INTRACLASS AND CONCORDANCE CORRELATION COEFFICIENTS AS A MEASURE OF RELIABILITY

Joe Matsuoka[1], Hiroshi Ohtsu[2], Tatsuo Kagimura[3], and Tadashi Kusunoki[4]

[1] Quintiles Transnational Japan K.K., Tokyo Japan
[2] Department of Clinical Trial Data Management, The University of Tokyo, Japan
[3] Nippon Boehringer Ingelheim Co., Ltd., Tokyo Japan
[4] Japanese Society for Pharmacoepidemiology, Tokyo Japan

The intraclass correlation coefficient (ICC) has long been used as a measure of reliability and different versions for two- and one-way layout are available corresponding to study design (Ebel (1951) [1], Shrout and Fleiss (1979) [2], etc.). While the concordance correlation coefficient (CCC) described by Lin (1989) [3] and Barnhart et al. (2002) [4] is a measure of agreement for two-way layout. We present here a "one-way layout version" of CCC. Letting Y_{ij}, j=1,…,m, be continuous data for subject i, i=1,…,n, and plotting (Y_{ij}, Y_{ik}) (j≠k) twice on a scatter plot with horizontal and vertical axes of x and y, as (x,y)=(Y_{ij}, Y_{ik}) and reciprocally as (Y_{ik},Y_{ij}), we have a dispersion with a total of nm(m-1)points symmetric along the 45°line y=x, thus Fisher (1925) [5] called it "symmetrical table." If data for every subject have a same value, correlation coefficient between x and y and regression line become r_{xy}=1 and y=x, respectively. Since r_{xy}=1 directly means complete agreement/stability of data, we call r_{xy} "one-way layout version" of CCC, and correction for bias, Cb for usual two-way CCC, is unnecessary. This measure for one-way layout is also applicable to two-way layout. We write various ICC and CCC including one-way layout versions in terms of two-way ANOVA, V_S, V_R, and V_e for the subjects, raters, and error mean squares. When interrater variance is ignored, corresponding ICC and CCC coincide with $(V_S - V_e)/\{ V_S +(m-1) V_e \}$, which is the upper limit of usual ICC and CCC and they are expressible as modification on the upper limit. Examining the modifiers on numerator and/or denominator, we can see that usual ICCs are smaller than the upper limit and CCCs are further smaller than ICCs under a normal condition of reliability studies, $V_S \geq V_R \geq V_e$. However, since every modifier is divided by n or n-1, modification is not so influential if n is large and all measures approach to common upper limit as $n \to \infty$. Using the F ratio $F=V_S/ V_e$, common upper limit is written as (F-1)/(F-1+m) thus a high reliability is hopeless without a high F value or small V_e relative to V_S even though interrater bias V_R is not effective. On the other hand, appropriate measure for one-way layout is limited to one-way version of ICC and CCC. We will present illustrative examples of reliability studies conducted by our Scaling Keys of Evaluation Techniques for Cerebrovascular patients Heterogeneity (SKETCH) study group to examine the above-mentioned relationship amongst various ICCs and CCCs.

References

[1]Ebel RL (1951). Estimation of the reliability of ratings. Psychometrika 16: 407-424.
[2]Shrout PE and Fleiss JL (1979). Intraclass correlations: use in assessing rater reliability. Psychological Bulletin 86: 420-428.
[3]Lin LI (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics 45: 255-268.
[4]Barnhart HX et al. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics 58: 1020-1027.
[5]Fisher RA (1925). Statistical methods for research workers. Oliver & Boyd.