Alternative Agreement Approaches in Testing the Agreement of Three-dimensional Computerized Tomography Measurements of Styloid Processes by Two Observers at Two Different Times*

Stiloid Proseslerin İki Farklı Zamanda İki Değerlendirici Tarafından Alınan Üç Boyutlu Bilgisayarlı Tomografi Ölçümlerinin Uyumunun Test Edilmesinde Alternatif Uyum Yaklaşımları

Abstract

Aim: This study aimed to evaluate the intraobserver and interobserver agreement of styloid process (SP) measurements on three-dimensional (3D) computerized tomography (CT) images between two experienced radiologists by using the concordance correlation coefficient (CCC) and coefficient of individual agreement (CIA) agreement statistical methods.

Materials and Methods: Contrast-enhanced carotid CT angiography images of 68 patients, performed between June 2015 and December 2015, were evaluated retrospectively. The length between the attachment point of the SP to the temporal bone and the distal end was measured on 3D CT images performed with a 64-slice CT scanner. Both the agreements between the two radiologists and the replicated measurements of each radiologist were calculated with the help of CIA and CCC agreement indexes.

Results: When CIA was used for agreement statistics, the two radiologists disagreed in the right and left measurements of the individual. When the CCC agreement statistic was used, there was a perfect agreement between the measurements of the two radiologists.

Discussion and Conclusion: The variance values of between-subject and within-subject should be taken into consideration for each observer in cases of two replicated measurements. If these values show very large differences from each other, CCCtotal values may have been calculated larger than the real value. However, the CIA statistics value is more stable and when such a case is encountered, researchers are advised that between-subject and within-subject variances should be calculated due to the differences between the two replicated measurements of each observer. Therefore, the $\sigma_{\mathbf{B}}^2/\sigma_W^2$ rate should absolutely be taken into consideration.

Keywords: disagreement; interobserver agreement; intraobserver agreement; replicated measurement; styloid processes

Öz

Amaç: Bu çalışmada üç boyutlu bilgisayarlı tomografi (BT) görüntüleri üzerinde iki deneyimli radyolog tarafından alınan stiloid proses (SP) ölçümlerinin değerlendiriciler arası ve değerlendirici içi uyumunu, uyum istatistik yöntemleri olan konkordans korelasyon katsayısı (concordance correlation coefficient—CCC) ve birey uyum katsayısı *(coefficient of individual agreement—CIA)* metotları kullanılarak değerlendirilmesi amaçlanmıştır.

Semra Erdogan[1], Kaan Esen[2], Yuksel Balci[2], Sermin Tok Umay[2], Gulhan Temel[3], Anil Ozgur[2]

[1] Mersin University, Faculty of Medicine, Department of Biostatistics and Bioinformatics

[2] Mersin University, Faculty of Medicine, Department of Radiology

[3] Mersin University, Faculty of Medicine, Department of Biostatistics and Bioinformatics

Geliş/*Received* : 12.07.2018
Kabul/*Accepted*: 11.01.2019

DOI: 10.21673/anadoluklin.443110

Yazışma yazarı/*Corresponding author*
Semra Erdogan
Mersin University, Faculty of Medicine, Department of Biostatistics and Bioinformatics, 33343 Mersin, Turkey
E-mail: semraerdogann@gmail.com

ORCID
Semra Erdogan: 0000-0002-2528-0585
Kaan Esen: 0000-0002-8404-7484
Yuksel Balci: 0000-0003-1758-9600
Sermin Tok: 0000-0002-7161-016X
Gulhan Temel: 0000-0002-2835-6979
Anil Ozgur: 0000-0003-4463-1923

* This study was presented as a poster at the 19th International and 2nd National Congress of Biostatistics held in Antalya in October 2017.

Gereç ve Yöntemler: Haziran 2015—Aralık 2015 döneminden, 68 hastaya ait kontrastlı karotis BT anjiyografi görüntüleri geriye dönük olarak incelendi. SP temporal kemiğe tutunma noktası ve distal ucu arasındaki mesafe, 64 kesitli bir BT cihazı ile gerçekleştirilen üç boyutlu BT görüntülerinde ölçüldü. Radyologlar arasındaki ve radyologların kendi tekrarlı ölçümleri arasındaki uyum, CIA ve CCC uyum indekslerinden yararlanılarak hesaplandı.
Bulgular: Uyum istatistikleri olarak CIA kullanıldığında, bireylerin sağ ve sol ölçümlerinde iki radyolog arasında uyum bulunmadı. CCC uyum istatistiği kullanıldığında ise iki radyoloğun tekrarlı ölçümleri arasında mükemmel bir uyum bulundu.
Tartışma ve Sonuç: Her bir değerlendiriciden iki tekrarlı ölçüm alındığında, denekler arası ve denekler içi varyans değerlerinin dikkate alınması gerekmektedir. Eğer bu değerler birbirinden oldukça büyük farklılıklar gösteriyorsa, CCCtotal değerleri gerçek değerden daha büyük bir değer olarak hesaplanacak ve elde edilen bu sonuçlar araştırmacıları yanıltacaktır. Böyle bir durumla karşılaşıldığında, CIA istatistiğinin çok daha kararlı olduğu ve denekler arası ve denekler içi varyans değerlerinin, her bir değerlendiricinin iki tekrarlı ölçümleri arasındaki farklılıklar üzerinden hesaplanması gerektiği araştırmacılara önerilir. Böylece, σ_B^2/σ_W^2 oranı mutlaka dikkate alınmalıdır.
Anahtar Sözcükler: değerlendirici içi uyum; değerlendiriciler arası uyum; stiloid proses; tekrarlı ölçümler; uyumsuzluk

INTRODUCTION

Accuracy and precise measurements make up a quite important component for any appropriate study design. Ideally, a measurement or variable should be measured without an error. However, in many cases, it is impossible to reveal the true value of the quantity of what is to be measured with one measurement. In such situations, where it is impossible or very difficult to identify the true value, more than one measurement can be taken from each subject or more than one observer/measurement tool/measurement method can be preferred (1–3). In a situation where multiple measurements have been taken from each subject, two different situations are handled in testing the agreement between the measurements. The first of these is that the replicated measurements taken from the same subject do not change related to time, and the second is that the replicated measurements obtained from the same subject are stable, i.e., the measurement values do not show any difference (4,5).

The styloid process (SP) is a cartilaginous bone located on the inferior side of the temporal bone (6). SP is a part of the stylohyoid complex that is composed of the cornu minus of the hyoid bone, the stylohyoid ligament and the styloid process. The elongated styloid process or calcified stylohyoid ligament causes symptoms described by Eagle, known as Eagle's syndrome (7). Dysphagia, otalgia, hemifacial or neck pain, change in the voice, foreign object sensation, pain in opening the mouth, discomfort during chewing are major symptoms (8). However, the symptoms may mimic other diseases in this region and clinical findings may be insufficient for the diagnosis. Radiologic evaluation is needed to support the diagnosis and for the management of the disease. Therefore, the length and variations of styloid processes have been evaluated in various studies in the literature (7,9). Nevertheless, the attachment point of the SP to the temporal bone may not always be observed clearly, and the measurements of SP may vary between radiologists.

The biggest problem in agreement analyses is to decide on the statistical method to be used. In many agreement studies, it has been observed that tests like the Pearson correlation coefficient, the regression analysis or dependent samples t test are used for continuous measurements and that classical statistics methods like the chi-squared test and the Cohen's kappa are used in categorical measurements. The correlation coefficient gives the amount of the relationship in continuous measurements, not the agreement between two variables. Also, a change in the scale of the measurement does not affect the correlation, but it certainly affects the agreement (10). Studies by Stralen et al. (2012) have revealed that systematic error is ignored when the Pearson correlation coefficient is used while testing the agreement between two continuous measurement methods. Additionally, the effects of prevalence and bias effects are not removed as a result of using the Cohen's kappa coefficient while testing the agreement between the two categorical measurement methods, and the different weighting calculations for disagreement cells are ignored (11).

Agreement analyses are generally evaluated by using scaled or unscaled agreement measurements. If the result variable is in a continuous measurement, the mean squared deviation (MSD), the coverage probability (CP) and the total deviation index (TDI) are given as unscaled agreement measurements, and the intraclass correlation coefficient (ICC), the concordance correlation coefficient (CCC) and the coefficient of individual agreement (CIA) are given as scaled agreement measurements (2,12).

Two popular methods used to evaluate the agreement between the quantitative measurements taken from different observations are ICC and CCC. However, although CCC and ICC agreement statistics are similar statistics, there are some differences between them. One of these differences is that while the choice of observers in CCC is fixed, ICC is suggested to be fixed or random and requires ANOVA model assumptions, while CCC does not. Together with this, in situations where there are no replicated measurements, ICC and CCC give very close or the same results. ICC is generally described in the one- way or two-way variance analysis structure in case of interobserver changeability assumption, meaning that the observers are interchangeable. In such a situation, there is the necessity that all the correlations between pairs are within error variance of all the observers and the replicated measurements of the observers are equal. King et.al. (2007) suggested CCC to evaluate the interobserver agreement in cases where there are replicated measurements when the observer and the measure time is taken as fixed effect in the model (2,13–17).

CIA is a scaled agreement index based on an acceptable disagreement idea. Furthermore, most of the time, with the evaluation of disagreement, both interobserver and intraobserver disagreements are evaluated. The intraobserver disagreement measures the consistency between the repeated measurements of the observer and the interobserver disagreement measures the consistency of the real differences between the measurements taken by the observers (18, 19). The agreement coefficient takes a value between 0 and 1. If the agreement coefficient takes a value close to 1, it means that there is a perfect interobserver agreement, if it takes a value close to 0, this means that there is almost no interobserver agreement (12,20).

Accordingly, in this study we aimed to evaluate the intraobserver and interobserver agreement of SP measurements on three-dimensional (3D) CT images between the replicated measurements of two experienced radiologists by using CCC and CIA agreement statistical methods.

MATERIAL AND METHODS

Study population

In this study, contrast-enhanced carotid CT angiography images of a total of 68 patients that were performed between June 2015 and December 2015 were evaluated retrospectively. CTA examinations were performed for different indications such as diagnosis of vascular stenosis, vascular malformations, or follow-up after previous treatments. However, none of the patients were evaluated for clinical findings of the Eagle syndrome.

The approval for the study was obtained from the ethical committee of our institute (Mersin University Clinical Research Ethical Committee; protocol number: 2016/237; date of approval: 14/07/2016).

Imaging protocol and image analysis

CT examinations were performed with a 64-slice CT scanner (Toshiba Aquilion 64, Toshiba Medical Systems, Tokyo, Japan). Eighty cc nonionic contrast medium was administered via an antecubital vein. The Bolus tracking technique was used. CT parameters were as follows: tube voltage, 120 kVP; effective mAs, 220; slice thickness, 0.5 mm. The images were transferred to the workstation and three-dimensional (3D) images were created with a DICOM imaging program (Vitrea; Vital Images) on the workstation. The length of SP was measured through the 3D images on the coronal plan (Figure 1). The length between the attachment point of the SP to the temporal bone and the distal end was measured. The ossification of stylohyoid ligament connecting with SP at the distal end was also measured. All of the measurements were performed by two experienced radiologists on the same workstation at standard CT parameters and the radiologists were unaware of the patient information. There was a 3-week period between the replicated measurements of the same radiologist.

Concordance correlation coefficient (CCC)

Barnhart et.al. (2007) were the first to present the inter-CCC, intra-CCC, and total-CCC formulations for situations with multiple observers, where none of the observers is accepted as reference and where there are replicated measurements of each observer. Here, in cases where observers have taken only one measurement from each individual, the obtained CCC value equals the total CCC. Replicated data are not necessary in the description of total-CCC. Although both between-subject (σ_{Bj}^2) and within-subject (σ_{Wj}^2) variabilities are estimated for replicated data, total variability ($\sigma_j^2 = \sigma_{Bj}^2 + \sigma_{Wj}^2$) can only be estimated for cases without replicated data. So, inter-CCC or intra-CCC cannot be estimated for cases without replicated data (2,3).

$$(1) \quad CCC_{total} = \rho_c = \frac{2\sigma_{B1}\sigma_{B2}\rho_{\mu12}}{2\sigma_{B1}\sigma_{B2} + (\mu_1-\mu_2)^2 + (\sigma_{B1}-\sigma_{B2})^2 + \sigma_{W1}^2 + \sigma_{W2}^2}$$

$$(2) \quad CCC_{inter} = \rho_c(\mu) = \frac{2\sigma_{B1}\sigma_{B2}\rho_{\mu12}}{2\sigma_{B1}\sigma_{B2} + (\mu_1-\mu_2)^2 + (\sigma_{B1}-\sigma_{B2})^2}$$

$$(3) \quad CCC_{j,intra} = \rho_j^I = ICC_{1j} = \frac{\sigma_{Bj}^2}{\sigma_{Bj}^2 + \sigma_{Wj}^2}$$

In cases where there are two observers and two repeated measurements, total-CCC, inter-CCC, and intra-CCC are calculated as in the following Equations 1 to 3, respectively. Intra-CCC is equivalent with ICC (one-way random effect model) formulation for each observer (2,3).

Here ($\mu_1 - \mu_2$) is the difference between the means and $\rho_{\mu12}$ is the pairwise correlation between measurements from observer 1 and observer 2.

Coefficient of individual agreement (CIA)

In continuous measurements, for the cases where there are two observers/methods and none of the observers/methods constitutes a gold standard, CIA is stated as in Equation 4 (2,3).

$$(4) \quad CIA^N = \psi^N = \frac{\sigma_{W1}^2 + \sigma_{W2}^2}{2(1-\rho_{\mu12})\sigma_{B1}\sigma_{B2} + (\mu_1-\mu_2)^2 + (\sigma_{B1}-\sigma_{B2})^2 + \sigma_{W1}^2 + \sigma_{W2}^2}$$

In the case where the disagreement between the measurements taken from different observers are similar and the disagreement between the replicated measurements of the same observer are similar, it can only be said that there is a perfect agreement between the two observers. In other words, the individual difference between the measurements taken by different observers is relatively small and thus, this difference is close to the individual difference between the replicated measurements of both observers. If the individual difference between the measurements taken by different observers is relatively large and exceeds the difference between the replicated measurements of the same observer, it can be concluded that there is a poor agreement between the observers (2).

The relationship between CIA and CCC

In the presence of only two observers and with the assumption that within-subject ($\sigma_{W1}^2 = \sigma_{W2}^2 = \sigma_W^2$) and between-subject ($\sigma_{B1}^2 = \sigma_{B2}^2 = \sigma_B^2$) variances of the two observers are equal, Barnhart et. al (2007) put forward the similar and different sides of these two coefficients. The difference between the means ($\mu_1 - \mu_2$), within-subject variance, between-subject variance, and correlation coefficient ($\rho_{\mu12}$) with CIA and total-CCC is reformulated as in Equations 5 and 6 (3).

$$(5) \quad \psi^N = \frac{2\sigma_W^2}{(\mu_1-\mu_2)^2 + 2(1-\rho_{\mu12})\sigma_B^2 + 2\sigma_W^2}$$

$$(6) \quad \rho_c = \frac{2\sigma_B^2\rho_{\mu12}}{(\mu_1-\mu_2)^2 + 2(\sigma_B^2 + \sigma_W^2)}$$

When the formulations belonging to the two coefficients are studied, it can be observed that as the correlation coefficient increases, both coefficient values increase and these two coefficients decrease due to the increase of the difference between the means. However, these two coefficients show differences in within-subject and between-subject variances. Namely, when the within-subject variance increases (σ_W^2), CIA increases while CCC decreases. According to the increase in the value of between-subject variance (σ_B^2), CIA decreases while CCC increases. CIA is less affected by the rate of within-subject variability to between-subject variability (σ_B^2/σ_W^2) compared to CCC (2,3). CIA is measured with the difference between the means ($\mu_1 - \mu_2$), between-subject variance and within-subject variance, and in such a case $\psi^N \geq \rho_c$ (3).

Statistical analysis

Normality controls belonging to continuous measurements were tested by the Shapiro–Wilk test. The differences between the male and female mean ages and the difference between the radiologists related to each measurement was tested by the Student t test. The differences between right and left measurements of each radiologist were examined by the paired samples t test. Mean and standard deviation values were given as descriptive statistics. The relationships belonging to continuous measurements were tested by the Pearson correlation coefficient. Furthermore, the agreement between the radiologists and the replicated measurements of the radiologists were calculated with the help of CIA and CCC agreement indexes. Other statistics were calculated by using a demo version of the SPSS 21 software package. The significant level was taken as p<0.05.

RESULTS

A total of 68 patients –22 (32.4%) females and 46 (67.6%) males– were included. The mean age for the female and male patients was 58.1±14.2 and 60.6±13.7, respectively, and there was no statistically significant difference between the two sexes in terms of age (p=0.485). Additionally, right and left measurements of the patients were measured at two different times by two different radiologists. The descriptive statistics and p values belonging to these are given in Table 1. In the right and left measurements, a statistically significant difference was not observed between the measurements of each radiologists (p>0.05 for all), and we also studied whether there was a difference between the radiologists for each measurement taken at different times. Again, no significant difference was observed (p>0.05 for all). In conclusion, it was determined that there was no difference both between the measurements taken by the radiologists and between the measurements taken at different times. Besides, in the right and left measurements of the patients, the relationship between the radiologists for each measurement taken at different times was studied, and a statistically linear relationship between the radiologists was identified (p<0.001 for all).

The agreement statistics showing the agreement between the two replicate measurements of the two radiologists for the right and left measurements of the individual are presented in Table 2. In order to calculate the agreement statistics according to the data, the between-subject and within-subject variance values of each radiologist were used. When variance values were examined, between-subject variability values were observed to be quite larger than the within-subject values for both radiologists. So, when compared with all other CCC values, CIA value was calculated to be quite small. As the CIA value considers the within-subject variability rather than the between-subject variability, it is expected that this value is different from and even smaller than other agreement statistics. Besides, it was observed that between-subject variance values and within-subject variance values were quite close to each other for both radiologists.

In the existence of two observers and from the assumption of the equality of between-subject and within-subject variants of the two observers, CIA and CCC_{total} statistics were calculated. Therefore, the differences between the two measurements of both radiologists were obtained. Within-subject and between-subject variances were calculated from these differences. Later, CIA and CCC_{total} calculations were made for the right and left measurements. The results are shown in Table 3. There were no great differences between the within-subject and between-subject variance values. It is well known that depending on the increase of the within-subject variance, CIA value increases, CCC_{total} value decreases, and depending on the increase of the between-subject variance, CCC_{total} value increases,

Table 1. The descriptive statistics taken at two different times by two different radiologists and the Pearson correlation coefficient between the radiologists for each measurement.

	Right			**Left**		
	1st Measurement	**2nd Measurement**	**P^a**	**1st Measurement**	**2nd Measurement**	**P^a**
Radiologist 1	26.31±9.59	26.14±9.41	0.127	26.92±10.03	26.86±9.86	0.691
Radiologist 2	26.03±9.39	25.98±9.36	0.505	27.13±10.06	27.03±10.03	0.155
P^b	0.863	0.916		0.901	0.920	
r	0.992	0.991		0.994	0.990	
(p*)	(**<0.001**)	(**<0.001**)		(**<0.001**)	(**<0.001**)	

a: p values belonging to the differences between 1st and 2nd measurements; b: p values belonging to the differences between the radiologists; r: Pearson correlation coefficient, p*: p values belonging to correlation coefficients

Table 2. The agreement statistics for right and left measurement

	CIA (ψ^N)	CCC$_{total}$ (ρ_c)	CCC$_{inter}$ ($\rho_c(\mu)$)	CCC$_{intra1}$ ($\rho_1^I = ICC_{11}$)	CCC$_{intra2}$ ($\rho_2^I = ICC_{12}$)	Correlation ($\rho_{\mu 12}$)
Right	0.195	0.991	0.993	0.998	0.999	0.993
Left	0.281	0.992	0.994	0.996	0.999	0.994

while CIA value is known to decrease. Furthermore, CIA is known to be less affected from σ_B^2/σ_W^2 value compared to the CCC$_{total}$ value. From this, we can say that there is quite a high agreement between the radiologists for the right and left measurements by looking at the CIA values, and that there is a medium level agreement between the radiologists by looking at the CCC$_{total}$ values.

DISCUSSION AND CONCLUSION

Radiographs, CT and cone-beam CT examinations, dried skulls were used to measure the length of SP in different populations. Panoramic radiographs are low-cost and provide lower radiation exposure (21). However, on panoramic radiographs the origin of the SP at the lower part of the temporal bone may be hidden by shadows of the base of the skull and it may be difficult to detect (22). CT is generally the first choice imaging modality in case of clinical suspicion (23). Three-dimensional reconstruction is also accepted as the gold standard imaging modality (24). Nevertheless, even in CT examinations, the nonvisualization of the attachment point of the SP may cause incorrect assessments. The interobserver and intraobserver agreement of SP measurements was evaluated especially in the studies performed with panoramic radiographs. Vieira et al. (2015) reported an interobserver agreement with an excellent (0.89) kappa coefficient related to the presence of belonged SP that was accepted as 30 mm or more in length (21). However, studies by Scaf et al. (2003) reported a moderate (r=0.52) intraclass correlation coefficient value for interobserver agreement related to the measurement of belonged SP on panoramic radiographs (23).

Although CT is a cross-sectional imaging modality, to the best of our knowledge, the information about the interobserver and intraobserver agreement on the evaluation of SP with 3D CT is limited. In our study, when CIA was used for agreement statistics, the two radiologists disagreed in the right and left measurements of the individual. When the CCC agreement statistic was used, there was a perfect agreement between the measurements of the two radiologists.

While the interobserver agreement is examined, different sources of disagreement between the observers have to be taken into consideration. The source of disagreement may arise from differing population means, differing between-subject variability, differing within-subject variability among observers, poor correlation between measurements by observers (2). Erdogan et al. (2016) examined the results of CIA in the presence of repeated measurements of two observers at low, medium and high within-subject variability for different sample sizes. They observed that CIA was not affected by sample size and repeated measurements, but affected by the magnitude of within-subject variability (25). In using the CCC and CIA for assessing agreement, one needs to consider the magnitude of the between-subject variability and within-subject variability. Barnhart et al. (2007) have stated that it has to be decided whether the within-subject variability is acceptable based on the subject matter for the considered measurement range. If it is in an acceptable case, especially when the between-subject variability is larger than the within-subject variability,

Table 3. Within-subject and between-subject variance for right and left measurement and agreement statistics

	σ_B^2	σ_W^2	σ_B^2/σ_W^2	CIA	CCC
Right	0.415	0.317	1.309	0.973	0.558
Left	0.500	0.606	0.825	0.985	0.446

σ_B^2 : between-subject variance; σ_W^2 : within-subject variance; CIA: coefficient of individual agreement; CCC: concordance correlation coefficient

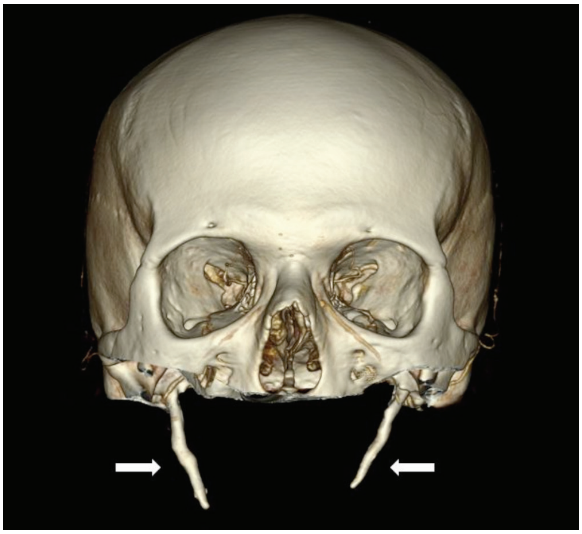

Figure 1. Bilateral elongated styloid process on 3D CT image

they support the use of CIA agreement statistic, and if it is not an acceptable value or cannot be decided, in order to comment they support the use of CCC and CIA agreement statistics together (3). Besides, studies by Pan et al. (2012) support that in the cases where the between-subject variability is large, the value of CCC is calculated to be inflated compared to the real value, and that in such cases CIA is more stable (26). Besides, studies by Barnhart et al. (2007) point out that if the between-subject variability varies greatly from a different population of the subject, the CCC values from these populations cannot be compared. In addition, they support use of the CIA agreement statistic if the magnitude of the between-subject variability relative to the within-subject variability is similar across these populations (3).

As a result, when there are two replicated measurements of two observers, the variance values of between-subject and within-subject should be taken into consideration for each observer. If these values show very large differences from each other, it has been put forward that CCC_{total} values may have been calculated larger than the real value and that the results may mislead the researchers; and that in such cases, CIA statistics is more stable (25). When such a case is encountered, researchers are advised to calculate the between-subject and within-subject variances due to the differences between the two replicated measurements of each observer and that it is necessary to calculate agreement statistics starting from this point. For the comment of the results obtained to be more reliable, σ_B^2/σ_W^2 the rate should absolutely be taken into consideration. Thus, the most correct agreement statistic results should be used.

Acknowledgements

This paper was linguistically supported by the Mersin Technology Transfer Office Academic Writing Center of Mersin University.

REFERENCES

1. Gao J, Pan Y, Haber M. Assessment of observer agreement for matched repeated binary measurements. Computational Statistics & Data Analysis. 2012;56(5):1052–60.
2. Barnhart HX, Haber MJ, Lin LI. An overview on assessing agreement with continuous measurements. J Biopharm Stat. 2007;17(4):529–69.
3. Barnhart HX, Lokhnygina Y, Kosinski AS, Haber M. Comparison of concordance correlation coefficient and coefficient of individual agreement in assessing agreement. J Biopharma Stat. 2007;17(4):721–38.
4. Bland J, Altman D. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60.
5. Bland J, Altman D. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat. 2007;17(4):571–82.
6. Başekim CC, Mutlu H, Güngör A, Silit E, Pekkafali Z, Kutlay M, et al. Evaluation of styloid process by three-dimensional computed tomography. Eur Radiol 2005;15(1):134–9.
7. Kosar MI, Atalar MH, Sabancioğullari V, Tetiker H, Erdil FH, Cimen M, et al. Evaluation of the length and angulation of the styloid process in the patient with prediagnosis of Eagle syndrome. Folia Morphol (Warsz). 2011;70(4):295–9.
8. Oztunç H, Evlice B, Tatli U, Evlice A. Cone-beam computed tomographic evaluation of styloid process: a retrospective study of 208 patients with orofacial pain. Head Face Med. 2014;10(5):1–7.
9. Okur A, Ozkırış M, Serin HI, Gencer ZK, Karaçavuş S, Karaca L, et al. Is there a relationship between symptoms of patients and tomographic characteristics of styloid process? Surg Radiol Anat. 2014;36(7):627–32.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. International Journal of Nursing Studies. 2010;47(8):931–6.

11. Van Stralen KJ, Dekker FW, Zoccali C, Jager KJ. Measuring agreement, more complicated than it seems. Nephron Clin Pract. 2012;120(3):c162–7.

12. Lin, L, Hedayet AS, Wu W. A unified model for continuous and categorical data. In: Statistical Tools for Measuring Agreement, 1. ed. New York: Springer; 2012.

13. Vangeneugden T, Laenen A, Geys H, Renard D, Molenberghs G. Applying linear mixed model to estimate reliability in clinical trials with repeated measurement. Controlled Clinical Trials. 2004;25:13–30.

14. Molenbergs G, Vangeneugden T, Laenen A. Estimating reliability and generalizability from hierarchical biomedical data. J Biopharm Stat. 2007;17(4):595–627.

15. King TS, Chinchilli VM, Carrasco JL, Wang K. A class of repeated measures concordance correlation coefficients. J Biopharm Stat. 2007;17(4):653–72.

16. Quiroz J. Assessment of equivalence using a concordance correlation coefficient in repeated measurement design. J Biopharm Stat. 2005;15:913–28.

17. Chen CC, Barnhart HX. Assessing agreement with intraclass correlation coefficient and concordance correlation coefficient for data with repeated measures. Comput Statist Data Anal. 2013;60:132–45.

18. Haber M, Barnhart HX. A general approach to evaluating agreement between two observers or methods of measurement. Stat Methods Med Res. 2008;17(2):151–69.

19. Haber M, Barnhart HX, Song J, Gruden J. Observer variability: a new approach in evaluating interobserver agreement. J Data Sci. 2005;3:69–83.

20. Lin L, Hedayat AS, Wu W. A unified approach for assessing agreement for continuous and categorical data. J Biopharm Stat. 2007;17(4):629–52.

21. Vieira EM, Guedes OA, Morais SD, Musis CR, Albuquerque PA, Borges AH. Prevalence of elongated styloid process in a Central Brazilian population. J Clin Diagn Res. 2015;9(9):ZC90–2.

22. Jung T, Tschernitschek H, Hippen H, Schneider B, Borchers L. Elongated styloid process: when is it really elongated? Dentomaxillofac Radiol. 2004;33(2):119–24.

23. Scaf G, Freitas DQ, Loffredo Lde C. Diagnostic reproducibility of the elongated styloid process. J Appl Oral Sci. 2003;11(2):120–4.

24. Nayak DR, Pujary K, Aggarwal M, Punnoose SE, Chaly VA. Role of three-dimensional computed tomography reconstruction in the management of elongated styloid process: a preliminary study. J Laryngol Otol. 2007;121(4):349–53.

25. Erdoğan S, Orekici Temel G, Selvi H, Ersöz Kaya I. An alternative method used in evaluating agreement among repeat measurement by two raters in education. Educational Sciences: Theory & Practice. 2017;17:217–36.

26. Pan Y, Haber M, Gao J, Barnhart HX. A new permutation-based method for assessing agreement between two observers making replicated quantitative readings. Stat Med. 2012;31(20):2249–61.