



LETTER TO THE EDITORS

Negative Values of the Intraclass Correlation Coefficient Are Not Theoretically Possible

In their methodological review of the indices of reproducibility, Tammemagi *et al.* [1] mentioned the intraclass correlation coefficient, which, they say, varies “from -1 for perfect disagreement, to 0 for random agreement, and to $+1$ for perfect agreement.” However, such an interpretation of negative values of this statistic, although previously asserted by Deyo *et al.* [12], is misleading. We argue that negative values are not theoretically possible although possibly observable as explained below.

In the framework of a two-way layout model, we define

$$X_{ij} = \mu + S_i + R_j + E_{ij} \quad i = 1 \dots N, \quad j = 1 \dots k$$

where X_{ij} is the random variable associated to the j th measure of the i th subject, μ is a constant equal to the grand mean of all the observations in the population, the subject effects S_i are independent random variables normally distributed with mean 0 and variance σ_S^2 , the rater effects R_j are independent random variables normally distributed with mean 0 and variance σ_R^2 , and the residual errors E_{ij} are independent random variables normally distributed with mean 0 and variance σ_E^2 ,

$$\text{cov}(S_i, E_{ij}) = 0, \quad \text{cov}(R_j, E_{ij}) = 0,$$

and

$$\text{cov}(S_i, R_j) = 0$$

which implies that

$$\sigma_X^2 = \sigma_S^2 + \sigma_R^2 + \sigma_E^2.$$

The intraclass correlation coefficient is defined as

$$\rho = \frac{\sigma_S^2}{\sigma_X^2} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_R^2 + \sigma_E^2}$$

Then, when

$$\sigma_S^2 = 0$$

(the whole variance is attributable to the rater effects and the random errors, i.e., no agreement) $\rho = 0$. Conversely, when

$$\sigma_R^2 = \sigma_E^2 = 0,$$

$\rho = 1$, which means perfect agreement (the whole variance is attributable to the variance between subjects).

The estimator of the intraclass correlation coefficient is then derived from an analysis of variance (ANOVA) table. Using the unbiased estimator of the three components of variance

$$\hat{\sigma}_E^2 = \text{MSE}$$

(MSE = mean sum of squares error),

$$\hat{\sigma}_R^2 = (\text{MSR} - \text{MSE})/N$$

(MSR = mean sum of squares raters), and

$$\hat{\sigma}_S^2 = (\text{MSS} - \text{MSE})/k$$

(MSS = mean sum of squares subjects), the intraclass correlation coefficient is then estimated, as indicated in the article by Tammemagi *et al.*, by

$$\hat{\rho} = \frac{N(\text{MSS} - \text{MSE})}{[N(\text{MSS}) + k(\text{MSR}) + (Nk - N - k)\text{MSE}]}$$

This estimator can indeed lead to negative values since if $\text{MSS} = 0$ and $\text{MSR} = 0$ the estimate is equal to

$$-1/(k - 1 - k/N),$$

which is the lowest value. In the case of two measures by subject, this lowest bound reduces to $-N/(N - 2)$, which is not very far from -1 . Negative values are then possibly observable but there is no other possible interpretation but poor agreement. In the same way, if $\text{MSS} = 0$ and $\text{MSR} = 0$ this leads to negative values for $\hat{\sigma}_R^2$ and $\hat{\sigma}_S^2$, and nevertheless, no one would state that a variance could be negative. These negative values of $\hat{\sigma}_R^2$ and $\hat{\sigma}_S^2$ are no longer possible as both N and p tend to $+\infty$, because of the consistency of the estimators. It is the same for $\hat{\rho}$, which is known to be biased but consistent [3,4].

Finally, it should be noted that the authors only consider a two-way layout model (with random or fixed effects) while the one-way layout model (no rater effects) is more widely used in assessing agreement. In this usual framework, the lower bound of the estimate is known to be $-1/(k - 1)$ [3], which then leads to the value -1 when only two measures by subject are considered. Nevertheless, here again, negative values of the estimate have no theoretical legitimacy.

BRUNO GIRAudeau

Département de Biostatistique et Informatique Médicale

Hôpital Saint-Louis

75475 Paris, France

and Epidémiologie et Sciences de l'Information

INSERM Uhhh

75571 Paris, France.

References

1. Tammemagi MC, Frank JW, Leblanc M, Artsob H, Streiner DL. Methodological issues in assessing reproducibility—A comparative study of various indices of reproducibility applied to repeat ELISA serologic tests for Lyme disease. *J Clin Epidemiol* 1995; 48: 1123–1132.
2. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. *Controlled Clin Trials* 1991; 12: 142S–158S.
3. Olkin I, Pratt WJ. Unbiased estimation of certain correlation coefficients. *Ann Math Stat* 1958; 29: 201–211.
4. Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* 1978; 43: 259–262.