Title: Establishment of cutpoints to categorize the severity of chronic pain using composite ratings with Rasch analysis

Running title: Optimal cutpoint for chronic pain classification

Authors: Chi-Wen Chien[a,b,*], Karl S. Bagraith[a,c], Asaduzzaman Khan[d], Michael Deen[e], Jia-Jia Syu[f], and Jenny Strong[a]

Authors' Affiliations:

[a] Occupational Therapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia.

[b] Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (SAR), China.

[c] Interdisciplinary Persistent Pain Centre, Gold Coast Hospital and Health Service, Gold Coast, Queensland, Australia.

[d] School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia.

[e] Metro South Persistent Pain Management Service, Princess Alexandra Hospital, Woolloongabba, Queensland, Australia.

[f] School of Public Health, The University of Queensland, Brisbane, Queensland, Australia.

Correspondence: Dr. Chi-Wen Chien, Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (SAR), China.

Tel: +852 2766 6703

Fax: +852 2330 8656

E-mail: Will.Chien@polyu.edu.hk

Category of the submission: Original article

Funding source: The first author was supported by a University of Queensland Postdoctoral Research Fellowship.

Conflicts of interest: The authors have no financial or other relationships that can lead to conflicts of interest.

'What's already known about this topic?'

- Single worst or average pain rating is commonly used to classify mild, moderate, and severe pain, but the optimal cutpoint is highly varied.
- The optimal cutpoint based on composite pain seems to offer more utility in pain classification.

'What does this study add?'

- Using composite pain, optimal classification for mild, moderate, and severe pain exhibited better discriminant ability than using single worst/average pain.
- The difficulty hierarchy of the least, worst, average, and current pain helps to screen people with irregular responses.

Abstract

Background: Establishment of cutpoints for classifying mild, moderate, and severe pain is commonly based on single rating of worst or average pain. However, single pain measure may serve as a brief and partial surrogate for composite pain ratings. This study aimed to base composite pain ratings to establish optimal cutpoint that maximized the difference of pain interference on daily function and compare its utility with those based on single worst and average pain.

Methods: Data was from a cohort study of 322 chronic pain patients. Brief Pain Inventory (including four items measuring the least, worst, average, and current pain) was administered. Rasch analysis and Serlin et al.'s (1995) method were used to derive optimal cutpoint.

Results: Rasch analysis calibrated the least, worst, average, and current pain items into a unidimensional hierarchy and produced composite pain measurement. The optimal cutpoint for composite pain (mild, ≤4; moderate, >4−6; severe, >6−10 on the 0−10 numeric rating scale) differed from those cutpoints for worst (≤6; >6−8; >8−10) and average pain (≤5; >5−7; >7−10). The optimal cutpoint for composite pain was better able than those for worst and average pain to distinguish among groups on patient-rated pain quality and quality of life. The optimal cutpoint for average pain had better discriminant ability than that for worst pain.

Conclusion: The results suggest that using optimal cutpoint for composite pain may be useful to classify clinically important groups in chronic pain patients and that average pain may be an alternative choice if a single item is used.

Intensity ratings for chronic pain are often categorized as mild, moderate, and severe to make clinical treatment decisions or provide a meaningful outcome in research (Anderson, 2005). This threefold pain categorization can be derived by designating cutpoints (CPs) on an 11-point numerical rating scale (NRS), where 0 indicates no pain and 10 represents worst possible pain. Many studies have established CPs on the 11-point NRS for the threefold pain categorization (Serlin et al., 1995; Jensen et al., 2001; Zelman et al., 2003; Hanley et al., 2006), but there is limited investigation in the impact of single versus composite pain ratings on CP establishment. Commonly a single rating of worst pain is used to establish CPs as its reduction is clinically important (Paul et al., 2005; Alschuler et al., 2012). On the other hand, average pain is considered to better reflect the breadth of chronic pain experience (Dworkin et al., 2005) and has been increasingly used in pain severity classification. However, single measures for worst or average pain serve merely as a brief and partial surrogate for composite measures that include more comprehensive aspects of pain (e.g., the least, worst, average, and current pain). This may result in a high variability of optimal CPs when using a single pain rating (Hirschfeld and Zernikow, 2013). In only two previous studies that focused on composite pain ratings by averaging several single pain ratings (Fejer et al., 2005; Dihle et al., 2006), the optimal CPs seemed to be more stable across groups with varied characteristics, and to yield higher statistics in discriminating between different severity groups on pain interference, compared to the CPs generated using single pain ratings. More investigations into the utility of composite scores used to categorize the severity of pain are necessary.

Rasch analysis staging approach, as proposed by Jette et al. (2008), offers a sophisticated method to establish CPs based on composite scores and has not yet been applied to pain severity classification. In this approach, individual items are calibrated by Rasch analysis onto the same unidimensional hierarchy based on their difficulty of endorsement. A-priori knowledge of clinically relevant criterion behaviours is subsequently used to identify CPs for distinct ‘stages’ along the established hierarchical continuum. Rasch analysis produces model-driven, interval-level measurement of composite scores, which may provide statistically greater precision in differentiating patient groups (Khan et al., 2013). The established item-difficulty hierarchy may also help to decide which stages (or severity groups) respondents belong according to the logicality of their responses to individual items.

The purpose of this study was to apply Rasch analysis staging approach to the Brief Pain Inventory (BPI) to derive optimal CPs, based on composite pain ratings, for mild, moderate, and severe pain in adults with chronic pain. The determination of optimal CPs was based on the pain severity classifications (for mild, moderate and severe pain) that maximized the difference of pain interference on daily function. We also sought to investigate the variability and discriminant ability of optimal composite pain CPs, in comparison with the CPs based on single item ratings of worst and average pain.

Methods

Source data

The current analysis was based on existing clinical data that consisted of 322 patients who attended a pain management program provided by a Multidisciplinary Pain Centre (MPC) in metropolitan Brisbane, Australia, between 2005−2009. To be eligible for the program, patients needed to: a) be ≥18 years old; b) have nonmalignant pain; and c) experience pain for >3 months. As part of the MPC's quality assurance procedure, patients were asked to complete a battery of self-report outcome measures on the first day of the program. These baseline responses in the de-identified form were analyzed in this study. The hospital's Institutional Review Board for Low Risk Research approved the study.

Table 1 shows demographic and health status information of the patients. These patients were nearly equally balanced between males and females, and their age ranged from 18 to 80 years. Almost half (47.5%) of the patients had senior high school or higher education qualification and 50% were unemployed due to pain. Their mean (SD) pain duration was

136.5 (118.0) months. Using the tick-box approach, approximately 73.3% reported pain in more than one bodily area (average=3.9 pain locations), in which lower back (71.7%), shoulder/arms/hands (51.9%), neck (46.7%), and calves/ankles/feet (42.5%) were the four most frequently reported pain locations. The remaining patients reported a single region of pain in the head/face/mouth (1.9%), upper back (0.6%), upper shoulder and upper limbs (4.0%), lower back and lower spine (3.7%), lower limbs (3.1%), lower back and lower limbs (9.0%), and other locations (2.8%).

Measures

Brief Pain Inventory (BPI)

The BPI (Cleeland and Ryan, 1994) is comprised of 11 items that can be divided into one pain severity scale and one pain interference scale. The severity scale has 4 items assessing pain intensity at its least during the last 24 hours, worst during the last 24 hours, average and current level, while the interference scale includes 7 items assessing pain's interference with daily functions (i.e., general activity, mood, walking ability, normal work, relations with others, sleep, and enjoyment of life). Each item is rated on a 0−10 NRS, where 0 indicates 'no pain' or 'does not interfere' and 10 indicates 'pain as bad as you can imagine' or 'completely interferes'. The responses to the 4 severity items and 7 interference items are typically averaged to derive distinct pain severity and interference scale totals, respectively. Acceptable psychometric properties of the BPI have been reported in adults with chronic nonmalignant pain (Cleeland and Ryan, 1994; Dworkin et al., 2005).

Short-Form McGill Pain Questionnaire (SF-MPQ)

The SF-MPQ (Melzack, 1987) consists of 15 descriptors (11 sensory and 4 affective) that are rated on an intensity scale (0=none, 1=mild, 2=moderate, and 3=severe). The sensory and affective scores are calculated; higher scores indicate more severe sensory and affective pain qualities, respectively. The SF-MPQ has shown acceptable psychometric properties in different groups of pain patients (Strand et al., 2008).

Depression, Anxiety, Stress Scale-21 (DASS-21)

The DASS-21 (Lovibond and Lovibond, 1995) is comprised of 3 scales that assess negative emotional states of depression, anxiety, and stress. Each of these scales has 7 items that ask respondents to indicate the extent to which each statement applied to them over the last week. Four response categories can be chosen for each item. A total score (0−21) is obtained for each DASS-21 subscale; higher scores indicate greater emotional distress. The DASS-21 has demonstrated satisfactory psychometric properties in patients with chronic pain (Taylor et al., 2005).

World Health Organization Quality of Life-Brief version (WHOQOL-BREF)

The WHOQOL-BREF (The WHOQOL Group, 1998) includes 2 generic items (overall quality of life [QOL] and general health) and 24 items that can be further classified into 4 domains: physical, psychological, social relationships, and environmental domains. Each item is rated on a 5-point scale. Responses from the 2 generic items are calculated as a single score with a range of 1−5. Domain total scores are calculated and potential scores for each domain range from 4−20. Higher scores indicate better QOL as reflected by the items or domains. The psychometric properties of the WHOQOL-BREF have been supported across various clinical groups (The WHOQOL Group, 1998).

Data analysis

In this study, the Rasch analysis staging approach (Jette et al., 2008) was applied to establish optimal CPs based on BPI pain severity scale for classifying mild, moderate, and severe pain. First, the 4 BPI pain severity items (i.e., the worst, least, average, and current pain) were calibrated into a unidimensional and hierarchical order using Rasch analysis. Second, optimal composite pain CPs were derived using Rasch-based scoring and, for comparison, we also derived optimal CPs based on the single ratings of worst and average pain. Finally, we investigated the variability and discriminant abilities of optimal CPs based on composite pain ratings as well as single worst and average pain ratings. Each step is detailed as follows.

Pain severity items calibration using Rasch analysis

Rasch analysis is one type of Item Response Theory models that investigates the probabilistic relationship between participants' underlying latent trait (e.g., pain intensity in this study) and their responses to assessment items (Rasch, 1960). Initially, a prior Rasch analysis was conducted to confirm that the 4 BPI pain severity items constituted a unidimensional construct (i.e., whether they measured a single construct) (Bagraith et al., submitted for publication). The results were summarized in Results S1.

Once the unidimensionality was confirmed, the BPI pain severity items were placed in a continuum from easy to difficult in a 'keyform' generated by Rasch analysis software, WINSTEPS (Linacre, 2011). The keyform (see Figure 1) can provide immediate, useful information about the relationship between respondents' underlying latent trait (i.e., pain severity) and the difficulty levels of each BPI item. It looks similar to a checklist, with all the BPI items on one side and the numbers corresponding to the 0-10 NRS of each item placed on the other side. On the left-hand side, the BPI items are ordered on the basis of Rasch item-difficulty calibrations, in which easier items (i.e., those on which patients are likely to report a larger magnitude of pain) are towards the bottom of the keyform and harder items (i.e., those on which patients are likely to report lesser magnitude of pain) are at the top. On the right-hand side, the 0-10 NRS of each item are presented like stairsteps from left to right. Each NRS rating for each item is associated with a difficulty calibration. The boundaries between each of 2 adjacent ratings are also provided in the keyform to denote the probabilistic midpoints. In Rasch analysis, the difficulty measures are expressed as 'logits' which represents the log-odds ratio of the probability of a patient experiencing pain at a particular NRS rating and are considered as an interval-level unit of measurement.

Derivation of optimal CPs

Given that the relationship between pain and its associated disabilities or performance limitations appears to be non-linear (Turner et al., 2004), it is unlikely to use clinically relevant criterion behaviours to differentiate from the stages of mild, moderate, and severe pain. Therefore, in this study, we followed the statistical strategy described by Serlin et al. (1995) to determine optimal CPs for mild, moderate, and severe pain in the Rasch keyform. This method has been used widely in previous studies for establishment of optimal pain CPs (Jensen et al., 2001; Zelman et al., 2003; Fejer et al., 2005; Zelman et al., 2005; Dihle et al., 2006; Hanley et al., 2006; Alschuler et al., 2012).

To enable Serlin et al.'s method (1995), the BPI Rasch-derived composite scores (based on logits) were transformed to a 0−10 range. The transformed composite pain scale was then classified into mild, moderate, and severe stages for 10 possible CP sets between the ratings of 3 and 8, including CP35, CP36, CP37, CP38, CP46, CP47, CP48, CP57, CP58, and CP68. Each CP label denotes the upper values in the mild and moderate categories (e.g., for CP36, mild: ≤3, moderate: >3–6, and severe: >6–10; see Figure S1 for illustration). We subsequently used the 7 BPI interference items as external reference measures (i.e., the dependent variables) in multivariate analysis of variance (MANOVA) for each of the proposed pain CP sets (as the group variable). The criterion used to determine the optimal CPs was the largest F value for the overall between-group variance as indicated by Pillai's trace, Wilks' lambda, and Hotelling's trace F statistics in the MANOVA. The CPs with the largest F statistics maximized the pain interference differences between the pain severity groups and hence provided the optimal classification for mild, moderate, and severe pain (Serlin et al., 1995). This MANOVA analytic approach was also employed to derive optimal CPs based on the single item ratings of worst and average pain.

Examinations of the variability and discriminant ability of optimal CPs

The variability of optimal CPs based on the composite or single worst/average pain ratings were further investigated using the bootstrap re-sampling procedure proposed by Hirschfeld and Zernikow (2013). In this procedure, 1000 pseudosamples with the same number of participants from the study sample were drawn at random (with replacement). For each pseudosample, optimal CPs were determined using the aforementioned approach by Serlin et al. (1995), among the 10 examined sets. The number of occasions that each CP set was deemed optimal across the 1000 pseudosamples, was considered an indicator of its variability. The goodness-of-fit tests were performed to examine whether there were significant differences in the frequency among the 10 CP sets being deemed optimal within the 1000 pseudosamples.

We also compared the discriminant ability of the optimal CPs (based on the composite, worst, or average pain ratings) in relation to pain quality, emotional functioning, and QOL. Simple linear regression analysis was performed individually by using either one of the SF-MPQ subscales (sensory and affective), the DASS-21 scales (depression, anxiety, and stress), or the WHOQOL-BREF outcomes (2 generic items and physical, psychological, social relationships, and environmental domains) as the dependent variable. The independent variable was the optimal CP based on the composite, worst, or average pain ratings. The extent to which the optimal CP explained the variance of each dependent variable was reported, and the significance level was set at P-value<0.05.

Results

<u>Calibration of pain severity items in Rasch-derived keyform</u>

While the 4 BPI pain severity items were confirmed by Rasch analysis to form a substantial unidimensional construct, 63 patients of this study sample were identified to exhibit misfit (i.e., infit and outfit mean square > 1.4) with respect to Rasch model's hierarchical expectations. These 63 patients were subsequently omitted from further analyses due to their likely spurious influence on the keyform generation and CP derivation.

Figure 1 illustrates the final Rasch-derived keyform, in which the 4 BPI items are placed in a hierarchical order according to their difficulty of endorsement. Least pain was found to be the most difficult item (i.e., having a lower probability to be reported with higher intensity than other BPI items), while patients most easily reported higher intensity ratings on the worst pain item (i.e., the easiest item). The difficulty continuum for Rasch-based composite pain scores (made up of the 4 BPI items) ranged over 19.61 logits (from -9.90 to 9.71). For ease of interpretation, the logit scores were linearly transformed into a 0−10 scale (with 1 decimal allowed to provide more precise differentiation) in accordance with the original BPI NRS format (see Figure 1).

<u>Derivation of optimal CPs for mild, moderate, and severe pain</u>

Table 2 shows the *F* values for the MANOVAs, calculated for the 10 proposed CP sets based on the BPI composite pain scores (using the 0−10 transformed scale), worst pain ratings, and average pain ratings relative to the BPI interference scale. Different optimal CPs were identified for each of the pain ratings, based on the results of the largest *F* statistics. According to the optimal CPs for composite pain (i.e., CP46), 7.7% of the sample (n=20) had mild pain, 56.8% (n=147) had moderate pain, and 35.5% (n=92) had severe pain. By comparison, when classified according to the worst pain, CP68 was optimal, in which 21.6% of the sample (n=56) had mild pain, 47.9% (n=124) had moderate pain, and 30.5% (n=79) had severe pain. For average pain, CP57 was optimal, which classified 41.7% (n=108), 40.9% (n=106), and 17.3% (n=45) of the sample into mild, moderate, and severe pain groups, respectively. The optimal CP values for the different pain ratings were embedded together in the Rasch-derived keyform (Figure 2, Part A) to provide a visual illustration for the classification of mild, moderate, and severe pain.

<u>Variability of optimal CPs in bootstrapping samples</u>

The results from the bootstrap re-sampling analysis indicated that CP46, which was deemed optimal for the composite pain scores, was identified as optimal in 328 (32.8%) of the 1000 pseudosamples. The goodness-of-fit analysis (chi-square=977, df=8, p<0.001) confirmed the significant difference between its frequency and those of the other 9 CP sets, particularly for the second most frequent CP set (i.e., CP47) which appeared in 27.4% of the pseudosamples. As for the single average pain rating, CP57 was identified as optimal in 452 (45.2%) of the bootstrapping pseudosamples, which was far more than the second most frequent CP set (i.e., CP58, optimal in 25.1% of the pseudosamples). However, CP68 based on the worst pain was identified as optimal in only 26.5% of the pseudosamples, and this proportion was less than the CP57 (optimal in 32.1% of the pseudosamples).

Comparison of optimal CPs in discriminating pain quality, emotional functioning, and QOL

For the CP46 classification based on composite pain, it was found to significantly explain 1.8−20.5% of the total variance of all outcome domains, with the exception of the WHOQOL-BREF social relationship and environmental domains (see Table S1). Similarly the CP57 classification based on average pain significantly explained 1.6−17.4% of the total variable of all but not the WHOQOL-BREF general health item as well as social relationship and environmental domains. However, the CP68 classification based on worst pain was found to significantly discriminate only the scores of the SF-MPQ sensory and affective subscales as well as the DASS-21 depression and anxiety scales, with 1.6−18.5% of the total variance explained.

Discussion

This study is the first to combine Rasch analysis with Serlin et al.’s method (1995) to derive optimal CPs for composite pain ratings in people with chronic pain. The CP46 for composite pain ratings (mild, ≤4; moderate, >4−6; severe, >6−10), in comparison to the other 9 CP sets, exhibited the largest intergroup differences in pain interference. The composite pain CP46 also appeared to be more able than the optimal CPs for worst pain (CP68) and average pain (CP57) to distinguish among pain severity groups on patients-rated outcomes (particularly for pain quality and QOL). Taken together, this study provides evidence for using the composite pain CP46 to classify mild, moderate, and severe pain in people with chronic pain.

The present evidence for composite pain ratings when used to classify pain is consistent with Dihle et al.'s (2006) and Fejer et al.'s (2005) findings, however, the optimal CP results differed. In Dihle et al.'s study (2006), the CP35 was found to be optimal for the mean scores of the worst and average pain ratings. Fejer et al.'s (2005) found the optimal CP36 for the composite scores that average three pain ratings (i.e., worst, average, and current pain). In contrast, the present study utilized Rasch analysis to calibrate the worst, least, average, and current pain ratings onto a unidimensional and hierarchical continuum that was subsequently used to generate model-driven estimates of each participant's composite pain. Rasch-derived estimates are purportedly interval-level (Bond and Fox, 2007) and may be superior when distinguishing between groups (Khan et al., 2013) or detecting treatment-related changes (Norquist et al., 2004). Furthermore, the etiologies of the pain problem in patient populations are also different between this study (i.e., those who had nonmalignant chronic pain of a heterogenous nature) and previous studies [e.g., those with neck pain (Fejer et al., 2005) or acute postoperative pain (Dihle et al., 2006)]. These methodological variations (i.e., the scoring approaches to composite pain rating or participants' characteristics) may account for the differential results of the optimal composite pain CP between the present study and previous studies.

The Rasch analysis staging approach offers further methodological advantages, including robust detection of participants with unusual or haphazard response patterns (Meijer, 2003). In this study, approximately one in five participants (19.6%) provided response patterns that were incongruent with that expected by the model (i.e., misfit). This high misfitting rate may have resulted from the stricter criterion (Kottorp et al., 2003) that we used to enhance the confidence in the accuracy of the CP analysis, compared to more lenient criterion (e.g., infit and outfit MnSQ >2.0) as used by other studies (Chien and Bond, 2009). However, a detailed investigation on those misfitting participants revealed that they experienced significantly lower pain interference and depression as well as higher QOL in physiological, social relationships, and environmental domains (see Table 1), although they had similar intensity on both single and composite pain to well-fit participants. From a clinical point of view, this implies that those misfitting participants may be better able to cope with the impact of the pain on their physical and emotional functioning. It is accordingly possible that some may have responded unexpectedly irregularly on pain intensity items (e.g., endorsing the 'same' response category for their worst and least pain) due to their fewer perceived impacts related to pain. Conversely, some may have weighed certain types of pain strongly (e.g., worst or current pain) even though their 'average' pain is much lower. For the participants whose responses misfit, unfortunately, there seem no simple approaches currently available to generate appropriate scores by adjusting their misfitting responses. Therefore, those misfitting participants were omitted from the study sample. This reiterates that identification of patients with aberrant pain response patterns is important before using composite pain ratings to classify their pain intensity in clinical and research settings.

In practice it is common that clinicians and researchers categorize patients into discrete categories (e.g., mild, moderate, and severe) from 0-10 NRS pain ratings for everyday clinical practice or for interpreting epidemiological studies. The present study suggests that the use of composite pain ratings (made up of the least, worst, average, and current pain ratings) with the CP46 to classify patients into mild, moderate, and severe pain may be more useful. In particular, the keyform provides a simple, visual tool that records patients' responses to the BPI items and determines their pain classification without the need for cumbersome software. For example, clinicians can circle the patients' responses in the keyform and evaluate whether at least 3 items belong to the same pain classification zone and these items are also close to each other (see Part B of Figure 2 for illustration). If all conditions are fulfilled, the patient's pain severity can be classified into that category. On the other hand, the keyform also allows ready identification of misfitting responses that are not amenable to the composite pain CP46 classification (e.g., Part C of Figure 2) and the individuals for whom pain ratings should be interpreted with caution (e.g., due to likely scale misinterpretation). An instruction guide about how to use the keyform is provided for clinicians and researchers of interests (see Appendix S1).

In addition to the optimal CP46 based on composite pain, this study found different optimal CP68 and CP57 for the single ratings of worst and average pain, respectively. Such difference in optimal CPs for the 2 pain references has been recognized (Fejer et al., 2005; Paul et al., 2005; Hanley et al., 2006), but the issue was limitedly explored from the perspective of item difficulty and coverage (Paul et al., 2005). In this study, the Rasch-based keyform reveals that worst pain is the easiest item and has slightly skewed difficulty coverage (i.e., more towards the patients who would be classified with mild pain if using the composite pain CPs, see Figure 2). By contrast, the average pain item quite evenly represents the full spectrum of participants' pain intensity. Furthermore, the optimal average pain CP57 was found to have the most stability in bootstrapping samples. It was also able to discriminate patients' pain quality, emotional functioning, and most QOL domains, compared to the worst pain CP68 that differentiated poorly the severity groups in terms of emotional stress and all QOL domains. It is thus suggested that, when employing a single item for pain classification, average pain could be a better choice than worst pain in people with chronic pain. In particular, the average pain CP57 could be used when patients exhibit irregular pain response patterns and the optimal composite pain CP46 is not suitable for use.

This study has several limitations that should be considered. First, the study included patients with heterogeneous chronic nonmalignant pain and most of them reported multiple pain locations, as is typical of many patients with chronic pain (Croft et al., 2007). Variations in optimal CPs have been seen across people with different pain problems or pain locations

(Hirschfeld and Zernikow, 2013). The optimal CPs found in the current study may be therefore not readily generalizable to homogenous groups of chronic pain patients with similar pain etiology or particular pain location. Second, this study sample was limited by a completion rate of 64.9%, and it is possible that exclusion of those patients who did not complete all the self-report measures may affect the study findings. Third, there were a considerable number of patients with aberrant response patterns who were unable to be classified using the Rasch-derived composite pain CP46. While we have suggested the average pain CP57 to be an alternative classification option, it is necessary to further investigate misfitting patients and ascertain the reasons behind their aberrant response patterns if this occurs in clinical practice. Fourth, the approach developed by Serlin et al. (1995) to determine the optimal CPs based on the largest amount of explained variance in the MANOVA has been criticized (Hirschfeld and Zernikow, 2013). However, there is still no existing statistic to determine how large the MANOVA value must be in order to find the optimal CPs (Zelman et al., 2005). This study has used the bootstrapping technique suggested by Hirschfeld and Zernikow (2013), but the variability was relatively large for the optimal composite pain CP46. Furthermore, the 7 pain interference items (used as external reference measures) or the single worst/average pain item are ordinal-level measurement and may be therefore not valid for the MANOVA. Future studies that adopt other approaches to development of the optimal pain classification are warranted.

Conclusion

This study established the optimal cutpoint based on composite pain for mild (≤4), moderate (>4−6), and severe (>6−10) categories among people with chronic pain. This optimal cutpoint for composite pain demonstrated the ability to discriminate levels of pain severity associated with pain quality, emotional functioning, and QOL. In addition, Rasch analysis calibrated the 4 BPI pain severity items in a progressive hierarchical map to enable identification of patients with irregular response patterns prior to classification of their composite pain ratings. Further studies are needed to replicate Rasch analysis with Serlin et al.'s approach (1995) or other approaches to development of the classification system based on composite pain ratings in different populations of chronic pain patients.

Acknowledgements

The authors would like to thank the patients for completing the measures and the MPC clinicians for the delivery of the program and diligent administration and collection of the questionnaire battery.

Authors Contributions

Each of the listed authors has confirmed their contribution with respect to the journal's authorship guidelines. All authors contributed to the study conception and design, data analysis, results interpretation, and manuscript drafting. Dr. Bagraith and Mr. Deen additionally contributed to acquisitions of data. All authors have approved the final version to be published.

References

Alschuler, K.N., Jensen, M.P., Ehde, D.M. (2012). Defining mild, moderate, and severe pain in persons with multiple sclerosis. *Pain Med* **13**, 1358-1365.

Anderson, K.O. (2005). Role of cutpoints: Why grade pain intensity? *Pain* **113**, 5-6.

Bond, T.G., Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (Mahwah, NJ: Lawrence Erlbaum Associates).

Chien, C.W., Bond, T.G. (2009). Measurement properties of fine motor scale of Peabody developmental motor scales-second edition: A Rasch analysis. *Am J Phys Med Rehabil* **88**, 376-386.

Cleeland, C.S., Ryan, K.M. (1994). Pain assessment: Global use of the Brief Pain Inventory. *Ann Acad Med Singapore* **23**, 129-138.

Croft, P., Dunn, K.M., Von Korff, M. (2007). Chronic pain syndromes: You can't have one without another. *Pain* **131**, 237-238.

Dihle, A., Helseth, S., Paul, S.M., Miaskowski, C. (2006). The exploration of the establishment of cutpoints to categorize the severity of acute postoperative pain. *Clin J Pain* **22**, 617-624.

Dworkin, R.H., Turk, D.C., Farrar, J.T., Haythornthwaite, J.A., Jensen, M.P., Katz, N.P., Kerns, R.D., Stucki, G., Allen, R.R., Bellamy, N., Carr, D.B., Chandler, J., Cowan, P., Dionne, R., Galer, B.S., Hertz, S., Jadad, A.R., Kramer, L.D., Manning, D.C., Martin, S., McCormick, C.G., McDermott, M.P., McGrath, P., Quessy, S., Rappaport, B.A., Robbins, W., Robinson, J.P., Rothman, M., Royal, M.A., Simon, L., Stauffer, J.W., Stein, W., Tollett, J., Wernicke, J., Witter, J. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* **113**, 9-19.

Fejer, R., Jordan, A., Hartvigsen, J. (2005). Categorising the severity of neck pain: Establishment of cut-points for use in clinical and epidemiological research. *Pain* **119**, 176-182.

Hanley, M.A., Masedo, A., Jensen, M.P., Cardenas, D., Turner, J.A. (2006). Pain interference in persons with spinal cord injury: Classification of mild, moderate, and severe pain. *J Pain* **7**, 129-133.

Hirschfeld, G., Zernikow, B. (2013). Variability of "optimal" cut points for mild, moderate, and severe pain: Neglected problems when comparing groups. *Pain* **154**, 154-159.

Jensen, M.P., Smith, D.G., Ehde, D.M., Robinsin, L.R. (2001). Pain site and the effects of amputation pain: Further clarification of the meaning of mild, moderate, and severe pain. *Pain* **91**, 317-322.

Jette, A.M., Norweg, A., Haley, S.M. (2008). Achieving meaningful measurements of ICF concepts. *Disabil Rehabil* **30**, 963-969.

Khan, A., Chien, C.W., Brauer, S.G. (2013). Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *J Clin Epidemiol* **66**, 681-687.

Kottorp, A., Bernspång, B., G., F.A. (2003). Validity of a performance assessment of activities of daily living for people with developmental disabilities. *J Intellect Disabil Res* **47**, 597-605.

Linacre, J.M. (2011). *User's guide to WINSTEPS* (Chicago, IL: MESA Press).

Lovibond, P.F., Lovibond, S.H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* **33**, 335-343.

Meijer, R.R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychol Methods* **8**, 72-87.

Melzack, R. (1987). The short-form McGill Pain Questionnaire. *Pain* **30**, 191-197.

Norquist, J.M., Fitzpatrick, R., Dawson, J., Jenkinson, C. (2004). Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* **42**, I25-36.

Paul, S.M., Zelman, D.C., Smith, M., Miaskowski, C. (2005). Categorizing the severity of cancer pain: Further exploration of the establishment of cutpoints. *Pain* **113**, 37-44.

Rasch, G. (1960). *Probabilistic Models for Some Intelligent and Attainment Tests* (Copenhagen, Dermark: Dermarks Paedagogiske Institute).

Serlin, R.C., Mendoza, T.R., Nakamura, Y., Edwards, K.R., Cleeland, C.S. (1995). When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain* **61**, 277-284.

Strand, L.I., Ljunggren, A.E., Bogen, B., Ask, T., Johnsen, T.B. (2008). The Short-Form McGill Pain Questionnaire as an outcome measure: Test-retest reliability and responsiveness to change. *Eur J Pain* **12**, 917-925.

Taylor, R., Lovibond, P.F., Nicholas, M.K., Cayley, C., Wilson, P.H. (2005). The utility of somatic items in the assessment of depression in patients with chronic pain: A comparison of the Zung Self-Rating Depression Scale and the Depression Anxiety Stress Scales in chronic pain and clinical and community samples. *Clin J Pain* **21**, 91-100.

The WHOQOL Group (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med* **28**, 551-558.

Turner, J.A., Franklin, G., Heagerty, P.J., Wu, R., Egan, K., Fulton-Kehoe, D., Gluck, J.V., Wickizer, T.M. (2004). The association between pain and disability. *Pain* **112**, 307-314.

Zelman, D.C., Dukes, E., Brandenburg, N., Bostrom, A., Gore, M. (2005). Identification of cut-points for mild, moderate and severe pain due to diabetic peripheral neuropathy. *Pain* **115**, 29-36.

Zelman, D.C., Hoffman, D.L., Seifeldin, R., Dukes, E.M. (2003). Development of a metric for a day of manageable pain control: Derivation of pain severity cut-points for low back pain and osteoarthritis. *Pain* **106**, 35-42.

0-10 numeric rating scale	Item Description
0 : 1 : 2 : 3 :4 : 5: 6 : 7 : 8 : 9 : 10	Least pain
0 : 1 : 2 : 3 : 4: 5 : 6: 7 : 8 : 9 : 10	Average pain
0 : 1 : 2 : 3 : 4 :5 : 6 : 7 : 8 : 9 : 10	Current pain
0 : 1 : 2 : 3 : 4 :5 : 6 :7 : 8 : 9 : 10	Worst pain
−10 −8 −6 −4 −2 0 2 4 6 8 10	Measure (based on logits)
0 1 2 3 4 5 6 7 8 9 10	Measure (based on 0-10 transformed scale)

Figure 1 Keyform for the BPI pain severity scale.

On the left-hand side, the BPI items are ordered according to Rasch item-difficulty calibrations, in which easier items are towards the bottom and harder items are at the top. On the right-hand side, the 0-10 NRS of each item increase from left to right, in correspondence with increased difficulty calibrations. The difficulty calibrations are based on Rasch logit scores in the keyform and is then transformed scores on a 0−10 scale.

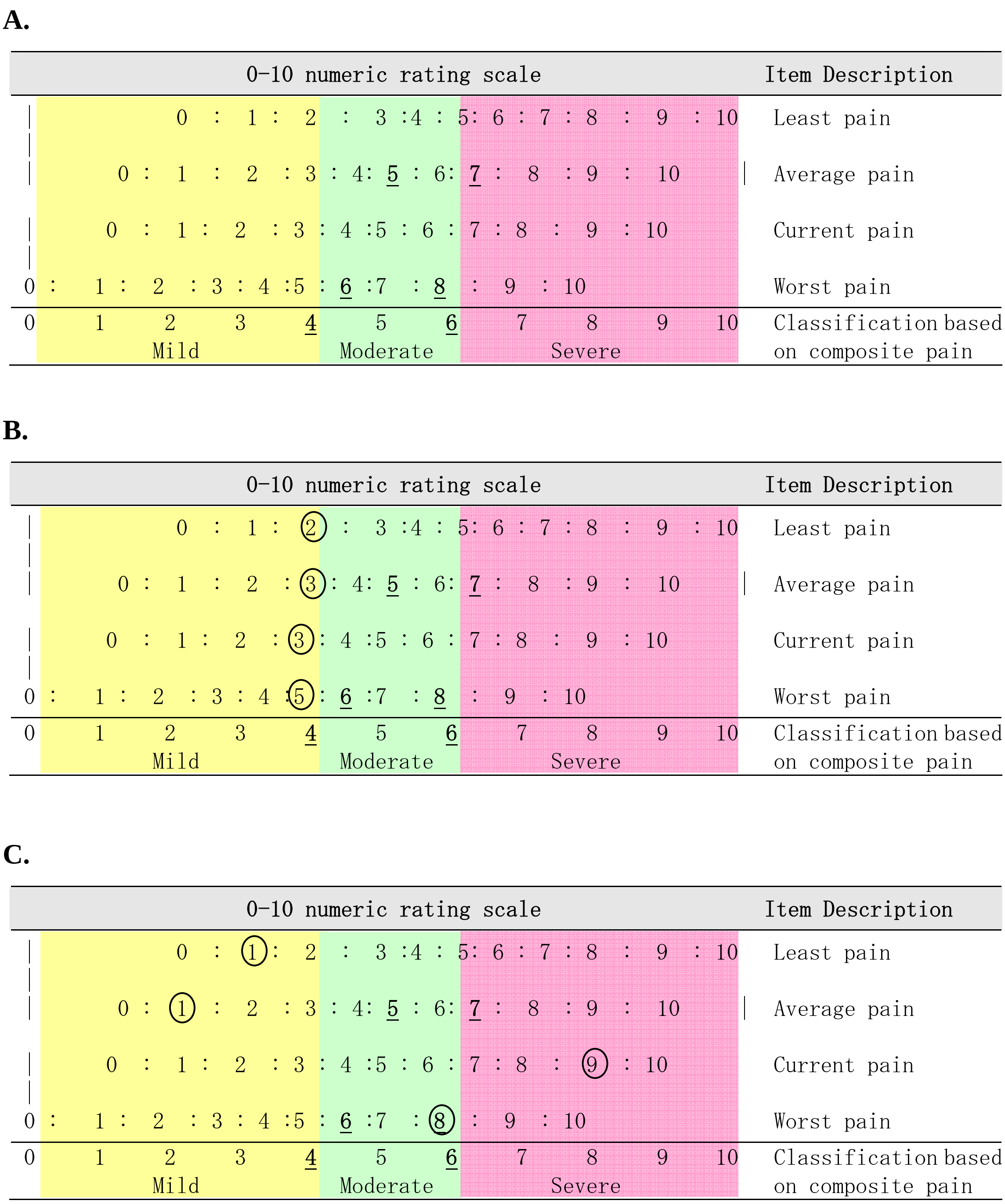


Figure 2 Keyforms with optimal cutpoints for the worst, average, and composite pain. (A) Blank Keyform. (B) Keyform for a person with well-fit response pattern. (C) Keyform for a person with misfit response pattern. Underlined values are the optimal cutpoints between mild and moderate pain or between moderate and severe pain. The 3 highlighted zones were based on the classification using the composite pain. Circled values are the patients pain responses retrieved from real data.

Table 1 Demographic characteristics and assessment results of participants

	Total group (n=322)	Well-fit group[a] (n=259)	Misfit group[a] (n=63)
Gender, male (female)[b]	160 (161)	132 (127)	38 (34)
Age (years), mean ± SD	48.2 ± 12.4	47.6 ± 12.4	50.8 ± 12.2
Education level, n (%)			
Primary school	40 (12.4)	32 (12.4)	8 (12.7)
Junior high school	126 (39.1)	100 (38.6)	26 (41.3)
Senior high school	63 (19.6)	49 (18.9)	14 (22.2)
Tertiary non-university	57 (17.7)	50 (19.3)	7 (11.1)
Tertiary - university	33 (10.2)	25 (9.7)	8 (12.7)
Unreported	3 (0.9)	3 (1.2)	0
Employment status, n (%)			
Full-time employment	23 (7.1)	18 (6.9)	5 (7.9)
Part-time employment	26 (8.1)	21 (8.1)	5 (7.9)
Retired	58 (18.0)	45 (17.4)	13 (20.6)
Home duties	34 (10.6)	25 (9.7)	9 (14.3)
Unemployed due to pain	161 (50.0)	135 (52.1)	26 (41.3)
Unemployed due to other reasons	15 (4.7)	12 (4.6)	3 (4.8)
Unreported	5 (1.5)	3 (1.2)	2 (3.2)
Pain duration (months), mean ± SD	136.5 ± 118.0	133.6 ± 115.9	148.2 ± 126.6
BPI pain severity index, mean ± SD	5.9 ± 1.6	5.9 ± 1.6	5.8 ± 1.5
BPI interference index, mean ± SD	6.0 ± 2.0	6.2 ± 2.0	5.6 ± 1.9*
SF-MPQ Sensory, mean ± SD	13.6 ± 6.9	13.8 ± 6.8	12.8 ± 7.3
SF-MPQ Affective, mean ± SD	4.5 ± 3.2	4.5 ± 3.1	4.5 ± 3.5
DASS-21 Depression, mean ± SD	7.9 ± 5.9	8.3 ± 5.9	6.3 ± 5.6*
DASS-21 Anxiety, mean ± SD	6.0 ± 4.6	6.1 ± 4.7	5.5 ± 4.1
DASS-21 Stress, mean ± SD	9.1 ± 5.5	9.4 ± 5.5	8.3 ± 5.6
WHOQOL-BREF Overall, mean ± SD	2.9 ± 0.9	2.8 ± 0.9	3.1 ± 0.9
WHOQOL-BREF Health, , mean ± SD	2.2 ± 1.0	2.1 ± 1.0	2.3 ± 1.0
WHOQOL-BREF Physical, mean ± SD	9.4 ± 2.3	9.3 ± 2.4	9.6 ± 2.0
WHOQOL-BREF Psychological, mean ± SD	11.8 ± 3.1	11.5 ± 3.1	13.2 ± 3.2*
WHOQOL-BREF Social, mean ± SD	12.4 ± 3.6	12.2 ± 3.6	13.2 ± 3.6*
WHOQOL-BREF Environmental, mean ± SD	13.3 ± 2.7	13.2 ± 2.7	14.0 ± 2.7*

[a] The division between well-fit and misfit groups was based on Rasch-based goodness-of-fit.

[b] One participant did not report his/her gender.

* indicates P-value < 0.05 when comparing the well-fit and misfit groups

BPI, Brief Pain Inventory; SF-MPQ, short form McGill Pain Questionnaire; DASS-21, short form of the Depression, Anxiety and Stress Scale; WHOQOL-BREF, World Health Organization Quality of Life-Brief version.

Table 2 Multivariate analysis of variance to determine optimal cutpoints using the worst, average, and composite pain scores and Brief Pain Inventory interference items

	Pillai's trace	Wilks' lambda	Hotelling's trace
Worst pain			
CP35	1.478	1.485	1.492
CP36	2.208	2.226	2.245
CP37	3.500	3.585	3.669
CP38	3.588	3.641	3.693
CP46	2.113	2.140	2.166
CP47	3.360	3.460	3.558
CP48	3.189	3.245	3.302
CP57	3.899	3.958	4.017
CP58	3.848	3.887	3.925
CP68	4.310	4.414	4.517
Average pain			
CP35	4.644	4.937	5.230
CP36	4.421	4.686	4.951
CP37	3.975	4.184	4.392
CP38	2.876	2.929	2.982
CP46	4.403	4.587	4.771
CP47	3.870	4.069	4.268
CP48	2.870	2.911	2.952
CP57	5.275	5.625	5.975
CP58	4.953	5.101	5.250
CP68	3.584	3.612	3.641
Composite pain[a]			
CP35	3.504	3.625	3.744
CP36	3.587	3.738	3.888
CP37	3.165	3.245	3.324
CP38	1.873	1.888	1.904
CP46	4.870	5.134	5.396
CP47	4.851	5.097	5.343
CP48	3.684	3.822	3.960
CP57	4.674	4.882	5.090
CP58	4.522	4.726	4.929
CP68	4.211	4.355	4.499

[a] Transformed scores (from Rasch-based logit scores to the 0−10 scale) were used for composite pain scores.

Note: Underlined values were the largest *F* statistics in each type of pain.

CP, cutpoint

Table S1 Comparison of pain quality, emotional functioning, and quality of life among the three pain severity groups based on optimal cutpoints for the worst, average, and composite pain

Type of pain (optimal cutpoint)	Classification			R^2	P-value
	Mild (mean)	Moderate (mean)	Severe (mean)		
Based on Worst pain (CP68)					
SF-MPQ Sensory	9.72	14.02	18.13	0.185	<0.001*
SF-MPQ Affective	3.12	4.53	5.91	0.103	<0.001*
DASS-21 Depression	7.41	7.78	9.69	0.021	0.019*
DASS-21 Anxiety	5.81	5.39	7.29	0.016	0.041*
DASS-21 Stress	9.34	8.58	10.46	0.008	0.157
WHOQOL-BREF Overall	2.96	2.80	2.67	0.013	0.073
WHOQOL-BREF Health	2.27	2.18	1.96	0.013	0.069
WHOQOL-BREF Physical	9.68	9.42	8.91	0.015	0.053
WHOQOL-BREF Psychological	11.46	11.64	11.18	0.002	0.534
WHOQOL-BREF Social	11.92	12.55	11.91	<0.001	0.846
WHOQOL-BREF Environmental	12.96	13.20	13.15	0.001	0.718
Based on Average pain (CP57)					
SF-MPQ Sensory	10.69	16.06	18.28	0.174	<0.001*
SF-MPQ Affective	3.57	5.05	6.02	0.090	<0.001*
DASS-21 Depression	7.31	8.16	10.88	0.040	0.001*
DASS-21 Anxiety	5.13	6.00	8.46	0.055	<0.001*
DASS-21 Stress	8.42	9.28	11.71	0.040	0.001*
WHOQOL-BREF Overall	2.95	2.72	2.62	0.020	0.023*
WHOQOL-BREF Health	2.25	2.08	2.00	0.009	0.120
WHOQOL-BREF Physical	9.98	9.04	8.46	0.060	<0.001*
WHOQOL-BREF Psychological	11.87	11.35	10.08	0.016	0.039*
WHOQOL-BREF Social	12.44	12.47	11.14	0.011	0.090
WHOQOL-BREF Environmental	13.44	12.99	12.82	0.009	0.136
Based on Composite pain[a] (CP46)					
SF-MPQ Sensory	8.16	12.66	18.14	0.205	<0.001*
SF-MPQ Affective	2.42	4.14	5.92	0.114	<0.001*
DASS-21 Depression	6.85	7.55	9.77	0.033	0.003*
DASS-21 Anxiety	5.20	5.43	7.29	0.032	0.004*
DASS-21 Stress	8.85	8.70	10.48	0.018	0.030*
WHOQOL-BREF Overall	3.20	2.86	2.62	0.030	0.006*
WHOQOL-BREF Health	2.55	2.17	1.99	0.019	0.025*
WHOQOL-BREF Physical	10.63	9.53	8.73	0.050	<0.001*
WHOQOL-BREF Psychological	12.73	11.72	10.79	0.035	0.003*
WHOQOL-BREF Social	12.20	12.67	11.53	0.013	0.068
WHOQOL-BREF Environmental	14.57	13.09	12.92	0.014	0.056

[a] Rasch-based logit scores were transformed to the 0−10 scale for the composite pain.
* P-value < 0.05
CP, cutpoint; SF-MPQ, short form McGill Pain Questionnaire; DASS-21, short form of the Depression, Anxiety and Stress Scale; WHOQOL-BREF, World Health Organization Quality of Life-Brief version.

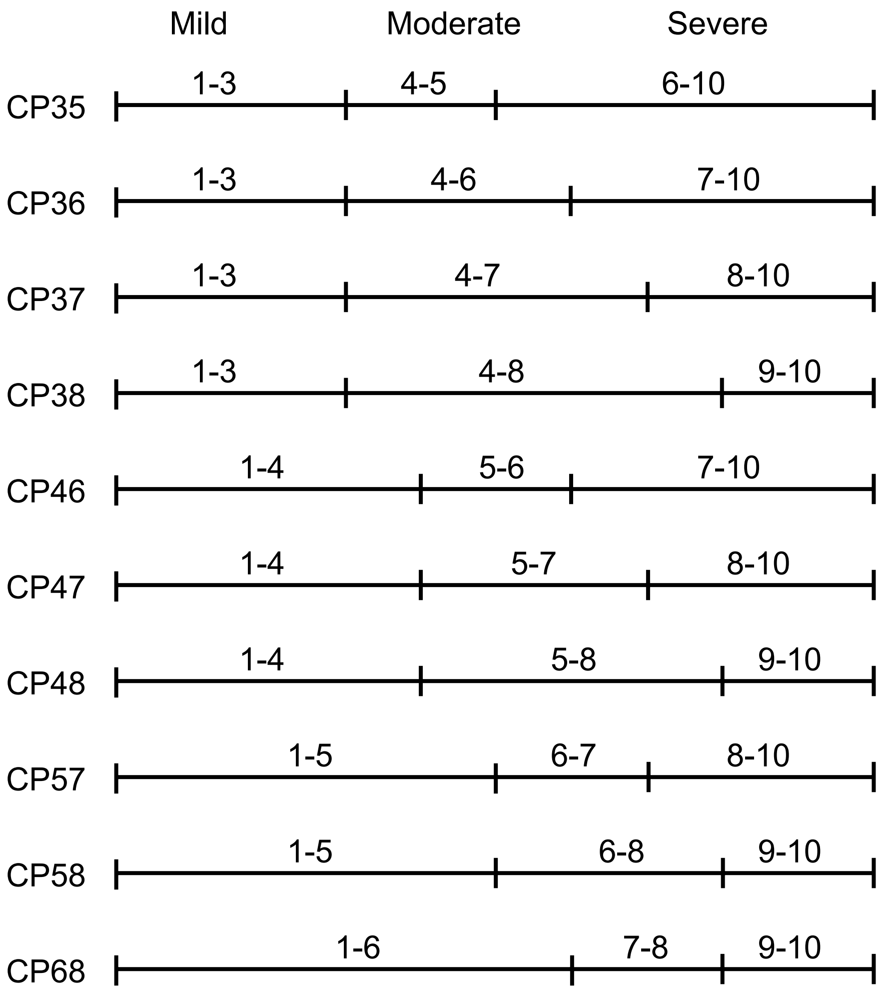


Figure S1. Mild, moderate, and severe pain cutpoint sets tested in this study. Number refers to pain severity ratings on a 0-10 numeric rating scale.

Appendix S1. Guide for Using the Composite Pain Keyform to Classify People with Chronic Pain

1st step: Circle the patient's responses to each BPI pain severity item in the keyform.

2nd step: Evaluate whether at least 3 items belong to the same pain classification zone and they are also close to each other. The classification zones are highlighted by yellow (for mild pain), green (for moderate pain), and pink (for severe pain).

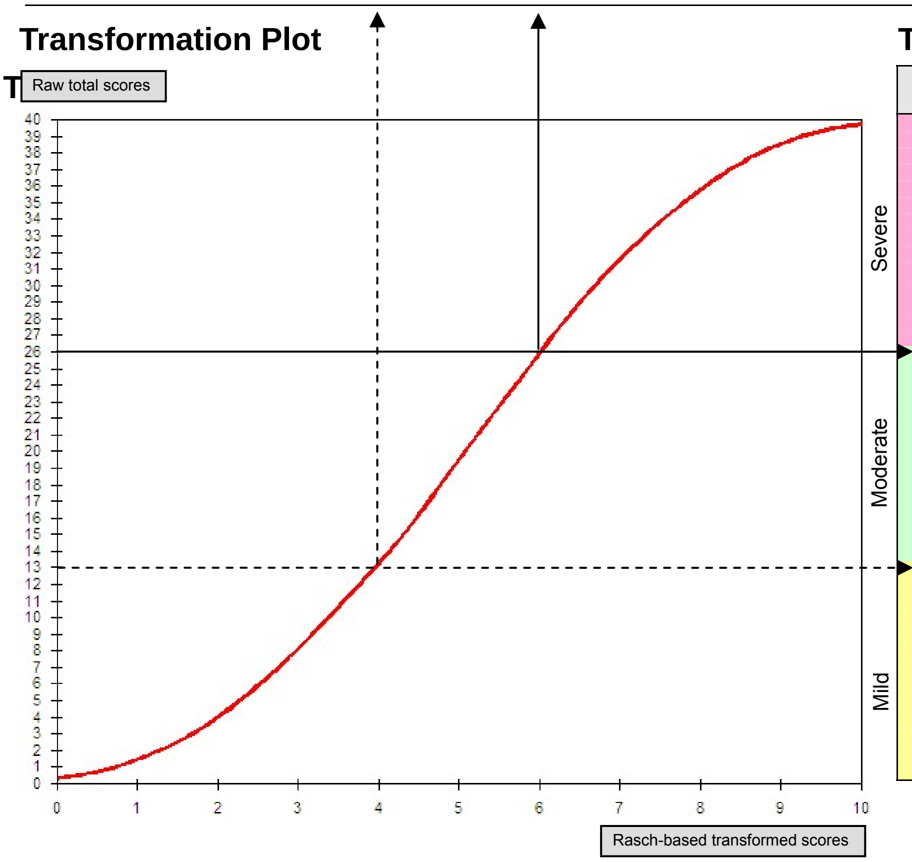
3rd step: If yes for the last step, calculate the total score based on the 4 BPI items and use the transformation table/plot to derive the patient's pain classification.

3rd step: If not* for the last step, you can classify the patient's pain based on average pain (≤5=mild, >5-7=moderate, and >7-10=severe).

Keyform

0-10 numeric rating scale	Item Description
0 : 1 : 2 : 3 :4 : 5: 6 : 7 : 8 : 9 : 10	Least pain
0 : 1 : 2 : 3 : 4: <u>5</u> : 6: <u>7</u> : 8 : 9 : 10	Average pain
0 : 1 : 2 : 3 : 4 :5 : 6 : 7 : 8 : 9 : 10	Current pain
0 : 1 : 2 : 3 : 4 :5 : 6 :7 : 8 : 9 : 10	Worst pain
0 1 2 3 <u>4</u> 5 <u>6</u> 7 8 9 10	Classification based on composite pain
Mild Moderate Severe	

Transformation Plot



Transformation

Raw total scores	Rasch-based logit scores	Transformed scores
40	-9.90	10
39	-8.47	9.3
38	-7.38	8.8
37	-6.59	8.4
36	-5.97	8.1
35	-5.43	7.8
34	-4.95	7.5
33	-4.50	7.3
32	-4.07	7.1
31	-3.66	6.9
30	-3.27	6.7
29	-2.89	6.5
28	-2.51	6.3
27	-2.15	6.2
26	-1.80	**6**
25	-1.46	5.9
24	-1.14	5.7
23	-0.84	5.5
22	-0.54	5.4
21	-0.24	5.2
20	0.06	5.1
19	0.35	4.9
18	0.66	4.8
17	0.96	4.6
16	1.27	4.5
15	1.58	4.3
14	1.89	4.1
13	2.21	**4**
12	2.54	3.8
11	2.87	3.6
10	3.23	3.4
9	3.61	3.2
8	4.01	3
7	4.43	2.8
6	4.89	2.5
5	5.38	2.3
4	5.92	2
3	6.53	1.7
2	7.26	1.3
1	8.29	0.7

* The following examples would be considered as aberrant item response patterns, including that the least pain is higher than the worst pain or the average pain, 2 items are located in one severity zone different from the zone for the other 2 items, or all the 4 pain items are located differently across 3 severity zones.