World Journal of Business Research
ISSN 2770-9078
Vol. 3, No. 3, 2023
Published by STSL Press

Original Paper

Methodological Issues of Quality of Life Assessment

Satyendra Nath Chakrabartty[1]

[1] Indian Statistical Institute, Indian Maritime University, Indian Ports Association, India

Abstract

Quality of Life Scales (QOLS) differ in dimensions, number and format of items, score ranges. Different QOLS in same sample gave different conclusions. Thus, QOLS's are not comparable.

The paper aims at reviewing methodological issues of QOLS and suggesting transformations as remedial actions to limitations of scoring of QOLS and address psychometric properties of transformed scores. The proposed approach converts ordinal item score to equidistant score (E) using different weights to response-categories of different items followed by standardizing to Z-scores and converting Z-scores to Y-scores in [1, 100]. Scale score is taken as sum of item-wise Y-scores following Normal distribution. Methods described to obtain reliability as per theoretical definition, validity using eigenvalues, discriminating values of items and scale and equivalent scores of QOLS. The paper involves no collection of data using a specified QOLS. The proposed scores offer meaningful arithmetic aggregations, meaningful comparison, ranking and classification of individuals, assessment of progress/deterioration and conducting statistical tests either for longitudinal or snap-shot data. Such scores help to find reliability as per definition, validity avoiding criterion scale, discriminating value. Normality facilitates estimation of population parameters and finding equivalent score combinations to integrate two QOLS.

Proposed score facilitating analysis in parametric set up satisfying desired properties and computing psychometric concepts has theoretical advantages and wide applicability including meaningfulness of operations, better comparison. Use of such method of transforming QOLS scores contributes to the existing methods.

Keywords: Quality of Life, Likert item, Normal distribution, Reliability, Validity, Discriminating value

Introduction

Large numbers of Quality of Life scales (QOLS) are there to assess quality of life (QOL) covering relevant domains like physical, mental, social wellbeing, jobs, communications, etc. Health is an important domain of QOL. Concept of health-related quality of life (HRQOL) and its determinants have evolved emphasizing on health-related measures to evaluate and to decide action for treating and managing undesirable symptoms, functional status, disease progressions, treatment effects, etc. to improve wellbeing or survival (Berzon et al., 1995).

QOLS differ on dimensions covered; item formats (number of items and response-categories per item). Flanagan QOLS with 15-items cover areas like Physical and Material Well-being, Relations with other people, Social, Community and Civic Activities, Personal Development, Fulfillment and Recreation using 5-point format (Flanagan, 1982). Terrible-delighted scale with 16 items, each in 7-point scale measures satisfaction where 1 indicates "Terrible" and 7 indicates "Delighted" (Andrews & Crandall, 1976). None of the two scales is symmetric. Empirically, the terrible-delighted scale was less negatively skewed than the 5-point QOLS.

Due to significant variations in dimensions, formats, method of scoring for different QOLS, Hunt et al. (1985) observed that it is difficult to know what is being measured by them in absence of agreed criteria for what constitutes quality-of-life and such instruments lack validity. Lack of methodological unanimity as far as measuring OOL is concerned was also observed (Boixadós et al., 2009). Use of different QOLS to the same sample may result in different conclusions and thus, such instruments are

not comparable.

Meaningful inter-country/inter-regional/inter-sample comparisons, classification, assessment of progress, statistical testing of equality of QOL means, etc. can be done if methodologically sound QOL scoring system is followed by transformations to facilitate all above mentioned uses along with assessment of responsiveness, reliability, validity, etc.

QOL scales will continue to be evolved with newer dimensions and items to suit the purpose. For example, new scales may include items relating to easiness/difficulties in use of smart-phone for credit/debit cards, inter-banking, digital communication, etc. Domains and items may be chosen depending on the purpose so that the chosen set represents a fair summary of the whole.

Ignoring the issues of selection of dimensions and items, the paper gives a method of transforming scores of Likert items to continuous scores following normal distribution for better and meaningful uses of QOLS.

QOL questionnaire

Cieza and Stucki (2005) linked 148 items of six HRQOL instruments (SF-36, the NHP, the QL-I, the WHOQOL-BREF, the WHODASII and the EQ-5D) involving 226 concepts to the International Classification of Functioning Disability and Health (ICF) (WHO, 2001). But 12 concepts could not be linked to ICF. Illustrative QOL scales and their features are as follows:

- *SF-36* with 36 items (combination of "Yes-No" type, 3-point, 5-point and 6-point items) covers 8 dimensions viz. physical health, mental health, social functioning, role functioning, general health, pain, and vitality (Ware & Sherbourne, 1992). The rescaled item score ranges between 0 to 100, where 100 indicates the highest level of functioning. Instead of a single measure like $SF\ 36_{Total}$, summary score of Physical Components and Mental Components are obtained. SF-36 has been used for general population and also with patients suffering from breast cancer, lung cancer, etc. It does not consider an important health variable called "sleep".

- *Nottingham Health Profile(NHP)* with "Yes – No" type 38 items covers six domains viz. physical abilities, pain, sleep, social isolation, emotional reactions, and energy level (Hunt et al., 1985). The optional Part II contains seven items reflecting how health problems affect occupation, jobs around the house, personal relationships, social life, sex life, hobbies, and holidays. NHP items are weighted to range between 0 to 100 and NHP score is average of the domain scores. It is difficult to compare the domains or to evaluate changes in pre- and post-intervention studies. Improvements for those with zero score in pre-administration cannot be evaluated, as zero scores may not indicate total absence of distress.

- *Spitzer Quality of Life Index—Patient Version* (*QL-I*) considers domains like health, daily living, activity, support and outlook using five number of 3-point items, one for each domain and visual-analogue scale (VAS). QOL index as summative score has been used for cancer and chronically-ill patients before and after therapy (Spitzer et al., 1981).

- *WHOQOL-BREF* has 26 items (one for general quality of life, one for health-related quality of life, and 24 items for the remaining four domains) (WHOQOL Group, 1994). Domain score, obtained as mean scores of items are transferred to have comparability with WHOQOL-100.

- World Health Organization Disability Assessment Schedule (*WHODASII*) with 36 items aims at assessing limitations of activities and inabilities in participation in six domains namely self-care, understanding and communicating, getting around, getting along with people, life activities, and participation in society (WHO, 2000).

- European Quality of Life instrument (*EQ-5D-5L*) has 5 dimensions viz. Mobility, Self-care, Usual activities, Pain & discomfort, Anxiety and depression (Euroqol Group, 1990). Levels of each of 5- item are marked as 1, 2, 3, 4 and 5 where "1" means no problem and "5" means extreme problems. Health-profile of a person is a 5-digit number, minimum being 1-1-1-1-1(no problem in any dimension) and maximum 5-5-5-5-5 (max. problem in each dimension). Instead of score, a person is categorized in

one of the possible 3125 = 5^5 categories. Frequency count of each such category is admissible. EQ-5D-5L showed higher responsiveness than the EQ-5D-3L system (Jin, 2019).

- *Duke-UNC Functional Social Support Questionnaire* (*FSSQ*) measures perceived functional support for patients with 8-items, each in 5-point format, where higher scores imply higher perceived support (Broadhead et al., 1988). FSSQ has been used for patients with breast cancer (Green et al., 2001).

- *Katz Index of Independence in Activities of Daily Living* (*KI-ADL*) measures performance in 6-areas of physical functioning viz. bathing, dressing, toileting, transferring, continence, and feeding by one "Yes-No" type item per area. A score of 6 indicates full functioning; 4 indicates moderate impairment and score $\leq$ 2 indicates severe impairment of functioning. However, KI-ADL scores with ceiling effects are less sensitive with cancer populations (Katz et al., 1963).

- *Functional Assessment of Cancer Therapy—General, Version 4* (*FACT-G 27)*: A 27-items (5-point scales) measure four HRQL domains: Physical Well-Being, Social/Family Well-Being, Emotional Well-Being, and Functional Well-Being. Number of items in domains varies (Cella et al., 1993). Validation of FACT-G in chronic diseases allowed evolution of multiple diseases, treatment, condition, and non-cancer-specific subscales (over 40 different Functional Assessment of Chronic Illness Therapy (FACIT). The FACT-G and FACIT scales have been used in studies like head and neck cancer, oral cancer, breast cancer, Hodgkin's disease, Hypogonadism, etc.

- *European Organization for Research and Treatment of Cancer Quality of Life*

Questionnaire (*EORTC QLQ-C30*) (www.eortc.be) covers dimensions like physical, role, emotional, and social functioning, along with disease-specific symptoms, financial impact, and global QOL. This has been used to assess QOL in cancer patients receiving chemotherapy, patients surviving with breast cancer, Hodgkin lymphoma, colorectal, prostate, lung cancer, etc.

Barring EQ-5D, most of the other instruments consider summative Likert score to find QOL score, despite the fact that ordinal Likert scores are not equidistant and addition is not meaningful (Bastien et al., 2001). Negative weights from principal component analysis (PCA) if any, indicate negative contribution of physical-health scores in the combined mental-health scores i.e. for high score on combined mental-health, one needs to have worse physical-health and vice versa. For EQ-5D, responses to the items are scored by a utility-weighted algorithm (Williams, 1995), to find single-index measures of health, in the score range 0 to 100 (Brooks, 1996).

Limitations and Remedial actions

Shortcomings and remedial actions are illustrated with the example of the *QOL questionnaire version II* to measure QOL of cancer patients in Indian scenario (Vidhubala et al., 2005). The questionnaire with 41 items covering 11 dimensions has the following major features:

- 39 items are in 4-point scale where 1 stands for "not at all", 2 for "A little", 3 for "Moderate" and 4 for "very much". Two other items on are in 10-point scale ranging from 1 (Very Poor) to 10 (Excellent). Maximum and minimum possible scores were 176 and 41 respectively.

- Suggested five categorization of total scores are: Above 165: very high QOL; 147–165: high QOL; 118–146: average QOL; 99–117: low QOL and below 99: very low QOL. PCA resulted in 10 independent factors explaining 62.6% of variance with different factor loadings for different dimensions and items. 10 independent factors out of 11 dimensions considered in the scale, tends to indicate poor correlations between a pair of dimensions. Cronbach alpha was 0.90. The scale was used by Nayak et al. (2017) on a sample of 768 cancer patients suffering from Stage III or IV of various types of cancer in India and found Mean = 105.32 and standard deviation (SD) = 12.93; mean and SD for Stage III cancer patients exceeded the same for those in Stage IV; 82.3% of the sample had QOL scores in the categories "Very low" and "Low" as per categorization suggested by Chang (1994).

Major limitations

- Assumes item scores are equidistant i.e. distance between "Not at all" and "A little"(d_{12}) = distance between "Moderate" and "A little" (d_{23}) = distance between "Very much" and "Moderate"

(d_{34}). However, ordinal scores do not satisfy $d_{12} = d_{23} = d_{34}$ and thus, arithmetic aggregation/averaging are not meaningful (Reeves et al., 2020). Moreover, the subjects may not perceive items as equidistant.

- Assigning equal importance to the items and dimensions are contradictory to different factor loadings for dimensions and items observed from PCA.

- Distributions of item/dimension scores are not considered. If X and *Y* have different distributions, it is difficult to interpret and find joint distribution of $X \pm Y$.

- Distributions of 10-point item with higher mean and SD are different from the 4-point items. Thus, addition of scores of 10-point items and 4-point items are not meaningful. Scales with few response-categories tend to result in lower value of mean and SD which may distort correlations with other scales (Martin, 1978; Chang, 1994).

- Summative Likert score giving equal importance to the items usually gives rise to a number of tied scores and reduces discriminating power of the scale.

The above said problem areas can be resolved by adopting the following steps suggested by (Chakrabartty, 2021):

I. Consider each dimension as a sub-test

II. Convert discrete raw score of each *k*-point item to continuous scores through weighted sum so that $W_1, 2W_2, 3W_3, \ldots\ldots, kW_k$ forms an arithmetic progression and ensures satisfaction of equidistant property.

For a 5-point item, II is achieved by weighted sum where different positive weights are assigned to levels of different items satisfying $\sum_{j=1}^{5} W_{ij} = 1$ and $5W_5 - 4W_4 = 4W_4 - 3W_3 = 3W_3 - 2W_2 = 2W_2 - W_1$ = constant > 0 which ensures item scores are monotonic. Two procedures to obtain such cardinal, continuous, monotonic, equidistant scores were given by Chakrabartty (2019).

III. Standardize item-wise equidistant score (E) by $Z = \frac{E-\bar{E}}{SD(E)}$ following $N(0,1)$

IV. Transform Z-score of each item to $Y \in$ [1,100] by the following linear transformation

$$Y = (99)\left[\frac{Z_{ij} - Min_{Z_{ij}}}{Max_{Z_{ij}} - Min_{Z_{ij}}}\right] + 1 \quad (1)$$

V. For the *i*-th individual, dimension score $D_i = \sum Y_j$ where summation is taken over the items under the dimension and proposed scale scores(P_i) is sum of scores of all dimensions. Clearly, each of D_i and P_i follows normal and parameters can be estimated from the data.

Properties

1. E-scores satisfy equidistant property. Normally distributed *Y*-scores taking values 1 to 100 is independent of number of response categories. Similarly, dimension score (D_i) and proposed scale scores(P_i) are continuous, monotonic, normal and facilitate undertaking parametric analysis including estimation of population mean (μ), population variance (σ^2), confidence interval of μ, testing statistical hypothesis like $H_0: \mu_1 = \mu_2$ or $H_0: \sigma_1^2 = \sigma_2^2$ etc. either for longitudinal data or snap-shot data.

2. *Y*-score reduces drastically number of tied scores. Thus, most of the individuals can be given unique ranks.

3. Effect of small change in *i*-th dimension (ΔP_i) to scale score (P_{Scale}) can be quantified in terms of elasticity i.e. percentage change of due to small change in P_i. The dimensions can be ranked in terms of elasticity.

4. Percentage progress/deterioration of the *i*-th patient between two successive time-periods can be

assessed by $\frac{P_{it}-P_{i(t-1)}}{P_{i(t-1)}} \times 100$, which reflects responsiveness of the scale and quantifies effectiveness of a treatment plan. Deterioration may be probed to find extent of deterioration in dimension scores to identify the dimension(s) requiring attention and possible modification of treatment or care plan for the patient. Similarly, progress for a group of patients can be assessed considering $\overline{P_t} > \overline{P_{(t-1)}}$ or rejection of $H_0: \mu_t = \mu_{(t-1)}$.

5. Plotting of progress/deterioration of a patient or a sample across time can be used to compare progress pattern i.e. response to treatments from the start i.e. t_0.

5. For two QOL scales X with normal pdf $f(x)$ and Y with normal pdf $g(y)$, one can find regression equation of the form $Y = \alpha_1 + \beta_1 X$ to predict Y from X or $X = \alpha_2 + \beta_2 Y$ to predict X with knowledge of Y. However, the two regression lines differ implying different values of equivalent score combinations (x_0, y_0). Better is to find values of (x_0, y_0) of the two QOL scales by

$$\int_{-\infty}^{x_0} f(x)dx = \int_{-\infty}^{y_0} g(y)dy \quad (2)$$

i.e. area of the curve $f(x)$ up to x_0= area of the curve $g(y)$ up to y_0.

The method of finding equivalent score-combinations is possible even if the scales have different number of items or dimensions.

Reliability

Cronbach α and test-retest reliability are commonly used to find reliability of QOL. Flanagan's QLOS had α from 0.82 to 0.92 and test-retest reliability(r) between 0.78 to 0.84 with time-gap of 3-weeks in stable chronic illness groups [26]. Similar results were reported for the terrible-delighted scale (Anderson, 1995; Neumann & Buskila, 1997)

Test-retest reliability is not preferred when patients undergo changes in the time-gap. Alpha works best for one-dimensional test. Computation of alpha despite obtaining 10 independent factors is not justified.

Remedial action

An assumption-free method of obtaining reliability as per the theoretical definition along with computation of error variance (S_E^2) and true score variance $({S_T}^2)$ from single administration of the test was proposed by Chakrabartty (2020) where the test is dichotomized to two parallel subtests g-th and h-th and using lengths $\|X_g\|$ and $\|X_h\|$ of the parallel subtests vectors and $Cos\theta_{gh}$ where $\theta_{gh}$is the angle between the two vectors. As per definition, $\|X_g\| = \sqrt{\sum_{i=1}^{n} X_{gi}^2}$

$\|X_h\| = \sqrt{\sum_{i=1}^{n} X_{hi}^2}$ and $Cos\theta_{gh} = \frac{\sum_{i=1}^{n} X_{gi}X_{hi}}{\|X_g\|\|X_h\|}$

Error variance of the entire test is

$${S_E}^2 = \frac{1}{N}[\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\| \ \|X_h\|Cos\theta_{gh}] \quad (3)$$

and test reliability as per theoretical definition is

$$r_{tt} = 1 - \frac{{S_E}^2}{{S_X}^2} = 1 - \frac{\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\| \|X_h\| Cos\theta_{gh}}{N{S_X}^2} \quad (4)$$

One can find reliability of each dimension using (4) and scale reliability as a battery reliability where the scale score/ battery score is equal to sum of scores of *m*-dimension by

$$r_{tt(scale)} = \frac{\sum_{i=1}^{m} r_{tt(i)} S_{D_i}^2 + \sum_{i=1, i\neq j}^{m} \sum_{j=1}^{m} 2Cov(D_i, D_j)}{\sum_{i=1}^{m} S_{D_i}^2 + \sum_{i=1, i\neq j}^{m} \sum_{j=1}^{m} 2Cov(D_i, D_j)} \quad (5)$$

where $r_{tt(i)}$ and S_{D_i} denote respectively reliability and sample SD of the *i*-th dimension.

Properties

1. Test reliability, isomorphic to the theoretical definition is possible.

2. If *X* follows normal, true score of an individual with observed score X_0 is $[X_0 \pm SEM]$, where *SEM* = value of sample S_E

3. Split-half reliability as correlation between parallel sub-tests r_{gh} is different from theoretical value of r_{tt} obtained from (4).

4. For a given data set, one can compute S_E^2, ${S_T}^2$, S_X^2, and r_{tt} using (3) and (4) and find population estimates. Unbiased and consistent estimate of S_X^2 is $\frac{1}{N-1}\sum(X_i - \bar{X})^2 = \frac{N}{N-1}S_X^2$.

5. Theoretically defined reliability by (4) helps to test whether the population reliability = 1 i.e. to test $H_0: \sigma_X^2 = \sigma_T^2$ against $H_1: \sigma_X^2 > \sigma_T^2$. This can be tested using test statistic $F = \frac{S_X^2}{S_T^2}$ and reject H_0 if $F > F_{\alpha,(N-1,N-1)}$.

Validity

Validity of a QOL instrument is often reported as construct validity i.e. correlation between the QOL score in question and criterion scale. Neither the 15-item Flanagan's QOLS nor the 16-item terrible-delighted scale exhibited psychometric attributes that support construct validity Construct validity vary for different choice of the criterion scale. For example, validity of QOLS was high with Life Satisfaction Index-Z ($r = 0.67$ to 0.75) (Wood et al., 1969); moderate with Duke-UNC Health Profile ($r = 0.25$ to 0.48) (Parkerson et al., 1981) and for the Arthritis Impact Measurement Scales (*r* = 0.28 to 0.44) (Meenan et al., 1980).

Factors influencing construct validity are: mismatch of dimensions covered and distributions of scores of the two scales, different score ranges, different sample types (say chronic pain patients and cancer patients), etc. Moreover, if r_{XY} is high (say) 0.75 for two scales X and Y, then 0.75 is the validity of X or Y? If r_{XY} is still more, need for two different scales may be questioned.

Remedial action: Consider factorial validity using PCA results of the scale in question. PCA of a multidimensional scale like QOL, results into a number of independent factors with eigenvalues >1. The first principal component with highest eigenvalue (λ_1) can be taken as the main factor for which the scale was developed. Thus, validity of the scale can be reflected by ratio of the first eigenvalue to the sum of all eigenvalues i.e. Factorial Validity $= \frac{\lambda_1}{\sum \lambda_i}$

Properties:

- Proposed measure is simple to comprehend and to calculate
- Item validity is given in terms of component loading = (the eigenvector) $\times \sqrt{\text{the eigenvalue}}$ which can be interpreted as the correlation of the item with the principal component or item validity.
- Sum of item validities $\neq$ Scale validity.

- Eigenvalue ≈ 0 indicates existence of multicolinearity among the items

Discriminating value

Discriminating value reflects ability of the scale to distinguish between individuals that have different degrees of the underlying construct (e.g. more or less severe disease). Discriminating value of Likert item ($Disc_i$) and test ($Disc_{Test}$) by various measures of dissimilarities using only the frequencies or probabilities of Item–Response categories were compared by Bastien, (2001) and found that Coefficient of variation (CV) is best where $Disc_i = \frac{SD_i}{mean_i}$ and $Disc_{Test} = \frac{SD_{Test}}{Mean_{Test}}$ and derived relationship between Cronbach α and $Disc_{Test}$ (with m-items) as

$$\alpha = (\frac{m}{m-1})(1 - \frac{\sum_{i=1}^{m} \bar{X}_i^2 . Disc_i^2}{\bar{X}^2 . {Disc_T}^2}) \quad (7)$$

Since, variance of the i-th item $S_{X_i}^2 = \bar{X}_i^2 . Disc_i^2 \ \forall \ i = 1, 2, \ldots, m \implies \sum_{i=1}^{m} S_{X_i}^2 = \sum_{i=1}^{m} \bar{X}_i^2 . Disc_i^2$ and Test variance $S_X^2 = \bar{X}^2 . {Disc_T}^2$

It can be proved that $(Disc_{Test})^2 = \frac{{CV\ True\ scores}^2}{r_{tt}}$ where $r_{tt} = \frac{S_T^2}{S_X^2}$ (8)

Thus, test reliability and $Disc_{Test}$ are related by a negative non-linear relationship.

Classifications

Classifications of individuals are often done by a recommended categorization of total scores to a finite number of categories. However, boundary points are to be decided ensuring that members within a class/cluster are similar (small within-group variance) and members between classes/clusters are dissimilar (high between-group variance). Quartile clustering classifies a group of individuals in four mutually exclusive classes viz. the quartiles Q_1, Q_2, Q_3, Q_4[34]. Quartile clustering of scale scores following normal distribution may be adopted for QOL scales because it is simple, appealing, adds clear meaning to the clusters, provides well-defined cut-off scores for the four mutually exclusive classes and assigns equal probability to each quartile/class i.e. $\int_0^{Q_1} f(x)dx = \int_{Q_1}^{Q_2} f(x)dx = \int_{Q_2}^{Q_3} f(x)dx = \int_{Q_3}^{Q_4} f(x)dx$

Discussions

Major limitations of QOL scales using summative Likert scores can be avoided by transforming item score (X) to equidistant score (E) by assigning different weights to response categories of different items $\rightarrow$ standardizing E-scores to Z-scores $\sim N(0,1) \rightarrow$ converting Z-scores to Y-scores for $1 \leq Y \leq 100 \rightarrow$ find dimension score (D_i) as sum of item-wise Y-scores $\rightarrow$ find scale scores (P_i) as sum of dimension scores. Y_i and P_i follow normal and parameters can be estimated from the data.

Such transformations generates continuous, monotonic, normally distributed data and facilitate inferences like estimation of population mean (μ), variance (σ^2), confidence interval of μ, testing statistical hypothesis like $H_0: \mu_1 = \mu_2$ or $H_0: \sigma_1^2 = \sigma_2^2$ etc. either for longitudinal or snap-shot data. In addition, it can assess progress/deterioration by a patient or a group of patients between two successive time periods, reflecting responsiveness of the scale and effectiveness of a treatment plan.

Equivalent score combinations (x_0, y_0) to integrate two QOL-scales X and Y were found where area under curve $f(x)$ up to x_0= area of the curve $g(y)$ up to y_0 where $f(x)$ and $g(y)$ represent respectively normal pdf of X and Y.

An assumption-free method is described to compute error variance (S_E^2), true score variance ($S_T{}^2$) and reliability of dimension and scale as per the theoretical definition. Normally distributed *Y*-scores can be used to estimate σ_X^2, σ_T^2 and r_{tt} and testing $H_0: r_{tt} = \frac{S_T^2}{S_X^2} = 1 \Leftrightarrow H_0: \sigma_X^2 = \sigma_T^2$ against $H_1: \sigma_X^2 > \sigma_T^2$.

A simple measure of validity of a multidimensional QOL scale is proposed as the ratio of the first eigenvalue to the sum of all eigenvalues. Item validity is proposed in terms of component loading.

Discriminating value of Likert item ($Disc_i$) and test discriminating value ($Disc_{Test}$) were defined as CV and relationship derived between Cronbach α and $Disc_{Test}$ and $Disc_{Test}$ and theoretically defined r_{tt}.

Advantages of quartile clustering discussed. It is simple, appealing, adds clear meaning to the clusters, provides well-defined cut-off scores for the four mutually-exclusive classes and assigns equal probability to each quartile. Quartile clustering using normally distributed *P*-scores is recommended.

The proposed measures improve quality of measurements of QOL scale, facilitate meaningful comparisons across groups and time and are critically relevant to policy makers and researchers in social and medical sciences.

Conclusions

Proposed method of transforming raw Likert scores to continuous, monotonic scores following normal distribution helps to avoid major limitations and undertake analysis under parametric set up. Suggested integration of several QOL scales has clear theoretical advantages. Assumption-free measures of reliability, validity, discriminating power, etc. may be used empirically to cover comprehensive areas of multidimensional QOL scales.

Future studies with multi-data sets involving more than one QOL scales are suggested along with issues relating to psychometric properties of the proposed transformation.

Declarations

Acknowledgement: Nil

Funding details: Nil

Conflict of interests: No potential conflict of interest is reported.

Data availability: No data used

Ethical statement: This is a methodological paper and no ethical approval is required

Authors' contributions: Sole author

References

Anderson, K. L. (1995). The effect of chronic obstructive pulmonary disease on quality of life**.** *Res Nurs Health, 18*, 547-556.

Andrews, F. M., & Crandall, R. (1976). The validity of measures of self-reported well-being. *Social Indicator Res, 3*, 1-19.

Bastien, C. H., Vallieres, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, *2*(4), 297-307.

Berzon, R. A., Donnelly, M. A., Simpson, R. L. Jr., Simeon, G. P., & Tilson, H. H. (1995). Quality of life bibliography and indexes:1994 update. *Qual Life Res, 4*, 547-569.

Boixadós, M., Pousada, M., Bueno, J., & Valiente, L. (2009). Quality of Life Questionnaire: Psychometric Properties and Relationships to Healthy Behavioral Patterns. *The Open Psychology Journal, 2*, 49-57.

Broadhead, W. E., Gehlbach, S. H., de Gruy, F V., & Kaplan, B. H. (1988). The Duke-UNC Functional Social Support Questionnaire. Measurement of social support in family medicine patients. *Med Care, 26*(7), 709-723.

Brooks, Richard AU. (1996). EuroQol: the current state of play. *Health Policy, 37*(1), 53-72.

Cella, D. F., Tulsky, D. S., Gray, G. et al. (1993). The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. *Journal of Clinical Oncology, 11*(3), 570-579.

Chakrabartty, Satyendra Nath. (2021). Integration of various scales for measurement of insomnia. *Research Methods in Medicine & Health Sciences, 2*(3), 102-111.

Chakrabartty, Satyendra Nath. (2020). Reliability of Test Battery. *Methodological Innovations*, 1-8.

Chakrabartty, Satyendra Nath. (2019). Limitations of insomnia severity index and possible remedies. *JSM Neurol Disorders Stroke, 5*, 1-9.

Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales In relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.

Cieza, A., & Stucki, G. (2005). Content comparison of health-related quality of life (HRQOL) instruments based on the international classification of functioning, disability and health (ICF), *Quality of Life Research, 14*, 1225-1237.

Euroqol Group. (1990). Euroqol-a facility for the measurement of health-related quality of life. *Health Policy, 16*, 199-208.

Flanagan, J. C. (1982). Measurement of the quality of life: Current state of the art. *Arch Phys Med Rehabil, 3*, 56-59.

Green, J. A., Kirwan, J. M., Tierney, J. F., Symonds, P., Fresco, L., Collingwood, M., & Williams, C. J. (2001). Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis. *Lancet, 8*, 358(9284), 781-6.

Hunt, S. M., McEwen, J., & McKenna, S. P. (1985). Measuring health status: a new tool for clinicians and epidemiologists. *Journal of the Royal College of General Practitioners, 35*, 185-188.

Jin, X., Al Sayah, F., Ohinmaa, A., Marshall, D. A., & Johnson, J. A. (2019). Responsiveness of the EQ-5D-3L and EQ-5D-5L in patients following total hip or knee replacement. *Qual Life Res, 28*(9), 2409-2417.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The Index of ADL: A Standardized Measure of Biologican and Psuchosocial Function. *JAMA, 185*, 914-919.

Martin, W. S. (1978). Effects of scaling on the correlation coefficient: additional considerations. *Journal of Marketing Rese arch, 15*, 314-318.

Meenan, R. F., Gertman, P. M., & Mason, J. H. (1980). Measuring health status in arthritis: The arthritis impact measurement scales. *Arthritis Rheum, 23*, 146-152.

Nayak, M. G., George, A., Vidyasagar, M. S., Mathew, S., Nayak, S., Nayak, B. S., Shashidhara, Y. S., & Kamath, Asha. (2017). Quality of Life among cancer patients. *Indian J Palliative Care, 23*(4), 445-450.

Neumann, L., & Buskila, D. (1997). Measuring the quality of life of women with fibromyalgia: a Hebrew version of the quality of life scale (QOLS). *J Musculoskel Pain, 5*, 5-17.

Parkerson, G. R., Gehbach, S H., Wagner, E. H., James, S. A., Clapp, N. E., & Muhlbaier, L. H. (1981). The Duke-UNC health profile: An adult health status instrument for primary care. *Med Care, 19*, 806-828.

Reeves, A. J., Baker, R. T., Casanova, M. P. et al. (2020). Examining the factorial validity of the Quality of Life Scale. *Health Qual Life Outcomes, 18*, 32.

Spitzer, W. O., Dobson, A. J., Hall, J., et al. (1981). Measuring the quality of life of cancer patients: A concise QL-index for use by physicians. *J Chronic Dis, 34*(12), 585-597.

Vidhubala, E., Laths, Kannan R. R., Mani, C. S., Karthikesh, K., Muthuvel, R., Surendran, V., Premkumari, R. (2005). Validation of quality of life questionnaire for patients with cancer - Indian scenario. *Indian Jr. Cancer, 42*, 138-144.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short form health survey (SF-36). A. Conceptual framework and item selection. *Medical Care, 30*, 473-483.

World Health Organization. (2001). International Classification of Functioning, Disability and Health: ICF. Geneva: WHO.

World Health Organization. (2000). World Health Organization Disability Assessment Schedule (WHODASII). Training Manual: a guide to administration. Geneva: WHO.

WHOQOL Group. (1994). *The development of the World Health Organization quality of life assessment instrument (the WHOQOL*). In J. Orley, & W. Kuyken (Eds.), *Quality of Life Assessment: International Perspectives*. Heidelberg: Springer Verlag.

Williams, A. (1995). The measurement and valuation of health: a chronicle. Discussion Paper 136. Centre for Health Economics, University of York.

Wood, V., Wylie, M. L., & Sheafor, B. (1969). An analysis of a short self-report measure of life satisfaction: Correlation with rater judgments. *J Gerontol, 24*, 465-469.