ScienceForecast
A digital science library

Research Article
Published: 13 Oct, 2020

SF Journal of Pain Medicine and Management

Equidistant Pain Scores with Normality

*Chakrabartty SN**

Indian Statistical Institute, Kolkata, West Bengal, India

Abstract

Context: Need is felt for meaningful computation of statistics of self-reported scales for measurement of pain and to ensure satisfaction of assumptions of techniques like linear regression, ANOVA, Factor Analysis(FA), Principal Component Analysis (PCA), t-test, Cronbach alpha, etc.

Aims: To describe methods of transforming ordinal pain scores to continuous equidistant scores to overcome limitations associated with multi-item scales for pain measurement and enabling parametric analysis without violation of assumptions.

Method: A non-parametric multi-staged method is described to transform ordinal raw scores of a Likert item ⇢ Continuous equidistant scores ⇢ Normalization of the equidistant scores ⇢ Proposed scores in the range [1,10]. Test score is taken as sum of item scores.

Statistical Analysis: Transformation at each stage described with the associated desired properties and empirical illustration to help clinicians to understand the main features of the proposed scores and to use them effectively.

Results: The proposed scores avoid major limitations of scoring existing pain scales, help in meaningful comparisons, quantifying effect of treatment plan and progress/deterioration of a patient or a group and facilitate application of statistical techniques in parametric set up.

Conclusions: Proposed scores reflecting intensity of pain help meaningful comparison in terms of pain intensity, change in pain intensity and drawing path of progress for better prognostication. Better methods for classification efficiency and reliability as per theoretical definition explained. Future studies suggested.

Keywords: Pain assessment; Equidistant scores; Normality; Weighted sum; Reliability

Introduction

In absence of objective biological markers for intensity of pain, large numbers of scales have been developed, based on Patient-Reported Outcomes (PROs) for pain assessment. Multidimensional aspects of pain include Sensory (Intensity, location, character of the pain sensation), Affective (Emotional and perceived components) and Impact related (Disability, dysfunctions, altered behaviour). Commonly used self-reported scales for pain measurement differ in formats, number of items, and also perceived factors of pain like physiological, psychological and emotional factors of pain or to assess impact of pain. Number of items used in multi-item scales ranges between 2 (SF-36 BPS) [1] to 78 (MPQ) [2]. Reliability of single pain ratings was inadequate unlike reliability of most of the composite scores [3]. Four important psychometric parameters of such scales were considered [4].

i) Responsiveness of a scale - Reflects the measure's sensitivity to change. It can be assessed in several ways [5].

ii) Minimal Clinically Important Difference (MCID) - The smallest score difference reported by patients that correlates with the patient stating that he/she is "slightly better" compared to his/her own state at an earlier point [6,7]. However, MICDs are context-specific and vary between samples [8]. There could be better way to see effect of treatment plan on an individual or a group.

iii) Validity - Extent to which a measurement scale agrees with clinical expectations about pain in the post-operative period like low pain before surgery, high following surgery, decreases with pain medication etc [9]. However, behavioral measure of pain was poorly correlated with two self-reported measures of pain intensity in 25 children in age group 3-7, following surgery [10] and raised question about the validity of current behavioral measures as indicators of pain intensity.

For the English version of Pain Catastrophizing Scale for Children (PCS-C), 2 factors with

OPEN ACCESS

Correspondence:
Satyendra Nath Chakrabartty, Indian Statistical Institute, Kolkata, West Bengal, India.
Tel: *+91-7838333652*
E-mail: *snc12@rediffmail.com*

Received Date: *01 Sep 2020*
Accepted Date: *01 Oct 2020*
Published Date: *13 Oct 2020*

Citation: *Chakrabartty SN. Equidistant Pain Scores with Normality. SF J Pain Med Manag. 2020; 1(1): 1001.*

eigen values exceeding one was found [11] against 3-factor solutions observed for children [12] and for adults [13].

iv) Test-retest reliability-Assumes that the sample has not undergone any changes in the time interval between two administrations. Test-retest reliability of Quebec Back Pain Disability Scale had time gap of 2 to 6 months [14]. Out of 746 articles, no article with excellent test-retest reliability was found [15]. Moreover, test-retest reliability may be high even if Cronbach alpha is low.

However, much attentions have not been given to the admissibility of operations like addition leading to computation of meaningful average and Standard Deviation (SD) of scores and also on verification of assumptions of normality, linearity, uni-dimensionality, etc. for application of techniques like linear regression, ANOVA, Factor Analysis (FA), Principal Component Analysis (PCA), t-test, Cronbach alpha, etc.

The paper attempts to address the issues of admissibility of operations and assumptions for parametric analysis and describes methods to overcome such limitations associated with multi-item scales for measurement of pain.

Methodological Limitations and Remedial Measures

Zero as anchor value

Format of items in many pain measuring scales attach numerical values to response alternatives as 0,1,2,3,…. and so on where higher value reflects higher pain intensity. Example: Numerical Rating Scale (NRS), Pain DETECT questionnaire (PD-Q) [16]; McGill Pain Questionnaire (MPQ) [17]; Neuropathic Pain Score (NPS) [18]; Pain Quality Assessment Scale (PQAS) [19], etc.

Use of zero as an anchor value unnecessary reduces scale mean and distorts variance, item-total correlations, etc. If each respondent chooses the alternative with zero value to an item indicating "no pain" then (i) mean=variance=0 for the "no pain" sub-group results in difficulties in computation of between group variance (ii) Correlation with that item is undefined (iii) Analysis involving expected values (value of the variable × probability of that value) is not meaningful. Transforming NRS scores and/or using ordinary least squares or dichotomizing and using logistic regression may be inappropriate due to presence of many zeroes [20]. Assigning values 1,2,3,….. and so on keeping the convention of higher numerical value ⇔ higher pain intensity, without changing structure of data is suggested.

Tied scores

A number of respondents may get tied score in a multi-item Likert scale, since summative Likert scores ignore the pattern of obtaining an individual score. Presence of tied scores implies that the scale fails to discriminate individuals with tied scores. Weighted sum with different weights to the response categories of different items can avoid tied scores and thus improves discriminating power of the scale.

Continuous and equidistant scores

If the alternatives of an item are marked as 1,2,3,…, meaningful addition demands that distance between alternative 1 and 2 $(d_{12})=d_{ij}\forall j=i+1$ and also $d_{13}=2d_{12}$ and so on. No pain measuring scale generating ordinal data satisfies this equidistant property.

Summative score used by such scales usually assign equal importance to the items which may not be justified because of different item-total correlation, different factor loadings, etc. However, S-LANSS score is a summative score assuming different importance to the items in terms of designed item score [21]. Neuropathic Pain Questionnaire (NPQ) is diagnostic and measurement tool also [22]. It assesses the intensity of 12 neuropathic symptoms and uses discriminate function coefficients to arrive at a total score. NPQ requires complex calculations to score and has not been validated against treatment changes. Power of discrimination between types of pain for NPQ was less in comparison to the same for LANSS pain scale [23].

Equidistant property can be achieved, if scores of an item are taken as weighted sum where $0<W_i<1$ are assigned to response categories marked as 1,2,3,…..5 (say) of different items so that $1W_1,2W_1,3W_1,4W_1$ and $5W_1$ form an arithmetic progression. A method to transfer ordinal scores (X) to continuous equidistant scores avoiding ties with a fixed zero point (Y) was given [24]. Such scores are also monotonic and thus indicate responsiveness of measurement.

Correlation and linearity

Correlation and linearity are two different concepts. High value of r_{XY} may not indicate that X is linearly related with Y and justify fitting regression equation of the form $= \alpha+\beta X+\epsilon$. For example, if X takes integer values between 1 to 30, $|r_{X,f(x)}| \geq 0.92$ for non-linear $f(X)=X^2,X^3$, $log_{10}{}^X$, *Cos X*, *Sin X*, etc. Thus, linearity implies high correlation but the converse is not true. To fit equation of the form $Y=\alpha+\beta X+\epsilon$, it is necessary to

i). Test goodness of fit by testing H_0: $S_E^2=0$, where $S_E^2 = 1/n\sum(Y_i-\hat{Y}_i)^2$ and

ii). Verify the assumption of normal distribution of error score $E=(Y-\hat{Y})$ where $\hat{Y}$ denotes the predicted values of Y from $Y=\alpha+\beta X+\epsilon$.

The above highlights the need of verification of associated assumptions of the techniques being used.

Normality

Statistical techniques like PCA, FA, SEM, t-test, ANOVA etc. assumes normal distribution of the variables. Continuous equidistant scores (Y) described above can be normalized by $z_i = \frac{Y_i-\bar{Y}}{SD(Y)}$ where $-\infty < Z_i < \infty$.

To avoid negative scores of Z_i, following linear transformation helps to get proposed scores (P) in a desired score range say 1 to 10:

$$P=(10-1)\left[\frac{(Z_{if})-Min(Z_{if})}{Max(Z_{if})-Min(Z_{if})}\right]+1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (1.1)$$

Converting item scores to similar distribution say, normal distribution avoids the problem of interpretation of added/subtracted variables.

The above procedure of converting raw scores of pain measuring scale (X) to equidistant, continuous, monotonic scores (Y), followed by normalization (Z) and converting to a desired score range (P) may also be used to:

Transform scores of items having different number of response categories and score-ranges like MPQ, Functional Assessment of Cancer Therapy Scale (FACT-G) with 33 items or Fact-G Version 4 with 27-items [24], etc. to P-scores.

Compare a number of pain measuring scales by converting scores of each scale to P-scores.

To find equivalent scores (X_0,Y_0) of two scales satisfying:

$$\int_{-\infty}^{X_0} f(X)dx = \int_{-\infty}^{Y_0} g(Y)dy$$

Where, $f(X)$ and $g(Y)$ are the density function of $N(0,1)$ and the (X_0, Y_0) combinations can be translated by (1.1) to Normal distribution with parameters which can be estimated from the data.

Properties

1. The proposed method is independent of distribution of underlying/observed variable and avoids major limitations of existing pain scales.

2. Generates continuous, monotonic scores satisfying equidistant property, normality and a desired positive score range with a fixed zero point. Test scores as a sum of *P*-scores of the items also follows normal distribution.

3. Facilitates computation of sample mean and SD and provides a platform to undertake parametric analysis.

4. Higher *P*-score indicates higher pain severity.

5. Correlation between Y and $Z(r_{YZ})$ will be nearly perfect like r_{ZP} due to linear relationships between Y & Z and also between P & Z. But, r_{XY} will fluctuate depending on weights as function of frequencies of different response categories of different items emerging from the data. r_{XP} is likely to have moderate value.

6. Helps in assessing extent of progress or deterioration of a patient over time points. The absolute value of Changed Score (CS)= $\frac{P_{it}-P_{i(t-1)}}{P_{i(t-1)}}\times 100$ gives percentage of deterioration or progress of the i^{th} patient in t^{th} time over (t-1)th time period, respectively for $P_{it}-P_{i(t-1)}>0$ or $P_{it}-P_{i(t-1)}<0$ where, P_{it} denotes transformed pain score of the i^{th} individual at time point t. Such CS can be examined and interpreted with the MCID.

7. Decreasing trend of plotting of *P*-scores of a patient over time periods implies steady progress of the patient. An increasing trend indicates steady deterioration of the patient over time, requiring attention and possible modification of treatment plan. Such *P*-score graphs can also be used to compare pattern of progress i.e., response to treatment plans between two patients or groups of patients with similar pain profile.

8. Clinicians can take advantage of the proposed method and rank patients uniquely avoiding ties; classify patients with respect to *P*-scores and also compare groups of patients either for longitudinal data or snap-shot data.

Classification

Each pain measuring scale suggest a set of cut-off scores to classify the individuals under several categories like "No pain", "Mild pain", "Moderate pain", "Severe pain" and "Worst pain"-Like Numerical Rating Scale (NRS), etc.

However, classification of individuals in mutually exclusives classes demands that individuals in a class will be similar among themselves (low value of within group variance) and dissimilar with individuals belonging to other classes (high value of between group variance). In addition, classifications need to be associated with clear clinical concepts of class labels.

For example, S-LANSS pain scale considers cut off score of 12 out of maximum possible score of 24 i.e., persons with score ≤12 are taken as those having no pain. In other words, sub-group of persons with score ≤12 are similar and will have low variance for the sub-group. However, a score of ≤12 can be achieved in different patterns of responses to the 7-items. Illustrative ways are shown in Table 1.

Table 1: S-LANSS score of 12 or less by different illustrative ways.

Patient	Item-1	Item-2	Item-3	Item-4	Item-5	Item-6	Item-7	Total
1	0	5	0	2	0	5	0	12
2	0	5	3	0	1	0	3	12
3	0	0	3	0	1	5	3	12
4	5	0	3	0	1	0	3	12
5	5	5	0	2	0	0	0	12
6	5	0	3	2	1	0	0	11
7	0	0	3	2	1	5	0	11
8	5	0	3	2	0	0	0	10
9	0	0	0	3	0	5	3	10
10	5	5	0	0	0	0	0	10
11	5	0	3	0	1	0	0	9
12	0	0	0	0	0	0	0	0
13	0	0	0	0	1	0	0	1
14	0	0	0	2	0	0	0	2
15	0	0	3	0	0	0	0	3
16	0	0	0	0	0	0	3	3
17	0	0	0	0	1	0	3	4
18	0	0	3	0	1	0	0	4
19	5	0	0	0	0	0	0	5
20	0	5	0	0	1	0	0	6
21	0	0	0	0	1	5	0	6
22	0	0	0	2	0	5	0	7
Mean								7.36
Variance								16.72

Value of variance tends to indicate dissimilarities in the group with score ≤12. Moreover; the illustrative patients may not be similar from clinical point of view. Clearly, the 12th patient and the 2nd patient may differ clinically.

Simple way of classifying a group of individuals to four classes is to find the quartiles Q_1, Q_2, Q_3, Q_4 and put the individuals with *P*-scores $\leq Q_1$ in class 1, Class 2nd if *P*-scores $\leq \Sigma$ and so on. However, this may not quantify efficiency of classification.

Better measure of classification efficiency involving K-number of mutually exclusive classes is Davies-Bouldin Index (DBI) [25] which is defined as:

$$DB_K = \frac{1}{K}\sum_{i=1}^{K} \underset{j=1.2.....K, i\neq j}{Max}\left[\frac{diam(C_i)+diam(C_j)}{\|C_i - C_j\|}\right] \quad \text{..... (1.2)}$$

Where, diameter of a cluster/class is defined as:

$$diam(C_i) = \sqrt{\frac{(\Sigma_{S\in C_i}\|x-C_i\|^2)}{n_i}} \quad \text{..... (1.3)}$$

n_i: Number of members in the i-th class, *Ci*: Centroid (or mean) of i-th cluster and *K*: Number of classes.

Upper limit of DBI is 1 and lower value implies better efficiency. DBI was found best among other cluster validity indices [26].

Discriminating value

Among the seven dissimilarity measures of discriminating value of Likert scale, Coefficient of Variation (CV) had maximum

Table 2: Weights to response categories and equidistant weighted scores.

Item	W_1	W_2	W_3	W_4	W_5	W_6	W_7	Common difference
5-point scale								
1	0.024944	0.186038	0.239737	0.266586	0.282695	-	-	0.347133
2	0.029529	0.186404	0.238696	0.264842	0.280529	-	-	0.343279
3	0.038382	0.18711	0.236686	0.261474	0.276347	-	-	0.335839
4	0.018437	0.185519	0.241214	0.269061	0.285769	-	-	0.352602
5	0.045498	0.187678	0.235071	0.258768	0.272986	-	-	0.329858
7-point scale								
1	0.027386	0.144537	0.183587	0.203112	0.214828	0.222638	0.003912	0.261688
2	0.021543	0.117886	0.150001	0.166058	0.175693	0.182116	0.186703	0.21423
3	0.015122	0.116565	0.150379	0.167286	0.177431	0.184193	0.189024	0.218008
4	0.014556	0.116448	0.150412	0.167394	0.177584	0.184377	0.189229	0.218341
5	0.006202	0.114729	0.150904	0.168992	0.179845	0.18708	0.192248	0.223256

Table 3: Mean variance of item scores and test scores.

	Raw Score (X)		Equidistant score (Y)		Normalized score (Z)		Proposed score (P)	
	Mean	**Var.**	**Mean**	**Var.**	**Mean**	**Var.**	**Mean**	**Var.**
5-point scale								
Item 1	3.85	1.15909	1.01427	0.13967	-0.00305	0.99605	7.40512	5.84474
Item 2	3.85	1.38131	0.99072	0.17905	-0.00347	0.99391	7.29001	7.64962
Item 3	3.81	1.36758	0.98209	0.15425	-0.00656	0.99055	7.30523	6.85791
Item 4	3.74	1.08323	0.98457	0.13468	-0.0016	0.99786	7.16127	5.47211
Item 5	3.73	1.33040	0.94601	0.14476	-0.00895	0.98664	7.11927	6.64519
Test	18.98	8.66626	4.91766	1.03901	-0.02363	6.90766	36.2809	44.76774
7-point scale								
Item 1	5.15	2.67424	0.73422	0.26651	-0.0016	0.99525	5.85617	12.54905
Item 2	5.21	2.61202	0.92297	0.11994	-0.0028	0.99638	7.30481	5.85904
Item 3	5.24	2.12364	0.93745	0.10066	-0.00153	0.99822	5.57443	4.75708
Item 4	5.17	2.12232	0.91688	0.10843	-0.00129	0.99816	7.19601	5.10928
Item 5	5.00	1.83838	0.89922	0.09163	-0.00025	0.99970	6.99949	4.13514
Test	25.77	17.9769	4.41074	0.92130	-0.00747	6.90482	32.9309	43.47208

theoretical advantages [27] where item discriminating value $(Disc_i)$ and test discriminating value $(Disc_{Test})$ are defined as $Disc_i=SD_i/mean_i$ and $Disc_{Test}=SD_{Test}/Mean_{Test}$. Note that for a scale with m-items, variance of the i-th item $S_{X_i}^2 = \bar{X}_i^2.Disc_i^2 \forall\, i=1,2,\ldots..m$ and sum of item variances $\sum_{i=1}^{m} S_{X_i}^2 = \sum_{i=1}^{m} \bar{X}_i^2.Disc_i^2$ and Test variance $S_X^2 = \bar{X}^2.Disc_T^2$. Thus, Cronbach α in terms of $Disc_i$'s and $Disc_{Test}$ is

$$\alpha=\left(\frac{m}{m-1}\right)\left(1-\frac{\sum_{i=1}^{m}\bar{X}_i^2.Disc_i^2}{\bar{X}^2.Disc_T^2}\right) \quad \ldots\ldots\ldots\ldots\ldots (1.4)$$

Non-linear relationship between $Disc_{Test}$ and test reliability, as per theoretical definition can be derived as $(Disc_{Test})^2=\frac{CV_{Truescores}^2}{r_{tt}}$ where, $r_{tt}=\frac{S_T^2}{S_X^2}$ ..... (1.5)

Thus, test reliability and $Disc_{Test}$ and are not independent but related by a negative relationship.

Reliability

Pain measurement scales often report test-retest reliability. However, question can be raised whether test-retest reflects reliability or agreement or both. Test reliability may reflect ability of a scale to produce the same rankings on both occasions; but agreement may require the scale to come out with identical values on both occasions [28]. Thus, interpretation of difference between two successive scores could be due to change of the respondents in the time gap or due to the characteristics of the scale.

Reliability in terms of Cronbach's alpha assumes among others continuous measurement, uncorrelated errors, normality, uni-dimensionality. Violation of assumption of continuous nature of data and normality may distort variance-covariance matrix and biased value of coefficient α [29-31]. Number of eigen-values >1 is 2 or more for a test, implies departure from uni-dimensionality and Cronbach α should not be used. Instead, test reliability as per the theoretical definition can be obtained following method suggested by [32], where the test is dichotomized in two parallel g^{th} and h^{th} sub-tests and reliability is given by

$$r_{tt}=1-\frac{\|X_g\|^2+\|X_h\|^2-2\|X_g\|\|X_h\|\cos\theta_{gh}}{NS_X^2} \quad \ldots\ldots\ldots\ldots\ldots (1.6)$$

where, θ_{gh} is the angle between the vectors X_g and X_h; $\|X_g\|$ is the length of Xg. $\|X_h\|$ is defined accordingly.

Empirical illustration

Empirical illustration of the proposed method is given using a 5-point and a 7-point scale, each with five items administered on the sample of 100 patients.

Table 4: Discriminating value of scales.

Discriminating value (CV)		
	Raw score(*X*)	Proposed score (*P*)
5-point scale	0.155103	0.184419
7-point scale	0.164529	0.200217

Equidistant scores

Item score of 5-point scale is taken as $\sum_{k=1}^{5} kW_k > 0$, $\sum_{k=1}^{5} W_k$ and common difference i.e. $(K+1)W_{(K+1)} - k.W_k$=constant for k=1,2,3,4,5. Similar procedure is adopted for 7-point scale, where k=7. Weights based on frequencies of different response categories of an item and common difference are shown in Table 2 for the 5-point and the 7-point scale.

Descriptive statistics

Mean variance of the raw score, equidistant score, normalized score and converted *Z*-score to score range 1-10 for the 5-point and 7-point scales are shown in Table 3.

Observations

Equidistant score (*Y*) made the data more homogeneous.

Proposed score (*P*) was normally distributed. However, parameters of the distribution of *P* scores depend on co-variances between pair of items were slightly different for 5-point and 7-point scale.

Tied score

20 individuals were tied with raw score of 19 in the 5-point scale. *Y*-scores and *P*-scores with five decimal places resulted in no tied scores. For the scale, SD of *P*-score corresponding to raw scale score of 19 was 0.3728. Avoidance of tied scores improved discriminating power of the scale in terms of CV.

Discriminating value of scales in terms of CV are shown in Table 4.

Correlations

Each of r_{YZ} and r_{ZP} is likely to be closed to 1.0 since they are related by linear functions. However, r_{XY} may fluctuate depending on the different weights assigned to different response categories of different items. Correlation matrix of test scores at various stages is given in Table 5.

Inter-item correlations and item-total correlations for raw scores and *P*-scores are shown in Table 6.

Observations

Item reliability in terms of item-total correlations of raw scores improved for 3 items and 2 items respectively for 5-point and 7-point scales when *P*-scores were used.

Negative correlation of raw scores between item 1 and 2 for 5-point scale changed to positive for *P*-scores. However, positive correlations of raw scores between item 1 and 2 of 7-point scale became negative for *P*-scores. Same is true for correlation between item 2 and item 4.

Generalization of improved inter-item correlation for *P*-scores cannot be made.

Table 5: Correlation matrix of test scores at stages.

	Raw Score (*X*)	Equidistant score (*Y*)	Normalized score (*Z*)	Proposed score (*P*)
Raw Score(X)	1	0.995656 (0.777428)	0.995693 (0.853015)	0.995904 (0.777703)
Equidistant score (Y)	-	1	0.99913 (0.98348)	0.999895 (0.99996)
Normalized score (Z)	-	-	1	0.998976 (0.983248)

Note: Figures without brackets relate to 5-point scale and figures within brackets are related to 7-point scale.

Table 6: Inter-item correlation matrix.

	Item 1	Item 2	Item 3	Item 4	Item 5	Test
5-point scale						
Item 1	1	-0.00998 (0.007738)	0.113524 (0.113524)	0.027945 (0.027945)	0.089069 (0.089069)	0.451606 (0.453044)
Item 2	-	1	0.059896 (0.079917)	0.190752 (0.189876)	0.01453 (0.002971)	0.492513 (0.514972)
Item 3	-	-	1	0.058591 (0.058591)	0.021492 (0.021492)	0.491812 (0.494212)
Item 4	-	-	-	1	0.386884 (0.386884)	0.614781 (0.610194)
Item 5	-	-	-	-	1	0.575504 (0.56236)
7-point scale						
Item 1	1	0.205808 (0.11399)	0.167002 (-0.0387)	0.082467 (0.094073)	0.22778 (0.078909)	0.62272 (0.622914)
Item 2	-	1	0.287178 (0.287576)	0.019005 (-0.01107)	0.129067 (0.130068)	0.607068 (0.559814)
Item 3	-	-	1	0.166147 (0.13965)	0.127805 (0.12321)	0.615539 (0.50146)
Item 4	-	-	-	1	0.025569 (0.04314)	0.44793 (0.448807)
Item 5	-	-	-	-	1	0.509551 (0.454112)

Note: Figures within brackets represent correlations for transformed P-scores.

Table 7: Eigen-values and percentage of variance explained.

Scale	Raw scores (*X*)		Normalized scores converted to (1, 10) (*P*)	
	No. of independent factors	Cumulative variance explained	No. of independent factors	Cumulative variance explained
5-point	3	71.44%	3	71.54%
7-point	2	52.70%	2	49.55%

Factor structure

High correlations between X and P (0.995904 for 5-point scale and 0.777703 for 7-point scale) did not change much the facture structure of raw scores and proposed scores, as can be seen from Table 7.

Observations

Factor structure of a scale remained unchanged for X and P.

Convergent validity of 5-point scale and 7-point scale were different.

3 factors for 5-point scale and 2 factors in 7-point scale may not imply same set of constructs.

Conclusions

A multi-staged method is described to transform ordinal discrete raw scores of a Likert item ⇢ Continuous equidistant scores ⇢ Normalized equidistant scores ⇢ Proposed scores in the range 1-10. Test score as sum of item scores follows normal distribution. The method is independent of distribution of underlying or observed variables.

The proposed scores (P) are continuous, monotonic (showing responsiveness of measurement) and satisfy equidistant property, normality with a desired positive score range having a fixed zero point and avoid major limitations of scoring existing pain scales. It helps in meaningful comparisons, quantifying extent of progress/deterioration of a patient or a group of patients over time i.e., Changed Score (CS) which can be examined and interpreted with the minimal clinically important change; For longitudinal data, graph of P-score over time periods can be used to find pattern of progress i.e., response to treatment plans for an individual patient or a groups of patients with similar pain profile. Provides a platform for undertaking analysis under parametric set up.

The method can be well used to (i) compare a number of pain measuring scales by converting scores of each scale with normality and fixed score range, (ii) transform scores of items having different number of response categories and score-ranges like MPQ, FACT-G, etc. to meaningful test scores with desired properties.

Significant values of r_{XP} did not change factor structure of a test. The same was confirmed by PCA. Better measures suggested for discriminating value, reliability of a scale and classification efficiency.

The proposed approach of converting X-score to P-scores is critically relevant to practitioners and researchers and is recommended for clear theoretical advantages and easiness in calculations.

Future studies may be undertaken with longitudinal data set for generalization of findings emerging from this study.

References

1. Ware JEJr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol. 1998; 51: 903-912.
2. McCaffery M, Beebe A. The Numeric Pain Rating Scale Instructions. J Holist Nurs. 1989; 81: 888-895.
3. Jensen MP, Castarlenas E, Tome-Pires C, Vega RDL, Sanchez-Rodriguez E, Miro J. The Number of Ratings Needed for Valid Pain Assessment in Clinical Trials: Replication and Extension. Pain Med. 2015; 16: 1764-1772.
4. Goldsmith ES, Taylor BC, Greer N, Murdoch M, MacDonald R, McKenzie L, et al. Focused Evidence Review: Psychometric Properties of Patient-Reported Outcome Measures for Chronic Musculoskeletal Pain. J Gen Inter Med. 2018.
5. Kroenke K, Theobald D, Wu J, Tu W, Krebs EE. Comparative Responsiveness of Pain Measures in Cancer Patients. J Pain. 2012; 13: 764–772.
6. Deyo RA, Ramsey K, Buckley DI, Michaels L, Kobus A, Eckstrom E, et al. Performance of a Patient Reported Outcomes Measurement Information System (PROMIS) Short Form in Older Adults with Chronic Musculoskeletal Pain. Pain Med. 2015.
7. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to Explain the Clinical Significance of Health Status Measures. Mayo Clin Proc. 2002; 77: 371-383.
8. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and Minimal Important Differences for Patient Reported Outcomes. Health Qual Life Outcomes. 2006; 4: 1-5.
9. Tyler DC, Tu A, Douthit J, Chapman CR. Towards Validation of Pain Measurement Tools for Children: A Pilot Study. Pain. 1993.
10. Beyer JE, McGrath PJ, Berde CB. Discordance Between Self -Report and Behavioral Pain Measures in Children Aged 3-7 Years After Surgery. J Pain Symptom Manage. 1990; 5: 350-356.
11. Parkerson HA, Noel M, Page MG, Fuss S, Katz J, Asmundson GJG. Factorial Validity of the English-language Version of the Pain Catastrophizing Scale-child Version. J Pain. 2013; 14: 1383-1389.
12. Crombez G, Bijttebier P, Eccleston C, Mascagni T, Mertens G, Goubert L. The Child Version of the Pain Catastrophizing Scale (PSC-C): A Preliminary Validation. Pain. 2003; 104: 639-646.
13. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and Validation. Psychol Assess. 1995; 7: 524-532.
14. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement Properties. Spine. 1995; 20: 341-352.
15. Paiva CE, Barroso EM, Carneseca EC, Souza CDP, Santos FTD, Lopez RVM, et al. A Critical Analysis of Test-Retest Reliability in Instrument Validation Studies of Cancer Patients Under Palliative Care: A Systematic Review. BMC Med Res Methodol. 2014; 14: 8.
16. Freynhagen R, Baron R, Gockel U, Tolle TR. Pain DETECT: A New Screening Questionnaire to Identify Neuropathic Components in Patients with Back Pain. Curr Med Res Opin. 2006; 22: 1911-1920.
17. Melzack R. The McGill Pain Questionnaire: Major Properties and Scoring Methods. Pain. 1975; 1: 277-299.
18. Galer BS, Jensen MP. Development and Preliminary Validation of a Pain Measure Specific to Neuropathic Pain: The Neuropathic Pain Scale. Neurol. 1997; 48: 332-338.
19. Jensen MP, Dworkin RH, Gammaitoni AR, Olaleye DO, Oleka N, Galer BS, et al. Assessment of Pain Quality in Chronic Neuropathic and Nociceptive Pain Clinical Trials with the Neuropathic Pain Scale. J Pain. 2005; 6: 98-106.
20. Goulet J, Buta E, Carroll C, Brandt C. Statistical Methods for the Analysis of NRS Pain Data. J Pain. 2015.
21. Bennett M. The LANSS Pain Scale: The Leeds Assessment of Neuropathic Symptoms and Signs. Pain. 2001; 92: 147-157.
22. Krause SJ, Backonja MM. Development of a Neuropathic Pain Questionnaire. Clin J Pain. 2003; 19: 306-314.
23. Chakrabartty SN. Limitations of Insomnia Severity Index and possible remedies, JSM Neurological Disorders and Stroke. 2019; 5: 1-9.
24. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: Properties, Applications and Interpretation. Health Qual Life Outcomes. 2003; 1: 79.

25. Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979; 1: 224-227.

26. Kim M, Ramakrishna R. New Indices for Cluster Validity Assessment. Pattern Recognition Letters. 2005; 26: 2353-2363.

27. Chakrabartty SN. Discriminating Value of Item and Test. International Journal of Applied Mathematics and Statistics. 2020; 59: 61-78.

28. Berchtold A. Test-retest: Agreement or Reliability?. Methodological Innovations. 2016; 9: 1-7.

29. Flora DB, Curran PJ. An Empirical Evaluation of Alternative Methods of Estimation Confirmatory Factor Analysis with Ordinal Data. Psychol Methods. 2004; 9: 466-491.

30. Green SB, Yang Y. Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. Psychometrika. 2009; 74: 155-167.

31. Sheng Y, Sheng Z. Is Coefficient Alpha Robust to Non-normal Data?. Front Psychol. 2012.

32. Chakrabartty SN. Reliability of Test Battery. Methodological Innovations. 2020; 13: 1-8.