

## Problems of Rating Scales in Health Measurements

Satyendra Nath Chakrabartty

Indian Ports Association, Indian Statistical Institute, Indian Maritime University.

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

Received date: November 05, 2024; Accepted date: February 13, 2025; Published date: February 20, 2025

Citation: Satyendra N. Chakrabartty, (2025), Problems of Rating Scales in Health Measurements, *Psychology and Mental Health Care*, 9(1): DOI:10.31579/2637-8892/314

Copyright: © 2025, Satyendra Nath Chakrabartty. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Patient-reported outcomes (PROs) using multi-item rating scales are not comparable due to different formats of the scales, different factors under consideration, etc.

**Objectives:** To discuss methodological limitations of PROs in health measurement and to provide a method for converting ordinal item-score of PROs to follow normal distribution.

**Method:** Converting raw item-score to equidistant score (E) followed by standardization to Z-scores  $\sim N(0,1)$  and converting Z-scores to proposed scores ( $P_i$ ) in the range 1 to 100. Scale scores ( $P_{Scale}$ ) as sum of  $P_i$ 's and battery scores (B-scores) as sum of scale scores follow normal distributions.

**Results:** Each of  $P_{Scale}$ -scores and B-scores satisfy desired properties, helps undertaking parametric analysis, facilitates comparing status and finding equivalent scores of two PROs having implications in classification and also to get reliability, validity in better fashion.

**Conclusion:** The suggested method contributing to improve scoring of PRO instruments with additional benefits of identification of poorly performing scales, assessment of progress across time is recommended.

**Keywords:** patient-reported scale; linear transformation; normal distribution; ability to detect changes; elasticity

### Introduction

Often subjective self-reported measures of illness are evaluated through rating scales to assess objective health (Bourne, 2009). Data resulting from such rating scales are categorical and in ordinal level. Large numbers of clinical researches use patient reported rating scales (PROs) to quantify clinical conditions like intensity of disease, effects of disease or treatment, health status, quality of life (QoL), pain, sleep disorders, depression, anxiety, stress and far beyond as part of the patient decision making process. The MAPI Research Trust, a nonprofit organization provides information for all stakeholders in the field of Patient Centered Outcomes, particularly for **Clinical Outcome Assessments (COAs)** (<https://c-path.org/programs/proc/>).

PROs consist of number of scales which vary in terms of features of the scales like number of items (scale length), number of levels (scale width), scoring methods, etc. and are not comparable. Scale length, scale width, frequencies of levels affect differential item functioning (DIF). Analysis of ordinal data emerging from PROs without satisfying the assumptions of statistical techniques used, may distort the results. Mokkink et al. (2010) suggested prior checking of measurement properties of PRO-instruments. Self-reported rating scale consisting of multi-point items

suffers from methodological limitations including not meaningful addition. If addition is not meaningful, computations like standard deviation (SD), correlation, Cronbach  $\alpha$ , etc. are meaningless. Statistical analysis like regression, Principal component analysis (PCA), Factor analysis (FA), testing equality of means by *t*-test or ANOVA assumes normal distribution of the variables under study. But, questionnaire scores with unknown distributions violate the assumption and may distort the results. Assigning equal importance to items and constituent scales in summative scoring of PROs is not justified since contributions of items or scales to total battery score, values of inter-item correlations, scale-battery correlations and factor loadings are different (Parkin et al. 2010). Mean, SD, Cronbach alpha tends to increase with increase in number of levels and items and may influence mean score more than the underlying variable (Lim, 2008). No consensus is there regarding number of levels per item in rating scales (Chakrabartty and Gupta, 2016). Studies attempting to evaluate effect of selenium supplementation on stroke used different definitions of stroke either by categorical variables or variables in ratio scale. While investigating dose-response correlation between dietary selenium intake and stroke risk, Shi et al. (2022) used self-reported single question "Has a doctor ever told you that you had a stroke?" to

define Stroke. Thus, stroke was taken here as a categorical variable and not in ratio scale. Zhang et al. (2023) asked each participant whether a doctor ever given a diagnosis of stroke (no, yes, unknown) and defined stroke as a self-reported physician diagnosis during follow-up. The follow-up time was the date of the first discovery of stroke. Sharifi-Razavi et al. (2022) included adults with accepted ischemic stroke by neuroimaging during the last 72hrs with a volume of at least one-third of middle *cerebral artery (MCA) territory* which is the most commonly affected territory in a cerebral infarction. Different inclusion criteria for stroke and different analysis resulted in different relationships between intake of selenium supplementation on stroke and conclusions and thus, effect of dietary selenium intake on stroke risk remains controversial. Beneficial effects of Selenium, on stroke risk have been found (Xiao et al. 2019; Hu et al. 2019). However, selenium at high levels is toxic (Hadrup & Ravn-Haren, 2020). Nutritional Prevention of Cancer (NPC) trial showed no benefit of selenium supplementation on the risk of stroke (Ding & Zhang, 2021; Shi et al. 2021). One possible reason of such differences could be consideration of benefits of circulating selenium levels and not on quantity of selenium intake, which probably had U-shaped relationship with stroke risks (Tan et al. 2021).

The paper suggests a method of transforming ordinal scores of  $i$ -th item of a PRO to normally distributed proposed scores ( $P_i$ -scores) facilitating meaningful addition and deriving scale score ( $P_{Scale}$ ) as arithmetic aggregation of  $P_i$ -scores satisfying desired properties, enabling assessment of progress and parametric analysis.

## 2. Problems Of Rating Scales:

If distance between two successive response-categories or levels of  $K$ -point items ( $K = 2, 3, 4, 5 \dots$ ) is denoted by  $d_{j,(j+1)}$  then  $d_{j,(j+1)} \neq d_{(j+1),(j+2)} \forall j = 1, 2, 3, 4 \dots$  i.e. scores are not equidistant (Rutter and Brown 2017). Thus, addition of ordinal item scores are not meaningful (Jamieson, 2004) and even  $\bar{X} >$  or  $< \bar{Y}$  is meaningless (Hand 1996). Despite this limitation, an individual score is taken as sum of item scores in ordinal scale (Kyte et al. 2015). Meaningful addition of two random variables,  $X + Y = Z$  requires similar probability distribution of  $X$  and  $Y$  and known distribution of  $Z$  for further uses. In terms of probability,  $X + Y = Z$  implies

$$P(Z = z) = P(X = x, Y = z - x) \text{ for discrete case and } P(Z \leq z) = P(X + Y \leq z) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^z f_{X,Y}(x, t - x) dt \right) dx \text{ for continuous case.}$$

Thus, knowledge of probability density function (pdf) of  $X$  and  $Y$  and their convolution are necessary. However, if each of  $X$  and  $Y$  follow log-normal distribution,  $X + Y$  cannot be obtained as such and require complex Lie-Trotter operator splitting method (Lo, 2012). Generic or disease-specific multidimensional rating scales for QoL may not consider all relevant constructs. For example, Disease-specific stroke adapted 30-item SIP version (SA-SIP30) with 8 subscales excludes domains like recreation, energy, pain, general health perceptions, overall quality of life or stroke symptoms (Golomb et al. 2001). Multidimensional rating scales may even fail to give global summary like 36-Item Short Form Health Survey questionnaire (SF-36)

(<http://www.webcitation.org/6cfeefPkf>). Multidimensional scale covers a number of sub-scales/dimensions where scale formats are different for different sub-scales. For example, SF-36 has 10 items (3-points) on Physical functioning, 3 items, each 6-point on Energy/Fatigue, 2 items on 5-point scale for Social functioning, 6 (6-points) items on Emotional well-being, 5 (5-point) items on General health, two items on Pain (one 6-point and one 5-point), seven binary items and another item regarding reported health transition over the last year. The set-up indicates (i) different distributions for binary items, 3-point, 5-point, 6-point items, (ii) higher mean, SD of sub-class containing 6-point items, (iii) different reliability,

validity, for different sub-classes (Preston and Colman, 2000). Two distinct concepts measured by the SF-36 are Physical Component Summary (PCS), and Mental Component Summary (MCS). Taft et al. (2001) found paradoxical inverse relationship between PCS and MCS which implies well physical condition pre-supposes poor mental health and vice versa. SF-36 was found to be negatively correlated with Patient Health Questionnaire (PHQ) and General anxiety disorder questionnaire (GAD-7) probably due to different factors measured by them (Johnson et al. 2019). Scoring methods of PROs are different. Dimension score of MacNew Heart Disease Health-Related Quality of Life Questionnaire (MacNew) is based on mean of the responses in items belonging to the dimension but, Cardiovascular Limitations and Symptoms Profile (CLASP) scores consider weights to find total for each subscale. Each dimension of Myocardial Infarction Dimensional Assessment Scale (MIDAS) is scored separately. No clear understanding of factors being measured is there. Against two factors proposed in the Hospital Anxiety and Depression Scale (HADS), factor structure of the instrument were found to be three in a range of clinical populations (Caci et al. 2003) against recommending HADS as a one-dimensional measure (Costantini et al. 1999) and statistical evidence for a three-factor structure (Strong et al. 2007). Similarly, for Psychological General Well-Being Index (PGWBI), Lundgren-Nilsson et al. (2013) found single construct of psychological wellbeing against underlying six factors of the scale raising questions about factor analytic interpretation in the presence of local dependency. Use of zero as an anchor value does not allow computation of expected value (value of the variable  $\times$  probability of that value). Responses to zero, reduces mean and SD of the scale, item-total correlations, affects regression or logistic regression, etc. If each respondent of a sub-group selects the level marked as "0" to an item then mean = variance = 0 for the sub-group for that item and correlation with that item is undefined. Stucki et al. (1995) found more than 40% of the patients scored zero in 10 subscales of Sickness Impact Profile (SIP) and in one subclass of SF-36. Better is to mark the anchor values as 1, 2, 3... and so on, keeping the convention of higher score  $\Leftrightarrow$  higher value of the variable being measured. Higher score in each of Nottingham Health Profile (NHP), Minnesota Living with Heart Failure (MLHF) indicate higher health problems, unlike Sickness impact Profile (SIP). Thus, directions of scores are different for different scales. Rating data with floor and ceiling effects follow unknown distribution and do not satisfy the assumption of PCA like bivariate normality for each pair of observed variables, normally distributed scores, etc. Test reliability by Cronbach alpha assumes one-dimensional scale and tau-equivalence (equality of all factor loadings). Multidimensional PROs like Insomnia Severity Index (ISI), Pittsburgh Sleep Quality Index (PSQI) and Insomnia Symptom Questionnaire (ISQ), etc. violate the assumption and underestimate the coefficient alpha (Daniel, 1990). The coefficient alpha is influenced by variance sources, sampling errors (Terry & Kelley, 2012), sample size (Charter, 1999) and even test length and test width (Luh, 2024). Validity of a multidimensional scale as correlation with criterion scores raises the question about the dimension /factor being reflected by the validity. It is desirable to find the validity of the main factor for which the scale was developed and also to derive relationship between test reliability and test validity. Vaughan, (1998) found lower validity where data contained predominantly high performers. To avoid such problems, structural validity of normally distributed transformed scores by PCA was preferred (Chakrabarty, 2020). Different cut-off scores are there for different PROs. For example, cut-off score of Sickness Impact Profile (SIP136) with 136 "Yes-No" type items distributed over 12 domains is  $\geq 22$  and for Stroke-Adapted Sickness Impact Profile (SA-SIP30) with 30 items covering 8 subscales is  $\geq 33$ . Natural question is whether score of 33 in SA-SIP30 is equivalent to the score of 22 in SIP136. Similarly, score of 14 in ISI indicating "no insomnia" is

equivalent to which score in *PSQI* or *ISQ*? Thus, finding equivalent scores of two scales can make better comparisons of the PROs and also help in classification of individuals. For QoL questionnaires, there could be no cut-off point to show better or worse QoL (Silva *et al.* (2014). Based on treatment status for Cancer Core Questionnaire(*EORTC QLQ-C30*), four different cut-off scores were found (Lidington *et al.* 2022). Intra- and-inter observer reliability of ordinal scale like Kessler Psychological Distress Scales (K 6 and K 10) are evaluated by Kappa and weighted Kappa. Major limitations in this context are:

- A low kappa does not imply low agreement (Bajpai *et al.* 2015). Confidence interval for Kappa  $\leq 0.60$  may indicate large volume of incorrect evaluation of data (Simundic, 2008).
- Methods of deciding weights for weighted kappa vary and may give different values of weighted kappa.
- Concepts of agreement in terms of  $\kappa$  or  $\kappa_{Weighted}$  are different from the concept of reliability of tests/scales.

### 3. Suggested method:

Let  $X_{ij}$  be the raw score of a respondent in the  $i$ -th item for choosing the  $j$ -th level where the levels are marked as 1, 2, 3, 4, ... avoiding zero and higher value of  $X_{ij}$  implies higher dysfunctions or impairments. The suggested method transforms ordinal item scores ( $X_i$ ) to equidistant scores ( $E_i$ ) and further transformation to proposed scores ( $P_i$ -scores) in the score range [1, 100] following  $(\mu_i, \sigma_i)$  facilitating meaningful addition to derive scale score ( $P_{Scale}$ ) as sum of  $P_i$ -scores. The method is described below.

For the  $i$ -th item, find maximum frequency  $f_{i,Max}$  and minimum frequency  $f_{i,Min}$ . For  $n$ -number of respondents in a 5-point item (say), find initial weights  $\omega_{i1} = \frac{f_{i,Min}}{n}$ , the common difference  $\alpha = \frac{5f_{i,Max} - f_{i,Min}}{4n}$  and other initial weights as  $\omega_{i2} = \frac{\omega_{i1} + \alpha}{2}$ ,  $\omega_{i3} = \frac{\omega_{i1} + 2\alpha}{3}$ ,  $\omega_{i4} = \frac{\omega_{i1} + 3\alpha}{4}$ , and  $\omega_{i5} = \frac{\omega_{i1} + 4\alpha}{5}$ .

Take final weights  $W_{ij} = \frac{\omega_{ij}}{\sum_{j=1}^5 \omega_{ij}}$  Here,  $\sum_{j=1}^5 W_{ij} = 1$ . Here,  $W_{ij}$ 's form an arithmetic progression. Generated scores  $E_{ij} = W_{ij}X_{ij}$  are continuous, monotonic and equidistant.

Standardized equidistant scores ( $E$ ) of each item as  $Z = \frac{E - \bar{E}}{SD(E)} \sim N(0, 1)$  and

$$P_i = \frac{(99) * (Z_i - \text{Min}(Z_i))}{\text{Max}(Z_i) - \text{Min}(Z_i)} + 1 \sim N(\mu_i, \sigma_i) \text{ where } 0 \leq P_i \leq 100 \text{ irrespective of length of scale and width of items.}$$

Normality of item scores ( $P_i$ 's) facilitates meaningful addition and the resultant scale scores  $P_{Scale} = \sum_i P_i$  as the convolution of  $P_i$ 's. Normally distributed  $P_{Scale}$  -scores can be added to get battery score ( $B$ -scores) also following normal.

Major properties of  $P_{Scale}$ -scores and  $B$ -scores are:

- Each avoids equal importance to items and dimensions and represents continuous, monotonically increasing scores.
- The zero point for scoring  $K$ -point items to get  $E$ -scores is obtained when  $f_{ij} = 0$ . Other items in ratio scales can be standardized and transformed to follow normal distribution in the range [1, 100] and added with  $P_i$ 's
- Contribution of  $j$ -th scale to the battery can be found by  $\frac{P_{jth\ scale\ score}}{B\text{-scores}}$ .

### 4. Benefits:

Parameters of distributions of  $P_{Scale}$ -scores and  $B$ -scores can be estimated from data. Normality enables estimation of population mean ( $\mu$ ), population variance ( $\sigma^2$ ), confidence interval of  $\mu$ , testing statistical hypothesis like  $H_0: \mu_1 = \mu_2$  or  $H_0: \sigma_1^2 = \sigma_2^2$  etc.

Based on battery scores, progress of  $i$ -th patient in  $t$ -th period over the previous period by  $\frac{B_{i(t)} - B_{i(t-1)}}{B_{i(t-1)}} \times 100$ . Decline is indicated in case of  $B_{i(t)} - B_{i(t-1)} < 0$ . For a group of patients,  $\overline{B_{i(t)}} > \overline{B_{i(t-1)}}$  indicates progress. Similarly, progress with respect to scores of  $P_{Scale}$  can be computed. Decline if any, may be probed to find the critical scale(s) where  $P_{Scale(t)} - P_{Scale(t-1)} < 0$  and initiate appropriate corrective actions in the treatment and management plan. Statistical test of significance of progress/deterioration can be made since ratio of two normally distributed variable follows  $\chi^2$  distribution.

Effect of small change in  $i$ -th scale ( $P_{i-th\ scale}$ ) to Battery score  $B$ -scores can be quantified by considering elasticity i.e. percentage change of  $B$ -scores due to small change in  $P_{i-th\ scale}$ . The scales can be ranked based on such elasticity. Elasticity studies in economics, reliability engineering, consider model like  $\log Q_{jt} = \alpha_j + \beta_j \log P_{jt}$  where  $Q_{jt}$  denotes the quantity demanded of  $j$ -th industry at time  $t$  and  $P_{jt}$  is industry price relative to the price index of the economy. However, for normally distributed  $P_{Scale}$ -scores and  $B$ -scores, logarithmic transformations are not required to fit regression equation of the form  $P_{Scale} = \alpha_i + \beta_i P_i + \epsilon_i$ . The coefficient  $\beta_i$  reflects the impact of a unit change in the independent variable ( $i$ -th dimension) on the dependent variable ( $P_{Scale}$ ). Policy makers can decide appropriate actions in terms of continuation of efforts towards the scales with high values of elasticity and corrective actions for the dimensions with lower elasticity i.e. areas of concern.

Normality of  $B$ -scores facilitates testing  $H_0: \mu_{B_t} = \mu_{B_{(t-1)}}$  reflecting effectiveness of the treatment plans and  $H_0: \text{Progress}_{(t+1)\text{ over } t} = 0$ , reflecting progression

Graph depicting progress/decline of one patient or a group of patients with similar socio-demographic profile is analogous to hazard function and helps to identify high-risk groups and compare response to treatments from the start.

For two scales  $X$  and  $Y$  with normal pdf  $f(x)$  and  $g(y)$  respectively, equivalent score  $y_0$  for a given value say  $x_0$  can be found by solving the equation  $\int_{-\infty}^{x_0} f(x) dx = \int_{-\infty}^{y_0} g(y) dy$  using standard normal table even if the scales have different lengths and widths (Chakrabarty, 2021).

$P$ -scores and  $B$ -scores following normal distributions satisfy the assumptions of PCA, FA and enable finding Factorial (FV)  $= \frac{\lambda_1}{\sum \lambda_i} = \frac{\lambda_1}{\sum S_{X_i}^2}$  where  $\lambda_1$  the highest eigenvalue indicating validity for the main factor being measured (Parkerson *et al.* 2013). The test significance of  $\lambda_1$  can be undertaken using the Tracy–Widom (TW) test statistic  $U = \frac{\lambda_1}{\sum \lambda_i}$  following TW-distribution (Nadler, 2011). Such FV avoids the problems of construct validity and selection of criterion scale ensuring matching constructs and two administrations of the scale and the criterion scale.

For standardized item scores,  $FV_{Z\text{-scores}}$  of a test with  $m$ -items is  $\frac{\lambda_1}{m}$  and the test variance  $S_X^2$  can be written as  $S_X^2 = \sum \lambda_i + 2 \sum_{i \neq j=1}^m Cov(X_i, X_j) = \frac{\lambda_1}{FV} + 2 \sum_{i \neq j=1}^m Cov(X_i, X_j)$  (1)

The equation (1) can be used to find the theoretical reliability

$$r_{tt(\text{theoretical})} = \frac{S_x^2}{S_x^2} = \frac{S_x^2}{FV + 2 \sum_{i \neq j}^m \text{Cov}(X_i, X_j)} \quad (2)$$

Equation (2) gives relationship between  $r_{tt(\text{theoretical})}$  and factorial validity, which is non-linear.

Ten Berge and Hofstee (1999) suggested maximum reliability of a test by  $\alpha_{PCA}$  which can be derived from the correlation matrix of  $m$ -number of items by

$$\alpha_{PCA} = \left( \frac{m}{m-1} \right) \left( 1 - \frac{1}{\lambda_1} \right) \quad (3)$$

$$\text{Relationship between FV and } \alpha_{PCA} \text{ can be derived as: } \alpha_{PCA} = \left( \frac{m}{m-1} \right) \left( 1 - \frac{1}{\lambda_1} \right) = \left( \frac{m}{m-1} \right) \left( 1 - \frac{1}{FV \cdot \sum \lambda_i} \right) = \left( \frac{m}{m-1} \right) \left( 1 - \frac{1}{m \cdot FV_{Z\text{-scores}}} \right) \quad (4)$$

As per (4), higher value of  $FV_{Z\text{-scores}}$  increases  $\alpha_{PCA}$

$$\text{Cronbach alpha of a battery consisting of } K\text{-scales can be obtained as a function of scale reliabilities by } \hat{\alpha}_{\text{Battery}} = \frac{\sum_{i=1}^K r_{tt(i)} S_{X_i} + \sum_{i=1, i \neq j}^K \sum_{j=1}^K 2 \text{COV}(X_i, X_j)}{\sum_{i=1}^K S_{X_i} + \sum_{i=1, i \neq j}^K \sum_{j=1}^K 2 \text{COV}(X_i, X_j)} \quad (5)$$

where  $r_{tt(i)}$  and  $S_{X_i}$  denote respectively reliability and SD of the  $i$ -th scale.

## 5. Discussion:

The suggested method defines meaningful scale scores and battery scores for each individual. Each of  $P_{\text{Scale}}$ -scores and  $B$ -scores satisfy desired properties, helps undertaking parametric analysis, comparing status and progression of patients including indication of effectiveness of treatment plans, finding equivalent scores of two patient reported scales (PROs) where area under normal curve corresponding to PRO-1 up to  $P_{\text{PRO-1}}^0$  = area under normal curve corresponding to PRO-2 up to  $P_{\text{PRO-2}}^0$ . For classification of individuals, equivalent cut-off scores of class boundaries may be found satisfying  $\frac{\text{Var. of group}_{\text{Score} \geq P_{\text{PRO-1}}^0}}{\text{Variance of PRO-1}} = \frac{\text{Var. of group}_{\text{Score} \geq P_{\text{PRO-2}}^0}}{\text{Variance of PRO-2}}$  which may facilitate to have similar efficiency of classification, in terms of within group variance and between group variance. Factorial validity (FV) reflecting the main factor being measured helps to have a clear understanding of the most important factor being measured. However, establishing clinically meaningful content validity is a vital step. Maximum value of test reliability  $\alpha_{PCA}$ , relationship between  $FV_{Z\text{-scores}}$  and  $\alpha_{PCA}$  and also between  $r_{tt(\text{theoretical})}$  and FV can be used effectively to compare scales. The scales with eigenvalues exceeding unity can be retained keeping in view that results may get distorted by wrong selection of constituent scales.

## 6. Conclusions:

The suggested  $B$ -scores reflecting disease severity with respect to the PRO measures is recommended with the scales chosen as per the selection criteria mentioned above. Future empirical investigations may be undertaken to evaluate properties of the suggested method and its clinical validation along with effects of socio-demographic factors.

## Declarations:

*Acknowledgement:* Nil

**Conflicts Of Interest/Competing Interests:** *The author has no conflicts of interest to declare*

*Funding:* Did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

*Ethical approval:* Not applicable since the paper does not involve human participants.

*Consent of the participants:* Not applicable since the paper does not involve data from human participants

**Data Availability Statement:** *The paper did not use any datasets*

**Code availability:** *No application of software package or custom code*

*CRedit statement:* Conceptualization; Methodology; Analysis; Writing and editing the paper by the **Sole Author**

## References:

- Bajpai, S., Bajpai, R. and Chaturvedi, HK. (2015). Evaluation of Inter-Rater Agreement and Inter-Rater Reliability for Observational Data: An Overview of Concepts and Methods, *Journal of the Indian Academy of Applied Psychology*, 41(3), 20-27
- Bourne, PA. (2009). The validity of using self-reported illness to measure objective health. *N Am J Med Sci.*; 1(5):232–238.
- Caci LA, Bayle FJ, Mattel V, Dossios C, Robert P, Boyer P. (2003). How does the Hospital Anxiety and Depression Scale measure anxiety and depression in healthy subjects? *Psychiatry Report*; 118: 89–99.
- Chakrabarty, Satyendra Nath (2021). Integration of various scales for Measurement of Insomnia. *Research Methods in Medicine & Health Sciences*, 2(3), 102-111.
- Chakrabarty, S. N. (2020). Improved Quality of Pain Measurement, *Health Science*, Vol. 1, 1 -6.
- Chakrabarty, SN and Gupta, R. (2016). Test Validity and Number of Response Categories: A Case of Bullying Scale, *Journal of the Indian Academy of Applied Psychology*; 42(2); 344-353
- Charter RA. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21(4), 559–566. <https://doi.org/10.1076/jcen.21.4.559.889>
- Costantini M, Musso M, Viterbori P. et al. (1999). Detecting psychological distress in cancer patients: validity of the Italian version of the Hospital Anxiety and Depression Scale. *Support Care Cancer*; 7: 121-127
- Daniel, Wayne W. (1990): *Friedman two-way analysis of variance by ranks. Applied Nonparametric Statistics (2nd ed.)*. Boston: PWS-Kent. 262–274.
- Ding J. and Zhang Y. (2021). Relationship between the circulating selenium level and stroke: a Meta-analysis of observational studies. *J Am Coll Nutr.*; 1–9.
- Golomb BA, Vickrey BG, Hays RD. (2001). A review of health-related quality-of-life measures in stroke. *Pharmacoeconomics*; 19(2):155-85
- Hadrup N, Ravn-Haren G. (2020). Acute human toxicity and mortality after selenium ingestion: a review. *J Trace Elem Med Biol.*; 58:126435
- Hand, D. J. (1996). Statistics and the Theory of Measurement, *J. R. Statist. Soc. A*; 159, Part 3, 445-492.
- Hu XF, Stranges S, Chan L. (2019). Circulating selenium concentration is inversely associated with the prevalence of stroke: results from the Canadian Health Measures Survey and the National Health and Nutrition Examination Survey. *J Am Heart Assoc.*; 8(10): e012290.



15. Jamieson, S. (2004). Likert scales: How to (ab) use them. *Medical Education*, 38, 1212-12
16. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A.(2019). Psychometric Properties of the General Anxiety Disorder 7-Item (GAD-7) Scale in a Heterogeneous Psychiatric Sample. *Front Psychol*;10:1713.
17. Kyte DG, Calvert M, Vander Wees PJ, Ten Hove R, Tolan S, Hill JC. (2015). An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*; 101 (2); 119-125
18. Lidington E, Giesinger JM, Janssen SHM, Tang S, Beardsworth S, Darlington AS *et al.* (2022). Identifying health-related quality of life cut-off scores that indicate the need for supportive care in young adults with cancer. *Qual Life Res.* 31, 2717–27.
19. Lim HE.(2008). The use of different happiness rating scales: bias and comparison problem? *Social Indicators Research*, 87, 259–267.
20. Lo CF (2012). The Sum and Difference of Two Lognormal Random Variables. *Journal of Applied Mathematics*, Article ID 838397,
21. Luh, Wei-Ming (2024). A General Framework for Planning the Number of Items/Subjects for Evaluating Cronbach's Alpha: Integration of Hypothesis Testing and Confidence Intervals. *Methodology*; Vol. 20(1), 1–21,
22. Lundgren-Nilsson, Å., Jonsdottir, I.H., Ahlborg, G. *et al.* (2013). Construct validity of the psychological general wellbeing index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a rasch analysis. *Health Qual Life Outcomes* 11.
23. Mokkink LB, Terwee CB, Patrick DL, *et al.* (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 19, 539–549.
24. Nadler, Boaz (2011): On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. *Journal of Multivariate Analysis*, 102; 363-371.
25. Parkerson, H. A., Noel, M., Gabrielle M. P., Fuss, S., Katz, J., Gordon J. G. Asmundson (2013): Factorial Validity of the English-Language Version of the Pain Catastrophizing Scale–Child Version, *The Journal of Pain*, 14(11),1383-1389.
26. 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
27. Preston, Carolyn C. and Colman, A. M. (2000): Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences, *Acta Psychologica* 104, 1-15
28. Rutter L. A. and Brown T. A. (2017): Psychometric properties of the generalized anxiety disorder scale-7 (GAD-7) in outpatients with anxiety and mood disorders. *J. Psychopathol. Behav. Assess.* 39 140–146.
29. Sharifi-Razavi A, Karimi N, Jafarpour H. (2022). Evaluation of Selenium Supplementation in Acute Ischemic Stroke Outcome: An Outcome Assessor Blind, Randomized, Placebo-Controlled, Feasibility Study. *Neurol India*;70(1):87-93.
30. Shi W, Su L, Wang J, Wang F, Liu X and Dou J. (2022). Correlation between dietary selenium intake and stroke in the National Health and Nutrition Examination Survey 2003–2018, *Annals of Medicine*, 54 (1), 1395–1402.
31. Shi L, Yuan Y, Xiao Y, *et al.*(2021). Associations of plasma metal concentrations with the risks of all-cause and cardiovascular disease mortality in Chinese adults. *Environ Int.*;157:106808.
32. Silva PAB, Soares SM, Santos JFG, *et al.*(2014). Cut-off point for WHOQOL-bref as a measure of quality of life of older adults. *Rev Saude Publica*; 48:390–397.
33. Simundic AM.(2008). Confidence interval. *Biochem Med*, 18:154–161.
34. Strong V, Waters R, Hibberd C, *et al.*(2007). Emotional distress in cancer patients: the Edinburgh Cancer Centre symptom study.*Br J Cancer*; 96: 868-874.
35. Stucki G, Liang MH, Phillips C, Katz JN. (1995): The Short Form-36 is preferable to the SIPas a generic health status measure in patients undergoing elective total hip arthroplasty. *Arthritis Care Res.*; 8(3):174-181.
36. Taft, Charles & Karlsson, Jan & Sullivan, Marianne. (2001). Do SF-36 Summary Component Scores Accurately Summarize Subscale Scores? *Quality of life research*, 10(5), 395-404.
37. Tan QH, Huang YQ, Liu XC, Liu L, Lo K, Chen JY and Feng YQ (2021) A U-Shaped Relationship Between Selenium Concentrations and All-Cause or Cardiovascular Mortality Patients With Hypertension. *Front. Cardiovasc. Med.* 8:671618.
38. Ten Berge JMF & Hofstee WK (1999): Coefficients alpha and reliabilities of unrotated and rotated components. *Psychometrika*, 64, 83–90.
39. Terry L, & Kelley K. (2012). Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals. *British Journal of Mathematical & Statistical Psychology*, 65(3), 371–401.
40. Xiao Y, Yuan Y, Liu Y, *et al.* (2019). Circulating multiple metals and incident stroke in Chinese adults. *Stroke*, 50(7):1661–1668.
41. Zhang H, Qiu H, Wang S and Zhang Y (2023). Association of habitually low intake of dietary selenium with new-onset stroke: A retrospective cohort study (2004–2015 China Health and Nutrition Survey). *Front. Public Health* 10:1115908.



This work is licensed under Creative Commons Attribution 4.0 License

To Submit Your Article Click Here:

**Submit Manuscript**

**DOI:10.31579/2637-8892/314**

**Ready to submit your research? Choose Auctores and benefit from:**

- fast, convenient online submission
- rigorous peer review by experienced research in your field
- rapid publication on acceptance
- authors retain copyrights
- unique DOI for all articles
- immediate, unrestricted online access

At Auctores, research is always in progress.

Learn more <https://auctoresonline.org/journals/psychology-and-mental-health-care>