Assessment tools in Rhinology and Otology and Quality of Life: Methodological Issues

Satyendra Nath Chakrabartty
Indian Statistical Institute, Indian Institute of Social Welfare and Business Management, Indian Ports Association, New Delhi, India

* Corresponding Author: **Satyendra Nath Chakrabartty**

Article Info

ISSN (online): 2582-8940
Volume: 04
Issue: 01
January-March 2023
Received: 15-02-2023;
Accepted: 02-03-2023
Page No: 17-23

Abstract
Background: A number of tools are used for measurement of severity of Rhinology and Otology related disorders and their impact on Quality of Life. The tools differ in terms of length (number of items), width (number of response-categories/levels), dimensions covered, psychometric documentation and scores are not comparable.
Objectives: To address methodological issues of tests and suggests remedial measures by transforming ordinal item scores to follow normal distribution for meaningful evaluation of measurement properties and better utilization of such tests.
Methods: Using data-driven weights to response-categories of different items, ordinal item scores are converted to equidistant score (*E*-scores) in ratio scale with fixed zero point. Proposed scores (*P*-scores) are obtained from *E*-scores via linear transformations lie between 1 and 100 following normal distribution. Dimension scores and test scores are taken as sum of item-wise *P*-scores which also follow normal with parameters obtained from data
Results: Normally distributed *P*-scores satisfy the basic assumption of statistical techniques and offer platform for parametric analysis facilitating meaningful arithmetic aggregation, meaningful comparisons and better utilization of such tests. It also help to find reliability as per theoretical definition, factorial validity avoiding criterion variable, assessment of responsiveness, optimal number of clusters and efficiency of classification, equivalent scores of two tests, etc.
Conclusion: Proposed scores following normal and satisfying desired properties of measurement is recommended. Experts and researchers can derive benefits of the proposed score for better comparisons and prognosis.

Keywords: Equivalent scores, Factorial validity, Normal distribution, Theoretical reliability, Quality of Life

Introduction

Large numbers of tools are there for diagnostic and measurement of severity of Rhinology and Otology related disorders and their impact on Quality of Life (QoL). In addition to bio-markers, disease-specific QoL scales are used containing Likert items, Numerical rating scale (NRS), "True-False" type binary items, Euroqol 5-Dimensions (EQ-5D-5L) scale, etc. Patient reported Outcome (PRO) instruments used for assessment of individual's perceptions of his/her physical and psychosocial health-status highlighting impact of illness in a patient's day-to-day life and also to adjust treatment strategies, improve delivery of care and optimize clinical outcomes. QoL scales differ with respect to dimensions covered; numbers of items, number of levels in items, scoring method, values of reliability, validity, responsiveness, etc. and scores are not comparable. Different experiences of patients are measured by different QoL tools [1]. Measurement issues and associated statistics and psychometrics are foundational elements of the scales. The tools do not consider distribution of scores and most of them use summative score for dimensions and the test.

Desired properties of scoring a scale to get a single value are:
P_1: To ensure meaningful arithmetic aggregation of item scores to get scale scores reflecting position of individuals by monotonically increasing continuous variables i.e., gain in a domain/item score to increase the scale score
P_2: Computation of mean and variance and other moments of scale scores
P_3: Same range of scores for each item
P_4: Finding relative importance of the domains
P_5: Quantification of progress made by one or a group of individuals over times i.e. responsiveness.

The paper describes methodological issues of scales used in disease severity and QoL measures and suggests transformation of item scores to continuous, monotonic and normally distributed scores in the range 1 to 100, satisfying the above said desired properties and facilitating meaningful application of statistical analysis under parametric set up and better utilization of such tests.

Major shortcomings of summative scores

- Levels of an item are ordered but not equidistant. Equidistant property requires constant value of distance between *j*-th and *(j+1)*-th levels $\forall\ j$ =1, 2, 3, 4 for a 5-point item. Distance between successive levels is unknown and not uniform [2].
- -. Equal importance to the items and subscales (dimensions) for summative score ignores different contributions of items and dimensions to total score, different item-total correlations, and different factor loadings [3]. Non-admissibility of addition means statistics like mean, standard deviation (SD), correlation, regression, analysis of variance (ANOVA), Principal component analysis (PCA), Factor analysis (FA), Cronbach alpha (which uses item variances and test variance), etc. are not meaningful and may produce strange results[4].
- Test results are sample based without throwing light to estimations and testing hypothesis in terms of population parameters.
- Rating data are skewed with floor and ceiling effects, and normality checks are necessitated [5]. Discrete ordinal data are non-normal and violate assumptions of many statistical procedures [6]. Features like metric, presence of zero point and clearly defined operational procedure of scoring of scales are the basics for measurement [7].
- -Summation of two random variables $X + Y = Z$ is meaningful if distributions of *X* and *Y* have same shape. It is necessary to know probability density function (pdf) of *Z* for application of parametric statistical analysis and better comparisons.
- Computation of mean, SD of test scores assume admissibility of addition of ordinal data and equidistant response-categories. Comparisons by *t*-test, ANOVA, finding independent factors by PCA, FA, assume normal distribution of scores, which are not usually tested in empirical investigations.
- Results may go wrong if assumptions of the techniques used are violated. For example, high correlation between two variables *X* and *Y* is taken as linear relationship between *X* and *Y* and regression of the form $Y = \alpha + \beta X + \epsilon$ is fitted. But, r_{XY} may be high even if *Y* is non-linearly related with *X*. If *X* takes integer values from 1 to 30, r_{X,X^2}= 0.97;$r_{X,X^3} = 0.92$; despite each of X^2, X^3 being non-linear function of *X*. Clearly, high correlation may not imply linearity.
- EQ-5D-5L gives pattern of health-status of persons. Calculation of value sets for EQ-5D-5L with upper bound 5-5-5-5-5 and lower bound 1-1-1-1-1, can be questioned on soundness of central estimates of each dimension–level combination and may result in different variance at different range of values i.e. heteroskedasticity [8].

Illustrative scales

Chronic rhinosinusitis (CRS) is a chronic disease of adults. Probability [persons with CRS to have poor QoL] = 9* Probability [persons without CRS][9]. CRS is a heterogeneous group of inflammatory diseases of nasal and para-nasal cavitie with polyp formation (CRSwNP) or without polyps (CRSsNP). CRS evaluations with multiple intrinsic and extrinsic factors, wide spectrum of disease variants with diverse characteristics on clinical and pathophysiologic level are complex [10]. Symptoms of CRS include nasal discharge, post-nasal drip, nasal congestion and/or obstruction, facial pain, pressure or fullness and decreased sense of smell for duration ≥12 weeks with objective findings on either computed tomography or nasal endoscopy [11]. Clearly, CRS disrupts day-to-day life of patients by affecting leisure and sleep [12].
QoL scores of CRS patients are significantly lower in comparison with the same in other common chronic diseases like congestive heart failure, angina, chronic obstructive pulmonary disease, back pain [13]. The etiology and pathophysiology of CRS is not known. Specialists differ in defining rhinitis and sinusitis as one clinical entity, or regard both as separate diseases. CRS affect negatively QoL, as emerged from rhinosinusitis QoL instruments [14]. Poor QoL scores from Sino-nasal Outcome Test 22 (SNOT-22) are due to functional, physical and psychological aspects unique to CRS. Factors affecting QoL in CRS patients include: symptom types, comorbidities like gastro-esophageal reflux disease (GERD), socio-bio-demographic factors like age, gender, behavioral factors including smoking habit, etc.[15-16].
Since treatment for rhinosinusitis is based on symptoms and their impact, it is important to quantify the sensitivity to change in health-status i.e. responsiveness. Out of 16 QoL instruments measuring sinusitis, following three met basic requirements of validity, reliability, and responsiveness [17]:

- Chronic Sinusitis Survey (CSS),
- Rhinosinusitis Outcome Measure-31(RSOM-31),
- SNOT-16.

CSS measures sinusitis-specific symptoms and medications during the previous 8 weeks using six items. Standardized total CSS-scores range between 0–100 where lower scores signify greater impact of disease on patients.
RSOM-31 is a broad-based tool covering disease-specific rhinosinusitis and also general QoL measurements. RSOM-31 was compacted into SNOT-22 to improve SNOT-16.
Other Rhinology and Otology related disorders having effects on QoL are:

- Tinnitus and hearing loss with or without mass are primarily due to otospongiosis, labyrinthitis ossificans, superior semicircular canal dehiscence, enlarged vestibular aqueduct syndrome, etc. While Otospongiosis affects the bony labyrinth, the labyrinthitis ossificans

affects the membranous labyrinth. The imaging for tinnitus and hearing loss in adults are undertaken by otoscopic exam and audiometry testing. Computed tomography and MR imaging have different and complementary roles in identifying causes and evaluating the disease. QoL scores of hearing loss group without tinnitus> QoL scores for normal hearing with tinnitus group [18].

Tinnitus Handicap Inventory (THI) using *EQ-5D-6L* quantifies impact of tinnitus on daily life indicated a single factor solution. Thus, separate analysis of subscales are invalid [19-20].

- Bilateral vestibular hypo-function (BVH) is characterized by reduced/absence of vestibular function, primarily due to ototoxic, infectious, traumatic, autoimmune resulting in vestibular end organ and/or vestibular nerve dysfunctions. Major symptoms of BVH are imbalance (vestibular ataxia), decreased dynamic visual acuity, oscillopsia during head and body movements due to instability of gaze. Evaluations of vestibulo-ocular function to diagnose BVH and assessing severity are usually undertaken using caloric testing, rotatory chair testing, and video-based quantitative head impulse testing (vHIT). However, no tool allows evaluation of frequency range of vestibular sensors in a continuous fashion unlike hearing loss. To assess QoL of BVH-patients, Oscillopsia Severity Questionnaire was developed for assessing oscillopsia severity with nine number of 5-point items(1 to 5). SF-36 and Dizziness Handicap Inventory (DHI) were used to assess QoL linked to vestibular dysfunctions [21].
- Meniere's disease (MD) is more common among women, and prevalence increases with increase in age. MD is a chronic and intermittent disorder with a variety of symptoms like vertigo, hearing loss, tinnitus, aural pressure, disequilibrium, etc.[22]. Changes in QoL for MD patients who had failed diuretic therapy and treated with migraine prophylactic medications were evaluated[23] by Meniere's Disease Outcomes Questionnaire–Retrospective (MDOQ-R) scale with 18 multiple-choice items (36 paired items) for pre- and post-treatment conditions[24]. Change in QoL by difference of mean scores of pre- and post-treatment conditions can be questioned since mean is not meaningful for ordinal scores [4] and non-normal distribution of MDOQ-R scores violate assumption of *t*-test to compare means.
- Dysphagia (swallowing difficulties) and dysphagia-related interventions affects QoL as measured through Swallowing Disorders questionnaire (SWAL-QOL) containing 44 items distributed over 11 domains (Burden; Eating desire; Feeding duration; Symptom frequency; Food selection; Communication; Fear; Mental health; Social Function; Sleep; and Fatigue), which are scored on a scale of 0 to 100 and can be cumbersome to complete.

Self-reported QoL Instruments:

SNOTT-22, a 22-item questionnaire with four domains (rhinologic symptoms, ear and facial symptoms, sleep disturbance and psychological symptoms) is designed to include symptoms intimately related with rhinosinusitis. Participants indicate their degree of agreement on each item in a 6-point scale marked from 0 to 5 where 0: no problem, 5: most serious problem. Domain scores and total scores for each participant is sum of item scores. A SNOT-22 score >7 is taken as poor QoL and a score ≤ 7 is considered as normal [25].SNOT-22 scores can differentially predict and guide treatment modality across different groups with CRS symptoms [26].

The 36-Item Short Form Survey (*SF-36*) is a generic QoL tool containing 36-item, and eight sub-scales that came out from the Medical Outcome Study [27]. SF-36 does not provide overall score like $SF36_{Total}$ due to several independent dimensions (http://www.webcitation.org/6cfeefPkf). SF-36 has been used to evaluate post-operative outcomes after Endoscopic Sinus Surgery (ESS), health and socio-economic burden of chronic CRS [28-29]. SF-36 with low response rate in older age groups [30] does not consider "sleep". However, CRS patients suffer from sleep disruption which may result in reduced QoL, impaired cognitive functions and mood disturbances [31].

Nasal Obstruction and Symptom Evaluation (*NOSE*) has been used to classify nasal obstruction severity [32]. However, it is not suitable for CRS severity and outcomes measurements [33].

EQ-5D-5L, commonly used generic health-status measurement tool has been used to study clinical outcomes in CRS [34]. A value set of EQ-5D-5L gives a pattern of health-status of a person. Value set 12345 is different from 54321 or any permutation of 1, 2, 3, 4 and 5. While 12345 indicate extremely poor health-state for the 5th dimension, the reverse is indicated by 54321 implying different clinical needs of the two persons. Clearly, summative scoring of dimensions of EQ-5D-5L is not valid. QoL as per the hearing status and the presence of tinnitus was evaluated using EQ-5D-5L[18] where dimensions were mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. However, the study did not consider relative severity or grade of tinnitus.

DHI is a handicap scale to assess effect of dizziness and unsteadiness on QoL. 25-items are grouped into 3 subscales (emotional, functional, and physical aspects of daily living). An item score could be 0 (no), 2 (sometimes) and 4 (yes). DHI score ranges between 0 implying absence of handicap and 100 indicating worst self-perceived disability handicap [35]. Total DIH-scores are used for interpretation and not the three subscales.

Rhinosinusitis disability index (*RSDI*) is a disease-specific QoL questionnaire containing 30 numbers of 5-point items (0 to 4) where 0: never, 1: almost never, 2: sometimes, 3: almost always, 4: always. Maximum RSDI-score 120 indicates worst QoL and the minimum RSDI-score 0 indicates better QoL. RSDI gives total score and domain scores for functional (9 items:1-5, 13, 23, 28, 29), emotional (10 items:12, 14-19, 21, 26, 27) and physical (11 items: 6-11, 20, 22, 24, 25, 30). Contributions of sub-scales to total RSI are clearly different [36].

Perceived Control of Rhinitis Questionnaire (*PCRQ*) contains 8 items where 8 ≤ Total scores≤ 40. Higher score indicates greater perceived control of rhinitis.

Observations:

- Discrepancies exist in assessment of disease burden by objective and subjectively perceived measures, especially for statistical and clinical importance sinus surgery results [37].
- Heterogeneity of outcome assessment methodologies poses obstacle on treatment effectiveness evaluation and

comparison. While RSDI and SNOT-22 are more sensitive to emotional impact of CRS, CSS examines medication use and symptoms [1].

- Directions of scales differ. While higher score of RSDI, DHI, SNOT-22, SWAL-QOL, etc. imply poor QoL, reverse is true for CSS-scores
- Domains with higher number of items contribute more to total test scores.
- Challenge is to find total score of an individual in SF-36, EQ-5D-5L
- Scales differed with respect to domains considered, dimensional structure, cut-off scores, sensitivity of changes, etc. Better is to convert discrete test scores to continuous variables enabling detection of small changes and to find equivalent scores of different scales for integration of scales.

Main Points

- Tools for measurement of severity of Rhinology and Otology related disorders and their impact on Quality of Life are not comparable since they differ in terms of length (number of items), width (number of response-categories/levels), dimensions covered, psychometric documentation and scoring methods.
- The paper addresses methodological issues of such tools and proposes remedial measures by transforming ordinal item scores to follow normal distribution for meaningful evaluation of measurement properties and better utilization of such tests
- Normally distributed scores facilitate meaningful aggregation, satisfy desired properties and offer platform for parametric analysis including statistical testing. In addition, the proposed method also helps to find reliability as per theoretical definition, factorial validity avoiding criterion variable, assessment of progress/deterioration of one or a group of patients, efficiency of classification, equivalent scores of two tests, etc.

Proposed method

Proposed scores by [38] are obtained by following stages:

1: Convert item scores to equidistant scores (*E*-scores) as weighted sum where weights $W_{ij}{}'s$ are different for different levels of different items, and W_1, $2W_2$, $3W_3$, $4W_4$ and $5W_5$ forms an arithmetic progression with common difference > 0. Find maximum (f_{max}) and minimum frequency (f_{min}) of the levels of each item. Take initial weights $\omega_{ij} = \frac{f_{ij}}{n}$. Arrange $\omega_{ij}{}'s$ so that $\omega_{i1} < \omega_{i2} < \omega_{i3} < \omega_{i4} < \omega_{i5}$ where $\omega_{i1} = \frac{f_{min}}{n}$ and $\omega_{i5} = \frac{f_{max}}{n}$. Consider intermediate weight $W_{i1} = \omega_{i1}$ and common difference α so that

$W_{i1} + 4\alpha = 5W_{i5} \Rightarrow \alpha = \frac{5f_{max} - f_{min}}{4n}$

Thus, $W_{i2} = \frac{\omega_{i1} + \alpha}{2}$, $W_{i3} = \frac{\omega_{i1} + 2\alpha}{3}$; $W_{i4} = \frac{\omega_{i1} + 3\alpha}{4}$; and $W_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$. Compute final weights $W_{ij(Final)} = \frac{W_{ij}}{\sum_{j=1}^{5} W_j}$ enabling $\sum W_{ij(Final)} = 1$ and $j.W_{j(Final)} - (j-1).W_{(j-1)(Final)}$ = constant, value of which are different for different items.

2: Standardize *E*-scores of the *i*-th item, $Z_{ij} = \frac{E_{ij} - \overline{E_i}}{SD(E_i)} \sim N(0,1)$

3: Transform Z_i to proposed score P_i in the score range [1,100] by

$$P_i = (100 - 1)\left[\frac{Z_i - MinZ_i}{MaxZ_i - MinZ_i}\right] + 1 \quad (1)$$

P_i-scores in the range [1,100] are continuous, normally distributed implying better admissibility of arithmetic aggregation and can be used for any number of items with different number of levels. Dimension scores and test scores taken as sum of $P_i{}'s$ will also follow normal. Such procedure also helps to find scale/battery score (say $SF36_{Total}$).

Procedure to obtain normally distributed score of EQ-5D-5L, proposed by [8] are:

I: Find proportion of responses in *j*-th level of *i*-th item as $p_{ij} > 0$ and $\sum_{j=1}^{5} p_{ij} = 1$. For

Item 1, proportions are $p_{11} = \frac{f_{11}}{n}, p_{12} = \frac{f_{12}}{n}, p_{13} = \frac{f_{13}}{n}, p_{14} = \frac{f_{14}}{n}$, and $p_{15} = \frac{f_{15}}{n}$ for *n*-respondents who completed the entire questionnaire.

II: Consider p_{ij} as data-driven weights and assign numerical values to a health-profile as weighted sum. For example, profile 12345 for *i*-th person (Y_i) can be expressed as an expected value = $1(p_{11}) + 2(p_{22}) + 3(p_{33}) + 4(p_{44}) + 5(p_{55})$ which is different from 54321 for *j*-th person (Y_j) = $5(p_{11}) + 4(p_{22}) + 3(p_{33}) + 2(p_{44}) + 1(p_{55})$. Following similar approach, dimension scores can also be obtained. Scores as weighted sum are expected values and are continuous.

III: Standardize by $Z_i = \frac{Y_i - \overline{Y}}{SD(Y)} \sim N(0,1)$ and transform Z_i to proposed score P_i such that $1 \leq P_i \leq 100$ by the linear transformation given in equation (1)

Major benefits of proposed scores:

1. Find single total score of an individual for a test or battery including EQ-5D-5L
2. Sub-class scores and total scores are continuous, monotonic, normally distributed with better admissibility of addition and facilitate parametric analysis including estimation of population mean (μ), population variance (σ^2), confidence interval of μ, testing hypothesis like $H_0: \mu_1 = \mu_2$ or $H_0: \sigma_1^2 = \sigma_2^2$ either for longitudinal data or snap-shot data.
3. Classify and rank group of persons.
4. Test effectiveness of treatments/cares by H_0: $\mu_{P_{pre-group}} = \mu_{P_{post-group}}$ using paired *t*-test since pre-treatment group and post-treatment group are not independent.
5. Percentage progress/deterioration of *i*-th patient in *t*-th time-period (P_{it}) by $\frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} \times 100$ reflects responsiveness of the scale and effectiveness of a treatment plan for better prognostication. $P_{it} - P_{i(t-1)} > 0$ implies progress in *t*-th period over (*t*-1)-th period. Deterioration in terms of *P*-scores indicated by $P_{it} - P_{i(t-1)} < 0$ may be probed to identify the sub-classes where deterioration occurred and extent of deterioration for possible corrective actions. Similarly, progress for a group of persons is reflected if $\overline{P_{it}} > \overline{P_{i(t-1)}}$
6. $H_0: \frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} = 0$ is possible since ratio of two normally distributed variables follows χ^2 distribution
7. Plotting of progress/deterioration of one or a group of

patients across time can help to compare progress pattern i.e. response to treatments from the beginning.

8. Normality helps to estimate variance of each item and the scale and thus enables estimation of Cronbach alpha at population level.
9. *P*-scores help to find Equivalency or integration of two scales. Let $f(X)$ and $g(Y)$ denote respectively normal density function of *P*-scores for Scale *X* and Scale *Y*. Equivalent score combinations P_{01} and P_{02} for Scale *X* and *Y* respectively can be found by solving the equation $\int_{-\infty}^{P_{01}} f(X)\,dx = \int_{-\infty}^{P_{02}} g(Y)\,dy$ using normal probability table for a known value of P_{01}.
10. Normality distributed *P*-scores enable PCA and computation of factorial validity as $\frac{\lambda_1}{\sum \lambda_i}$, where λ_1 is the highest eigenvalue associated with the first principal component. Factorial validity reflects the main factor for which the test was developed and accounts for $\frac{\lambda_1}{\sum \lambda_i} \times 100$ percent of overall variability. Such factorial validity avoids the problems of construct validity (administration of two tests to the same sample) and is independent of criterion scale [38]. However, factorial validity needs to tally with clinical findings.
11.
12. Test reliability as per theoretical definition $\frac{\text{True score variance } (S_T^2)}{\text{Observed score variance } (S_X^2)}$ was proposed [40]

By dichotomizing a test in two parallel subtests (*g*-th and *h*-th) and finding Error variance S_E^2 by

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

where n is the sample size; $\|X_g\| = \sqrt{\sum_{i=1}^{n} X_{ig}^2}$ denotes length of the *g*-th vector; $\|X_h\|$ is defined similarly and θ_{gh} is the angle between the *g*-th and *h*-th vectors given by

$$Cos\theta_{gh} = \frac{\sum_{i=1}^{N} X_{g_i} X_{h_i}}{\|X_g\|.\|X_h\|}$$

Thus, $r_{tt(Theoritical)} = \frac{S_T^2}{S_X^2} = 1 - \frac{S_E^2}{S_X^2} = 1 - \frac{\frac{1}{N}[\|X_g\|^2 + \|X_h\|^2 - 2\|X_g\|\|X_h\|Cos\theta_{gh}]}{NS_X^2}$ (3)

g-th and *h*-th subtests are parallel if $H_o: \overline{P_g} = \overline{P_h}$ by *t*-test and $H_o: \sigma_{P_g}^2 = \sigma_{P_h}^2$ by *F*-test are accepted. Both *t*-test and *F*-test assume normal distribution of variables.

Equation (3) helps to test $H_0: r_{tt(Thepretical)} = 1$ which boils down to test $H_0: \sigma_X^2 = \sigma_T^2$ by *F*-test.

$r_{tt(Theoritical)} \geq$ Split-half reliability (r_{gh}) [41] and reliability of a battery with *K*-subtests by

$$r_{tt\ (Battery)} = \frac{\sum_{i=1}^{K} r_{tt_i} S_{X_i}^2 + \sum_{i=1}^{K} \sum_{i \neq j} \sum_{j-1}^{K} 2\,Cov(X_i, X_j)}{\sum_{i=1}^{K} S_{X_i}^2 + \sum_{i=1}^{K} \sum_{i \neq j} \sum_{j-1}^{K} 2\,Cov(X_i, X_j)} \qquad (4)$$

where battery score is taken as sum the scores of the sub-tests (without weights)

If the battery score is defined as $b = \sum_{i=1}^{K} W_i X_i$ where X_i denotes the score of the *i*-th sub-test and W_i is the corresponding weight of the sub-test such that $W_i > 0$ and $\sum_{i=1}^{K} W_i = 1$, then

$$r_{tt\ (Battery)} = \frac{\sum_{i=1}^{K} r_{tt_i} W_i^2 S_{X_i}^2 + \sum_{i=1}^{K} \sum_{i \neq j} \sum_{j-1}^{K} 2 W_i W_j\, Cov(X_i, X_j)}{\sum_{i=1}^{K} W_i^2 S_{X_i}^2 + \sum_{i=1}^{K} \sum_{i \neq j} \sum_{j-1}^{K} 2\, W_i W_j Cov(X_i, X_j)} \qquad (5)$$

However, different selection of weights will give different values of $r_{tt\ (Battery)}$.

12.Efficiency of classification may be assessed by Davies-Bouldin Index (DBI) which is based on ratio of within-cluster and between-cluster distances[40]. For *K*-number of classes DBI is computed by

$$DBI_K = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1\ (i \neq j)}^{K} Max\left[\frac{DiamC_i - DiamC_j}{\|C_i - C_j\|}\right] \qquad (6)$$

where diameter of *i*-th class $DiamC_i = \sqrt{\frac{\sum_{x_i \in C_i} \|x_i - C_i\|^2}{n_i}}$

C_i: Centroid or mean of the *i*-th class; n_i: Number of members in the *i*-th class.

Upper limit of DBI is 1 and lower value implies better efficiency.

The optimal number of clusters has the smallest DBI value, which can be obtained from the graph of DBI and Number of clusters.

Limitations

The proposed method assumes no missing data. If cases with missing data are dropped, it produces systematic bias which may undermine the validity of the sample representativeness.

Discussion

Ordinal item scores and health-profiles of 5D-5L are converted to continuous score in [1, 100] following normal distribution. Dimension scores and test scores are obtained as sum of item-wise *P*-scores and are normally distributed, parameters of which can be estimated from data.

Advantages of Normally distributed *P*-scores include

- Satisfy the basic assumption of statistical techniques like PCA, FA, *t*-test, paired *t*-test, *F*-test, etc. and enables estimation of population parameters and testing statistical hypothesis.
- Meaningful arithmetic aggregation, meaningful comparisons and better utilization of such tests
- Test reliability as per theoretical definition from single administration of a test and battery reliability avoiding inherent problems of Test-retest reliability and Cronbach's alpha.
- -Testing hypothesis $H_0: r_{tt(Thepretical)} = 1$ and also to test whether two tests are parallel.
- Undertaking PCA and computation of factorial validity i.e. validity of the main purpose for which test was constructed.
- Assessment of responsiveness i.e. progress/deterioration of one or a group of patients across time. Significance of progress/deterioration can be tested by χ^2 test. The curve showing P_{i_t} against time-periods gives progress-path over time. Comparison of progress-paths can help to draw important inferences.
- Optimal number of clusters and efficiency of classification by Davies-Bouldin Index considering ratio of within-cluster and between-cluster distances.
- Equivalent scores of two tests avoiding problems of linear equating, IRT model, statistical prediction and

equipercentile equating. This helps to equate boundary points of classification based on scores of two tests including cut-off points (X_0) such that individuals with scores $\leq X_0$ are normal healthy persons and those with scores exceeding X_0 have the disease.

Conclusions

Normally distributed proposed scores, satisfying the desired properties of measurement, facilitating parametric analysis and better computation of psychometric qualities is recommended. Practicing experts and researchers can derive benefits of the proposed score for better comparisons and prognosis. Simulation studies may be undertaken to evaluate merits of the proposed approach.

Declaration

Declaration:
Acknowledgement: Nil
Funding details: No funds, grants, or other support was received
Conflict of interests: No potential conflict of interest is reported.
Informed Consent: Not applicable
Clinical trials: Not applicable
Data availability: No data used in this methodological paper
Authors' contributions: Sole author

References

1. Quintanilla-Dieck L, Litvack JR, Mace JC, Smith TL. Comparison of disease-specific quality-of-life instruments in the assessment of chronic rhinosinusitis. Int Forum Allergy Rhinol. 2012, 2(6):437-43.10.1002/alr.21057.
2. Munshi J. A Method for Constructing Likert Scales. SSRN Electronic Journal. 1990; 10.2139/ssrn.2419366.
3. Parkin D, Rice N, Devlin N. Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? Med Decis Making. 2010; 30(5):556-565.
4. Marcus-Roberts HM, Roberts FS. Meaningless statistics. Journal of Educational Statistics. 1987; 12:383-394.
5. Clason DL, Dormody TJ. Analyzing data measured by individual Likert-type items. J Agric Educ. 1994; 35(4):4
6. Harwell, MR, Gatti GG. Rescaling ordinal data to interval data in educational research. Review of Educational Research. 2001; 71:105-131.
7. Yusoff, Rohana, Janor RM. Generation of an Interval Metric Scale to Measure Attitude, SAGE Open, 2014; 1-16, DOI: 10.1177/2158244013516768
8. Chakrabartty SN. Alternate method of scoring Euroqol Five- Dimensional Scales, Health Science Journal. 2021; (15):6:847.
9. Nyaiteera V, Nakku D, Nakasagga E, *et al*. The burden of chronic rhinosinusitis and its effect on quality of life among patients re-attending an otolaryngology clinic in south western Uganda. BMC Ear Nose Throat Disord. 2018; 26(18):10. doi: 10.1186/s12901-018-0058-z.
10. Akdis CA, Bachert C, Cingi C. *et al*. Endotypes and phenotypes of chronic rhinosinusitis: A PRACTALL document of the European Academy of Allergy and Clinical Immunology and the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol. 2013; 131(6):1479-90.
11. Fokkens WJ, Lund VJ, Mullol J, *et al*. European position paper on rhinosinusitis and nasal polyps (EPOS 2012). A summary for otorhinolaryngologists. Rhinology. 2012; 50(1):1-12.
12. Metson RB, Gliklich RE. Clinical outcomes in patients with chronic sinusitis. Laryngoscope. 2000; 110(S94):24-28. doi: 10.1097/00005537-200003002-00007
13. Van Crombruggen K, Van Bruaene N, Holtappels G, Bachert C. Chronic sinusitis and rhinitis: clinical terminology" Chronic Rhinosinusitis" further supported. Rhinology. 2010; 48(1):54-58.
14. Rudmik L, Smith TL. Quality of life in patients with chronic rhinosinusitis. Curr Allergy Asthma Rep. 2011; 11(3):247-52. doi: 10.1007/s11882-010-0175-2.
15. Alobid I, Benitez P, Bernal-Sprekelsen M, *et al*. Nasal polyposis and its impact on quality of life: comparison between the effects of medical and surgical treatments. Allergy. 2005; 60(4):452-458. doi: 10.1111/j.1398-9995.2005.00725.x.
16. Baumann I, Blumenstock G. Impact of gender on general health-related quality of life in patients with chronic sinusitis. Am J Rhinol. 2005; 19(3):282-287.
17. Linder JA and Atlas SJ. Health-related quality of life in patients with sinusitis. Curr Allergy Asthma Rep. 2004; 4:490-495.
18. Joo YH, Han KD, Park KH. Association of Hearing Loss and Tinnitus with Health-Related Quality of Life: The Korea National Health and Nutrition Examination Survey. PLoS One. 2015; 29;10(6):e0131247.10.1371/journal.pone.0131247
19. Newman CW, Sandridge SA, Jacobson GP. Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. J Am Acad Audiol. 1998; 9(2):153-160.
20. Baguley DM, Andersson G. Factor analysis of the tinnitus handicap inventory. Am J Audiol. 2003; 12(1):31-34.
21. Hermann R, Ionescu EC, Dumas O, Tringali S, Truy E, Tilikete C. Bilateral Vestibulopathy: Vestibular Function, Dynamic Visual Acuity and Functional Impact. Front Neurol. 2018; 10; 9:555. doi: 10.3389/fneur.2018.00555
22. Gates GA. Ménière's disease review. J Am Acad Audiol. 2005; 17(1):16-26.
23. Ghavami Y, Haidar YM, Moshtaghi O, Lin HW, Djalilian HR. Evaluating Quality of Life in Patients With Meniere's Disease Treated as Migraine. Ann Otol Rhinol Laryngol. 2018; 127(12):877-887. doi: 10.1177/0003489418799107
24. Kato BM, LaRouere MJ, Bojrab DI, Michaelides EM. Evaluating quality of life after endolymphatic sac surgery: the Ménière's Disease Outcomes Questionnaire. Otol Neurotol. 2004; 25(3):339-344.
25. Gillett S, Hopkins C, Slack R, Browne J. A pilot study of the SNOT 22 score in adults with no sinonasal disease. Clin Otolaryngol, 2009; 34(5):467-469. doi: 10.1111/j.1749-4486.2009.01975.x.
26. DeConde AS, Mace JC, Bodner T, Hwang PH, Rudmik L, Soler ZM, *et al*. SNOT22 quality of life domains differentially predict treatment modality selection in chronic rhinosinusitis. Int Forum Allergy Rhinol. 2014; 4(12):972-9.
27. McHorney CA, Ware JE, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability

across diverse patient groups. Med Care. 1994; 32(1):40-66.
28. Van Der Veen J, Seys SF, Timmermans M, Levie P, Jorissen M, Fokkens WJ, *et al*. Real life study showing uncontrolled rhinosinusitis after sinus surgery in a tertiary referral centre. Allergy. 2017; 72(2):282-90.
29. Ference EH, Stubbs V, Lidder AK, Chandra RK, Conley D, Avila PC, *et al*. Measurement and comparison of health utility assessments in chronic rhinosinusitis. Int Forum Allergy Rhinol. 2015; (10):929-36.
30. Andresen EM, Gravitt GW, Aydelotte ME, Podgorski CA.: Limitations of the SF-36 in a sample of nursing home residents. Age Ageing. 1999; 28(6):562-6.
31. Banoub RG, Hoehle LP, Phillips KM, Schulman BJ, Caradonna DS, Gray ST, *et al*. Depressed Mood Modulates Impact of Chronic Rhinosinusitis Symptoms on Quality of Life. J Allergy Clin Immunol Pract. 2018; 6 (6):2098-105.
32. Lipan MJ, Most SP. Development of a severity classification system for subjective nasal obstruction. JAMA Facial Plast Surg. 2013; 15(5):358-61.
33. Thomas AJ, Orlandi RR, Ashby S, Mace JC, Smith TL, Alt JA. Nasal obstruction has a limited impact on sleep quality and quality of life in patients with chronic rhinosinusitis. Laryngoscope. 2016; 126(9):1971-6.
34. Remenschneider AK, Scangas G, Meier JC, Gray ST, Holbrook EH, Gliklich RE, *et al*. EQ-5D-derived health utility values in patients undergoing surgery for chronic rhinosinusitis. Laryngoscope. 2015; 125(5):1056-61.
35. Jacobson GP, Newman CW. The development of the Dizziness Handicap Inventory. Arch Otolaryngol Head Neck Surg. 1990; 116:424-7.
36. Gairola P, Bist SS, Mishra S, Agrawal V.: Assessment of Quality of Life in Patients of Chronic Rhinosinusitis. Clin Rhinol An Int J. 2014; 7(3):96-99.
37. Zhang Z, Adappa ND, Doghramji LJ, *et al*. Quality of life improvement from sinus surgery in chronic rhinosinusitis patients with asthma and nasal polyps. Int Forum Allergy Rhinol. 2014; 4(11):885-92.
38. Chakrabartty SN. Scoring of health state utility instruments for better comparability. IASR Journal of Medical and Pharmaceutical Science. 2021; 1 (3):14-21. https://iasrpublication.com/media/articles/IJMPS01030 02-14-21.pdf
39. Parkerson, HA. Noel, M., Pagé MG, Fuss S, Katz J, Asmundson GJ. Factorial Validity of the English-Language Version of the Pain Catastrophizing Scale–Child Version, The Journal of Pain, 2013; 14 (11), 1383-1389, https://doi.org/10.1016/j.jpain.2013.06.004
40. Chakrabartty, Satyendra Nath. Reliability of test battery. Methodological Innovations, 2020; 1-8. DOI: 10.1177/2059799120918340
41. Davies D, Bouldin W. A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979; 1(2):224-227.