Original Article

Reliability of lateral cephalometric radiographs in the assessment of the upper airway in children: *A retrospective study*

Fabio Savoldi[a]; Gou Xinyue[b]; Colman P. McGrath[c]; Yanqi Yang[d]; Shiu Cheuk Chow[e]; James K. H. Tsoi[f]; Min Gu[g]

ABSTRACT

Objectives: To test the reliability of Lateral cephalometric radiographs (LCRs) for use in the assessment of the upper airway, hyoid bone, soft palate, and tongue.

Materials and Methods: The records of 57 healthy Chinese children from a nonhospital population (mean age = 12.6 years, SD = 0.5, 28 males and 29 females) who received two consecutive LCRs in the natural head posture were retrospectively analyzed. Fifteen linear, angular, and area measurements were used to describe the airway, hyoid bone, soft palate, and tongue. The reliability between the two LCRs was assessed with the intraclass correlation coefficient (ICC) and *F*-test. Errors were estimated with the Dahlberg and Bland-Altman method, and intra- and inter-assessor agreements were determined.

Results: Measurements of upper airway and hyoid bone had excellent method reliability, intra-assessor reliability, and inter-assessor reliability (ICC $>$ 0.8). However, the method reliability and the inter-assessor reliability for soft palate and tongue was less favorable (ICC from 0.60 to 0.96). Soft palate area and thickness were the most critical parameters. Intra-assessor reliability was greater than both method reliability and inter-assessor reliability (which were similar).

Conclusions: The measurement of upper airway morphology, defined as the intramural space, and of the hyoid bone position were highly reliable on LCRs of children. However, the limited reliability in the assessment of tongue and soft palate area may compromise the diagnostic application of LCRs to these structures. (*Angle Orthod.* 2020;90:47–55.)

KEY WORDS: Cephalogram; Obstructive sleep apnea; Sleep-disordered breathing; Tongue; Hyoid bone

[a] Postgraduate student, Department of Orthodontics, Dental School, University of Brescia, Brescia, Italy.

[b] Research Assistant, Orthodontics, Faculty of Dentistry, The University of Hong Kong, Hong Kong SAR, and Attending Physician, Department of Orthodontics, Shenyang Stomatology Hospital, Shenyang, People's Republic of China.

[c] Clinical Professor, Dental Public Health, Faculty of Dentistry, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong SAR.

[d] Associate Professor, Orthodontics, Faculty of Dentistry, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong SAR.

[e] Honorary Research Associate, Orthodontics, Faculty of Dentistry, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong SAR.

[f] Assistant Professor, Dental Materials Science, Discipline of Applied Oral Sciences, Faculty of Dentistry, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong SAR.

[g] Clinical Assistant Professor, Orthodontics, Faculty of Dentistry, The University of Hong Kong, Prince Philip Dental Hospital, Hong Kong SAR.

Corresponding author: Dr Min Gu, Orthodontics, 2/F, Prince Philip Dental Hospital, 34 Hospital Road, Sai Ying Pun, Hong Kong SAR (e-mail: drgumin@hku.hk)

Accepted: June 2019. Submitted: February 2019.
Published Online: August 12, 2019

INTRODUCTION

Lateral cephalometric radiographs (LCRs) have been widely used as a screening tool for children with suspected sleep-disordered breathing,[1] which may be related to metabolic, cardiovascular, and neurocognitive morbidity in young people.[2] The diagnostic application of LCRs has been recognized by a meta-analysis concluding that there was reduced sagittal width of the upper airway in children with obstructive sleep apnea.[3]

LCRs have been used to investigate the intramural airway spaces,[1,4] tongue,[5] soft palate,[1,5] and supporting structures, such as the hyoid bone,[1,4] mandible,[1,4] and cervical vertebrae.[1] Some of these structures may be difficult to identify, and, for example, the use of a radiopaque paste has been advocated to highlight the tongue contour.[6] In addition, there are specific criticisms of the use of LCRs in upper airway assessment because images are obtained with subjects in an upright position, which obviously differs from the sleeping position. Nevertheless, it is acknowledged that LCRs can discriminate between obstructive sleep apnea and snoring independently from the position of the subject,[7] confirming its potential relevance as a screening method. In fact, it is recommended that mouth-breathing children be sent for a sleep assessment if their superior pharyngeal airway space appears small on an LCR.[4] Yet, although the intrinsic static nature of LCRs raises concerns about their reliability, there is limited exploration of this—a few reports among adults[8–10] and no previous study performed in children.

The present study aimed to determine the reliability of LCRs in the assessment of the upper airway in children in order to identify which, if any, variables were reliable for potential use in the clinical diagnosis and assessment of treatment outcomes in the management of sleep-disordered breathing.

MATERIALS AND METHODS

Subjects

In the study of Cooke,[11] published in 1986, stratified sampling was performed on 11 randomly selected schools in Hong Kong; 618 children were recruited to receive an LCR. A subgroup of 57 children (28 male, 29 female; mean age = 12.6 years; SD = 0.5), without previous orthodontic treatment and receiving LCRs with the same protocol, was included in the present study. LCRs showing mouth opening, swallowing action, tongue not in rest position, evident changes in head posture, or tissue falling outside the frame were excluded (Figure 1).

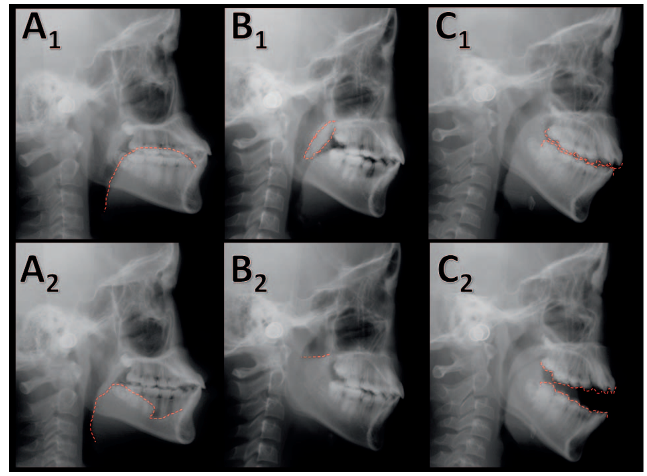

Figure 1. Examples of cases (A, B, C) excluded because the tongue was not in the rest position (A_1 and A_2), swallowing action was present (B_1 and B_2), or the mouth was open (C_1 and C_2). The images on top show the situation with larger upper airway (1), and the images at bottom show the same patient with the upper airway narrowed (2).

The present study was approved by the Institutional Review Board of the University of Hong Kong / Hospital Authority Hong Kong West Cluster (UW12-405).

Acquisition of LCRs

LCRs were taken with an analogic X-ray unit (GE1000, General Electric, Boston, Massachusetts, USA) with cephalometer (CI-2, Wehmer, Lombard, Illinois, USA). Subjects were in natural head posture (NHP),[12] and ear posts were used to stabilize the position. One LCR was taken at baseline (T_1), and a second LCR (T_2) was taken either after 5–10 minutes (subgroup A) or 60–100 minutes (subgroup B). Only LCRs taken with the described protocol were selected from the original sample collected by Cooke.[11]

Variables and Tracing

Cephalometric analysis was carried out with a software program (CASSOS, Soft Enable Technology Limited, Hong Kong SAR), and linear measurements were corrected according to a magnification of 8.75% for midsagittal structures. Area measurements were made with graphical software (ImageJ[13]) and corrected according to a magnification of 18.27%. Figure 2 and Table 1 illustrate the points and lines used to identify the variables, and Figure 2 and Table 2 illustrate the variables.[5,14,15]

Measurements were obtained by a primary assessor (G.X., an orthodontist) and a secondary assessor (G.M., an orthodontist) after an initial calibration on 10 LCRs. The primary assessor conducted assessments at T_1 and T_2. After a washout period of about 1

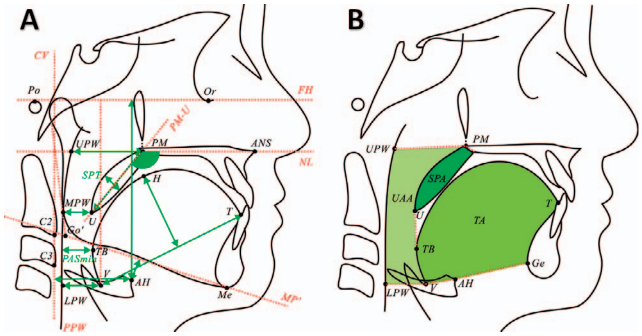

Figure 2. Points (black dots), construction lines (red, dashed), and variables (green arrows and areas) used for the analysis of the upper airway, hyoid bone, tongue, and soft palate. Linear and angular measurements are on the left (A) and area measurements on the right (B).

month, both the primary assessor (T_1') and the secondary assessor (T_1'') repeated 25 of the T_1 measurements.

Four data sets were analyzed: one including the first LCR measured by the primary assessor (T_1), one with the first LCR remeasured by the primary assessor (T_1'), one with the first LCR measured by the secondary assessor (T_1''), and one with the second LCR measured by the primary assessor (T_2). In the assessments, T_1-T_2 represents the airway change, T_1-T_1' the intra-assessor difference, and T_1-T_1'' the inter-assessor difference.

Sample-Size Calculation

The sample size was calculated allowing the intraclass correlation coefficient (ICC) to identify a significant agreement > 0.8, with a power of 80% and a significance level of 5% (two sided).[16] The required sample was $n = 49$ and, given the retrospective nature of the study, all 57 children were included.

Data Analysis

The normality of the data distribution was verified with the Shapiro-Wilk test. Differences in the T_1-T_2 changes between subgroup A and subgroup B were compared with the Student's *t*-test for independent samples.

First, for comparison analysis, the mean directional difference (DD)[17] was calculated ($DD = X_{T1} - X_{T2}$). For

Table 1. Cephalometric Landmarks and Lines

Abbreviation	Name	Description	From
Cephalometric landmarks			
AH	Anterior hyoid	Most anterior and superior point of the body of the hyoid bone	Shen et al. 1994
ANS	Anterior nasal spine	Median and most anterior point of the basal bone of the maxilla	Samman et al. 2003[c]
C2	Second cervical vertebra	Most anterior and inferior point of the second cervical vertebra	Gu et al. 2014[c]
C3	Third cervical vertebra	Most anterior and inferior point of the third cervical vertebra	Gu et al. 2014[c]
Go′	Gonion (geometric)	Intersection of the tangents of inferior and posterior borders of the mandible[a]	Gu et al. 2014
Ge	Genial tubercle	Most posterior point on the mandibular symphysis	Shen et al. 1994
H	Superior part of the tongue	Most distant point of the tongue respect to VT	Shen et al. 1994[c]
LPW	Lower pharyngeal wall	Intersection of the line passing through V and perpendicular to PPW[b]	Shen et al. 1994[c]
Me	Menton	Most inferior point on the body of the chin	Gu et al. 2014[c]
MPW	Middle pharyngeal wall	Intersection of the line passing through U and perpendicular to PPW[b]	Shen et al. 1994[c]
Or	Orbitale	Most inferior point of the lower border of the orbit[a]	Samman et al. 2003[c]
PM	Pterygo-maxillare	Projection of the tip of the pterygomaxillary fissure on the hard palate	Samman et al. 2003[d]
Po	Porion	Most superior point of the external auditory meatus[a]	Samman et al. 2003[c]
T	Tongue tip	Most anterior point of the tip of the tongue	Shen et al. 1994[c]
TB	Tongue base	Most posterior point of the base of the tongue	ex novo
U	Uvula	Tip of the uvula	Shen et al. 1994
UPW	Upper pharyngeal wall	Intersection of NL and PPW	Shen et al. 1994[c]
V	Vallecula	Intersection of the epiglottis and the base of the tongue	Shen et al. 1994
Cephalometric lines			
CV	Cervical plane	Line joining C2 to C3	Samman et al. 2003[c]
FH	Frankfort horizontal plane	Line joining Or to Po	Samman et al. 2003
MP′	Mandibular plane (geometric)	Line joining Me to Go′	Gu et al. 2014[c]
NL	Nasal line	Line joining ANS to PM	Shen et al. 1994
PPW	Posterior pharyngeal wall	Line contouring the posterior wall of the pharynx	ex novo
PM-U	Long axis of the soft palate	Line joining PM with U	Shen et al. 1994[c]
VT	Long axis of the tongue	Line joining V to T	Shen et al. 1994[c]

[a] Averaged between left and right.
[b] Estimate perpendicular.
[c] Modified wording compared with the original definition.
[d] Modified meaning compared with the original definition.

Table 2. Cephalometric Variables

Abbreviation	Unit	Name	Description	From
Upper airway				
PASmin	mm	Retroglossal oropharyngeal airway space	Distance from TB to PPW[a]	Shen et al. 1994[b]
PM-UPW	mm	Nasopharyngeal airway space	Distance from PM to UPW	Shen et al. 1994[b]
UAA	mm^2	Upper airway area	Area within PPW (posteriorly); PM-UPW (superiorly); posterior contour of the soft palate, U to TB, and the residual part of the tongue base from TB to V (anteriorly); and LPW-V (inferiorly)	Samman et al. 2003[c]
U-MPW	mm	Retropalatal oropharyngeal airway space	Distance from U to MPW	Shen et al. 1994[b]
V-LPW	mm	Hypopharyngeal airway space	Distance from V to LPW	Shen et al. 1994[b]
Soft palate				
NL/PM-U	°	Soft palate inclination	Angle between PM-U and NL	Shen et al. 1994[b]
PM-U(l)	mm	Soft palate length	Distance from PM to U	Shen et al. 1994[b]
SPA	mm^2	Soft palate area	Area within the contour of the soft palate (posteriorly); and the line passing through PM and perpendicular to PM-U (anteriorly)	Shen et al. 1994[b]
SPT	mm	Soft palate thickness	Maximum thickness of the soft palate perpendicular to PM-U[a]	Shen et al. 1994[b]
Hyoid bone				
AH-CV	mm	Horizontal position of the hyoid bone with respect to the vertebrae	Length from AH to CV (parallel to FH)	Shen et al. 1994[b]
AH-FH	mm	Vertical position of the hyoid bone with respect to the Frankfort plane	Distance from AH to FH[a]	Shen et al. 1994[b]
AH-MP	mm	Vertical position of the hyoid with respect to the mandible	Distance from AH to MP[a]	
Tongue				
H-VT	mm	Height of the tongue	Distance from H to VT	Shen et al. 1994[b]
TA	mm^2	Tongue area	Area within the contour of the tongue (superiorly and posteriorly); the line passing through V, AH, GE (inferiorly); the lingual contour of the chin and of the lower incisors (anteriorly)	Shen et al. 1994[c]
VT(l)	mm	Length of the tongue	Distance from V to T	Shen et al. 1994[b]

[a] Estimated perpendicular.
[b] Modified wording compared with the original definition.
[c] Modified meaning compared with the original definition.

normally distributed data, a one sample Student's *t*-test was used to compare mean DD to zero. For not normally distributed data, a one sample Wilcoxon signed-rank test was used to compare median DD to zero. Significant differences in DD indicate systematic bias between groups, and effect size was estimated through the standardized directional difference (SDD),[17] calculated by dividing DD by the standard deviation of the T_1 measurements ($SDD = DD / SD_{T1}$). SDD was considered small if close to ± 0.2, medium if close to ± 0.5, and large if close to ± 0.8 or above.[18]

For assessing differences not accounting for positive and negative signs, the absolute difference (AD) was calculated. Dahlberg's error was calculated,[19] and the Bland-Altman method[20] was used for graphical illustration of the agreements between T_1 and T_2 measurements.

Finally, for agreement analysis, the single measure ICC for absolute agreement was employed.[21] ICC was considered poor if < 0.5, fair from 0.5 to 0.7, good from 0.7 to 0.8, excellent if > 0.8, and perfect if $= 1.0$.[21] The *F*-test was used to assess if ICC was > 0.8 (T_1-T_2).[22]

Statistical analysis was performed with statistical software (SPSS Statistics 20, IBM, Armonk, New York, USA) at the significance level $\alpha = 0.05$.

RESULTS

From the initial pool of 618 patients, 550 were excluded because LCRs were taken with different protocols, resulting in 68 patients. Of these, five were excluded because the tongue was not in the rest position, four patients were not occluding, one was swallowing, and one had the hyoid bone out of the X-ray film, resulting in 57 patients included in the

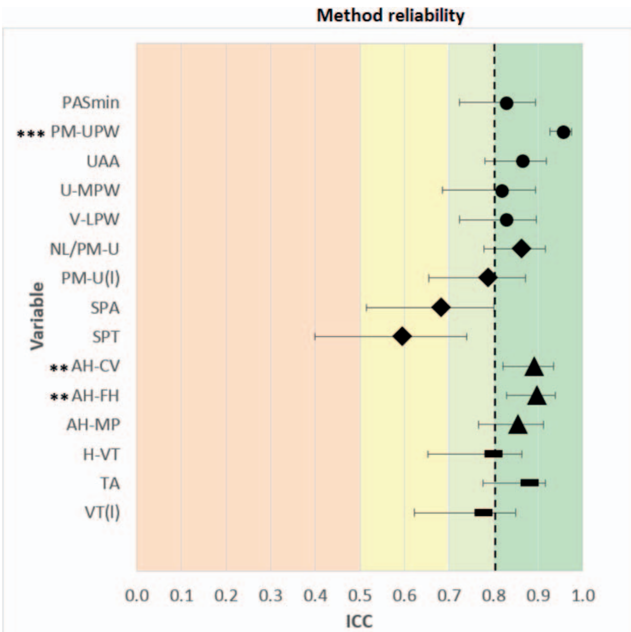

Figure 3. Confidence intervals (CIs) of the intraclass correlation coefficients (ICC) between the first (T_1) and second (T_2) lateral cephalometric radiograph. The threshold value of 0.8 is dashed, and asterisks indicate that the F-test reported the CI to be significantly higher than 0.8 (**P* < .05; ***P* < .01; ****P* < .001). Circles stand for the airway, diamonds for the soft palate, triangles for the hyoid bone, and dashes for the tongue.

analysis. No patient showed evident changes in the head posture.

No difference was present between first and second LCRs (T_1-T_2) between the subgroup A and B, meaning that taking LCRs at intervals of 5–10 minutes or 60–100 minutes was not relevant, and data sets were merged.

Method Reliability

The comparison of the upper airway between the two consecutive LCRs, and representing the method reliability (T_1 - T_2), is shown in Table 3. Only two measurements showed a DD statistically different from zero, but they were < 1.0 mm. The ICC ranged between 0.60 (fair) and 0.96 (excellent). Three linear parameters (AH-CV, AH-FH, and PM-UPW) showed ICC significantly > 0.8, and the upper airway area (UAA) was also worth mentioning ($P = 0.059$) (Figure 3) which showed few outsiders in the Bland-Altman plots.

Intra-assessor and Inter-assessor Reliability

The comparison of the two sets of measurements made by the same assessor on the same LCR, and representing the intra-assessor reliability (T_1 - T_1'), is shown in Table 4. The ICC ranged between 0.86 and 1.00 (excellent).

The comparison of the two sets of measurements made by the two assessors on the same LCR, and representing the inter-assessor reliability (T_1 - T_1''), is shown in Table 4. The ICC ranged between 0.66 (fair) and 0.99 (excellent).

Overall Reliability

Only the AD and the Dahlberg error of the method reliability and the inter-assessor reliability showed values which may have clinical relevance (Tables 3 and 4). Measurements of the upper airway and the hyoid bone had excellent method reliability (ICC from 0.82 to 0.96), intra-assessor reliability (ICC from 0.99 to 1.0), and inter-assessor reliability (ICC from 0.94 to 0.99). However, the assessment of the soft palate and the tongue area had somewhat lower method reliability (ICC from 0.60 to 0.86) and lower inter-assessor reliability (from 0.66 to 0.96). The soft palate area and its thickness was of particular concern (Figure 3).

DISCUSSION

Variations Due to the Assessors

The intra-assessor reliability was excellent, showing minimal differences and errors in all measurements. In fact, all measurements but one had a reliability close to perfect (Table 4). The only measurement with lower but still acceptable intra-assessor reliability was the soft palate area, whose anterior border is located where the soft palate contacts the tongue. As the two muscles have similar radiolucency and are usually in tight contact for sealing the oral cavity during nasal breathing, distinguishing their borders may be challenging. Accordingly, a small but increased systematic bias (SDD) was associated with measurement of the soft palate area and the tongue area, which were the two more critical (Table 4). These results were in agreement with Malkoc et al.,[10] who reported a high intra-assessor reliability in the linear measurements of the upper airway, hyoid bone, soft palate, and tongue. Similarly, Juliano et al.[4] reported a perfect or substantial intra-assessor reliability in the linear measurements of the upper airway and hyoid bone. Additionally, Pirila-Parkkinen et al.[1] showed excellent results for linear and angular measurements of the upper airway, hyoid bone, and soft palate. Thus, very little of the overall LCR reliability might be affected by intra-assessor variations.

Although the inter-assessor reliability of upper airway and hyoid bone were excellent, it decreased to good for the tongue area, and fair for the soft palate

Table 3. Method Reliability Assessed by Comparison Between First and Second LCR[a]

		Data Sets					
		T_1 (n = 57)			T_2 (n = 57)		
Variable	Unit	Mean	SD	CI	Mean	SD	CI
Upper airway							
PASmin	mm	8.2	2.7	7.5 to 9.0	8.3	2.8	7.6 to 9.0
PM-UPW	mm	20.8	3.6	19.9 to 21.8	20.9	3.4	20.0 to 21.8
UAA	mm²	486	129	453 to 520	500	127	467 to 533
U-MPW	mm	8.2	2.2	7.6 to 8.7	8.6	2.1	8.1 to 9.2
V-LPW	mm	10.9	3.6	10.0 to 11.8	10.8	3.8	9.8 to 11.8
Soft palate							
NL/PM-U	°	126.7	6.1	125.1 to 128.3	126.5	6.6	124.8 to 128.2
PM-U(l)	mm	29.4	2.7	28.7 to 30.1	28.9	2.8	28.1 to 29.6
SPA	mm²	179	33	170 to 187	177	31	169 to 185
SPT	mm	7.8	1.3	7.5 to 8.1	7.9	1.4	7.6 to 8.3
Hyoid bone							
AH-CV	mm	30.3	3.3	29.4 to 31.2	30.3	3.4	29.4 to 31.2
AH-FH	mm	76.2	5.7	74.7 to 77.7	76.9	6.2	75.3 to 78.5
AH-MP	mm	10.3	4.9	9.1 to 11.6	10.9	5.3	9.5 to 12.2
Tongue							
H-VT	mm	27.8	3.2	26.9 to 28.6	27.9	2.9	27.1 to 28.6
TA	mm²	1817	232	1757 to 1878	1841	216	1785 to 1897
VT(l)	mm	60.2	6.0	58.7 to 61.8	61.1	6.2	59.5 to 62.7

[a] AD indicates absolute difference; CI, 95% confidence interval; DD, directional difference; DE, Dahlberg's error; ICC, intraclass correlation coefficient; LCR, lateral cephalometric radiograph; SD, standard deviation; SDD, standardized directional difference.
[b] SDD was considered small if close to ±0.2, medium if close to ±0.5, and large if close to ±0.8 or above.
[c] ICC was considered poor if < 0.5, fair from 0.5 to 0.7, good from 0.7 to 0.8, and excellent if > 0.8.
[d] One-sample Student's *t*-test comparing mean DD to 0.
[e] One-sample Wilcoxon signed-rank test comparing median DD to 0.
[f] *F*-test verifying if ICC is > 0.8.
[g] Statistically significant *P*-values ($P < .05$) are boldfaced.

Table 4. Intra-assessor and Inter-assessor Reliability by Comparison of Repeated Measurements on the Same LCR[a]

		Data Sets								
		T_1 (n = 25)			T_1' (n = 25)			T_1'' (n = 25)		
Variable	Unit	Mean	SD	CI	Mean	SD	CI	Mean	SD	CI
Upper airway										
PASmin	mm	8.8	2.7	7.7 to 9.8	8.8	2.7	7.7 to 9.8	9.0	2.5	8.0 to 10.0
PM-UPW	mm	21.5	2.9	20.4 to 22.7	21.5	2.9	20.4 to 22.7	21.9	3.2	20.7 to 23.2
UAA	mm²	521	141	466 to 576	519	141	464 to 575	527	144	471 to 584
U-MPW	mm	8.4	2.2	7.5 to 9.3	8.4	2.3	7.5 to 9.3	8.8	2.5	7.8 to 9.8
V-LPW	mm	11.3	3.2	10.1 to 12.6	11.2	3.2	10.0 to 12.5	12.0	3.1	10.7 to 13.2
Soft palate										
NL/PM-U	°	127.5	5.6	125.3 to 129.7	127.5	5.4	125.4 to 129.6	127.2	6.3	124.7 to 129.6
PM-U(l)	mm	29.0	3.1	27.8 to 30.2	29.1	3.1	27.9 to 30.3	29.4	3.1	28.2 to 30.7
SPA	mm²	177	31	165 to 190	173	28	162 to 184	169	26	159 to 179
SPT	mm	7.2	1.0	6.8 to 7.7	7.4	1.1	6.9 to 7.8	8.0	1.2	7.5 to 8.4
Hyoid bone										
AH-CV	mm	30.9	3.5	29.6 to 32.3	31.0	3.4	29.6 to 32.3	30.4	3.4	29.1 to 31.8
AH-FH	mm	76.6	7.5	73.6 to 79.5	76.5	7.4	73.6 to 79.5	75.8	7.4	72.9 to 78.8
AH-MP	mm	11.5	5.2	9.4 to 13.5	11.5	5.2	9.4 to 13.5	11.2	5.2	9.1 to 13.2
Tongue										
H-VT	mm	27.3	3.3	26.0 to 28.6	27.3	3.3	26.0 to 28.6	27.7	3.3	26.4 to 28.9
TA	mm²	1835	239	1741 to 1929	1838	241	1743 to 1932	1913	243	1818 to 2008
VT(l)	mm	59.9	7.2	57.1 to 62.8	60.0	7.3	57.1 to 62.8	59.9	7.3	57.0 to 62.8

[a] AD indicates absolute difference; CI, 95% confidence interval; DD, differential difference; DE, Dahlberg's error; ICC, intraclass correlation coefficient; LCR, lateral cephalometric radiograph; SD, standard deviation; SDD, standardised differential difference.
[b] SDD was considered small if close to ±0.2, medium if close to ±0.5, and large if close to ±0.8 or above.
[c] ICC was considered poor if < 0.5, fair from 0.5 to 0.7, good from 0.7 to 0.8, and excellent if > 0.8.

Table 3. Extended

First LCR vs Second LCR (T_1-T_2)											
Group Comparison									Group Agreement		
DD				SDD[b]	AD			DE	ICC[c]		
Mean	SD	CI	*P*-value		Mean	SD	CI			CI	*P*-value[d]
–0.1	1.6	–0.5 to 0.4	.772[d]	–0.02	1.2	1.1	0.9 to 1.5	1.1	0.83	0.72 to 0.89	.275
–0.1	1.0	–0.4 to 0.2	.515[d]	–0.03	0.8	0.7	0.6 to 1.0	0.7	0.96	0.93 to 0.97	< **.001**
–14	66	–31 to 3	.113[e]	–0.11	49	46	37 to 61	47	0.86	0.78 to 0.92	.059
–0.5	1.2	–0.8 to –0.2	**.004**[d]	–0.22	1.0	0.9	0.8 to 1.2	0.9	0.82	0.68 to 0.89	.379
0.1	2.2	–0.5 to 0.7	.773[d]	0.03	1.6	1.4	1.2 to 2.0	1.5	0.83	0.72 to 0.90	.271
0.2	3.4	–0.6 to 1.1	.582[f]	0.04	2.8	1.8	2.3 to 3.3	2.4	0.86	0.78 to 0.92	.064
0.6	1.7	0.1 to 1.0	**.018**[d]	0.21	1.3	1.3	1.0 to 1.7	1.3	0.79	0.66 to 0.87	.539
2	26	–5 to 9	.543[d]	0.06	20	16	16 to 24	18	0.68	0.52 to 0.80	.975
–0.1	1.2	–0.4 to 0.2	.399[e]	–0.10	0.9	0.8	0.7 to 1.1	0.9	0.60	0.40 to 0.74	.999
0.0	1.6	–0.4 to 0.4	.954[d]	0.00	1.2	1.0	1.0 to 1.5	1.1	0.89	0.82 to 0.93	**.006**
–0.7	2.6	–1.4 to 0.0	.055[e]	–0.12	2.1	1.8	1.6 to 2.5	1.9	0.90	0.83 to 0.94	**.005**
–0.5	2.7	–1.2 to 0.2	.15[d]	–0.11	2.0	1.9	1.5 to 2.5	1.9	0.86	0.77 to 0.91	.095
–0.1	2.0	–0.6 to 0.4	.685[d]	–0.03	1.6	1.3	1.3 to 1.9	1.4	0.78	0.65 to 0.86	.653
–23	117	–54 to 7	.13[d]	–0.10	92	75	72 to 111	83	0.86	0.78 to 0.92	.068
–0.9	4.2	–2.0 to 0.2	.102[d]	–0.15	3.1	3.0	2.3 to 3.8	3.0	0.76	0.62 to 0.85	.786

Table 4. Extended

Intra-assessor (T_1-T_1')								Inter-assessor (T_1-T_1'')							
Group Comparison						Group Agreement		Group Comparison						Group Agreement	
DD			AD					DD			AD				
Mean	SD	SDD[b]	Mean	SD	DE	ICC[c]	CI	Mean	SD	SDD	Mean	SD	DE	ICC	CI
0.0	0.1	0.00	0.1	0.1	0.1	1.00	1.00 to 1.00	–0.2	0.5	–0.09	0.4	0.3	0.4	0.98	0.95 to 0.99
0.0	0.1	0.00	0.1	0.1	0.1	1.00	1.00 to 1.00	–0.4	0.8	–0.13	0.6	0.5	0.6	0.96	0.90 to 0.99
2	7	0.01	5	5	5	1.00	1.00 to 1.00	–6	46	–0.04	30	35	32	0.95	0.89 to 0.98
0.0	0.3	0.01	0.2	0.2	0.2	0.99	0.99 to 1.00	–0.4	0.7	–0.17	0.7	0.4	0.6	0.94	0.85 to 0.98
0.1	0.3	0.02	0.2	0.2	0.2	1.00	0.99 to 1.00	–0.7	0.7	–0.21	0.7	0.7	0.7	0.96	0.73 to 0.99
0.0	0.4	0.00	0.3	0.2	0.3	1.00	1.00 to 1.00	0.4	1.7	0.06	1.2	1.2	1.2	0.96	0.91 to 0.98
−0.1	0.3	−0.03	0.3	0.2	0.2	0.99	0.99 to 1.00	–0.4	1.0	–0.14	0.9	0.6	0.8	0.94	0.86 to 0.97
5	15	0.15	6	15	11	0.86	0.72 to 0.94	8	21	0.27	17	16	16	0.69	0.42 to 0.85
−0.1	0.2	−0.11	0.2	0.2	0.2	0.97	0.92 to 0.99	–0.7	0.7	–0.71	0.8	0.6	0.7	0.66	0.03 to 0.87
0.0	0.2	−0.01	0.2	0.2	0.2	1.00	0.99 to 1.00	0.5	0.9	0.15	0.9	0.6	0.8	0.95	0.87 to 0.98
0.0	0.3	0.01	0.2	0.2	0.2	1.00	1.00 to 1.00	0.8	0.8	0.10	0.9	0.6	0.8	0.99	0.92 to 1.00
0.0	0.2	0.00	0.2	0.1	0.1	1.00	1.00 to 1.00	0.3	0.7	0.06	0.6	0.4	0.5	0.99	0.98 to 1.00
0.0	0.4	−0.01	0.3	0.2	0.3	0.99	0.99 to 1.00	–0.4	0.9	–0.12	0.8	0.5	0.7	0.96	0.90 to 0.98
−3	8	−0.01	7	5	6	1.00	1.00 to 1.00	–78	161	–0.33	122	129	124	0.75	0.47 to 0.88
0.0	0.3	−0.01	0.2	0.2	0.2	1.00	1.00 to 1.00	0.0	1.0	0.01	0.7	0.7	0.7	0.99	0.98 to 1.00

thickness and area. Accordingly, the AD in the tongue area was relevant, and a "medium" to "large" systematic bias (SDD) was present in the tongue area and soft palate thickness, respectively. The study findings confirmed that the area of contact between the two structures was critical but that the airway patency, which is the area of greater clinical relevance, was not affected.

Factors Affecting the Morphology of the Analyzed Structures

In adults, changes in the NHP and cranio-cervical inclination affect the hyoid bone position[23] and upper airway morphology.[23,24] Thus, although using the NHP allows good reliability in cephalometric analysis in children,[25] minor variations are not controllable and may have influenced the measurements (Figure 4). In addition, although patients with mouth opening were excluded (Figure 1) in the present study, slight changes in the mandibular posture did not determine exclusion and may have affected the position of the hyoid bone.[9]

Tongue posture and swallowing primarily affect upper airway morphology. During swallowing, a chain of muscular activities is triggered[26] and has a generalized effect on the airway. For this reason, patients swallowing or showing evidence of tongue movement were excluded (Figure 1). However, some variation in the tongue position cannot be prevented and inevitably affected the results (Figure 4).

Overall, in order to control these confounding variables, LCRs used to assess the upper airway should be taken in NHP, natural neck posture, light dental contacts in centric occlusion, during normal inspiration, without swallowing, and with the tongue in the rest position.

Overall Reproducibility of the Analyzed Structures

In general, given the excellent intra-assessor agreement, the variations between the two LCRs reported in this study should be mainly attributable to real morphologic changes. Furthermore, since no differences were found in LCRs taken at a 5- to 10-minute interval or at a 60- to 100-minute interval, the time interval may not affect their reliability, in agreement with a previous study in adults.[10]

The upper airway and hyoid bone measurements showed excellent reliability on the two consecutive LCRs (Figure 3). In addition, the DD, although significant in the case of the retropalatal oropharyngeal airway space ($P = .004$), was never clinically relevant, and the systematic bias was low or medium (Table 3).

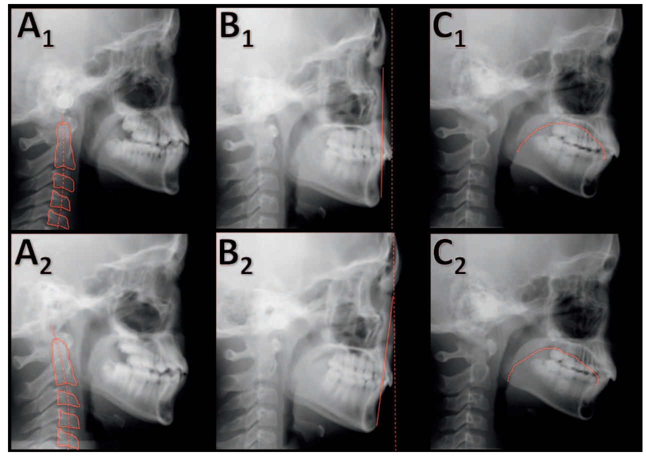

Figure 4. Examples of three nonexcluded cases (A, B, C) in which minor changes in neck posture (A_1 to A_2), head posture (B_1 to B_2), and tongue posture (C_1 to C_2) affected the upper airway size. The images on top show the situation with larger upper airway (1), and the images at the bottom show the same patient with the upper airway narrowed (2).

In particular, the variables in which reliability was significantly higher than excellent were the nasopharyngeal airway space, horizontal position of the hyoid bone with respect to the vertebrae, and vertical position of the hyoid bone with respect to the Frankfort plane (Figure 3). These findings were in disagreement with Stepovich,[8] who found the reliability of the position of the hyoid bone to be questionable. However, beside the very limited sample size of this previous study, the examinations were made when subjects were seated.[8] In fact, Malkoc et al.[10] by using the NHP in standing patients found good reproducibility of the hyoid bone position.

Conversely, the tongue measurements showed lower reliability, ranging from good to excellent, and the soft palate showed even poorer results (Figure 3). Accordingly, higher absolute differences and errors were present up to 3.1 mm and 92 mm^2, confirming these two movable structures to be the most critical. Although radiopaque pastes can enhance the visibility of the tongue borders,[6] the effect is limited to the tongue dorsum and might be necessary to adopt different imaging techniques, such as magnetic resonance imaging[27] for proper assessment of these structures.

This said, the poor reliability in the soft palate measurement was in contrast with a study by Malkoc et al.[10] However, this former study was performed in adults, and hyoid movements[9] and oropharyngeal reflexes[26] are different in children. Furthermore, although the Pearson correlation coefficient used by Malkoc et al.[10] is appropriate for detecting linear associations, it does not properly represent the reliability.

Limitations

Reliability is fundamental in order to use LCRs for diagnosis. If the LCR produces the same results multiple times and independently from the assessor, it can be considered reliable. However, reliability is independent from validity, which expresses, for example, whether the obstruction measured on an LCR is representative of the real obstruction of the patient. The present study aimed at investigating the reliability of LCRs and did not consider its validity for the diagnosis of upper airway obstruction. In particular, the diagnosis and treatment of upper airway disorders in children require a comprehensive medical approach,[2] and it is important to critically analyze the potential role of LCRs in such contexts.

CONCLUSIONS

- The measurement of the upper airway morphology, considered the intramural space, and the hyoid bone position were highly reliable on LCRs of children.
- The limited reliability in the assessment of tongue and soft palate may compromise the diagnostic application of these parameters on LCRs, and different imaging methods might be advisable.
- The reliability of LCRs taken in the NHP, under gentle occlusion, and by instructing the patient to refrain from swallowing, may be affected by such factors as minor variations in head posture, neck posture, and tongue movements. It is important to minimize these variables in order to improve its clinical relevance.

ACKNOWLEDGMENTS

The authors declare no conflict of interest. We are grateful to Ms Samantha K. Y. Li for her assistance with the statistical analysis. We also thank Dr Michael S. Cooke for the collection of lateral cephalograms during the preparation of his PhD thesis, which made possible the present retrospective study.

REFERENCES

1. Pirila-Parkkinen K, Lopponen H, Nieminen P, et al. Cephalometric evaluation of children with nocturnal sleep-disordered breathing. *Eur J Orthod*. 2010;32:662–671.
2. Katz ES, D'Ambrosio CM. Pathophysiology of pediatric obstructive sleep apnea. *Proc Am Thorac Soc*. 2008;5:253–262.
3. Katyal V, Pamula Y, Martin AJ, et al. Craniofacial and upper airway morphology in pediatric sleep-disordered breathing: systematic review and meta-analysis. *Am J Orthod Dentofacial Orthop*. 2013;143:20–30.e23.
4. Juliano ML, Machado MA, de Carvalho LB, et al. Polysomnographic findings are associated with cephalometric measurements in mouth-breathing children. *J Clin Sleep Med*. 2009;5:554–561.
5. Samman N, Mohammadi H, Xia J. Cephalometric norms for the upper airway in a healthy Hong Kong Chinese population. *Hong Kong Med J*. 2003;9:25–30.
6. Johal A, Conaghan C. Maxillary morphology in obstructive sleep apnea: a cephalometric and model study. *Angle Orthod*. 2004;74:648–656.
7. Pracharktam N, Hans MG, Strohl KP, et al. Upright and supine cephalometric evaluation of obstructive sleep apnea syndrome and snoring subjects. *Angle Orthod*. 1994;64:63–73.
8. Stepovich ML. A cephalometric positional study of the hyoid bone. *Am J Orthod*. 1965;51:882–900.
9. Ingervall B, Carlsson GE, Helkimo M. Change in location of hyoid bone with mandibular positions. *Acta Odontol Scand*. 1970;28:337–361.
10. Malkoc S, Usumez S, Nur M, et al. Reproducibility of airway dimensions and tongue and hyoid positions on lateral cephalograms. *Am J Orthod Dentofacial Orthop*. 2005;128: 513–516, 2005.
11. Cooke MS. *Cephalometric Analyses Based on Natural Head Posture of Chinese Children in Hong Kong* [doctoral dissertation]. Hong Kong: University of Hong Kong; 1986.
12. Solow B, Tallgren A. Natural head position in standing subjects. *Acta Odontol Scand*. 1971;29:591–607.
13. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9: 671–675.
14. Shen GF, Samman N, Qiu WL, et al. Cephalometric studies on the upper airway space in normal Chinese. *Int J Oral Maxillofac Surg*. 1994;23:243–247.
15. Gu M, McGrath CP, Wong RW, et al. Cephalometric norms for the upper airway of 12-year-old Chinese children. *Head Face Med*. 2014;10:38.
16. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101–110.
17. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378–382.
18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, Mahwah, New Jersey, USA: Lawrence Erlbaum Associates; 1988.
19. Dahlberg G. *Statistical Methods for Medical and Biological Students*. London: George Allen and Unwin; 1940.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
21. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–428.
22. Blacker D. Psychiatric rating scales. In Sadock BJ, Sadock V, eds: *Comprehensive Textbook of Psychiatry*. 8th ed, Philadelphia, PA: Lippincott Williams & Wilkins;2005:929–955.
23. Muto T, Takeda S, Kanazawa M, et al. The effect of head posture on the pharyngeal airway space (PAS). *Int J Oral Maxillofac Surg*. 2002;31:579–583.
24. Anegawa E, Tsuyama H, Kusukawa J. Lateral cephalometric analysis of the pharyngeal airway space affected by head posture. *Int J Oral Maxillofac Surg*. 2008;37:805–809.
25. Cooke MS. Five-year reproducibility of natural head posture: a longitudinal study. *Am J Orthod Dentofacial Orthop*. 1990; 97:489–494.
26. Ruark JL, McCullough GH, Peters RL, et al. Bolus consistency and swallowing in children and adults. *Dysphagia*. 2002;17:24–33.
27. Schwab RJ, Kim C, Bagchi S, Keenan BT, et al. Understanding the anatomic basis for obstructive sleep apnea syndrome in adolescents. *Am J Respir Crit Care Med*. 2015;191:1295–309.