

Validity and reliability of a method for assessment of cervical vertebral maturation

Xiao-Guang Zhao^a; Jiuxiang Lin^b; Jiu-Hui Jiang^c; Qingzhu Wang^a; Sut Hong NG^a

ABSTRACT

Objective: To evaluate the validity and reliability of the cervical vertebral maturation (CVM) method with a longitudinal sample.

Materials and Methods: Eighty-six cephalograms from 18 subjects (5 males and 13 females) were selected from the longitudinal database. Total mandibular length was measured on each film; an increased rate served as the gold standard in examination of the validity of the CVM method. Eleven orthodontists, after receiving intensive training in the CVM method, evaluated all films twice. Kendall's W and the weighted kappa statistic were employed.

Results: Kendall's W values were higher than 0.8 at both times, indicating strong interobserver reproducibility, but interobserver agreement was documented twice at less than 50%. A wide range of intraobserver agreement was noted (40.7%–79.1%), and substantial intraobserver reproducibility was proved by kappa values (0.53–0.86). With regard to validity, moderate agreement was reported between the gold standard and observer staging at the initial time (kappa values 0.44–0.61). However, agreement seemed to be unacceptable for clinical use, especially in cervical stage 3 (26.8%).

Conclusions: Even though the validity and reliability of the CVM method proved statistically acceptable, we suggest that many other growth indicators should be taken into consideration in evaluating adolescent skeletal maturation. (*Angle Orthod.* 2012;82:229–234.)

KEY WORDS: Growth and development; Cervical vertebral maturation; Reliability and validity

INTRODUCTION

Correction of malocclusion and treatment prognosis are strongly influenced by growth. The most favorable time to address many orthodontic problems with skeletal manifestations is during the period of pubertal growth. Assessment of skeletal maturation is essential in planning individual orthodontic treatment because

of marked individual variations in timing, duration, and intensity of pubertal growth.

Lamparski¹ noted that maturational changes in the cervical vertebrae on routine lateral cephalograms could be used as indicators to measure biologic age without additional radiographs. He developed the first method to assess skeletal age. Thereafter, several cervical vertebral maturation (CVM) methods were proposed, and some modifications were made and improvements achieved.^{2–5} However, all these methods depend on the observer's subjective judgment of morphologic changes in cervical vertebral bodies.¹ The purely subjective assessment is easy to learn and to carry out in clinic. However, the subjectivity will cause significant errors in clinical usage in terms of validity and reliability.

The CVM method proposed by Baccetti et al.⁵ in 2005 was used worldwide in orthodontic clinics. Two characteristics of this method set it apart from others. One is that longitudinal data from radiographs were used in its development while other CVM methods used cross-sectional materials. The other is that it was established according to changes in the annual rate of

^a Postgraduate student, Department of Orthodontics, Peking University School and Hospital of Stomatology, Beijing, China.

^b Professor, Department of Orthodontics, Peking University School and Hospital of Stomatology, Beijing, China.

^c Associate Professor, Department of Orthodontics, Peking University School and Hospital of Stomatology, Beijing, China.

Corresponding author: Dr Jiuxiang Lin, Department of Orthodontics, Peking University School and Hospital of Stomatology, 22 Zhongguancun Nandajie, Haidian District, Beijing, 100081 P.R. China
(e-mail: jxlin@pku.edu.cn)

Accepted: July 2011. Submitted: May 2011.

Published Online: August 29, 2011

© 2012 by The EH Angle Education and Research Foundation, Inc.

increase in total mandibular length (Co-Gn) instead of by comparison with hand-wrist skeletal age.

The validity of Baccetti's method⁵ has not been proved with a proper gold standard. Because this method was established on the basis of longitudinal yearly changes in mandibular length, similar series of materials should be available and should serve as the gold standard. In terms of its reliability, some controversy continues.^{6,7} Therefore, the objective of this study was to evaluate the validity and reliability of the CVM method in skeletal maturation assessment with longitudinal cephalograms.

MATERIALS AND METHODS

Longitudinal samples were obtained from the Research Center on Craniofacial Growth and Development at Peking University. More than 900 orthodontically untreated subjects participated in this longitudinal project from 1990 through 1996. Sequential lateral cephalograms were taken at yearly intervals for consecutive years. Informed consent was obtained from all subjects and their parents. The study protocol was reviewed and approved by the institutional review board of Peking University. General selection criteria for the project were put forth in the previous study.⁸

In Baccetti's study,⁵ the maximum annual increase in Co-Gn over 6 consecutive years was defined as the peak of mandibular growth at puberty. The cephalogram before the peak mandibular growth was defined as cervical stage 3 (CS3), and the one after the peak was defined as CS4.⁵ Therefore, to test its validity, maximum mandibular annual growth must be found in our longitudinal sample, and the increase in the rate of mandibular total length should serve as the gold standard. Inclusion criteria were as follows: (1) The subject was younger than 10 years when his first cephalogram was taken; (2) at least four cephalograms obtained over consecutive years were available; (3) the two consecutive cephalograms with maximum annual Co-Gn increase were located in the middle of the longitudinal cephalograms of an individual subject; in another case if the maximum annual Co-Gn increase was noted during the first year, it was greater than 5 mm and had to exceed the following period of yearly incremental growth by at least 2 mm^{5,9-11}; and (4) the lateral cephalograms at each period of time were of high clarity and good contrast. Finally, the longitudinal sample was limited to 18 subjects (5 males and 13 females) with 86 lateral cephalograms. The average age of the initial visit was 9 years 2 months.

Because the same cephalometer was used throughout this study, the magnification for each radiograph was identical. No corrections for magnification were made in linear measurement of cephalometric analysis.

Mandibular length (Go-Gn) was measured directly on the cephalograms with a micrometer caliper, accurate to within 0.01 mm, to detect the change in mandibular growth velocity. Gold standards of CVM staging were defined for each subject. The cephalogram before the peak mandibular growth was defined as CS3. Films taken before the CS3 at yearly intervals were CS2 and CS1; films taken thereafter were CS4, CS5, and CS6, respectively.

All lateral cephalograms were digitized to uncompressed TIFF images using a professional dental film scanner (VIDAR Dental Film Digitizer, Vidar Systems Corporation, Herndon, Va) at 300-dpi optical resolution and 8-bit depth grayscale to ensure high imaging quality. Eleven orthodontists with an average of 17.2 years (range, 13–21 years) of clinical experience were invited as observers to stage the 86 digital cephalograms. All were from the Department of Orthodontics at Peking University School and Hospital of Stomatology and did not participate in the design of this study nor in construction of the sample. Each observer was given sufficient training in the CVM method by one investigator (Dr Zhao). The CVM method was explained to observers, and 20 clinical adolescent cephalograms were used to practice the CVM method. Then a hard copy of the schematic representation of the CVM method and the two samples of each stage in Baccetti's original article⁵ were distributed to every observer as references. Digital films were sequenced randomly, and the same copy was provided for observers to judge. After a 4-week washout period, observers were asked to stage the resequenced films.

Statistical Analysis

Because cervical staging is an ordinal variable, the Kendall coefficient of concordance (Kendall's W) and the weighted kappa were employed to assess agreement.⁶

The reliability of the CVM method is based on agreement among observers (interobserver agreement) at each time and agreement between two times for each observer (intraobserver agreement). Kendall's W, ranging from 0 (no agreement) to 1 (complete agreement), was used to assess interobserver agreement. Intermediate values of W indicate a greater or lesser degree of agreement among the various responses of observers. Intraobserver agreement was calculated through weighted kappa analysis. A kappa value respectively indicates poor (≤ 0), slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00) agreement.¹² The validity of the CVM method was represented by agreement between the gold standard and estimated staging for the initial time if

Table 1. Kendall's W for Interobserver Agreement on Cervical Staging at the Initial Time and 1 Month Later

	Kendall's W	P	Statistical Level of Agreement
Initial	0.831	< .0001	Strong
One month later	0.838	< .0001	Strong

intraobserver agreement was acceptable. This was also calculated using the weighted kappa statistic. A *P* value less than .05 was considered significant. Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS), version 16.0 (SPSS Inc, Chicago, Ill) and Stata 9.0 (StataCorp, College Station, Tex).

In terms of the linear measurement of mandibular length, 20 randomly selected cephalograms were remeasured 2 weeks later by the same investigator. Method errors were assessed using Dahlberg's formula, and systematic errors were ascertained using paired *t*-tests, similar to the recommendations of Houston.¹³ Method errors did not exceed 0.2 mm and were negligible. Paired *t*-tests demonstrated no statistically significant differences in measurements (*P* > .05).

RESULTS

Table 1 depicts the agreement of 11 observers on cervical staging at two times with the use of Kendall's W statistic. For 86 adolescent cephalograms, Kendall's W values were 0.831 for the initial time and 0.838 a month later. Both values were greater than 0.8, which showed strong statistical agreement among observers regarding cervical staging. However, high Kendall's W values do not support good percent agreement. A total of 4730 interobserver observations were compared. In Table 2, the percent agreement is shown as 39.3% (1858/4730) for the initial time and 44.9% (2122/4730) 1 month later. Percent disagreement values one stage apart were 42.8% (2026/4730) and 41.7% (1974/4730), respectively; this was similar to the percent agreement. In other words, in the assessment 1 month later, observers' chances to judge the same cervical stage as the first time were similar to judging a previous or a later stage by mistake. Percent dis-

Table 2. Descriptive Analysis of Interobserver Agreement and Disagreement on Cervical Staging at the Initial Time and 1 Month Later

Cervical Staging	Initial	One Month Later
Complete agreement	39.3% (1858/4730)	44.9% (2122/4730)
One stage apart	42.8% (2026/4730)	41.7% (1974/4730)
Two stages apart	14.3% (678/4730)	10.8% (513/4730)
Three stages apart	2.9% (138/4730)	2.1% (98/4730)
Four stages apart	0.5% (25/4730)	0.4% (21/4730)
Five stages apart	0.1% (4/4730)	0% (2/4730)

Table 3. Weighted Kappa Coefficients for Intraobserver Agreement Between the Initial Time and the 1 Month Time

Observer	Weighted Kappa Coefficient	Standard Error
1	0.54	0.069
2	0.74	0.068
3	0.53	0.068
4	0.86	0.074
5	0.76	0.068
6	0.59	0.065
7	0.62	0.072
8	0.69	0.071
9	0.82	0.071
10	0.64	0.060
11	0.64	0.068

agreement at more than two stages apart was reduced significantly.

With regard to intraobserver agreement between the two time points, values of weighted kappa coefficients ranged from 0.53–0.86 (Table 3), statistically indicating substantial intraobserver agreement. This suggests that cervical staging by the CVM method can be repeatedly assessed with similar results and is thus reliable.

Table 4 shows that the intraobserver percent agreement of each observer ranged from 40.7% (35/86) to 79.1% (68/86). The wide range indicates great variation in intraobserver agreement among different orthodontists. Table 5 presents overall intraobserver percent agreement and disagreement when data provided by 11 observers are combined. A total of 946 intraobserver observations were analyzed. The 56.9% (538/946) intraobserver agreement was greater than the 34.4% (326/946) one-stage-apart disagreement.

Because statistically substantial intraobserver agreement existed between evaluations at the two times, indicating no significant difference between them, only the results of staging at the initial time were examined in the evaluation of validity. For 11 observers, the values of weighted kappa coefficients ranged from 0.44–0.61 (Table 6), which indicates

Table 4. Intraobserver Agreement Percentages Between Initial Time and 1 Month Time

Observer	Agreement Frequency	Agreement Percentage
1	40/86	46.5
2	55/86	64.0
3	35/86	40.7
4	68/86	79.1
5	62/86	72.1
6	45/86	52.3
7	41/86	47.7
8	44/86	51.2
9	60/86	69.8
10	46/86	53.5
11	42/86	48.8

Table 5. Descriptive Analysis of Intraobserver Agreement and Disagreement on Cervical Staging Between Initial Time and 1 Month Later

Cervical Staging	Agreement Percentage
Complete agreement	56.9% (538/946)
One stage apart	34.4% (326/946)
Two stages apart	6.9% (65/946)
Three stages apart	1.6% (15/946)
Four stages apart	0.2% (2/946)

moderate agreement. Therefore, the validity of the CVM method is statistically acceptable.

Table 7 shows the percent agreement and stage-apart disagreement of validity at various cervical vertebral stages. A relatively higher percent agreement is seen at cervical stage 6 (54.5%) of the gold standard, and the least agreement is noted at cervical stage 3 (26.8%), followed by cervical stage 2 (34.3%). Larger one-stage-apart disagreement was found at cervical stage 2 (51.7%), stage 3 (49.5%), and stage 5 (51.0%). Moreover, two-stage-apart disagreement at cervical stage 1 (25.8%), stage 3 (22.2%), and stage 4 (20.2%) of the gold standard is a little higher compared with the others.

DISCUSSION

Longitudinal research is an essential method in the study of the growth and development of the body and the face.^{1,14–16} When the first CVM method was developed in 1972, Lamparski¹ suggested that a longitudinal sample should be employed to eliminate much of the individual variation that was present in his study. The CVM method examined in this study was proposed through longitudinal data with reference to mandibular growth.⁵ Therefore, to test the validity of this CVM method, longitudinal samples were used, and the yearly mandibular growth rate served as the gold standard, instead of the judgment of experts as suggested in another study.⁷

A limitation of this study was that consecutive cephalograms for each subject numbered fewer than six. Because the CVM method consists of six stages, ideally a longer period of longitudinal data should be used. Because of the limited source, only 18 subjects, each with more than four cephalograms, were finally enrolled.

The reliability of the CVM method was tested separately by interobserver and intraobserver agreement. Strong agreement was found at both times (Kendall's W values >0.8). The result was a little better than the finding in the previous study.⁶ A reasonable explanation as suggested by Baccetti et al.¹⁷ was that the observers had received more intensive training and practice. However, interobserver percent agreement among observers was still below 50%; this was similar

Table 6. Weighted Kappa Coefficients for Agreement Between the Gold Standard and Observers' Cervical Staging at Initial Time

Observer	Weighted Kappa Coefficient	Standard Error
1	0.44	0.066
2	0.55	0.067
3	0.46	0.062
4	0.55	0.066
5	0.57	0.067
6	0.61	0.066
7	0.52	0.070
8	0.51	0.066
9	0.50	0.065
10	0.55	0.065
11	0.50	0.069

to the results of the previous study.⁶ This meant that observers in this study had more opportunity to disagree with cervical staging as assessed by others. With respect to intraobserver agreement, the weighted kappa values showed substantial agreement, but the average percent agreement was only 56.9% with a wide range (40.7%–79.1%). This suggested that 1 month later, the observers had nearly half the opportunity to arrive at a different conclusion in cervical vertebral assessment, and great variation was noted in the response of observers.

The validity of this method was proved statistically by the moderate agreement between the gold standard and observers' staging at the initial time in this study, but the percent agreement of various stages seemed unacceptable for clinical use. Our area of greatest concern in skeletal evaluation is whether the adolescent pubertal growth spurt has arrived. Therefore, cervical stage 3, which represents the beginning of pubertal growth, is the most meaningful in CVM assessment. However, much to our disappointment, agreement for cervical stage 3 was only 26.8%, which is the lowest value among the six cervical stages.

The accuracy and reproducibility of the CVM method leave much to be desired. Factors that affect them are listed as follows.

- 1) The "measurements" are purely subjective.¹ The transition in the shape of cervical vertebral bodies is a consecutive and gradual process. It is difficult to exactly define and identify the gradual appearance of the concavity in the inferior border of the vertebral bodies. By definition, a square shape has equal length and height of the vertebral body. An arbitrarily set border of what is considered square must be applied. So, the difference between horizontally rectangular, square, and vertically rectangular shapes depends on the researcher's arbitrary decision.¹⁸
- 2) The shapes of cervical vertebrae show marked variation from subject to subject. Sometimes the

Table 7. Descriptive Analysis of Accuracy on Observers' Cervical Staging According to the Gold Standard

Cervical Staging	Gold Standard					
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Complete agreement	39.4% (26/66)	34.3% (49/143)	26.8% (53/198)	40.9% (81/198)	37.9% (75/198)	54.5% (78/143)
One stage apart	33.3% (22/66)	51.7% (74/143)	49.5% (98/198)	34.8% (69/198)	51.0% (101/198)	30.8% (44/143)
Two stages apart	25.8% (17/66)	13.3% (19/143)	22.2% (44/198)	20.2% (40/198)	8.6% (17/198)	13.3% (19/143)
Three stages apart	1.5% (1/66)	0	1.5% (3/198)	4.0% (8/198)	2.0% (4/198)	1.4% (2/143)
Four stages apart	0	0.7% (1/143)	0	0	0.5% (1/198)	0

shape and inferior borders of C2–C4 cannot fulfill the definition of a cervical stage at the same time. For example, in some cases (eg, subject #17 in this study), the shape of the C3 body has changed to horizontally rectangular or even square, but a distinct concavity of the inferior border of C3 does not appear. Furthermore, we noted in the clinic that the shapes of C3 and C4 bodies were still horizontally rectangular in some adult patients (Figure 1).

- 3) Could the mandibular growth rate be taken as a reference to establish the CVM method? When most CVM methods^{1–3,19,20} were developed, except those proposed by Baccetti et al.,^{4,5} hand-wrist skeletal ages (stages) were used as the gold standard to measure the general growth rate. Controversy about the relationship between facial (mandibular) growth and general bodily growth continues. Some studies^{9,21,22} support a high degree

of association between them. Other studies^{23,24} show a low degree of association. Scammon's growth curve²⁵ of different tissues may explain in part the weak association. Craniofacial growth is affected not only by the general bodily curve but also by the neural growth curve. Moreover, Mitani²⁴ found that the amount and timing of mandibular growth seem to be more variable than the other areas he studied. It is well known that evaluation of skeletal maturation for patients before orthodontic treatment is conducted to judge whether tooth movement and skeletal modification will be affected by their growth and development. Even though mandibular growth plays an important role in adolescent orthodontic treatment, the response of other parts of the face to adolescent growth should be taken into consideration. Therefore, we believe there is no positive answer yet to the question of whether the mandibular growth rate can be used as a reference to establish the CVM method. If it cannot be used, the validity of this method will be unreliable.

Although evaluating skeletal maturation with the CVM method can provide some useful information for orthodontic treatment planning, considerable variation in skeletal maturation and timing of craniofacial growth at puberty indicates that this method should be used only in conjunction with other indicators such as overall bodily growth, sexual maturation, and so forth. It must be remembered that errors can occur when one diagnostic test is relied on too heavily.²

CONCLUSIONS

- The reliability of the CVM method was proved statistically with strong interobserver agreement and substantial intraobserver agreement. However, the percent interobserver agreement was below 50% at both times. Percent intraobserver agreement varied widely among observers (range, 40.7%–79.1%).
- The validity of the CVM method was also supported statistically, with moderate agreement between the gold standard and observer staging at the initial time. However, percent agreement, especially at cervical stage 3, left much to be desired.
- The CVM method should be used with other growth indicators in the evaluation of skeletal maturation.



Figure 1. A clinical cephalogram of a 32-year-old male with horizontally rectangular C4 and square C3.

REFERENCES

1. Lamparski D. *Skeletal Age Assessment Utilizing Cervical Vertebrae* [thesis]. Pittsburgh, Pa: University of Pittsburgh; 1972.
2. Hassel B, Farman AG. Skeletal maturation evaluation using cervical vertebrae. *Am J Orthod Dentofacial Orthop.* 1995; 107:58–66.
3. San Roman P, Palma JC, Oteo MD, Nevado E. Skeletal maturation determined by cervical vertebrae development. *Eur J Orthod.* 2002;24:303–311.
4. Baccetti T, Franchi L, McNamara JA Jr. An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. *Angle Orthod.* 2002;72: 316–323.
5. Baccetti T, Franchi L, McNamara JA Jr. The cervical vertebral maturation (CVM) method for the assessment of optimal treatment timing in dentofacial orthopedics. *Semin Orthod.* 2005;11:119–129.
6. Gabriel DB, Southard KA, Qian F, Marshall SD, Franciscus RG, Southard TE. Cervical vertebrae maturation method: poor reproducibility. *Am J Orthod Dentofacial Orthop.* 2009; 136:478.e471–e477; discussion 478–480.
7. Ballrick J, Fields H, Vig K, Beck F, Germack J, Baccetti T. Reliability and validity of cervical vertebral maturation and hand-wrist radiographs. In: *Proceedings of the 83rd General Session of the IADR/AADR/CADR.* Baltimore, MD. 2005:9–12. Available at: http://iadr.confex.com/iadr/2005Balt/techprogram/abstract_62129.htm. Accessed August 15, 2011.
8. Chen L, Liu J, Xu T, Lin J. Longitudinal study of relative growth rates of the maxilla and the mandible according to quantitative cervical vertebral maturation. *Am J Orthod Dentofacial Orthop.* 2010;137:736.e731–e738.
9. Bergersen EO. The male adolescent facial growth spurt: its prediction and relation to skeletal maturation. *Angle Orthod.* 1972;42:319–338.
10. Gu Y, McNamara JA. Mandibular growth changes and cervical vertebral maturation: a cephalometric implant study. *Angle Orthod.* 2007;77:947–953.
11. Franchi L, Baccetti T, McNamara JA Jr. Mandibular growth as related to cervical vertebral maturation and body height. *Am J Orthod Dentofacial Orthop.* 2000;118:335–340.
12. Blackman NJM, Koval JJ. Interval estimation for Cohen's kappa as a measure of agreement. *Stat Med.* 2000;19: 723–741.
13. Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod.* 1983;83:382–390.
14. Scammon RE. The first seriatim study of human growth. *Am J Phys Anthropol.* 1927;10:329–336.
15. Moyers RE. *Handbook of Orthodontics.* 4th ed. Chicago, Ill: Year Book Medical Publishers; 1988.
16. Franchi L, Baccetti T, McNamara JA Jr. Postpubertal assessment of treatment timing for maxillary expansion and protraction therapy followed by fixed appliances. *Am J Orthod Dentofacial Orthop.* 2004;126:555–568.
17. Baccetti T, Franchi L, McNamara JA Jr. Reproducibility of the CVM method: a reply. *Am J Orthod Dentofacial Orthop.* 2010;137:446–447.
18. Fudalej P, Bollen AM. Effectiveness of the cervical vertebral maturation method to predict postpeak circumpubertal growth of craniofacial structures. *Am J Orthod Dentofacial Orthop.* 2010;137:59–65.
19. Chen LL, Xu TM, Jiang JH, Zhang XZ, Lin JX. Quantitative cervical vertebral maturation assessment in adolescents with normal occlusion: a mixed longitudinal study. *Am J Orthod Dentofacial Orthop.* 2008;134:720.e721–e727.
20. Chatzigianni A, Halazonetis DJ. Geometric morphometric evaluation of cervical vertebrae shape and its relationship to skeletal maturation. *Am J Orthod Dentofacial Orthop.* 2009; 136:481.e481–489; discussion 481–483.
21. Singh IJ, Savara BS, Miller PA. Interrelations of selected measurements of the face and body in pre-adolescent and adolescent girls. *Growth.* 1967;31:119–131.
22. Rose GJ. A cross-sectional study of the relationship of facial areas with several body dimensions. *Angle Orthod.* 1960;30: 6–13.
23. Howells WW. A factorial study of constitutional type. *Am J Phys Anthropol.* 1952;10:91–118.
24. Mitani H, Sato K. Comparison of mandibular growth with other variables during puberty. *Angle Orthod.* 1992;62:217–222.
25. Scammon RE. The measurement of the body in childhood. In: Harris J, ed. *The Measurement of Man.* Minneapolis, Minn: University of Minnesota Press; 1930;171–215.