Original Article
IASR Journal of Medical and Pharmaceutical Science
Website: https://iasrpublication.com/ijmps

Scoring of health state utility instruments for better comparability

Satyendra Nath Chakrabartty
Indian Statistical Institute, Indian Maritime University, Indian Ports Association

ABSTRACT

Background: Health state utility (HSU) instruments differ in descriptive system, domains considered, scoring methods, analysis of ordinal data without checking limitations and assumptions of such analysis. A high score indicates higher favorable health state for SF 36, HUI 3, SF 6D, AQoL (response scores)unlike scales like EU-5D-5L, 15 D. **Objectives**: To provide method for converting item-response scores to continuous, monotonic, normally distributed scores in 1 to 100, avoiding limitations of existing HSU scales with different number of items and different number of levels. **Materials and Method**: The methodological paper involves no data from patients. Here, health state of an item score is transformed to equidistant-scores by weighted sum where weights are based on frequency of responses to each level of an item. Weights are different for different levels of different items. Standardize *Y*-scores to *Z*-scores $\sim N(0,1)$and convert *Z*-scores to proposed scores (P_i) in the range 1 to 100. Scale scores as sum of P_i's follows normal distribution. **Results:** *P*-scores facilitate inferences like estimation and testing of statistical hypothesis on equality of population parameters either for longitudinal or snap-shot data, assessment of progress/deterioration by a patient or a group of patients. Equivalent score combinations $(x_{0,}y_0)$ to integrate two HSUs were found. **Conclusions**: The proposed method avoids differences due to descriptive systems, scale effects, utility algorithms, etc. and contributes to improve scoring of HSU instruments avoiding limitations of ordinal scores and facilitating analysis under parametric set up for meaningful comparisons.

Keywords: *Utility; Normal distribution; Health states; Likert items; Responsiveness; Equivalent scores*

***Corresponding Author**
Satyendra Nath Chakrabartty
Indian Statistical Institute, Indian Maritime University, Indian Ports Association

INTRODUCTION

Health state utility (HSU) instruments differ in descriptive system (number and format of items, domains considered) and utility algorithm i.e. system of scoring (or valuing health states) of items and are not comparable. As a result, the best HSU instrument remains unanswered. Because of differences in envisaged domains of HSU instruments, they measure different things [1].Researchers differed in their choice of HSU scales. For example, [2] compared EQ-5D- 5L, SF-6D, HUI 3 and 15D and observed that the HSUs with non-linear relationships gave rise to different utilities for the same individual or same group of individuals. AQoL was chosen by [1] as an additional scale to investigate differences of descriptive systems of five HSU scales. [3]also included AQoL 8 and found that 66 % of the differences between utilities was attributable to the descriptive systems, 30.3 % to scale effects and 3.7 % to micro-utility effects. While 15 D and HUI 3 are symptoms oriented,EQ-5D and SF-6D are more oriented towards functioning. AQoL considers items on symptoms, functioning, social relationships and happiness is a hybrid of health and well-being instruments.

Usually, evaluations of health states from different HSUs are done by first calculating utilities and un-weighted values for each instrument and determining scale effects by linear relationship between utilities, then by comparing effect of the descriptive system of scale-adjusted values and 'micro-utility effects' using the unexplained difference between utilities and values [3]. However, methods used to find utility weights are different for different HSU instruments.

Questions can be raised on meaningfulness of the scoring in terms of the following:

- Calculating values/scores of an instrument when addition of item scores is not admissible for ordinal Likert scores [4], since levels of an item are not equidistant [5; 6] and respondents do not perceive the levels as equidistant [7].

- Likert scores are skewed and do not follow Normal distribution, which is a common assumption of Regression, ANOVA, Principal Component Analysis (PCA), Factor Analysis (FA), Structural Equation Modeling (SEM), t-test, F-test, Cronbach alpha, estimation of population parameters, etc. Thus, all conclusions are based on sample observations without throwing light on what is happening to the population.
- Addition of dimensions scores giving equal importance to the dimensions may not be justified since dimensions have different contributions to total score, different values of dimension-total correlations, different factor loadings, etc. [8]
- Item scores do not follow same or similar distribution. Unknown and different distributions of items and resultant domains/dimensions make it difficult to interpret X ± Y, further operations on X ± Y and joint distribution of X ± Y when X and Y follow two different distributions.
- Anchor value of zero in an item or zero value in rescaled scale is not desirable, since it tends to lower mean and variance. Too many zeros to a rescaled item will artificially lower the covariance and hence correlation with that item.
- EQ-5D -5L questionnaire results in categorical health-state or health-profile of the individual in terms of a 5-digited number, consisting of any permutation of 1, 2, 3, 4 and 5 where repetition of any number is allowed. How to assign a numerical value to a given health-state emerging from EQ-5D-5L so that score follow say Normal distribution?
- The SF-36has 28 items in Likert format (3-point, 5-point and 6-point), seven binary items and another item regarding reported health transition over the last year [9] and does not provide support to calculate a single measure of health-related quality of life, like "SF-36 Total/Global/Overall Score" since such overall scores are not appropriate. How to find $SF36_{Total}$ with better admissibility of arithmetic aggregation avoiding limitations of summative Likert scores?
- How to ensure scoring of a scale starting with item-response matrix of a scale satisfies desired properties of measurement like meaningful comparisons, ranking, classification of individuals, assessment of progress/deterioration across time for an individual or a group of individuals and drawing of progress-paths and parametric analysis including estimation and statistical test of hypothesis?
- How to integrate two scales by finding equivalent score combinations of threshold values (X_{0i}, X_{0j}) so that area under a given threshold value of X_{0i} for the i-th scale is equal to area under X_{0j} for the j-th scale ($i \neq j$)?

Methodological issues for assignment of single numerical value to a health state i.e. scoring health states demands that the proposed individual scores are additive, follow a known distribution, facilitate computation of mean, variance and correlations matrix of dimensions for a sample along with parametric analysis including estimation of population parameters and testing of statistical hypothesis, etc.

Ignoring the issues of selection of dimensions and items, the paper makes a brief review of major HSU instruments and proposes method of transferring item score (responses to the item) to continuous score following Normal distribution irrespective of number of items and number of levels in a sub-scale and enabling computation of dimension score as sum of normally distribute item scores and total score as sum of such dimension scores.

Brief description of HSU instruments:

EQ-5D-5L: Five dimensions, each with one item in 5-point format [10]. Value set of an individual is a permutation of 1, 2, 3, 4 and 5 where repetitions are allowed. Lowest value set is 1-1-1-1-1 implying no problem in any dimension and highest value set is 5-5-5-5-5 indicating maximum problem in each dimension. Clearly, a person's score is categorical since it categorize a person in one of the possible 3125 $=5^5$ categories. Here, frequency of each observed category is admissible. Summative Likert score of value set 12345 and 54321 are equal and cannot differentiate health status of persons at different dimensions with a particular total score. Attempts were made to convert5D-5L profiles into a single measure using preference-based weights[11].Methods to describe the profile data of 5D-5L at a given time point and changes in profiles between different time points were reviewed [12]. After reviewing psychometric properties of EQ-5D-5L, further and more rigorous exploration was suggested [13]. Researchers attempted to assess distributional characteristics of the EQ-5D-5L. Shannon index was used to find informational richness and evenness (rectangularity) of dimensions [14]; Health State Density Curve (HSDC) and Health State Density Index (HSDI) proposed by [15] used cumulative frequencies of health-states anddegree of concentration in self-reported health measures, etc. Each has limitations. Pertinent point is to find probability distribution of health-states of EQ-5D-5L or to transform the EQ-5D-5L data to a known distribution say Normal distribution which is frequently used in statistical analysis.

A single numerical value of a value set for an individual could help in assessment of illness severity, identification of areas requiring priorities, estimation of population parameters to go beyond sample based observations, comparisons between populations or one population across time, etc.

The 36-Item Short Form Health Survey questionnaire (*SF-36*) is a self-reported questionnaire with eight sub-scales for evaluating Health-Related Quality of Life. Original scores of an item are rescaled so that range of rescaled sores is 0 to100. Score of a sub-scale is sum of rescaled values of the items pertaining to the sub-scale. Descriptive statistics, reliability, validity, etc., are obtained separately for each sub-scale. Mean, Standard deviation (SD) and shape of distribution are different for “Yes-No” type, 3-point, 5-point, 6-point Likert items. Reliability, validity, and discriminating power are different for *K*- point scales for *K*= 2, 3, 5, 6, and so on [16]. Multidimensional structures of SF-36 have been confirmed using factor analysis [17], structural equation model analysis [18].

Rescaling involves reverse scoring for negatively worded items. For example, item 20 and 32, under social functioning, are in 5-point format. In the former, a choice of level 5 indicates maximum limitations in social functioning. But similar response-choice in item 32 indicates the reverse. To maintain parity of direction, choice of levels and corresponding recorded values are decided so that higher response choice is proportional to recorded values.

HUI 3: Generic HUI gives single summary scores of HRQL on a scale between 0 and 1 where 0 utility indicates death and 1 indicates perfect health (all attributes at level 1) (www.healthutilities.com). While HUI 2 with seven attributes has a score range in the interval -0.03 to 1.00; the HUI 3 with eight attributes is defined in the interval -0.36 to 1.00 where negative scores represent states worse than being dead. Instead of dimension or domain, HUI uses the term “attribute”. Each attribute has six levels (except “Speech”, “Emotion” and “Pain” each of which has five levels). However, psychological health at the level of symptoms or activity is not captured by any item of HUI 3. Example of marking the levels for the attribute “Vision” is shown below:

1. Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.
2. Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, but with glasses.
3. Able to read ordinary newsprint with or without glasses but unable to recognize a friend on the other side of the street, even with glasses.
4. Able to recognize a friend on the other side of the street with or without glasses but unable to read ordinary newsprint, even with glasses.
5. Unable to read ordinary newsprint and unable to recognize a friend on the other side of the street, even with glasses.
6. Unable to see at all.

Scoring of HUI 3 involves consideration of a Look – up Table published by the Health Utilities Inc. (www.*healthutilities.com)* with six rows (corresponding to six levels) and eight columns (one for each attribute arranged in the order of Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition and pain). Value in each cell is considered as the weight (W_{ij}) for the *i*-th level and the *j*-th attribute and computing formula is the weighted sum. For example, Utility value (U) of a response set 6-5-4-3-2-1-2-3 is

$1.371(W_{6Vision} * W_{5Hearing} * W_{4Speech} * W_{3Ambulation} * W_{2Dextirity} * W_{Emotion} * W_{2Cognition} * W_{3Pain}) - 0.371$

Here, sum of weights exceeds unity for a column as well as for a row. Thus, weighted sum violates convex property. Attributes of HUI 3 are mostly structurally independent i.e. attributes are poorly correlated (0.02 to 0.35) unlike EU-5D-5L (0.24 to 0.64) [19]. If Cronbach alpha of HUI 3 with *n*-attributes is taken as$(\frac{n}{n-1})(1-\frac{Sum of variance of attributes}{Variance of HUI\ 3})$, it will show poor reliability since poor correlations among the attributes will give poor covariance between pair of attributes and thus, make variance of HUI 3 almost equal to sum of variance of attributes.

*SF-6D*uses 11 items of the SF-36 and contains six domains. Items in a domain have different number of levels like Pain(6 levels), Mental health (5), Physical functioning (6), Social functioning (5), Role limitations (4) and Vitality

(5). Combining scores of domains containing different number of items and levels could be possible, if each item score is transferred to follow same distribution. Assignments of numerical values to health states of SF-6D are done using the algorithm by [20] which generates preference weights from a representative sample. Thus, the scoring is heavily dependent on the sample.Programs are available to convert SF 36 scores to SF-6D scores. SF-6D emphasis on psycho-social aspects of health also under a generic heading; ‘mental health’. But, the description of its various severity levels refers more specifically to being ‘depressed or very nervous’ at different frequencies. It measures limitations of a patient without considering trade-offs between different dimensions of health; say between pain and physical functioning, or between quality and length of life. SF-6D as an independent instrument was not favoured [21].

15D-instrument (http://15d-instrument.net/15d/) consist of 15 dimensions: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort/symptoms, depression, distress, vitality, and sexual activity. Each dimension has one item with five ordinal levels giving over 30 billion possible health states. Each of the 15 dimensions consists of one item with five levels marked as 1 to 5 where 1 represents “no problem” and 5 represents “maximum problem” in that dimension. The single index 15D score on a 0-1 scale is calculated from the health state descriptive system based on utility weights and uses arithmetic aggregation formula[22]. Calculation of 15 D score consider computation of $W_j(X_j)$ as the average value on various levels of j-th dimension and $I_j(X_j)$ as the average relative importance to various levels of j-th dimension and 15 D score as $\sum I_j(X_j)[W_j(X_j)]$.

Calculation of average values assumes admissibility of addition of ordinal scores and equal importance to the items, both is not justified. *AQoL* (https://www.monash.edu/business/che/aqol) has four different versions: The longest (8D)includes 35 items distributed among eight dimensions: Independent Living, Happiness, Mental Health, Coping, Relationships, Self-Worth, Pain, Senses. The shortest (4D) version has 12 items in four dimensions: Independent Living, Mental Health, Relationships, and Senses. Total number of health states is as high as 2.4×10^{23} for AQoL 8D. Items of AQoL attempt to measure global life satisfaction and happiness, social relationships and impact of health on social relationships. As a psychometric instrument, AQoL scores are addition of the unweighted response order of each question. However, utility scores are derived using preference weights.

Multi-dimensional HSU instruments differ in direction of scoring. For example, a high score indicates higher favorable health state for SF 36, HUI 3, SF 6D, AQoL (response scores) but reverse is true for scales like EU-5D-5L, 15 D. For comparison of HSU instruments, it will be necessary to rescale by reverse scoring for negatively worded items so that rescaled item indicates less problems and maximum rescaled score to indicate perfect heath.

PROPOSED METHOD

Ordinal responses of each item are proposed to be transferred to continuous scale following Normal distribution. By convolution property, normally distributed item scores can be added to get dimension/domain/attribute score following normal and sum of domain scores will give the single score of the instrument which also follows normal. Proposed transformations in stages are elaborated in detail separately for EU-5D-5L and scales with Likert items like SF 36 with K-numbers of levels, K= 2, 3, 4, 5, 6, etc. assuming no missing values.

EQ-5D-5L: Two methods of getting EQ-5D-5L scores using geometric aggregation and arithmetic aggregation after few linear transformations and preferred the later for higher theoretical advantages. The multi-stage method of arithmetic aggregation is described below.

Stage I: Find proportion of responses to each level of an item. 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. In general, if p_{ij} denotes proportion of responses in j-th level of i-th item, each $p_{ij} > 0$ and $\sum_{j=1}^{5} p_{ij} = 1$ and total of proportions of level 1 is $p_{01} = \frac{\sum_{i=1}^{5} f_{i1}}{5n}$. Similarly, one can find p_{02}, p_{03}, p_{04} and p_{05}.

Stage II: Consider p_{ij} as data-driven weights and assign numerical values to health profile of a person as weighted sum. For example, profile of 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 score of each dimension can also be obtained. Scores as weighted sum are expected values and is continuous.

Stage III: Standardize by $Z_i = \frac{Y_i - \bar{Y}}{SD(Y)} \sim N(0,1)$

Stage IV: Stage IV: Transform Z_i to proposed score P_i to have a score range [1,100] by the following linear transformation $P_i = (100-1)\left[\frac{Z_i - MinZ_i}{MaxZ_i - MinZ_i}\right] + 1$ (1)

SF 36: In line with method proposed by [24], requisite transformation of SF 36 items can be done as follows: Consider the eight sub-classes of SF 36, after taking reverse scoring of all "negatively phrased" items so that choice of higher level indicates more favorable health state i. e. response categories of each item are ordered from low to high where the lowest level is marked as 1. The sub-classes have different number of items and formats like binary, 3-point, 5-point and 6-point. The proposed method of converting raw item score (X) to equidistant sore (E) followed by standardization and further transformation to follow Normal in a desired score-range.

Equidistant scores

Let X_{ij} be the discrete raw score of the i-th item in the j-th response category of an individual. For a 5-point item, revised score (RS) is $W_{ij}X_{ij}$ where W_{ij}'s are weights to different levels of the item satisfying $W_{ij} > 0, \sum_{j=1}^{5} W_{ij} = 1$. RS will be equidistant and monotonic if $W_1, 2W_2, 3W_3, 4W_4$ and $5W_5$ forms an arithmetic progression with positive value of common difference. Such weights can be found using frequency of each level as follows:

For each item, find maximum (f_{max}) and minimum frequency (f_{min}) of the levels. Find initial weights $\omega_{ij} = \frac{f_{ij}}{n}$. Arrange the $\omega_{ij}^{'}s$ so that $\omega_{i1} < \omega_{i2} < \omega_{i3} < \omega_{i4} < \omega_{i5}$ where $\omega_{i1} = \frac{f_{min}}{n}$ and $\omega_{i5} = \frac{f_{max}}{n}$. Put intermediate weight $W_{i1} = \omega_{i1}$ Find the common difference α so that $W_{i1} + 4\alpha = 5W_{i5} \Rightarrow \alpha = \frac{5f_{max} - f_{min}}{4n}$. Define other intermediate weights as $W_{i2} = \frac{\omega_{i1} + \alpha}{2}$, $W_{i3} = \frac{\omega_{i1} + 2\alpha}{3}$; $W_{i4} = \frac{\omega_{i1} + 3\alpha}{4}$; and $W_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$. Divide each W_{ij} by $\sum_{j=1}^{5} W_{ij}$ and get 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 will be different for different items.

Observations

$W_{j(Final)}$ are based on empirical probabilities of basic Item-score matrix.

ii) If $f_{ij} = 0$ for a particular j-th level of an item, the method fails and can be taken as zero value for scoring Likert items as weighted sum

iii) Mean, variance and range of item scores and sub-test scores will get reduced in comparison to the same from usual summative scores.

iv) Generated scores (E) as weighted sum are continuous and equidistant implying better admissibility of arithmetic aggregation

v) The can be used for any item with k-number of levels

Follow Stages to standardize E-scores by Z-transformation for the i-th item, $Z_{ij} = \frac{E_{ij} - \overline{E_i}}{SD(E_i)} \sim N(0, 1)$ and convert the Z-scores to proposed P-scores by equation (1)

For the i-th item, let $P_i \sim N(\mu_i, \sigma_i^2)$ and $1 \leq P_i \leq 100$ where μ_i and σ_i^2 can be estimated from the data. It may be noted that P-score of an item can be obtained irrespective of number of items and number of response categories in a sub-class.

Sub-class and total SF 36 scores:

Sub-class score of an individual is sum of normalized item scores following normal with mean $\sum_i \mu_i$ and SD = $\sqrt{\sum \sigma_i^2 + 2\sum_{i \neq j} Cov(P_i, P_j)}$

Scale/battery score or total SF-36 score ($SF36_{Total}$) is similarly taken as sum of sub-class scores, which also follows normal.

Properties and Benefits

The proposed method converts health state score or ordinal item-wise scores of each item to continuous, monotonic scale following normal distribution. Sum of normally distributed item-wise scores is taken as the sub-scale score and sum of sub-scale scores is taken as the single measure of total score (P) following normal distribution, parameters of which can be estimated from the data. Major benefits are:

1. Find single total score of an individual
2. Transferred item score, sub-class score sand total scores are continuous, monotonic, normally distributed with better admissibility of arithmetic aggregation. They 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 $H_0: \sigma_1^2 = \sigma_2^2$ etc. either for longitudinal data or snap-shot data.
3. Classifying and ranking group of persons answering the instrument
4. Testing effectiveness of treatments/cares by testing 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 the i-th person between two successive time-periods can be assessed by $\frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} \times 100$ where P_{it} denotes P-score of the i-th patient in t-th time period. The ratio $\frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}}$ 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. The reverse is true for $P_{it} - P_{i(t-1)} < 0$. Deterioration in terms of P-scores 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 $\overline{P_{it}} > \overline{P_{i(t-1)}}$. Problem of large difference in measuring changes over time periods by various scales can be avoided by $\frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} \times 100$
6. Statistical test of significance of progress/deterioration can be made since ratio of two normally distributed variable follows χ^2 distribution
7. Plotting of progress or deterioration of a patient or a group of patients across time can help to compare progress pattern i.e. response to treatments from the start.
8. Normality will help to find estimate of variance of each item and variance of the scale and thus enables estimation of Cronbach alpha for the population.
9. High correlation between raw score of and P-scores obtained by linear transformations indicate that proposed transformations did not change much the structure of the data.
10. Equivalency or integration of two scales can be studied by transferring responses of each scale to normally distributed P-scores say $f(X)$ and $g(Y)$ denoting respectively normal density function of P-scores for Scale 1 and Scale 2 and finding equivalent score combinations P_{01} for Scale 1 and P_{02} for Scale 2 by solving the equation $\int_{-\infty}^{P_{01}} f(X)\,dx = \int_{-\infty}^{P_{02}} g(Y)\,dy$ which can be solved for a known value of P_{01} using normal probability table.

CONCLUSIONS

Two methods are proposed for transferring configural scores like health states and item score of Likert-type items to continuous score following Normal distribution irrespective of number of items and number of levels in a sub-scale. Depending on the descriptive system, amulti-dimensional HSU instrument can be scored by one of the methods suggested to get dimension scores and single total score of an individual. Normality of the proposed scores with better admissibility of arithmetic aggregation gives significant benefits like undertaking parametric analysis, estimation of population parameters and testing of hypothesis for longitudinal data or snap-shot data.

In addition, for better classification and ranking of persons, the method helps to assess responsiveness of the instrument, effectiveness of treatments/cares, drawing progress-path from the start for a patient or group of patients which may have bearing with survival analysis. Normality also helps to find equivalent score combinations of two or more HSU instruments.

Considering theoretical advantages, the proposed method of transforming raw scores of HSU instruments to follow normal distribution is recommended for more meaningful comparisons and inferences. Future studies with longitudinal data can be undertaken to find sensitivity of the proposed score over time with emphasis on progression of disease and to different therapeutic interventions, etc.

REFERENCES

1. Olsen, J.E. and Mission, R.A.: A conceptual map of health‐ related quality of life dimensions: key lessons for a new instrument, *Quality of Life Research*; 2020; 29:733–74 https://doi.org/10.1007/s11136-019-02341-3
2. Gamst-Klaussen, T., Chen, Gang, Lamu, AN and Olsen, JA: Health state utility instruments compared: inquiring into nonlinearity across EQ-5D-5L, SF-6D, HUI-3 and 15D. Quality of life research; 2016, 25(7):1667-78. doi: 10.1007/s11136-015-1212-3.
3. Richardson, J., Iezzi, A., & Khan, M. A: Why do multiattribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects.*Quality of Life Research*, 2015; 24,2045–2053.
4. Jamieson, S.: Likert scales: How to (ab) use them. *Medical Education*,2004,38, 1212 -1218
5. Bastien, C. H., Vallieres, A., & Morin, C. M.: Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine*, 2001; 2(4), 297–307
6. Ferrando, P. J.: A Kernel density analysis of continuous typical-response scales. *Educational and Psychological Measurement*, 2003, 63, 809-824
7. Lee JA and Soutar G.: Is Schwartz's value survey an interval scale and does it really matter? *J Cross-Cultural Psychol*; 2010, 41: 76–86.
8. 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
9. Ware JE, Kosinski MA, Gandek B: *SF-36 Health Survey: Manual and interpretation guide*. 2005; Lincoln: Quality Metric Inc.
10. Brooks R. EuroQol: the current state of play. *Health Policy.* 1996;37(1):53–72. doi: 10.1016/0168-8510(96)00822-6
11. Stolk, E., Ludwig, K., Rand, K., van Hout, B., & Ramos-Goni, J. M.: Overview, update, and lessons learned from the international EQ-5D-5L valuation work: Version 2 of the EQ-5D-5L valuation Protocol. *Value in Health*, 2019, 22(1), 23–30.
12. Devlin N., Parkin D., Janssen B. Analysis of EQ-5D Profiles. In: Methods for Analysing and Reporting EQ-5D Data, 2020, Springer, https://doi.org/10.1007/978-3-030-47622-9_2
13. Feng, You‐ Shan, Kohlmann, Thomas, Mathieu F. Janssen, MF and Buchholz, I.: Psychometric properties of the EQ‐ 5D‐ 5L: a systematic review of the literature, Quality of Life Research (2021) 30:647–673, https://doi.org/10.1007/s11136-020-02688-y
14. Buchholz I, Janssen MF, Kohlmann T, Feng YS.: Asystematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoecon*, 2018, 36(6):645–661

15. Zamora B, Parkin D, Feng Y, Bateman A, Herdman M, Devlin N.: New methods for analyzing the distribution of EQ-5D observations. OHE Research Paper, 2018 www.ohe.org/publications/newmethods- analysing-distribution-eq-5d-observations
16. Preston, Carolyn C. and Colman, A. M.: Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences, *Acta Psychologica*, 2000, 104, 1-15
17. Guermazi M, Allouch C, Yahia M, et al. Translation in Arabic, adaptation and validation of the SF-36 Health Survey for use in Tunisia. *Ann Phys Rehabil Med*, 2012; 5: 388–403.
18. Anagnostopoulos F, Niakas D, Pappa E. Construct validation of the Greek SF-36 Health Survey. *Qual Life Res.*, 2005*;* 14: 1959–1965
19. Feeny, D., Furlong, W., Torrance, G., Goldsmith, C., Zhu, Z., DePauw, S., et al. Multi attribute and single attribute utility functions for the Health Utilities Index Mark 3 System, *Medical Care*, 2002, 40(2), 113–128
20. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ.* 2002;21(2):271–92
21. Ferreira, LN., Ferreira,PL., Pereira,LN., Rowen, Donna and Brazier, John E. Exploring the Consistency of the SF-6D, *Value in Health*, 2013, 16 (6), 1023 - 1031 https://doi.org/10.1016/j.jval.2013.06.018.
22. Sintonen, Harri: The 15D instrument of health-related quality of life: properties and applications, *Annals of Medicine*, 2001; 33:5, 328-336, DOI: 10.3109/07853890109002086
23. Chakrabartty, S. N.: Alternate method of scoring Euroqol Five- Dimensional Scales, Health Science Journal, 2021, Vol (15), No. 6 :847
24. Chakrabartty, Satyendra Nath: Limitations of insomnia severity index and possible remedies. *JSM Neurol Disorders Stroke*, 2019; 5: 1–9.