Clinical outcome assessment in clinical trials of chronic pain treatments

Kushang V. Patel[a],[*], Dagmar Amtmann[b], Mark P. Jensen[b], Shannon M. Smith[c], Christin Veasley[d], Dennis C. Turk[a]

Abstract

Clinical outcome assessments (COAs) measure outcomes that are meaningful to patients in clinical trials and are critical for determining whether a treatment is effective. The objectives of this study are to (1) describe the different types of COAs and provide an overview of key considerations for evaluating COAs, (2) review COAs and other outcome measures for chronic pain treatments that are recommended by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) or other expert groups, and (3) review advances in understanding pain-related COAs that are relevant to clinical trials. The authors reviewed relevant articles, chapters, and guidance documents from the European Medicines Agency and U.S. Food and Drug Administration. Since the original core set of outcome measures were recommended by IMMPACT 14 years ago, several new advancements and publications relevant to the measurement or interpretation of COAs for chronic pain trials have emerged, presenting new research opportunities. Despite progress in the quality of measurement of several outcome domains for clinical trials of chronic pain, there remain some measurement challenges that require further methodological investigation.

Keywords: Clinical trials, Chronic pain, Outcome assessment, Patient-reported outcomes

1. Introduction

Clinical outcome assessments (COAs) measure outcomes that are meaningful to patients in clinical trials and are critical for determining whether a treatment is effective. Unlike a biomarker (eg, hemoglobin A_{1C}), COAs measure how a patient feels or the impact of a health condition on how the patient functions in daily life.[145,155] Understanding the benefits and risks of a treatment and how regulatory agencies decide on labeling treatment claims depends, in part, on the characteristics of the COA used in clinical trials. For chronic pain treatments, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended 6 core outcome domains that should be considered when designing clinical trials: (1) pain, (2) physical functioning, (3) emotional functioning, (4) participant ratings of improvement and satisfaction with treatment, (5) symptoms and adverse events (AEs), and (6) participant disposition.[137] The systematic collection and reporting of outcomes in these domains are valuable for adjudicating the efficacy of treatments and comparing results across trials, although not all of these domains are always measured with COAs per se. The objectives of this review article are to (1) describe the different types of COAs and provide an overview of key considerations for evaluating COAs, (2) review COAs and other outcome measures for chronic pain treatments that are recommended by IMMPACT or other expert groups for phase 2 and 3 clinical trials, and (3) review advances in understanding pain-related COAs that are relevant to clinical trials. In addition to relevant book chapters and peer-reviewed publications identified through PubMed and Google Scholar searches, guidance documents from the European Medicines Agency and the U.S. Food and Drug Administration (FDA) were reviewed because they aim to integrate best practices for patient-centered outcome assessment. Furthermore, the bibliographies of key publications were examined along with citation searches to identify additional studies. Note that the target patient populations for the selected set of outcome measures reviewed herein are adults with chronic pain conditions; however, the article by Palermo et al.[102] included in this supplement provides a comprehensive review of trial design considerations for pediatric patient populations. Outcome assessment in adults with cognitive disorders, such as developmental disabilities or dementia, will not be reviewed, although pain assessment in older adults with dementia has been reviewed previously.[64] The reader is referred to that review for information about the COA tools currently available to assess pain in this population.

2. Types of clinical outcome assessments

There are 4 major types of COAs: patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), observer-reported outcomes (ObsROs), and performance-based outcomes (PerfOs).[144,145] **Table 1** presents the FDA's definition for each COA type. The first 3 types differ according to who is rating the concept that is being measured. A PRO records a research participant's response to questions about their own health or functioning that is independent of a clinician's or anyone else's judgement. This definition is similar to the European Medicines Agency description of PROs.[44,45] By contrast, ClinROs rely on the judgment of a trained professional about the health or functioning of a patient based on observation or examination. Observer-reported outcomes do not involve direct reporting from either patients or clinicians, but rather are observations of patients made by a person who generally does not have specialized clinical training (eg, family member or caregiver). This person can, however, receive training on what to observe. In contrast to the COAs that rely on a rater's judgement, PerfOs measure a patient's performance of a discrete task following an established protocol (eg, 6-minute walk test). Although PerfOs usually involve trained professional staff who systematically record a patient's or study participant's performance, they do not incorporate the assessor's judgement or interpretation of performance. Importantly, outcome measures that are assessed with mobile health technologies (eg, wearable physical activity or sleep monitors) do not necessarily fit within the 4 major COA types as currently defined by the FDA,[144] although they do provide a unique source of information that may be important in the evaluation of treatment effects (see section 4.2.4); whether such outcome approaches are ultimately classified as a subtype of ObsRO or will be classified as a distinct COA type remains to be seen.

3. Considerations for evaluating clinical outcome assessments

The selection of COAs as primary or secondary endpoints should be based on the anticipated effects of the intervention (eg, drug or nonpharmacological treatment) on specific aspects of how a patient feels or functions in daily life.[155] The primary endpoint is used to determine treatment efficacy and/or safety, whereas the purpose of secondary endpoints is to enhance the interpretation of the primary endpoint result and/or to provide evidence for a specific mechanism underlying the treatment effect.[148] Therefore, the measurement properties of COAs play a critical role in not only detecting a treatment effect but also in establishing treatment benefit for improving meaningful aspects of the study participant's health. The key measurement properties for evaluating COAs include content validity, reliability, construct validity, responsiveness, and interpretability. A comprehensive discussion of these properties is beyond the scope of this article, but each are reviewed briefly below and have been discussed at length in numerous publications.[25,26,87,96,97,107,108,114,116,138]

3.1. Content validity

Content validity is the degree to which the COA measures the concept of interest (ie, the domain or factor that the COA is purported to measure).[96,147] Evidence for content validity is based primarily on both expert opinion and qualitative research engaging patients from the target population. A critical component of content validity is ensuring that the COA comprehensively

Definitions of clinical outcome assessments.[144]

COA type	Definition
Patient-reported outcome	A PRO is a measurement based on a report that comes from the patient (ie, study subject) about the status of a patient's health condition without amendment or interpretation of the patient's report by a clinician or anyone else. A PRO can be measured by self-report or by interview, provided that the interviewer records only the patient's response. Symptoms or other unobservable concepts known only to the patient (eg, pain severity or nausea) can only be measured by PRO measures. PROs can also assess the patient perspective on functioning or activities that may also be observable by others.
Performance outcome	A PerfO is a measurement based on a task(s) performed by a patient according to instructions that is administered by a health care professional. Performance outcomes require patient cooperation and motivation.
Clinician-reported outcome	A ClinRO is based on a report that comes from a trained health care professional after observation of a patient's health condition. A ClinRO measure involves a clinical judgment or interpretation of the observable signs, behaviors, or other physical manifestations thought to be related to a disease or condition. ClinRO measures cannot directly assess symptoms that are known only to the patient (eg, pain intensity).
Observer-reported outcome	An ObsRO is a measurement based on an observation by someone other than the patient or a health professional. This may be a parent, spouse, or other nonclinical caregiver who is in a position to regularly observe and report on a specific aspect of the patient's health. An ObsRO measure does not include medical judgment or interpretation.

COA, clinical outcome assessments; PROs, patient-reported outcomes.

covers the entire range of health issues (eg, symptoms or functional limitations) experienced by the patient population that is relevant to the concept of interest. Generally, content validity should be established before evaluating other measurement properties. Expert methodologists have published detailed guides on establishing the content validity of existing or newly developed PROs that can be adapted for other COA types.[87,107,108,120]

3.2. Reliability

Reliability is the extent to which the COA is free of measurement error.[96] In clinical trials, the reproducibility of a treatment effect depends, in part, on the reliability of the COA. Test–retest reliability is critical for PROs and PerfOs, whereas intrarater reliability and interrater reliability are particularly relevant for ClinROs and ObsROs. For test–retest reliability, the period between assessments should be carefully considered in the context of the population being studied and measurement construct—a prolonged time interval will increase the probability that a study participant might experience a real change in their health status or functioning, whereas too short an interval can result in participants simply recalling their initial (baseline) responses. For COAs with an interval scale of measurement, the intraclass correlation coefficient (ICC) is a statistic used to measure the degree of agreement between either repeated administrations (test–retest reliability) or multiple observers/raters (interrater reliability) of an outcome measure in the same set of

patients. The ICC can also be used to estimate intrarater reliability. By contrast, the kappa statistic is commonly used to measure agreement (test–retest, interrater, and intrarater reliability) for COAs with a nominal measurement scale. In addition to the reproducibility aspect of reliability, the internal consistency of PROs with multiple questionnaire items is also an important consideration because it reflects the interrelatedness (correlations) among the items, providing an indication of the extent to which the PRO items measure a single, latent (unobservable) construct.[148] The Cronbach alpha coefficient is the most widely used statistic to assess internal consistency of multi-item PROs.[27]

In addition to the traditional metrics based on classical test theory (CTT) that assess the reliability of the total score from a PRO or other type of COA, the precision of individual measurement items and the total score in distinguishing study participants across different levels of a latent construct is also important. Item response theory (IRT) is a family of statistical models that aim to explain the relationship between study participants' responses to individual items on a measurement scale and the latent construct. The item information derived from IRT models can be summed to provide an overall assessment of the scale's precision (for detailed discussion of IRT, see Ref. 114). Unlike the CTT-based reliability, the IRT-based information (the IRT equivalent of CTT reliability) is estimated at each point along the latent construct, recognizing that the amount of error is typically greater at the extremes of the continuum and providing important information for powering clinical trials.[114]

3.3. Construct validity

Construct validity is the extent to which the COA demonstrates relationships with other measures, outcome domains, or patient characteristics that are consistent with a priori hypotheses.[147] For example, to evaluate the convergent validity (a type of construct validity) of a new PRO of pain intensity among patients with chronic low back pain (CLBP), we would expect greater pain intensity would correlate positively with higher levels of emotional distress and pain-related interference with physical function. Discriminant (or divergent) validity is demonstrated when the COA is less strongly associated with measures of constructs that are theoretically more distal or unrelated. The extent to which a COA of interest correlates with a gold standard measure or legacy measure is criterion validity, a subtype of convergent validity.[147] Known groups validity is another form of construct validity in which comparisons of a COA are made between different groups of patients that are known to differ in an expected way on the concept of interest. Generally, multiple studies are needed to cumulate evidence of construct validity for a COA, requiring time and resources for measurement development.

3.4. Responsiveness

Responsiveness is the ability to detect change over time in the concept of interest.[147] For clinical trials to establish treatment benefit, it is critical for COAs to be able to detect improvement and worsening in the concept of interest across the entire measurement range for a given target patient population.[147] Evidence for responsiveness of a COA can be generated from either clinical trials or observational studies by examining the relationships of change in COA scores with changes in other relevant outcome measures.[116] Sample size calculations for clinical trials depend, in part, on the responsiveness of the COA selected as the primary endpoint.[116,145]

3.5. Interpretability

In chronic pain treatment trials, PROs are the primary type of COA used as pain and other related symptoms are subjective experiences that are generally not known by anyone other than the patient or study participant. Unlike common clinical signs, such as systolic blood pressure, it is often difficult to meaningfully interpret change in PRO score values.[25] In addition, it is widely recognized that a statistically significant between-group difference in PRO scores does not necessarily equate to a clinically meaningful difference.[26,38,41] Furthermore, statistical significance does not indicate the extent to which individual patients experienced clinically meaningful improvement.[25,26,38,41] Therefore, to enrich the interpretation of PRO and other COA endpoint results, appropriate responder definitions should be established that identify clinically meaningful within-person change for the target patient population (it is important to note that clinically important within-person change does not equate to clinically important group difference because larger magnitudes of change are generally required for the former than the latter.[38,41])

Although there are a variety of anchor-based and distribution-based methods for identifying responder definitions, recent guidelines recommend triangulating from multiple methods and analyses (including use of multiple anchors) to derive either a single threshold value or a range of values for interpreting meaningful changes in COA scores.[26] Based on several methodological studies, IMMPACT recommends reporting the proportion of patients with reductions in pain intensity of ≥30% and ≥50% in clinical trials of chronic pain treatment.[41] Because responder analyses play an increasingly important role in regulatory decision-making,[26] more methodological research is needed to define clinically meaningful change of COAs used in chronic pain trials.

4. Review of clinical outcome assessments for chronic pain

As noted earlier in section 1.0, IMMPACT recommends the assessment of 6 core outcome domains for clinical trials of chronic pain treatments.[137] In the subsequent sections, we review the specific COAs recommended by IMMPACT[37] and other expert groups, including the Outcome Measures in Rheumatology (OMERACT), Osteoarthritis Research Society International (OARSI), Neuropathic Pain Special Interest Group (NeuPSIG), National Institutes of Health (NIH) Task Force on Research Standards for Chronic Low Back Pain, and Veterans Health Administration Pain Measures Work Group. Relevant COAs that are increasingly used in the field will also be reviewed.

4.1. Core outcome domain of pain

Treatments for chronic pain can target and affect multiple aspects of the pain experience, including pain intensity, pain quality, and the use of rescue analgesic medications.[36] Given that pain is experienced privately, PROs are generally used to assess pain intensity and quality. Reliable and valid pain assessment is critical for not only determining the effectiveness of chronic pain treatments, but also for diagnosing chronic pain conditions. Indeed, accurate pain assessment is a core component of dimension 1 of the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION)-American Pain Society Pain Taxonomy (AAPT).[48]

4.1.1. Pain intensity

Pain intensity is likely the most commonly assessed aspect of chronic pain in clinical practice and the most commonly

prespecified primary outcome in chronic pain trials.[37] As defined by Jensen and Karoly,[74] pain intensity is "how much a person hurts," signifying the magnitude of pain (by contrast, pain affect is more complex, reflecting the emotional distress associated with the pain experience.[48,74]) In adults without cognitive impairment, pain intensity is typically assessed with either the verbal rating scale (VRS), visual analogue scale (VAS), or numerical rating scale (NRS). These scales may incorporate a variety of time frames (eg, past day, last 7 days) and anchor terms (eg, upper limit equals "worst pain imaginable," "worst pain possible," or "worst pain ever experienced"). Each of each of these types of scales has strengths and weaknesses that should be considered when designing a chronic pain trial.[74]

4.1.1.1. Verbal rating scale for pain intensity

There are multiple versions of the VRS; but, in general, the VRS for pain intensity is a rank-ordered list of words describing different levels of pain that include extreme anchors and intermediate adjectives (eg, no pain, mild pain, moderate pain, and severe pain). Each descriptor is coded with a numerical value for data analysis (eg, 0, 1, 2, and 3 for no pain, mild pain, moderate pain, and severe pain, respectively). Respondents are asked to select the single descriptor that best represents their level of pain. Although the VRS is easy to administer and has demonstrated construct validity,[68,74,123] the scoring method assumes that the VRS has an interval scale of measurement. However, it is unlikely that the magnitude of difference in pain intensity between "no pain" and "mild pain" is the same as the difference between "moderate" and "severe." This can pose a challenge for interpreting change in VRS scores over time.

4.1.1.2. Visual analogue scale for pain intensity

In contrast to the VRS, the VAS has measurement scale properties approaching a ratio scale.[111,113] The VAS asks respondents to mark their level of pain on a line (horizontal or vertical) that is usually 10 centimeters long with ends that are labeled with anchors such as "No pain" and "Worst imaginable pain." The distance in millimeters (mm) from the beginning of the line ("No pain" side) to the marking is the participant's pain level. There are paper-based, mechanical (with a sliding marker), and electronic (with automated data capture) versions of the VAS. Numerous studies have demonstrated the reliability, validity, and responsiveness of the VAS for pain intensity assessment in various chronic pain patient populations[68,74,123]; however, compliance with the VAS is lower than with VRS and NRS, particularly among older adults,[68] potentially reflecting the greater motor, visual, and cognitive demands required for completing the VAS.

4.1.1.3. Numerical rating scale for pain intensity

The NRS is likely the most commonly used measure of pain intensity in clinical settings because it is easy to administer and score. Research participants are asked to rate their pain intensity on a 0 to 10 (or 20 or 100) scale where 0 usually represents "No pain" and 10 represents a descriptor indicating an extreme level of pain (eg, "Worst imaginable pain"). As with the VRS and VAS, there are several versions of the NRS. Some instruct the study participant to verbally report or write down the number that best reflects their pain intensity, whereas others present the numbers or boxes with numbers in ascending order and participants circle/mark their pain level. The measurement properties of the NRS for

pain intensity are robust and well established across multiple patient populations,[68,74,123] although recent research suggests that the NRS may not have utility in populations of individuals with chronic pain from developing countries with low literacy rates.[106] A clear strength of the NRS is its versatility with different modes of data collection, ranging from in-person interviews to telephone calls and interactive voice recording to mobile technology and wearable devices. However, in contrast to the VAS, it is unlikely that the 0 to 10 NRS has a ratio scale of measurement. For instance, a 3-point decline on the 0 to 10 NRS starting from a baseline of 8 probably does not have the same meaning as a 3-point decline starting at 4.

4.1.1.4. Picture or face scales for pain intensity

In addition to the NRS, VAS, and VRS, pain intensity can also be assessed by using drawings or photographs of different facial expressions of pain.[74] Although the image or face scales were originally developed for use in adult populations with impairment in cognitive function or communication,[74] they also have the potential to standardize pain assessment in the general adult population from across different countries and cultures. Indeed, in a recent study of adults with musculoskeletal pain in Nepal,[106] the Faces Pain Scale-Revised (FPS-R)[67] demonstrated adequate construct validity and was the most preferred scale followed by the VRS, VAS, and NRS. This study also highlighted the challenges of assessing pain in a non-Western country with a relatively low literacy rate and limited exposure to rating scales in daily life.[106] For example, approximately a quarter of participants did not respond to the NRS or VAS for pain intensity despite receiving verbal instructions that were repeated up to 3 times, and a third responded to the NRS with a range of pain intensities rather than selecting a single value.[106] In addition, 15% of participants volunteered different adjectives to describe their pain intensity rather than selecting one of the response options from the VRS (translated from English to Nepali).[106] By contrast, the FPS-R had fewer problematic responses than the 3 other measures of pain intensity, suggesting that the FPS-R might be a more appropriate measure for use in multisite trials involving non-Western countries with low literacy. However, further research in such settings is required that evaluates different pain intensity assessments and participant training programs to enhance the reliability and validity of pain intensity scores.

4.1.1.5. Recommendations and advances for pain intensity assessment

Although the 4 types of pain intensity measures are highly correlated with one another,[68] there are tradeoffs to consider when selecting one as a trial endpoint. Based on the findings of a literature review and a consensus meeting of experts, IMMPACT recommended the 0 to 10 NRS to assess pain intensity in chronic pain treatment trials.[37] More recently, the Critical Path's PRO Consortium completed a literature review and has also concluded that the 0 to 10 NRS is currently the most optimal pain intensity measure for clinical trials in adults without cognitive impairment.[123] Despite these recommendations, there remain important challenges to standardized collection and reporting of pain intensity data in clinical trials.

A systematic review by Smith et al.[128] identified underreporting of critical elements of pain intensity assessment in clinical trials. For example, 20% of clinical trials published in leading pain journals between January 2011 and July 2012 had not clearly identified the pain intensity measure used, and 12% had not

reported the anchors for the response options. Interestingly, another systematic review observed that among clinical research studies that used the 0 to 10 NRS, there were 14 different anchors used for the upper end of the scale (eg, "Worst pain imaginable," "Unbearable pain," and "Maximum pain") and 11% of studies had not reported the response anchors.[68] It is unclear to what extent pain intensity measures with different anchors have equivalent measurement properties (reliability, validity, and responsiveness). The review by Smith et al. also identified that only half of the trials reported the type of pain intensity rated (ie, average, least, worst, and current) and a third did not report the period rated (eg, past 12 or 24 hours). Based on these findings, **Table 2** presents recommendations for standardizing the reporting of pain intensity assessments.[128]

In efforts to improve the assay sensitivity of chronic pain trials (ie, the ability to detect a true treatment effect), investigators have sought to better understand and improve the measurement properties of pain intensity outcome measures. For example, in a secondary data analysis of a randomized controlled trial (RCT) of oxymorphone extended-release for CLBP, Jensen et al.[72] sought to determine whether increasing the reliability of the pain intensity assessment (0–100 mm VAS of average pain in the last 24 hours collected in clinic) by increasing the number of ratings in composite outcome measures would increase assay sensitivity. Consistent with psychometric theory and previous research,[76] the internal consistency of the composite measures increased as the number of ratings increased, although the incremental gain was small (the Cronbach alpha increased from 0.94 to 0.96 for composites with 2 and 9 ratings, respectively); however, there was no improvement in assay sensitivity (Cohen's d = 0.57 for composites with 2 and 9 ratings).[72] Indeed, a single pain rating had almost the same assay sensitivity (Cohen's d = 0.52) as the composite measures of multiple ratings.[72] This provocative finding suggested that the collection of multiple ratings may not be necessary, but the authors emphasized the need for replication and additional investigation.[72] Accordingly, Stone et al.[132] retrospectively analyzed data from a clinical trial of self-management for adults with osteoarthritis (OA) that collected 0 to 10 NRS average pain intensity data at home using interactive voice recording (recall period was the day of the call, "today"). Similar to the findings of Jensen et al.,[72] the internal consistency of pain intensity composites increased as the number of ratings increased, but the amount of gain was greater than that observed by Jensen et al. probably because the starting Cronbach alpha was much lower in the OA self-management trial than in the CLBP drug trial. However, unlike Jensen et al., improvements in pain intensity reliability were associated with increased assay sensitivity.[132] Importantly, there are several methodological differences between these studies that can potentially explain differences in the study findings. Accordingly, additional studies are needed to further investigate the effect of increasing the internal consistency (reliability) of pain intensity assessment on the assay sensitivity of chronic pain trials.

More recent efforts by ACTTION to understand and ultimately improve assay sensitivity include the evaluation of composite responder outcomes of pain intensity and physical functioning in clinical trials of neuropathic pain. Patel et al.[103] conducted an individual patient data analysis of 15 RCTs of duloxetine, gabapentin, and pregabalin for diabetic peripheral neuropathy (DPN) and postherpetic neuralgia (PHN). A composite responder outcome of ≥50% reduction in pain intensity, or a ≥20% reduction in pain intensity and ≥30% improvement in physical functioning was validated and cross-validated in different neuropathic pain conditions and treatments. Notably, this responder outcome had slightly lower number needed to treat than a standard responder outcome of ≥50% reduction in pain intensity, and was more favorably associated with patient global ratings of improvement.[103] This initial study exploring composite responder outcomes of pain intensity and physical functioning in clinical trials of DPN and PHN requires replication in different contexts of use (eg, different patient populations and treatments).

In addition to harmonizing individual-level data from RCTs and conducting secondary analyses of outcome measures in chronic pain trials, ACTTION has published several systematic reviews and meta-analyses relevant to the design, analysis, and reporting of chronic pain treatment trials. Most recently, Smith et al.[129] compared the assay sensitivity of change in average pain intensity vs change in worst pain intensity in RCTs of efficacious treatments for several chronic pain conditions. Previous research had suggested that peak (worst) pain and end (current) pain might influence overall (average) pain ratings, indicating that worst pain ratings might provide a more accurate assessment than average pain ratings.[79,115] Furthermore, the FDA's draft guidance on analgesic drug development recommends worst pain intensity ratings as a primary endpoint.[146] However, the meta-regression analyses by Smith et al.[129] did not show any consistent difference in assay sensitivity between average vs worst pain intensity ratings.

Finally, recognizing that patients often have idiosyncratic interpretations of PROs[159] and are only given minimal instruction on how to respond or interpret PRO questionnaire items, ACTTION developed a program to train patients to be more accurate in their pain ratings.[126] The program guided study participants to identify personal anchors for 0 to 10 NRS for pain intensity scale and educated participants on the meaning of "average" pain and its relationship to "least" and "worst" pain. Participants were also instructed to make their pain ratings in a quiet place, free of distractions. The training emphasized that participants should focus only on pain intensity when making their ratings and not on other symptoms (eg, fatigue or stress). In an

---

Table 2

Recommendations for reporting pain intensity assessments.[128]

Report the type of pain intensity assessment used (eg, VAS, NRS) with a description that clearly distinguishes the assessment from others (eg, for a VAS, describing the length of the line; for an NRS, reporting the range of possible ratings)
Report the definitions of anchors, except in cases where a well-known assessment is used verbatim and the anchors are easily referenced (eg, anchors for the 4 BPI pain intensity NRS items)
Report the frequency of administering the pain intensity assessment
Report the period to be rated, except in cases where a well-known assessment is used verbatim and the period is easily referenced (eg, SF-MPQ PPI assesses present pain)
Report the type of pain intensity rated by participants (eg, average, usual, least, worst, current, and present)
Report the specific bodily area or pain condition to be rated; if none was specified, this should be stated

BPI, Brief Pain Inventory; NRS, numerical rating scale; VAS, visual analogue scale.

initial exploratory study, the training did not improve the reliability or validity of pain intensity assessment in adults with CLBP, hip/ knee OA, or DPN.[126] However, the program had remitted benefits for improving discriminant validity, reducing missing data, and reducing errors in rank ordering least, average, and worst pain intensity ratings.[126] Future studies are needed to examine the effects of this training program and others[135] in prospective analgesic clinical trials.

4.1.2. Pain quality

In addition to the intensity or magnitude of pain, there are affective and sensory qualities of pain that can be modulated with therapeutics;[35,61,121] therefore, it might be appropriate to measure pain affect or quality as an endpoint, depending on the intervention and patient population.[37]

4.1.2.1. Pain affect

As described earlier, pain affect reflects the extent to which pain is disturbing, unpleasant, or causes distress.[37,74] Based in part on experimental pain protocols developed by Price and Harkins,[112] the following analogy can be used to help research participants differentiate pain affect from intensity: "When thinking about listening to music, pain intensity can be thought of as the volume of the music, and pain unpleasantness can be likened to how much one likes or dislikes the music."[66] However, from a measurement perspective, distinguishing pain affect from intensity can be challenging.[47,74]

The pain affect subscale of the McGill Pain Questionnaire is the most commonly used assessment with well-established measurement properties.[39,74,88,93] (for a historical perspective on over 40 years of research involving the MPQ and subsequent short forms, see the comprehensive review by Professor Chris Main[88]). The original MPQ consisted of 20 classes of pain/ symptom descriptors (2–6 descriptors in each class) that were used to identify different qualities of pain (eg, affective and sensory). The first short form-MPQ (SF-MPQ) improved the scaling and formatting of the assessment, reducing the number of pain descriptors to 15 that are each rated by patients or research participants using a VRS (response options: none, mild, moderate, and severe).[94] The second short form (SF-MPQ-2) retained these original 15 descriptors, added 7 other descriptors to improve the assessment of neuropathic pain, and rescaled response options from the VRS to a 0 to 10 NRS (anchors: "none" and "worst possible"; symptom recall period: past week).[39] Exploratory and confirmatory factor analyses of the SF-MPQ and SF-MPQ-2 identify an affective scale with the following 4 descriptors: "tiring-exhausting," "sickening," "fearful," and "punishing-cruel." More methodological studies have evaluated the SF-MPQ than the SF-MPQ-2, but there are increasing numbers of studies demonstrating acceptable to excellent measurement properties of the SF-MPQ-2 affective subscale.[39,40,57,78,86,93]

Although there are other single-item and multi-item VRS and VAS pain affect scales, the psychometric properties are limited.[74] As noted by Jensen and Karoly,[74] differentiating pain affect from sensory components of pain is a measurement challenge requiring further theoretical development and methodological investigation.[47] Indeed, in a sample of U.S. veteran patients with chronic pain, the correlations between the SF-MPQ-2 pain affect subscale and sensory subscales ranged from 0.74 to 0.85, and the correlation of pain affect with pain severity (Multidimensional Pain Inventory [MPI]) was 0.50.[86] Similar results have been observed in younger and older adults with cancer pain.[57]

4.1.2.2. Sensory quality

The sensory qualities of pain can provide clues about the potential biopsychosocial mechanisms underlying the painful condition and help guide treatment decisions.[63] There are several multi-item questionnaires with different strengths and weaknesses that can be used to assist with phenotyping research participants or, potentially, assess treatment efficacy.[6,43,63,73] In this section, we review PROs with established measurement properties that assess pain sensory qualities as well as instruments that specifically aim to assess and screen for neuropathic pain.

The SF-MPQ-2 is designed to provide an overall assessment of pain-related symptoms by including a range of neuropathic and nonneuropathic pain descriptors.[39] In addition to assessing pain affect, multiple studies on adults with different pain conditions have identified 3 other subscales: *continuous pain* ("throbbing pain," "cramping pain," "gnawing pain," "aching pain," "heavy pain," and "tender"); *intermittent pain* ("shooting pain," "stabbing pain," "sharp pain," "splitting pain," "electric-shock pain," and "piercing"); and *neuropathic pain* ("hot-burning pain," "cold-freezing pain," "pain caused by light touch," "itching," "tingling or 'pins and needles,'", and "numbness").[39,40,57,78,86] These subscales have demonstrated internal consistency, convergent and divergent validity, and responsiveness to treatment in a variety of pain conditions.[39,40,57,78,86] Researchers have also computed ICCs for test–retest reliability of the subscales that range from a low of 0.4 to 0.7 over a 3-month test interval in adults with knee OA[142] to a high of 0.9 to 1.0 over a 3-day test interval in adults with chronic visceral pain.[156] Although the factor structure of the SF-MPQ-2 is well established, the moderate-to-high correlations (r's = 0.7–0.9) between the subscale scores raises some concern about whether they represent unique latent constructs.[57,86] In future research, it would be valuable to determine whether the subscales discriminate between different pain populations (eg, determine differences in the neuropathic pain subscale in adults with and without neuropathic pain conditions) or predict treatment response (ie, effect modification).

The Pain Quality Assessment Scale (PQAS) is the only other rigorously evaluated PRO that assesses a broad array of sensory qualities of pain.[71,75] Originally based on the Neuropathic Pain Scale (NPS)—a brief (11-item), validated measure of neuropathic pain that discriminates between different pain conditions and detects treatment effects[53]—the PQAS added 10 more items to improve the content validity for assessing neuropathic pain as well as nonneuropathic pain symptoms.[71] In total, there are 21 items in the PQAS that assess pain intensity (1 item), unpleasantness (1 item), sensory qualities (16 items), spatial characteristics (2 items rating intensity of deep and surface pain), and temporal pattern (1 item identifying intermittent, variable, or stable pain).[71] With exception to the last categorical item measuring the temporal pattern of pain, all other items are assessed on a 0 to 10 NRS with a 1-week recall period and anchors of "No pain" and "The most intense pain sensation imaginable" (note: anchors vary depending on the pain descriptor/symptom being assessed). The wording of the PQAS (instructions and certain items) was revised slightly to improve understandability based on cognitive testing with chronic pain patients.[75]

In adults diagnosed with carpal tunnel syndrome (CTS; n = 138), knee OA (n = 368), and LBP (n = 455), exploratory and confirmatory factor analyses of the spatial and sensory pain quality items identified 3 PQAS subscales: *paroxysmal pain* (shooting, sharp, electric, hot, and radiating), *superficial pain* (itchy, cold, numb, sensitive, and tingling), and *deep pain* (aching,

heavy, dull, cramping, and throbbing).[151] The internal consistencies of the subscales ranged from 0.69 to 0.85 within the different pain populations.[151] Correlations between the subscales were not reported. Notably, study participants with CTS (ie, neuropathic pain) had statistically significant lower scores on the paroxysmal and deep pain subscales and higher surface pain scores than those with knee OA.[151] The surface pain scores were also significantly higher in the CTS group compared with the LBP group.[60] In terms of responsiveness, the individual items of PQAS were shown to improve with lidocaine patch 5% in patients with CTS.[71] Among patients with moderate-to-severe neuropathic pain (peripheral neuropathy) who participated in an enriched enrollment randomized withdrawal trial of pregabalin, all 3 PQAS pain quality subscales improved from pretitration to posttitration phases; however, pregabalin had a greater effect on the PQAS paroxysmal pain subscale than on the deep or surface pain scales, or on pain intensity.[70] During the withdrawal phase, pregabalin (vs placebo) had the largest effect on paroxysmal pain followed by surface pain, but there were no statistically significant effects on deep pain or pain intensity.[70] A similar pattern of results was observed with the PQAS subscales in an enriched enrollment randomized withdrawal trial of an extended-release formulation of oxymorphone for LBP, although there was a high discontinuation rate in the withdrawal phase of this trial.[60] Taken together, these studies demonstrate that analgesics can have differential effects on pain quality as measured by PQAS.[60,70] Furthermore, subsequent analysis of the pregabalin trial showed the paroxysmal and deep pain subscales at baseline were associated with treatment response but not with placebo response,[54] illustrating the potential predictive value of phenotyping self-reported pain qualities.[6,43,63]

Both the SF-MPQ-2[39] and PQAS[71,75] provide a relatively brief (21–22 items), overall assessment of the sensory qualities of pain; however, each has unique strengths and weaknesses. A clear strength of the SF-MPQ-2 is that it reliably assesses pain affect with 4 items, whereas the PQAS has a single-item rating of pain unpleasantness. However, the PQAS has greater content validity, covering a wider range of pain qualities reported by patients with neuropathic and nonneuropathic pain conditions.[73,85] It should be noted that the Patient-Reported Outcome Measurement Information System (PROMIS) has developed a pain quality item bank[117] and validated a neuropathic pain quality PRO,[3] but the responsiveness of PROMIS pain quality measures to pain treatment has not been established. The Multidimensional Affect and Pain Survey is another multi-item assessment of pain quality, but it has not been studied extensively in different pain populations and has less content validity than the SF-MPQ-2 and PQAS.[73] Currently, the SF-MPQ-2 and PQAS are recommended by IMMPACT for use in chronic pain treatment trials because they are the most well-established pain quality assessments with acceptable to excellent measurement properties.[43]

4.1.2.2.1. Neuropathic pain assessment and screening

Comprehensive reviews on self-reported measures of neuropathic pain have been published previously.[6,63,89] There are several assessment questionnaires specifically designed to monitor neuropathic pain symptoms as well as screening tools to assist clinicians identify neuropathic pain. Both assessment and screening questionnaires can also be used to phenotype potential treatment responders/nonresponders. We briefly review both types of self-reported measures for use in clinical trials of neuropathic pain. There are also ClinROs that can be used in combination with the self-reported measures to improve diagnostic accuracy of neuropathic pain.[6]

As described earlier, the NPS, PQAS, and SF-MPQ-2 are assessments consisting of neuropathic pain quality descriptors that have been used previously in neuropathic pain treatment trials. The Neuropathic Pain Symptom Inventory (NPSI) contains 10 symptom descriptors that are assessed on a 0 to 10 NRS with a 24-hour recall period and descriptor-specific anchors (eg, "No burning" to "Worst burning imaginable").[14] In addition, there are 2 temporal items that assess pain duration and the number of pain paroxysms in the past 24 hours.[14] The factor structure of the NPSI has ranged from 3 to 5 dimensions, depending on the pain populations studied and analytic techniques, although *evoked pain* and *deep pain* were factors identified consistently across studies.[7,14,30,49] Several studies have demonstrated that the NPSI has acceptable to excellent measurement properties.[6,7,14,63] Furthermore, IMMPACT has recommended the NPSI for phenotyping patients with neuropathic pain.[43]

Several brief and easy-to-use screening tools for neuropathic pain are available,[6] including the Leeds Assessment of Neuropathic Symptoms and Signs (LANSS),[11,12] Douleur neuropathique en 4 questions (DN4),[13,15] painDETECT,[50] and ID Pain.[110] Similar to neuropathic pain assessments, these screening measures use symptom descriptors such as burning or electric shock to identify neuropathic pain. In general, these measures have adequate sensitivities and specificities, but some perform better depending on the characteristics of pain population and study goal.[6] Importantly, a systematic review has identified limitations in the reliability and validity of these screening measures, particularly in their translations into different languages as well as their lack of diagnostic specificity.[89] As noted by IMMPACT, however, self-report screening tools should not replace a comprehensive clinical examination for diagnosing a neuropathic pain condition.[6,43] PainDETECT[50] was recommended by IMMPACT to screen for neuropathic pain phenotypes.[43]

4.1.3. Rescue analgesics and concomitant pain treatments

Systematic monitoring of rescue medication use and concomitant use of pain treatments during the course of a chronic pain clinical trial is necessary for appropriately interpreting trial results and is recommended as a core outcome measure by IMMPACT.[37] As noted by Rowbotham and McDermott[122] in their article in this supplement, pain relief derived from rescue medication can reduce assay sensitivity. Therefore, the amount (dose), date taken, and reason for the rescue medication use (ie, to differentiate relieving pain for the condition being investigated or some other reason such as a transient headache or dental pain) should be systematically recorded (ideally at the time the medication is taken). These data can then be analyzed as a secondary outcome (eg, time to initial rescue medication use or total amount taken through the course of the trial). Further research is needed to explore methods of integrating rescue medication use with other outcome measures such as pain intensity ratings.[28,124]

4.2. Core outcome domain of physical functioning

For many adults, chronic pain negatively impacts their physical functioning,[100,105] and improvement in physical functioning is an important treatment goal.[101,139] However, treatments designed to improve pain may not necessarily improve physical functioning. For example, in pooled analyses of 15 clinical trials involving first-line treatments for neuropathic pain, the correlation between

change in pain intensity from baseline to posttreatment and change in physical functioning was small (−0.22 in DPN trials and −0.08 in PHN trials).[103] For these reasons, IMMPACT recommends physical functioning as a core outcome domain.[37]

In pain research, the terms "physical functioning" and "pain interference" are often used interchangeably; however, it is important to distinguish between them to appropriately interpret trial results. Physical functioning refers to one's ability to perform activities that require physical action, such as self-care (eg, bathing, dressing, and eating), walking indoors or outdoors, or climbing stairs. Pain interference, however, is the extent to which pain impedes one's ability to perform or participate in activities, ranging from basic physical activities to more complex social activities. Some pain interference measures cover a wider range of activity domains, whereas others specifically assess pain interference with physical functioning. Therefore, when considering outcome measures, it is important to not only differentiate measures of physical functioning from pain interference, but also general measures of pain interference from specific measures of pain interference with physical functioning.

4.2.1. Types of clinical outcome assessments for physical functioning

Multiple measurement approaches can be used to assess physical functioning. Patient-reported outcome measures assess the study participant's perception of their own abilities to perform activities, whereas PerfOs objectively assess a participant's physical capacity to perform a standardized, discrete task. Although PROs and PerfOs can provide complementary information on a research participant's physical functioning, neither of these types of COAs objectively measure the participant's physical functioning in their home and/or work environments. However, wearable devices with accelerometers can be used to collect the research participant's daily activity pattern in their real-world environment. It is likely that each type of assessment provides a different perspective on the participant's functioning, and triangulating from them can provide a more comprehensive assessment. In the sections below, we review examples of different types of COAs for physical functioning and pain interference with physical functioning.

4.2.2. Patient-reported outcomes of physical functioning and pain interference

There are several generic and pain condition-specific PROs of physical functioning and pain interference. Generic measures facilitate the comparison of results across different clinical trials and pain populations as well as facilitate cost-effectiveness evaluation, whereas condition-specific measures have greater content validity for the population of interest and therefore will be more responsive to treatment effects. For these reasons, IMMPACT recommends the collection of both generic and condition-specific measures, if available.[37] Because several groups have published reviews or recommendations on PROs of physical functioning and pain interference, we will highlight recommended measures in the following 2 sections.[18,23,52,65,91,134,141,152]

4.2.2.1. Generic measures of physical functioning and pain interference

The Short Form-36 (SF-36) is one of the most commonly used PROs of health-related quality of life that measures 8 domains, including physical functioning.[157] The physical function subscale

consists of 10 items that ask participants, "Does your health now limit you in these activities? If so, how much?" Activities range from bathing or dressing oneself to running. Each item has 3 response options: "Yes, limited a lot"; "Yes, limited a little"; and "No, not limited at all." Scores from each item are summed and transformed to a 0 to 100 scale, where higher scores indicate better physical functioning. The measurement properties of the SF-36 physical functioning subscale have been examined in multiple pain populations and are generally acceptable to excellent.

Both the Brief Pain Inventory (BPI)[21] and MPI[80] have pain interference scales with well-established measurement properties and are recommended by IMMPACT as core outcome measures.[37] The MPI has 9 pain interference items with a 0 to 6 NRS, whereas the BPI has 7 items with a 0 to 10 NRS. Interestingly, although the BPI pain interference scale is typically analyzed with a single score based on factor analyses supporting unidimensionality,[22,133] more recent studies using Rasch analysis[153] and confirmatory factor analysis[154] suggest that there are 2 separate subscales: "physical interference" (items: general activity, walking ability, and normal work) and "affective interference" (items: mood, relationships with others, and enjoyment of life), leaving out the item on sleep. Further analysis of the BPI factor structure with clinical trial data are needed to determine whether only the 3 physical/activity-related items are sufficient for measuring pain interference with physical functioning.

Building off of the SF-36, BPI, MPI, and many other legacy measures, the NIH PROMIS has developed item banks using IRT to comprehensively measure a variety of health-related domains, including physical functioning[119] and pain interference.[1] The IRT-based item banks can be administered by computer adaptive tests that assess the full spectrum of physical functioning and pain interference, from low to high. Published or custom short forms can also be used to measure physical functioning and pain interference. The reliability and validity of PROMIS physical functioning and pain interference measures have been established in several pain populations,[1,2,4,5,24,51] and emerging studies are demonstrating the responsiveness of the short-form versions of these measures.[16,17,84] Furthermore, there is now a crosswalk table available to transform the BPI pain interference scores to PROMIS pain interference short-form scores[5] (crosswalks for scores from other PROMIS and legacy measures are available at http://www.prosettastone.org/Pages/default.aspx.)

4.2.2.2. Condition-specific measures of physical functioning and pain interference

In addition to generic measures, there are several condition-specific pain interference PROs, primarily for musculoskeletal conditions. For example, a recent systematic review of PROs for physical functioning or pain interference in adults with nonspecific LBP had identified 17 measures.[18] After applying COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology[95–97] and undergoing a 2-round Delphi survey process with experts, the Oswestry Disability Index version 2.1a[46] and 24-item Roland-Morris Disability Questionnaire[118] were recommended as a core outcome measure for clinical trials on LBP.[118] Importantly, this systematic review also highlighted that the content validity of many commonly used PROs of physical functioning and pain interference in adults with LBP is inadequate,[19] which can potentially decrease the responsiveness of an outcome measure and reduce assay sensitivity.

Similar findings have been reported for PROs of physical functioning and pain interference in hip or knee OA.[52,65] For example, although the Western Ontario and McMaster Universities Arthritis Index (WOMAC)[10] is a widely used and recommended PRO in adults with hip or knee OA, systematic reviews find that most studies evaluating the WOMAC do not meet the COSMIN criteria for modern measurement standards.[52,55,65] This may, in part, reflect poor quality of study reporting rather than poor measurement qualities of the PRO. Future research evaluating the measurement properties of a new or existing COA should attend to the COSMIN criteria. Finally, in contrast to the abundance of condition-specific measures of physical functioning or pain interference for chronic musculoskeletal pain, recent systematic reviews have reported a dearth of measures for neuropathic pain.[91,92] Thus, there is a need to either modify existing PROs or develop new ones that assess functional limitations that are attributable to specific neuropathic pain conditions because generic measures (eg, SF-36 physical function subscale) may not capture them.

4.2.3. Performance-based outcomes

Both IMMPACT and OMERACT recognize that PerfOs can provide complementary information to PROs.[134] An advantage of PerfOs is that they assess one's physical capacity in a standardized manner, eliminating behavioral adaptations (eg, changes in the frequency or how a person performs a task) or environmental modifications that can influence a person's report of their ability to perform activities. There are a number of PerfOs used in clinical pain research and trials.[134,141] A systematic review applied COSMIN methodology to assess the measurement properties of 21 PerfOs of physical functioning in adults diagnosed with hip or knee OA.[33] Based on the results of the systematic review[33] and the consensus of expert clinicians and researchers, OARSI recommends the following PerfOs assessments for hip or knee OA: the 30-second chair-stand test, 40-m fast-paced walk test, a stair-climb test, timed up-and-go test, and 6-minute walk test.[34] Of these, the first 3 were recommended as the minimal core set of PerfO assessments for hip or knee OA or after arthroplasty.[34] Subsequent methodological investigation of these PerfOs generally supports the OARSI recommendations.[32] It should be noted that the OARSI-recommended set of PerfO assessments and other PerfOs might be appropriate for other chronic pain conditions, but their measurement properties must be established in the populations of interest, as recommended by IMMPACT and OMERACT.[134]

4.2.4. Accelerometer-based outcomes

As described earlier, there are strengths and weaknesses to PROs and PerfOs for assessing physical functioning. A shared weakness is that neither of these COAs provides an objective, time-stamped assessment of the daily activities a person actually does in their own real-world environment. With advances in mobile technology and greater acceptance of wearable devices, accelerometry can provide a unique window to understand how individuals with chronic pain function physically and potentially respond to treatments. Although a number of observational studies have examined accelerometer-measured physical activity in adults with chronic pain,[29,98,99,131], relatively few clinical trials have collected accelerometer data to assess physical activity. One exception is a randomized, placebo-controlled, crossover trial of celecoxib for knee OA in which the accelerometer data were combined with pain intensity ratings as a composite outcome measure.[136] There was a greater difference in responder rates between celecoxib and placebo using the pain-activity composite outcome than using pain ratings alone.[136] Although this finding is exciting and provocative, there is a need to further develop accelerometry assessment protocols in clinical trials of chronic pain as well as develop and validate novel outcome measures that capitalize on the richness of the raw signal data.[104,136]

4.3. Core outcome domain of emotional functioning

It is widely recognized that pain and psychological distress often co-occur.[56] Indeed, the co-occurrence of pain and depression likely ranges between 30% and 50%.[8] Furthermore, longitudinal studies have demonstrated bidirectional effects between pain and depression,[83] and psychological distress or psychiatric comorbidity can reduce the effectiveness of pain treatments.[42,158] Accordingly, it is important to measure emotional functioning in chronic pain trials as a potential endpoint or as a treatment effect modifier.

The Beck Depression Inventory (BDI)[9] and the Profile of Mood States[90] have been recommended as core outcome measures for chronic pain trials.[37] Both measures have well-established reliability and validity and have been used extensively in clinical trials,[37] but these 2 measures have some weaknesses, including the length of the Profile of Mood States (65 items) and the BDI does not measure anxiety. However, the Hospital Anxiety and Depression Scale (HADS) provides a more efficient alternative with a total of 14 items (7 for anxiety and 7 for depression).[161] In addition, the HADS was designed for use in a general medical setting rather than for a psychiatric patient population, excluding somatic symptoms that can be confounded by illness and side effects of treatment.[161] The measurement properties of the HADS were recently evaluated in a randomized, double-blind, parallel-group study of tapentadol immediate-release vs oxycodone immediate-release in 666 adults with acute low back pain.[140] The HADS demonstrated good-to-excellent internal consistency, convergent validity, predictive validity, and responsiveness to treatment, and the data fit the prespecified factor structure.[140] Accordingly, these results indicate that the HADS is a valid PRO assessment of anxiety and depression in adults with acute low back pain. Recently, it was recommended by IMMPACT as a core phenotyping measure to assess general negative affect.[43] Future research should replicate these findings and extend them to other pain populations.

The Patient Health Questionnaire-9 (PHQ-9) is a 9-item PRO that is used to screen for depression, assess the severity of depression, and monitor depression treatment in clinical settings.[81] Each item measures 1 of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition's (DSM-IV's) 9 criteria for depressive disorder diagnoses.[81] The psychometric properties of the PHQ-9 are well established.[125] The PHQ-9 has been used in several clinical trials, and the 2-item version (PHQ-2) is recommended by the Veterans Health Administration Pain Measures Work Group, which prioritized strong psychometric properties and pragmatic aspects (brevity) in recommending outcome measures.[82]

The NIH PROMIS has developed item banks for depression and anxiety based on items from legacy measures, such as the BDI, HADS, and PHQ-9.[109] These item banks can be administered using computerized adaptive testing, or fixed-length short forms are available. As with pain interference and physical functioning, the PROMIS depression and anxiety measures have demonstrated excellent measurement properties[109] and

crosswalks with legacy measures have been established.[20] The NIH Task Force on Research Standards for Chronic Low Back Pain recommends the 4-item PROMIS short form for depression.[31]

4.4. Core outcome domain of participant global ratings improvement

Global ratings of improvement and treatment satisfaction are self-reported measures that provide research participants the opportunity to integrate multiple aspects of their treatment experience into a single assessment, weighing treatment benefits (eg, improvements in pain, physical functioning, and emotional functioning) and harms (ie, side effects and AEs).[37] In addition, these ratings can be used to gauge what study participants consider a meaningful treatment response. Indeed, responses to the global improvement or treatment satisfaction assessment serve as anchors for establishing the clinically important difference in pain intensity and other outcomes. Notably, improvements in pain intensity and physical functioning are 2 consistent aspects of the treatment response that patients with OA, CLBP, and FM consider personally meaningful.[36,58,77] There are several single-item ratings (Likert scale) used at the end of clinical trials to assess patient improvement or satisfaction with treatment, including the Patient Global Assessment of Treatment Satisfaction, the Patient-Rated Global Assessment of Response to Therapy, and the Patient Global Impression of Change.[62] The Patient Global Impression of Change with 7 ordinal response options (ranging from "Very much worse" to "Very much improved") is recommended by IMMPACT as a core outcome measure.[37]

4.5. Core outcome domain of reporting adverse events

In any type of clinical trial, it is critically important to systematically assess, analyze, and report AEs. This information is essential to determine the research participant's overall treatment experience during the course of the clinical trial as well as to inform clinical decision-making. Therefore, ACTTION has conducted a series of systematic reviews to make evidence-based recommendations on methods of assessing, analyzing, and reporting AEs in pain treatment trials.[69,127,130,160] Despite the expansion in 2004 of the Consolidated Standards of Reporting Trials (CONSORT) statement with a set of standard harms-reporting recommendations, the reporting of AEs after 2005 has remained inadequate in pharmacologic,[127,130] nonpharmacologic,[69] and intravenous and invasive[160] pain treatment trials. In an effort to improve AE reporting, Smith et al.[130] developed a comprehensive checklist of reporting items that captures clinically relevant AE information that are necessary to draw conclusions about the participants' AE experience as well as the trial's overall AE assessment (see Table 6 in Ref. 130).

4.6. Core outcome domain of participant disposition

In accordance with IMMPACT recommendations[37] and CONSORT guidelines, the disposition of study participants should be systematically recorded to facilitate the transparent reporting of trial results. This includes detailed information on participants' progression throughout the trial, from screening for eligibility to reasons for withdrawal to follow-up status. The article by Gewandter et al.[59] in this supplement provides a clear description of the reporting requirements of pain clinical trials as well as provides a pain-specific CONSORT checklist.

5. Development of new clinical outcome assessments

Under the Prescription Drug User Fee Act (PDUFA) VI, the FDA is preparing guidance on Patient-Focused Drug Development, which is defined as, "a systematic approach to help ensure that patients' experiences, perspectives, needs, and priorities are captured and meaningfully incorporated into the development and evaluation of medical products throughout the medical product life cycle."[150] The central role of patients throughout the drug development process has implications for COA.[143] It will be critical to engage patients and define the intended treatment benefit as a specific aspect of how a patient feels or functions.[155] In the case that no appropriate COA sufficiently measures this meaningful health aspect, then the FDA has a COA Qualification Program for applicants to enter into a multi-stage process of developing and validating a new COA that measures a well-defined COI in a specific context of use (ie, the intended population, setting, and manner of use). If approved, then the "COA qualification represents a conclusion that within the stated context of use, results of the assessment can be relied upon to measure a specific concept and have a specific interpretation and application in drug development and regulatory decision-making."[149] Currently, ACTTION is pursuing the COA qualification process for a new PRO measure of pain intensity (QUALITE-Pain: QUALIfied for Therapeutic Evaluations of Pain) and a new accelerometer-based physical activity tool (PAACT: Physical Activity Accelerometry Assessment for Analgesic Clinical Trials).

6. Summary and recommendations for future research

In summary, COAs play an important role in the evaluation of treatment efficacy and safety. We have reviewed several outcome domains and measures for clinical trials of chronic pain treatments that were recommended by expert groups. This review also identified opportunities to improve the quality of outcome assessment. For example, greater efforts are needed to standardize the reporting of critical elements of pain intensity assessment (see **Table 2** and Ref. 128) as well as AEs.[130] In addition, there remain opportunities to further evaluate the effects of training research participants to improve the accuracy of pain ratings in clinical trials. Finally, although the outcome measures reviewed in this article generally have acceptable-to-excellent measurement properties based on CTT, several measures have not been evaluated using modern psychometric methods such as IRT. In future research, the COSMIN checklists and taxonomy of measurement properties[95–97] should be considered when evaluating existing, modified, or newly developed COAs.

Disclosures

measures. The remaining authors have no conflict of interest to declare.

References

[1] Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, Cella D, Rothrock N, Keefe F, Callahan L, Lai JS. Development of a PROMIS item bank to measure pain interference. PAIN 2010;150:173–82.

[2] Amtmann D, Kim J, Chung H, Askew RL, Park R, Cook KF. Minimally important differences for Patient Reported Outcomes Measurement Information System pain interference for individuals with back pain. J Pain Res 2016;9:251–5.

[3] Askew RL, Cook KF, Keefe FJ, Nowinski CJ, Cella D, Revicki DA, DeWitt EM, Michaud K, Trence DL, Amtmann D. A PROMIS measure of neuropathic pain quality. Value Health 2016;19:623–30.

[4] Askew RL, Cook KF, Revicki DA, Cella D, Amtmann D. Evidence from diverse clinical populations supported clinical validity of PROMIS pain interference and pain behavior. J Clin Epidemiol 2016;73:103–11.

[5] Askew RL, Kim J, Chung H, Cook KF, Johnson KL, Amtmann D. Development of a crosswalk for pain interference measured by the BPI and PROMIS pain interference short form. Qual Life Res 2013;22:2769–76.

[6] Attal N, Bouhassira D, Baron R. Diagnosis and assessment of neuropathic pain through questionnaires. Lancet Neurol 2018;17:456–66.

[7] Attal N, Fermanian C, Fermanian J, Lanteri-Minet M, Alchaar H, Bouhassira D. Neuropathic pain: are there distinct subtypes depending on the aetiology or anatomical lesion? PAIN 2008;138:343–53.

[8] Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163:2433–45.

[9] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–71.

[10] Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.

[11] Bennett M. The LANSS Pain Scale: the leeds assessment of neuropathic symptoms and signs. PAIN 2001;92:147–57.

[12] Bennett MI, Smith BH, Torrance N, Potter J. The S-LANSS score for identifying pain of predominantly neuropathic origin: validation for use in clinical and postal research. J Pain 2005;6:149–58.

[13] Bouhassira D, Attal N, Alchaar H, Boureau F, Brochet B, Bruxelle J, Cunin G, Fermanian J, Ginies P, Grun-Overdyking A, Jafari-Schluep H, Lantéri-Minet M, Laurent B, Mick G, Serrie A, Valade D, Vicaut E. Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). PAIN 2005;114:29–36.

[14] Bouhassira D, Attal N, Fermanian J, Alchaar H, Gautron M, Masquelier E, Rostaing S, Lanteri-Minet M, Collin E, Grisart J, Boureau F. Development and validation of the neuropathic pain symptom inventory. PAIN 2004;108:248–57.

[15] Bouhassira D, Lanteri-Minet M, Attal N, Laurent B, Touboul C. Prevalence of chronic pain with neuropathic characteristics in the general population. PAIN 2008;136:380–87.

[16] Chen CX, Kroenke K, Stump T, Kean J, Krebs EE, Bair MJ, Damush T, Monahan PO. Comparative responsiveness of the PROMIS pain interference short forms with legacy pain measures: results from three randomized clinical trials. J Pain 2019;20:664–675.

[17] Chen CX, Kroenke K, Stump TE, Kean J, Carpenter JS, Krebs EE, Bair MJ, Damush TM, Monahan PO. Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. PAIN 2018;159:775–82.

[18] Chiarotto A, Boers M, Deyo RA, Buchbinder R, Corbin TP, Costa LO, Foster NE, Grotle M, Koes BW, Kovacs FM, Lin CW. Core outcome measurement instruments for clinical trials in nonspecific low back pain. PAIN 2018;159:481–95.

[19] Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. J Clin Epidemiol 2018;95:73–93.

[20] Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. Psychol Assess 2014;26:513–27.

[21] Cleeland CS, Ryan KM. Pain assessment: global use of the brief pain inventory. Ann Acad Med Singap 1994;23:129–38.

[22] Cleeland CS. The brief pain inventory user guide, 2009.

[23] Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: international knee documentation committee (IKDC) subjective knee evaluation form, knee injury and osteoarthritis outcome score (KOOS), knee injury and osteoarthritis outcome score physical function short form (KOOS-PS), knee outcome survey activities of daily living scale (KOS-ADL), lysholm knee scoring scale, oxford knee score (OKS), western Ontario and McMaster Universities osteoarthritis Index (WOMAC), activity rating scale (ARS), and tegner activity score (TAS). Arthr Care Res 2011;63:S208–28.

[24] Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S, Dewalt DA, Fries JF, Pilkonis PA, Reeve BB, Stone AA, Weinfurt KP, Cella D. PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. J Clin Epidemiol 2016;73: 89–102.

[25] Coon CD, Cappelleri JC. Interpreting change in scores on patient-reported outcome instruments. Ther Innov Regul Sci 2016;50:22–9.

[26] Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 2018;27:33–40.

[27] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297–334.

[28] Dai F, Silverman DG, Chelly JE, Li J, Belfer I, Qin L. Integration of pain score and morphine consumption in analgesic clinical studies. J Pain 2013;14:767–77.

[29] Dansie EJ, Turk DC, Martin KR, Van Domelen DR, Patel KV. Association of chronic widespread pain with objectively measured physical activity in adults: findings from the National Health and Nutrition Examination survey. J Pain 2014;15:507–15.

[30] de Andrade DC, Ferreira KA, Nishimura CM, Yeng LT, Batista AF, de Sá K, Araujo J, Stump PR, Kaziyama HH, Galhardoni R, Fonoff ET, Ballester G, Zakka T, Bouhassira D, Teixeira MJ. Psychometric validation of the Portuguese version of the neuropathic pain symptoms inventory. Health Qual Life Outcomes 2011;9:107.

[31] Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, Carrino J, Chou R, Cook K, DeLitto A, Goertz C, Khalsa P, Loeser J, Mackey S, Panagis J, Rainville J, Tosteson T, Turk D, Von Korff M, Weiner DK. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014;15:569–85.

[32] Dobson F, Hinman RS, Hall M, Marshall CJ, Sayer T, Anderson C, Newcomb N, Stratford PW, Bennell KL. Reliability and measurement error of the Osteoarthritis Research Society International (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. Osteoarthritis Cartilage 2017; 25:1792–6.

[33] Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthr Cartil 2012;20:1548–62.

[34] Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, Buchbinder R, Snyder-Mackler L, Henrotin Y, Thumboo J, Hansen P, Bennell KL. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthr Cartil 2013;21:1042–52.

[35] Dworkin RH, Corbin AE, Young JP, Sharma U, LaMoreaux L, Bockbrader H, Garofalo EA, Poole RM. Pregabalin for the treatment of postherpetic neuralgia A randomized, placebo-controlled trial. Neurol 2003;60:1274–83.

[36] Dworkin RH, Jensen MP, Gould E, Jones BA, Xiang Q, Galer BS, Gammaitoni AR. Treatment satisfaction in osteoarthritis and chronic low back pain: the role of pain, physical and emotional functioning, sleep, and adverse events. J Pain 2011;12:416–24.

[37] Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, Kerns RD, Stucki G, Allen RR, Bellamy N, Carr DB, Chandler J, Cowan P, Dionne R, Galer BS, Hertz S, Jadad AR, Kramer LD, Manning DC, Martin S, McCormick CG, McDermott MP, McGrath P, Quessy S, Rappaport BA, Robbins W, Robinson JP, Rothman M, Royal MA, Simon L, Stauffer JW, Stein W, Tollett J, Wernicke J, Witter J. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. PAIN 2005;113:9–19.

[38] Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, Farrar JT, Hertz S, Raja SN, Rappaport BA, Rauschkolb C,

Sampaio C. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. PAIN 2009; 146:238–44.

[39] Dworkin RH, Turk DC, Revicki DA, Harding G, Coyne KS, Peirce-Sandner S, Bhagwat D, Everton D, Burke LB, Cowan P, Farrar JT, Hertz S, Max MB, Rappaport BA, Melzack R. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). PAIN 2009;144:35–42.

[40] Dworkin RH, Turk DC, Trudeau JJ, Benson C, Biondi DM, Katz NP, Kim M. Validation of the Short-form McGill Pain Questionnaire-2 (SF-MPQ-2) in acute low back pain. J Pain 2015;16:357–66.

[41] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.

[42] Edwards RR, Cahalan C, Calahan C, Mensing G, Smith M, Haythornthwaite JA. Pain, catastrophizing, and depression in the rheumatic diseases. Nat Rev Rheumatol 2011;7:216–24.

[43] Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, Hansson P, Haroutounian S, Arendt-Nielsen L, Attal N, Baron R, Brell J, Bujanover S, Burke LB, Carr D, Chappell AS, Cowan P, Etropolski M, Fillingim RB, Gewandter JS, Katz NP, Kopecky EA, Markman JD, Nomikos G, Porter L, Rappaport BA, Rice AS, Scavone JM, Scholz J, Simon LS, Smith SM, Tobias J, Tockarshewsky T, Veasley C, Versavel M, Wasan AD, Wen W, Yarnitsky D. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. PAIN 2016;157:1851–71.

[44] European Medicines Agency. Committee for medicinal products for human use (CHMP): reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. European Medicines Agency: London, United Kingdom, 2005.

[45] European Medicines Agency. Appendix 2 to the guideline on the evaluation of anticancer medicinal products in man. The use of patient-reported outcome (PRO) measures in oncology studies. European Medicines Agency: London, United Kingdom, 2016.

[46] Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980;66:271–3.

[47] Fernandez E, Turk DC. Nociception and emotion: separation and synthesis. Psychol Bull 1992;112:205–17.

[48] Fillingim RB, Loeser JD, Baron R, Edwards RR. Assessment of chronic pain: domains, methods, and mechanisms. J Pain 2016;17:T10–20.

[49] Freeman R, Baron R, Bouhassira D, Cabrera J, Emir B. Sensory profiles of patients with neuropathic pain based on the neuropathic pain symptoms and signs. PAIN 2014;155:367–76.

[50] Freynhagen R, Baron R, Gockel U, Tölle TR. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006;22:1911–20.

[51] Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in assessing physical function in arthritis: PROMIS short forms and computerized adaptive testing. J Rheumatol 2009;36:2061–6.

[52] Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, Mirza F, Bayona M. A systematic review of measurement properties of patient-reported outcome measures used in patients undergoing total knee arthroplasty. J Arthroplasty 2017;32:1688–97.

[53] Galer BS, Jensen MP. Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. Neurol 1997;48:332–8.

[54] Gammaitoni AR, Smugar SS, Jensen MP, Galer BS, Bolognese JA, Alon A, Hewitt DJ. Predicting response to pregabalin from pretreatment pain quality: clinical applications of the pain quality assessment scale. Pain Med 2013;14:526–32.

[55] Gandek B. Measurement properties of the western Ontario and McMaster Universities osteoarthritis Index: a systematic review. Arthr Care Res 2015;67:216–29.

[56] Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007;133:581–624.

[57] Gauthier LR, Young A, Dworkin RH, Rodin G, Zimmermann C, Warr D, Librach SL, Moore M, Shepherd FA, Pillai Riddell R, Macpherson A, Melzack R, Gagliese L. Validation of the short-form McGill pain questionnaire-2 in younger and older people with cancer pain. J Pain 2014;15:756–70.

[58] Geisser ME, Clauw DJ, Strand V, Gendreau RM, Palmer R, Williams DA. Contributions of change in clinical status parameters to Patient Global Impression of Change (PGIC) scores among persons with fibromyalgia treated with milnacipran. PAIN 2010;149:373–8.

[59] Gewandter JS, Eisenach JC, Gross RA, Jensen MP, Keefe FJ, Lee DA, Turk DC. Checklist for the preparation and review of pain clinical trial publications: a pain-specific supplement to CONSORT. PAIN Rep 2019;4:e621.

[60] Gould EM, Jensen MP, Victor TW, Gammaitoni AR, White RE, Galer BS. The pain quality response profile of oxymorphone extended release in the treatment of low back pain. Clin J Pain 2009;25:116–22.

[61] Gracely RH, McGrath P, Dubner R. Validity and sensitivity of ratio scales of sensory and affective verbal pain descriptors: manipulation of affect by diazepam. PAIN 1978;5:19–29.

[62] Guy W. Public health service alcohol, drug abuse, and mental health administration. ECDEU assessment manual for psychopharmacology, revised, 1976. Rockville: US Department of Health, Education, and Welfare, 1976. pp. 217–22. DHEW Publication no. [ADM] 76–338.

[63] Haanpää M, Attal N, Backonja M, Baron R, Bennett M, Bouhassira D, Cruccu G, Hansson P, Haythornthwaite JA, Iannetti GD, Jensen TS, Kauppila T, Nurmikko TJ, Rice AS, Rowbotham M, Serra J, Sommer C, Smith BH, Treede RD. NeuPSIG guidelines on neuropathic pain assessment. PAIN 2011;152:14–27.

[64] Hadjistavropoulos T, Herr K, Prkachin KM, Craig KD, Gibson SJ, Lukas A, Smith JH. Pain assessment in elderly adults with dementia. Lancet Neurol 2014;13:1216–27.

[65] Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, Price AJ. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. Patient Rep Outcome Meas 2016;7:101.

[66] Hemington KS, Cheng JC, Bosma RL, Rogachov A, Kim JA, Davis KD. Beyond negative pain-related psychological factors: resilience is related to lower pain affect in healthy adults. J Pain 2017;18:1117–28.

[67] Hicks CL, von Baeyer CL, Spafford P, van Korlaar I, Goodenough B. The Faces Pain Scale—revised: toward a common metric in pediatric pain measurement. PAIN 2001;93:173–83.

[68] Hjermstad MJ, Fayers PM, Haugen DF, Caraceni A, Hanks GW, Loge JH, Fainsinger R, Aass N, Kaasa S; European Palliative Care Research Collaborative (EPCRC). Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. J Pain Symptom Manage 2011;41:1073–93.

[69] Hunsinger M, Smith SM, Rothstein D, McKeown A, Parkhurst M, Hertz S, Katz NP, Lin AH, McDermott MP, Rappaport BA, Turk DC, Dworkin RH. Adverse event reporting in nonpharmacologic, noninterventional pain clinical trials: ACTTION systematic review. PAIN 2014;155:2253–62.

[70] Jensen MP, Gammaitoni AR, Bolognese JA, Alon A, Smugar SS, Galer BS, Hewitt DJ. The pain quality response profile of pregabalin in the treatment of neuropathic pain. Clin J Pain 2012;28:683–6.

[71] Jensen MP, Gammaitoni AR, Olaleye DO, Oleka N, Nalamachu SR, Galer BS. The pain quality assessment scale: assessment of pain quality in carpal tunnel syndrome. J Pain 2006;7:823–32.

[72] Jensen MP, Hu X, Potts SL, Gould EM. Single vs composite measures of pain intensity: relative sensitivity for detecting treatment effects. PAIN 2013;154:534–8.

[73] Jensen MP, Johnson LE, Gertz KJ, Galer BS, Gammaitoni AR. The words patients use to describe chronic pain: implications for measuring pain quality. PAIN 2013;154:2722–8.

[74] Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Turk DC, Melzack R, editors. Handbook of pain assessment. New York: Guilford Press, 2011.

[75] Jensen MP, Lin CP, Kupper AE, Galer BS, Gammaitoni AR. Cognitive testing and revision of the pain quality assessment scale. Clin J Pain 2013;29:400–10.

[76] Jensen MP, McFarland CA. Increasing the reliability and validity of pain intensity measurement in chronic pain patients. PAIN 1993;55:195–203.

[77] Jensen MP, Wang W, Potts SL, Gould EM. The meaning of global outcome measures in pain clinical trials: more than just change in pain intensity. Clin J Pain 2013;29:289–95.

[78] Kachooei AR, Ebrahimzadeh MH, Erfani-Sayyar R, Salehi M, Salimi E, Razi S. Short form-McGill pain questionnaire-2 (SF-MPQ-2): a cross-cultural adaptation and validation study of the Persian version in patients with knee osteoarthritis. Arch Bone Jt Surg 2015;3:45–50.

[79] Kahneman D, Fredrickson BL, Schreiber CA, Redelmeier DA. When more pain is preferred to less: adding a better end. Psychol Sci 1993;4: 401–5.

[80] Kerns RD, Turk DC, Rudy TE. the West haven-yale multidimensional pain inventory (WHYMPI). PAIN 1985;23:345–56.

[81] Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606–13.

[82] Kroenke K, Krebs EE, Turk D, Von Korff M, Bair MJ, Allen KD, Sandbrink F, Cheville AL, DeBar L, Lorenz KA, Kerns RD. Core outcome measures for chronic musculoskeletal pain research: recommendations from a Veterans Health Administration work group. Pain Med 2019. doi: 10.1093/pm/pny279. [Epub ahead of print].

[83] Kroenke K, Wu J, Bair MJ, Krebs EE, Damush TM, Tu W. Reciprocal relationship between pain and depression: a 12-month longitudinal analysis in primary care. J Pain 2011;12:964–73.

[84] Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and minimally important differences for 4 patient-reported outcomes measurement information system short forms: physical function, pain interference, depression, and anxiety in knee osteoarthritis. J Pain 2017;18:1096–110.

[85] Linl CP, Kupperl AE, Gammaitonil AR, Galerl BS, Jensen MP. Frequency of chronic pain descriptors: implications for assessment of pain quality. Eur J Pain 2011;15:628–33.

[86] Lovejoy TI, Turk DC, Morasco BJ. Evaluation of the psychometric properties of the revised short-form McGill Pain Questionnaire. J Pain 2012;13:1250–7.

[87] Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, Snyder C, Boers M, Cella D. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. Qual Life Res 2012; 21:739–46.

[88] Main CJ. Pain assessment in context: a state of the science review of the McGill pain questionnaire 40 years on. PAIN 2016;157:1387–99.

[89] Mathieson S, Maher CG, Terwee CB, de Campos TF, Lin CW. Neuropathic pain screening questionnaires have limited measurement properties. A systematic review. J Clin Epidemiol 2015;68:957–66.

[90] McNair DM, Lorr M, Droppleman LF. Profile of mood states. San Diego, CA: Educational and Industrial Testing Service, 1971.

[91] Mehta P, Claydon L, Hendrick P, Cook C, Baxter GD. Pain and physical functioning in neuropathic pain: a systematic review of psychometric properties of various outcome measures. Pain Prac 2016;16:195–508.

[92] Mehta P, Claydon L, Hendrick P, Winser S, Baxter GD. Outcome measures in randomized-controlled trials of neuropathic pain conditions: a systematic review of systematic reviews and recommendations for practice. Clin J Pain 2015;31:169–76.

[93] Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. PAIN 1975;1:277–99.

[94] Melzack R. The short-form McGill pain questionnaire. PAIN 1987;30:191–7.

[95] Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HCW. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10:22.

[96] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. International consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes: results of the COSMIN study. J Clin Epidemiol 2010;63:737–45.

[97] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW. 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;19:539–49.

[98] Murata S, Sawa R, Nakamura R, Isa T, Ebina A, Kondo Y, Tsuboi Y, Torizawa K, Fukuta A, Ono R. Association between objectively measured physical activity and the number of chronic musculoskeletal pain sites in community-dwelling older adults. Pain Med 2019;20:717–23.

[99] Murphy SL, Niemiec SS, Lyden AK, Kratz AL. Pain, fatigue, and physical activity in osteoarthritis: the moderating effects of pain-and fatigue-related activity interference. Arch Phys Med Rehabil 2016;97:S201–9.

[100] Nahin RL. Estimates of pain prevalence and severity in adults: United States, 2012. J Pain 2015;16:769–80.

[101] O'Brien EM, Staud RM, Hassinger AD, McCulloch RC, Craggs JG, Atchison JW, Price DD, Robinson ME. Patient-centered perspective on treatment outcomes in chronic pain. Pain Med 2010;11:6–15.

[102] Palermo TM, Kashikar-Zuck S, Friedrichsdorf SJ, Powers SW. Special considerations in conducting clinical trials of chronic pain management interventions in children and adolescents and their families. PAIN Rep 2019;4:e649.

[103] Patel KV, Allen R, Burke L, Farrar JT, Gewandter JS, Gilron I, Katz NP, Markman JD, Marshall SF, Resnick M, Rice AS. Evaluation of composite responder outcomes of pain intensity and physical function in neuropathic pain clinical trials: an ACTTION individual patient data analysis. PAIN 2018;159:2245–54.

[104] Patel KV, Dansie EJ, Turk DC. Impact of chronic musculoskeletal pain on objectively measured daily physical activity: a review of current findings. Pain Manage 2013;3:467–74.

[105] Patel KV, Guralnik JM, Dansie EJ, Turk DC. Prevalence and impact of pain among older adults in the United States: findings from the 2011 National Health and Aging Trends Study. PAIN 2013;154:2649–57.

[106] Pathak A, Sharma S, Jensen MP. The utility and validity of pain intensity rating scales for use in developing countries. Pain Rep 2018;3:e672.

[107] Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. Value Health 2011;14:967–77.

[108] Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. Value Health 2011;14:978–88.

[109] Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D; PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assess 2011;18:263–83.

[110] Portenoy R; ID Pain Steering Committee. Development and testing of a neuropathic pain screening questionnaire: ID pain. Curr Med Res Opin 2006;22:1555–65.

[111] Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. PAIN 1994;56:217–26.

[112] Price DD, Harkins SW. Combined use of experimental pain and visual analogue scales in providing standardized measurement of clinical pain. Clin J Pain 1987;3:1–8.

[113] Price DD, McGrath PA, Rafii A, Buckingham B. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. PAIN 1983;17:45–56.

[114] Price LR. Psychometric methods: theory and practice. New York: Guildford Publications, 2017.

[115] Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. PAIN 1996;6:3–8.

[116] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61:102–9.

[117] Revicki DA, Cook KF, Amtmann D, Harnam N, Chen WH, Keefe FJ. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. Qual Life Res 2014;23:245–55.

[118] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine 1983;8:141–4.

[119] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol 2014;67:516–26.

[120] Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. Value Health 2009;12:1075–83.

[121] Rowbotham M, Harden N, Stacey B, Bernstein P, Magnus-Miller L; Gabapentin Postherpetic Neuralgia Study Group. Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial. JAMA 1998;280:1837–42.

[122] Rowbotham MC, McDermott MP. Ethical considerations in the design, execution, and analysis of clinical trials of chronic pain treatments. PAIN Rep 2019;4:e646.

[123] Safikhani S, Gries KS, Trudeau JJ, Reasner D, Rüdell K, Coons SJ, Bush EN, Hanlon J, Abraham L, Vernon M. Response scale selection in adult pain measures: results from a literature review. J Patient Rep Outcomes 2018;2:40.

[124] Silverman DG, O'Connor TZ, Brull SJ. Integrated assessment of pain scores and rescue morphine use during studies of analgesic efficacy. Anesth Analg 1993;77:168–70.

[125] Smarr KL, Keefer AL. Measures of depression and depressive symptoms: beck depression inventory-II (BDI-II), center for epidemiologic studies depression scale (CES-D), geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), and patient health questionnaire-9 (PHQ-9). Arthritis Care Res 2011;63:S454–66.

[126] Smith SM, Amtmann D, Askew RL, Gewandter JS, Hunsinger M, Jensen MP, McDermott MP, Patel KV, Williams M, Bacci ED, Burke LB, Chambers CT, Cooper SA, Cowan P, Desjardins P, Etropolski M, Farrar JT, Gilron I, Huang IZ, Katz M, Kerns RD, Kopecky EA, Rappaport BA, Resnick M, Strand V, Vanhove GF, Veasley C, Versavel M, Wasan AD, Turk DC, Dworkin RH. Pain intensity rating training: results from an exploratory study of the ACTTION PROTECCT system. PAIN 2016;157: 1056–64.

[127] Smith SM, Chang RD, Pereira A, Shah N, Gilron I, Katz NP, Lin AH, McDermott MP, Rappaport BA, Rowbotham MC, Sampaio C, Turk DC, Dworkin RH. Adherence to CONSORT harms-reporting recommendations in publications of recent analgesic clinical trials: an ACTTION systematic review. PAIN 2012;153:2415–21.

[128] Smith SM, Hunsinger M, McKeown A, Parkhurst M, Allen R, Kopko S, Lu Y, Wilson HD, Burke LB, Desjardins P, McDermott MP, Rappaport BA, Turk DC, Dworkin RH. Quality of pain intensity assessment reporting: ACTTION systematic review and recommendations. J Pain 2015;16: 299–305.

[129] Smith SM, Jensen MP, He H, Kitt R, Koch J, Pan A, Burke LB, Farrar JT, McDermott MP, Turk DC, Dworkin RH. A comparison of the assay sensitivity of average and worst pain intensity in pharmacologic trials: an ACTTION systematic review and meta-analysis. J Pain 2018;19: 953–960.

[130] Smith SM, Wang AT, Katz NP, McDermott MP, Burke LB, Coplan P, Gilron I, Hertz SH, Lin AH, Rappaport BA, Rowbotham MC, Sampaio C, Sweeney M, Turk DC, Dworkin RH. Adverse event assessment, analysis, and reporting in recent published analgesic clinical trials: ACTTION systematic review and recommendations. PAIN 2013;154: 997–1008.

[131] Smuck M, Kao MC, Brar N, Martinez-Ith A, Choi J, Tomkins-Lane CC. Does physical activity influence the relationship between low back pain and obesity? Spine J 2014;14:209–16.

[132] Stone AA, Schneider S, Broderick JE, Schwartz JE. Single-day pain assessments as clinical outcomes: not so fast. Clin J Pain 2014;30: 739–43.

[133] Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the brief pain inventory for chronic nonmalignant pain. J Pain 2004;5:133–7.

[134] Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, Clauw DJ, Gignac MA, Markman JD, Williams DA, Bujanover S, Burke LB, Carr DB, Choy EH, Conaghan PG, Cowan P, Farrar JT, Freeman R, Gewandter J, Gilron I, Goli V, Gover TD, Haddox JD, Kerns RD, Kopecky EA, Lee DA, Malamut R, Mease P, Rappaport BA, Simon LS, Singh JA, Smith SM, Strand V, Tugwell P, Vanhove GF, Veasley C, Walco GA, Wasan AD, Witter J. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. PAIN 2016;157:1836–50.

[135] Treister R, Lawal OD, Shecter JD, Khurana N, Bothmer J, Field M, Harte SE, Kruger GH, Katz NP. Accurate pain reporting training diminishes the placebo response: results from a randomised, double-blind, crossover trial. PLoS One 2018;13:e0197844.

[136] Trudeau J, Van Inwegen R, Eaton T, Bhat G, Paillard F, Ng D, Tan K, Katz NP. Assessment of pain and activity using an electronic pain diary and actigraphy device in a randomized, placebo-controlled crossover trial of celecoxib in osteoarthritis of the knee. Pain Pract 2015;15:247–55.

[137] 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; 106:337–45.

[138] Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, Allen RR, Atkinson HJ, Chandler J, Cleeland C, Cowan P. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. PAIN 2006;125:208–15.

[139] Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, Cleeland CS, Cowan P, Farrar JT, Hertz S, Max MB, Rappaport BA. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. PAIN 2008;137:276–85.

[140] Turk DC, Dworkin RH, Trudeau JJ, Benson C, Biondi DM, Katz NP, Kim M. Validation of the hospital anxiety and depression scale in patients with acute low back pain. J Pain 2015;16:1012–21.

[141] Turk DC, Fillingim RB, Ohrbach R, Patel KV. Assessment of psychosocial and functional impact of chronic pain. J Pain 2016;17:T21–49.

[142] Turner KV, Moreton BM, Walsh DA, Lincoln NB. Reliability and responsiveness of measures of pain in people with osteoarthritis of the knee: a psychometric evaluation. Disabil Rehabil 2017;39:822–9.

[143] US Food and Drug Administration. Center for drug evaluation and research patient-focused drug development. Available at: https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm579400.htm. Accessed September 1, 2018.

[144] US Food and Drug Administration. Clinical outcome assessment (COA): glossary of terms. Available at: http://wayback.archive-it.org/7993/20180424212148/https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm#pro. Accessed September 1, 2018.

[145] US Food and Drug Administration. Guidance for industry and FDA staff: qualification process for drug development tools. 2014. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf. Accessed September 1, 2018.

[146] US Food and Drug Administration. Guidance for industry: analgesic indications: developing drug and biological products. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM384691.pdf. Accessed December 19, 2018.

[147] US Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. Available at: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed September 1, 2018.

[148] US Food and Drug Administration. Multiple endpoints in clinical trials: draft guidance. 2017. Available at: https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM536750.pdf. Accessed December 15, 2018.

[149] US Food and Drug Administration. Clinical outcome assessment (COA) program. Available at: https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm. Accessed December 31, 2018.

[150] US Food and Drug Administration. Patient focused drug development. Glossary. Available at: https://www.fda.gov/Drugs/DevelopmentApprovalProcess/ucm610317.htm. Accessed December 31, 2018.

[151] Victor TW, Jensen MP, Gammaitoni AR, Gould EM, White RE, Galer BS. The dimensions of pain quality: factor analysis of the Pain Quality Assessment Scale. Clin J Pain 2008;24:550–5.

[152] Visser AW, Bøyesen P, Haugen IK, Schoones JW, van der Heijde DM, Rosendaal FR, Kloppenburg M. Instruments measuring pain, physical function, or patient's global assessment in hand osteoarthritis: a systematic literature search. J Rheumatol 2015, 42:2118–34. jrheum-141228.

[153] Walton DM, Beattie T, Putos J, MacDermid JC. A Rasch analysis of the Brief Pain Inventory Interference subscale reveals three dimensions and an age bias. J Clin Epidemiol 2016;74:218–26.

[154] Walton DM, Putos J, Beattie T, MacDermid JC. Confirmatory factor analysis of 2 versions of the Brief Pain Inventory in an ambulatory population indicates that sleep interference should be interpreted separately. Scand J Pain 2016;12:110–6.

[155] Walton MK, Powers JH III, Hobart J, Patrick D, Marquis P, Vamvakas S, Isaac M, Molsen E, Cano S, Burke LB. Clinical outcome assessments: conceptual foundation—report of the ISPOR clinical outcomes assessment–emerging good practices for outcomes research task force. Value Health 2015;18:741–52.

[156] Wang JL, Zhang WJ, Gao M, Zhang S, Tian DH, Chen J. A cross-cultural adaptation and validation of the short-form McGill Pain Questionnaire-2: Chinese version in patients with chronic visceral pain. J Pain Res 2017; 10:121.

[157] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 30; 1992:473–83.

[158] Wasan AD, Michna E, Edwards RR, Katz JN, Nedeljkovic SS, Dolman AJ, Janfaza D, Isaac Z, Jamison RN. Psychiatric comorbidity is prospectively associated with diminished opioid analgesia and increased opioid misuse in patients with chronic low back pain. Anesthesiol 2015;123:861–72.

[159] de C Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. PAIN 2000;85:457–63.

[160] Williams MR, McKeown A, Pressman Z, Hunsinger M, Lee K, Coplan P, Gilron I, Katz NP, McDermott MP, Raja SN, Rappaport BA. Adverse event reporting in clinical trials of intravenous and invasive pain treatments: an ACTTION systematic review. J Pain 2016;17:1137–49.

[161] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70.