CLINICAL STUDY

Reliability of the cervical vertebrae maturation (CVM) method

Predko-Engel A[1], Kaminek M[1], Langova K[2], Kowalski P[1], Fudalej PS[1,3]

Department of Orthodontics, Medical Faculty of Palacký University, Olomouc, Czech Republic.
agnieszkapredko@gmail.com

Abstract: *Objective*: To assess the reliability of the cervical vertebrae maturation method (CVM).
Background: Skeletal maturity estimation can influence the manner and time of orthodontic treatment. The CVM method evaluates skeletal growth on the basis of the changes in the morphology of cervical vertebrae C2, C3, C4 during growth. These vertebrae are visible on a lateral cephalogram, so the method does not require an additional radiograph.
Methods: In this website based study, 10 orthodontists with a long clinical practice (3 routinely using the method – "Routine user – RU" and 7 with less experience in the CVM method – "Non-Routine user – nonRU") rated twice cervical vertebrae maturation with the CVM method on 50 cropped scans of lateral cephalograms of children in circumpubertal age (for boys: 11.5 to 15.5 years; for girls: 10 to 14 years). Kappa statistics (with lower limits of 95% confidence intervals (CI)) and proportion of complete agreement on staging was used to evaluate intra- and inter-assessor agreement.
Results: The mean weighted kappa for intra-assessor agreement was 0.44 (range: 0.30–0.64; range of lower limits of 95% CI: 0.12–0.48) and for inter-assessor agreement was 0.28 (range: –0.01–0.58; range of lower limits of 95% CI: –0.14–0.42). The mean proportion of identical scores assigned by the same assessor was 55.2 % (range: 44–74 %) and for different pairs of assessors was 42 % (range: 16–68 %).
Conclusions: The reliability of the CVM method is questionable and if orthodontic treatment should be initiated relative to the maximum growth, the use of additional biologic indicators should be considered *(Tab. 4, Fig. 1, Ref. 24)*. Text in PDF *www.elis.sk*.
Key words: skeletal age; cervical vertebrae maturation (CVM), reliability of the method.

Introduction

The choice of the moment of initiation of orthodontic therapy may affect the final outcome. For example, skeletal Class II Division 1 malocclusion is frequently associated with hypoplastic mandible, retruded chin (1) and, consequently, unaesthetic profile of the face. As shown in the randomized controlled clinical trials (2, 3, 4), if treatment of Class II is started early, i.e. before pubertal growth spurt, it results mainly in dento-alveolar changes with little alterations of the facial skeleton. This situation is undesirable because the underlying problem is not corrected and facial profile may not sufficiently improve. Some evidence (5, 6) suggested, however, that deferring treatment with functional appliances until a growth spurt might result in a more favourable skeletal response. It has been hypothesized that functional jaw orthopaedics at adolescent growth spurt induced considerably greater skeletal effect than treatment at a preadolescence. Although the RCT is required to definitely prove this hypothesis, it is often recommended that patients with Class II Division 1 commence treatment at the maximal pubertal growth (7).

A practical problem associated with the timing of therapy is an identification of the period of maximum growth. Two popular methods are used in orthodontics: assessment of hand-wrist radiographs (HWR) and evaluation of cervical vertebrae (CVM). In the HWR method, skeletal maturation is determined based on the stages of ossification of the bones of hand and wrist (8, 9), whereas changes of the cervical vertebrae morphology are used in the CVM method (10, 11, 12). Both methods relate maturational stages in the corresponding areas with general and facial growth. Because the CVM method does not require an additional radiograph for assessment, admittedly, it has become the most widely method used by clinicians worldwide.

A method can be used in research and clinics provided it is valid and reliable. Validity refers to an accuracy of the method, i.e. how well the method measures what it should measure. Reliability refers to reproducibility, i.e. how reproducible the findings would be if the same measurement was repeatedly made on the same subject (13). Many studies showed an excellent reproducibility of the CVM method (11, 12, 14, 15, 16). In the investigations by Franchi et al, Baccetti et al, Uysal et al, Lai et al, and Soegiharto et al (11, 12, 14, 15, 16), the observers, who staged a maturation of cervical vertebrae with the CVM method, had a 'research level' knowledge of the CVM system because they participated in the investigations and finally co-authored the publications. In other words, they assessed efficacy of the CVM method. Several recent publications of Gabriel et al, Nestman et al and Zhao et al (17, 18, 19), questioning the superb repeatability of the CVM, seem to

[1]Department of Orthodontics, Medical Faculty of Palacky University, Olomouc, Czech Republic, [2]Institute of Medical Biophysics, Medical Faculty of Palacky University, Olomouc, Czech Republic, and [3]Department of Orthodontics and Dentofacial Orthopedics, University of Bern, Bern, Schwitzerland

Address for correspondence: A. Predko-Engel, MD, Department of Orthodontics, Medical Faculty of Palacky University, Palackeho 12, CZ-772 00 Olomouc, Czech Republic

imply that reliability of the method may be lower when applied by the practice-based orthodontist. A dichotomy between reliability of the method under the ideal conditions and when it is used in everyday, routine practice resembles to some degree the difference between the outcomes of explanatory and pragmatic trials (20, 21). Explanatory (or efficacy) trials are designed to measure the efficacy of an intervention – to find out if the treatment exerts a biologic effect in a research setting under *ideal* conditions. Pragmatic (or effectiveness) trials are designed to measure the effectiveness of an intervention – to explore whether an intervention, as administered in routine practice, is effective in *routine* and *everyday* practice settings (22). Because the treatment results obtained in an explanatory trial are not always confirmed in a pragmatic trial, the latter one may be carried out to finally recommend a new treatment method. Using this perspective, the reliability of the CVM method has not been fully explored. The use of the panels of assessors comprised of practicing orthodontists (17, 18, 19) allowed presenting the reliability scores that are closer to the results seen in everyday practice. However, orthodontists participating in the rating panels received an extensive training prior to scoring sessions. Calibration exercises are rarely available for practice-based orthodontists and, when used in the aforementioned publications, they might have increased reliability coefficients. Therefore, the aim of this study was to evaluate the reliability of the CVM method in assessing cervical vertebrae maturation by a non-calibrated panel of orthodontists with a long clinical practice.

Materials and methods

The patient files at the Department of Orthodontics, Palacký University, Olomouc, Czech Republic, were analysed to identify healthy subjects who met the following inclusion criteria:
– Caucasians,
– circumpubertal age (for boys: 11.5 to 15.5 years; for girls: 10 to 14 years),
– lateral cephalogram with a good representation of 2nd, 3rd, and 4th cervical vertebrae (C2, C3, C3),
– absence of visible pathology of cervical vertebrae. No other inclusion criteria were used.

Fifty randomly selected cephalograms were scanned at 300 dpi resolution. Nowadays , estimation of the digitized images obtained from analogue x-rays is comparatively reliable with an estimation of the original x-rays (23). The images were cropped to visualize only cervical vertebrae; thus, the dentition was not visible on any of the images. Subsequently, the scans were loaded into Power-Point to prepare a presentation for rating. A PowerPoint presentation consisted of a detailed description of the CVM method along with the instructions how to rate, examples of all stages of skeletal maturity, and 50 images to be assessed. The file with presentation was posted on a prepared website.

In this study, a 6-stage modification of the CVM method (12) was used. The summary of the method is given in the Table 1.

Totally, 10 practicing orthodontists (designated as assessors A through J) with the mean practice time of 12.3 years (range from 4 to 27 years) were asked to log into the website and to rate the im-

Table 1. Definitions of the CVM stages (according to Baccetti et al, 2005; Nestman et al, 2012)

Stage	Description
1	The lower borders of all 3 vertebrae (C2–C4) are flat. The bodies of C3 and C4 are trapezoid in shape.
2	A concavity is present at the lower border of C2 in 80 % of cases. The bodies of both C3 and C4 are trapezoid in shape.
3	Concavities at the lower borders of both C2 and C3 are present. The bodies of C3 and C4 can be either trapezoid or rectangular horizontal in shape.
4	Concavities at the lower borders of C2, C3, and C4 now are present. The bodies of C3 and C4 are rectangular horizontal in shape.
5	The concavities at the lower borders of C2, C3, and C4 still are present. At least 1 body of C3 or C4 is square in shape. If not square, the body of the other cervical vertebrae still is rectangular horizontal.
6	The concavities at the lower borders of C2, C3, and C4 still are present. At least 1 body of C3 or C4 is rectangular vertical in shape. If not rectangular vertical, the body of the other cervical vertebrae is square.

ages using the CVM method. Three of 10 orthodontists routinely (i.e. at least once a week) used the CVM method in their practice (*Experienced* assessors: A, B, and E), whereas remaining 7 orthodontists either did not use the CVM or used it less frequently than once a week (*Inexperienced* assessors: C, D, F, G, H, I, and J). The orthodontists assessed the images twice, but the order of images was changed in the second rating session. The interval between the first and the second rating was at least 3 weeks. No calibration was done between the assessors. Instead, each assessor was asked to follow the instructions based on the publications of Baccetti et al (12), which were included in the website presentation.

Statistical analysis
The Spearman's correlation coefficients were computed to evaluate the correlation between the first and second rating of an individual assessor. The kappa statistics was used to assess intra- and inter-assessor agreement in assigning CVM scores. The interpretation of the kappa values was done according to Altman (21), who defined agreement as *poor* for kappa ≤ 0.20; *fair* for kappa from 0.21 to 0.40; *moderate* for kappa from 0.41 to 0.60; *good* for kappa from 0.61 to 0.80; and *very good* for kappa from 0.81 to 1.00.

The one-way random intraclass correlation coefficient (ICC) was used to compare the performance of assessors routinely using the CVM method in private practices (*Experienced* assessors) *versus* those who rarely (or never) use the CVM (*Inexperienced* assessors). All statistical analyses were performed with a software package (SPSS for Windows, version 11.0, SPSS, Chicago, Ill).

Results

Intraobserver agreement
The mean correlation coefficient between the first and second rating was 0.704 (range: 0.54 for observer C to 0.79 for observer A; p < 0.05). The mean weighted kappa was 0.44 (range: 0.30 for assessor G to 0.64 for assessor F) (Tab. 2). Lower limit

Tab. 2. Kappa values for intra-observer agreement (in bold) and inter-observer agreement in the 1st rating session. Lower limit of 95% confidence interval in the brackets.

Observers	A	B	C	D	E	F	G	H	I	J
A	**0.40 (0.24...)**	0.36 (0.20...)	-0.01 (-0.14...)	0.35 (0.18...)	0.45 (0.28...)	0.42 (0.25...)	0.22 (0.06...)	0.38 (0.22...)	0.43 (0.26...)	0.24 (0.08...)
B	–	**0.53 (0.36...)**	0.22 (0.06...)	0.46 (0.29...)	0.35 (0.19...)	0.54 (0.37...)	0.19 (0.03...)	0.47 (0.31...)	0.23 (0.07...)	0.33 (0.16...)
C	–	–	**0.36 (0.20...)**	0.19 (0.04...)	0.17 (0.03...)	0.09 (-0.05...)	0.05 (-0.10...)	0.17 (0.02...)	0.04 (-0.10...)	0.25 (0.10...)
D	–	–	–	**0.36 (0.19...)**	0.20 (0.06...)	0.37 (0.18...)	0.15 (0...)	0.50 (0.33...)	0.25 (0.09...)	0.37 (0.21...)
E	–	–	–	–	**0.43 (0.26...)**	0.35 (0.19...)	0.23 (0.08...)	0.37 (0.21...)	0.25 (0.09...)	0.30 (0.13...)
F	–	–	–	–	–	**0.64 (0.48...)**	0.30 (0.13...)	0.58 (0.42...)	0.21 (0.04...)	0.23 (0.06...)
G	–	–	–	–	–	–	**0.30 (0.12...)**	0.25 (0.08...)	0.17 (0...)	0.27 (0.10...)
H	–	–	–	–	–	–	–	**0.56 (0.44...)**	0.23 (0.07...)	0.40 (0.22...)
I	–	–	–	–	–	–	–	–	**0.37 (0.21...)**	0.13 (-0.02...)
J	–	–	–	–	–	–	–	–	–	**0.43 (0.25...)**

Tab. 3. Percentage of *identical* scores assigned by the same observer in two rating sessions (in bold) and by different pairs of observers in the 1st rating session.

Observers	A	B	C	D	E	F	G	H	I	J
A	**50**	48	16	48	54	54	36	50	54	38
B	–	**62**	36	58	46	64	34	58	38	46
C	–	–	**48**	34	30	24	22	32	20	40
D	–	–	–	**50**	34	52	32	62	40	52
E	–	–	–	–	**52**	46	36	48	38	42
F	–	–	–	–	–	**74**	44	68	38	40
G	–	–	–	–	–	–	**44**	40	32	42
H	–	–	–	–	–	–	–	**68**	38	54
I	–	–	–	–	–	–	–	–	**48**	30
J	–	–	–	–	–	–	–	–	–	**56**

Tab. 4. Mean percentage intra-assessor agreement (out of total 500 intra-assessor observations) and inter-assessor agreement (out of total 2250 inter-assessor observations) on CVM staging.

Cervical staging	Intra-assessor	Inter-assessor
Complete agreement	55.2	42.0
1 stage apart	27.4	36.5
2 stages apart	12.2	14.8
3 stages apart	2.6	4.1
4 stages apart	2.2	2.0
5 stages apart	0.4	0.6

of 95% confidence interval for kappa was from 0.12 (assessor G) to 0.48 (assessor H). *Experienced* assessors demonstrated similar intra-assessor agreement as *Inexperienced* ones: 0.45 versus 0.44, respectively.

The mean percentage of identical scores assigned by the same assessor was 55.2 % (range: 44 % for assessor G to 74 % for assessor F) (Tab. 3). In 44.8 % observations, there were discrepancies between the first and second ratings (Tab. 4). Although most differences were 1 stage apart, more than 17 % divergences were 2 or more stages apart. Occasionally, disagreement was particularly large reaching 5 stages apart. *Experienced* assessors showed 54.7 % agreement, whereas *Inexperienced* assessors demonstrated 55.4 % agreement between the first and second scoring.

Interobserver agreement

The mean weighted kappa was 0.28 (range: –0.01 for assessors A vs C to 0.58 for assessors F vs H) (Tab. 2). Lower limit of 95% confidence interval for kappa was from –0.14 (assessor A vs C) to 0.42 (assessor F vs H).

The mean percentage of identical scores assigned by a pair of assessors was 42 % (range: 16 % for assessors A vs C to 68 % for assessors F vs H) (Tab. 3). In 58 % observations, there were discrepancies between the assessors (Tab. 4). Most disagreements (36.5 %) were 1 stage apart but more than 20 % divergences were 2 or more stages apart. Occasionally, assessors differed considerably among themselves. For example, Figure 1 presents the image that was rated as 1, 2, 5, or 6 by different assessors.

Experienced versus Inexperienced assessors

Experienced assessors demonstrated a higher consistency in scoring than their Inexperienced colleagues – ICC for the former group was 0.796 (95% CI: 0.697–0.871) and for the latter group was 0.617 (95% CI: 0.506–0.727). The difference was statistically significant (p = 0.011).

Discussion

The aim of this study was to evaluate reliability of the CVM method. Reliability is a measure of the reproducibility, i.e. the extent to which a method gives consistent results that. Thus, if the method is reliable, results obtained by the same observer on various occasions or by different observers should be similar (13, 24).

Our findings suggested that reliability of the CVM method was modest. The interpretation of the mean kappa according to Altman (21) implied that the intra-assessor concordance (kappa = 0.44) was *moderate* and inter-assessor concordance (kappa = 0.28) was *fair*. Also, the range of lower limits of 95% confidence interval for both the intra- and inter-assessor agreement underscores a

Fig. 1. Example of the image used for staging with the Cervical Vertebrae Maturation (CVM) method.

relatively low reproducibility of the CVM ratings. Furthermore, the mean percentage intra- and inter-assessor complete agreement on staging (55.2 % and 42 %, respectively) demonstrated a considerable discordance within and between the observers. Overall, the present results supported the findings of Gabriel et al (17) and Zhao et al (19). These studies aimed specifically to evaluate the reproducibility of the CVM stage determination and used similar methodology as in our investigation. Both research groups used the rating panels comprising orthodontists with a long clinical experience who did not participate in selection/preparation of images used for the CVM staging. In contrast with our study, however, the assessors in the investigations by Gabriel et al and Zhao et al were calibrated in the use of the CVM method before the rating sessions (17, 19). The mean intra-assessor complete agreement on staging was 62.3 % and the mean intra-assessor kappa was 0.60 in the work of Gabriel et al (17) and 56.9 % and 0.68, respectively, in the study by Zhao et al (19). For the inter-assessor agreement, there were 45 % complete agreements in the 1st rating session and 49 % in the 2nd session in the study by Gabriel et al (17) and 39 % (1st session) and 45 % (2nd session) in the study by Zhao et al (kappa for inter-assessor agreement was not reported) (19). A poorer concordance observed in the current study in comparison with aforementioned works likely results from the lack of pre-rating calibration exercises. We purposefully did not plan the training for assessors to mimic a typical situation at the practice where an orthodontist using the CVM method can only rely on his/her own experience or can refer to publications explaining the CVM system. Thus, in our opinion, the present results relatively closely reflect

the reliability of the CVM method in the "real" clinical setting.

Several publications, however, reported a high reliability of the CVM method (11, 12, 14, 15, 16). All these research groups using the various statistical methods demonstrated almost perfect intra- and/or inter-observer reproducibility of the CVM staging. For example, Franchi et al (11) reported 100 % intra- and 98.6 % inter-operator agreement; Baccetti et al (12) demonstrated 96.7 % inter-assessor concordance in the CVM staging; Soegiharto et al (16) showed with the aid of kappa statistics a very high reproducibility of the CVM method (kappa from 0.85 to 0.97, depending on sex and race). A possible explanation of the difference between findings of our group and other researchers is the separation of two research phases in the current investigation: preparations of cephalograms for rating session and rating itself. In our study, a researcher involved in selection of radiographs, cropping the images, and preparation of a Power Point presentation, did not participate in the rating session. In contrast, the authors of aforementioned publications selected radiographs, made them ready for the CVM staging and also themselves rated them and performed intra- and inter-assessor agreement tests. A participation of a future assessor in preparation of the rating probably leads to a situation where he/she can extensively familiarize with morphology of cervical vertebrae on cephalograms to be rated and he/she may develop a lasting impression of which CVM stage is to be assigned later on. Even a considerable washout period may not be sufficient to cancel this memory effect. Furthermore, if two or more authors prepare the rating sessions, mutual presentations of radiographs with not-clear-cut morphology of cervical vertebrae, explanations, and discussions are very likely. This may also affect inter-assessor reliability assessment if the panel of independent judges is not used for the CVM staging.

Another possible cause for a high reproducibility of the CVM method is the wide age range in the assessed samples (14, 15, 16). In a situation when the sample comprises patients aged, for example, 5 to 24 years (14) or 8 to 18 years by Lai et al (15), there are many both very young subjects with immature cervical vertebrae and postadolescent patients with mature cervical vertebrae. Because both extreme maturational stages (i.e. very immature *vs* very mature) are relatively easy to identify, intra-assessor and inter-assessor agreement can be excessively high.

We found that clinicians using routinely the CVM method (*Experienced*) performed better, i.e. were more consistent, than clinicians with little or no experience in the use of the CVM method (*Inexperienced*) – the ICC for the former group was significantly higher than for the latter group. The ICC gives a composite picture of intra-observer and inter-observer variability. As a result, if an inter-group difference was found, it might be due to lower intra-assessor or inter-assessor agreement in one of the groups. In our study, intra-assessor agreement among *Experienced* observers was comparable with the intra-assessor agreement among *Inexperienced* observers (kappa was 0.45 and 0.44 for *Experienced* and *Inexperienced* assessors, respectively). Therefore, it is likely that better consistency in the *Experienced* group resulted from higher inter-assessor agreement there. The comparison of the mean kappa for the inter-assessor concordance among *Experienced* assessors

(i.e. A vs B, A vs E, and B vs E; kappa = 0.39) and *Inexperienced* assessors (i.e. C vs D, C vs F, C vs G, etc.; kappa = 0.25) seemed to confirm this hypothesis.

Baccetti et al (12) presented five prerequisites that ideal indicator of skeletal maturity must possess. Among others, they listed "Consistency in the interpretation of the data. The inter-examiner error in the appraisal of the defined stages or phases should be as low as possible". Our findings imply that the requirement of high consistency is not fulfilled during application of the CVM method by an orthodontist with a long clinical practice.

Conclusions

1) Reliability of the CVM method is questionable. The mean proportion of identical scores assigned by the same assessor was 55.2 % (range: 44–74 %). For different pairs of assessors it was 42% (range: 16–68 %).

2) If orthodontic treatment should be initiated relative to maximum growth, the use of additional biologic indicators should be considered.

References

1. McNamara JA. Components of Class II malocclusion in children 8–10 years of age. Angle Orthodont 1981; 51: 177–202.

2. Tulloch JF, Phillips C, Koch G, Proffit WR. The effect of early intervention on skeletal pattern in Class II malocclusion: a randomized clinical trial. Amer J Orthodont Dentofacial Orthop 1997; 111: 62–74.

3. O'Brien K. Is early treatment for Class II malocclusion effective? Results from a randomized controlled trial. Amer J Orthodont dentofacial Orthop 2006; 129 (Suppl 4): 64–65.

4. O'Brien K, Wright J, Conboy F, Sanjie Y, Mandall N, Chadwick S, Connolly I, Cook P, Birnie D, Hammond M, Harradine N, Lewis D, McDade C, Mitchell L, Murray A, O'Neill J, Read M, Robinson S, Roberts-Harry D, Sandler J, Shaw I. Effectiveness of treatment for Class II malocclusion with the Herbst or twin-block appliances: a randomized, controlled trial. Amer J Orthodont dentofacial Orthop 2003; 124: 128–137.

5. Pancherz H, Hägg U. Dentofacial orthopedics in relation to somatic maturation. An analysis of 70 consecutive cases treated with the Herbst appliance. Amer J Orthodont dentofacial Orthop 1985; 88: 273–287.

6. Baccetti T, Franchi L, Kim LH. Effect of timing on the outcomes of 1–phase nonextraction therapy of Class II malocclusion. Amer J Orthodont dentofacial Orthop 2009; 136: 501–509.

7. McNamara JA, McNamara L, Graber LW. Optimizing orthodontic and Dentofacial orthopedic treatment timing. 477–516. In: Graber LW, Vanarsdall RL Jr., Vig KW (Editors). Orthodontics: Current Principles and Techniques. 5th Edition, Philadelphia ; Elsevier Mosby, 2012.

8. Houston WJB, Miller JC, Tanner JM. Prediction of the timing of the adolescent growth spurt from ossification events in hand–wrist films. Brit J Orthodont 1979; 6: 145–152.

9. Fishman LS. Radiographic evaluation of skeletal maturation. A clinically oriented method based on hand–wrist films. Angle Orthodont 1982; 52: 88–112.

10. Lamparski DG. Skeletal age assessment utilizing cervical vertebrae. Thesis, University of Pittsburgh, 1972.

11. Franchi L, Baccetti T, McNamara JA. Mandibular growth as related to cervical vertebral maturation and body heigh. Amer J Orthodont dentofacial Orthop 2000; 118: 335–340.

12. Baccetti T, Franchi L, McNamara JA. The Cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. Seminars in Orthodontics 2005; 11: 119–129.

13. Dawson B, Trapp RG. Basic and clinical biostatistics. Lange Medical Books/McGraw–Hill. Medical Publishing Division. New York; 2004.

14. Uysal T, Ramoglu SI, Basciftci FA, Sari Z. Chronologic age and skeletal maturation of the cervical vertebrae and hand– wrist: is there a relationship? Amer J Orthodont Dentofacial Orthop 2006; 130: 622–628.

15. Lai EH, Liu JP, Chang JZ, Tsai SJ, Yao CC, Chen MH, Chen YJ, Lin CP. Radiographic assessment of skeletal maturation stages for orthodontic patients: hand–wrist bones or cervical vertebrae? J Formos Med Assoc 2008; 107: 316–25.

16. Soegiharto BM, Cunningham SJ, Moles DR. Skeletal maturation in Indonesian and white children assessed with hand–wrist and cervical vertebrae methods. Amer J Orthodont Dentofacial Orthop 2008; 134: 217–26.

17. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RD, Southard TE. Cervical vertebrae maturation method: poor reproducibility. Amer J Orthodont Dentofacial Orthop 2009; 136: 478–480.

18. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphologic criteria: Poor reproducibility. Amer J Orthodont Dentofacial Orthop 2011; 140: 182–188.

19. Zhao XG, Lin J, Jiang JH, Wang Q, Ng SH. Validity and reliability of a method for assessment of cervical vertebral maturation. Angle Orthodont 2012; 82: 229–234.

20. Cook JA. The challenges faced in the design, conduct and analysis of surgical randomised controlled trials. 2009; Trials.10: 9. doi: 10.1186/1745–6215–10–9.

21. Altman DG. 404–408. In: Practical Statistics for Medical Research. London: Chapman and Hall, 1991.

22. Farrokhyar F, Karanicolas PJ, Thoma A, Simunovic M, Bhandari M, Devereaux PJ, Anvari M, Adili A, Guyatt G. Randomized controlled trials of surgical interventions. Ann Surg 2010; 251: 409–416.

23. Thurzo A, Javorka V, Stanko P, Lysy J, Suchancová B, Lehotska V, Valkovic L, Makovnik M. Digital and manual cephalometric analysis. Bratisl Med J 2010; 111 (2): 97–100.

24. Higgins PA, Straub AJ. Understanding the error of our ways: mapping the concepts of validity and reliability. Nurs Outlook 2006; 54: 23–29.