Diagnostic testing of cervical vertebral maturation staging: An independent assessment

Kara M. Morris,[a] Henry W. Fields, Jr,[b] F. Michael Beck,[c] and Do-Gyoon Kim[d]
Columbus, Ohio

Introduction: The reliability of the cervical vertebral maturation (CVM) method has been questioned. The objective of this research was to evaluate the diagnostic reliability of the CVM method to diagnose the mandibular growth spurt using longitudinal records from an alternative database (Iowa Facial Growth Study [IFGS]) using established diagnostic testing methods. **Methods:** Cephalometric films from 43 subjects (males = 20, females = 23) with Class I or Class II skeletal pattern from the IFGS were scanned, digitized, and adjusted for magnification. At least 5 consecutive, annual films were digitized. For each subject, mandibular length (Co-Gn) was measured for each film, and the growth increment between films was calculated. The largest growth increment was the growth spurt. For each subject, the film displaying CVM stage 3 was identified by a blinded examiner viewing the films in random order. Interrater and intrarater repeatability for Co-Gn (intraclass correlation) and CVM staging (weighted kappa) were calculated. Diagnostic tests, including sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were carried out. The present data were compared with data previously derived from samples of the University of Michigan, Oregon, and Burlington Growth studies (UMGS, OGS, and BGS, respectively). A multilevel logistic regression analysis was also run with the mandibular growth peak as the response variable. **Results:** Interrater repeatability for mandibular measurements (intraclass correlation coefficient [ICC] = 0.91) and CVM staging (k = 0.88) were excellent. Intrarater repeatability for mandibular measurements (ICC = 0.98) and CVM staging (kw = 0.55) were excellent to moderate. The UMGS data demonstrated higher sensitivity with comparable specificity. Accuracy was largely similar. Their PPV and NPV had larger ranges. The OGS and BGS data, compared with the IFGS data, showed that our sensitivity and PPVs were higher, that their specificity was higher, and that the NPV and accuracy were very similar. The regression analysis was applied to age groups 10-11 years through 13-14 years. Only chronological age was significant ($P = 0.04$). **Conclusions:** Agreement between CVM stage 3 and the maximum mandibular growth spurt is inconsistent. The diagnostic capability of CVM for the mandibular growth spurt is questionable. (Am J Orthod Dentofacial Orthop 2019;156:626-32)

Orthodontic practitioners often use growth modification strategies for treatment of Class II skeletal malocclusion. Most studies have found that these methods produce modest changes.[1] Some recent data appear to document more robust, long-term improvement in Class II skeletal correction when the treatment coincides with peak mandibular and facial growth.[2] This concept has been supported by systematic reviews and meta analyses,[3,4] and, in these studies, the authors used the cervical vertebral maturation (CVM) method to predict the onset of pubertal growth spurts.

The *repeatability* of the provider to accurately stage CVM has been investigated. Gabriel et al[5] and Nestman et al[6] made the case that insufficient repeatability in stage identification disqualifies CVM from being a clinically useful diagnostic tool. Perinetti et al[7] studied the repeatability of the 6-stage CVM system with unbiased raters, and they noted satisfactory statistical outcomes. Ballrick et al[8] concurred.

Another issue relates to the *validity* of the cervical vertebral staging as it relates to peak mandibular growth. Research has supported the use of CVM in determining

[a]Formerly, Division of Orthodontics, The Ohio State University, Columbus, Ohio; currently, Private practice, Columbus, Ohio.
[b]Division of Orthodontics, The Ohio State University College of Dentistry, Columbus, Ohio.
[c]Division of Biosciences, The Ohio State University College of Dentistry, Columbus, Ohio.
[d]Division of Orthodontics, The Ohio State University, Columbus, Ohio.
All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.
Address correspondence to: Henry W. Fields, Division of Orthodontics, The Ohio State University College of Dentistry, 4088F Postle Hall, 305 W 12th Ave, Columbus, OH; e-mail, fields.31@osu.edu.
Submitted, March 2018; revised and accepted, November 2018.
0889-5406/$36.00

https://doi.org/10.1016/j.ajodo.2018.11.016

skeletal maturation as comparable with hand-wrist film analysis.[9] A systematic review of the predictive ability of the CVM method for determining the growth peak found most research to have methodologic flaws.[10] Assessing CVM stage has been attempted using a variety of means—qualitative, linear, morphometric, and mathematic—leading to great variability in the types of data collected and difficulty when comparing outcomes.

More recently, Gray et al[11] concluded that quantitative assessment of vertebral maturation stage could not determine the peak in mandibular growth in the study of records derived from the Burlington Growth Study. With reasonable repeatability in mandibular length determination and good repeatability regarding CVM stage, there was no significant association between peak mandibular growth and vertebral maturation stage. Mellion et al[12] used subjects from the Bolton-Brush Growth Study and compared multiple maturation indexes to predict the onset and peak of mandibular growth. Among hand-wrist, age, and vertebral stage as predictors, CVM performed worst for predicting peak mandibular growth.

Engel et al[13] found poor results when regression analysis was used to determine the ability of the CVM method to predict the mandibular growth spurt, although their sample was limited to Class II females. The use of this particular demographic by Engel et al was criticized because it was one in which an understandably smaller mandibular growth spurt would be more difficult to detect.

Most orthodontic research suffers from small sample size, and much of the current research regarding the reliability of the CVM method suffers from the same shortcomings.[11,14,15] The annual growth increment for the peak of the mandibular spurt varies from 4.5 to 8 mm, and the measurement of annual mandibular growth during preadolescence is certainly less, approximately 1-3 mm.[16,17] Statistical science theory cautions against the overapplication of studies that interpret observations based solely on *P* values, especially those with small effect sizes.

Perinetti et al[18] took a different approach to investigating the relationship of CVM staging to mandibular growth spurt. They proposed a novel *diagnostic reliability* model hinging on sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV). This approach focused on the CVM method as a diagnostic tool. All tests were performed using data from the same, previously quantified University of Michigan Growth Study (UMGS) subjects used to describe the CVM method. This study demonstrated "satisfactory diagnostic reliability …when the mandibular growth spurt was identified of the basis of the [mandibular growth] increments." Specifically, PPVs of 0.20-1 were noted, with wide confidence intervals.

Later, Perinetti et al[19] used the same approach to further test the diagnostic reliability using an independent sample from the Oregon and Burlington Growth studies (OGS and BGS, respectively). The results of that study suggested that, "none of the CVM stages 2, 3, or 4 alone reached satisfactory diagnostic reliability in the identification of imminent mandibular growth peak."

The objective of this study was to evaluate the ability of CVM stage 3 (Cvs3) to diagnose the mandibular growth spurt using independent longitudinal records obtained from the Iowa Facial Growth Study (IFGS),[20] and Perinetti et al's[18] methodology for diagnostic reliability testing. The IFGS data were used as a similar and largely contemporaneous sample to the UMGS. This should expand our understanding of the diagnostic reliability of the cervical vertebral staging method using another unrelated sample.

MATERIAL AND METHODS

Subjects

Lateral cephalometric films were obtained from the IFGS. In brief, the IFGS contains anthropomorphic and radiographic information collected annually for white subjects from 1940 to 1960 in Iowa City, Iowa. Records were collected beginning in early childhood and extending for some patients into adulthood. The IFGS files are accessible through the AAO Foundation Craniofacial Growth Legacy Collections[20] and have a large total sample (n = 183) with good-quality films and key anatomic structures that are largely unobstructed by cephalostat components.

Inclusion criteria for the sample were subjects with publicly available annual cephalometric films (n = 100). Exclusion criteria included subjects displaying unknown or Class III malocclusions (n = 4), those with <11 months or >13 months between annual films (n = 41), and subjects with orthodontic appliances placed before the growth spurt (n = 12). The final sample size for this study included 43 total subjects (20 males and 23 females). At least 5 annual films were digitized for all subjects, with the longitudinal measurement sequence beginning between 8 and 11 years and extending for some subjects to 16 years. Because these cephalometric films were deidentified and publicly available, the institutional review board determined the study was exempt.

Data collection

For each subject, digital files of scanned films from the IFGS were downloaded as .tiff files using

Firefox FTP software (Mountain View, Calif). Each film was originally scanned at a known resolution of dots per inch, which allowed control of magnification specific to each subject. The films were subsequently loaded into Dolphin software (Chatsworth, Calif). All films were digitized by the same examiner (K.M.).

In keeping with the study by Perinetti et al[18], mandibular growth was assessed as total mandibular length—defined by condylion (Co)-gnathion (Gn). The specific location of the mandibular points is consistent with previous studies.[21] The Co is the most superior-posterior point on the tip of the condyle; the Gn is the midpoint between the anterior and inferior point on the bony chin.

Growth increments between annual age intervals for Co-Gn were calculated from the raw measures. Because exclusion criteria allowed only films between 11 and 13 months, and subsequent subjects demonstrated <1-month variability within the series, no annualization was performed. The peak pubertal growth for each subject was defined by the maximum increase in total mandibular length between consecutive films.

Subsequently, the initial Cvs3 for the series of films for each subject was identified. The examiner (K.M.) was blinded to the sex and age of the subject, and films for each subject were shown in a random order. In keeping with typical clinical practice, the entire film, including soft tissue profile, jaws, and dentition, was displayed. All films were assigned a stage 3 or not. If multiple Cvs3 stages were noted, the earliest stage was used to calculate the diagnostic reliability and all subsequent stage 3 films were noted but not counted as a true positive. When stage 3 preceded the mandibular peak, the conversion was tracked.

Cervical maturation stages were classified, as described below, in keeping with methodology from previous studies.[14,18] The classification relied on the presence or absence of the lower border concavity of CV 2, 3, and 4. This was common among the methods of Gray et al,[11] Mellion et al,[12] Engel et al,[13] and Perinetti al[18,19] and avoided the use of vertebral shape, which has been shown to be a source of additional error.[5,6]

Stage 1 (Cvs1): The inferior borders of the bodies of all (CV 2-4) are flat. The superior borders are tapered from posterior to anterior.

Stage 2 (Cvs2): A concavity develops in the inferior border of the second vertebra. The anterior vertical height of the bodies increases.

Stage 3 (Cvs3): A concavity develops in the inferior border of the third vertebra.

Stage 4 (Cvs4): A concavity develops in the inferior border of the fourth vertebra. Concavities in the lower borders of the fifth and of the sixth vertebrae are beginning to form. The bodies of all cervical vertebrae are rectangular in shape.

Statistical analysis

Intrarater repeatability statistics for Co-Gn length and Cvs3 selection were calculated after the same examiner re-evaluated a randomly selected subset of 20 radiographs 2 weeks later using intraclass correlation and weighted kappa statistics, respectively. Interrater repeatability for Co-Gn length and Cvs3 were assessed by having an additional examiner evaluate the same films (H.W.F.) using the same method and statistics.

Co-Gn was used as the gold standard for mandibular growth spurt, and the diagnostic reliability tests were completed for each age interval. The analysis for each age interval determined the capability of the initial Cvs3 to identify the mandibular growth spurt as measured by Co-Gn. The overall sensitivity, specificity, accuracy, PPV, and NPV were calculated for each age interval by sex (because of known differences in maturation rates) and for the group as a whole.

A multilevel logistic regression analysis, in which the outcome variable was the presence or absence of the pubertal growth spurt (measured as the maximum increment of Co-Gn), and the predictive variables were (1) presence or absence of Cvs3, (2) sex, (3) chronological age, and (4) interactions among these variables, was also completed.

RESULTS

Intrarater diagnostic repeatability for Co-Gn (intraclass correlation coefficient [ICC] = 0.98, 0.93-0.99) and CVM staging (kw = 0.55, 0.26-0.84) was excellent and moderate, respectively. Interrater diagnostic repeatability for Co-Gn (ICC = 0.91, 0.71-0.97) and CVM staging (kw = 0.88, 0.77-1) was generally excellent.[22]

All subjects: IFGS

Diagnostic reliability analysis results for all subjects (n = 43) are given in Table I, which shows results for ages 9–15 years because age groups 8-9 years and 15-16 years had numerous values that could not be calculated because of cell sizes. Sensitivity (0.25-1.0) and specificity outcomes (0.65-1) were highly variable, with the best sensitivity seen at the tails of the age spectrum, where smaller sample sizes are noted. PPV was lowest (≤0.50) when the growth spurt prevalence was the greatest. NPV was higher throughout the data set (0.74-0.95).

Table I. IFGS diagnostic testing analysis values (95% confidence interval; lower-upper)

Diagnostic variables	*8-9*	*9-10*	*10-11*	*11-12*	*12-13*	*13-14*	*14-15*	*15-16*
All subjects (N = 43)								
Sensitivity	0.33 (0.01-1)	1 (0.03-1)	0.25 (0.01-0.81)	0.69 (0.39-0.91)	0.31 (0.09-0.61)	1 (0.48-1)	1 (0.29-1)	*
Specificity	1 (0.29-1)	0.85 (0.65-0.96)	0.77 (0.61-0.89)	0.63 (0.44-0.80)	0.87 (0.69-0.96)	1 (0.90-1)	1 (0.88-1)	*
Accuracy	0.67 (0.22-0.96)	0.81 (0.62-0.94)	0.72 (0.56-0.85)	0.65 (0.49-0.79)	0.70 (0.54-0.83)	0.88 (0.74-0.96)	0.91 (0.75-0.98)	*
PPV	1 (0.03-1)	1 (0.40-1)	1 (0.00-0.45)	0.45 (0.23-0.68)	0.50 (0.16-0.84)	*	*	*
NPV	0.60 (0.15-0.95)	0.96 (0.78-1)	0.91 (0.76-0.98)	0.83 (0.61-0.95)	0.74 (0.57-0.88)	0.88 (0.74-0.96)	0.91 (0.75-0.98)	*
Peaks (total films)	3 (6)	1 (27)	4 (43)	13 (43)	13 (43)	5 (41)	3 (32)	1 (9)
Females (n = 23)								
Sensitivity	0.50 (0.01-0.99)	1 (0.03-1)	0.33 (0.01-0.91)	0.67 (0.30-0.93)	1 (0.40-1)	1 (0.29-1)	*	1 (0.03-1)
Specificity	1 (0.29-1)	0.85 (0.55-0.98)	0.75 (0.51-0.91)	0.50 (0.23-0.77)	0.95 (0.74-1)	1 (0.81-1)	1 (0.79-1)	1 (0.48-1)
Accuracy	0.80 (0.28-0.99)	0.79 (0.49-0.95)	0.70 (0.47-0.87)	0.57 (0.34-0.77)	0.78 (0.56-0.93)	0.86 (0.64-0.97)	1 (0.79-1)	0.83 (0.36-1)
PPV	1 (0.03-1)	1 (0.16-1)	0.17 (0.00-0.64)	0.46 (0.19-0.75)	1 (0.03-1)	*	*	*
NPV	0.75 (0.19-0.99)	0.92 (0.62-1)	0.88 (0.64-0.99)	0.70 (0.35-0.93)	0.82 (0.60-0.95)	0.86 (0.64-0.97)	1 (0.79-1)	0.83 (0.36-1)
Peaks (total films)	2 (5)	1 (14)	3 (23)	9 (23)	4 (23)	3 (21)	0 (16)	1 (6)
Males (n = 20)								
Sensitivity	1 (0.03-1)	*	1 (0.03-1)	0.75 (0.19-0.99)	0.44 (0.14-0.79)	1 (0.16-1)	1 (0.29-1)	*
Specificity	*	0.85 (0.55-0.98)	0.79 (0.54-0.94)	0.75 (0.48-0.93)	0.73 (0.39-0.94)	1 (0.81-1)	1 (0.75-1)	1 (0.29-1)
Accuracy	1 (0.03-1)	0.85 (0.55-0.98)	0.75 (0.51-0.91)	0.75 (0.51-0.91)	0.60 (0.36-0.81)	0.90 (0.68-0.99)	0.81 (0.54-0.96)	1 (0.29-1)
PPV	*	1 (0.16-1)	1 (0.40-1)	0.43 (0.10-0.82)	0.57 (0.18-0.90)	*	*	*
NPV	1 (0.03-1)	1 (0.72-1)	0.94 (0.70-1)	0.92 (0.64-1)	0.62 (0.32-0.86)	0.90 (0.68-0.99)	0.81 (0.54-0.96)	1 (0.29-1)
Peaks (total films)	1 (1)	0 (13)	1 (20)	4 (20)	9 (20)	2 (20)	3 (16)	0 (3)

*Could not calculate.

Females: IFGS

The female (n = 23) data tracked the *All Subject* data in most cells. Many small cell sizes yielded high values and poor PPV, even in the age group with a high number of peaks (11-12 years age group) (Table I).

Males: IFGS

For males (n = 20), the trend was the same, but the greatest number of growth spurts moved to the 12-13 years age group. PPV was near the chance level (Table I).

For each age group, the total number of measured mandibular growth spurts are presented in Table I. Most growth peaks were seen in the 11-12 years age group for females, and the 12-13 years age group for males. For 3 subjects, 1 female and 2 males, Cvs3 began in 1 age stage and extended into the next when the mandibular growth spurt occurred. Seventeen other subjects converted from Cvs3 to Cvs4 or beyond before the mandibular growth spurt (8 females and 9 males).

Logistic regression

A multilevel logistic regression analysis was completed. Our initial analysis included all age groups. However, when we attempted to assess potential interactions, the model would not converge, most probably because of low relative numbers in the age group extremes. Consequently, we restricted our analysis to the

Table II. Results of the multilevel logistic regression analysis

Effect	Num DF	Den DF	Chi-square	Pr > ChiSq
Type III test of fixed effects				
CV	1	36	2.20	0.14
CAGE	1	120	4.25	0.04
CAGE × CV	1	120	3.03	0.08
Sex	1	40	1.01	0.32
CV × sex	1	36	0.02	0.90
CAGE × sex	1	20	0.93	0.34
CAGE × CV × sex	1	120	0.02	0.89

CAGE, chronological age; *CV*, cervical vertebra stage; *DF*, degrees of freedom.

following age groups: 10-11 years, 11-12 years, 12-13 years, and 13-14 years. This analysis included 80% of the available data. In this analysis, only chronological age was significant ($P = 0.04$) (Table II).

DISCUSSION

This study attempted to validate CVM diagnostic reliability by replicating the Perinetti et al[18] UMGS study on a larger, comparable, and independent sample of patients (Table III). This seemed appropriate, given that much treatment data have been generated based on the small UMGS sample used to derive the original relationship between Cvs3 and Co-Gn growth. Studies using other growth data sets, including the Burlington and Bolton-Brush Growth collections, failed to detect clinically useful interactions, but used different methods.[12,13]

When our data for all subjects are compared with the UMGS data,[18] it is clear that their data demonstrate higher sensitivity with comparable specificity. Accuracy was largely similar. Their PPV and NPV had larger ranges, but our values for these statistics were lower in the most populous peak groups (11-12 years and 12-13 years age groups) (Table I).

The present data for IFGS (Table III) were analyzed for all subjects, but then also segmented into female-only and male-only groupings. This was a logical decision based on widely accepted differences in maturation timing.[23] Our sex-segmented data displayed expected differences in timing for the onset of the growth peak and small samples (Table I).

Perinetti et al[19] found unsatisfactory diagnostic reliability results using the CVM method to detect the mandibular growth peak using data from the OGS and BGS. The overall data set for Cvs3 shows that our sensitivity (0.25-1.0) and PPV (0.45-1) were higher; their specificity was higher (0.74-0.96); and the NPV and accuracy were very similar. The usual dynamics of these types of results—high sensitivity and NPV along with high specificity and high PPV—was absent in both studies. This is most likely a reflection of the prevalence of the growth spurts in the age groups.

In general, the PPV for Cvs3 predicting the mandibular growth spurt measured by Co-Gn was highly variable. The OGS and BGS data and the IFGS data showed the PPV to be lower than 0.5 for the age groups with the most growth spurts. This suggests that the chance a positive Cvs3 resulted in a positive growth spurt

Table III. IFGS subject data set

Subject ID	Sex*	Co-Gn peak growth age group	Cvs3 age group
6 M	1	14-15	12-13
13 M	1	10-11	12-13
15 F	0	11-12	11-12
22 F	0	11-12	12-13
33 F	0	15-16	11-12
33 M	1	11-12	11-12
34 F	0	12-13	11-12
39 F	0	12-13	11-12
32 F	0	10-11	11-12
40 F	0	11-12	11-12
40 M	1	13-14	11-12
43 F	0	13-14	10-11
44 F	0	8-9	9-10
45 F	0	8-9	8-9
46 F	0	11-12	10-11
50 F	0	11-12	11-12
52 F	0	11-12	11-12
50 M	1	13-14	10-11
57 M	1	13-14	10-11
51 M	1	14-15	12-13
53 M	1	12-13	12-13
56 M	1	12-13	11-12
59 F	0	11-12	10-11
60 F	0	12-13	11-12
61 F	0	10-11	10-11
62 F	0	9-10	10-11
63 M	1	11-12	10-11
64 M	1	12-13	9-10
65 M	1	14-15	9-10
66 F	0	13-14	11-12
66 M	1	12-13	12-13
67 M	1	12-13	10-11
69 M	1	12-13	12-13
70 F	0	11-12	11-12
72 F	0	11-12	11-12
73 M	1	12-13	12-13
74 M	1	12-13	11-12
76 F	0	10-11	11-12
75 M	1	11-12	11-12
79 M	1	11-12	11-12
78 F	0	12-13	9-10
82 M	1	12-13	11-12
83 M	1	8-9	10-11

*0 = female; 1 = male.

was low. This type of result undermines a clinician's confidence in the diagnostic test and would lead to early and overtreatment of patients—most probably beginning treatment before it was indicated. For instance, the growth spurt would not be missed, but potential treatment would be longer than necessary if begun at Cvs3 (3 subjects), or in another scenario, treatment could start and end before the spurt (11 subjects, based on a treatment period of less than 24 months).

It must be noted that the PPV is a combination of prevalence, sensitivity, and specificity, and it can be argued that the low prevalence of Cvs3 throughout these studies influenced the statistical outcomes.[24] The prevalence of a growth spurt in an age group was 0.08-0.29 in the UMGS, 0.04-0.3 in the OGS, and 0.02-0.3 in the IFGS data. These prevalence levels can make even strong sensitivity and specificity values of greater than 0.8 (higher than these studies) decline to below the chance level.[24]

Most positives are false positives for tests of Cvs3 with low prevalence. Small cell sizes provide misleading results, often a value of 1 or return no result at all, depending on the calculation. Given the small cell sizes when the sexes were separated, it is tempting to combine them. Even with the OGS and BGS and IFGS presenting twice the sample size as the UMGS, the problem was still present for several cells when all subjects were combined (Table I).

Overall, for the IFGS data, accuracy ranged from 0.65 to 0.91; for females 0.57-1.0 and for males 0.6-1. The overall UMGS data showed 0.67-1. For the OGS and BGS data, it was 0.66-0.94. One might argue that the accuracy component of diagnostic reliability testing is a better gauge of the real diagnostic reliability when compared with PPV or NPV. Fletcher et al[24] noted accuracy is a summary term that is not appropriate to be useful in most clinical situations because the component information is lost.

These results, and those of the OGS and BGS, fail to support a strong diagnostic reliability for the CVM method uniformly for males, females, or total samples in these age groups, when PPV and NPV should provide the necessary guidance for a diagnostic method. Perinetti et al[3,4] suggest that the large variability, and that the peak mandibular growth spurt may occur well after the Cvs3 stage (40% in the current study), may make this diagnostic tool inadequate to be used alone for growth assessment.

Logistic regression, which has been completed by others, did not find a significant relationship or predictive value for the cervical vertebral status with mandibular peak growth. Chronological age was a better predictor. This is consistent with several other findings.[11-13] Although there seems to be a positive relationship between mandibular peak growth and age, if one takes this literally and does not use other available clinical growth information the growth peak may be missed (our data show a steep decline for the 13-14 years age group used in our regression). Noninvasive, conventional approaches to supplementary clinical information to detect growth status would include documented height changes, weight changes, and secondary sex characteristics. One could add a hand-wrist film.

The overall usefulness of the CVM method as shown by the regression may be related to the ability to define and identify the stages or that the CVM method may be ineffective, as suggested by the diagnostic reliability results reported here and elsewhere.[19,25] The problems have been summarized as follows: identification of the beginning of the growth spurt is difficult, and some cervical vertebral stages last longer than the growth spurt.[25]

Our repeatability was moderate for the cervical vertebra staging and is a limitation of the study. This is lower than some reported repeatability.[8,18] This may be a reflection of the exceptional cases (reported to be approximately 13%)[26] that make classification and diagnostic reliability lower. Our method also reflected non–computer-assisted determination of the cervical stage similar to that used in clinical practice. Because these data were generated from historic samples and there are most likely differences because of secular trends compared with data from today's youth, the influence on variability is uncertain. The generalizability to today's youth should be viewed with caution for all these samples.[27]

Overall, our findings from the diagnostic reliability testing, regardless of grouping, produced less robust support for the use of Cvs3 as a predictor for the mandibular growth spurt than that of Perinetti et al[18] UMGS data, especially in age groups where most growth peaks were located. Our data appear to support the general findings of the Perinetti et al[19] OGS and BGS data. Both of these studies used independent data sets to test a method generated on a third data set (UMGS). Given small differences in the methods, the results are markedly similar.

In view of these results, practitioners are advised to use multiple methods, with their known shortcomings to help inform growth modification timing.

CONCLUSIONS

Based on diagnostic reliability testing, the diagnostic capability of CVM for diagnosing mandibular growth spurt is questionable.

ACKNOWLEDGMENTS

The authors wish to acknowledge Sean Curry, AAO Foundation Legacy Collection Web site and Database Developer, and Nathan Holton, PhD, Assistant Professor, Department of Orthodontics, College of Dentistry, University of Iowa.

REFERENCES

1. Baccetti T, Franchi L, Toth LR, McNamara JA Jr. Treatment timing for Twin-block therapy. Am J Orthod Dentofacial Orthop 2000; 118:159-70.
2. Franchi L, Pavoni C, Faltin K Jr, McNamara JA Jr, Cozza P. Long-term skeletal and dental effects and treatment timing for functional appliances in Class II malocclusion. Angle Orthod 2013; 83:334-40.
3. Perinetti G, Primožič J, Furlani G, Franchi L, Contardo L. Treatment effects of fixed functional appliances alone or in combination with multibracket appliances: a systematic review and meta-analysis. Angle Orthod 2015;85:480-92.
4. Perinetti G, Primožič J, Franchi L, Contardo L. Treatment effects of removable functional appliances in pre-pubertal and pubertal Class II patients: a systematic review and meta-analysis of controlled studies. PLoS One 2015;10:e0141198.
5. 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.e1-7: discussion 478-480.
6. Nestman TS, Marshall SD, Qian F, Holton N, Franciscus RG, Southard TE. Cervical vertebrae maturation method morphology criteria: poor reproducibility. Am J Orthod Dentofacial Orthop 2011;140:182-8.
7. Perinetti G, Caprioglio A, Contardo L. Visual assessment of the cervical vertebral maturation stages: a study of diagnostic accuracy and repeatability. Angle Orthod 2014;84:951-6.
8. Ballrick JW, Fields HW, Beck FM, Sun Z, Germak J. The cervical vertebrae staging method's reliability in detecting pre and post mandibular growth. Orthod Waves 2013;72:105-11.
9. Stiehl J, Müller B, Dibbets J. The development of the cervical vertebrae as an indicator of skeletal maturity: comparison with the classic method of hand-wrist radiograph. J Orofac Orthop 2009; 70:327-35.
10. Santiago RC, de Miranda Costa LF, Vitral RW, Fraga MR, Bolognese AM, Maia LC. Cervical vertebral maturation as a biologic indicator of skeletal maturity. Angle Orthod 2012;82:1123-31.
11. Gray S, Bennani H, Kieser JA, Farella M. Morphometric analysis of cervical vertebrae in relation to mandibular growth. Am J Orthod Dentofacial Orthop 2016;149:92-8.
12. Mellion ZJ, Behrents RG, Johnston LE Jr. The pattern of facial skeletal growth and its relationship to various common indexes of maturation. Am J Orthod Dentofacial Orthop 2013;143: 845-54.
13. Engel TP, Renkema AM, Katsaros C, Pazera P, Pandis N, Fudalej PS. The cervical vertebrae maturation (CVM) method cannot predict craniofacial growth in girls with Class II malocclusion. Eur J Orthod 2016;38:1-7.
14. Franchi L, Baccetti T, McNamara JA Jr. Mandibular growth as related to cervical vertebral maturation and body height. Am J Orthod Dentofacial Orthop 2000;118:335-40.
15. Baccetti T, Franchi L, Mc Namara JA Jr. An improved version of the cervical vertebral maturation (CVM) method for the assessment of mandibular growth. Angle Orthod 2002;72:316-23.
16. Björk A. Variations in the growth pattern of the human mandible: longitudinal radiographic study by the implant method. J Dent Res 1963;42:400-11.
17. Stahl F, Baccetti T, Franchi L, Mcnamara JA Jr. Longitudinal growth changes in untreated subjects with Class II Division 1 malocclusion. Am J Orthod Dentofacial Orthop 2008;134:125-37.
18. Perinetti G, Contardo L, Castaldo A, McNamara JA Jr, Franchi L. Diagnostic reliability of the cervical vertebral maturation method and standing height in the identification of the mandibular growth spurt. Angle Orthod 2016;86:599-609.
19. Perinetti G, Primozic J, Sharma B, Cioffi I, Contardo L. Cervical vertebral maturation method and mandibular growth peak: a longitudinal study of diagnostic reliability. Eur J Orthod 2018;40: 666-72.
20. AAO Foundation Craniofacial Growth Legacy Collections: Available at: http://www.aaoflegacycollection.org/aaof_home.html. Accessed July 31, 2019.
21. 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;11:119-29.
22. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284-90.
23. Marshall WA, Tanner JM. Puberty. In: Falkner F, Tanner JM, editors. Human Growth, vol. 2. New York: Plenum Publishing; 1986. p. 171-210.
24. Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: the essentials. 5th ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2012.
25. Perinetti G, Contardo L, Primozic J. Diagnostic accuracy of the cervical vertebral maturation method. Eur J Orthod 2018;40:453-4.
26. Perinetti G, Bianchet A, Franchi L, Contardo L. Cervical vertebral maturation: an objective and transparent code staging system applied to a 6-year longitudinal investigation. Am J Orthod Dentofacial Orthop 2017;151:898-906.
27. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocr Rev 2003;24:668-93.