

Cephalometric landmark variability among orthodontists and dentomaxillofacial radiologists: a comparative study

Ana Paula Reis Durão^{1,*}, Aline Morosolli², Pisha Pittayapat³, Napat Bolstad⁴, Afonso P. Ferreira⁵, Reinhilde Jacobs⁶

¹Department of Dental Radiology, Faculty of Dental Medicine, University of Porto, Porto, Portugal

²Department of Surgery, Dentistry School, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

³Department of Radiology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand

⁴Department of Clinical Dentistry, Faculty of Health Science, UiT The Arctic University of Norway, Tromsø, Norway

⁵Department of Orthodontics, Faculty of Dental Medicine, University of Porto, Portugal

⁶Oral Imaging Center, OMFS-IMPACT Research Group, Department of Imaging and Pathology, Faculty of Medicine, University of Leuven, Leuven, Belgium

ABSTRACT

Purpose: The aim this study was to compare the accuracy of orthodontists and dentomaxillofacial radiologists in identifying 17 commonly used cephalometric landmarks, and to determine the extent of variability associated with each of those landmarks.

Materials and Methods: Twenty digital lateral cephalometric radiographs were evaluated by two groups of dental specialists, and 17 cephalometric landmarks were identified. The *x* and *y* coordinates of each landmark were recorded. The mean value for each landmark was considered the best estimate and used as the standard. Variation in measurements of the distance between landmarks and measurements of the angles associated with certain landmarks was also assessed by a subset of two observers, and intraobserver and interobserver agreement were evaluated.

Results: Intraclass correlation coefficients were excellent for intraobserver agreement, but only good for interobserver agreement. The least reliable landmark for orthodontists was the gnathion (Gn) point (standard deviation [SD], 5.92 mm), while the orbitale (Or) was the least reliable landmark (SD, 4.41 mm) for dentomaxillofacial radiologists. Furthermore, the condylion (Co)-Gn plane was the least consistent (SD, 4.43 mm).

Conclusion: We established that some landmarks were not as reproducible as others, both horizontally and vertically. The most consistently identified landmark in both groups was the lower incisor border, while the least reliable points were Co, Gn, Or, and the anterior nasal spine. Overall, a lower level of reproducibility in the identification of cephalometric landmarks was observed among orthodontists. (*Imaging Sci Dent* 2015; 45: 213-20)

KEY WORDS: Cephalometry; Analysis; Anatomic Landmarks; Orthodontics; Reliability and Validity

Introduction

Since its introduction by Broadbent¹ in 1931, lateral cephalometric radiography has been widely used in orthodontics. It is used to characterize facial morphology, to

predict the growth of the facial skeleton, to plan orthodontic treatment, and to evaluate treatment outcomes.² Cephalometric analyses also provide angular and linear measurements useful for diagnostic purposes and planning orthodontic treatment. Errors in cephalometric analysis may occur for numerous reasons. One of the most important types of errors involves inconsistent and imprecise landmark identification. Inaccurate landmark identification may lead to erroneous diagnoses and treatment plans.³⁻⁶ The identification of certain anatomical landmarks, such

Received April 23, 2015; Revised May 21, 2015; Accepted May 27, 2015

*Correspondence to : Dr. Ana Reis Durão

Faculdade de Medicina Dentária da Universidade do Porto, Rua Dr. Manuel Pereira da Silva, 4200-393 Porto, Portugal

(Tel) 351-91-9991914, (Fax) 351-22-20173013, (E-mail) paula.o.reis@gmail.com

Copyright © 2015 by Korean Academy of Oral and Maxillofacial Radiology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Imaging Science in Dentistry · pISSN 2233-7822 eISSN 2233-7830

as the porion (Po), condylion (Co), orbitale (Or), basion, gonion (Go), anterior nasal spine (ANS), posterior nasal spine (PNS), and lower inferior apex (LIA), may be more prone to error due to overlapping structures superimposed on the landmark and its location.² Likewise, the quality of radiographic images can interfere with the identification of some landmarks, such as Po, Co, Or, ANS, point B, the pogonion (Pog), Go, and the glabella.^{7,8} Moreover, some authors have argued that the level of an observer's knowledge and his or her professional background play an important role in landmark identification.⁷⁻¹⁰ Other authors have considered the possibility that errors could be caused by diverse individual conceptions of how landmarks are defined, rather than by discrepancies in education and training.^{8,11} Inconsistency in the identification of multiple landmarks could further increase the magnitude of the error.^{3,11,12} Interobserver reproducibility of landmark identifications was found to be very low among dentomaxillofacial radiologists.¹² Some dentomaxillofacial radiologists, as well as orthodontists, are trained to perform two-dimensional (2D) cephalometric analyses. No previous reports have compared orthodontists and dentomaxillofacial radiologists regarding the reliability of landmark identification. Therefore, the aim of the present study was to evaluate the reproducibility of 17 commonly used cephalometric landmarks by orthodontists and dentomaxillofacial radiologists.

Materials and Methods

Twenty digital lateral cephalometric radiographs were selected from the database of the Oral Imaging Center, University of Leuven. Lateral cephalograms were acquired by positioning the patients in a standard digital cephalometric device, using a charge-coupled device sensor (Veraviewepocs 2D[®], J. Morita, Kyoto, Japan). The exposure values were set at 77 kV and 7.2 mA, with an exposure time of approximately 1.6 s, depending on the patient. The inclusion criteria were 1) no evidence of current orthodontic treatment; 2) a sufficiently high-quality digital cephalometric image for landmark identification, with the ruler clearly visible on the film, allowing image calibration in the cephalometric analysis software program; 3) no unerupted or partially erupted teeth that could compromise landmark identification; 4) no gross skeletal asymmetry. All selected images were exported in the TIFF format and subsequently imported into the computer program used for cephalometric analysis (Radiocef Studio 2, Radio Memory Ltd., Belo Horizonte, Brazil).

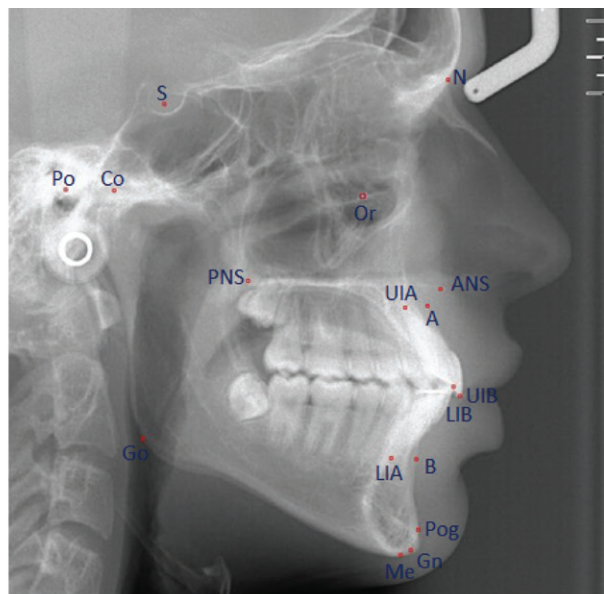


Fig. 1. Cephalometric landmarks used in the study. N, nasion; Or, orbitale; S, sella; Co, condylion; Po, porion; PNS, posterior nasal spine; ANS, anterior nasal spine; A, point A; UIA, upper incisor apex; UIB, upper incisor border; LIB, lower incisor border; LIA, lower incisor apex; B, point B; Pog, pogonion; Gn, gnathion; Me, menton; Go, gonion.

Seventeen commonly used cephalometric landmarks were included in this analysis (Fig. 1). Landmark identification was carried out on the digital image using a mouse-driven cursor in a predetermined sequence.

Eight experienced observers, including four orthodontists and four dentomaxillofacial radiologists, performed this study. The experience of the observers ranged from eight to 15 years. An initial training and calibration session was attended by all eight observers, including an explanation of the anatomical structures and the landmarks they were required to identify. At the end of the session, the main author (ARD) responded to any remaining questions. Thus, all observers followed the same landmark definitions in the identification process. For optimal visualization, landmark identification was performed in a dimly lit room without any interruptions. Interobserver reliability was evaluated. The same procedure was repeated three months later by all eight observers. Intraobserver agreement was also assessed based on the performance of one dentomaxillofacial radiologist who repeated this procedure six months after the first observation.

After selecting a landmark with the mouse cursor, a dot on the image indicated the position of the landmark. The landmark position could be corrected until the operator was satisfied. The vertical and horizontal positions of

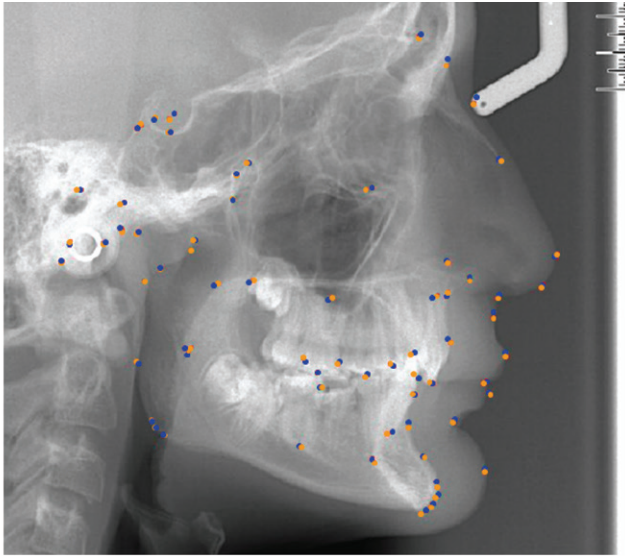


Fig. 2. Example of a lateral cephalometric radiograph with the identification of landmarks by two observers.

each landmark were recorded as x and y coordinates.

The landmarks' digitized coordinates were then imported into Excel (version 2003; Microsoft, Redmond, WA, USA). Statistical analysis was performed using SPSS version 20.0 for Windows (IBM Corp., Armonk, NY, USA). The level of statistical significance for all tests was set at $\alpha=0.05$. We determined which group was closer to the standard measurement, as defined by the average of all measurements. In addition, some linear and angular measurements used in Ricketts¹³ and McNamara's¹⁴ cephalometric analysis were performed with the aid of the cephalometric software. Three radiographs were classified as borderline cases, in between orthognathic surgery and orthodontics. In general, the variability in landmark identification ranged between 1 mm and 2 mm (Fig. 2).

The same computer software was used to access the variability in the angular and linear measurements. The angular and linear measurements used were the following: A-nasion (N), Co-Gn (mandibular unit length), Co-point A (A), Po-Or, Go-menton (Me), Pog-N (facial plane), lower incisor border (LIB) (A-Pog), convexity of A; Go-Me (mandibular plane) and sella (S)-Go.

The mean, standard deviation, and measurements of dispersion were calculated for each landmark in order to assess the extent of variation. Intraobserver and interobserver variation for each landmark in the x and y axes were evaluated, using intraclass correlation coefficients (ICCs) with a confidence interval of 95%. According to the general guidelines for this measure, an ICC >0.90 indicates excellent agreement, an ICC of 0.75-0.90 reflects

good agreement, and an ICC <0.75 represents poor to moderate reliability.¹⁵

The best estimate of the location of each landmark was defined as the mean value of the x and y coordinates of each landmark as identified by the eight observers, and this estimate was used as the standard for assessing variability. The average distance between the mean positions identified by each observer was calculated to identify interobserver error. Differences in the location of landmarks were analyzed using the Student's t -test with a significance level of $p<0.05$. Interobserver reliability was assessed using Euclidean distances.

Results

ICCs were calculated to assess intraobserver and interobserver variation in each group (Table 1), and these findings were compared between groups. In general, the intraobserver ICC was >0.90 , indicating excellent agreement. Exceptions were the x component of Po, Me, and point B and the y component of N, Or, and S, which showed good agreement (ICC between 0.75 and 0.90). Furthermore, the vertical components of Go and point B demonstrated poor or moderate agreement (ICC <0.75) in the intraobserver evaluation. The most variation was associated with the vertical component of Go (1.73 mm), and the least was seen in the vertical component of Po (0.04 mm).

In general, the ICC was >0.90 for interobserver error, with exception of the x components of N and Or, which showed good agreement (Table 1). The ICC was lower for orthodontists, with good agreement (ICC between 0.75 and 0.90) regarding Or, Po, Gn, point B, and UIA. Likewise, the x coordinates of N, Me, Pog, A, PNS, LIA, and LIB also showed good agreement. Poor or moderate agreement was only found in the x coordinates of ANS. Dentomaxillofacial radiologists demonstrated higher ICCs, with ICCs over 0.90 for most landmarks. The exceptions were the x coordinates of Or and Po and the y coordinates of Go and point B, for which good agreement was observed. Poor or moderate agreement was observed for the y component of Or. Generally, interobserver reliability was excellent among dentomaxillofacial radiologists.

Overall, in both groups, Co was associated with high variation in the x dimension. The largest difference was of 5.05 mm observed between two dentomaxillofacial observers, in contrast to a difference of 3.56 mm observed among orthodontists. The horizontal component of Or

Table 1. Intraclass correlation coefficients (ICCs) reflecting interobserver and intraobserver variability.

	Interobserver*		Intraobserver**	
	ICC	95% CI	ICC	95% CI
N: nasion				
<i>x</i>	0.78	0.310-0.848	0.99	0.985-0.998
<i>y</i>	0.99	0.985-0.998	0.85	0.301-0.845
Or: orbitale				
<i>X</i>	0.88	0.726-0.951	0.91	0.797-0.965
<i>Y</i>	0.96	0.923-0.987	0.78	0.202-0.812
S: sella				
<i>x</i>	0.99	0.984-0.997	0.99	0.976-0.996
<i>y</i>	0.98	0.962-0.994	0.87	0.701-0.946
Po: porion				
<i>x</i>	0.97	0.916-0.986	0.80	0.569-0.917
<i>y</i>	0.95	0.889-0.982	0.94	0.845-0.974
Co: condyilion				
<i>x</i>	0.91	0.797-0.965	0.94	0.856-0.976
<i>y</i>	0.97	0.917-0.986	0.96	0.899-0.983
Go: gonion				
<i>x</i>	0.98	0.956-0.993	0.98	0.964-0.994
<i>y</i>	0.93	0.830-0.971	0.71	0.394-0.872
Me: menton				
<i>x</i>	0.97	0.918-0.987	0.84	0.631-0.932
<i>y</i>	0.99	0.970-0.995	0.97	0.932-0.989
Pog: pogonion				
<i>x</i>	0.98	0.941-0.990	0.98	0.949-0.992
<i>y</i>	0.98	0.951-0.992	0.98	0.951-0.992
Gn: gnathion				
<i>x</i>	0.97	0.924-0.988	0.97	0.936-0.990
<i>y</i>	0.99	0.963-0.994	0.99	0.954-0.979
Point B				
<i>x</i>	0.97	0.930-0.989	0.89	0.754-0.957
<i>y</i>	0.96	0.894-0.983	0.72	0.418-0.879
Point A				
<i>x</i>	0.97	0.916-0.986	0.92	0.804-0.966
<i>y</i>	0.98	0.960-0.994	0.95	0.890-0.982
ANS				
<i>x</i>	0.93	0.823-0.970	0.95	0.868-0.978
<i>y</i>	0.99	0.968-0.995	0.97	0.939-0.990
PNS				
<i>x</i>	0.98	0.952-0.992	0.94	0.852-0.975
<i>y</i>	0.99	0.972-0.996	0.99	0.969-0.995
LIA				
<i>x</i>	0.97	0.932-0.989	0.97	0.917-0.987
<i>y</i>	0.97	0.937-0.990	0.99	0.969-0.995
LIB				
<i>x</i>	0.98	0.961-0.994	0.98	0.960-0.994
<i>y</i>	0.99	0.968-0.995	0.99	0.981-0.997
UIB				
<i>x</i>	0.98	0.951-0.992	0.98	0.960-0.994
<i>y</i>	0.98	0.947-0.991	0.98	0.963-0.994
UIA				
<i>x</i>	0.97	0.925-0.988	0.98	0.956-0.993
<i>y</i>	0.97	0.924-0.988	0.98	0.944-0.991

* Among the eight observers, **A dentomaxillofacial radiologist.
CI, confidence interval; UIA, upper incisor apex; UIB, upper incisor border; LIB, lower incisor border; LIA, lower incisor apex.

was less reproducible among the dentomaxillofacial radiologists. Orthodontists demonstrated lower reproducibility of Go in the *x* and *y* axes, Me and the posterior nasal spine in the *x* axis, and point B in the *y* axis.

The Euclidean distance was used to assess variability in landmark identification between each observer and in comparison to the standard measurement. The mean location differences of all landmarks among orthodontists ranged from 0.99 mm to 5.92 mm. The landmark with the lowest variability was LIB (0.99 mm; SD, 0.65 mm), while the highest variability was found for Gn (5.92 mm; SD, 4.59 mm). The minimal and maximal extent of horizontal variation were associated with ANS (minimum 0.53 mm; SD, 3.74 mm and maximum 2.97 mm; SD, 2.02 mm). Regarding reproducibility of the vertical component of the location, A presented the minimum variation (0.10 mm; SD, 1.86 mm), while Gn was the most variable (4.60 mm; SD, 3.67 mm). Table 2 shows the Euclidean distances between the best estimate of each landmark and the locations identified by orthodontists, which was defined as the interobserver error of landmark identification. In general, orthodontists demonstrated less than 1 mm of error in S, Pog, LIB, and the upper incisor border in both the horizontal and vertical directions.

Dentomaxillofacial radiologists displayed less than 1 mm of error in both the horizontal and vertical directions for N, S, A, and LIA. Less variation was seen between the best estimate of each landmark and the points identified by dentomaxillofacial radiologists (Table 3). The greatest average Euclidean distance was observed for Or (4.41 mm; SD, 2.04 mm) and the lowest for LIB (0.84 mm; SD, 0.46 mm). The landmarks with the minimal and the maximal horizontal variation were LIB (0.08 mm; SD, 0.81) and Or (3.94 mm; SD, 2.51 mm), respectively. The landmark with the least variability in the vertical component of its location was LIB (0.10 mm; SD, 1.08 mm), and the landmark with the most variability was Gn (2.28 mm; SD, 1.65 mm). Only the errors associated with S, LIB, and the A point were, overall, less than 1 mm in both directions. The best estimate for each landmark was defined as the mean position identified by eight observers.

Despite an overall level of variation lower than the acceptable value of 2 mm, some landmarks presented higher levels of variation. Some dentomaxillofacial radiologists had errors of more than 2 mm in the horizontal dimension for Or, Po, ANS, Go, and Gn. Orthodontists demonstrated more than 2 mm of error for the *x* coordinates of Or, Po, Co, Go, Gn, ANS, PNS, and the upper incisor apex (UIA). Statistically significant differences ($p < 0.05$) in the in-

Table 2. Minimum and maximum Euclidean distances (in mm) for orthodontists, defined as the absolute difference in millimeters between the mean value and standard deviation of each landmark, averaged across all observers.

Landmark	Horizontal component (x)			Vertical component (y)		
	Mean	SD	<i>p</i>	Mean	SD	<i>p</i>
N: nasion						
Minimum	0.19	3.30	0.804	-0.34	0.74	0.054
Maximum	1.43	2.74	0.031	0.21	1.89	0.631
Or: orbitale						
Minimum	0.24	1.46	0.650	0.20	1.74	0.902
Maximum	2.04	1.85	<0.001	-1.29	2.13	0.014
S: sella						
Minimum	0.24	0.92	0.263	-0.32	0.67	0.048
Maximum	-0.42	0.80	0.029	-0.76	1.09	0.006
Po: porion						
Minimum	0.53	1.59	0.155	0.44	1.47	0.194
Maximum	-2.60	2.08	<0.001	-1.42	2.28	0.012
Co: condyilion						
Minimum	0.23	1.69	0.549	-2.61	2.11	<0.001
Maximum	-2.66	1.13	<0.001	1.15	2.68	0.069
Go: gonion						
Minimum	0.50	1.34	0.110	-1.09	2.47	0.063
Maximum	2.20	1.83	<0.001	2.14	2.36	0.001
Me: menton						
Minimum	0.73	3.77	0.395	-0.42	1.84	0.323
Maximum	-1.57	2.38	0.008	1.44	2.64	0.025
Pog: pogonion						
Minimum	0.27	2.30	0.430	-0.16	1.99	0.984
Maximum	0.42	3.31	0.576	0.61	2.23	0.238
Gn: gnathion						
Minimum	-0.31	2.22	0.025	0.21	2.00	0.847
Maximum	2.27	4.11	0.023	4.60	3.67	0.000
Point B						
Minimum	0.20	1.84	0.725	0.21	2.87	0.759
Maximum	0.87	3.79	0.319	3.38	2.73	<0.001
Point A						
Minimum	0.17	2.38	0.754	0.10	1.32	<0.001
Maximum	1.27	2.40	0.029	2.08	2.25	0.001
ANS						
Minimum	0.16	3.06	0.897	0.22	1.73	0.599
Maximum	2.97	3.74	0.002	-0.81	1.10	0.004
PNS						
Minimum	0.54	1.44	0.113	0.34	1.26	0.240
Maximum	-2.66	1.43	<0.001	-0.80	1.23	0.009
LIA						
Minimum	-0.38	2.38	0.495	-0.37	2.51	0.516
Maximum	-1.44	2.01	0.005	2.61	2.53	0.010
LIB						
Minimum	-0.10	2.01	0.781	0.24	2.19	0.631
Maximum	0.40	2.39	0.461	1.00	2.59	0.048
UIB						
Minimum	-0.23	1.71	0.556	-0.21	2.13	0.659
Maximum	-0.68	1.59	0.069	1.39	2.16	0.010
UIA						
Minimum	0.33	2.50	0.566	0.21	2.41	0.896
Maximum	-1.20	1.87	0.010	-2.41	1.61	0.000

SD, standard deviation; ICC, intraclass correlation; CI, confidence interval; PNS, posterior nasal spine; ANS, anterior nasal spine; UIA, upper incisor apex; UIB, upper incisor border; LIB, lower incisor border; LIA, lower incisor apex.

Table 3. Minimum and maximum Euclidean distances (in mm) for dentomaxillofacial radiologists, defined as the absolute difference in millimeters between the mean value and standard deviation of each landmark, averaged over all observers.

Landmark	Horizontal component (x)			Vertical component (y)		
	Mean	SD	<i>p</i>	Mean	SD	<i>p</i>
N: nasion						
Minimum	-1.94	2.20	0.001	0.19	0.50	0.101
Maximum	0.15	2.30	0.952	0.33	0.97	0.148
Or: orbitale						
Minimum	-0.47	1.78	0.248	0.20	0.71	0.737
Maximum	3.94	2.51	<0.001	-2.20	6.73	0.160
S: sella						
Minimum	-0.20	0.63	0.022	-0.66	1.02	0.009
Maximum	0.45	0.29	<0.001	1.90	3.24	0.017
Po: porion						
Minimum	0.26	1.07	0.284	0.21	0.89	0.453
Maximum	2.08	2.69	0.003	1.65	1.13	<0.001
Co: condyilion						
Minimum	0.48	1.69	0.216	-0.18	1.34	0.545
Maximum	1.66	1.30	<0.001	0.80	1.39	0.018
Go: gonion						
Minimum	-1.19	0.71	<0.001	-2.14	1.95	0.000
Maximum	0.25	0.97	0.751	0.33	1.97	0.965
Me: menton						
Minimum	-0.37	1.45	0.268	-1.55	1.67	0.001
Maximum	1.51	1.21	<0.001	0.49	1.67	0.205
Pog: pogonion						
Minimum	0.60	0.73	0.961	0.19	1.26	0.608
Maximum	-1.10	1.40	0.002	-1.65	1.57	<0.001
Gn: gnathion						
Minimum	-0.20	0.94	0.353	0.17	1.59	0.880
Maximum	-1.36	1.57	0.001	-2.28	1.65	<0.001
Point B						
Minimum	0.39	1.97	0.086	0.99	2.21	0.060
Maximum	-1.33	1.43	0.001	-1.50	2.15	0.005
Point A						
Minimum	-0.32	0.72	0.706	0.21	1.28	0.898
Maximum	-1.00	1.39	0.001	-1.00	1.03	<0.001
ANS						
Minimum	0.25	1.37	0.421	0.57	0.91	0.012
Maximum	-2.62	1.14	<0.001	-0.63	1.71	0.115
PNS						
Minimum	0.16	1.31	0.818	0.14	0.83	0.448
Maximum	-0.17	1.51	0.616	-0.81	1.05	0.003
LIA						
Minimum	0.93	1.42	0.008	-0.39	1.29	0.189
Maximum	-1.66	1.39	<0.001	-2.09	1.37	<0.001
LIB						
Minimum	0.08	0.81	0.664	0.10	1.08	0.691
Maximum	-1.03	1.51	0.002	-1.02	1.30	0.002
UIB						
Minimum	0.30	1.06	0.215	-0.19	1.08	0.431
Maximum	-1.21	1.41	0.001	-1.25	1.47	0.001
UIA						
Minimum	0.67	1.06	0.011	0.48	0.88	0.025
Maximum	-0.98	1.29	0.003	0.76	1.06	0.005

SD, standard deviation; ICC, intraclass correlation; CI, confidence interval; PNS, posterior nasal spine; ANS, anterior nasal spine; UIA, upper incisor apex; UIB, upper incisor border; LIB, lower incisor border; LIA, lower incisor apex.

Table 4. Standard deviation for each linear and angular measurement, as performed by two observers in radiographs from 20 patients.

Patient	A-N (SD± 2.70 mm)	Co-Gn	Co-A (SD± 6.00 mm)	(Po-Or). (Go-Me) (SD±3.90°)	Pog-N (SD± 3.80 mm)	A-Pog (SD± 2.40 mm)	Convexity of A	Go-Me (SD± 5.00 mm)	S-Go (SD± 6.00 mm)
P1	0.05	0.47	1.19	1.86	1.03	0.93	0.04	3.61	2.52
P2	0.78	1.90	0.58	0.18	0.71	0.31	0.41	3.26	1.63
P3	0.31	0.97	1.75	0.91	0.29	0.38	0.07	1.03	0.40
P4	0.03	0.71	1.84	0.87	0.02	0.89	0.14	2.07	0.17
P5	1.20	0.25	0.45	1.84	1.68	0.15	0.23	1.12	0.08
P6	1.23	0.34	1.42	0.29	0.13	1.28	1.09	0.54	1.22
P7	3.13	0.68	2.49	0.60	2.11	0.25	0.42	0.54	0.76
P8	0.49	1.33	0.08	1.66	2.06	0.96	0.61	0.78	0.01
P9	0.93	1.10	0.08	1.82	0.26	0.09	0.95	0.69	1.73
P10	0.46	0.23	0.78	1.04	1.40	0.43	0.27	1.81	1.36
P11	0.01	0.35	1.34	0.99	0.52	0.18	0.29	0.08	0.79
P12	0.41	3.73	1.14	0.42	0.23	0.31	0.30	4.18	2.60
P13	0.54	4.43	3.16	0.94	1.44	0.09	0.36	1.85	2.90
P14	0.51	2.60	0.52	0.64	1.03	0.18	0.01	2.15	0.88
P15	1.82	3.45	0.59	1.96	0.01	0.58	0.57	0.20	1.34
P16	2.65	0.30	0.85	2.36	0.27	0.37	0.72	3.74	2.76
P17	1.50	3.45	1.72	1.15	1.53	0.01	1.15	0.45	2.14
P18	1.07	1.15	1.35	1.15	1.43	0.35	0.37	2.93	1.67
P19	0.63	3.15	1.76	1.55	2.20	0.21	0.77	0.78	3.78
P20	0.16	1.47	2.34	0.26	0.64	0.10	0.49	1.46	1.26

N, nasion; Or, orbitale; S, sella; Co, condylion; Po, porion, A, point A; Gn, gnathion; Me, menton; Go, gonion.

terobserver error in the identification of certain landmarks were observed.

In both groups, the reliability of N, Or, Me, ANS, and UIA was better in the vertical direction. In contrast, the consistency for Go was greater in the horizontal direction. Additionally, orthodontists displayed less variance in the vertical component of Gn, point B and Pog, whereas S was more reproducible in the horizontal dimension. When linear and angular measurements were evaluated, the SDs were relatively small and never exceeded the SDs, that have been previously reported in the literature (Table 4). The largest SD was observed for the linear measurement of Co-Gn (mandibular unit length) (4.43 mm) and the lowest range of variation was observed for A-Pog (0.10 mm). Co and Gn were the least reliable landmarks. We also drifted significant changes in the SNA angle for three patients. However, the identification of the skeletal class of each image was consistent for both observers.

Discussion

Projection and tracing errors are a major class of errors that occur in 2D cephalometric analysis, and the most important source of tracing errors is landmark identification and measuring.^{2,3} Landmark identification is significantly affected by operator experience. It is known that intraob-

server error is generally lower than interobserver error.^{2,3,6} We found and high intraobserver agreement in landmark identification (ICC > 0.90). Previous studies have shown a relatively high rate of interobserver errors in landmark identification. Da Silveira and Silveira¹⁰ found a very low level of reproducibility among dentomaxillofacial radiologists in landmark identification. Similarly, we found significant interobserver variation in landmark identification. Some authors have proposed that individual perceptions of the definition of each landmark could lead to variations in angular and linear measurements.^{7,11} Nonetheless, even in severe patient conditions cases, the accuracy of cephalometric analysis was not affected.¹⁴ Some authors have argued that landmark identification errors of less than 1 mm are clinically acceptable.^{3,15} It has also been suggested that errors of less than 2° or 2 mm would most likely not make a significant difference in treatment.^{3,5,6,8,11}

This study attempted to evaluate the occurrence of discrepancies in landmark identification between orthodontists and dentomaxillofacial radiologists. The average value of the measurements performed by all observers was used as the standard for a specific landmark to quantify the error.³ The validity of any measure obtained by cephalometric radiography depends primarily on the reproducibility of cephalometric landmarks. Our study included 20 cephalometric lateral radiographs, which was a sufficient-

ly large sample to ensure the credibility of our statistical analyses.

Interobserver error was used as a variable to determine the reliability of measurements, as reflected by the dispersion of error around the best estimate for each landmark. Extensive differences in landmark identification were observed. Nevertheless, these differences would probably have had a low clinical impact. In general, we found statistically significant differences in the horizontal components of many landmarks in both groups. The dentomaxillofacial radiologists group showed greater variation regarding the horizontal components of Or, Me, ANS, and LIA. Orthodontists showed significant variability in ANS, Or, Po, Co, and Me. Reliability is an important aspect of measurement. If a metric cannot be reproduced consistently, then the value of the methodology in terms of cost, time, and patient treatment decisions is questionable.¹⁶ Landmark identification has been shown to be significantly affected by operator experience, which may be as important as the tracing method itself.¹⁷

Many factors can interfere with the reliability of cephalometric landmark identification, including the nature of cephalometric landmarks, the resolution and quality of digital images, and the training level or experience of the observers.^{3,18} In this study, all observers had significant experience in cephalometric analysis. Studies have shown that observer experience leads to a wide range of variation, but the degree of error is similar among observers with the same training background.^{11,19} Although dentomaxillofacial radiologists are not trained to clinically evaluate patients, they are trained to evaluate radiographic images. It is possible that even trained professionals who have completed calibration programs report incorrect values of cephalometric parameters. In fact, the reference points that were identified inconsistently have been described in a vague way in the literature, which might contribute to the uncertainty of the localization of these points (Co, Gn, Or, and ANS). One of the major causes of error in cephalometric landmark identification is the specific features of individual landmarks. The superimposition of adjacent structures can make it more difficult to identify certain landmarks, such as Co and Po, on radiographs. In a 2009 study, Chien et al.²⁰ found a high degree of variation for the vertical component of Go. In our study, the error in the identification of this landmark was only less than 1 mm for one orthodontist and one dentomaxillofacial radiologist. These findings can be attributed to the difficulty of establishing the landmark associated with broadly curved structures.

Some cephalometric landmarks are more reliable in either the horizontal or vertical plane, meaning that the distribution of error is asymmetric.¹ Differences in landmark identification were found between both groups along the horizontal and vertical axes. Overall, the differences on the horizontal axis were greater than those on the vertical axis in both groups.

Despite the low reproducibility of some major landmarks, only the x and y components of Or, Go, Gn, and LIA, the x coordinates of Po, ANS, Co, PNS, and the y component of point B showed a mean value of intraobserver error higher than 2 mm. In contrast, Gn, showed low intraobserver and interobserver variation with respect to the x coordinates. This result differed from the findings of Medelnic et al.,²¹ who found high variation in the identification of the x and y coordinates of Gn. A possible explanation proposed by Baumrind et al.² is that reference points located on a prominence or curvature, such as Gn, may have higher variability than points in flat areas.

We found differences in the SNA angle in three patients. Apart from that, we found no differences in the skeletal classification associated with a variation of 1-2 mm in these landmark identification. Higher levels of error would most likely affect the diagnosis and, consequently, treatment planning. This is an important consideration, since operator variations exceeding the SD normally associated with a given linear or angular measurement may reveal lack of knowledge and/or experience. We would argue that errors greater than 2 mm reflect the observer's lack of knowledge and/or experience. Based on this criterion, the reproducibility of the identification of certain landmarks is quite low, and the reliability of cephalometric analysis should be questioned. Depending on the observer and on the error, it is possible that different results could have no impact on diagnosis and treatment planning. The existing literature suggests that lateral cephalometric radiographs have been used before treatment without adequate scientific evidence of their utility. Limited evidence is present regarding the usefulness of this radiographic technique in orthodontics.²²

This study evaluated intraobserver and interobserver variability in cephalometric landmark identification carried out by two groups of dental specialists (orthodontists and radiologists). No such comparisons have been made in previous studies.

Many variables contribute to the final diagnosis and treatment plan in orthodontics, such as face-bow recordings, clinical examinations, and intraoral and extraoral photographs. Therefore, it is difficult to predict if a single

error in landmark identification would have an impact on clinical practice. Errors in both dental casts and cephalometric analysis may lead to erroneous decisions about teeth extraction.¹² Each component of diagnosis and treatment planning should be performed with maximum precision.

In conclusion, we established that some landmarks were less reproducible than others in the horizontal and/or vertical axes. The most consistent landmark identified in both groups was LIB, while the least reliable points were Co, Gn, Or, and ANS. Our results indicated a lower degree of reproducibility in the identification of cephalometric landmarks among orthodontists. Further studies focusing on the impact of erroneous cephalometric analyses in a larger sample and in borderline cases (between orthodontics and orthognathic surgery) may be needed to determine the real-world clinical impact of such variation.

References

1. Broadbent BH. A new x-ray technique and its application to orthodontia. *Angle Orthod* 1931; 1: 45-66.
2. Baumrind S, Frantz RC. The reliability of head film measurements. 1. Landmark identification. *Am J Orthod* 1971; 60: 111-27.
3. Chen YJ, Chen SK, Huang HW, Yao CC, Chang HF. Reliability of landmark identification in cephalometric radiography acquired by a storage phosphor imaging system. *Dentomaxillofac Radiol* 2004; 33: 301-6.
4. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. *Eur J Orthod* 2001; 23: 569-78.
5. Miloro M, Borba AM, Ribeiro-Junior O, Naclério-Homem MG, Jungner M. Is there consistency in cephalometric landmark identification amongst oral and maxillofacial surgeons? *Int J Oral Maxillofac Surg* 2014; 43: 445-53.
6. Tng TT, Chan TC, Hägg U, Cooke MS. Validity of cephalometric landmarks. An experimental study on human skulls. *Eur J Orthod* 1994; 16: 110-20.
7. Gravely JF, Benzie PM. The clinical significance of tracing error in cephalometry. *Br J Orthod* 1974; 1: 95-101.
8. Kvam E, Krogstad O. Variability in tracings of lateral head plates for diagnostic orthodontic purposes. A methodologic study. *Acta Odontol Scand* 1969; 27: 359-69.
9. Lau PY, Cooke MS, Hägg U. Effect of training and experience on cephalometric measurement errors on surgical patients. *Int J Adult Orthodon Orthognath Surg* 1997; 12: 204-13.
10. da Silveira HL, Silveira HE. Reproducibility of cephalometric measurements made by three radiology clinics. *Angle Orthod* 2006; 76: 394-9.
11. Proffit WR, Fields HW, Sarver DM. Contemporary orthodontics. 4th ed. St. Louis: Mosby Elsevier; 2006.
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-8.
13. Ricketts RM. Analysis - the Interim. *Angle Orthod* 1970; 40: 129-37.
14. McNamara JA Jr. A method of cephalometric evaluation. *Am J Orthod* 1984; 86: 449-69.
15. McClure SR, Sadowsky PL, Ferreira A, Jacobson A. Reliability of digital versus conventional cephalometric radiology: a comparative evaluation of landmark identification error. *Semin Orthod* 2005; 11: 98-110.
16. Shaheed S, Iftikhar A, Rasool G, Bashir U. Accuracy of linear cephalometric measurements with scanned lateral cephalograms. *Pak Oral Dental J* 2011; 31: 68-72.
17. Murali RV, Sukumar MR, Tajir TF, Rajalingam S. Comparative study of manual cephalometric tracing and computerized cephalometric tracing in digital lateral cephalogram for accuracy and reliability of landmarks. *Indian J Multidiscip Dent* 2011; 1: 126-34.
18. Durão AR, Pittayapat P, Rockenbach MI, Olszewski R, Ng S, Ferreira AP, et al. Validity of 2D lateral cephalometry in orthodontics: a systematic review. *Prog Orthod* 2013; 14: 31.
19. Houston WJ, Maher RE, McElroy D, Sherriff M. Sources of error in measurements from cephalometric radiographs. *Eur J Orthod* 1986; 8: 149-51.
20. Chien PC, Parks ET, Eraso F, Hartsfield JK, Roberts WE, Ofner S. Comparison of reliability in anatomical landmark identification using two-dimensional digital cephalometrics and three-dimensional cone beam computed tomography in vivo. *Dentomaxillofac Radiol* 2009; 38: 262-73.
21. Medelnic J, Hertrich K, Steinhäuser-Andresen S, Hirschfelder U, Hofmann E. Accuracy of anatomical landmark identification using different CBCT- and MSCT-based 3D images: an in vitro study. *J Orofac Orthop* 2011; 72: 261-78.
22. Lagravère MO, Low C, Flores-Mir C, Chung R, Carey JP, Heo G, et al. Intraexaminer and interexaminer reliabilities of landmark identification on digitized lateral cephalograms and formatted 3-dimensional cone-beam computerized tomography images. *Am J Orthod Dentofacial Orthop* 2010; 137: 598-604.