



HHS Public Access

Author manuscript

J Pain. Author manuscript; available in PMC 2025 January 01.

Published in final edited form as:

J Pain. 2024 January ; 25(1): 142–152. doi:10.1016/j.jpain.2023.07.028.

Comparable Minimally Important Differences and Responsiveness of Brief Pain Inventory and PEG Pain Scales across Six Trials

David E. Reed II, PhD^{1,2}, Timothy E. Stump³, Patrick O. Monahan, PhD³, Kurt Kroenke, MD^{4,5}

¹Center of Innovation for Veteran-Centered and Value-Driven Care, VA Puget Sound Health Care

²Department of Health Systems and Population Health, School of Public Health, University of Washington

³Department of Biostatistics and Health Data Science, Indiana University Fairbanks School of Public Health and School of Medicine, Indianapolis, IN, USA

⁴Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA

⁵Regenstrief Institute, Inc., Indianapolis, IN, USA

Abstract

The 3-item pain intensity (P), interference with the enjoyment of life (E), and interference with general activity (G), or PEG, has become one of the most widely used measures of pain severity and interference. The minimally important differences (MID) and responsiveness of the PEG are essential metrics for solidifying its role in research and clinical care. The current study aims to establish the MID and responsiveness of the PEG by synthesizing data from 1,710 participants across 6 controlled trials. MIDs were estimated using absolute score changes among individuals reporting their pain was “a little better” on a retrospective global change anchor as well as distribution-based estimates using standard deviation thresholds and 1 and 2 standard errors of measurement. Responsiveness was assessed using standardized response means, area under the curve, and treatment effect sizes. MID estimates for the PEG ranged from 0.60 to 1.1 when using 0.35 SD, and 0.78 to 1.22 using 1 standard error of measurement. MID estimates using the global anchor had somewhat more variability but most estimates ranged from 1.0 to 1.75. Responsiveness effect sizes were generally large (> .80) for standardized response means and moderate (> .50) for treatment effect. Similarly, the most area under the curve values demonstrated an acceptable level of scale responsiveness (.70). Importantly, MID estimates and responsiveness of the PEG and BPI scales were largely comparable when aggregating data across trials. Our synthesis indicates

Corresponding Author: David E. Reed, II, PhD, Center of Innovation for Veteran-Centered and Value-Driven Care, VA Puget Sound Health Care, 1660 South Columbian Way, S-152, Seattle, WA 98108, Phone: 206-716-5889.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The authors have no conflicts of interest to disclose.

that 1 point is a reasonable MID estimate on these 0- to 10-point pain scales, with 2 points being an upper bound.

Keywords

PEG; Brief Pain Inventory; pain; psychometrics; measurement

Chronic pain remains one of the most significant and debilitating conditions in the United States, affecting millions of individuals.^{17,19} Pain is reported across a variety of patient populations, including in primary care^{24,47} and oncology¹ clinics, where clinicians are tasked to efficiently assess pain severity (ie, the intensity of pain) and interference (ie, the degree to which pain disrupts daily activities). Due to the need for a brief measure that assesses both pain severity and interference, the 3-item pain intensity (P), interference with enjoyment of life (E), and interference with general activity (G), or PEG²⁴, has become one of the most widely used ultra-brief measures used to assess pain severity and interference. However, establishing the minimally important difference (MID) of the PEG as well as its responsiveness are critical steps to support its increasing use.

The PEG was derived from the longer 11-item Brief Pain Inventory (BPI) legacy scale⁶ which includes assessments of pain severity (4 items), in addition to pain interference (7 items), across several health domains (eg, mood, walking ability, and sleep). The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations¹¹ and guidelines from the Veterans Health Administration Pain Measures Work Group²⁸ include the assessment of both pain severity and interference. Recommendations specifically mention the BPI-Interference subscale and pain numeric scale (which is included in both the PEG and BPI-Severity subscale). An MID of 1 point for the BPI-Interference subscale and the numeric rating scale (NRS) has been proposed.¹² While assessing both pain intensity and interference broadens the scope of information gathered in research and clinical settings, administering all items is less efficient. Comparatively, the PEG has only 3 items, efficiently integrating the assessment of both pain severity and interference into a single composite score.

The MID is defined as the least required amount of reported change on an assessment to be considered important for an individual.³⁶ Responsiveness is defined as a measure's capability to detect meaningful change over time due either to an intervention or change that occurs naturalistically.³⁴ A variety of methods has been established to determine MIDs and responsiveness¹⁵, but they broadly encompass two strategies. Distribution-based methods use psychometric properties of the assessments (eg, standard deviations and reliability) to establish an MID. In contrast, anchor-based methods analyze change relative to a retrospective or prospective item that participants answer about self-perceived improvement.

Because different methods may produce somewhat different estimates of MID, some experts recommend triangulating both anchor-based and distribution-based methods when estimating MID.^{40,45} Common anchors include the patient-rated global impression of change and comparison with absolute change on a legacy measure of the same domain. Common distribution-based methods include 0.2 to 0.5 standard deviations (SDs) or

1 to 2 standard errors of measurement (SEMs) as the lower and upper bounds of an MID.^{5,26,30,39,40,42,51} Anchor-based methods have important advantages⁵⁰ and are recommended by the FDA as a preferred method⁴⁶; however, supplementing anchor-based with distribution-based approaches may be informative, especially when different methods demonstrate reasonable convergence.

Although the PEG has had relatively rapid clinical and research uptake, its use and interpretation will be substantially enhanced by further evidence regarding its MID and responsiveness. A recent push to ensure randomized controlled trials (RCTs) are adequately powered to detect clinically, as opposed to merely statistically, meaningful changes, further emphasizes the importance of establishing MIDs for the PEG.¹⁴ Moreover, the movement towards using measurement-based care to monitor and adjust treatment for physical and psychological symptoms¹³ highlights the importance of determining the responsiveness of patient-reported outcome measures. Thus, further establishing the MID and responsiveness of the PEG will enable clinicians and researchers to assess whether patients or participants are improving meaningfully over time, and will serve as accurate benchmarks to determine within and between group effect sizes and power calculations for clinical trials.

The current study is a psychometric synthesis of the PEG and aims to assess the MID and responsiveness of the PEG and compare its psychometric properties to the BPI legacy measure across 6 randomized clinical trials in a variety of clinical settings and patients with pain. The PEG was compared to the total BPI score, in addition to the BPI-Severity and BPI-Interference subscales. In integrating the findings across the 6 trials, we hypothesized that the ultra-brief PEG would have generally comparable MID estimates and responsiveness compared to its legacy, but longer, parent measure, the BPI.

Methods

We examined data across 6 randomized clinical trials to compare the BPI and PEG. These trials were chosen because they included the requisite data for calculating the MIDs and responsiveness for the PEG and 3 BPI scales. Some of the MID and responsiveness metrics have previously been published for the individual trials.^{4,5,19,22,32} In the current paper, however, we summarize and synthesize results across all 6 trials. Additionally, the integration of PEG findings is especially important given its emergence as an ultra-brief pain measure with accelerating clinical and research use.

The Stepped Care for Affective disorders and Musculoskeletal Pain trial (SCAMP; [NCT00118430](#)) included 427 adults diagnosed with co-occurring musculoskeletal pain and depression recruited from primary care clinics.^{22,25} The Indiana Cancer Pain and Depression trial (INCPAD; [NCT00313573](#)) included 274 adults with cancer diagnosed with pain and/or depression recruited from outpatient oncology clinics.^{31,32} The Stepped Care Optimizing Pain Care Effectiveness trial (SCOPE; [NCT00926588](#)) included 250 adults who had chronic musculoskeletal pain recruited from primary care clinics.^{19,29} The Care Management for the Effective use of Opioids trial (CAMEO; [NCT01236521](#)) included 261 adults with low back pain receiving long-term opioid therapy recruited from multiple Veteran Affairs (VA) primary care clinics.^{2,4,5} The Strategies for Prescribing Analgesics Comparative

Effectiveness trial (SPACE; [NCT01583985](#)) included 240 adults with chronic back pain, or hip or knee osteoarthritis pain recruited from multiple VA primary care clinics.^{4,5,23} The Stroke Survivors Self-Management trial (SSM; [NCT01507688](#)) included 258 adults with stroke recruited from several hospitals, including VA hospitals.^{4,5} Table 1 summarizes the trial information.

Measures

Brief Pain Inventory.

The Brief Pain Inventory^{6,33} includes both a total score as well as 2 subscales for severity and interference. The severity subscale includes 3 items assessing pain at its worst, least, and average over the past 24 hours, and 1 item assessing current pain level. Participants respond on a scale ranging from 0 (no pain) to 10 (pain as bad as you can imagine). The Interference subscale has 7 items assessing how pain has interfered with general activity, mood, walking ability, work/housework, relationships, sleep, and enjoyment of life. Participants respond on a scale ranging from 0 (does not interfere) to 10 (completely interferes). Unweighted means of items are used to calculate the total scale score and subscales, and higher scores represent more pain severity and/or interference.

PEG.

The PEG²⁴ is a 3-item assessment of pain severity and interference. Participants respond to the item “What number best describes your pain on average in the past week” on a 0 (no pain) to 10 (pain as bad as you can imagine) scale. Participants respond to the items “What number best describes how, during the past week, pain has interfered with your enjoyment of life?” and “What number best describes how, during the past week, pain has interfered with your general activity” on a 0 (does not interfere) to 10 (completely interferes) scale. The total scale score equals an unweighted mean of items, and higher scores represent more pain severity and/or interference.

Retrospective Global Change.

Individuals are asked on a 7-point Likert scale whether, since their last assessment: “*Overall would you say your pain is ...*”. In SCAMP and INCPAD, the 7 response options are *worse, about the same, a little better, somewhat better, moderately better, a lot better, and completely better (pain is gone)*. In the other 4 trials, the 7 response options are *much better, moderately better, a little better, no change, a little worse, moderately worse, much worse*.

Data Analysis

Using established methods to estimate MIDs and responsiveness^{18,27,39}, we synthesized previous psychometric work^{4,5,19,22,32} coupled with new analyses of the PEG to estimate its MID and determine its responsiveness. Distribution-based approaches used baseline data (T1) from the full study sample without comparing subgroups. Anchor-based approaches used data from either the follow-up assessment (T2) for the retrospective anchor or from two timepoints – baseline (T1) and follow-up (T2) – for the longitudinal application of cross-sectional anchors, and compared subgroups according to patient-reported global change.

One responsiveness metric (between-group treatment effect size) compared longitudinal change according to intervention versus control group status.

MID Estimates using an Anchor-Based Approach

Absolute changes in scale scores from baseline to follow-up (T1 to T2) were determined for individuals who reported they were “a little better” at follow-up (T2) on the 7-point retrospective global anchor.

MID Estimates using Distribution-Based Approaches

SD Thresholds.

Proportional values of the SD represent one type of effect size and are a common distribution-based method for estimating a MID. Small, medium, and large ES are considered .2 SD, .5 SD., and .8 SD, respectively.⁷ Therefore, as in previous work^{5,39}, we considered .35 SD to be a reasonable point estimate of MID as it is midway between a small and medium ES. We also report .5 SD as a more conservative upper bound of the MID.^{5,19,22}

Standard Error of Measurement.

The SEM is another distribution-based method of estimating an MID. The SEM is calculated by multiplying the standard deviation by the square root of 1-reliability. We used Cronbach’s alpha as the reliability estimate. Prior studies have found that 1 SEM corresponds to anchor-based MIDs.^{5,49,51} Depending on the context, 2 SEM can also be an appropriate approach to estimate the MID.⁴⁸ Thus, we considered 1 SEM to be a reasonable point estimate of MID, with 2 SEM representing an upper bound. Of note, when reliability = .75, 1 SEM = .50 SD (See Table 2 footnote for SEM formula).

Responsiveness using Anchor-Based Approaches

Standardized Response Mean (SRM).

The SRM is calculated as a standardized difference in scale scores between 2-time points (ie, $T1 - T2 / SD$ of change scores) that corresponds to participants’ response to a global anchor item (improved, unchanged, or worse). Our analyses focused on the SRM for the patient group that reported improvement compared to those that did not improve. All 6 trials used a retrospective global anchor wherein respondents at the follow-up time point (T2) reported their change in pain compared to the initial time point (T1). Additionally, 3 trials used a prospective global anchor wherein respondents provided cross-sectional global pain estimates at two time points (T1 and T2) and the difference in T1 and T2 global pain estimates were calculated to classify individuals as improved, unchanged or worse.⁴ The SRM is a type of effect size for assessing responsiveness, and SRMs of .2, .5, and .8 are considered small, medium, and large effect sizes respectively.⁷ Although these SRM thresholds were originally derived for Cohen’s d effect sizes,³⁷ differences between SRM and Cohen’s d are generally quite small; therefore, Cohen’s d thresholds can serve as a

reasonable approximation of thresholds for SRM²⁰ as well as between-group treatment effect sizes (described below).¹⁹

Area Under the Curve.

The area under the curve (AUC) is an anchor-based method of assessing responsiveness and is determined by a receiver operator characteristic (ROC) analysis.⁴¹ The discriminatory strength of the scale for determining any improvement and moderate improvement using the retrospective global anchor was estimated by the AUC. Some experts recommend an AUC .70 as a threshold for responsiveness when using a criterion standard anchor but also acknowledge that criterion standards often do not exist for patient-reported outcomes.^{38,43} Comparable AUCs suggest similar responsiveness of scales.

Responsiveness to Treatment

Between Group Treatment Effect Sizes.

This responsiveness metric was calculated by subtracting the control group mean change from the intervention group mean change and dividing this difference by the standard deviation of the pooled change score. This metric could be calculated for the 3 trials that had a true control (usual care) group rather than an active comparator. Treatment effect sizes of .2, .5, and .8 represent small, medium, and large intervention effects, respectively.⁷

Results

Sample Characteristics

Study and participant characteristics are summarized in Table 1. Table 1 also summarizes the type of intervention or control exposure each group received in the two-arm trials as well as the retrospective global change anchor used. Patients were recruited from either primary care clinics (n=4), an oncology clinic (n=1), or after being diagnosed with stroke (n=1). The 6 trials included a total of 1,710 participants, with the sample size across trials ranging from 240 to 427. The average age of participants in each trial ranged from 55.2 to 61.7. Men constituted 67.7% of the total sample, ranging from 33.9% to 92.3% across the trials. Overall, 71.2% of participants were white (range across trials, 58.3%–86.2%) and 24.4% were black (range, 7.5%–38.2%)

In all but one trial (post-stroke), participants on average endorsed moderate pain severity and interference (Table 2). Specifically in these 5 trials, PEG scores ranged from 5.4 to 6.5; BPI-Severity, from 5.1 to 6.8; BPI-Interference, from 5.3 to 6.4; and BPI-Total, from 5.2 to 6.5. In the post-stroke trial, scale scores reflected milder pain.

Minimally Important Differences

Table 2 summarizes the means and internal consistency reliability of the scales, as well the distribution-based and anchor-based estimates of MIDs. MID estimates using .35 SD ranged from .60 to 1.05 for the PEG, .49 to .98 for BPI severity, .67 to 1.10 for BPI interference, and .56 to .98 for BPI total. MID estimates using 1 SEM ranged from .78 to 1.22 for the

PEG, .57 to 1.04 for BPI severity, .74 to .87 for BPI interference, and .54 to .70 for BPI total. Data for .5 SD and 2 SEM is also summarized in the table.

MID estimates using change scores from individuals reporting being “a little better” on the retrospective global anchor revealed somewhat greater variability. For 3 trials (SCOPE, CAMEO, and SPACE), most MID estimates for the 4 scales using the global anchor were in the 1.0 to 1.75 range. Conversely, the SSM trial showed unexpectedly small negative MIDs using this global anchor, whereas the SCAMP trial showed slightly larger MIDs (possibly because “a little better” and “somewhat better” were collapsed for the global anchor in SCAMP which assessed 5 levels of improvement (rather than the 3 levels of improvement assessed in the other 4 trials).

Responsiveness

Table 3 summarizes data regarding responsiveness using anchor-based approaches (SRM and AUC) and treatment response. The SRM for improvement with the retrospective global anchor for 5 of the 6 trials (excluding the post-stroke trial which was an extreme outlier) ranged from .77 to 1.43 for the PEG, .71 to 1.18 for BPI severity, .76 to 1.20 for BPI interference, and .83 to 1.29 for BPI total. For the 3 trials which had a prospective global anchor, SRM estimates were generally similar to estimates using a retrospective anchor in 2 trials, but higher for the post-stroke trial. For 5 of the 6 trials (excluding the post-stroke trial), AUC values of the PEG demonstrated an acceptable level of scale responsiveness (.70).

The between-group treatment effect size in the 3 trials where this responsiveness metric could be calculated showed a moderate treatment effect in 2 trials (SCAMP and INCPAD) and a small to moderate treatment effect in 1 trial (SCOPE). Importantly, both the AUC values and treatment effect sizes were generally comparable for the 4 scales within each trial.

Synthesis of Metric Data

Table 4 provides a synthesis of the MID and responsiveness metrics across the 6 trials. Several important findings should be noted. First, the results were similar whether using the median or the weighted mean to integrate metrics across the 6 trials. Second, most metrics are relatively similar for the PEG and BPI scales, except for a somewhat higher SEM for the PEG (an expected consequence of shorter scales usually having a lower Cronbach's alpha). Third, most MID estimates are around 1 point ($\pm .3$) on these 0 to 10-point scales. Fourth, responsiveness as assessed by the SRM revealed large effects sizes ($> .80$) and acceptable AUC ($> .70$) and was generally comparable for all 4 scales. Fifth, between-group (control vs treatment) effect sizes were in the .5 SD range, which is consistent with a moderate responsiveness to treatment. Sixth, many of the differences between the PEG and BPI scales in MID values were $< .20$ which is considered a lower threshold for a small difference.³⁵ The most notable exception was the SEM having somewhat higher PEG-BPI differences due a lower Cronbach's alpha for the PEG, which is expected for a shorter scale.

Because SSM was the only trial that did not explicitly enroll patients with elevated pain and because several of its psychometric findings differed substantially from the other 5 trials,

we performed a sensitivity analysis by comparing the synthesis results with and without the SSM trial. As summarized in the Supplemental Table, results are generally similar with and without the SSM trial. Specifically, differences between the PEG and the 3 BPI scales fell into a similarly narrow range whether including or excluding the SSM trial. Also, changes in the values of specific metrics were typically quite small ($< .10$) when excluding the SSM trial.

Discussion

The current research examined MIDs and responsiveness of the PEG and compared these estimates to the legacy BPI subscales and total score across 6 clinical trials. The aggregate data across trials supports a 1-point difference in the PEG as being clinically meaningful, which is generally consistent with other literature within the context of a 0 to 10-point scale.^{5,12,35} Indeed, when comparing group differences, evidence-based reviews use .5, 1 and 2 point changes on a 0 to 10-point numeric rating scale as indicative of small, moderate, and large treatment effects, respectively.³⁵ Results also suggest that MID estimates and responsiveness of the PEG and BPI scales are largely comparable when using multiple psychometric approaches. Study strengths include the synthesis of data across more than 1,700 patients from 6 trials, heterogeneity of clinical settings and patient samples which enhances generalizability, scoring of all 4 pain measures on a similar 0 to 10 point scale, and triangulation of psychometric estimates using a variety of accepted methods.

Our findings comprise the most comprehensive synthesis of empiric data aimed at systematically establishing an MID for the 3-item PEG, which has gained substantial uptake since its development over a decade ago. In 2016, the United States Surgeon General initiated a Turn the Tide opioid campaign and sent a letter to more than 2.3 million health care practitioners and public health leaders across the country to seek help in addressing the prescription opioid crisis. The campaign advised using a validated pain scale before prescribing and highlighted the PEG as a specific example. Consequently, the U.S. Centers for Disease Control and Prevention included the PEG in its Centers for Disease Control pocket guide.³

Comparable results across trials and measures suggest that clinicians and researchers should feel confident using the 3-item PEG to detect between- and within-group change over time using the 1-point (best estimate) to 2-point (upper bound) threshold. The current research suggests clinical trials should be powered to detect a 1- to 2-point threshold. From a power perspective, 1 point is the more conservative threshold because larger sample sizes are needed to detect a smaller population difference. In addition, a 1-point difference is in line with the common usage of a 1-point change in the NRS as being considered a meaningful difference. Using a 2-point threshold to define “treatment responder” in the sample data increases certainty that meaningful change occurred for that individual but increases the likelihood of false negatives. Nevertheless, using a 1-point change over time to define a “treatment responder” in the sample data also increases the likelihood (compared to using a 2-point change) that false positives are categorized as treatment responders.

The magnitude of an MID may vary depending upon whether one is measuring a difference or change at the level of an individual person versus using aggregated individual-level data to compare differences between groups in research or clinical populations.^{15,21} To be considered meaningful, change within an individual may need to be larger than differences that are detected between groups.¹⁰ Thus, a 1-point change may be appropriate as the MID for group changes in research studies whereas a larger change (1 to 2 points) may be considered when clinically monitoring individual patients.¹² It should be pointed out that the majority of MID estimates across the trials are around 1 point or less (Table 4), supporting this as the best MID point estimate. Finally, we note there is debate on how best to use MIDs in relation to categorizing treatment responders.¹⁵

Comparable MIDs and responsiveness of the PEG and BPI scales across trials also allow researchers and clinicians to choose the measure based on the needs and implications of the clinical or research question being asked. The PEG is less burdensome and provides a broad snapshot of symptoms, while the BPI subscales focus on a specific aspect of the pain experience while taking more time to complete. Of note, both the 3-item PEG and 11-item BPI total score integrate both pain severity and interference into a single composite score rather than 2 separate domain scores. A single score may have advantages when choosing a single primary outcome in research or when monitoring and adjusting pain treatment in clinical practice. The current research thus expands the options within clinical research settings. Clinicians now have a benchmark for using the PEG in settings where pain severity and interference is of interest. Importantly, because the pain numeric rating scale is part of the PEG, both can easily be used with a patient or within the same study using pain NRS-established benchmarks if the NRS is the principal outcome of interest.¹²

Results in the single trial of post-stroke patients, whose pain levels were mild, differed from the more consistent results across the other 5 trials. Thus, we conducted a sensitivity analysis by synthesizing results with and without the stroke trial. It is reassuring that the results were generally similar in that the differences between the synthesis of 6 versus 5 trials were relatively small. Nonetheless, further research in populations with different levels of pain, as well as different health conditions, is warranted.

Strengths

There are several noted strengths of the current research. Data were compiled from 6 separate clinical trials in a variety of settings. These heterogeneous samples increase the generalizability of the results. We also used a variety of methods to establish MIDs and responsiveness, increasing confidence in the results. Finally, patients in all but the SSM post-stroke trial required at least moderate pain for inclusion, suggesting the 1-point best estimate for MID and the 2-point upper bound are appropriate for patients who meet common inclusion criteria for chronic pain trials, as well as patients being treated for pain in practice.

Limitations and Future Directions

Results are presented with limitations. Half of the trials were limited to a retrospective anchor in calculating the SRM and between-group comparisons. Our use of only six trials with heterogeneous samples limits the results' generalizability. Most patients were recruited from primary care. Therefore, different MID values may be found in samples of patients with more severe chronic pain or patients that have a more extensive history of chronic pain treatments. For example, a 1-point change may be more clinically meaningful for patients established in a pain clinic with consistent severe pain and who are considering a complicated neck surgery, compared to someone with moderate knee pain 60% of the time seeking treatment through primary care. Finally, the noted differences in the post-stroke trial suggest that results may be less generalizable to post-stroke patients specifically, or possibly non-pain samples, more broadly.

Current results suggest potentially productive lines of future research. More research is needed to determine whether the current estimates are consistent among other populations of individuals with pain, including patients who receive their care in specialty pain clinics. Future research should examine how an MID on the PEG corresponds to other emotional well-being constructs relevant to chronic individuals with chronic pain, such as depressive affect,¹² self-concept,⁴⁴ or meaning and purpose.^{8,9} In addition, the prevalence of co-occurring mental health conditions among individuals with chronic pain is high.¹⁶ Ensuring that the 1 to 2-point benchmark remains consistent among individuals with co-occurring chronic pain and, for example, depression or posttraumatic stress disorder will provide valuable insights into how different patient populations consider meaningful change.

Conclusions

The accurate and efficient assessment of chronic pain across a variety of settings has become an important component of the clinical encounter. The PEG was developed to assess both pain severity and interference using only 3 items. Examining data from 6 randomized clinical trials, we used several distribution- and anchor-based methods to establish 1 to 2 points as an MID for the PEG. Moreover, the PEG and BPI scales demonstrated comparable responsiveness. Results allow clinicians and researchers to assess whether patients are making meaningful improvements over time, categorize treatment responders, and power randomized clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Disclosures

The studies from which data are derived were funded by the VA Office of Research and Development (NCT01507688, NCT01583985, NCT01236521, NCT00926588), the National Cancer Institute (NCT00313573), and the National Institute of Mental Health (NCT00118430).

References

1. MHJ van den Beuken-van Everdingen, de Rijke JM, Kessels AG, Schouten HC, van Kleef M, Patijn J. Prevalence of pain in patients with cancer: a systematic review of the past 40 years. *Annals of Oncology*. 2007;18(9):1437–1449. <https://linkinghub.elsevier.com/retrieve/pii/S0923753419421741>. doi:10.1093/annonc/mdm056 [PubMed: 17355955]
2. Bushey MA, Slaven JE, Outcalt SD, et al. Effect of Medication Optimization vs Cognitive Behavioral Therapy Among US Veterans With Chronic Low Back Pain Receiving Long-term Opioid Therapy. *JAMA Network Open*. 2022;5(11):e2242533. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2798621>. doi:10.1001/jamanetworkopen.2022.42533 [PubMed: 36394874]
3. Centers for Disease Control and Prevention. Prescribing Opioids for Chronic Pain. [accessed 2023 Feb 13]. https://www.cdc.gov/drugoverdose/pdf/turnthetide_pocketguide-a.pdf
4. Chen CX, Kroenke K, Stump T, et al. Comparative Responsiveness of the PROMIS Pain Interference Short Forms With Legacy Pain Measures: Results From Three Randomized Clinical Trials. *The Journal of Pain*. 2019;20(6):664–675. <https://linkinghub.elsevier.com/retrieve/pii/S1526590018309313>. doi:10.1016/j.jpain.2018.11.010 [PubMed: 30529442]
5. Chen CX, Kroenke K, Stump TE, et al. Estimating minimally important differences for the PROMIS pain interference scales: results from 3 randomized clinical trials. *Pain*. 2018;159(4):775–782. <https://journals.lww.com/00006396-201804000-00017>. doi:10.1097/j.pain.0000000000001121 [PubMed: 29200181]
6. Cleeland CS. Brief Pain Inventory (Short Form). 1991. https://www.mdanderson.org/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf
7. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge; 1988. <https://www.taylorfrancis.com/books/9781134742707>. doi:10.4324/9780203771587
8. Dezutter J, Luyckx K, Wachholtz A. Meaning in life in chronic pain patients over time: associations with pain experience and psychological well-being. *Journal of Behavioral Medicine*. 2015;38(2):384–396. doi:10.1007/s10865-014-9614-1 [PubMed: 25537924]
9. Dezutter J, Offenbaecher M, Vallejo MA, Vanhooren S, Thauvoye E, Toussaint L. Chronic pain care: The importance of a biopsychosocial-existential approach. *International Journal of Psychiatry in Medicine*. 2016;51(6):563–575. doi:10.1177/0091217417696738. doi:10.1177/0091217417696738 [PubMed: 28629291]
10. Donaldson G. Patient-reported outcomes and the mandate of measurement. *Quality of Life Research*. 2008;17(10):1303–1313. <http://link.springer.com/10.1007/s11136-008-9408-4>. doi:10.1007/s11136-008-9408-4 [PubMed: 18953670]
11. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1):9–19. <https://journals.lww.com/00006396-200501000-00005>. doi:10.1016/j.pain.2004.09.012 [PubMed: 15621359]
12. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. *Journal of Pain*. 2008;9(2):105–121. doi:10.1016/j.jpain.2007.09.005 [PubMed: 18055266]
13. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, Harbin HT. A Tipping Point for Measurement-Based Care. *Psychiatric Services*. 2017;68(2):179–188. <http://psychiatryonline.org/doi/10.1176/appi.ps.201500439>. doi:10.1176/appi.ps.201500439 [PubMed: 27582237]
14. Gianola S, Castellini G, Corbetta D, Moja L. Rehabilitation interventions in randomized controlled trials for low back pain: proof of statistical significance often is not relevant. *Health and Quality of Life Outcomes*. 2019;17(1):127. <https://hqlo.biomedcentral.com/articles/10.1186/s12955-019-1196-8>. doi:10.1186/s12955-019-1196-8 [PubMed: 31331343]
15. Hays RD, Peipert JD. Between-group minimally important change versus individual treatment responders. *Quality of Life Research*. 2021;30(10):2765–2772. <https://link.springer.com/10.1007/s11136-021-02897-z>. doi:10.1007/s11136-021-02897-z [PubMed: 34129173]

16. Hooten WM. Chronic Pain and Mental Health Disorders: Shared Neural Mechanisms, Epidemiology, and Treatment. *Mayo Clinic Proceedings*. 2016;91(7):955–970. doi:10.1016/j.mayocp.2016.04.029. doi:10.1016/j.mayocp.2016.04.029 [PubMed: 27344405]
17. Institute of Medicine (US) Committee on Advancing Pain Research Care and Education. *Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research*. Washington, D.C.: National Academies Press; 2011. <http://www.nap.edu/catalog/13172>. doi:10.17226/13172
18. Johns SA, Kroenke K, Krebs EE, Theobald DE, Wu J, Tu W. Longitudinal Comparison of Three Depression Measures in Adult Cancer Patients. *Journal of Pain and Symptom Management*. 2013;45(1):71–82. <https://linkinghub.elsevier.com/retrieve/pii/S0885392412002114>. doi:10.1016/j.jpainsymman.2011.12.284 [PubMed: 22921152]
19. Kean J, Monahan PO, Kroenke K, Wu J, Yu Z, Stump TE, Krebs EE. Comparative Responsiveness of the PROMIS Pain Interference Short Forms, Brief Pain Inventory, PEG, and SF-36 Bodily Pain Subscale. *Medical Care*. 2016;54(4):414–421. <https://journals.lww.com/00005650-201604000-00012>. doi:10.1097/MLR.0000000000000497 [PubMed: 26807536]
20. Kim S, Hays RD, Birbeck GL, Vickrey BG. Responsiveness of the Quality of Life in Epilepsy Inventory (QOLIE-89) in an Antiepileptic Drug Trial. *Quality of Life Research*. 2003;12:147–155. doi:10.1023/A:1022209105926 [PubMed: 12639061]
21. King MT, Dueck AC, Revicki DA. Can Methods Developed for Interpreting Group-level Patient-reported Outcome Data be Applied to Individual Patient Management? *Medical Care*. 2019;57(Suppl 1):S38–S45. <https://journals.lww.com/00005650-201905001-00008>. doi:10.1097/MLR.0000000000001111 [PubMed: 30985595]
22. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative Responsiveness of Pain Outcome Measures Among Primary Care Patients With Musculoskeletal Pain. *Medical Care*. 2010;48(11):1007–1014. <https://journals.lww.com/00005650-201011000-00010>. doi:10.1097/MLR.0b013e3181eaf835 [PubMed: 20856144]
23. Krebs EE, Gravely A, Nugent S, et al. Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain the SPACE randomized clinical trial. *JAMA - Journal of the American Medical Association*. 2018;319(9):872–882. doi:10.1001/jama.2018.0899 [PubMed: 29509867]
24. Krebs EE, Lorenz KA, Bair MJ, et al. Development and Initial Validation of the PEG, a Three-item Scale Assessing Pain Intensity and Interference. *Journal of General Internal Medicine*. 2009;24(6):733–738. <http://link.springer.com/10.1007/s11606-009-0981-1>. doi:10.1007/s11606-009-0981-1 [PubMed: 19418100]
25. Kroenke K, Bair MJ, Damush TM, Wu J, Hoke S, Sutherland J, Tu W. Optimized Antidepressant Therapy and Pain Self-management in Primary Care Patients With Depression and Musculoskeletal Pain. *JAMA*. 2009;301(20):2099. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>. doi:10.1001/jama.2009.723 [PubMed: 19470987]
26. Kroenke K, Baye F, Lourens SG. Comparative Responsiveness and Minimally Important Difference of Common Anxiety Measures. *Medical Care*. 2019;57(11):890–897. <https://journals.lww.com/10.1097/MLR.0000000000001185>. doi:10.1097/MLR.0000000000001185 [PubMed: 31415337]
27. Kroenke K, Baye F, Lourens SG. Comparative validity and responsiveness of PHQ-ADS and other composite anxiety-depression measures. *Journal of Affective Disorders*. 2019;246:437–443. <https://linkinghub.elsevier.com/retrieve/pii/S016503271832041X>. doi:10.1016/j.jad.2018.12.098 [PubMed: 30599366]
28. Kroenke K, Krebs EE, Turk D, et al. Core Outcome Measures for Chronic Musculoskeletal Pain Research: Recommendations from a Veterans Health Administration Work Group. *Pain Medicine*. 2019;20(8):1500–1508. <https://academic.oup.com/painmedicine/article/20/8/1500/5274160>. doi:10.1093/pm/pny279 [PubMed: 30615172]
29. Kroenke K, Krebs EE, Wu J, Yu Z, Chumbler NR, Bair MJ. Telecare collaborative management of chronic pain in primary care a randomized clinical trial. *JAMA - Journal of the American Medical Association*. 2014;312(3):240–248. doi:10.1001/jama.2014.7689 [PubMed: 25027139]
30. Kroenke K, Stump TE, Chen CX, et al. Minimally important differences and severity thresholds are estimated for the PROMIS depression scales from three randomized clinical trials.

- Journal of Affective Disorders. 2020;266:100–108. <https://linkinghub.elsevier.com/retrieve/pii/S0165032719322761>. doi:10.1016/j.jad.2020.01.101 [PubMed: 32056864]
31. Kroenke K, Theobald D, Wu J, Norton K, Morrison G, Carpenter J, Tu W. Effect of Telecare Management on Pain and Depression in Patients With Cancer. *JAMA*. 2010;304(2):163. <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2010.944>. doi:10.1001/jama.2010.944 [PubMed: 20628129]
 32. Kroenke K, Theobald D, Wu J, Tu W, Krebs EE. Comparative Responsiveness of Pain Measures in Cancer Patients. *The Journal of Pain*. 2012;13(8):764–772. <https://linkinghub.elsevier.com/retrieve/pii/S1526590012006566>. doi:10.1016/j.jpain.2012.05.004 [PubMed: 22800982]
 33. Kumar S, Rastogi S, Kumar S, Mahendra P, Bansal M, Chandra L. Pain in trigeminal neuralgia: neurophysiology and measurement: a comprehensive review. *Journal of medicine and life*. 2013;6(4):383–8. <http://www.ncbi.nlm.nih.gov/pubmed/24701256> [PubMed: 24701256]
 34. Liang MH. Longitudinal Construct Validity. *Medical Care*. 2000;38:II-84–II-90. <http://journals.lww.com/00005650-200009002-00013>. doi:10.1097/00005650-200009002-00013 [PubMed: 10982093]
 35. McDonagh MS, Selph SS, Buckley DI, et al. Nonopioid Pharmacologic Treatments for Chronic Pain. 2020. <https://effectivehealthcare.ahrq.gov/products/nonopioid-chronic-pain/research>. doi:10.23970/AHRQEPCCER228
 36. McGlothlin AE, Lewis RJ. Minimal Clinically Important Difference. *JAMA*. 2014;312(13):1342. <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2014.13128>. doi:10.1001/jama.2014.13128 [PubMed: 25268441]
 37. Middel B, Van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*. 2002;2(4). <http://www.ijic.org/article/10.5334/ijic.65/>. doi:10.5334/ijic.65
 38. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*. 2010;10(1):22. <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-10-22>. doi:10.1186/1471-2288-10-22 [PubMed: 20298572]
 39. Mosher CE, Secinti E, Johns SA, Kroenke K, Rogers LQ. Comparative responsiveness and minimally important difference of Fatigue Symptom Inventory (FSI) scales and the FSI-3 in trials with cancer survivors. *Journal of Patient-Reported Outcomes*. 2022;6(1):82. <https://jpro.springeropen.com/articles/10.1186/s41687-022-00488-1>. doi:10.1186/s41687-022-00488-1 [PubMed: 35870034]
 40. Mouelhi Y, Jouve E, Castelli C, Gentile S. How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health and Quality of Life Outcomes*. 2020;18(1):136. <https://hqlo.biomedcentral.com/articles/10.1186/s12955-020-01344-w>. doi:10.1186/s12955-020-01344-w [PubMed: 32398083]
 41. Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman HS, Klerman GL. Performance of screening and diagnostic tests: Application of receiver operating characteristic analysis. *Archives of General Psychiatry*. 1987;44(6):550–555. [PubMed: 3579501]
 42. Norman GR, Sloan JA, Wyrwich KW. Interpretation of Changes in Health-related Quality of Life. *Medical Care*. 2003;41(5):582–592. <http://journals.lww.com/00005650-200305000-00004>. doi:10.1097/01.MLR.0000062554.74615.4C [PubMed: 12719681]
 43. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1147–1157. <http://link.springer.com/10.1007/s11136-018-1798-3>. doi:10.1007/s11136-018-1798-3 [PubMed: 29435801]
 44. Reed DE, Cobos B, Nagpal AS, Eckmann M, McGeary DD. The role of identity in chronic pain cognitions and pain-related disability within a clinical chronic pain population. *The International Journal of Psychiatry in Medicine*. 2021 Jan 24:009121742198914. <http://journals.sagepub.com/doi/10.1177/0091217421989141>. doi:10.1177/0091217421989141
 45. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal*

- of Clinical Epidemiology. 2008;61(2):102–109. <https://linkinghub.elsevier.com/retrieve/pii/S0895435607001199>. doi:10.1016/j.jclinepi.2007.03.012 [PubMed: 18177782]
46. U.S. Food & Drug Administration. Patient-Focused Drug Development Guidance Public Workshop: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. 2019. <https://www.fda.gov/media/132505/download>
47. Upshur CC, Luckmann RS, Savageau JA. Primary care provider concerns about management of chronic pain in community clinic populations. *Journal of General Internal Medicine*. 2006;21(6):652–655. <http://link.springer.com/10.1111/j.1525-1497.2006.00412.x>. doi:10.1111/j.1525-1497.2006.00412.x [PubMed: 16808752]
48. Wyrwich KW. Minimal Important Difference Thresholds and the Standard Error of Measurement: Is There a Connection? *Journal of Biopharmaceutical Statistics*. 2004;14(1):97–110. <https://www.tandfonline.com/doi/full/10.1081/BIP-120028508>. doi:10.1081/BIP120028508 [PubMed: 15027502]
49. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking Clinical Relevance and Statistical significance in Evaluating Intra-Individual Changes in Health-Related Quality of Life. *Medical Care*. 1999;37(5):469–478. <http://journals.lww.com/00005650-199905000-00006>. doi:10.1097/00005650-199905000-00006 [PubMed: 10335749]
50. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S. Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*. 2013;22(3):475–483. <http://link.springer.com/10.1007/s11136-012-0175-x>. doi:10.1007/s11136-012-0175-x [PubMed: 22528240]
51. Wyrwich KW, Tierney WM, Wolinsky FD. Further Evidence Supporting an SEM-Based Criterion for Identifying Meaningful Intra-Individual Changes in Health-Related Quality of Life. *Journal of Clinical Epidemiology*. 1999;52(9):861–873. <https://linkinghub.elsevier.com/retrieve/pii/S0895435699000712>. doi:10.1016/S0895-4356(99)00071-2 [PubMed: 10529027]

Perspective:

This article synthesizes data from 6 clinical trials to establish the minimally important difference (MID) and responsiveness of the 3-item PEG pain scale. The PEG demonstrated good responsiveness, and 1 to 2 points proved to be reasonable estimates for the lower and upper bounds of the MID.

Highlights:

- Six RCTs were used to establish 1 to 2 points on the PEG as the lower and upper bounds of an MID.
- MID estimates and responsiveness of the PEG and BPI scales were largely comparable.
- Results provide guidance in determining meaningful improvement for patient care and future trials.

Table 1.

Key Characteristics of the Six Randomized Clinical Trials (n = 1710)

Trial	SCAMP (n=427)	INCPAD (n=274)	SCOPE (n=250)	CAMEO (n=261)	SPACE (n=240)	SSM (n=258)
Clinical Population	Co-occurring chronic musculoskeletal pain and depression	Cancer with pain and/or depression	Chronic musculoskeletal pain	Low Back Pain	Chronic back, hip, or knee pain	Post-Stroke
Setting	Primary Care	Oncology	Primary Care	Primary Care	Primary Care	Neurology
Age, mean (SD)	59.1 (13.0)	58.1 (10.5)	55.2 (8.5)	57.9 (9.5)	58.3 (13.7)	61.7 (10.8)
Sex, n (%)						
Men	199 (46.6)	93 (33.9)	207 (82.8)	241 (92.3)	208 (86.7)	209 (81.0)
Women	228 (53.4)	181 (66.1)	43 (17.2)	20 (7.7)	32 (13.3)	49 (19.0)
Race, n (%)						
White	249 (58.3)	212 (77.4)	192 (76.8)	191 (73.2)	207 (86.2)	166 (64.3)
Black	163 (38.2)	57 (20.8)	48 (19.2)	54 (20.7)	18 (7.5)	78 (30.2)
Other	15 (3.5)	5 (1.8)	10 (4.0)	16 (6.1)	15 (6.3)	14 (5.4)
Intervention group	Optimized Antidepressants and Pain Self-Management	Optimized Antidepressants and Analgesics	Optimized Analgesics	Optimized Analgesics	Optimized Opioid Analgesics	Stroke Self-Management Program
Control group	Usual Care	Usual Care	Usual Care	Cognitive Behavioral Therapy	Optimized Non-Opioid Analgesics	Usual Care
Retrospective global anchor [†]	7-item Likert Version A	7-item Version A	7-item Version B	7-item Version B	7-item Version B	7-item Version B

* Data in all studies represent sex assigned at birth.

[†] The participant is asked whether, since the last assessment: “Overall would you say your pain is ...”. The 7 response options in Version A are *worse, about the same, a little better, somewhat better, moderately better, a lot better, completely better (pain is gone)*. The 7 response options in Version B are *much better, moderately better, a little better, no change, a little worse, moderately worse, much worse*.

Table 2.

Means, Reliability, and Minimally Important Differences of Brief Pain Inventory and PEG Pain Scales in 6 Trials (n=1,710)

Variable	SCAMP (n=427)	INCPAD (n=274)	SCOPE (n=250)	CAMEO (n=261)	SPACE (n=240)	SSM (n=258)
Population	Primary care	Oncology	Primary care	Primary care	Primary care	Post-Stroke
Mean (SD)						
• BPI Severity	5.7 (1.8)	5.2 (1.8)	5.1 (1.7)	6.8 (1.6)	5.6 (1.4)	2.9 (2.8)
• BPI Interference	5.8 (2.4)	5.7 (2.6)	5.3 (2.2)	6.4 (2.1)	5.5 (1.9)	2.7 (3.0)
• BPI Total	5.7 (2.0)	5.5 (2.1)	5.2 (1.8)	6.5 (1.8)	5.5 (1.6)	2.8 (2.8)
• PEG	6.0 (2.2)	5.9 (2.2)	5.4 (2.0)	6.5 (1.9)	5.8 (1.7)	2.9 (3.0)
Cronbach's alpha internal reliability						
• BPI Severity	0.83	0.79	0.87	0.79	0.84	0.86
• BPI Interference	0.87	0.89	0.88	0.86	0.85	0.94
• BPI Total	0.88	0.89	0.91	0.87	0.87	0.94
• PEG	0.73	0.69	0.76	0.72	0.79	0.85
Anchor-Based MIDs						
A Little Better [*]						
• BPI Severity	1.36	--	0.73	0.87	1.17	-0.39
• BPI Interference	2.02	--	1.34	1.41	1.67	-0.52
• BPI Total	1.82	--	1.04	1.26	1.48	-0.55
• PEG	2.05	--	1.25	1.18	2.02	-0.52
Distribution-Based MIDs						
0.35 standard deviation (SD) [†]						
• BPI Severity	0.63	0.63	0.60	0.56	0.49	0.98
• BPI Interference	0.84	0.91	0.77	0.74	0.67	1.05
• BPI Total	0.70	0.74	0.63	0.63	0.56	0.98
• PEG	0.77	0.77	0.70	0.67	0.60	1.05
0.50 standard deviation (SD) [†]						
• BPI Severity	0.90	.90	0.85	0.80	0.70	1.40
• BPI Interference	1.20	1.3	1.10	1.05	0.95	1.50
• BPI Total	1.00	1.1	0.90	0.90	0.80	1.40
• PEG	1.10	1.1	1.00	0.95	0.85	1.50
Variable	SCAMP	INCPAD	SCOPE	CAMEO	SPACE	SSM
1 SEM [‡]						
• BPI Severity	0.74	0.82	0.61	0.72	0.57	1.04
• BPI Interference	0.87	0.86	0.76	0.77	0.75	0.74
• BPI Total	0.69	0.70	0.54	0.63	0.58	0.69
• PEG	1.14	1.22	0.98	1.01	0.78	1.15

Variable	SCAMP (n=427)	INCPAD (n=274)	SCOPE (n=250)	CAMEO (n=261)	SPACE (n=240)	SSM (n=258)
Population	Primary care	Oncology	Primary care	Primary care	Primary care	Post-Stroke
2 SEM [‡]						
• BPI Severity	1.48	1.65	1.23	1.44	1.13	2.08
• BPI Interference	1.73	1.72	1.52	1.53	1.49	1.48
• BPI Total	1.39	1.39	1.08	1.26	1.15	1.38
• PEG	2.29	2.45	1.96	2.02	1.55	2.30

* Absolute change in score from baseline to follow-up in individuals who reported on the retrospective global anchor that they were “a little better” at follow-up (4 trials). For the SCAMP that used Version A of the retrospective global anchor with 5 categories (rather than 3 categories) of improvement, score changes are for those reporting “a little” or “somewhat” better. Data not available for INCPAD.

[†] Difference of 0.2 SD is often considered a small effect, and 0.5 SD is considered a moderate effect, with 0.35 midway between small and moderate. Some consider 0.35 to 0.50 SD differences as one distribution-based method for estimating minimally important difference (MID)

[‡] SEM = standard error of measurement = $SD \times \sqrt{1 - \alpha}$, where α = Cronbach’s alpha. The SEM is a second distribution-based method of estimating an MID, for which 1 to 2 SEM are often considered the lower and upper bounds.

Table 3.

Responsiveness of Brief Pain Inventory and PEG Pain Scales in 6 Trials

Variable*	SCAMP (n=427)	INCPAD (n=274)	SCOPE (n=244) [†]	CAMEO (n=261)	SPACE (n=240)	SSM (n=258)
Population	Primary care	Oncology	Primary care	Primary care	Primary care	Post-Stroke
Anchor-Based Responsiveness						
SRM for Improvement, retrospective*						
• BPI Severity	1.00	1.13	0.71	0.72	1.18	0.17
• BPI Interference	0.86	0.91	0.94	0.76	1.20	0.17
• BPI Total	1.02	1.10	0.93	0.83	1.29	0.17
• PEG	0.99	1.08	0.86	0.77	1.43	0.18
SRM for Improvement, prospective*						
• BPI Severity	--	--	--	0.67	1.35	0.72
• BPI Interference	--	--	--	0.65	1.21	0.75
• BPI Total	--	--	--	0.73	1.31	0.76
• PEG	--	--	--	0.85	1.45	0.77
AUC for any improvement [‡]						
• BPI Severity	.82	.78	.73	.72	.76	.55
• BPI Interference	.74	.73	.68	.73	.77	.53
• BPI Total	.80	.79	.73	.75	.79	.54
• PEG	.76	.74	.71	.72	.79	.54
AUC for moderate improvement [‡]						
• BPI Severity	.83	.81	.74	.74	.78	.60
• BPI Interference	.72	.73	.69	.69	.76	.59
• BPI Total	.79	.80	.74	.73	.80	.60
• PEG	.75	.75	.72	.73	.75	.59
Responsiveness to Treatment						
Between-group Treatment Effect Size [‡]						
• BPI Severity	.56	.58	.38	--	--	--
• BPI Interference	.59	.46	.37	--	--	--
• BPI Total	.64	.58	.42	--	--	--
• PEG	.58	.52	.37	--	--	--

* SRM = standardized response mean, which is the within-group change effect size between two time points calculated as (T1 mean score – T2 mean score) / SD of change score). For each trial, SRM was calculated for three global change groups (improved, unchanged, worse). All trials used a retrospective global anchor question to estimate SRM, and three trials also used a prospective global anchor. The SRM is a type of effect size and therefore, SRMs of 0.2, 0.5, and 0.8 represent small, moderate, and large effect sizes respectively. In this table, the SRM is reported for the improved group only.

[‡]AUC = area under the curve as determined by ROC analysis. The discriminatory strength of the scale for determining any improvement and moderate improvement using the retrospective global anchor was estimated by the AUC. Whereas good AUCs for diagnostic tests are often > 0.80, AUCs for scales in determining improvement are typically lower, and what is more important is determining whether scales have comparable AUCs (i.e., similar responsiveness).

[‡]The between-group treatment effect size is calculated as: (intervention group mean change – control group mean change) / pooled change score SD. Treatment effect sizes of 0.2, 0.5, and 0.8 represent small, moderate, and large intervention effects, respectively. End-of-trial treatment effect size data was available for 3 of the 4 trials that had a usual care (rather than active comparator) control group.

[¶]Data in this column is from the 244 SCOPE participants who completed both baseline and 3-month assessments.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Synopsis of Minimally Important Difference and Responsiveness Metrics Across Six Trials

Variable	PEG	BPI Total	BPI Interference	BPI Severity	PEG-BPI Difference*
Cronbach's alpha					
Median	.73	.89	.88	.84	-.11 to -.16
Weighted Mean	.75	.89	.88	.83	-.08 to -.14
Minimally Important difference (MID)					
Global anchor – A Little Better					
Median	1.25	1.26	1.41	0.87	-.16 to +.38
Weighted Mean	1.28	1.06	1.28	0.81	.00 to +.47
0.35 standard deviation (SD)					
Median	0.74	0.67	0.81	0.62	-.07 to +.12
Weighted Mean	0.77	0.71	0.84	0.65	-.07 to +.12
0.50 standard deviation (SD)					
Median	1.05	0.95	1.15	0.88	-.10 to +.17
Weighted Mean	1.09	1.03	1.16	0.92	-.07 to +.17
1 SEM					
Median	1.08	0.66	0.77	0.73	+.31 to +.42
Weighted Mean	1.06	0.64	0.80	0.75	+.26 to +.42
2 SEM					
Median	2.16	1.33	1.53	1.46	+.63 to +.83
Weighted Mean	2.12	1.29	1.59	1.50	+.53 to +.83
Responsiveness					
SRM for improvement, retrospective					
Median	0.93	0.98	0.89	0.86	-.05 to +.07
Weighted Mean	0.89	0.90	0.81	0.84	-.01 to +.08
SRM for improvement, prospective					
Median	0.85	0.76	0.75	0.72	+.09 to +.13
Weighted Mean	1.01	0.92	0.86	0.90	+.09 to +.15
AUC for any improvement					
Median	0.73	0.77	0.73	0.75	-.04 to +.00
Weighted Mean	0.71	0.74	0.70	0.74	-.03 to +.01
AUC for moderate improvement					
Median	0.74	0.77	0.71	0.76	-.03 to +.04
Weighted Mean	0.72	0.75	0.70	0.76	-.04 to +.02
Between-group Treatment Effect Size					
Median	0.52	0.58	0.46	0.56	-.06 to +.06
Weighted Mean	0.51	0.57	0.50	0.52	-.06 to +.01

* Range of differences between the PEG and the 3 BPI MID/responsiveness metrics calculated as PEG metric minus BPI metric. For example, the PEG-BPI difference in the 6 trials for the 0.50 standard deviation weighted mean is $1.09 - 1.03 = +.06$ for the PEG-BPI Total difference; $1.09 - 1.16 = -.07$ for the PEG-BPI Interference difference; and $1.09 - 0.92 = +.17$ for the PEG-BPI Severity difference. Thus, the range is $-.07$ to $+.17$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript