Article:
Langford, D.J., Mark, R.P., France, F.O. et al. (11 more authors) (2024) Use of patient-reported global assessment measures in clinical trials of chronic pain treatments: ACTTION systematic review and considerations. Pain, 165 (11). pp. 2445-2454. ISSN 0304-3959

https://doi.org/10.1097/j.pain.0000000000003270

Use of patient-reported global assessment measures in clinical trials of chronic pain treatments: ACTTION systematic review and considerations

Dale J Langford, PhD[1,2,3]*

Remington P. Mark, DO[2]

Fallon O. France, MD[2]

Mahd Nishtar, MD[2]

Meghan Park, MD[2]

Sonia Sharma, PhD[4]

Isabel C. Shklyar[1,9]

Thomas J. Schnitzer, MD, PhD[5]

Philip G. Conaghan, MBBS, PhD, FRACP, FRCP[6]

Dagmar Amtmann, PhD[7]

Bryce B. Reeve, PhD[8]

Dennis C. Turk, PhD[3]

Robert H. Dworkin, PhD[1,2]

Jennifer S. Gewandter, PhD, MPH[2]

[1]Department of Anesthesiology, Critical Care & Pain Management, Pain Prevention Research Center at Hospital for Special Surgery, New York, NY

[2]Department of Anesthesiology & Perioperative Medicine, University of Rochester, Rochester, NY

[3]Department of Anesthesiology & Pain Medicine, University of Washington, Seattle, WA, USA

[4]Neuro Pain Management Center, Department of Neurosurgery, University of Rochester, Rochester, NY

[5]Department of Physical Medicine & Rehabilitation, Northwestern University, Chicago, IL, USA

[6]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, and NIHR Leeds Biomedical Research Centre, Leeds, United Kingdom

[7]Department or Rehabilitation Medicine, University of Washington, Seattle, WA, USA

[8]Center for Health Measurement, Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, USA

[9]The University of Texas at Austin

***Corresponding Author:**
Dale J. Langford, PhD
Department of Anesthesiology, Critical Care & Pain Management
Pain Prevention Research Center at Hospital for Special Surgery
Hospital for Special Surgery
535 East 70th Street, New York, NY 10021
langfordd@hss.edu
(212) 774-7583

Number of manuscript pages: 21**; Number of figures:** 3**; Number of tables:** 5

Data availability statement: Data available upon reasonable request.

Abstract

Establishing clinically meaningful changes in pain experiences remains important for clinical trials of chronic pain treatments. Regulatory guidance and pain measurement initiatives have recommended including patient-reported global assessment measures (e.g., Patient-Global Impression of Change [PGIC]) to aid interpretation of within-patient differences in domain-specific clinical trial outcomes (e.g., pain intensity). The objectives of this systematic review were to determine the frequency of global assessment measures inclusion, types of measures, domains assessed, number and types of response options, and how measures were analyzed. Of 4,172 abstracts screened across 6 pain specialty journals, we reviewed 96 clinical trials of chronic pain treatments. Fifty-two (54.2%) studies included a global assessment measure. The PGIC was most common (n=28; 53.8%), with relatively infrequent use of other measures. The majority of studies that used a global assessment measure (n=31; 59.6%) assessed change or improvement in an unspecified domain. Others assessed overall condition severity (n=9; 17.3%), satisfaction (n=8; 15.4%), or overall health status/recovery (n=5; 9.6%). The number, range, and type of response options were variable and frequently not reported. Response options and reference periods even differed within the PGIC. Global assessment measures were most commonly analyzed as continuous variables (n=24; 46.2%) or as dichotomous variables with positive categories combined to calculate the proportion of participants with a positive response to treatment (n=18; 34.6%). This review highlights the substantial work necessary to clarify measurement and use of patient global assessment in chronic pain trials, and provides short and longer-term considerations for measure selection, reporting and analysis, and measure development.

1. Introduction

Establishing clinically meaningful changes in pain experiences remains an important and inadequately defined issue in clinical trials of chronic pain treatments. Regulatory guidance emphasizes the importance of using the patient perspective to establish meaningful clinical outcomes and further recommends the use of anchor-based methods, in which a global assessment measure, such as the Patient Global Impression of Change (PGIC)[37] or the Patient Global Impression of Severity (PGIS)[45], is used to define whether a patient has experienced a change in their condition (i.e., within-participant change) [41]. Moreover, patient-reported measures of global assessment have long been recommended as core outcome measures for clinical trials of pain-related treatments [14; 17]. However, whether and how patient-reported global assessment measures are typically used in clinical trials of chronic pain treatments remains unclear.

There are a multitude of patient global assessment measures, including (but not limited to) the PGIC, PGIS, Global Perceived Effect Scale (GPE) (reviewed in [27] and [4; 5]), Global Rating of Change (GROC) (reviewed in [4; 5]), Patient Global Assessment (PGA) (reviewed in [32]), and the Patient Acceptable Symptom State (PASS) [28]. Yet, variability can exist in question stems, reference periods, and response options; for example, instructions may direct patients to consider their overall health status, overall pain, quality of life, function, condition severity, and so on. Furthermore, response options may vary in number, type, and content. Reference periods also vary, especially for PGIC-type questions for which the respondent compares their current state versus their state prior to treatment. This lack of standardization makes it difficult to establish valid and informative conclusions about meaningful change in pain both within and across studies. A 2016 review of patient global assessment of rheumatoid arthritis disease activity highlighted variability in the content of various assessment measures and the potential impact on interpretation and resultant validity and utility of the outcome (e.g., achieving treat-to-target goals) [32]. A systematic review of the use of investigator global assessment measures in trials of atopic dermatitis treatment also described lack of standardization in the types of measures used and how the measures were analyzed [19]. Typically, the intent of global assessment measures is to capture the degree of participant-perceived change in a broader, more nuanced, construct that can allow researchers to (1) determine whether individual participants have experienced a perceptible or

important change in their condition and (2) support interpretation of observed changes in outcome scores. Draft regulatory guidance outlines an approach for incorporating global assessment measures into trial analysis to demonstrate "meaningful within-patient change" [43] ("meaningful" or "clinically important" between-group differences is a separate and challenging issue [13] and not the focus herein). However, it is not clear whether this approach is routinely adopted in existing chronic pain trials.

Given the important role that patient global assessment can play in clinical trials, drug approvals, and cost-effectiveness analyses, clarifying how global assessment measures are used in clinical trials of chronic pain treatments, and ensuring measures are valid and reliable, is critical.

The primary objectives of this review were to describe the (1) proportion of trials that incorporated a patient-global assessment measure; (2) types of measures used; (3) domains assessed (e.g., pain, overall health status), and (4) how global assessment measures were operationalized. Our hope is that this review serves to improve the standardized use of current global assessment measures, provides the impetus for establishing psychometric properties of commonly used measures, and lays the groundwork for developing novel content-valid global assessment measures that validly capture domains important to patients.

2. Methods

This targeted systematic review (targeted given its focus on six pain specialty journals) was registered in Open Science Framework (08/02/2023).

2.1 Article selection

Articles that reported efficacy results of randomized clinical trials (RCTs) of at least one investigational treatment for chronic pain, published over between January of 2016 and September of 2023, were identified across six major pain specialty journals, *PAIN, Pain Reports, Clinical Journal of Pain*, *European Journal of Pain*, *Journal of Pain,* and *Pain Medicine,* identified using a highly sensitive hedge for RCTs and filtered by year of publication and journal (see Appendix 1 for search strategy). Articles were included if pain was a main outcome and there were at least 50 participants per comparison group [31]. Articles that described a secondary analysis or a protocol of an RCT, did not include an active arm (i.e., comparing different placebo conditions), focused on

acute pain or experimental pain in healthy participants, or investigated a preventive treatment (e.g., preoperative intervention to prevent persistent postsurgical pain) were not included in this review.

2.2 Data extraction

A coding template was used to extract both trial characteristics and information about the use of global assessment measures. The coding template was developed and pretested by 2 authors (J.S.G. and D.J.L.). Pretesting included independent coding of 7 randomly selected articles and discussion of discrepancies, in 3 rounds. This pretesting informed modifications to the template in order to improve clarity and content (see Appendix 2 for final coding manual). Six coders (R.P.M., F.O.F., M.N., M.P., S.S., I.C.S.) were trained on data extraction and the use of the online coding template and underwent 2 rounds of practice coding, comparing their coding to consolidated coding of J.S.G. and D.J.L. Each round of practice coding was followed by an overview of discrepancies (compared to consolidated coding of J.S.G. and D.J.L.) and discussion with J.S.G. and/or D.J.L. Additional explanatory statements were incorporated into the coding template based on training experience.

Trial characteristics extracted included: trial design, blinding, type of intervention, sample size, number of groups, targeted pain condition(s), and study sponsor. Coders recorded whether a study incorporated a global assessment measure(s) as a primary or secondary outcome. Global assessment measures were defined to reviewers as measures that capture a patient's overall impression/rating of change or improvement in health status, function, etc. They may also measure a patient's impression of overall severity, tolerability or acceptability of pain, or overall satisfaction with treatment or their current condition. Some examples were provided and reviewers were reminded that global assessment measures do not include pain intensity rating scales, other symptom-specific measures, or measures with multiple subscales, and that most existing measures were single-items. If so, the type of measure(s), domains specified by the measure(s), and the number and type of response options were recorded. After training and practice, pairs of coders were randomly assigned a subset of articles, in a randomly selected order. Data were collected using REDCap (Research Electronic Data capture hosted at University of Rochester [22; 23]). After all articles were coded independently, D.J.L. reviewed the data

for discrepancies. If the discrepancy was due to obvious oversight, it was noted and corrected. If the discrepancy was due to differences in interpretation, it was reviewed and resolved in consult with J.S.G. D.J.L. performed a post hoc review of articles that used a global assessment measure in order to determine how the measure was analyzed (e.g., continuous, categorical, dichotomized to identify responders). Descriptive statistics were used to summarize the review findings. No formal statistical comparisons were conducted.

3. Results

Figure 1 displays the PRISMA flow diagram for article selection (see Appendix 3 for list of reviewed studies). Ultimately, after removing duplicates and screening titles/abstracts, and full texts, 96 studies met inclusion criteria and were reviewed.

3.1 Coder discrepancies

A total of 1,254 items were coded in duplicate. There were 177 (14.1%) coding differences between the coders among trial characteristics and global assessment measure coding items. Of these differences, 172 (97.2%) were due to oversight by one of the coders and 5 (2.8%) were due to differences in interpretation.

3.2 Trial characteristics

Table 1 shows the trial characteristics of the 96 reviewed studies. The majority of the reviewed trials used a parallel group design to evaluate superiority of an intervention. Interventions were most commonly non-pharmacologic (55.2%) or pharmacologic (31.3%), with an average sample size of 112 per group. Targeted pain conditions were variable, with chronic low back pain (24.0%), diverse chronic pain conditions (19.8%), or peripheral neuropathic pain (14.6%) being the most common.

3.3 Global assessment measures

As shown in Table 2, 52 (54.2%) of 96 included studies incorporated at least one patient-reported global assessment measure, most commonly as a secondary outcome (94.2%). Most of these studies (76.9%) used one global assessment measure; others that used 2 (11.5%) or 3 or more (11.5%) evaluated more than one domain separately (e.g., 3 single-item numeric rating scales for perceived improvement in pain, function, and overall quality of life [3] or pain intensity, pain interference, and ability to cope with pain [46], or separate global

assessment measures for change and satisfaction [8] . The PGIC was most common, used in 28 studies (53.8%), followed by the Global Perceived Effect (GPE) scale (n=7, 13.5%). Other measures were variably and infrequently used, and generally captured patient-perceived change (worsening or improvement) (15.4%), improvement only (5.8%), or satisfaction with study treatment (21.2%). More than half of the studies (n=31; 59.6%) asked participants to rate overall change or improvement without specifying domains to consider in their ratings. Other study measures asked about illness/condition severity (n=9, 17.3%), satisfaction with treatment (n=8, 15.4%), or overall health status or recovery (n=5, 9.6%).

All measures were comprised of a single item, with Likert type, verbal response, or numeric rating scales. Table 3 outlines the number, range, and lower and upper anchor descriptions for the response options of the global measures. The number of response options was variable (ranging from 3 to 101 [e.g., 0-100% satisfaction scale]), and the number, range, and anchors often not reported (n=15, 28.8%; n=10, 19.2%, and n=19, 36.5%, respectively). Most commonly, studies included a global assessment measure with 7 response options (n=20, 38.5%), ranging from 1 to 7 (n=13, 25.0%), with lower and upper anchors of “very much worse” to “very much improved” (n=14; 26.9%) – all typical characteristics of the PGIC; however, lower and upper anchors for the commonly used PGIC varied across studies (Table 4), with four differently worded combinations of lower and upper anchors.

As presented in Table 5, global assessment measures were most commonly analyzed as a continuous outcome (n=24, 46.2%) or as a dichotomous outcome with positive categories combined (e.g., “much improved” and “very much improved”) to calculate the proportion of participants with a positive response to treatment (n=18, 34.6%). Other approaches, such as calculating the proportion of participants endorsing each possible response option or that combined negative response options to calculate the proportion of participants with a negative response to treatment, were rare (n=8, 15.4% and n=3, 7.3%, respectively).

4. Discussion

These findings demonstrate substantial variability in the application of global assessment measures, even for a single, commonly used measure (e.g., PGIC). The domains assessed by the global measures were also largely unspecified, and the measures were often not clearly described.

Just over half of the studies reviewed included a global assessment measure, indicating that despite regulatory guidance [41] and measurement initiative recommendations [12; 14; 17], these measures were not consistently included in chronic pain trials published in six major pain specialty journals. For the trials that did include a global assessment, it was not surprising that the PGIC was the most commonly used measure, given that existing regulatory guidance and IMMPACT recommendations specifically endorse the PGIC. It is important to note, however, that we did not observe uniformity in the details of the PGIC (e.g., number, range, response options, analysis), suggesting the need for improved clarity and standardization. In fact, the FDA's Patient-Focused Drug Development workshop guidance discussion document provides an example of the PGIC that includes 5 response options "much better", "a little better", "no change", "a little worse" and "much worse", and allows investigators to specify the domain ("symptom, overall status, etc."). Yet, we more commonly observed a 7-point PGIC ranging from "very much improved" to "very much worse." Moreover, the number and types of response options across all measures were infrequently reported (n = 15, 28.8%; n = 19, 36.5%, respectively). Unreported response options present a particular challenge for understanding the respondent's assessment; for example, if no lower anchor is reported, it cannot be determined whether the lower anchor denotes worsening or no change/stability. A post hoc review of studies that used the PGIC revealed that instructions about what starting point participants should evaluate change from was most often not reported (n=19; 67.9%). For those studies that reported the time frame to consider change, all used variably worded instructions (e.g., "since baseline", "since starting study drug", "since beginning treatment", "since the start of the study"). Importantly, there was also substantial variability in the time points at which assessment was completed (i.e., from 5 days to 12 months), and almost half of these studies (n=12; 42.9%) involved PGIC completion at multiple time points.

The majority of studies used measures that asked participants to rate their overall change or improvement without specifying what domains to consider in their rating (e.g., pain overall, function, quality of life, combination thereof). This lack of specification may be purposeful in order to allow the participant to conceptualize their experience in their own individual and reflexively meaningful way. It is also possible that domains were specified to participants directly and that this lack of information was due to deficiencies in reporting. Nonetheless, these abstract question stems call into question the content validity of the PGIC. What domains one considers when rating making a global assessment are likely highly individual and variable, and may reflect changes in other domains (e.g., physical activity, mood) over and above pain [39]. However, changes in these domains may be meaningful even within the context of stable pain intensity.

Overall health status and illness/condition severity were the next most common domains assessed. These constructs are quite distinct (i.e., overall health vs. specific illness/condition), making it difficult to even understand the domains being assessed or to aggregate findings across studies. A review of the use of PGA in rheumatoid arthritis noted that patient assessment of "global health" and "disease activity" (each commonly used) capture different information about an individual's status and suggest that the context should be considered, noting that global health can encapsulate comorbid conditions and psychological distress, whereas disease activity aims to capture more specific disease processes [32].

Considerable variability in terms of how global assessment measures were analyzed was also observed. Issues with the various approaches to analysis and considerations for future research are outlined below.

4.1 Considerations

Given that regulatory guidance and IMMPACT recommend the PGIC, coupled with its already widespread use, a short-term recommendation is to include the PGIC in trials of chronic pain treatments, but with standardized and/or clearly reported question stems, reference periods, response options, and analytic approaches. However, it should be acknowledged that the retrospective nature of the PGIC presents some concerning challenges, including (1) the possibility of inaccurately recalling a prior state, particularly in trials of longer duration and (2) that respondents' state at the time of completion may unduly influence their rating of

change. A longitudinal study on the validity of the PGIC in clinical care of fibromyalgia demonstrated that longer follow-up duration (compared to shorter duration with similar change in other specific disease measures) was associated with lower PGIC ratings (i.e., less improvement) [37]. Of note, FDA guidance acknowledges these challenges and proposes an alternative measure, the Patient Global Impression of Severity (PGIS), which assesses the participant's current state ("symptom/overall status/etc" [42]) at two time points (i.e., baseline and post-treatment) [41]. However, as discussed above, what constitutes a meaningful change in severity presents its own challenge.

It is clear from these findings of infrequent and variable use of patient-reported global assessment measures that a considerable amount of work is needed to ensure that measures are valid, reliable, interpretable and meaningful to patients and clinicians. Adhering to regulatory guidance on the development of patient-reported outcome measures requires that the measure captures a specific "concept of interest" for a specific "context of use," that the measure is developed using both qualitative and quantitative methodologies, and that the measure undergoes rigorous evaluation [33; 34; 41]; however, the paradoxical issue is that such guidance has not generally been applied to or required of existing patient-reported global assessment measures.

It will be important to establish interpretability and content validity of global assessment measures from the patient's perspective, including what domains are important and how these domains should be described (e.g., "overall change in pain, function, and quality of life", "tolerability of pain", "satisfaction with pain relief", current status or global severity of the pain condition). While efforts have been made to evaluate construct validity of global assessment measures of changes in specific symptoms (e.g., hand [2], hip, knee, and back [1] in the context of OA), little research has focused on establishing psychometric properties of more general global assessment measures, like the PGIC, across chronic pain disorders. In fact, of the 28 studies that used the PGIC, only 3 papers noted its psychometric properties [24; 44; 46]. One of these papers included a blanket statement regarding "adequate psychometric properties" of all measures used in the study, without specific references [24]. Another paper [46] stated that psychometric data were purposefully not provided because the PGIC was not designed to measure a single underlying construct, citing Scott & McCracken . A third paper [44] reported

that the PGIC was validated among chronic pain populations, citing: (1) a conference abstract, which found that only 31% of the variance in post-treatment PGIC scores was accounted for by changes in pain intensity, mood, or function, highlighting the important contribution of additional factors [18], (2) the Scott & McCracken paper that was cited above as rationale for not reporting psychometric properties , and (3) two studies that evaluated convergent validity of the PGIC in the setting of clinical care for fibromyalgia [37] and peripheral neuropathic pain [36]. These two studies[36; 37] examined the correlation of the PGIC scores with changes in relevant outcome measures (e.g., pain intensity, pain interference, fatigue, mood, quality of life) over the course of routine clinical care; while correlations were statistically significant and in the expected (positive) direction, associations were moderate at best ($r < 0.5$). Moreover, these findings may not necessarily generalize to the clinical trial setting. Scott and McCracken [39] used the PGIC to evaluate patients' impression of change overall and separately for specific domains (e.g., mood, pain, physical functioning) following Acceptance and Commitment Therapy (ACT) for chronic pain, and found that overall ratings were associated with improvement in mood and physical activities, rather than pain [39]. While this finding might be related to the type of intervention (i.e., ACT encourages acceptance of pain and activity engagement despite pain), it further highlights the need to understand what domains are relevant to the target indication/outcome and the intervention under study, and/or what domains are meaningful to the patient in deriving global assessments that are clinically relevant and informative. A recently developed multidimensional global impression of change (MGIC) measure expanded the PGIC to include 7 additional specified domains: pain, mood, sleep, physical functioning, coping with pain, managing pain flare-ups, medication effectiveness [20]. The MPIC accounted for a greater proportion of variance in treatment outcomes (e.g., pain intensity, disability, depression) than the traditional PGIC, suggesting that specifying domains may be advantageous.

Figure 2 summarizes the continuum of specificity in domains, but also conveys the challenge of selecting a domain that is both clinically meaningful and sensitive to change (i.e., can show treatment effect if one exists). It could reasonably be debated whether broader measures are more meaningful given that they encapsulate a more holistic experience that includes pain, function, quality of life, and so on, or whether specific measures are

more meaningful given they capture the target indication/reason for treatment. Perhaps, incorporating both broad and specific global measures (e.g., as in [35]) or making a context-informed decision about what measure to use (and providing rationale for that choice) would be a worthwhile approach. Similar rationale may apply to sensitivity to change.

Another consideration is whether and how participants incorporate treatment-related adverse effects (e.g., sedation, constipation, dizziness, nausea) in their global assessments. If harms are not naturally and consistently considered by patients, it may prove useful to include and report instructions about making assessments that incorporate both benefits and harms of a treatment. This information would be useful, in particular, for comparative effectiveness trials, and if standardized, for aggregating findings across studies to support meta-analyses and to inform treatment guidelines. While not explicit, a couple of interesting measures identified in this review that may naturally capture both benefit and harms (and potentially convenience) included participant ratings of their likelihood of undergoing treatment again [10] and whether they would recommend the treatment to someone with similar difficulties [6]. However, it would still be necessary to learn how people interpret these questions, what factors impact their responses, and in particular whether and how they would consider benefit and harms in their responses. A study that evaluated 5 lidocaine patch clinical trials for knee OA and chronic low back pain found that pain intensity, relief, and interference with physical function, and *not* occurrence or severity of adverse events (or sleep or interference with emotional function), made significant independent contributions to patient global assessment of treatment satisfaction scores [11], suggesting either minimal/mild harms in this context (i.e., 5% lidocaine patch) that did not affect participants overall experience or the need to provide explicit instructions to participants to consider both benefit and harms in their global assessments.

Development of novel global assessment measures may also be warranted. An ideal measure would adequately encapsulate and specify the domains that patients should include in their assessment of change and include response options that are differentiable by patients with clearly interpretable levels that lend themselves well to establishing meaningful change. Systematic evaluation of psychometric properties for

existing or newly developed global assessment measures should be a priority, as inadequate reliability and validity of global assessment tools can have a substantial impact on the informativeness of a clinical trial.

Understanding the advantages and disadvantages of various analytic approaches and in which settings each are appropriate will be important. We found that the most common approach was to compare mean global assessment scores between treatment groups, but whether and how a statistically significant difference in these measures is clinically meaningful remains unclear. In fact, it might be argued that this approach offers no measurable information beyond the analysis of the primary outcome itself, as it presents a similar challenge of defining a meaningful difference. However, having multiple measures that are different between groups, including one that asks about global or overall status, may add to the robustness of the conclusion of a treatment effect.

Another issue is that many measures like the PGIC are ordinal in nature, not continuous; therefore, comparing mean scores on such global assessment measures (which was done by six of the studies under review) may not yield valid or interpretable findings. In the case of collapsing positive categories to identify and compare the proportion of responders, further work is necessary to define a positive response. For example, is it appropriate (and when) to combine "minimally improved", "much improved" and "very much improved" to reflect achievement of a meaningful score difference or is it most appropriate to combine "much improved" and "very much improved" responses to denote more substantial "clinically meaningful" improvement? The PGIC was used to estimate "clinically meaningful" change in pain intensity by evaluating the association between absolute and percent change in pain intensity with PGIC responses among 10 placebo-controlled clinical trials. A change of 2-points on the 0-10 NRS or 30% was determined to be clinically meaningful defined by a PGIC score of "much improved" or better [16]. However, it is also possible that even slight improvement ("minimally improved") constitutes a meaningful change for an individual living with chronic pain, perhaps particularly when there are no, or minimal, side effects.

Regardless, the approach of comparing the proportion of participants who report improvement between study groups is easily interpretable and clinically relevant [15]. However, as with dichotomizing any variable,

some information is lost and there may be less power to detect a statistically significant difference. Only a small number of studies collapsed negative categories to identify the proportion of participants who experienced a negative response to treatment [26; 30; 40]; each of these studies also reported the proportion of positive responders. This approach may be particularly valuable for interventions with some anticipated harms or for comparative effectiveness trials (e.g., [30]).

FDA draft guidance outlines a more comprehensive approach for using global assessment measures to provide support that observed differences in a primary outcome are meaningful [43]. This FDA guidance encourages evaluation of (1) the distribution of global assessment response options (i.e., proportion of respondents who endorsed each of the response options), an approach adopted by eight of the reviewed studies [7-10; 25; 35; 38; 46]), and (2) the distribution of change in primary outcome scores among each of the response options by treatment group (an approach that none of the reviewed studies used) as a means of interpreting within-patient change in the outcome among treatment groups.

Figure 3 summarizes short- and longer-term considerations for global assessment measure selection, reporting and analysis, and development of new measures.

4.2 Limitations

Limitations of this review should be noted. We limited our search to six pain specialty journals and so our findings may not generalize to all chronic pain trials; for example, trials focused on arthritis published in rheumatology specialty journals may be more consistent given OMERACT's published guidelines on how and what global assessment measures to include in arthritis trials. However, the objective of this review was to efficiently and effectively assess the landscape of chronic pain trials across diverse pain conditions published among major pain journals. We also limited our search to a 7-year period to capture recent trends in the use of global assessment measures. It is possible that articles published prior to 2016 may show different results, particularly those published prior to 2005, when IMMPACT published recommendations on core outcome measures to include in clinical trials of chronic pain treatments [12]. Nonetheless, the current review describes

recent trends in the use of global assessment measures across a range of interventions evaluated amongst a diverse sample of chronic pain conditions.

4.3 Conclusion

Overall, the findings of this review highlight a substantial amount of work to be done to clarify the development and application of patient-reported global assessment measures in chronic pain clinical trials. Efforts to establish psychometric properties of existing measures, to understand domains important to patients, to accurately and consistently describe those domains, and to address shortcomings in existing measures are warranted. Given the important role that patient global assessment can play in the outcome of clinical trials, regulatory approvals, and clinical practice, combined with limited treatment options for chronic pain, such efforts are critical.

Acknowledgements

The authors wish to acknowledge University of Rochester Miner Library's Reference and Education Services Librarian, Ms. Linda Hasman, for her assistance with the article search strategy, as well as Soroush Besherat, Raissa Lou, and Rachel De Guzman for help with article screening.

This study was supported the Hospital for Special Surgery Department of Anesthesiology, Critical Care, and Pain Management's Research and Education Fund, and by funding from the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks (ACTTION; 1U01FD007702) public-private partnership, and by. The views expressed in this article are those of the authors, and no official endorsement by the Food and Drug Administration (FDA) or the companies that provided support for the ACTTION public-private partnership should be inferred. The ACTTION public-private partnership has received research contracts, grants, or other revenue from the FDA, multiple pharmaceutical and device companies, philanthropy, royalties, and other sources (a list of industry sponsors can be found at https://www.acttion.org/partners). During than the past 3 years, DJL has received funding from the US Food and Drug Administration. DCT has received in the past 5 years research grants and contracts from the US Food and Drug Administration and the US National Institutes of Health, U.S. Patient-Centered Outcome Research Institute, and US National Center for Occupational Health and Safety; received compensation for serving on advisory boards from Eli Lilly, GlaxoSmithKline, Novartis, and Pfizer. He received and honorarium from Wolters Kluwer in his role as editor-in-chief of The Clinical Journal of Pain. PGC is funded in part through the UK National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. TJS has received grants or contracts from Pfizer, Regeneron, Eli Lilly, Amgen, Grunenthal, Galapagos, Novartis, Techfields, and Paradigm, and consulting fees from PFizer, Lilly, Vertex, GlaxoSmithKline, Biosplice, Collegium, Acadia, Grunenthal, Techfields, Galapagos, Unity. Moebius, Paradigm, Moebius, and Paradigm. RHD has received in the past 5 years research grants and contracts from the US Food and Drug Administration and the US National Institutes of Health, and compensation for serving on advisory boards or consulting on clinical trial methods from Abide, Acadia, Adynxx, Analgesic Solutions, Aptinyx, Aquinox, Asahi Kasei, Astellas, Beckley, Biogen, Biohaven, Biosplice, Boston Scientific, Braeburn, Cardialen, Centrexion, Chiesi, Chromocell, Clexio, Collegium, CoimbiGene, Confo, Decibel, Editas, Eli Lilly, Endo, Ethismos (equity), Eupraxia, Exicure, GlaxoSmithKline, Glenmark, Gloriana, Hope, Kriya, Lotus, Mainstay, Merck, Mind Medicine (also equity), Neumentum, Neurana, NeuroBo, Novaremed, Novartis, OliPass, Orion, Oxford

Cannabinoid Technologies, Pfizer, Q-State, Reckitt Benckiser, Regenacy (also equity), Rho, Sangamo, Sanifit, Scilex, Semnur, SIMR Biotech, Sinfonia, SK Biopharmaceuticals, Sollis, SPM Therapeutics, SPRIM Health, Teva, Theranexus, Vertex, Vizuri, and WCG. JSG has received (unrelated to this work) grant support from the NIH and Neurometrix, consulting income from AlgoTX, GW pharmaceuticals, Eikonozo, and Saluda, and shares for consulting from Eisana Corp. The remaining authors have no conflicts of interest related to this work to disclose.

References

[1] Allen KD, Jordan JM, Doherty M, Renner JB, Kraus VB. Performance of global assessments of hip, knee, and back symptom change. Clin Rheumatol 2011;30(3):331-338.

[2] Allen KD, Jordan JM, Renner JB, Kraus VB. Relationship of global assessment of change to AUSCAN and pinch and grip strength among individuals with hand osteoarthritis. Osteoarthritis Cartilage 2006;14(12):1281-1287.

[3] Bennell KL, Nelligan RK, Rini C, Keefe FJ, Kasza J, French S, Forbes A, Dobson F, Abbott JH, Dalwood A, Harris A, Vicenzino B, Hodges PW, Hinman RS. Effects of internet-based pain coping skills training before home exercise for individuals with hip osteoarthritis (HOPE trial): a randomised controlled trial. Pain 2018;159(9):1833-1842.

[4] Bobos P, MacDermid J, Nazari G, Furtado R, Catwad. Psychometric properties of the global rating of change scales in patients with neck disorders: a systematic review with meta-analysis and meta-regression. BMJ Open 2019;9(11):e033909.

[5] Bobos P, Ziebart C, Furtado R, Lu Z, MacDermid JC. Psychometric properties of the global rating of change scales in patients with low back pain, upper and lower extremity disorders. A systematic review with meta-analysis. J Orthop 2020;21:40-48.

[6] Boersma K, Sodermark M, Hesser H, Flink IK, Gerdle B, Linton SJ. Efficacy of a transdiagnostic emotion-focused exposure treatment for chronic pain patients with comorbid anxiety and depression: a randomized controlled trial. Pain 2019;160(8):1708-1718.

[7] Casey MB, Smart KM, Segurado R, Hearty C, Gopal H, Lowry D, Flanagan D, McCracken L, Doody C. Exercise combined with Acceptance and Commitment Therapy compared with a standalone supervised exercise programme for adults with chronic pain: a randomised controlled trial. Pain 2022;163(6):1158-1171.

[8] Castro J, Correia L, Donato BS, Arruda B, Agulhari F, Pellegrini MJ, Belache FTC, de Souza CP, Fernandez J, Nogueira LAC, Reis FJJ, Ferreira AS, Meziat-Filho N. Cognitive functional therapy compared with core exercise and manual therapy in patients with chronic low back pain: randomised controlled trial. Pain 2022;163(12):2430-2437.

[9] Dear BF, Karin E, Fogliati R, Dudeney J, Nielssen O, Gandy M, Staples L, Scott AJ, Heriseanu AI, Bisby MA, Hathway T, Titov N, Schroeder L. The Pain Course: a randomised controlled trial and economic evaluation of an internet-delivered pain management program. Pain 2022;163(7):1388-1401.

[10] Deer TR, Levy RM, Kramer J, Poree L, Amirdelfan K, Grigsby E, Staats P, Burton AW, Burgher AH, Obray J, Scowcroft J, Golovac S, Kapural L, Paicius R, Kim C, Pope J, Yearwood T, Samuel S, McRoberts WP, Cassim H, Netherton M, Miller N, Schaufele M, Tavel E, Davis T, Davis K, Johnson L, Mekhail N. Dorsal root ganglion stimulation yielded higher treatment success rate for complex regional pain syndrome and causalgia at 3 and 12 months: a randomized comparative trial. Pain 2017;158(4):669-681.

[11] 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(4):416-424.

[12] 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, Immpact. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113(1-2):9-19.

[13] 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(3):238-244.

[14] Dworkin RH, Turk DC, Peirce-Sandner S, Baron R, Bellamy N, Burke LB, Chappell A, Chartier K, Cleeland CS, Costello A, Cowan P, Dimitrova R, Ellenberg S, Farrar JT, French JA, Gilron I, Hertz S, Jadad AR, Jay GW, Kalliomaki J, Katz NP, Kerns RD, Manning DC, McDermott MP, McGrath PJ, Narayana A, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Reeve BB, Rhodes T, Sampaio C, Simpson DM, Stauffer JW, Stucki G, Tobias J, White RE, Witter J. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. Pain 2010;149(2):177-193.

[15] Farrar JT. What is clinically meaningful: outcome measures in pain clinical trials. Clin J Pain 2000;16(2 Suppl):S106-112.

[16] Farrar JT, Young JP, Jr., LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149-158.

[17] Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, Furst D, Goldsmith C, Kieszak S, Lightfoot R, Paulus H, Tugwell P, Weinblatt M, Widmark R, H.J. W, Wolfe F. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36(6):729-740.

[18] Ferguson L, Scheman J. Patient global impression of change scores within the context of a chronic pain rehabilitation program. J Pain 2009;10(4).

[19] Futamura M, Leshem YA, Thomas KS, Nankervis H, Williams HC, Simpson EL. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: Many options, no standards. J Am Acad Dermatol 2016;74(2):288-294.

[20] Gagnon CM, Scholten P, Atchison J. Multidimensional Patient Impression of Change Following Interdisciplinary Pain Management. Pain Pract 2018;18(8):997-1010.

[21] Gurrell R, Dua P, Feng G, Sudworth M, Whitlock M, Reynolds DS, Butt RP. A randomised, placebo-controlled clinical trial with the alpha2/3/5 subunit selective GABAA positive allosteric modulator PF-06372865 in patients with chronic low back pain. Pain 2018;159(9):1742-1751.

[22] Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, Duda SN, Consortium RE. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019;95:103208.

[23] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381.

[24] Hedman-Lagerlof M, Hedman-Lagerlof E, Axelsson E, Ljotsson B, Engelbrektsson J, Hultkrantz S, Lundback K, Bjorkander D, Wicksell RK, Flink I, Andersson E. Internet-Delivered Exposure Therapy for Fibromyalgia: A Randomized Controlled Trial. Clin J Pain 2018;34(6):532-542.

[25] Hosomi K, Sugiyama K, Nakamura Y, Shimokawa T, Oshino S, Goto Y, Mano T, Shimizu T, Yanagisawa T, Saitoh Y, investigators T-P. A randomized controlled trial of 5 daily sessions and continuous trial of 4 weekly sessions of repetitive transcranial magnetic stimulation for neuropathic pain. Pain 2020;161(2):351-360.

[26] Huffman CL, Goldenberg JN, Weintraub J, Sanin L, Driscoll J, Yang R, Chew ML, Scavone JM. Efficacy and Safety of Once-Daily Controlled-Release Pregabalin for the Treatment of Patients With Postherpetic Neuralgia: A Double-Blind, Enriched Enrollment Randomized Withdrawal, Placebo-Controlled Trial. Clin J Pain 2017;33(7):569-578.

[27] Kamper SJ, Ostelo RW, Knol DL, Maher CG, de Vet HC, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 2010;63(7):760-766 e761.

[28] Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66 Suppl 3:iii40-41.

[29] Langford DJ, Sharma S, McDermott MP, Beeram A, Besherat S, France FO, Mark R, Park M, Nishtar M, Turk DC, Dworkin RH, Gewandter JS. Covariate adjustment in chronic pain Trials: An oft-missed opportunity. J Pain 2023.

[30] Lumley MA, Schubiner H, Lockhart NA, Kidwell KM, Harte SE, Clauw DJ, Williams DA. Emotional awareness and expression therapy, cognitive behavioral therapy, and education for fibromyalgia: a cluster-randomized controlled trial. Pain 2017;158(12):2354-2363.

[31] Moore A, Fisher E, Eccleston C. Flawed, futile, and fabricated-features that limit confidence in clinical research in pain and anaesthesia: a narrative review. Br J Anaesth 2023;130(3):287-295.

[32] Nikiphorou E, Radner H, Chatzidionysiou K, Desthieux C, Zabalan C, van Eijk-Hustings Y, Dixon WG, Hyrich KL, Askling J, Gossec L. Patient global assessment in measuring disease activity in rheumatoid arthritis: a review of the literature. Arthritis Res Ther 2016;18(1):251.

[33] 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(8):967-977.

[34] 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(8):978-988.

[35] Perez-Aranda A, Feliu-Soler A, Montero-Marin J, Garcia-Campayo J, Andres-Rodriguez L, Borras X, Rozadilla-Sacanell A, Penarrubia-Maria MT, Angarita-Osorio N, McCracken LM, Luciano JV. A randomized controlled efficacy trial of mindfulness-based stress reduction compared with an active control group and usual care for fibromyalgia: the EUDAIMON study. Pain 2019;160(11):2508-2523.

[36] Perrot S, Lanteri-Minet M. Patients' Global Impression of Change in the management of peripheral neuropathic pain: Clinical relevance and correlations in daily practice. Eur J Pain 2019;23(6):1117-1128.

[37] Rampakakis E, Ste-Marie PA, Sampalis JS, Karellis A, Shir Y, Fitzcharles MA. Real-life assessment of the validity of patient global impression of change in fibromyalgia. RMD Open 2015;1(1):e000146.

[38] Sanabria-Mazo JP, Colomer-Carbonell A, Borras X, Castano-Asins JR, McCracken LM, Montero-Marin J, Perez-Aranda A, Edo S, Sanz A, Feliu-Soler A, Luciano JV. Efficacy of Videoconference Group Acceptance and Commitment Therapy (ACT) and Behavioral Activation Therapy for Depression (BATD) for Chronic Low Back Pain (CLBP) Plus Comorbid Depressive Symptoms: A Randomized Controlled Trial (IMPACT Study). J Pain 2023;24(8):1522-1540.

[39] Scott W, McCracken LM. Patients' impression of change following treatment for chronic pain: global, specific, a single dimension, or many? J Pain 2015;16(6):518-526.

[40] Tchivileva IE, Hadgraft H, Lim PF, Di Giosia M, Ribeiro-Dasilva M, Campbell JH, Willis J, James R, Herman-Giddens M, Fillingim RB, Ohrbach R, Arbes SJ, Jr., Slade GD. Efficacy and safety of propranolol for treatment of temporomandibular disorder pain: a randomized, placebo-controlled clinical trial. Pain 2020;161(8):1755-1767.

[41] U.S. Food & Drug Administration. Patient-Focused Drug Development Guidance Public Workshop: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. Silver Spring, MD: Center for Drug Evaluation and Research, 2019.

[42] U.S. Food and Drug Administration. Attachment to discussion document for patient-focused drug development public workshop on Guidance 3: select, develop or modify fit-for-purpose clinical outcome assessments, Appendix 4. Silver Spring, MD: U.S. Food and Drug Administration, , 2018.

[43] U.S. Food and Drug Administration. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making guidance for industry, Food and Drug Administration staff, and other stakeholders. In: U.S. Department of Health and Human Services editor. Silver Spring, MD: U.S. Food and Drug Administration, 2023.

[44] Vandermost M, Bagraith KS, Kennedy H, Doherty D, Kilner S, Sterling M, Henry D, Jones M. Improvement in pain interference and function by an allied health pain management program: Results of a randomized trial. Eur J Pain 2021;25(10):2226-2241.

[45] Viktrup L, Hayes RP, Wang P, Shen W. Construct validation of patient global impression of severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract symptoms secondary to benign prostatic hyperplasia. BMC Urol 2012;12:30.

[46] Williams RM, Day MA, Ehde DM, Turner AP, Ciol MA, Gertz KJ, Patterson D, Hakimian S, Suri P, Jensen MP. Effects of hypnosis vs mindfulness meditation vs education on chronic pain intensity and secondary outcomes in veterans: a randomized clinical trial. Pain 2022;163(10):1905-1918.

Figure Legend

Figure 1. PRISMA flow diagram. *Reproduced from Langford, Lou et al with permission* [29].

Figure 2. Continuum of domains that respondents may consider in global assessments from broad domains (e.g., unspecified domains, overall health status) to more specific domains (e.g., Patient Global Assessment of Low Back Pain [21]). Whether a global assessment measure's clinical meaningfulness or sensitivity to change differ as a function of the specific nature of the domain(s) assessed is unclear and warrants investigation. Regardless, the clinical context in which a treatment is being studied should be considered when selecting a domain and the measure should be clearly and thoroughly described.

Figure 3. Short and longer-term considerations for measure selection, reporting and analysis, and development of global assessment measures for use in clinical trials of chronic pain treatments. Acronyms: NRS = numeric rating scale; PGIC = Patient Global Impression of Change; PGIS = Patient Global Impression of Severity.