total)= 0.86, SHS(factors)= 0.62-0.88 (?)	N/A	N/A	N/A		>25% of the results were not aligned with the hypotheses (-)	N/A
Novel self-report instrument (215)	N/A		Acceptable factor loading (+)	Cronbach's α: factor 1 = 0.94, factor 2 = 0.90 (+)	N/A	N/A	N/A	N/A	100% of the result aligned with the hypothesis (+)	N/A
L-VISS and VDS (216)	N/A		N/A	Cronbach's α: L-VISS = 0.85, VDS = 0.94 (?)	N/A	N/A	N/A	N/A	100% of the result aligned with the hypothesis (+)	N/A

** + = sufficient, - = insufficient, ± = inconsistent, ?= indeterminate; PROM= Patient-reported outcome measure, CSI=Central sensitization Inventory, PSQ= Pain Sensitivity Questionnaire, FSQ/FSDC= Fibromyalgia Survey Questionnaire/ Fibromyalgia Survey diagnostic criteria, NFF= Nociplastic-based Fibromyalgia Feature, GPQ= Generalized pain questionnaire, SHS= Sensory Hypersensitivity Scale, L-VISS and VDS= Leiden Visual Sensitivity Scale and Visual Discomfort Scale, N/A= not applicable.

Table 8. Quality of evidence for measurement properties of PROMs

PROM	Measurement properties										
	Content validity			Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Hypothesis testing	Responsiveness
	Relevance	Comprehensiveness	Comprehensibility								
CSI (159-195)	**Moderate**	**Moderate**	**Moderate**	**Moderate**	**High**	N/A	**High**	N/A	N/A	**High**	**High**
PSQ (175, 196-202)	**Moderate**	**Moderate**	**Moderate**	**Low**	**High**	N/A	**Low**	N/A	N/A	**High**	N/A
FSQ (203-211)	**Moderate**	**Moderate**	**Moderate**	N/A	**High**	N/A	**Moderate**	N/A	N/A	**High**	N/A
NFF (212)	N/A	N/A	N/A	N/A	**High**	N/A	N/A	N/A	**High**	**High**	N/A
GPQ (213)	N/A	N/A	N/A	**Very low**	**High**	N/A	N/A	N/A	N/A	**Moderate**	N/A
SHS (214)	N/A	N/A	N/A	N/A	**High**	N/A	N/A	N/A	N/A	**Moderate**	N/A
Novel self-report instrument (215)	N/A	N/A	N/A	**Low**	**High**	N/A	N/A	N/A	N/A	**Low**	N/A
L-VISS and VDS (216)	N/A	N/A	N/A	N/A	**High**	N/A	N/A	N/A	N/A	**Low**	N/A

PROM= Patient-reported outcome measure, CSI=Central sensitization Inventory, PSQ= Pain Sensitivity Questionnaire, FSQ/FSDC= Fibromyalgia Survey Questionnaire/ Fibromyalgia Survey diagnostic criteria, NFF= Nociplastic-based Fibromyalgia Feature, GPQ= Generalized pain questionnaire, SHS= Sensory Hypersensitivity Scale, L-VISS and VDS= Leiden Visual Sensitivity Scale and Visual Discomfort Scale, NA= not applicable.

Table 9. Agreements/disagreements of quality rating of measurement properties of PROMs

PROM	Measurement Properties	Study	Issues	Rater 1 assessment	Rater 2 assessment	Consensus	Expert opinion
CSI	Structural validity	Mayer et al., 2012	Factor analysis: Acceptable factor loading with 3 missing items	Rating for structural validity of CSI: Insufficient (-)	Sufficient(+)	Not achieved	Sufficient(+) because more than 85% of total items loaded on factor
		Liang et al., 2022	Factor analysis: Good factor loading with 4 missing items	Insufficient (-)	Sufficient(+)	Not achieved	Sufficient(+)
		Tanaka et al., 2017	Factor analysis: Factor loading of all items was acceptable except for seven items, which had loadings less than 0.40.	Insufficient (-)	Sufficient(+)	Not achieved	Insufficient (-) due to its misalignment with the predetermined criterion for acceptable factor loading, which was set as ≥ 0.40.
		Feng et al., 2022	Factor analysis: 10 items were not loaded on the factor	Insufficient (-)	Sufficient(+)	Not achieved	Insufficient (-) because 40% (10 items out of 25 items) of total items were not loaded on factor
		All studies which conducted factor analysis for structural validity	**Overall rating** of summarized results for structural validity of CSI	Sufficient(+)	Sufficient(+) / insufficient (-)	Not achieved	**Overall rating:** sufficient (+) because a higher percentage (60%) of results met the criteria of sufficient (+) rating
PSQ		Latka et al., 2019	Factor analysis: two-factor model with acceptable factor loading but three items were loaded on both factors with the load value was relatively low.	Insufficient (-)	Sufficient(+)	Not achieved	Sufficient (+) because the factor loading was acceptable indicating the majority of items adequately capture the underlying structure of the factors despite very few overlaps in item loading with low load values
All instruments	Cross-cultural validity	All studies that validated the translated version of the original questionnaire	Measurement invariance: Multiple group factor analysis OR differential item functioning (DIF)	Not reported	Not reported	Consensus achieved	
CSI	Measurement error	Bid et al., 2016 Bakhtadze et al., 2022 Wiangkham et al., 2022 Bakhtadze et al., 2021 Sharma et al., 2020 Cuesta-Vargas et al., 2016 Knezevic et al., 2018 Mikkonen et al., 2021 Kosińska et al., 2021 Bilika et al., 2020 Madi et al., 2021 Klute et al., 2021	SDC / MDC was calculated but MIC value wasn't determined	Indeterminate (?)	Indeterminate (?)	Consensus was achieved because without specifying the MIC value, it wasn't possible to evaluate whether MIC was greater than SDC	

PROM= Patient-reported outcome measure, CSI=Central sensitization Inventory, PSQ= Pain Sensitivity Questionnaire, NFF= Nociplastic-based Fibromyalgia Feature, MIC= minimal important change, SDC= Smallest detectable change, MDC= Minimum detectable change

The Central Sensitization Inventory (CSI) is a tool specifically designed to assess symptoms associated with central sensitization in patients with central sensitivity syndrome (CSS) [159]. As previously mentioned, central sensitivity syndrome, also known later as Chronic Overlapping Pain Conditions (COPCs), represents a group of chronic pain disorders without a clear medical explanation, where central sensitization may be a prominent contributing factor [172]. Part A of the CSI evaluates 25 physical and psychological symptoms that are typically indicative of central sensitization. Part B gathers data on 10 pre-existing specific conditions including seven COPC and three CS-related disorders. Part B is for information only and is not scored [159]. A score of 40 or higher has been regarded as a reasonable cutoff to notify healthcare providers that a patient's symptoms may suggest the presence of CS [172]. CSI was rated 'sufficient' for content validity, structural validity, internal consistency, reliability, construct validity and responsiveness. However, CSI was rated 'indeterminate' for measurement error because some studies assessed measurement error, but they failed to calculate the minimal important change (MIC) value, which complicates the determination of whether it exceeded the smallest detectable change. Additionally, CSI obtained a 'sufficient' rating for hypothesis testing for construct validity, as 83% of the results aligned with the hypotheses. CSI received a 'high' quality of evidence score for most of the reported measurement properties (internal consistency, reliability, construct validity and responsiveness) except structural validity and content validity. Due to a lack of clear reporting regarding the assessment of the content validation process, the quality of the studies received doubtful rating. Therefore, the level of evidence for content validity of CSI was downgraded to 'moderate' for risk of bias. Again, the results of available studies on structural validity were rated as sufficient (60%) and insufficient (40%). Therefore, the level of evidence was downgraded to 'moderate' for inconsistency. Furthermore, the quality of evidence for measurement error was not graded due to the absence of MIC value [141]. None of the studies included in the review evaluated the cross-cultural validity or criterion validity of CSI. As a result, there was a lack of available evidence on these aspects [159-195].

The Pain Sensitivity Questionnaire (PSQ) was developed to assess a patient's subjective perception of pain intensity and their pain threshold in response to different stimuli [217]. This 17-item questionnaire was designed to investigate self-reported pain sensitivity as a supplement or alternative to experimental pain testing. PSQ has been validated in patients with chronic pain [196] to assist in identifying pain sensitization and/or increased pain sensitivity because higher-than-average experimental pain sensitivity has been found in

several chronic pain disorders [128,221,222,223]. Fourteen items of PSQ relate to situations that are assessed as painful. These items address a spectrum of pain intensities to various types of pain (such as hot, cold, sharp, and dull), and describe painful situations occurring in different parts of the body (including the head, upper, and lower extremities). Three items (items 5, 9, and 13) are not normally rated as painful and are not involved in the scoring. PSQ was rated 'sufficient' for content validity, structural validity, internal consistency, and reliability. However, PSQ was rated 'insufficient' for construct validity because more than 25% of study results were not aligned with the hypotheses. No assessment of cross-cultural validity, measurement error, criterion validity, or responsiveness were found. The quality of evidence of PSQ was scored 'high' for internal consistency and construct validity. However, the quality of evidence of the other measurement properties varied from 'low' for structural validity and reliability to 'moderate' for content validity. Low evidence was found because the included studies were of 'inadequate' methodological quality. A moderate level of evidence for content validity was found due to inappropriate descriptions of the content validation process in included studies [175,196-202].

The Fibromyalgia Survey Questionnaire (FSQ): The American College of Rheumatology developed new preliminary diagnostic criteria for fibromyalgia in 2010 based on 1990 criteria [218] as well as a self-report Fibromyalgia Survey Questionnaire (FSQ) for patient surveys and clinical research in 2011 [224]. In 2016, a revised version of FSQ was published [225], to improve validity and enhance utility. FSQ has been proposed as a surrogate tool for identifying significant subgroups of patients experiencing centralized pain [226,227]. FSQ is a combination of the Widespread Pain Index (WPI) and Symptom Severity (SS) scale [211] which evaluates the severity of symptoms linked to fibromyalgia and the distribution of pain across various locations. A score of 13 or higher on this measure signifies the presence of fibromyalgia [224]. Again, when assessing centralized pain using the FSQ, it's utilized as a continuous measure, where higher scores denote a greater intensity of CS- presence [228]. FSQ was rated 'sufficient' for content validity, reliability, and construct validity. However, due to the absence of evidence for sufficient structural validity, internal consistency for FSQ was rated 'indeterminate'. None of the included studies assessed structural validity, cross-cultural validity, measurement error, criterion validity, or responsiveness. FSQ received a 'high' quality of evidence score for internal consistency and hypothesis testing, but 'moderate' evidence was found for content validity and reliability [203-211].

The Nociplastic-based Fibromyalgia Features (NFF) tool was developed and validated to detect fibromyalgia in patients with chronic pain [212]. The NFF tool is comprised of seven binary items designed to assess key features of fibromyalgia, including generalized pain, localized pain, migratory pain, worsening of pain due to physical or emotional stress, intense nature of pain, morning fatigue, and tender points. The established cutoff point is 4. NFF was rated 'sufficient' for criterion validity and construct validity. Again, the internal consistency of NFF received an 'indeterminate' rating as no factor analysis was performed to test the hypothesis of unidimensional factor structure. NFF quality of evidence was scored 'high' for internal consistency, criterion validity and construct validity. Additionally, there was no evaluation of content validity, cross-cultural validity, measurement error, criterion validity, or responsiveness [212].

The Generalized Pain Questionnaire (GPQ) was formulated to evaluate the presence and intensity of diverse symptoms often linked to generalized hypersensitivity to pain [213]. Generalized pain hypersensitivity is prevalent among individuals experiencing different chronic pain conditions and represents a manifestation of central sensitization [17, 12]. A suggested GPQ threshold score of >10 is recommended to identify potential instances of generalized heightened sensitivity to pain. GPQ was rated 'sufficient' for structural validity, internal consistency, and construct validity. However, assessments of content validity, cross-cultural validity, reliability, measurement error, criterion validity, or responsiveness were absent. GPQ received a 'high' quality of evidence score for internal consistency and 'moderate' for construct validity. However, the quality evidence of structural validity was scored 'very low'. This score was given because there was only one methodologically 'inadequate' study (due to insufficient sample size) [213].

The Sensory Hypersensitivity Scale (SHS) was developed to evaluate the sensory dimensions of hypersensitivity [214]. Sensory hypersensitivity is indicative of central sensitization. SHS measures sensitivity in two aspects: general sensitivity and sensitivity to specific modalities. SHS was rated 'insufficient' for construct validity because 60% of study results didn't correspond with the hypotheses. However, SHS was rated 'indeterminate' for internal consistency because of the absence of factor analysis to provide evidence for unidimensional factor structure resulting in downgrading of the COSMIN rating. SHS quality of evidence was scored 'high' for internal consistency, and 'moderate' for construct validity. There were no evaluations conducted regarding content validity, cross-cultural validity, reliability, measurement error, criterion validity, or responsiveness [214].

The Novel Self-report Questionnaire was developed to evaluate symptoms shown to be indicative of altered central pain processing (altered spinal and supraspinal pain processing) to identify the presence of nociplastic contributors to pain supporting the classification of primary chronic pain [215]. Novel Self-report Questionnaire was rated 'sufficient' for structural validity, internal consistency, and construct validity. The remaining measurement properties were not evaluated. Novel Self-report Questionnaire received 'low' for the quality of evidence of structural validity and construct validity because of the insufficient sample size (<100) and only one study with 'doubtful' quality respectively. However, this questionnaire was rated 'high' for its internal consistency [215].

The Leiden Visual Sensitivity Scale (L-VISS) was developed based on responsiveness to light and pattern sensitivity in people with migraine [219]. The Visual Discomfort Scale (VDS) was developed to measure visual discomfort [220]. Both scales measure possible somatic, perceptual and performance difficulties resulting from exposure to various light sources or patterns. L-VISS and VDS were validated in specific chronic pain conditions related to central sensitization to confirm the effectiveness of this combined scale for quantifying visual sensitivity in chronic pain and add to evidence highlighting the significant role of central sensitization as a mechanism behind visual discomfort [216]. L-VISS & VDS was rated 'sufficient' for construct validity. Internal consistency was rated 'indeterminate', because of the absence of structural validity assessment. L-VISS and VDS quality of evidence was scored 'High' for internal consistency. However, 'low' evidence was found for construct validity due to only one study with 'doubtful' quality. Other measurement properties were not evaluated [216].

5.5. Meta-analysis

Reliability estimates (intraclass correlation coefficient) for the CSI and PSQ were quantitatively pooled in a meta-analysis. All of these studies assessed test-retest reliability (ICC) using mostly two-way random effects models with absolute agreement, and single-rater type. Quantitative pooling of intraclass correlation coefficient (ICC) for CSI was conducted for both overall chronic pain conditions (17 studies: n=1825) and subtypes of chronic pain: chronic musculoskeletal pain (5 studies: n=684) and chronic neck pain (3 studies: n=354). Based on pooled data, the test-retest reliability of the CSI yielded a mean of 0.93 (95% CI: 0.91- 0.95), I^2 = 91.9% in overall chronic pain conditions (Figure 2); ICC of 0.90 (95% CI: 0.87- 0.93), I^2 = 70.7% in chronic musculoskeletal pain (Figure 3); and ICC of 0.93 (95% CI:

0.88- 0.99), I^2 = 94.4% in chronic neck pain (Figure 4). For the PSQ, based on two studies (n=216) the test-retest reliability (ICC) yielded a mean of 0.86 (95% CI: 0.72- 0.99), I^2 = 87.4% in chronic pain conditions (Figure 5). Meta-analysis of reliability estimates for other instruments was not conducted due to insufficient data for pooling. Additional factors, such as the duration since the onset of the condition and the anticipated stability of the pain experience, could not be considered due to insufficient information. Pooled results of the test-retest reliability of CSI and PSQ are presented in Table 10.

Table 10. Pooled results of reliability estimates in a meta-analysis

PROM	Conditions	Studies	ICC	95% CI	Cochrane's Q	P-value	I^2	τ^2
CSI	Overall chronic pain	[160-162,164,165,167,168,171, 173,174,178,179,181,182,184, 188,191]	0.93	0.91-0.95	196.92	< .0001	91.90%	0.0012
	Chronic musculoskeletal pain	[171,173,179,184,191]	0.9	0.87-0.93	13.64	0.009	70.70%	0.0007
	Chronic non-specific neck pain	[164,165,168]	0.93	0.88-0.99	35.67	< .0001	94.40%	0.0018
PSQ	Chronic pain	[200,202]	0.86	0.72-0.99	7.96	0.005	87.40%	0.0086

PROM=Patient reported outcome measures, CSI= Central sensitization Inventory, PSQ= Pain Sensitivity Questionnaire, ICC= Intraclass correlation coefficient

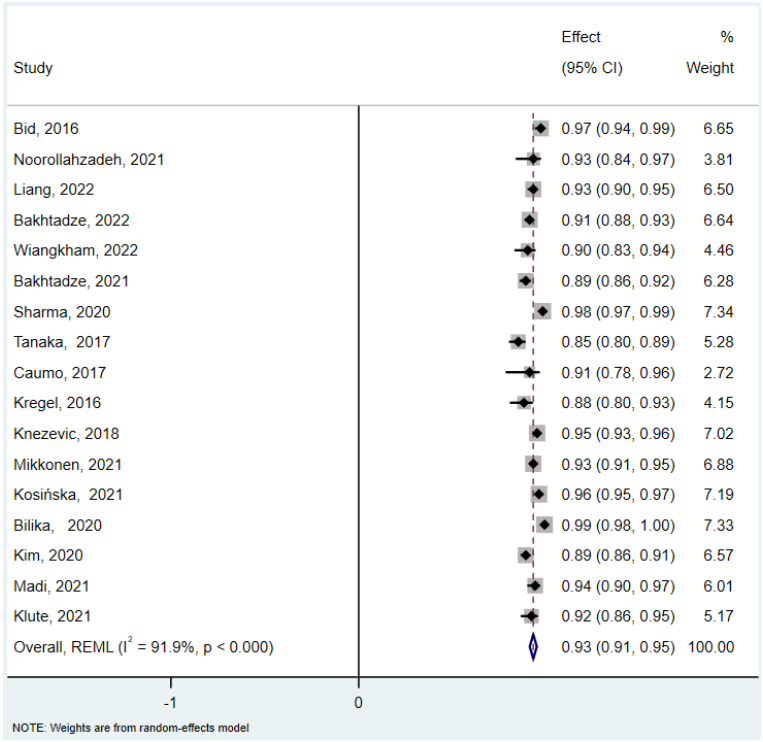

Figure 2: Forest Plot of pooled Intraclass Correlation Coefficients (ICC) of CSI in overall chronic pain conditions

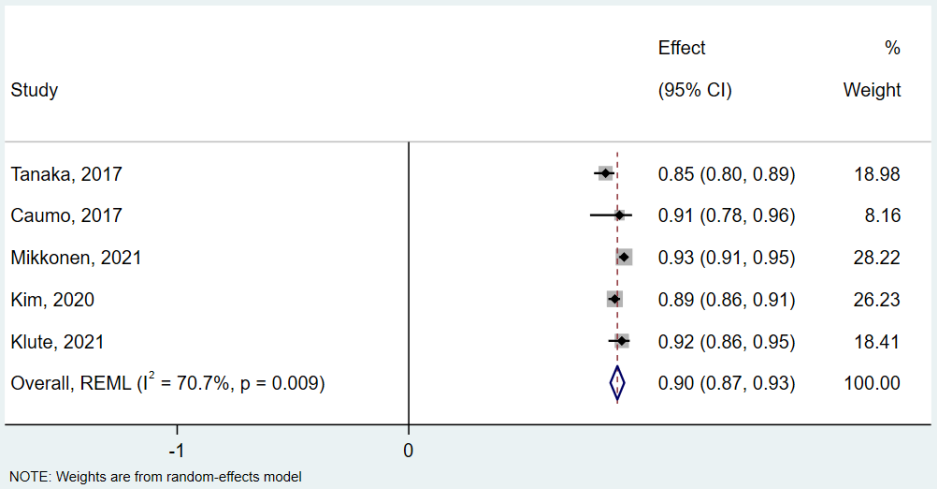

Figure 3: Forest Plot of pooled Intraclass Correlation Coefficients (ICC) of CSI in chronic musculoskeletal pain

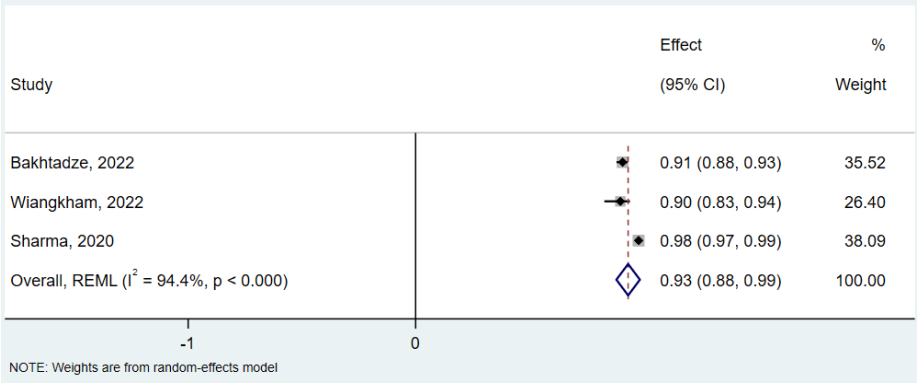

Figure 4: Forest Plot of pooled Intraclass Correlation Coefficients (ICC) of CSI in chronic neck pain

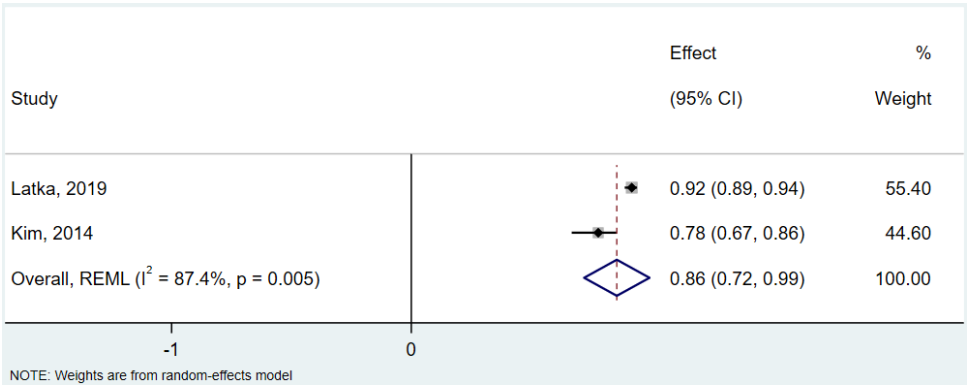

Figure 5: Forest Plot of pooled Intraclass Correlation Coefficients (ICC) of PSQ in chronic pain

Chapter 5

5 Discussion

This systematic review identified eight patient-reported outcome measures (PROMs) used to assess central sensitization manifestations in patients with chronic pain conditions. Some tools were not originally developed to assess central sensitization pain like L-VISS & VDS [219, 220]. However, these tools were validated in chronic pain patients and could be used to assess CS-related symptoms (e.g., sensory hyperresponsiveness to light and patterns) [216]. Not all the instruments were designed to assess all the core features (pain intensity, sensory hyperresponsiveness, physical well-being, pain interference with functionality, emotional/psychological distress, fatigue) associated with central sensitization. For example, PSQ [217] focuses only on pain intensity and pain threshold to external stimuli. However, it is recommended to incorporate all PROMs that assess one or more specific constructs of interest, instead of solely prioritizing the most frequently utilized PROMs [158]. The findings of this review showed that most identified PROMs have limited evidence regarding their psychometric properties. Again, the lack of sufficient evidence for most instruments does not necessarily indicate poor instrument quality. Instead, it predominantly reflects the limited availability of published articles as well as the poor methodological quality of studies examining these properties.

For many of the included studies, the rating of methodological quality of studies was doubtful or inadequate. This is due to inadequate sample size or due to unclear reporting of important information which downgraded their evaluations. For example, almost all the studies assessing content validity [160-163,165,168,173,178179,186,188,192,198,200,202,203,209] got a doubtful rating due to a lack of adequate information. In a few included studies, factor analyses were absent, despite the recommendation to assess the unidimensionality of the instrument before computing internal consistency values (i.e. Cronbach’s alpha) [160,168,182,203-207,209,211]. The time interval between test and retest is an important factor for evaluating reliability and a period of around two weeks is often regarded as a suitable interval between two administrations. However, it depends on the specific construct being assessed and the underlying condition [141, 145]. In this review, a timeframe ranging from 5 days to 2 weeks was selected, which was considered appropriate for addressing both the requirement to minimize recall bias and to ensure the stability of patients' conditions. The

time gap between the test and retest did not align with the predetermined criteria for assessing reliability in certain studies [164,165,167,174,176,196,200,202].

A total of Fourty-two studies [160-162,164-171,173,174,176-179,181,182-186,188-192,195,197-207,209] evaluated the translated version of the PROMs, but cross-cultural validity has not yet been assessed. Assessment of cross-cultural validity is crucial in translation studies to determine if the translated versions evaluate culturally different populations similarly to their original counterparts. Moreover, this review found that studies translating and validating outcome instruments into other languages have employed different translation guidelines. The evaluation of measurement error is fundamental to differentiate actual changes from both systematic and random errors. In this review, twelve studies were found to assess measurement error [160,164,165,167,168,177-179,181,182,188,191] although the MIC (minimal important change) value was not determined which signifies the meaningful change in a patient's condition. Without specifying the MIC, evidence of measurement error remains indeterminate, regardless of the proper calculation of the smallest detectable change. The COSMIN user manual reported the consensus about the nonexistence of "gold standard" for PROMs [141]. Therefore, criterion validity was not found to be assessed for most of the included instruments. Also, inadequate, and doubtful ratings of quality of some studies for hypothesis testing resulted from the absence of a clear description of the comparator instruments [164,167,176, 200,215,216]. For responsiveness, only two studies were identified [194,195] to detect the change in the assessment score following the interventions received by patients. A few studies were found to report missing items [159,162,173,174,176,184]. When there were a greater number of missing items (more than 20% of total items), we downgraded the rating of the quality of measurement property (structural validity) of instruments. However, the COSMIN steering committee reached a consensus that the absence of reporting regarding the quantity and management of missing items would not inevitably result in biased study outcomes. Moreover, there is no agreement on the optimal approach to managing missing items in studies focusing on measurement properties [141].

Among the eight instruments assessed, CSI was assessed most frequently and was identified as having the most favourable overall psychometric properties. In particular, seven out of nine measurement properties of CSI were evaluated [159–195]. The overall ratings of six measurement properties were sufficient with moderate to high evidence (content validity,

structural validity, internal consistency, reliability, hypothesis testing for construct validity, and responsiveness). Confirmatory factor analysis and/or exploratory factor analysis confirmed the adequacy of structural validity of CSI with acceptable factor loadings. The internal consistency of CSI was excellent, with Cronbach's alpha value ranging from 0.87 to 0.99. The findings from hypothesis testing indicated that CSI demonstrates a strong correlation with other PROMs assessing related but dissimilar constructs. Moreover, CSI was found to be responsive to changes following interventions. Based on 17 studies, pooled results of the test-retest reliability of the CSI demonstrated excellent reliability, with an ICC of 0.93 (95% CI: 0.91- 0.95). Despite this high reliability, there was substantial heterogeneity, as indicated by an I^2 statistic of 91.9% for overall chronic pain conditions. The variability in the observed heterogeneity could be attributed to several factors, including differences in the study populations, such as the types and severity of chronic pain conditions, as well as variations in the study's time and location. However, subgroup analysis based on different types of chronic pain conditions showed that the I^2 statistic was lower in one subgroup compared to others. This observation is noteworthy because it suggests that the subgroup analysis may have accounted for some of the heterogeneity present in the overall data. In meta-analysis, results from studies on content validity, structural validity, and cross-cultural validity often cannot be pooled [141].

The PSQ was found to assess four measurement properties having sufficient overall rating (content validity, structural validity, internal consistency, reliability) with low to high levels of evidence and one property with insufficient rating (hypothesis testing for construct validity) [196–202]. Based on the pooled result (2 studies), PSQ was found to have excellent reliability with an ICC of 0.86 (95% CI: 0.72- 0.99) and high heterogeneity I^2 = 87.4% in chronic pain conditions. The FSQ was found to have sufficient overall rating for three properties (content validity, reliability and hypothesis testing for construct validity) and an indeterminate rating for one measurement property (internal consistency) despite having high-level evidence [203–211]. While 7 studies (203-207, 209, 211) assessing FSQ reported internal consistency values, a quantitative meta-analysis to pool Cronbach's alpha was not conducted. These results cannot be considered evidence of internal consistency because the assessment of structural validity was not performed in these studies. For GPQ and novel self-report instruments, sufficient overall ratings were found for three measurement properties (structural validity, internal consistency, and hypothesis testing for construct validity) with different levels of evidence (very low to high) [213,215]. The NFF had high evidence with

sufficient rating for two measurement properties (criterion validity and hypothesis testing for construct validity) and high evidence with an indeterminate rating for one property (internal consistency) [212]. Again, two outcome measures, SHS and L-VISS & VDS, were found to assess two measurement properties (internal consistency, and hypothesis testing for construct validity) but only L-VISS & VDS had only one measurement property with an overall sufficient rating (construct validity) [216].

Selection of an appropriate PROM also involves considering other aspects like feasibility and interpretability. These are considered non-quantifiable attributes since they do not directly assess the quality of a PROM. Therefore, they are typically discussed descriptively rather than subject to evaluation [141]. Both factors are crucial considerations when determining the most suitable questionnaire, as difficulties in understanding by both patients and clinicians may suggest insufficient content validity, while the presence of floor and ceiling effects could undermine reliability. This review encountered difficulties in comparing the interpretability of the identified PROMs because the included studies lacked sufficient information. However, few studies have reported the absence of floor and ceiling effects for CSI [162,164,165,167,179,182,186,191], PSQ [200,202] and L-VSS &VDS [216].

Strengths and Limitations

To date, the present review is the first to identify available PROMs used to assess central sensitization manifestations in patients with chronic pain. A thorough and systematic evaluation of both the methodological quality of studies and the quality of psychometric properties of PROMs was conducted with the results of studies quantitatively pooled where possible. Thus this review provides a broad overview of the quality of available PROMs. While Scerbo and colleagues [138] assessed only central sensitization inventory (CSI) in a systematic review, the present review encompasses a wide variety of instruments and uses the most updated recommendations (COSMIN) of quality assessment.

There were also some limitations of this review. It's expected that the rating system utilized for assessing instruments may be prone to subjective biases (e.g. reviewers' personal preferences or interpretation of evidence) which could potentially influence the rating and consistency of the assessment. In this review, two reviewers separately participated in the assessment process, and conflicts were resolved by experts. This approach ensured that decisions were based on consensus rather than individual biases indicating both the integrity

and trustworthiness of the assessments. An additional drawback is the risk of language bias. this review only included studies published in the English language. As a result, three studies assessing CSI and PSQ, published in languages other than English were excluded during the selection process. Thus, the exclusion of those studies could have potentially influenced the assessment of those two instruments. Furthermore, publication bias was not assessed because there are no accepted methods for doing so yet in studies of systematic reviews of measurement properties of PROMs.

Recommendations for future research

This review highlights a significant research gap in assessing measurement properties of included PROMs, as none of the PROMs have been assessed in all nine measurement properties. Specifically, there is a remarkable absence of publications examining content validity, structural validity, reliability, cross-cultural validity, measurement error, and responsiveness. Content validity, in particular, is an important measurement property, and PROMs lacking it are advised to be excluded from further assessment and not recommended for use [141]. Only three instruments (CSI, PSQ, and FSQ) were found to have assessed content validity in this review. Further studies are needed to prioritize the comprehensive evaluation of the fundamental psychometric properties of all these instruments, ensuring adherence to recommended validation guidelines. Furthermore, future studies focusing on measurement properties should aim for higher methodological quality by employing adequate sample sizes, appropriate statistical methods, and comprehensive reporting of the relevant information necessary for assessing each measurement property.

In this review, there is evidence of poor reporting of many of the included studies. Insufficient reporting obstructs the precise evaluation of methodologies within the studies and the trustworthy application of their findings. Therefore, we suggest that the studies assessing the psychometric properties of PROMs place greater emphasis on following standardized reporting guidelines such as COSMIN reporting guidelines [229] to ensure that all relevant information regarding the methodology and results of these studies is transparent and easily accessible to readers, researchers and clinicians.

Numerous studies have identified core features of central sensitization [10,12,39,57,85]. The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has established six core outcome domains for chronic pain clinical trials [230]. However, if future

studies focus on the development of a standardized core outcome domain set designed to assess chronic pain with central sensitization, that will not only improve the comprehensiveness of assessments but also promote greater consistency and comparability among studies within this area. Additionally, future development of PROMs should prioritize the inclusion of robust validity checks designed to minimize the risk of malingering by the patients.

Furthermore, to facilitate the quantitative meta-analysis of the psychometric values of PROMs, the development of future guidelines is necessary to shed light on reporting parameter estimates of measurement properties, especially for those properties where quantitative pooling of study findings is possible (e.g., reliability, internal consistency, measurement error, construct validity) and on outlining the required steps for conducting the meta-analysis.

Conclusion

This systematic review provided a summary of the psychometric quality of eight identified instruments based on available publications. Additionally, this review assessed the methodological quality of the identified studies. The findings of this review could provide valuable guidance to healthcare providers and researchers to select appropriate tools for assessing treatment outcomes and making clinical decisions accordingly for providing patient-centred care in chronic pain patients with central sensitization. The CSI was found to receive the highest overall ratings with moderate to high levels of quality of evidence among all the included instruments, followed by PSQ and FSQ. The other PROMs included did not achieve similarly high and sufficient ratings due to numerous incomplete evaluations as well as a lack of publications. Although not all properties have been studied, the CSI could serve as a useful PROM for chronic pain associated with central sensitization. Careful consideration is advised when selecting an instrument, as none of the included PROMs had sufficient evidence across all nine measurement properties.

References

1. Reid KJ, Harker J, Bala MM, Truyers C, Kellen E, Bekkering GE, Kleijnen J. Epidemiology of chronic non-cancer pain in Europe: narrative review of prevalence, pain treatments and pain impact. Curr Med Res Opin. 2011 Feb;27(2):449-62.
2. Johannes CB, Le TK, Zhou X, Johnston JA, Dworkin RH. The prevalence of chronic pain in United States adults: Results of an Internet-based survey. J Pain. 2010; 11:1230–1239.
3. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet. 2021 May 29;397(10289):2082-2097.
4. Vellucci R. Heterogeneity of chronic pain. Clin Drug Investig. 2012 Feb 22;32 Suppl 1:3-10.
5. Crofford LJ. Chronic Pain: Where the Body Meets the Brain. Trans Am Clin Climatol Assoc. 2015; 126:167–83.
6. Menefee LA, Frank ED, Doghramji K, Picarello K, Park JJ, Jalali S, Perez-Schwartz L. Self-reported sleep quality and quality of life for individuals with chronic pain conditions. Clin J Pain. 2000; 16:290-297.
7. Fine PG. Long-term consequences of chronic pain: Mounting evidence for pain as a neurological disease and parallels with other chronic disease states. Pain Med. 2011;12: 996–1004.
8. Vartiainen P, Heiskanen T, Sintonen H, Roine RP, Kalso E. Health-related quality of life and burden of disease in chronic pain measured with the 15D instrument. Pain. 2016; 157:2269-2276
9. Tracey I, Mantyh PW. The cerebral signature for pain perception and its modulation. Neuron. 2007; 55:377–91.
10. Nijs J, Van Houdenhove B, Oostendorp RA. Recognition of central sensitization in patients with musculoskeletal pain: Application of pain neurophysiology in manual therapy practice. Manual Therapy. 2010; 15:135-141.
11. Latremoliere A, Woolf CJ. Central sensitization: a generator of pain hypersensitivity by central neural plasticity. J Pain. 2009; 10:895–926.
12. Woolf CJ. Central sensitization: implications for the diagnosis and treatment of pain. Pain. 2011;152(3 Suppl): S2–15.
13. Aguilar Ferrándiz ME, Nijs J, Gidron Y, Roussel N, Vanderstraeten R, Van Dyck D, Huysmans E, De Kooning M. Auto-Targeted Neurostimulation Is Not Superior to Placebo

in Chronic Low Back Pain: A Fourfold Blind Randomized Clinical Trial. Pain Physician. 2016 Jul; 19(5): E707-19.

14. Coombes BK, Bisset L, Vicenzino B. Cold hyperalgesia associated with poorer prognosis in lateral epicondylalgia: a 1-year prognostic study of physical and psychological factors. Clin J Pain. 2015; 31:30-35.
15. Kim SH, Yoon KB, Yoon DM, Yoo JH, Ahn KR. Influence of Centrally Mediated Symptoms on Postoperative Pain in Osteoarthritis Patients Undergoing Total Knee Arthroplasty: A Prospective Observational Evaluation. Pain Pract. 2015 Jul;15(6): E46-53.
16. Jull G, Sterling M, Kenardy J, Beller E. Does the presence of sensory hypersensitivity influence outcomes of physical rehabilitation for chronic whiplash? --A preliminary RCT. Pain. 2007 May; 129(1-2):28-34.
17. Sterling M, Jull G, Kenardy J. Physical and psychological factors maintain long-term predictive capacity post-whiplash injury. Pain. 2006 May;122(1-2):102-8.
18. Mills S, Torrance N, Smith BH. Identification and Management of Chronic Pain in Primary Care: a Review. Curr Psychiatry Rep. 2016 Feb;18(2):22.
19. Karcioglu O, Topacoglu H, Dikme O, Dikme O. A systematic review of the pain scales in adults: Which to use? Am J Emerg Med. 2018 Apr;36(4):707-714.
20. Dydyk AM, Givler A. Central Pain Syndrome. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan- [updated 2023 Feb 19].
21. Macfarlane GJ, Kronisch C, Dean LE, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017; 76(2):318–328.
22. den Boer C, Terluin B, van der Wouden JC, Blankenstein AH, van der Horst HE. Tests for central sensitization in general practice: a Delphi study. BMC Fam Pract. 2021 Oct 19;22(1):206.
23. Fillingim RB, Loeser JD, Baron R, Edwards RR. Assessment of Chronic Pain: Domains, Methods, and Mechanisms. J Pain. 2016 Sep; 17(9 Suppl):10-20.
24. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights. 2013 Aug 4;6:61-8
25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737-45
26. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use. (5th ed.). New York: Oxford University Press; 2015.

27. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006;4:54.
28. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118(8):622-629.]
29. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. Nat Rev Rheumatol. 2010;6(10):599-606.
30. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, Cohen M, Evers S, Finnerup NB, First MB, Giamberardino MA, Kaasa S, Kosek E, Lavand'homme P, Nicholas M, Perrot S, Scholz J, Schug S, Smith BH, Svensson P, Vlaeyen JWS, Wang SJ. A classification of chronic pain for ICD-11. Pain. 2015 Jun;156(6):1003-1007.
31. Verhaak PF, Kerssens JJ, Dekker J, Sorbi MJ, Bensing JM. Prevalence of chronic benign pain disorder among adults: A review of the literature. Pain. 1998; 77:231-239.
32. Scholz J, Woolf CJ. Can we conquer pain? Nat Neurosci. 2002 Nov;5 Suppl:1062-7.
33. Tracey I, Bushnell MC. How neuroimaging studies have challenged us to rethink: is chronic pain a disease? J Pain. 2009 Nov;10(11):1113-20.
34. Kroenke K, Krebs EE, Bair MJ. Pharmacotherapy of chronic pain: a synthesis of recommendations from systematic reviews. Gen Hosp Psychiatry. 2009; 31:206–219.
35. Fitzcharles MA, Cohen SP, Clauw DJ, Littlejohn G, Usui C, Häuser W. Nociplastic pain: towards an understanding of prevalent pain conditions. Lancet. 2021; 397(10289):2098–2110.
36. Staud R. Abnormal endogenous pain modulation is a shared characteristic of many chronic pain conditions. Expert Rev Neurother. 2012; 12:577–85.
37. Price DD, Staud R, Robinson ME, Mauderli AP, Cannon R, Vierck CJ. Enhanced temporal summation of second pain and its central modulation in fibromyalgia patients. PAIN. 2002; 99:49–59.
38. Apkarian VA, Hashmi JA, Baliki MN. Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. Pain. 2011 Mar;152(3 Suppl):S49-S64.
39. Nijs J, Torres-Cueco R, van Wilgen CP, Girbes EL, Struyf F, Roussel N, van Oosterwijck J, Daenen L, Kuppens K, Vanwerweeen L, Hermans L, Beckwee D, Voogt L, Clark J, Moloney N, Meeus M. Applying modern pain neuroscience in clinical practice: criteria for the classification of central sensitization pain. Pain Physician. 2014 Sep-Oct;17(5):447-57.

40. Fingleton C, Smart K, Moloney N, Fullen BM, Doody, C. Pain sensitization in people with knee osteoarthritis: A systematic review and meta-analysis. Osteoarthritis Cartilage. 2015; 23: 1043–1056.
41. Arendt-Nielsen L, Graven-Nielsen T. Central sensitization in fibromyalgia and other musculoskeletal disorders. Curr Pain Headache Rep. 2003;7(5):355–361.
42. Van Oosterwijck J, Nijs J, Meeus M, et al. Evidence for central sensitization in chronic whiplash: A systematic literature review. Eur J Pain. 2013;17(3):299–312.
43. Lluch E, Torres R, Nijs J, Van Oosterwijck J. Evidence for central sensitization in patients with osteoarthritis pain: a systematic literature review. Eur J Pain. 2014 Nov;18(10):1367-75.
44. Lluch E, Dueñas L, Falla D, Baert I, Meeus M, Sánchez-Frutos J, Nijs J. Preoperative Pain Neuroscience Education Combined With Knee Joint Mobilization for Knee Osteoarthritis: A Randomized Controlled Trial. Clin J Pain. 2018 Jan;34(1):44-52.
45. Sanchis MN, Lluch E, Nijs J, Struyf F, Kangasperko M. The role of central sensitization in shoulder pain: a systematic literature review. Semin Arthritis Rheum. 2015; 44:710–16.
46. Burstein R, Levy D, Jakubowski M. Effects of sensitization of trigeminovascular neurons to triptan therapy during migraine. Rev Neurol (Paris). 2005;161(6–7):658–660.
47. Chalaye P, Goffaux P, Bourgault P, Lafrenaye S, Devroede G, Watier A, Marchand S. Comparing pain modulation and autonomic responses in fibromyalgia and irritable bowel syndrome patients. Clin J Pain. 2012 Jul;28(6):519-26.
48. Nijs J, Meeus M, Van Oosterwijck J, Ickmans K, Moorkens G, Hans G, De Clerck LS. In the mind or in the brain? Scientific evidence for central sensitization in chronic fatigue syndrome. Eur J Clin Invest. 2012 Feb;42(2):203-12.
49. Roussel NA, Nijs J, Meeus M, Mylius V, Fayt C, Oostendorp R. Central sensitization and altered central pain processing in chronic low back pain: fact or myth? Clin J Pain. 2013 Jul;29(7):625-38.
50. Meeus M, Vervisch S, De Clerck LS, Moorkens G, Hans G, Nijs J. Central sensitization in patients with rheumatoid arthritis: A systematic literature review. Semin Arthritis Rheum. 2012;41:556-567.
51. Nijs J, Leysen L, Adriaenssens N, Aguilar Ferrándiz ME, Devoogdt N, Tassenoy A, Ickmans K, Goubert D, van Wilgen CP, Wijma AJ, Kuppens K, Hoelen W, Hoelen A, Moloney N, Meeus M. Pain following cancer treatment: Guidelines for the clinical classification of predominant neuropathic, nociceptive and central sensitization pain. Acta Oncol. 2016 Jun;55(6):659-63.

52. Baker JF, Conaghan PG, Emery P, Baker DG, Ostergaard M. Relationship of patient-reported outcomes with MRI measures in rheumatoid arthritis. Ann Rheum Dis. 2017; 76: 486–90.
53. Heisler AC, Song J, Dunlop DD, Wohlfahrt A, Bingham CO III, Bolster MB, Clauw DJ, Marder W, Phillips K, Neogi T, Lee YC. Association of Pain Centralization and Patient-Reported Pain in Active Rheumatoid Arthritis. Arthritis Care Res (Hoboken). 2020 Aug;72(8):1122-1129.
54. De Oliveira Silva D, Rathleff MS, Petersen K, Azevedo FM, Barton CJ. Manifestations of pain sensitization across different painful knee disorders: a systematic review including meta-analysis and metaregression. Pain Med. 2019; 20: 335–58.
55. Maixner W, Fillingim RB, Williams DA, Smith SB, Slade GD: Overlapping chronic pain conditions: Implications for diagnosis and classification. J Pain. 17:T93-T107, 2016
56. Veasley C, Clare D, Clauw DJ, Cowley T, Nguyen RHN, Reinecke P, Vernon SD, Williams DA. [Accessed July 26, 2016] Impact of chronic overlapping pain conditions on public health and the urgent need for safe and effective treatment: 2015 analysis and policy recommendations. Chronic Pain Research Alliance.
57. Yunus MB. Role of central sensitization in symptoms beyond muscle pain, and the evaluation of a patient with widespread pain. Best Pract Res Clin Rheumatol. 2007 Jun;21(3):481-97.
58. Yunus MB. Fibromyalgia and overlapping disorders: the unifying concept of central sensitivity syndromes. Semin Arthritis Rheum. 2007 Jun; 36(6):339-356.
59. Gold MS, Gebhart GF. Nociceptor sensitization in pain pathogenesis. Nature Medicine. 2010;16(11):1248-1257.
60. Tracey WD Jr. Nociception. Current Biology. 2017;27(4): R129-R133
61. Raja SN, Meyer RA, Campbell JN. Peripheral mechanisms of somatic pain. Anesthesiology. 1988; 68(4):571-590.
62. Perl E, Kumazawa T, Lynn B, Kenins P. Sensitization of high threshold receptors with unmyelinated (C) afferent fibers. Progress in Brain Research. 1976; 43, 263-277.
63. Berta T, Qadri Y, Tan PH, Ji RR. Targeting dorsal root ganglia and primary sensory neurons for the treatment of chronic pain. Expert Opin Ther Targets. 2017 Jul;21(7):695-703.
64. Basbaum AI, Bautista DM, Scherrer G, Julius D. Cellular and molecular mechanisms of pain. Cell. 2009;139(2):267-284.

65. Arendt-Nielsen L, Skou ST, Nielsen TA, Petersen KK. Altered Central Sensitization and Pain Modulation in the CNS in Chronic Joint Pain. Curr Osteoporos Rep. 2015 Aug;13(4):225-34.
66. Edwards RR. Individual differences in endogenous pain modulation as a risk factor for chronic pain. Neurology. 2005 Aug 9; 65(3):437-43.
67. Yarnitsky D. Conditioned pain modulation (the diffuse noxious inhibitory control-like effect): its relevance for acute and chronic pain states. Curr Opin Anaesthesiol. 2010 Oct; 23(5):611-5.
68. Arendt-Nielsen L, Graven-Nielsen T. Translational musculoskeletal pain research. Best Pract Res Clin Rheumatol. 2011 Apr;25(2):209-26.
69. Gebhart GF. Descending modulation of pain. Neuroscience and Biobehavioral Reviews 2004; 27, 729-737.
70. Porreca F, Ossipov MH, Gebhart G. Chronic pain and medullary descending facilitation. Trends in Neurosciences. 2002; 25, 319-325.
71. Fields HL, Basbaum AI, Heinricher MM. Central nervous system mechanisms of pain modulation. In: McMahon S, Koltzenburg M, eds. *Textbook of Pain*. 5th ed. Burlington, Massachusetts, USA: Elsevier Health Sciences; 2005:125–142.
72. Fields HL, Heinricher MM, Mason P. Neurotransmitters in nociceptive modulatory circuits. Annual Review of Neuroscience. 1991; 14, 219-245.
73. Tracey I, Dunckley P. Importance of anti- and pro-nociceptive mechanisms in human disease. Gut. 2004 Nov; 53(11):1553-5.
74. Ren K, Dubner R. Descending modulation in persistent pain: an update. Pain. 2002 Nov; 100(1-2):1-6.
75. Nijs J, Van Houdenhove B. From acute musculoskeletal pain to chronic widespread pain and fibromyalgia: application of pain neurophysiology in manual therapy practice. Man Ther. 2009; 14(1): 3–12.
76. Yunus MB. Editorial review: an update on central sensitivity syndromes and the issues of nosology and psychobiology. Curr Rheumatol Rev. 2015; 11(2):70–85.
77. Nijs J, George SZ, Clauw DJ, Fernández-de-Las-Peñas C, Kosek E, Ickmans K, Fernández-Carnero J, Polli A, Kapreli E, Huysmans E, Cuesta-Vargas AI, Mani R, Lundberg M, Leysen L, Rice D, Sterling M, Curatolo M. Central sensitisation in chronic pain conditions: latest discoveries and their potential for precision medicine. Lancet Rheumatol. 2021 May;3(5):e383-e392.

78. Harte S, Harris R, Clauw D. The neurobiology of central sensitization. J Appl Biobehav Res. 2018;23.
79. Staud R, Craggs JG, Perlstein WM, Robinson ME, Price DD. Brain activity associated with slow temporal summation of C-fibre evoked pain in fibromyalgia patients and healthy controls. European Journal of Pain (London, England). 2008; 12:1078-1089.
80. Meeus M, Nijs J, Van de Wauwer N, Toeback L, Truijen S. Diffuse noxious inhibitory control is delayed in chronic fatigue syndrome: An experimental study. Pain. 2008; 139:439-448.
81. Filatova E, Latysheva N, Kurenkov A. Evidence of persistent central sensitization in chronic headaches: A multi-method study. The Journal of Headache and Pain. 2008; 9:295-300.
82. Raphael KG, Janal MN, Anathan S, Cook DB, Staud R. Temporal summation of heat pain in temporomandibular disorder patients. Journal of Orofacial Pain. 2009; 23:54-64.
83. Smart KM, Blake C, Staines A, Doody C. Self-reported pain severity, quality of life, disability, anxiety and depression in patients classified with ‘nociceptive’, ‘peripheral neuropathic’ and ‘central sensitization’ pain. The discriminant validity of mechanisms-based classifications of low back (leg) pain. Man Ther. 2012; 17:119-125.
84. Staud R, Robinson ME, Price DD. Temporal summation of second pain and its maintenance are useful for characterizing widespread central sensitization of fibromyalgia patients. J Pain. 2007 Nov; 8(11):893-901.
85. Kosek E, Clauw D, Nijs J, Baron R, Gilron I, Harris RE, Mico JA, Rice ASC, Sterling M. Chronic nociplastic pain affecting the musculoskeletal system: clinical criteria and grading system. Pain 2021 Nov 1;162(11):2629-2634.
86. (IASP). International Association for the Study of Pain. IASP terminology. December 14, 2017. https://www.iasppain.org/terminology?navItemNumber=576#Centralsensitization (Accessed 28-3-2019).
87. Murphy AE, Minhas D, Clauw DJ, Lee YC. Identifying and Managing Nociplastic Pain in Individuals with Rheumatic Diseases: A Narrative Review. Arthritis Care Res. 2023; 75: 2215-2222.
88. Wolfe F, Häuser W, Hassett AL, Katz RS, Walitt BT. The development of fibromyalgia--I: examination of rates and predictors in patients with rheumatoid arthritis (RA). Pain. 2011 Feb;152(2):291-299.

89. Merskey H, BNE. IASP Task Force on Taxonomy. Part III: Pain Terms, A Current List with Definitions and Notes on Usage Classification of Chronic Pain. 2nd ed. Seattle, USA: International Association of the Study of Pain (IASP); 1994.
90. Scholz J, Finnerup NB, Attal N, Aziz Q, Baron R, Bennett MI, Benoliel R, Cohen M, Cruccu G, Davis KD, et al. The IASP classification of chronic pain for ICD-11: chronic neuropathic pain. Pain. 2019 Jan;160(1):53-59.
91. Fleming KC, Volcheck MM. Central sensitization syndrome and the initial evaluation of a patient with fibromyalgia: a review. Rambam Maimonides Med J. 2015; 6(2): e0020.
92. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA. 1998 Jul 8;280(2):147-51.
93. Dydyk AM, Givler A. Central Pain Syndrome. 2023 Feb 19. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 31971703.
94. Skevington SM. Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. Pain. 1998; 76:395–406.
95. Coghill RC, McHaffie JG, Yen YF. Neural correlates of interindividual differences in the subjective experience of pain. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8538-42.
96. Apkarian AV, Bushnell MC, Treede RD, Zubieta JK. Human brain mechanisms of pain perception and regulation in health and disease. Eur J Pain. 2005; 9:463-484.
97. Giesecke T, Gracely RH, Grant MA, Nachemson A, Petzke F, Williams DA, Clauw DJ. Evidence of augmented central pain processing in idiopathic chronic low back pain. Arthritis Rheum. 2004; 50:613–23.
98. De Lange FP, Kalkman JS, Bleijenberg G, Hagoort P, van der Werf SP, van der Meer JW, Toni I. Neural correlates of the chronic fatigue syndrome—an fMRI study. Brain 2004; 127:1948-1957.
99. Yuan YZ, Tao RJ, Xu B, Sun J, Chen KM, Miao F, Zhang ZW, Xu JY. Functional brain imaging in irritable bowel syndrome with rectal balloon-distention by using fMRI. World J Gastroenterol. 2003 Jun;9(6):1356-60.
100. Kwiatek R, Barnden L, Tedman R, Jarrett R, Chew J, Rowe C, Pile K. Regional cerebral blood flow in fibromyalgia: single-photon-emission computed tomography evidence of reduction in the pontine tegmentum and thalami. Arthritis Rheum. 2000 Dec;43(12):2823-33.
101. Arendt-Nielsen L, Yarnitsky D. Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. J Pain. 2009 Jun;10(6):556-72.

102. Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, et al. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): Standardized protocol and reference values. Pain. 2006;123:231–243.
103. Arant KR, Katz JN, Neogi T. Quantitative sensory testing: identifying pain characteristics in patients with osteoarthritis. Osteoarthritis Cartilage. 2022 Jan;30(1):17-31.
104. Mackey IG, Dixon EA, Johnson K, Kong JT. Dynamic Quantitative Sensory Testing to Characterize Central Pain Processing. J Vis Exp. 2017 Feb 16;(120):54452.
105. O'Brien AT, Deitos A, Pego YT, Fregni F, Carrillo-de-la-Peña MT. Defective endogenous pain modulation in fibromyalgia: a meta-analysis of temporal summation and conditioned pain modulation paradigms. J Pain. 2018; 19:819–36.
106. Costa LdaC, Maher CG, McAuley JH, Hancock MJ, Herbert RD, Refshauge KM, Henschke N. Prognosis for patients with chronic low back pain: inception cohort study. BMJ. 2009; 339:b3829–8.
107. Ryan A, Healey M, Cheng C, Dior U, Reddington C. Central sensitisation in pelvic pain: a cohort study. Aust N Z J Obstet Gynaecol. 2022;62(6):868–74.
108. Den Boer C, Terluin B, van der Wouden JC, Blankenstein AH, van der Horst HE. Tests for central sensitization in general practice: a Delphi study. BMC Fam Pract. 2021;22(1):206.
109. Guyatt GH, Ferrans CE, Halyard MY, Revicki DA, Symonds TL, Varricchio CG, Kotzeva A, Valderas JM, Alonso J; Clinical Significance Consensus Meeting Group. Exploration of the value of health-related quality-of-life information from clinical research and into clinical practice. Mayo Clin Proc. 2007 Oct;82(10):1229-39.
110. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. N Engl J Med. 1996; 334:835-40.
111. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. Science. 1998; 282:245e6.
112. Sajid MS, Tonsi A, Baig MK. "Health-related quality of life measurement". International Journal of Health Care Quality Assurance. 2008; Vol. 21 No. 4, pp. 365-373.
113. Chassany O, Sagnier P, Marquis P, Fullerton S, Aaronson N. Patient-Reported Outcomes: The Example of Health-Related Quality of Life—A European Guidance Document for the Improved Integration of Health-Related Quality of Life Assessment in the Drug Regulatory Process. Drug Inf J. 2002;36(1):209-238.
114. Pascoal C, Brasil S, Francisco R, Marques-da-Silva D, Rafalko A, Jaeken J, Videira PA, Barros L, Dos Reis Ferreira V. Patient and observer reported outcome measures to evaluate

health-related quality of life in inherited metabolic diseases: a scoping review. Orphanet J Rare Dis. 2018 Nov 28;13(1):215.

115. Devlin, N. and Appleby, J. Getting the Most out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making. 2010.
116. Ferrans CE. Differences in what quality-of-life instruments measure. J Natl Cancer Inst Monogr. 2007; 37:22-6.
117. Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, Rothman M; PRO Harmonization Group. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health. 2003 Sep-Oct;6(5):522-31.
118. Nunnally JC, Bernstein IR. Psychometric theory. 3rd ed. 1994. New York: McGraw-Hill.
119. Likert R. A technique for the measurement of attitudes. Archives of Psychology. 1932;22(140):55.
120. Revicki DA, Ehreth JL. Health-related quality-of-life assessment and planning for the pharmaceutical industry. Clin Ther. 1997; 19: 1101–1115.
121. Basch E, Bennett AV. Patient-reported outcomes in clinical trials of rare diseases. J Gen Intern Med. 2014;29 SUPPL. 3:801–803.
122. Zaugg, M., Baur, H. & Schmitt, KU. Applying patient-reported outcome measures (PROMs) in physiotherapy: an evaluation based on the QUALITOUCH Activity Index. Arch Physiother 12, 27 (2022).
123. Greenhalgh J, Dalkin S, Gooding K, Gibbons E, Wright J, Meads D, Black N, Valderas JM, Pawson R. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library; 2017 Jan.
124. Nguyen H, Butow P, Dhillon H, Sundaresan P. A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs) in routine cancer care. J Med Radiat Sci. 2021 Jun;68(2):186-195.
125. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007–2013). Contemporary Clinical Trials. 2015;43:1-9.
126. Holmes MM, Lewith G, Newell D, Field J, Bishop FL. The impact of patient-reported outcome measures in clinical practice for pain: a systematic review. Qual Life Res. 2017 Feb;26(2):245-257.

127. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018 May;27(5):1147-1157.
128. Gagnier JJ, Johnston BC. Poor quality patient reported outcome measures bias effect estimates in orthopaedic randomized studies. J Clin Epidemiol. 2019 Dec;116:36-38..
129. Suiter AM, Sarli CC. Selecting a Journal for Publication: Criteria to Consider. Mo Med. 2019 Nov-Dec;116(6):461-465.
130. Powell C. The Delphi technique: myths and realities. J Adv Nurs. 2003;41:376e82.
131. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias Checklist for Systematic Reviews of Patient-Reported Outcome Measures. Qual. Life Res 2018, 27, 1171–1179.
132. Higgins J, Green S, Eds. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011].
133. Deeks, JJ, Bossuyt, PM, Gatsonis, Eds. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Reviews; London, UK, 2013; pp. 3–15.
134. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015 Jan 1;4(1):1.
135. Eden J, Levit L, Berg A, Morton S. Finding What Sorts in Health Care: Standards for Systematic Reviews; National Academic Press: Washington, DC, USA, 2011.
136. Schunemann H, Brozek J, Guyatt G, Oxman G. GRADE Handbook. Handbook for Grading the Quality of Evidence and the Strength Recommendation Using GRADE Approach; Eds.; 2013.
137. Rutter-Locher Z, Arumalla N, Norton S, Taams LS, Kirkham BW, Bannister K. A systematic review and meta-analysis of questionnaires to screen for pain sensitisation and neuropathic like pain in inflammatory arthritis. Semin Arthritis Rheum. 2023 Aug; 61:152207.
138. Scerbo T, Colasurdo J, Dunn S, Unger J, Nijs J, Cook C. Measurement Properties of the Central Sensitization Inventory: A Systematic Review. Pain Pract. 2018 Apr;18(4):544-554.
139. Middlebrook N, Rushton AB, Abichandani D, Kuithan P, Heneghan NR, Falla D. Measures of central sensitization and their measurement properties in musculoskeletal trauma: A systematic review. Eur J Pain. 2021 Jan;25(1):71-87.

140. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res. 2009 Oct;18(8):1115-23.
141. Mokkink LB, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, De Vet HCW, Terwee CB. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User Manual Version 1 Feb 2018. 2018
142. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol, DL, et al. COSMIN checklist manual. COSMIN manual [Internet]. 2012.
143. Polit DF, Yang FM. Measurement and the measurement of change. China: Wolters Kluwer; 2016.
144. Waltz CF, Strickland OL, Lenz ER. Measurement in Nursing and Health Research. 5 ed. New York (US): Springer Publishing Company, LLC; 2017.
145. Streiner DL, Norman G. Health Measurement Scales. A practical guide to their development and use. 4th edition ed. New York: Oxford University Press; 2008.
146. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin. 1968; 70:213-20.
147. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996; 1:30–46.
148. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284-290.
149. Cortina JM. What is coefficient alpha? An examination of theory and applications. Appl Psychology. 1993; 78:98-104.
150. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49.
151. Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. Psychometrika. 1951; 16:297-334.
152. Streiner DL. Figuring out factors: The use and misuse of factor analysis. Can J Psychiatry. 1994; 39:135-40.
153. de Vet HC, Terwee CB, Mokkink L, Knol DL. Measurement in Medicine: a practical guide: Cambridge University Press; 2010.
154. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. Journal of Clinical Epidemiology. 2006; 59:1033-9.

155. Floyd FJ, Widaman FK. Factor analysis in the development and refinement of clinical assessment instruments. Psychological Assessment. 1995; 7:286-99.
156. Teresi JA, Ocepek-Welikson K, Kleinman M, Eimicke JP, Crane PK, Jones RN, et al. Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. Psychol Sci Q 2009;51(2):148-80.
157. Embretson SE & Reise SP. Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates, Inc., Publishers. 2000.
158. Cochrane Training. Chapter analyzing data and undertaking meta-analyses. 2022 https://training.cochrane.org/handbook/current/chapter-10#section-10-10.
159. Mayer TG, Neblett R, Cohen H, Howard KJ, Choi YH, Williams MJ, Perez Y, Gatchel RJ. The development and psychometric validation of the central sensitization inventory. Pain Pract. 2012 Apr;12(4):276-85.
160. Bid D, Neela S, Rathod P, Ramalingam T. Content Validity and Test-Retest Reliability of the Gujarati Version of the Central Sensitization Inventory. NJIRM. 2016;7:18-24.
161. Noorollahzadeh K, Kahrizi S, Fesharaki MG, Heidarian M, Neblett R, Behzadipour S. Cross-cultural adaptation and psychometric validation of the Persian version of the central sensitization inventory. Musculoskelet Sci Pract. 2021 Feb; 51:102314.
162. Liang D, Yu X, Guo X, Zhang J. Adaptation and validation of the Chinese version of the Central Sensitisation Inventory in patients with chronic pain. Gen Psychiatr. 2022 Dec 23;35(6):e100919.
163. Roby NU, Packham TL, MacDermid JC, Carlesso LC. Validity of the Central Sensitization Inventory (CSI) through Rasch analysis in patients with knee osteoarthritis. Clin Rheumatol. 2022; 41(10):3159–68.
164. Bakhtadze M, Kukushkin ML, Filatova EG, Latysheva NV, Proskuryakov KV, Berdnikova AV. Russian language version of the Central Sensitization Inventory: A study of validity and reliability in nonspecific neck pain, accompanied by headaches. Russian Journal of Pain. 2022;20(1):12
165. Wiangkham T, Phungwattanakul N, Rinpol T, Somsiri A, Mainoiy S, Bannathong T, Sanso S, Kongmee I, Jaiyote S, Kuesoongnern S, Sookcharoen P, Laoruengthana A, Suwanmongkhon W, Luciano JV, Neblett R. Central Sensitisation Inventory-Thai version: translation, cross-cultural adaptation, and psychometric evaluation in chronic non-specific neck pain. Disabil Rehabil. 2022 Nov 28:1-8.

166. Hendriks E, Voogt L, Lenoir D, Coppieters I, Ickmans K. Convergent Validity of the Central Sensitization Inventory in Chronic Whiplash-Associated Disorders; Associations with Quantitative Sensory Testing, Pain Intensity, Fatigue, and Psychosocial Factors. Pain Med. 2020 Dec 25;21(12):3401-3412.
167. Bakhtadze M, Kukushkin ML, Churyukanov M, Davidov O, Proskuryakov KV, Kachanovsky M. Russian language version of the Central Sensitization Inventory: validity and reliability of the questionnaire for chronic nonspecific neck and back pain. Russian J Pain. 2021;19(1):12.
168. Sharma S, Jha J, Pathak A, Neblett R. Translation, cross-cultural adaptation, and measurement properties of the Nepali version of the central sensitization inventory (CSI). BMC Neurol. 2020 Jul 27;20(1):286.
169. Düzce Keleş E, Birtane M, Ekuklu G, Kılınçer C, Çalıyurt O, Taştekin N, Is EE, Ketenci A, Neblett R. Validity and reliability of the Turkish version of the central sensitization inventory. Arch Rheumatol. 2021 Oct 18;36(4):518-526.
170. Salaffi F, Farah S, Mariani C, Sarzi-Puttini P, Di Carlo M. Validity of the Central Sensitization Inventory compared with traditional measures of disease severity in fibromyalgia. Pain Pract. 2022 Nov; 22(8):702-710.
171. Tanaka K, Nishigami T, Mibu A, Manfuku M, Yono S, Shinohara Y, Tanabe A, Ono R. Validation of the Japanese version of the Central Sensitization Inventory in patients with musculoskeletal disorders. PLoS One. 2017 Dec 7;12(12):e0188719.
172. Neblett R, Cohen H, Choi Y, Hartzell MM, Williams M, Mayer TG, Gatchel RJ. The Central Sensitization Inventory (CSI): establishing clinically significant values for identifying central sensitivity syndromes in an outpatient chronic pain sample. J Pain. 2013 May;14(5):438-45.
173. Caumo W, Antunes LC, Elkfury JL, Herbstrith EG, Busanello Sipmann R, Souza A, Torres IL, Souza Dos Santos V, Neblett R. The Central Sensitization Inventory validated and adapted for a Brazilian population: psychometric properties and its relationship with brain-derived neurotrophic factor. J Pain Res. 2017 Sep 1; 10:2109-2122.
174. Kregel J, Vuijk PJ, Descheemaeker F, Keizer D, van der Noord R, Nijs J, Cagnie B, Meeus M, van Wilgen P. The Dutch Central Sensitization Inventory (CSI): Factor Analysis, Discriminative Power, and Test-Retest Reliability. Clin J Pain. 2016 Jul;32(7):624-30.
175. Coronado RA, George SZ. The Central Sensitization Inventory and Pain Sensitivity Questionnaire: An exploration of construct validity and associations with widespread pain

sensitivity among individuals with shoulder pain. Musculoskelet Sci Pract. 2018 Aug; 36:61-67.

176. Feng B, Hu X, Lu WW, Wang Y, Ip WY. Cultural Validation of the Chinese Central Sensitization Inventory in Patients with Chronic Pain and its Predictive Ability of Comorbid Central Sensitivity Syndromes. J Pain Res. 2022 Feb 15; 15:467-477.
177. Cuesta-Vargas AI, Roldan-Jimenez C, Neblett R, Gatchel RJ. Cross-cultural adaptation and validity of the Spanish central sensitization inventory. Springerplus 2016 Oct 21;5(1):1837.
178. Knezevic A, Neblett R, Jeremic-Knezevic M, Tomasevic-Todorovic S, Boskovic K, Colovic P, Cuesta-Vargas A. Cross-Cultural Adaptation and Psychometric Validation of the Serbian Version of the Central Sensitization Inventory. Pain Pract. 2018 Apr; 18(4):463-472.
179. Mikkonen J, Luomajoki H, Airaksinen O, Neblett R, Selander T, Leinonen V. Cross-cultural adaptation and validation of the Finnish version of the central sensitization inventory and its relationship with dizziness and postural control. BMC Neurol. 2021 Mar 31; 21(1):141.
180. Neblett R, Hartzell MM, Mayer TG, Cohen H, Gatchel RJ. Establishing Clinically Relevant Severity Levels for the Central Sensitization Inventory. Pain Pract. 2017 Feb;17(2):166-175.
181. Kosińska B, Tarnacka B, Turczyn P, Gromadzka G, Malec-Milewska M, Janikowska-Hołowenko D, Neblett R. Psychometric validation of the Polish version of the Central Sensitization Inventory in subjects with chronic spinal pain. BMC Neurol. 2021 Dec 11;21(1):483.
182. Bilika P, Neblett R, Georgoudis G, Dimitriadis Z, Fandridis E, Strimpakos N, Kapreli E. Cross-cultural Adaptation and Psychometric Properties of the Greek Version of the Central Sensitization Inventory. Pain Pract. 2020 Feb;20(2):188-196.
183. Kregel J, Schumacher C, Dolphens M, Malfliet A, Goubert D, Lenoir D, Cagnie B, Meeus M, Coppieters I. Convergent Validity of the Dutch Central Sensitization Inventory: Associations with Psychophysical Pain Measures, Quality of Life, Disability, and Pain Cognitions in Patients with Chronic Spinal Pain. Pain Pract. 2018;18.
184. Kim MS, Koh IJ, Kim CK, Choi KY, Kim CY, In Y. Cross-cultural adaptation and validation of the Korean version of the Central Sensitization Inventory in patients undergoing total knee arthroplasty for knee osteoarthritis. Plos one. 2020 Dec 1;15(12):e0242912.

185. van der Noord R, Paap D, van Wilgen CP. Convergent validity and clinically relevant categories for the Dutch Central Sensitization Inventory in patients with chronic pain. J Appl Biobehav Res. 2018;23(2):Article 12119.
186. Chiarotto A, Viti C, Sulli A, Cutolo M, Testa M, Piscitelli D. Cross-cultural adaptation and validity of the Italian version of the Central Sensitization Inventory. Musculoskelet Sci Pract. 2018 Oct;37:20-28.
187. Valera-Calero JA, Úbeda-D'Ocasar E, Arias-Buría JL, Fernández-de-Las-Peñas C, Gallego-Sendarrubias GM, Cigarán-Méndez M. Convergent validity of the central sensitization inventory in women with fibromyalgia: Association with clinical, psychological and psychophysical outcomes. Eur J Pain. 2022 Nov; 26(10):2141-2151.
188. Madi M, Hamzeh H, Abujaber S, Altubasi I. Cross cultural adaptation, validity, and reliability of Central Sensitization Inventory in Arabic language. Disabil Rehabil. 2022 Dec; 44(25):8075-8083.
189. Knezevic A, Neblett R, Colovic P, Jeremic-Knezevic M, Bugarski-Ignjatovic V, Klasnja A, Pantelinac S, Pjevic M. Convergent and Discriminant Validity of the Serbian Version of the Central Sensitization Inventory. Pain Pract. 2020 Sep; 20(7):724-736.
190. Van Wilgen CP, Vuijk PJ, Kregel J, Voogt L, Meeus M, Descheemaeker F, Keizer D, Nijs J. Psychological distress and widespread pain contribute to the variance of the central sensitization inventory: a cross-sectional study in patients with chronic pain. Pain Practice. 2018 Feb;18(2):239-46.
191. Klute M, Laekeman M, Kuss K, Petzke F, Dieterich A, Leha A, Neblett R, Ehrhardt S, Ulma J, Schäfer A. Cross-cultural adaptation and validation of the German Central Sensitization Inventory (CSI-GE). BMC Musculoskelet Disord. 2021 Aug 18;22(1):708.
192. Holm LA, Nim CG, Lauridsen HH, Filtenborg JB, O'Neill SF. "Convergent validity of the central sensitization inventory and experimental testing of pain sensitivity". Scand J Pain. 2021 Oct 19;22(3):597-613.
193. Neblett R, Hartzell MM, Cohen H, Mayer TG, Williams M, Choi Y, Gatchel RJ. Ability of the central sensitization inventory to identify central sensitivity syndromes in an outpatient chronic pain sample. Clin J Pain. 2015 Apr; 31(4):323-32.
194. Neblett R, Hartzell MM, Williams M, Bevers KR, Mayer TG, Gatchel RJ. Use of the Central Sensitization Inventory (CSI) as a treatment outcome measure for patients with chronic spinal pain disorder in a functional restoration program. Spine J. 2017 Dec;17(12):1819-1829.

195. Bid DD, Soni NC, SYadav A, Rathod PV. A study on central sensitization in chronic non-specific low back pain. Indian Journal of Physiotherapy and Occupational Therapy. 2017;11(4):166.
196. Ruscheweyh R, Verneuer B, Dany K, Marziniak M, Wolowski A, Çolak-Ekici R, Schulte TL, Bullmann V, Grewe S, Gralow I, Evers S, Knecht S. Validation of the pain sensitivity questionnaire in chronic pain patients. Pain. 2012 Jun;153(6):1210-1218.
197. Sellers AB, Ruscheweyh R, Kelley BJ, Ness TJ, Vetter TR. Validation of the English language pain sensitivity questionnaire. Reg Anesth Pain Med. 2013 Nov-Dec; 38(6):508-14.
198. Ibancos-Losada MDR, Osuna-Pérez MC, Cortés-Pérez I, Montoro-Cárdenas D, Díaz-Fernández Á. Validation and Cross-Cultural Adaptation of the Spanish Version of the Pain Sensitivity Questionnaire (PSQ-S). J Clin Med. 2021 Dec 28;11(1):151.
199. Inal FY, Gul K, Yilmaz Camgoz Y, Daskaya H, Kocoglu H. Validation of the Turkish version of the Pain Sensitivity Questionnaire in patients with chronic pain. J Int Med Res. 2021 Dec;49(12):3000605211060158.
200. Latka D, Miekisiak G, Kozlowska K, Olbrycht T, Chowaniec J, Latka K, Niedzwiecki M, Ruscheweyh R, Jarmuzek P. Translation, validation, and cross-cultural adaptation of the Polish version of the pain sensitivity questionnaire. J Pain Res. 2019 Mar 14; 12:969-973.
201. Grundström H, Larsson B, Arendt-Nielsen L, Gerdle B, Kjølhede P. Associations between pain thresholds for heat, cold and pressure, and Pain Sensitivity Questionnaire scores in healthy women and in women with persistent pelvic pain. Eur J Pain. 2019; 23: 1631–1639.
202. Kim HJ, Ruscheweyh R, Yeo JH, Cho HG, Yi JM, Chang BS, Lee CK, Yeom JS. Translation, cross-cultural adaptation, and validity of the Korean version of the pain sensitivity questionnaire in chronic pain patients. Pain Pract. 2014 Nov;14(8):745-51.
203. Carrillo-de-la-Peña MT, Triñanes Y, González-Villar A, Romero-Yuste S, Gómez-Perretta C, Arias M, Wolfe F. Convergence between the 1990 and 2010 ACR diagnostic criteria and validation of the Spanish version of the Fibromyalgia Survey Questionnaire (FSQ). Rheumatol Int. 2015 Jan;35(1):141-51.
204. Häuser W, Jung E, Erbslöh-Möller B, Gesmann M, Kühn-Becker H, Petermann F, Langhorst J, Weiss T, Winkelmann A, Wolfe F. Validation of the Fibromyalgia Survey Questionnaire within a cross-sectional survey. PLoS One. 2012;7(5):e37504.

205. Fors EA, Wensaas KA, Eide H, Jaatun EA, Clauw DJ, Wolfe F, Helvik AS. Fibromyalgia 2016 criteria and assessments: comprehensive validation in a Norwegian population. Scand J Pain. 2020 Oct 25;20(4):663-672.
206. Jiao J, Cheng ZY, Xiao YY, Wang H, Zhang YF, Zhao YY, Jia Y. Validation of the 2011 and 2016 American college of rheumatology diagnostic criteria for fibromyalgia in a Chinese population. Ann Med. 2023;55(2):2249921.
207. Kang JH, An M, Choi SE, Xu H, Park DJ, Lee JK, Lee SS. Performance of the revised 2016 fibromyalgia diagnostic criteria in Korean patients with fibromyalgia. Int J Rheum Dis. 2019 Sep;22(9):1734-1740.
208. Moore MN, Wallace BI, Song J, Muhammad LN, Heisler AC, Clauw DJ, Bolster MB, Marder W, Neogi T, Wohlfahrt A, Dunlop DD, Lee YC. Correlation of Fibromyalgia Survey Questionnaire and Quantitative Sensory Testing Among Patients With Active Rheumatoid Arthritis. J Rheumatol. 2022 Sep;49(9):1052-1057.
209. Aguirre Cárdenas C, Oñederra MC, Esparza Benavente C, Durán J, González Tugas M, Gómez-Pérez L. Psychometric Properties of the Fibromyalgia Survey Questionnaire in Chilean Women With Fibromyalgia. J Clin Rheumatol. 2021 Sep 1;27(6S): S284-S293.
210. Neville SJ, Clauw AD, Moser SE, Urquhart AG, Clauw DJ, Brummett CM, Harte SE. Association Between the 2011 Fibromyalgia Survey Criteria and Multisite Pain Sensitivity in Knee Osteoarthritis. Clin J Pain. 2018 Oct;34(10):909-917.
211. Bidari A, Ghavidel-Parsa B, Amir Maafi A, Montazeri A, Ghalehbaghi B, Hassankhani A, Aarabi Y, Haghdoost A. Validation of fibromyalgia survey questionnaire and polysymptomatic distress scale in a Persian population. Rheumatol Int. 2015 Dec;35(12):2013-9.
212. Ghavidel-Parsa B, Bidari A, Atrkarroushan Z, Khosousi MJ. Implication of the Nociplastic Features for Clinical Diagnosis of Fibromyalgia: Development of the Preliminary Nociplastic-Based Fibromyalgia Features (NFF) Tool. ACR Open Rheumatol. 2022 Mar;4(3):260-268.
213. van Bemmel PF, Voshaar MAO, Klooster PMT, Vonkeman HE, van de Laar MA. Development and preliminary evaluation of a short self-report measure of generalized pain hypersensitivity. J Pain Res. 2019 Jan 17;12:395-404.
214. Dixon EA, Benham G, Sturgeon JA, Mackey S, Johnson KA, Younger J. Development of the Sensory Hypersensitivity Scale (SHS): a self-report tool for assessing sensitivity to sensory stimuli. J Behav Med. 2016 Jun;39(3):537-50.

215. Austin PD, Asghari A, Costa DSJ, Siddall PJ. The development of a novel questionnaire assessing alterations in central pain processing in people with and without chronic pain. Scand J Pain. 2020 Apr 28;20(2):407-417.
216. Ten Brink AF, Proulx MJ, Bultitude JH. Validation of the Leiden Visual Sensitivity Scale and Visual Discomfort Scale in Chronic Pain Conditions. Perception. 2021 May;50(5):399-417.
217. Ruscheweyh R, Marziniak M, Stumpenhorst F, Reinholz J, Knecht S. Pain sensitivity can be assessed by self-rating: Development and validation of the Pain Sensitivity Questionnaire. Pain. 2009 Nov;146(1-2):65-74.
218. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB, Yunus MB. The American College of Rheumatology Preliminary Diagnostic Criteria for Fibromyalgia and Measurement of Symptom Severity. Arthritis Care Res. 2010; 62:600-610.
219. Perenboom MJL, Zamanipoor Najafabadi AH, Zielman R, Carpay JA, Ferrari MD. Quantifying visual allodynia across migraine subtypes: the Leiden Visual Sensitivity Scale. Pain. 2018 Nov;159(11):2375-2382.
220. Conlon E, Lovegrove W, Chekaluk E, Pattison P. Measuring Visual Discomfort. Visual Cognition. 1999;6.
221. Maixner W, Fillingim R, Booker D, Sigurdsson A. Sensitivity of patients with painful temporomandibular disorders to experimentally evoked pain. Pain. 1995; 63:341–51.
222. Lautenbacher S, Rollman GB, McCain GA. Multi-method assessment of experimental and clinical pain in patients with fibromyalgia. Pain. 1994; 59:45–53.
223. Schoenen J, Bottin D, Hardy F, Gerard P. Cephalic and extracephalic pressure pain thresholds in chronic tension-type headache. Pain. 1991; 47:145–9.
224. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Hauser W, Katz RS, Mease P, Russell AS, Russell IJ, Winfield JB. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR Preliminary Diagnostic Criteria for Fibromyalgia. The Journal of rheumatology. 2011; 38:1113–1122.
225. Wolfe F, Clauw DJ, Fitzcharles MA, Goldenberg DL, Häuser W, Katz RL, Mease PJ, Russell AS, Russell IJ, Walitt B. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum. 2016 Dec;46(3):319-329.
226. Wilson AT, Razzell C, Hanney WJ. The Association Between the Patient Self-Report Survey for the Assessment of Fibromyalgia with Pain Sensitivity and Psychological Factors in Individuals with Musculoskeletal Pain. J Pain Res. 2023; 16:3297-3308

227. Wolfe F. New American College of Rheumatology criteria for fbromyalgia: a twenty-year journey. Arthritis Care Res (Hoboken). 2010;62(5):583–4.
228. Wolfe F. Fibromyalgianess. Arthritis Rheum. 2009;61(6):715–6.
229. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021 Aug;30(8):2197-2218.
230. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, Cleeland C, Dionne R, Farrar JT, Galer BS, Hewitt DJ, Jadad AR, Katz NP, Kramer LD, Manning DC, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robinson JP, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Witter J. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain. 2003 Dec;106(3):337-345.

Appendix A: Search Strategy

PubMed

Measurement properties filter:

(instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR

IRT[tiab] OR Rasch[tiab] OR “Differential item functioning”[tiab] OR DIF[tiab] OR “computer adaptive testing”[tiab] OR “item bank”[tiab] OR “cross-cultural equivalence”[tiab])

AND

("Central sensitization") OR ("central pain") OR ("centralized pain") OR ("nociplastic pain") OR ("pain sensitization") OR ("pain hypersensitivity")OR ("central hyperexcitability") OR ("heightened pain sensitivity")

AND

("chronic pain") OR ("chronic primary pain condition")) OR ("central pain syndrome")) OR ("central sensitivity syndrome")) OR ("central sensitization syndrome")) OR ("Functional somatic syndrome")) OR ("chronic pain condition")) OR ("chronic musculoskeletal pain")

AND

("patient reported") OR ("self reported")) OR ("self administered")) OR ("patient administered")) OR ("measure")

AND

("Questionnaire") OR ("tool")) OR ("scale")) OR ("instrument")) OR ("measurement")) OR ("Assessment")) OR ("index") OR ("central sensitization inventory") OR ("CSI")) OR ("Pain sensitivity questionnaire")) OR ("PSQ")) OR ("Fibromyalgia survey questionnaire"))

Exclusion filter

(“addresses”[Publication Type] OR “biography”[Publication Type] OR “case reports”[Publication Type] OR “comment”[Publication Type] OR “directory”[Publication Type] OR “editorial”[Publication Type] OR “festschrift”[Publication Type] OR “interview”[Publication Type] OR “lectures”[Publication Type] OR “legal cases”[Publication Type] OR “legislation”[Publication Type] OR “letter”[Publication Type] OR “news”[Publication Type] OR “newspaper article”[Publication Type] OR “patient education handout”[Publication Type] OR “popular works”[Publication Type] OR “congresses”[Publication Type] OR “consensus development conference”[Publication Type] OR “consensus development conference, nih”[Publication Type] OR “practice guideline”[Publication Type]) NOT (“animals”[MeSH Terms] NOT “humans”[MeSH Terms])

MEDLINE (Ovid)

(Central sensitization* or centralized pain* or central pain* or nociplastic pain* or pain sensitization* or pain hypersensitivity* or central hyperexcitability* or heightened pain sensitivity*).tw,kf.

(chronic pain* or chronic primary pain* or central pain syndrome* or central sensitivity syndrome* or central sensitization syndrome* or Functional somatic syndrome* or chronic pain condition* or chronic musculoskeletal pain*).tw,kf.

(patient reported or self reported or self administered or patient administered or measure).tw,kf.

(Questionnaire or tool or scale or instrument or measurement or index).tw,kf.

(central sensitization inventory or CSI).tw,kf.

"pain sensitivity questionnaire*".tw,kf.

"Fibromyalgia survey questionnaire*".tw,kf.

(Widespread Pain Index and Symptom Severity Scale).tw,kf.

"Pain instruments*".tw,kf.

Pain Measurement*.mp.

Pain Assessment*.mp.

Pain scale*.mp.

(Pain adj3 tool*).tw,kf.

(pain adj3 questionnaires*).tw,kf.

(checklist adj3 pain*).tw,kf.

(instrumentation or methods).mp.

outcome measure*.mp.

outcome assessment*.tw.

exp Psychometrics/

"psychometr*".tw,kf.

"Measurement properties*".tw,kf.

(clinimetr* or clinometr*).tw,kf.

"Measurement error*".tw,kf.

Internal consistency*.mp.

Reproducibility of result*.mp.

reproducib*.tw.

Reliability*.mp.

(reliab* or unreliab* or valid* or coefficient of variation or coefficient or homogeneity or homogeneous or internal consistency).tw,kf.

(cronbach* and (alpha or alphas)).tw.

validity*.mp.

validation studies*.tw,kf.

comparative study*.mp.

(item and (correlation* or selection* or reduction*)).tw.

(precision or imprecision or precise values).mp.

(test-retest or (test and retest) or (reliab* and (test or retest))).tw,kf.

(interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intra-observer).tw,kf.

(intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or inter-assay or intraassay or intra-assay).tw,kf.

(interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant).tw.

(kappa or kappas).tw,kf.

repeatab*.mp.

((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).tw.

hypothesis testing*.mp.

Generalizability*.mp.

(generaliza* or generalisa*).tw.

(intraclass and correlation*).tw.

(factor analysis or factor analyses or factor structure or factor structures).tw.

(dimension or subscale).mp.

(multitrait and scaling and (analysis or analyses)).tw.

(item discriminant or interscale correlation).mp.

interscale correlation*.tw.

(error or errors).tw.

(individual variability or interval variability).tw.

rate variability.tw.

(variability and (analysis or values)).tw.

(uncertainty and (measurement or measuring)).tw.

standard error of measurement.tw.

(sensitiv or responsive).tw.

(limit and detection).mp.

minimal detectable concentration.tw.

interpretab*.tw.

((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).tw.

(small* and (real or detectable) and (change or difference)).tw.

meaningful change.tw.

(ceiling effect or floor effect).tw.

item response model.tw.

IRT.tw.

rasch.tw.

differential item functioning.tw.

DIF.tw.

computer adaptive testing.tw.

item bank.tw.

cross-cultural equivalence.tw.

EMBASE

('Validation study' or 'feasibility study' or 'intermethod comparison' or 'data collection method' or 'psychometry' or 'reproducibility' or reproducib*:ab,ti or 'audit':ab,ti or psychometr*:ab,ti or clinimetr*:ab,ti or clinometr*:ab,ti or 'observer variatio' or 'observer variation':ab,ti or 'discriminant analysis' or 'validity' or reliab*:ab,ti or valid*:ab,ti or 'coefficient':ab,ti or 'internal consistency':ab,ti or (cronbach*:ab,ti and ('alpha':ab,ti or 'alphas':ab,ti)) or 'item correlation':ab,ti or 'item correlations':ab,ti or 'item selection':ab,ti or 'item selections':ab,ti or 'item reduction':ab,ti or 'item reductions':ab,ti or 'agreement':ab,ti or 'precision':ab,ti or 'imprecision':ab,ti or 'precise values':ab,ti or 'test-retest':ab,ti or ('test':ab,ti and 'retest':ab,ti) or (reliab*:ab,ti and ('test':ab,ti or 'retest':ab,ti)) or 'stability':ab,ti or 'interrater':ab,ti or 'inter-rater':ab,ti or 'intrarater':ab,ti or 'intra-rater':ab,ti or 'intertester':ab,ti or 'inter-tester':ab,ti or 'intratester':ab,ti or 'interobeserver':ab,ti or 'inter-observer':ab,ti or 'intraobserver':ab,ti or 'intertechnician':ab,ti or 'inter-technician':ab,ti or 'intratechnician':ab,ti or 'interexaminer':ab,ti or 'inter-examiner':ab,ti or 'intraexaminer':ab,ti or 'interassay':ab,ti or 'inter-assay':ab,ti or 'intraassay':ab,ti or 'intra-assay':ab,ti or 'interindividual':ab,ti or 'inter-individual':ab,ti or 'intraindividual':ab,ti or 'intra-individual':ab,ti or 'interparticipant':ab,ti or 'interparticipant':ab,ti or 'intraparticipant':ab,ti or 'kappa':ab,ti or 'kappas':ab,ti or 'coefficient of variation':ab,ti or repeatab*:ab,ti or ((replicab*:ab,ti or 'repeated':ab,ti) and ('measure':ab,ti or 'measures':ab,ti or 'findings':ab,ti or 'result':ab,ti or 'results':ab,ti or 'test':ab,ti or 'tests':ab,ti)) or generaliza*:ab,ti or generalisa*:ab,ti or 'concordance':ab,ti or ('intraclass':ab,ti and correlation*:ab,ti) or 'discriminative':ab,ti or 'known group':ab,ti or 'factor analysis':ab,ti or 'factor analyses':ab,ti or 'factor structure':ab,ti or 'factor structures':ab,ti or 'dimensionality':ab,ti or subscale*:ab,ti or 'multitrait scaling analysis':ab,ti or 'multitrait scaling analyses':ab,ti or 'item discriminant':ab,ti or 'interscale correlation':ab,ti or 'interscale correlations':ab,ti or (('error':ab,ti or 'errors':ab,ti) and (measure*:ab,ti or correlat*:ab,ti or evaluat*:ab,ti or 'accuracy':ab,ti or 'accurate':ab,ti or 'precision':ab,ti or 'mean':ab,ti)) or 'individual variability':ab,ti or 'interval variability':ab,ti or 'rate variability':ab,ti or 'variability analysis':ab,ti or ('uncertainty':ab,ti and ('measurement':ab,ti or 'measuring':ab,ti)) or 'standard error of measurement':ab,ti or sensitiv*:ab,ti or responsive*:ab,ti or ('limit':ab,ti and 'detection':ab,ti) or 'minimal detectable concentration':ab,ti or interpretab*:ab,ti or (small*:ab,ti and ('real':ab,ti or 'detectable':ab,ti) and ('change':ab,ti or 'difference':ab,ti)) or 'meaningful change':ab,ti or 'minimal important change':ab,ti or 'minimal important difference':ab,ti or 'minimally important change':ab,ti or 'minimally important difference':ab,ti or 'minimal detectable change':ab,ti or 'minimal detectable difference':ab,ti or 'minimally detectable change':ab,ti or 'minimally detectable difference':ab,ti or 'minimal real change':ab,ti or 'minimal real difference':ab,ti or 'minimally real change':ab,ti or 'minimally real difference':ab,ti or 'ceiling effect':ab,ti or 'floor effect':ab,ti or 'item response model':ab,ti or 'irt':ab,ti or 'rasch':ab,ti or 'differential item functioning':ab,ti or 'dif':ab,ti or 'computer adaptive testing':ab,ti or 'item bank':ab,ti or 'cross-cultural equivalence':ab,ti

AND

("chronic pain condition*" or "central sensitivity syndrome*" or "central pain syndrome*" or "chronic primary pain condition*" or "chronic pain*" or "chronic overlapping pain condition*" or "Functional somatic syndrome*" or "chronic musculoskeletal pain*").mp.

AND

("Central sensitization*" or "Central sensitisation*" or "centralized pain*" or "central pain*" or "nociplastic pain*" or "pain sensitization*" or "pain sensitivity*" or "pain hypersensitization*" or "central hyperexcitability*").mp.

AND

("pain assessment*" or "pain measurement*" or "pain rating*" or "disability*" or "quality of life*" or "outcome assessment*" or "quality-of-life*" or "functional assessment*" or "self report*" or "self evaluation*" or "self assessment*" or "self-assessment*" or "self-report*" or "self-reported*" or "self reported*" or "patient reported outcome*").mp.

OR

("Questionnaire*" or "tool*" or "scale*" or "instrument*" or "measurement*" or "index*").mp.

OR

("central sensitization inventory*" or "CSI*").mp.

OR

"pain sensitivity questionnaire*".mp.

OR

"Fibromyalgia survey questionnaire*".mp.

OR

"Widespread Pain Index and Symptom Severity Scale*".mp.

Scopus

TITLE-ABS-KEY(psychometr*) OR TITLE-ABS-KEY(clinimetr*) OR TITLE-ABS-KEY(clinometr*) OR TITLE-ABS-KEY('observer') OR TITLE-ABS-KEY('observer variation') OR TITLE-ABS-KEY(reliab*) OR TITLE-ABS-KEY(valid*) OR TITLE-ABS-KEY('imprecision') OR TITLE-ABS-KEY('precise values') OR TITLE-ABS-KEY('coefficient') OR TITLE-ABS-KEY(reproducib*) OR TITLE-ABS-KEY('internal consistency') OR TITLE-ABS-KEY((cronbach*) AND ('alpha' OR 'alphas')) OR TITLE-ABS-KEY ('item correlation') OR TITLE-ABS-KEY('item correlations') OR TITLE-ABS-KEY('item selection') OR TITLE-ABS-KEY('item selections') OR TITLE-ABS-KEY('item reduction') OR TITLE-ABS-KEY('item reductions') OR TITLE-ABS-KEY('agreement') OR TITLE-ABS-KEY('precision') OR TITLE-ABS-KEY('test-retest') OR TITLE-ABS-KEY('test' AND 'retest') OR TITLE-ABS-KEY(reliab*) AND ('test' OR 'retest') OR TITLE-ABS-KEY('stability') OR TITLE-ABS-KEY('interrater') OR TITLE-ABS-KEY('inter-rater') OR TITLE-ABS-KEY('intrarater') OR TITLE-ABS-KEY('intra-rater') OR TITLE-ABS-KEY('intertester') OR TITLE-ABS-KEY('inter-tester') OR TITLE-ABS-KEY('intratester')

OR TITLE-ABS-KEY('intratester') OR TITLE-ABS- KEY('interobserver') OR TITLE-ABS-KEY('intertechnician') OR TITLE-ABS-KEY('inter-technician') OR TITLE-ABS-KEY('intratechnician') OR TITLE-ABS-KEY('intratechnician') OR TITLE-ABS-KEY('inter-observer') OR TITLE-ABS-KEY('intraobserver') OR TITLE-ABS-KEY('intraobserver') OR TITLE-ABS-KEY('interexaminer') OR TITLE-ABS-KEY('inter-examiner') OR TITLE-ABS-KEY('intraexaminer') OR TITLE-ABS-KEY('intraexaminer') OR TITLE-ABS-KEY('interassay') OR TITLE-ABS-KEY('inter-assay') OR TITLE-ABS-KEY('intraassay') OR TITLE-ABS-KEY('intra-assay') OR TITLE-ABS-KEY('interindividual') OR TITLE-ABS-KEY('interindividual') OR TITLE-ABS-KEY('intraindividual') OR TITLE-ABS-KEY('intra-individual') OR TITLE-ABS-KEY('interparticipant') OR TITLE-ABS-KEY('inter-participant') OR TITLE-ABS-KEY('intraparticipant') OR TITLE-ABS-KEY('intraparticipant') OR TITLE-ABS-KEY('kappa') OR TITLE-ABS-KEY('kappas') OR TITLE-ABS-KEY('coefficient of variation') OR TITLE-ABS-KEY(repeatab*) OR TITLE-ABS-KEY((replicab* OR 'repeated') AND ('measure' OR 'measures' OR 'findings' OR 'result' OR 'results' OR 'test' OR 'tests')) OR TITLE-ABS-KEY(generaliza* OR generalisa*) OR TITLE-ABS-KEY('concordance') OR TITLE-ABS-KEY('intraclass' AND correlation) OR TITLE-ABS-KEY('discriminative') OR TITLE-ABS-KEY('known group') OR TITLE-ABS-KEY('factor analysis') OR TITLE-ABS-KEY('factor analyses') OR TITLE-ABS-KEY('factor structure') OR TITLE-ABS-KEY('factor structures') OR TITLE-ABS-KEY('dimensionality') OR TITLE-ABS-KEY(subscale*) OR TITLE-ABS-KEY('multitrait scaling analysis') OR TITLE-ABS-KEY('multitrait scaling analyses') OR TITLE-ABS-KEY('item discriminant') OR TITLE-ABS-KEY('interscale correlation') OR TITLE-ABS-KEY('interscale correlations') OR TITLE-ABS-KEY(('error' OR 'errors') AND (measure OR correlat* OR evaluat* OR 'accuracy' OR 'accurate' OR 'precision' OR 'mean')) OR TITLE-ABS-KEY('individual variability') OR TITLE-ABS-KEY('interval variability') OR TITLE-ABS-KEY('rate variability') OR TITLE-ABS-KEY('variability analysis') OR TITLE-ABS-KEY('uncertainty' AND ('measurement' OR 'measuring')) OR TITLE-ABS-KEY('standard error of measurement') OR TITLE-ABS-KEY(sensitiv*) OR TITLE-ABS-KEY(responsive*) OR TITLE-ABS-KEY('limit' AND 'detection') OR TITLE-ABS-KEY('minimal detectable concentration') OR TITLE-ABS-KEY(interpretab*) OR TITLE-ABS-KEY(small* AND ('real' OR 'detectable') AND ('change' OR 'difference')) OR TITLE-ABS-KEY('meaningful change') OR TITLE-ABS-KEY('minimal important change') OR TITLE-ABS-KEY('minimal important difference') OR TITLE-ABS-KEY('minimally important change') OR TITLE-ABSKEY('minimally important difference') OR TITLE-ABS-KEY('minimal detectable change') OR TITLE-ABS-KEY('minimal detectable difference') OR TITLE-ABS-KEY('minimally detectable change') OR TITLE-ABS-KEY('minimally detectable difference') OR TITLE-ABS-KEY('minimal real change') OR TITLE-ABS-KEY('minimal real difference') OR TITLE-ABS-KEY('minimally real change') OR TITLE-ABS-KEY('minimally real difference') OR TITLE-ABS-KEY('ceiling effect') OR TITLE-ABS-KEY('floor effect') OR TITLE-ABS-KEY('item response model') OR TITLE-ABS-KEY('irt') OR TITLE-ABS-KEY('rasch') OR TITLE-ABS-KEY('differential item functioning') OR TITLE-ABS-KEY('dif') OR TITLE-ABS-KEY('computer adaptive testing') OR TITLE-ABS-KEY('item bank') OR TITLE-ABSKEY('cross-cultural equivalence')

AND

TITLE-ABS-KEY ("chronic pain condition*" OR "central sensitivity syndrome*" OR "central pain syndrome*" OR "chronic primary pain condition*" OR "chronic pain*" OR "chronic overlapping pain condition*" OR "Functional somatic syndrome*" OR "chronic musculoskeletal pain*")

AND

TITLE-ABS-KEY ("Central sensitization*" OR "Central sensitisation*" OR "centralized pain*" OR "central pain*" or "nociplastic pain*" OR "pain sensitization*" OR "pain sensitivity*" OR "pain hypersensitization*" OR "central hyperexcitability*")

AND

TITLE-ABS-KEY ("Questionnaire*" OR "tool*" OR "scale*" OR "instrument*" or "measurement*" OR "index*" OR "central sensitization inventory*" OR "CSI*" OR "pain sensitivity questionnaire*" OR "Fibromyalgia survey questionnaire*" OR "Widespread Pain Index and Symptom Severity Scale*")

Web of Science

TS=(("Central sensitization*" OR "Central sensitisation*" OR "centralized pain*" OR "central pain*" or "nociplastic pain*" OR "pain sensitization*" OR "pain sensitivity*" OR "pain hypersensitization*" OR "central hyperexcitability*")))

AND

TS=((chronic pain OR central pain syndrome OR chronic primary pain condition OR central sensitivity syndrome OR chronic pain condition OR chronic overlapping pain condition OR Functional somatic syndrome OR chronic musculoskeletal pain)))

AND

TS=((Questionnaire OR tool OR scale OR instrument OR measurement OR index OR central sensitization inventory OR CSI OR pain sensitivity questionnaire OR Fibromyalgia survey questionnaire OR Widespread Pain Index and Symptom Severity Scale))

AND

(TS="psychometric properties" OR TS="measurement properties" OR TS=psychometr* OR TS="outcome assessment" OR TS= "validation studies" OR TS="observer variation" OR TS=reproducib* OR TS=reliab* OR TS=unreliab* OR TS=valid* OR TS=coefficient OR TS=homogeneity OR TS=homogeneous OR TS="internal consistency" OR TS=(cronbach* AND (alpha OR alphas)) OR TS=(item AND (correlation* OR selection* OR reduction*)) OR TS=precision OR TS=imprecision OR TS=(precise values) OR TS=test-retest OR TS=(test AND retest) OR TS=(reliab* AND (test OR retest)) OR TS=stability OR TS=interrater OR TS=inter-rater OR TS=intrarater OR TS=intra-rater OR TS=intertester OR TS=inter-tester OR TS=intratester OR TS=intra-tester OR TS=interobserver OR TS=inter-observer OR TS=intraobserver OR TS=intra-observer OR TS=intertechnician OR TS=intertechnician OR TS=intratechnician OR TS=intra-technician OR TS=interexaminer OR TS=interexaminer OR TS=intraexaminer OR TS=intra-examiner OR TS=interassay OR TS=inter-assay OR TS=intraassay OR TS=intra-assay OR TS=interindividual OR TS=inter-

individual OR TS=intraindividual OR TS=intra-individual OR TS=interparticipant OR TS=inter-participant OR TS=intraparticipant OR TS=intra-participant OR TS=kappa OR TS=kappa's OR TS=kappas OR TS=repeatab* OR TS=((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR TS=generaliza* OR TS=generalisa* OR TS=concordance OR TS=(intraclass AND correlation*) OR TS=discriminative OR TS=(known group) OR TS="factor analysis" OR TS="factor analyses" OR TS=dimension* OR TS=subscale* OR TS=(multitrait AND scaling AND (analysis OR analyses)) OR TS="item discriminant" OR TS="interscale correlation*" OR TS=(error OR errors) OR TS="individual variability" OR TS=(variability AND (analysis OR values)) OR TS=(uncertainty AND (measurement OR measuring)) OR TS="standard error" OR TS="of measurement" OR TS=sensitiv* OR TS=responsive* OR TS=((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR TS=(small* AND (real OR detectable) AND (change OR difference)) OR TS="meaningful change" OR TS="minimal important change" OR TS="minimal important difference" OR TS="minimally important change" OR TS="minimally important difference" OR TS="minimal detectable change" OR TS="minimal detectable difference" OR TS="minimally detectable change" OR TS="minimally detectable difference" OR TS="minimal real change" OR TS="minimal real difference" OR TS="minimally real change" OR TS="minimally real difference" OR TS="ceiling effect" OR TS="floor effect" OR TS="Item response model" OR TS=IRT OR TS=Rasch OR TS="Differential item functioning" OR TS=DIF OR TS="computer adaptive testing" OR TS="item bank" OR TS="cross-cultural equivalence")

CINAHL

(TI psychometr* OR TI observer variation OR TI reproducib* OR TI reliab* OR TI unreliab* OR TI valid* OR TI coefficient OR TI homogeneity OR TI homogeneous OR TI "internal consistency" OR AB psychometr* OR AB observer variation OR AB reproducib* OR AB reliab* OR AB unreliab* OR AB valid* OR AB coefficient OR AB homogeneity OR AB homogeneous OR AB "internal consistency" OR (TI cronbach* OR AB cronbach* AND (TI alpha OR AB alpha OR TI alphas OR AB alphas)) OR (TI item OR AB item AND (TI correlation* OR AB correlation* OR TI selection* OR AB selection* OR TI reduction* OR AB reduction*)) OR TI agreement OR TI precision OR TI imprecision OR TI "precise values" OR TI test-retest OR AB agreement OR AB precision OR AB imprecision OR AB "precise values" OR AB test-retest OR (TI test OR AB test AND TI retest OR AB retest) OR (TI reliab* OR AB reliab* AND (TI test OR AB test OR TI retest or AB retest)) OR TI stability OR TI interrater OR TI interrater OR TI intrarater OR TI intra-rater OR TI intertester OR TI inter-tester OR TI intratester OR TI intra-tester OR TI interobserver OR TI inter-observer OR TI intraobserver OR TI intra-observer OR TI intertechnician OR TI inter-technician OR TI intratechnician OR TI intra-technician OR TI interexaminer OR TI inter-examiner OR TI intraexaminer OR TI intraexaminer OR TI interassay OR TI inter-assay OR TI intraassay OR TI intra-assay OR TI interindividual OR TI inter-individual OR TI

intraindividual OR TI intra-individual OR TI interparticipant OR TI inter-participant OR TI intraparticipant OR TI intra-participant OR TI kappa OR TI kappa's OR TI kappas OR TI repeatab* OR AB stability OR AB interrater OR AB inter-rater OR AB intrarater OR AB intra-rater OR AB intertester OR AB inter-tester OR AB intratester OR AB intra-tester OR AB interobserver OR AB inter-observer OR AB intraobserver OR AB intra-observer OR AB intertechnician OR AB inter-technician OR AB intratechnician OR AB intra-technician OR AB interexaminer OR AB inter-examiner OR AB intraexaminer OR AB intra-examiner OR AB interassay OR AB inter-assay OR AB intraassay OR AB intra-assay OR AB interindividual OR AB inter-individual OR AB intraindividual OR AB intra-individual OR AB interparticipant OR AB inter-participant OR AB intraparticipant OR AB intra-participant OR AB kappa OR AB kappa's OR AB kappas OR AB repeatab* OR ((TI replicab* OR AB replicab* OR TI repeated OR AB repeated) AND (TI measure OR AB measure OR TI measures OR AB measures OR TI findings OR AB findings OR TI result OR AB result OR TI results OR AB results OR TI test OR AB test OR TI tests OR AB tests)) OR TI generaliza* OR TI generalisa* OR TI concordance OR AB generaliza* OR AB generalisa* OR AB concordance OR (TI intraclass OR AB intraclass AND TI correlation* or AB correlation*) OR TI discriminative OR TI "known group" OR TI factor analysis OR TI factor analyses OR TI dimension* OR TI subscale* OR AB discriminative OR AB "known group" OR AB factor analysis OR AB factor analyses OR AB dimension* OR AB subscale* OR (TI multitrait OR AB multitrait AND TI scaling OR AB scaling AND (TI analysis OR AB analysis OR TI analyses OR AB analyses)) OR TI item discriminant OR TI interscale correlation* OR TI error OR TI errors OR TI "individual variability" OR AB item discriminant OR AB interscale correlation* OR AB error OR AB errors OR AB "individual variability" OR (TI variability OR AB variability AND (TI analysis OR AB analysis OR TI values OR AB values)) OR (TI uncertainty OR AB uncertainty AND (TI measurement OR AB measurement OR TI measuring OR AB measuring)) OR TI "standard error of measurement" OR TI sensitiv* OR TI responsive* OR AB "standard error of measurement" OR AB sensitiv* OR AB responsive* OR ((TI minimal OR TI minimally OR TI clinical OR TI clinically OR AB minimal OR AB minimally OR AB clinical OR AB clinically) AND (TI important OR TI significant OR TI detectable OR AB important OR AB significant OR AB detectable) AND (TI change OR AB change OR TI difference OR AB difference)) OR (TI small* OR AB small* AND (TI real OR AB real OR TI detectable OR AB detectable) AND (TI change OR AB change OR TI difference OR AB difference)) OR TI meaningful change OR TI "ceiling effect" OR TI "floor effect" OR TI "Item response model" OR TI IRT OR TI Rasch OR TI "Differential item functioning" OR TI DIF OR TI "computer adaptive testing" OR TI "item bank" OR TI "cross-cultural equivalence" OR TI outcome assessment OR AB meaningful change OR AB "ceiling effect" OR AB "floor effect" OR AB "Item response model" OR AB IRT OR AB Rasch OR AB "Differential item functioning" OR AB DIF OR AB "computer adaptive testing" OR AB "item bank" OR AB "cross-cultural equivalence" OR AB outcome assessment)

AND

TX(Central sensitization OR centralized pain OR central pain OR nociplastic pain OR pain sensitization OR pain sensitivity OR pain hypersensitization OR central hyperexcitability

AND

TX (chronic pain OR central pain syndrome OR chronic primary pain condition OR central sensitivity syndrome OR chronic pain condition OR chronic overlapping pain condition OR Functional somatic syndrome OR chronic musculoskeletal pain)

AND

TX (Questionnaire OR tool OR scale OR instrument OR measurement OR index OR central sensitization inventory OR CSI OR pain sensitivity questionnaire OR Fibromyalgia survey questionnaire OR Widespread Pain Index and Symptom Severity Scale)

Appendix B: “Do file”- Codes for meta-analysis

```
generate z = atanh(r) // r-to-z = inverse hyperbolic tangent
generate sez = sqrt(1/(n - 3))

*admetan z sez

metan z sez, label(namevar = study)


display _newline ///
" Pooled estimate of r = " tanh(r(eff)) _newline ///
"Lower Limit of 95% CI = " tanh(r(eff) - (1.96 * r(se_eff))) _newline ///
"Upper Limit of 95% CI = " tanh(r(eff) + (1.96 * r(se_eff)))


// Prepare data for -forestplot-

generate _USE = 1

// Generate CI's for r
gen effect = tanh(r(eff))
generate lb = tanh(_LCI)
generate ub = tanh(_UCI)


label var n "Sample size"

*Restricted maximum liklihood method

metan r lb ub, label(namevar = study) random(reml)
```

Web link: Do File Codes for meta-analysis(reliability estimate).do

Curriculum Vitae

Name: Mst Farjana Akhter

Post-secondary Education and Degrees:

Master of Science, Epidemiology and Biostatistics (Candidate), 2024
Western University, London, Ontario

Doctor of Medicine (MD), Pharmacology and Therapeutics
Bangabandhu Sheikh Mujib Medical University, Bangladesh
2015-2018

Bachelor of Medicine & Bachelor of Surgery (MBBS)
Rangpur Medical College & Hospital, Bangladesh
Passing year:2010
Internship- 2011

Related Work experiences:

Graduate Teaching Assistant (2024)
Department of Epidemiology and Biostatistics,
Western University, London, Ontario

Research Co-assistant
Western University, London, Ontario

Worked as a Peer reviewer of 'The Asian Journal of Medicine and Health'.

Assistant Professor (May 2018 – August 2022)
Department of Pharmacology,
Green Life Medical College, Bangladesh

Lecturer (January 2012 - February 2015)
Department of Physiology,
Ibn Sina Medical College, Bangladesh

Guest lecturer (2015 -2018)
Nursing institution of BSMMU, Bangladesh

Publications:

1. Akhter F, Dewan Z, Hasnat M, Akhter S. Levocarnitine in the management of fatigue in levothyroxine-treated hypothyroid patients. *IMC Journal of Medical Science*, 2019; 13(2), 45-52.

2. Akhter F, Dewan Z, Salma U, Roy J. Effect of Levocarnitine in the management of Dyslipidemia in Hypothyroid patients. *GMC Journal of Medical Science 2020;* 5(1):8-14.

3. Salma U, Sultana S, Akhter F, Sarker M, Nishi MK. Antimicrobial activity of aqueous extracts of Cinnamomum zeylanicum bark against Staphylococcus aureus and Escherichia coli. *GMC Journal of Medical Science* 2020; 5(1):31-35.

4. Roy J, Dewan ZF, Dey BP, Akhter F, Haider Z, Das B, Basak PK, Barman BK, Bhowmik MC, Jebunnahar. Effect of Ethanol Extract of Musa sapientum Flowers on Alloxan induced Type 2 Diabetes Mellitus in the Rat. *Journal of Comilla Medical College Teachers' Association* 2021; 25(2):51-57.