Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2023;9(1):1-10

Original Article

A Modification on Intra Class Correlation Estimation for Ordinal Scale Variable Using Latent Variable Model

Samira Chaibakhsh[1*], Asma Pourhoseingholi[2]

[1]Eye Research Center, the Five Senses Institute, Rassoul Akram Hospital, Iran University of Medical Sciences, Tehran, Iran.
[2]Prevention of Cardiovascular Disease, Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

ARTICLE INFO

Received 01.10.2022
Revised 01.12.2022
Accepted 10.01.2023
Published 15.03.2023

Key words:
Intraclass correlation;
Reliability;
Latent variable;
Multivariate model;
Test-retest

ABSTRACT

Introduction: A common way for computing test-retest reliability is Intra Class Correlation which was developed for continuous variables. But it widely used to assess test-retest reliability in questionnaires with Likert scales. Most of the time consecutive numbers regarded as option labels of a question. If the probability of choosing options be the same, using this method is logic, otherwise it is not. Therefore, in this study a modified estimator of ICC is proposed to improve the estimation of ICC for ordinal scale by using latent variable model.

Methods: In this method test-retest answers were considered as bivariate variables and cumulative Probit latent variable model was fitted. A simulation study with N=1500 replicates was conducted to compare the ICC estimations of Likert scale approach with a latent variable approach. Different sample sizes (n=20, 30) was generated with different correlation parameters. The simulations were repeated for questions with 3 and 5 options with different probability of selecting options of a question. After that the two approaches were run on Beck for suicidal ideation questionnaire.

Results: In general the difference between Likert scale approach and latent variable approach were higher in 3 question options compared to 5 and also by increasing sample size and correlation between bivariate data, Root Mean Square Errors and bias were decreased. Assuming different probabilities for options, there was a considerably difference between Root Mean Square Errors, bias and standard deviation of estimation of ICC in two models. Using latent variable approach resulted less bias, SD and Root Mean Square Errors especially in lower sample sizes.

Conclusion: Simulations showed when the probability of choosing options of a question are skewed, using this method reduced Root Mean Square Errors especially when the options are less. This method was affected more on standard deviation compare to bias of estimations.

*.Corresponding Author: smr.chaibakhsh@gmail.com

Introduction

Accuracy and consistency of measurements play an important role in evaluating observations (Mehta et al., 2018; Yen & Lo, 2002). Although there is not an instrument with complete, precise measurements, but the errors should minimize to accomplish reliable results (Barnhart, Haber, & Lin, 2007).

One of the components of the reliability is repeatability which is known as test-retest reliability. Test-retest shows that how close are the measurements of a sample unit in similar situations. Thus, the more agreement between measurements of a unit in similar situations indicate higher reliability (Metrology, 2008). In this context, one of the most famous tools for assessing test-retest consistency between quantitative variables is intra class correlations (ICC) which is commonly used to evaluate test-retest reliability of a questionnaire(Ren, Yang, & Lai, 2006) .

ICC was first introduced by Fisher in 1925 and after that it was used widely to test the agreement between similar situations and raters (Chen & Barnhart, 2008). After proposing ICC by Fisher, this index expanded by other researchers in accordance with The data (Bartko, 1966; McGraw & Wong, 1996; Shrout & Fleiss, 1979). So depend on the problem, different version of ICCs can be used (Liljequist et al., 2019). For example, Qin et al. In 2019 recommended that when test-retest is assessed in time points, using two-way ANOVA model is more appropriate. Using one-way model would under estimate ICC because tie is a design factor (Qin et al., 2019). Shan G, in 2020 improved the ICC estimation in the cluster data for situation with non-normal distribution(shan. et. Al., 2020). The most ordinary ICC was defined as a one-way ANOVA random effect model, including random error and random effect of subjects. This ICC is useful when sample unit measurements are replicated. Thus, it can be used for calculating test-retest reliability of the questionnaires. (Ren et al., 2006). That is why it has been used for assessing questionnaire psychometrics in many researches.

However, ICC refers to continuous variables, but it commonly utilizes for assessing test retest reliability in Likert scale questionnaires. In these questionnaires, the options of questions are ordinal. One of the important differences between continuous and ordinal scales is that unlike continuous scale, in ordinal scale the distance between categories are not necessary to be the same (Agresti, 2013).Thus, it is obvious that the analysis of ordinal variables is different with continuous ones and it seems that if the ordinal variable has categories with unequal distances, choosing an appropriate analysis that is specific for analyzing ordinal data becomes more important.

In recent decades, many approaches have proposed to analyze ordinal data. One of these approaches is a latent variable model. In this model it is assumed that ordinal variable has an underlyng continuous distribution. This approach was proposed by McCullagh in 1980(McCullagh, 1980) for the first time. After that in 1995 kim (Kim, 1995) extended the model for correlated outcomes and utilized this model for modelling paired organs. Then this model became more popular, so it was extended and utilized by researchers to analyze bivariate outcomes. For example, Catalano used this model for correlated continuous and ordinal outcomes. Also Todem et al. (Todem, Kim, & Lesaffre, 2007) utilized this model for bivariate ordinal data with repeated outcomes.

In this model a random variable was added to the model to control the correlation between repeated outcomes. The Bayesian approach of kim model was performed by Biswas and Dos in 2002(Biswas & Das, 2002). It would be better to notice that in all models the underlying latent variable distribution was bivariate normal distribution.As mentioned before, the most common index that is used to assess the test retest reliability is the ICC that is the index of evaluating the agreement of quantitative variables. Since one of the basic properties of ordinal variables is that the distance between categories of the variables are unknown (unlike quantitative variables), so it is expected that using this index for computing test-retest reliability in Likert scale questionnaire specially when the probability of selecting options of a question are not the same and they are skewed to one of the sides of positive or negative positions, is not adequate.

Method: In this context, this research aimed to propose a modified ICC based on a latent variable approach for assessing test retest reliability in Likert scale questioners. In this proposed approach instead of using consecutive numbers as labels of options of a question (common method) we want to modify the distance between option labels by using the probability of selecting each option.

Method

Definitions and notations

Suppose that Y_{ij} is the response of the i[th] subject (i=1,...,N) and j=1,2 shows the time (1: test, 2: retest). Thus $Y_{ij} = (Y_{i1}, Y_{i2})$ is the bivariate response vector of the subject i on question j. the ICC can be defined in the form of ANOVA:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (1)$$

where , $\mu = E\left(Y_{ij}\right), \alpha_i$ is the random effect of the subject i which is i.i.d. with N(0, σ_α^2) and ε_{ij} is the random error which is i.i.d. with N(0, σ_ε^2) and also Y_{ij} is a continuous variable. So the variance of the Y_{ij} is $\sigma_\alpha^2 + \sigma_\varepsilon^2$. Thus:

$$ICC = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}. \quad (2)$$

By increasing σ_α^2 and decreasing σ_ε^2, ICC tends to 1 and ICC=1 shows the complete agreement. As mentioned before ICC can use for continuous variable, but in Likert scale questions we face ordinal scale. To solve this problem we can use latent variable models by assuming that has an underling bivariate continuous distribution.

latent variable model

Let Y_{ij}^* has a continuous distribution. In latent variable approach it is assumed that Y_{ij} =c because the latent variable Y_{ij}^* is between the specific interval (θ_{c-1} θ_c) and we cannot observe the amount of Y_{ij}^* but Y_{ij} is observable. Thus, if $Y_{ij} \in \{1, 2, \ldots, C\}$ there are C-1 cut-off points that define the intervals.

Now consider the vector ${}_i j^* = (Y_{i1}^*, Y_{i2}^*)$ as a bivariate latent variable model. Thus:

$$\begin{aligned} &\pi_{ab} = Pr(Y_{i1} = a, Y_{i2} = b) = \\ &\Pr\left(\theta_{a-1} < Y_{i1}^* < \theta_a, \theta_{b-1} < Y_{i2}^* < \theta_b\right) = \\ &F_\rho\left(\theta_a, \theta_b\right) - F_\rho\left(\theta_a, \theta_{b-1}\right) - \\ &F_\rho\left(\theta_{a-1}, \theta_b\right) + F_\rho(\theta_a, \theta_b) \end{aligned} \quad (3)$$

Where F_ρ is a bivariate distribution of the latent variable and ρ is the correlation parameter. If F is considered as a bivariate normal distribution

then:

$$\pi_{ab} = \psi_\rho(\theta_a, \theta_b) - \psi_\rho(\theta_a, \theta_{b-1}) - \psi_\rho(\theta_{a-1}, \theta_b) + \psi_\rho(\theta_a, \theta_b) \quad (4)$$

where ψ_ρ is the bivariate normal standard distribution.

modification on intra class correlation

In the original method, ICC is computed using the scores labeled to each option of a question. So Y_{ij} are the amounts that the researcher related to the options of a question. Most of the time labeled numbers are consecutive or have a same interval. In this study this approach (assuming the same interval between options) is named Likert approach. In most of the applied examples this assumption is not valid. So in this section we want to improve the labeled numbers of the options of a question by considering the rational interval between labeled numbers. In this study this approach is called latent variable because the latent variable model was used to control the option intervals.

Estimation of the cut-off points

By using latent variable models, cut-off points can be estimated using equation (3). Let $\gamma = (\rho, \theta)$ is the parameter vector. And the number of options of the question is q. The likelihood function is:

$$L(\gamma; X) = \sum_{i=1}^{n}\sum_{j=1}^{2} I_{i1} I_{i2} Ln(\pi_{ij}(\gamma, X)) \quad (5)$$

Where I_{ij} =1 if the jth response time in ith sample is j (Y_{ij}=m) and otherwise I_{ij}=0. For obtaining parameter estimations (ρ and q-1 cut-off points) $\partial L(\gamma)/\partial\gamma$ must be solved. The mvtnorm package in R software to calculate the joint normal probabilities and nlm function to estimate the parameters and heir standard errors. This function utilizes a Newton-type algorithm for estimation.

After solving (5), the Likert answers would substitute by the related cut-off point estimations. The difference between estimated cut-off points with Likert answers is that the interval between options will be modified so it seems that the estimated ICC become more efficient.

Simulation study

A simulation study with N=1500 replicates was conducted to compare the ICC estimations of Likert scale approach (LSA) with a latent variable approach (LVA). A sample with different sample sizes (n=20, 30) was generated from bivariate uniform distribution with different correlation parameters (0.6, 0.7, 0.8). Theses correlations were selected because the correlation below 0.5 is not practically remarkable (Maxwell, 1977). Also simulates repeated for questions with 3 and 5 options. These bivariate samples were considered as true test-retest answers. To distinguish values with each other, U_{ij} is considered as continuous true values from bivariate uniform distribution and LSA_{ij} , LVA_{ij} are ordinal responses related to U_{ij} which are obtained from LSA and LVA approaches. The estimated ICC from LSA and LVA were compared with the true ICC by standard deviation (SD), absolute bias and RMSE.

Likert scale approach

This approach is related to the common way for calculating ICC in Likert scale questionnaire. It means that for computing

ICC, the options of a question which are labeled as consecutive numbers are considered as Y_{ij} For generating consecutive numbers, q-1 (q=3,5 is the number of options of a question) samples were selected from the true values (U_{ij}). Then in accordance with sorted q samples, the bivariate data (were divided into q groups. For instance, if $U_{ijk} < q_{(1)}$ then LSA_{ij}=1. Thus the probability of choosing an option in large iterations will be the same. Then the ICC was computed for each sample using $LSA_{ij}=(LSA_{i1}, LSA_{i2})$. For considering different probability for choosing options, this time 2(q-1) samples were selected from and after sorting them, after sorting 2(q-1) samples the first q-1 samples were selected to transform in to ordinal data (LSA_{ij}). Thus the probability of the last option could be more than other options.

Latent variable approach

In this approach LSA_{ij} considered as bivariate response variable and as express in section 2.2 and then the cut-off points were estimated. The latent bivariate distribution was bivariate normal distribution. The estimated cut-off points were used to estimate the cumulative probability of choosing each option. After that, the points with the same cumulative probabilities were found for the uniform distribution and these points considered as $LVA_{ij}=(LVA_{i1},LVA_{i2})$ and ICC was calculated using LVA_{ij} .

Results

Simulation results

Table 1 and table 2 show the results of the simulation for the situation with 3 and 5 options of a question respectively. In general the difference between LVA and LSA were higher in 3 question options compared to 5 and also by increasing sample size and correlation between bivariate data, RMSE (figure 1-4) and bias were decreased.

Considering equal probability for options, however the difference between SD and bias were low, but the results showed that it is better

Table 1. Simulation Results Of Estimating Icc Using Likert Scale Approach (Lsa) And Latent Variable Approach (Lva) When The Probabilities Of Selecting 3 Options Are Equal Or Skewed With Different Correlation Coefficients (P)

Considering 3 options for questions			The mean of ICC	Likert scale approach		Latent variable approach	
Equal probability for options	N=20	ρ=0.6	0.581	0.000	0.172	0.003	0.164
		ρ=0.7	0.692	0.003	0.150	0.007	0.137
		ρ=0.8	0.811	0.000	0.120	0.004	0.129
	N=30	ρ=0.6	0.612	0.004	0.132	0.003	0.139
		ρ=0.7	0.696	0.003	0.118	0.007	0.130
		ρ=0.8	0.801	0.003	0.096	0.004	0.099
Skew probability for options	N=20	ρ=0.6	0.579	0.059	0.277	0.001	0.149
		ρ=0.7	0.713	0.056	0.260	0.004	0.137
		ρ=0.8	0.791	0.041	0.225	0.000	0.118
	N=30	ρ=0.6	0.612	0.037	0.227	0.001	0.121
		ρ=0.7	0.698	0.028	0.214	0.004	0.110
		ρ=0.8	0.800	0.024	0.163	0.000	0.089

Table 2. Simulation results of estimating ICC using likert scale approach (LSA) and latent variable approach (LVA) when the probabilities of selecting 5 options are equal or skewed with different correlation coefficients (ρ)

Considering 5 options for questions			The mean of ICC	Likert scale approach		Latent variable approach	
Equal probability for options	N=20	ρ=0.6	.0.598	0.003	0.123	0.023	0.163
		ρ=0.7	0.710	0.006	0.107	0.009	0.147
		ρ=0.8	0.810	0.001	0.088	0.014	0.123
	N=30	ρ=0.6	0.595	0.001	0.050	0.003	0.065
		ρ=0.7	0.697	0.001	0.030	0.002	0.047
		ρ=0.8	0.798	0.004	0.068	0.013	0.098
Skew probability for options	N=20	ρ=0.6	0.592	0.047	0.249	0.017	0.193
		ρ=0.7	0.689	0.049	0.234	0.021	0.187
		ρ=0.8	0.799	0.030	0.194	0.015	0.144
	N=30	ρ=0.6	0.601	0.032	0.204	0.015	0.159
		ρ=0.7	0.699	0.030	0.187	0.020	0.147
		ρ=0.8	0.801	0.022	0.158	0.008	0.122

to use LSA method specially in lower sample sizes (n=20). The correlation coefficient of bivariate data (test and retest) didn't affect the difference between bias and SD in two models remarkably.

Assuming different probabilities for options, there was a considerably difference between RMSE (figure1-4), bias and SD of estimation of ICC in two models. Using LVA resulted less bias, SD and RMSE especially in lower sample sizes (n=20).

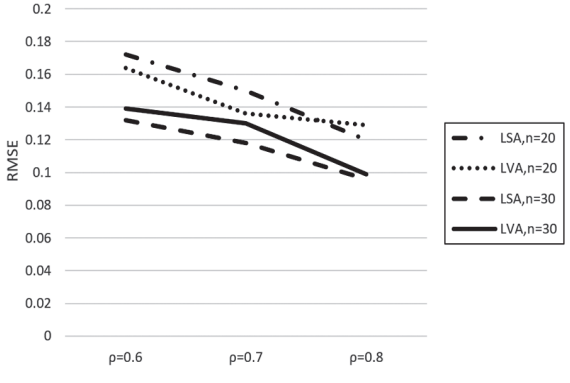

Figure 1. Comparing RMSE of estimating ICC in likert scale approach (LSA) with latent variable approach (LVA) when the probabilities of selecting 3 options are equal (ρ: the correlation between bivariate data, n: sample size)

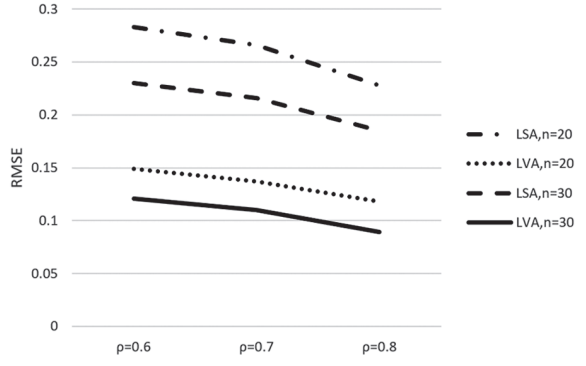

Figure 2. Comparing RMSE of estimating ICC in likert scale approach (LSA) with latent variable approach (LVA) when the probabilities of selecting 3 options are skewed (ρ: the correlation between bivariate data, n: sample size)

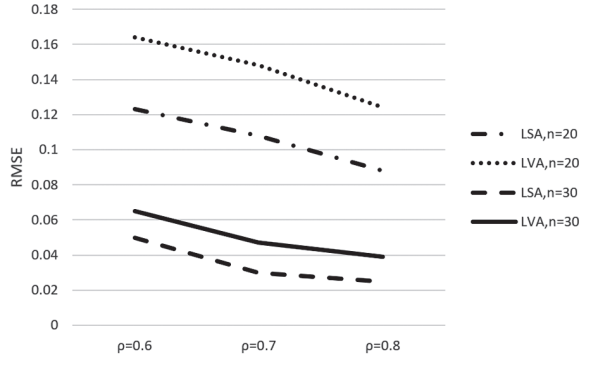

Figure 3. Comparing RMSE of estimating ICC in likert

scale approach (LSA) with latent variable approach (LVA) when the probabilities of selecting 5 options are equal (ρ: the correlation between bivariate data, n: sample size)

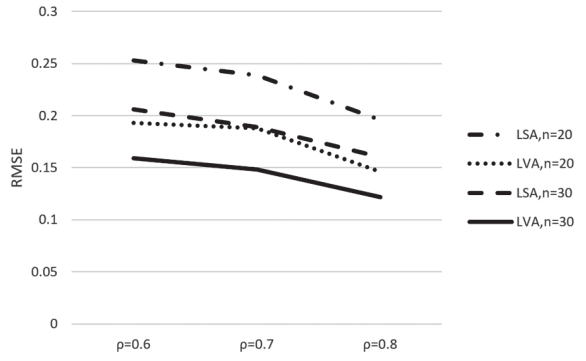

Figure 4. Comparing RMSE of estimating ICC in likert scale approach (LSA) with latent variable approach (LVA) when the probabilities of selecting 5 options are skewed (ρ: the correlation between bivariate data, n: sample size)

Beck scale for suicidal ideation example

Beck for suicidal ideation scale is a 19- item measurement for evaluating suicide risk and the severity of suicidal intent. It was developed about 20 years ago by Beck et al. Each of the 19 items has 3 options scores from 0 to 2. The first 5 items shows the attitudes of living and dying and the participants who report suicide attempts (question number 5) continuous answering items number 6 to 19. The total score obtained by summing the scores of 19 items so the total score is ranged from 0 to 38.(Beck, Brown, & Steer, 1997).

A study conducted to compare 3 psychological methods for improving patients who subsided thus they were hospitalized. Beck scale for suicidal ideation was used to evaluate the

Table 3. Obtained results of using likert scale approach (LSA) and latent variable approach (LVA) for Beck for suicidal ideation scale

No. of question	ICC/ Likert (SD)	ICC/ Normal (SD)	n
1	0.851 (0.142)	0.846 (0.105)	22
2	0.730 (0.199)	0.724 (0.149)	22
3	0.730 (0.194)	0.724 (0.158)	22
4	0.649 (0.243)	0.611 (0.180)	22
5	0.686 (0.281)	0.758 (0.191)	22
6	1 (0)	1 (0)	6
7	1 (0)	1 (0)	6
8	0.706 (0.206)	0.706 (0.162)	6
9	0.706 (0.214)	0.706 (0.159)	6
10	1 (0)	1 (0)	6
11	0.853 (0.105)	0.878 (0.087)	6
12	0.865 (0.139)	0.777 (0.128)	6
13	0.414 (0.238)	0.371 (0.248)	6
14	0.918 (0.120)	0.986 (0.045)	6
15	1 (0)	1 (0)	6
16	0.899 (0.066)	0.963 (0.018)	6
17	1 (0)	1 (0)	6
18	1 (0)	1 (0)	6
19	1 (0)	1 (0)	6

improvement of patients across the 3 methods. Before using this questionnaire, researchers wanted to assess internal and external consistency by alpha chronbach and ICC respectively.

After the test retest, it was clear that the probability of choosing options were not equal, thus, the proposed method in this study was used for calculating ICC of items. The results are shown in table 3. A total of 22 patients were filled the questionnaire thus there were 22 samples of the first 5 items, but because of the structure of the questionnaire (as mentioned before) only 6 samples were existed for question 6-19. For assessing the standard deviation of ICC estimates in both methods, bootstrap was utilized. Obtained results showed that the standard deviation in Likert scale approach was higher than latent variable approach in all questions. This result was in the line of the simulation. The estimates of ICC in first 5 items were more similar compared to the other items.

Discussion

In this study an alternative method was proposed to improve the estimation ICC for evaluating external consistency in Likert scale questionnaires. This method was developed by using a latent variable model thus the options of a question were regarded as a continuous latent variable and the latent distribution of test-rest answers are a continuous bivariate distribution. Therefore the labels of the options (usually they labeled as 1, 2,…,n / n=number of the options) can be estimated by latent variable approach using the probability of choosing each option. By getting away from equality of the probabilities, the proposed method (LVA) works better than the common method (LSA). This result was not out of mind, because of by labeling the options of questions by consecutive numbers, it means that the probability of choosing options considered the same and if this assumption was not correct, therefore the adequacy of the ICC estimation would be less.

As the first step, results showed that, there was not any remarkable difference between Pearson correlation nd ICC. This result was concluded by Raadt et at.(Raadt et at. 2021).

The Simulation shows that when the probabilities of choosing options were the same and there were 3 options to select, although there were not considerable differences between two models, but, LVA had less RMSE in lower sample sizes (n=20) but in higher sample size (n=30) LSA was more adequate. When we deal with 5 options to choose for a question, LSA had less RMSE. The reason is that when we deal with equal probabilities, the best label for options will be consecutive numbers and in the LSA model we do so. But in LVA the probabilities of choosing an option and labels will be estimated and because they are approximations, thus the labels will not be exactly consecutive numbers.

The RMSE comparatively higher in questions with 5 options to 3 options (LSA was more adequate). This may arise from that by utilizing LVA, we had to estimate 4 and 2 cut points in questions with 5 and 3 options respectively. The estimated parameters in 5 options question were more, thus the error of estimation would be higher. Another reason is that as mentioned before, the ICC is an index for continuous data and 5 option is closer to continuous data compared to 3 options thus using LSA would have a better estimation in 5 options.

In data with skewed probabilities LVA had less RMSE compare to LSA in all simulations

with different sample sizes and correlation coefficients. The obtained results were expected because in LSA equal probability of choosing an option in the LSA was assumed so using LSA estimator is not logic. But in LVA the probabilities would be estimated based on observed data. The diversity of the models was higher in 3 options vs. 5 options because as mentioned before, data with 5 option is closer to continuous data compared to 3 options.

Conclusion

Totally it seems that when the probability of choosing options are not the same (skew to an option) it is better to use LVA to estimate ICC instead of the common estimator (LSA) because it had a less standard deviation (SD) and bias compare to LSA. Specially using LVA would reduce SD and the difference between the two models will be more considerable when the options of the questions are lower.
Although we have focused on test-retest but this approach can be used in assessing inter rater consistency in two raters with Likert evaluations. The usefulness of this method would be bolder when the true value of the ICC is near the threshold of adequacy (most of the time 0.7).

References

1. Agresti, A. (2013). Categorical data analysis: John Wiley & Sons.

2. Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. Journal of biopharmaceutical statistics, 17(4), 529-569.

3. Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. Psychological reports, 19(1), 3-11.

4. Beck, A. T., Brown, G. K., & Steer, R. A. (1997). Psychometric characteristics of the Scale for Suicide Ideation with psychiatric outpatients. Behaviour research and therapy, 35(11), 1039-1046.

5. Biswas, A., & Das, K. (2002). A Bayesian analysis of bivariate ordinal data: Wisconsin epidemiologic study of diabetic retinopathy revisited. Statistics in Medicine, 21(4), 549-559.

6. Chen, C.-C., & Barnhart, H. X. (2008). Comparison of ICC and CCC for assessing agreement for data without and with replications. Computational Statistics & Data Analysis, 53(2), 554-564.

7. Kim, K. (1995). A bivariate cumulative probit regression model for ordered categorical data. Statistics in Medicine, 14(12), 1341-1352.

8. Liljequist D, Elfving B, Skavberg Roaldsen K (2019) Intraclass correlation – A discussion and demonstration of basic features. PLoS ONE 14(7): e0219854McCullagh, P. (1980). Regression models for ordinal data. Journal of the Royal Statistical Society. Series B (Methodological), 109-142.

9. Maxwell, A. E. (1977). Multivariate analysis in behavioural research (pp. 164p-164p). London: Chapman and Hall.

10. McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass

correlation coefficients. Psychological methods, 1(1), 30.

11. Mehta, S., Bastero-Caballero, R. F., Sun, Y., Zhu, R., Murphy, D. K., Hardas, B., & Koch, G. (2018). Performance of intraclass correlation coefficient (ICC) as a reliability index under various distributions in scale reliability studies. Statistics in medicine.

12. Metrology, J. (2008). Evaluation of measurement data—Guide to the ex pression of uncertainty in measurement. Bureau International des Poids et Mesures.

13. Qin. S., Nelson. L., McLeod. L., Eremenco. S., Coons. S.J. (2019). Quality of Life Research 28, 1029–1033.

14. de Raadt, A., Warrens, M. J., Bosker, R. J., & Kiers, H. A. (2021). A comparison of reliability coefficients for ordinal rating scales. Journal of Classification, 38(3), 519-543.

15. Ren, S., Yang, S., & Lai, S. (2006). Intraclass correlation coefficients and bootstrap methods of hierarchical binary outcomes. Statistics in medicine, 25(20), 3576-3588.

16. shan. G. (2020). Estimation of bias-corrected intraclass correlation coefficient for unbalanced clustered studies with continuous outcomes. Communications in Statistics - Simulation and Computation, DOI:10.1080/03610918.2020.1811332

17. Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. Psychological bulletin, 86(2), 420.

18. Todem, D., Kim, K., & Lesaffre, E. (2007). Latent-variable models for longitudinal data with bivariate ordinal outcomes. Statistics in Medicine, 26(5), 1034-1054.

19. Yen, M., & Lo, L.-H. (2002). Examining test-retest reliability: an intra-class correlation approach. Nursing research, 51(1), 59-62.