Original Article

Assessment of Reliability of Three Different Computer-Assisted Analysis Programs

Ebubekir Toy, DDS, PhD;[1,*] Sıddık Malkoç, DDS, PhD;[2] Sedat Altındiş, DDS;[3] Sertaç Aksakallı, DDS, PhD[4]

ABSTRACT

Objective: The aim of this study was to assess the reliability of cephalometric analysis using 3 different digital analysis programs.

Methods: For this study, a dry human skull with the entire mandible, maxilla, and teeth was used. Fifteen lateral and 15 posteroanterior cephalometric digital images were taken by rotating the skull from 0° to ±14° at 2° intervals to obtain different images. Two researchers located the landmarks on the digital images independently using 3 computer-assisted analysis programs, Dolphin Image Software (Dolphin Imaging and Management Solutions), Quick Ceph Image (Quick Ceph Systems Inc), and Vistadent OC (GAC Int Inc). Following the first measurements (T1), all landmarks were relocated within a 2-week interval (T2) by each examiner. A paired and the independent Student *t* tests were used for intraexaminer and interexaminer measurements, and Pearson correlations were obtained. Intraclass correlation coefficients (ICC) were calculated to determine intraexaminer repeatability.

Results: For the repeated measurements, mean differences were statistically insignificant, and high correlations for the repeated measurements were found, and the intraexaminer correlations were significant for each examiner ($p<.001$). When the interexaminer correlations of 3 analyses were compared, interexaminer correlations showed high consistency and the lowest Pearson *r* value was the same angular measurement (S ant-n-ss) ($p<.001$). ICC values demonstrated high intraexaminer repeatability. The highest value of ICC was the mandibular body length (go-me) for both examiners ($p<.001$).

Conclusion: The 3 tested analysis programs may be accepted as reliable for clinical use. (*Turkish J Orthod* 2013;26:134–142)

KEY WORDS: Computer-assisted analysis, Digital images, Reliability, Repeatability

INTRODUCTION

In orthodontics, cephalometric radiographs are used for the interpretation of hard and soft tissue changes due to growth or orthodontic treatment. Cephalometric analysis is an indispensable method for establishing orthodontic diagnoses and treatment planning. These tools are also used for prediction of growth and control of treatment results.[1] However, cephalometric measurements can be misinterpreted due to the film magnifications and distortions. Errors of landmark identification, inaccurate tracings, and miscalculations are other main reasons behind misdiagnoses.[2]

Traditional cephalometric radiographs were traced and analyzed manually until recently. Rapid development of computer technology has allowed clinicians to trace the cephalograms digitally, which has provided a wide range of practices with cephalometry since the early 1980s.[3–6] In addition, several software programs have been introduced to perform computer-assisted cephalometric analysis. These programs benefit image storage, allowing for more accurate assessment at poorly defined areas, rapid superimposition of serial radiographs, and data sharing with other clinicians. By digitizing the landmarks, computer-aided cephalometric analysis programs have made it possible to perform and reproduce complete analyses and high-speed data processing and interpretation. This also means that

[1]İnönü University, Faculty of Dentistry, Department of Orthodontics, Malatya, Turkey
[2]İnönü University, Faculty of Dentistry, Department of Orthodontics, Malatya, Turkey
[3]İnönü University, Faculty of Dentistry, Department of Orthodontics, Malatya, Turkey
[4]Bezmialem Vakıf University, Faculty of Dentistry, Department of Orthodontics, Istanbul, Turkey

Corresponding author: Dr Ebubekir Toy, İnönü Universitesi Dişhekimliği Fakültesi Ortodonti AD 44280 Malatya Turkey. Tel: +90 422-341 01 06 / 6154 Fax: +90 442-341 1107 E-mail: ebubekirtoy@gmail.com

To cite this article: Toy E, Malkoç S, Altındiş S, Aksakallı S. Assessment of reliability of three different computer-assisted analysis programs. *Turkish J Orthod* 2013;26:134–142 (DOI: http://dx.doi.org/10.13076/TJO-D-13-00013)
Date Submitted: February 2013. Date Accepted: April 2013.

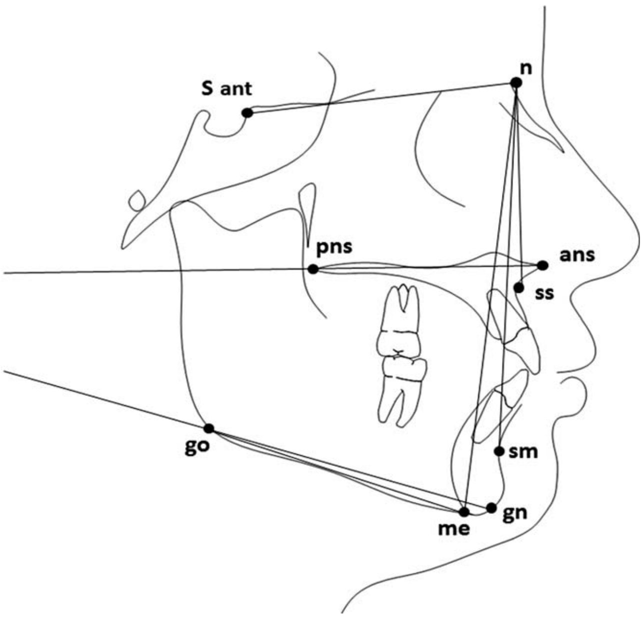

Figure 1. Cephalometric landmarks and measurements of lateral cephalograms used in this study. S ant: Anterior wall of sella turcica; n: nasion; ans: anterior nasal spine; pns: posterior nasal spine; ss: subspinale; sm: supramentale; gn: gnathion; me: menton; go: gonion. Angular measurements: S ant-n-ss (°); S ant-n-sm (°); go-gn/s ant–n (°); ans-pns/go-gn (°). Linear measurements: S ant-n (mm); go-me (mm); ans-me (mm); n-me (mm).

computerized measurements provide the clinician with a significant advantage in terms of time.[7] In digital radiographs, clinicians can also modify the brightness of an image so that anatomic landmarks can be effectively determined and more accurate results obtained.[8] However, operator errors continue in landmark location so that digitization of the image carries some risks.

In orthodontic literature, reliability of the lateral cephalometric analysis has been examined in many reports.[9–12] Moreover, the projection errors caused by head rotation on different axes have been investigated on both lateral[10–12] and posteroanterior (PA) cephalometric radiographs.[13] A large number of investigators have studied the comparison between manually and computer analyzed techniques for speed, repeatability, and reproducibility,[2,7,14] as well as the comparison between 2-dimensional radiography and 3-dimensional computed tomography.[15,16] However, there are few studies in the literature regarding the comparison among different computer-assisted cephalometric programs in terms of accuracy.[17,18] Therefore, the aim of this study was to assess the reliability of cephalometric analysis using 3 different digital analysis programs.

MATERIALS AND METHODS

For this study, a dry human skull with the entire mandible, maxilla, and teeth was used. The skull had no evident asymmetry and was in good condition. Nine anatomic landmarks were determined on the skull for lateral cephalometric digital radiographs and 6 for PA digital radiographs. Reference steel balls 1 mm in diameter were glued onto the landmarks. The Frankfort horizontal plane of the skull was placed parallel to the floor and securely fixed with an ear rod, headrest, and rubber pads. We aimed to obtain different images by rotating the skull from 0° to ±14° at 2° intervals. The radiologic images of the craniofacial structures changed at times, and some of them elongated or shortened after rotation. Therefore, we had 15 different lateral and 15 different posteroanterior cephalometric digital images to investigate. The vertical axis was used as the rotational axis to

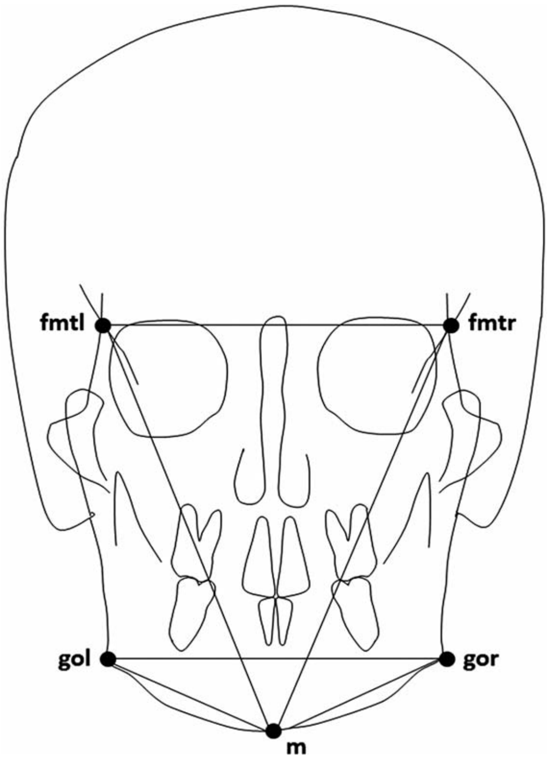

Figure 2. The landmarks and measurements of posteroanterior cephalograms used in this study. fmtl: left front malar temporale; fmtr: right front malar temporale; gol: left gonion; m: mandibular midpoint; gor: right gonion. Angular measurements: gol-m-gor (°); fmtl-m-fmtr (°). Linear measurements: fmtl-fmtr (mm); gol-gor (mm); gol-m (mm); gor-m (mm).

Table 1. Error study for the repeated measurements (T1 and T2) and intraexaminer correlations of each examiner[a]

	Examiner 1														
	Dolphin					Vistadent					Quick Ceph				
	Mean Diff			Correlation		Mean Diff			Correlation		Mean Diff			Correlation	
Variables	T1-T2	SD	*p*	*r*	*P**	T1-T2	SD	*p*	*r*	*P**	T1-T2	SD	*p*	*r*	*P**
Lateral cephalogram															
S ant-n-ss, degrees	0.18	0.45	ns	0.94	***	0.02	0.09	ns	0.98	***	−0.01	0.06	ns	0.98	***
S ant-n-sm, degrees	0.11	0.39	ns	0.83	***	0.00	0.07	ns	0.99	***	0.01	0.06	ns	0.99	***
go-gn/s ant–n, degrees	−0.03	0.34	ns	0.99	***	−0.02	0.14	ns	0.99	***	−0.06	0.12	ns	0.99	***
ans-pns/go-gn, degrees	0.00	0.17	ns	0.99	***	0.00	0.07	ns	0.99	***	−0.03	0.08	ns	0.99	***
S ant-n, mm	−0.10	0.45	ns	0.96	***	0.02	0.11	ns	0.99	***	−0.04	0.15	ns	0.99	***
go-me, mm	0.02	0.15	ns	1.00	***	0.06	0.28	ns	0.99	***	0.01	0.09	ns	1.00	***
ans-me, mm	0.11	0.21	ns	0.94	***	0.08	0.26	ns	0.97	***	0.02	0.08	ns	0.99	***
n-me, mm	−0.03	0.17	ns	0.98	***	−0.03	0.10	ns	0.99	***	−0.02	0.11	ns	0.99	***
PA cephalogram															
fmtl-fmtr, mm	−0.05	0.60	ns	0.97	***	−0.01	0.39	ns	0.94	***	0.05	0.28	ns	0.98	***
gol-gor, mm	−0.02	0.64	ns	0.94	***	0.04	0.20	ns	0.99	***	0.01	0.35	ns	0.99	***
gol-m-gor, degrees	−0.18	0.43	ns	0.96	***	0.02	0.10	ns	1.00	***	−0.10	0.26	ns	0.98	***
fmtl-m-fmtr, degrees	−0.17	0.52	ns	0.94	***	−0.01	0.15	ns	0.98	***	0.02	0.07	ns	0.99	***
gol-m, mm	0.02	0.53	ns	0.99	***	0.02	0.19	ns	1.00	***	0.02	0.07	ns	1.00	***
gor-m, mm	−0.41	0.89	ns	0.99	***	−0.18	0.42	ns	0.99	***	0.02	0.38	ns	0.99	***

[a] Mean diff indicates mean difference; *p*, paired *t* test significance; ns, not significant; *r*, Pearson correlation; *P**, Pearson correlation significance; PA, posteroanterior. Note that all *p* values are not significant and all *P** are ***.
* $p < .05$; ** $p < .01$; *** $p < .001$.

expose as applied by Yoon *et al.*[10] and Malkoc *et al.*[13]

All digital images were obtained using a Planmeca ProMax x-ray unit (Planmeca Oy, Helsinki, Finland). We imported all images into the tested analysis programs. There was no need to adjust images because all digital images had the same magnification. Two researchers located the landmarks on the cephalometric digital images independently using 3 computer-assisted analysis programs, Dolphin Image Software 10.5 (Dolphin Imaging and Management Solutions, Los Angeles, CA, USA), Quick Ceph Image (Quick Ceph Systems Inc, San Diego, CA, USA), and Vistadent OC version 4.2.30 (GAC Int Inc, Bohemia, New York, NY, USA). Four angular and 4 linear parameters for lateral images and 2 angular and 4 linear parameters for PA images were measured on each radiograph (Figures 1 and 2). Following the first measurements (T1), all landmarks were relocated within a 2-week interval (T2) by each examiner to determine intraexaminer repeatability and using different sequences for each evaluation. While the landmarks were marked on digital images using a mouse-driven cursor, image improvements, including embossing, brightness, contrast, and magnification, were applied when necessary in order to identify particular landmarks as clearly as possible. All tested calculations were generated automatically by the software.

Statistical Analysis

All statistical analyses were performed using the Statistical Package for Social Sciences for Windows, version 20.0 (SPSS Inc, Chicago, IL, USA). Descriptive statistical data including means and standard deviations were calculated for all numerical data. Normal distribution was confirmed by a Kolmogorov Smirnov test. A paired Student *t* test was used to compare the mean differences of the same examiner between the first (T1) and second

Table 1. Extended.

Examiner 2														
Dolphin					Vistadent					Quick Ceph				
Mean Diff			Correlation		Mean Diff			Correlation		Mean Diff			Correlation	
T1-T2	SD	*p*	*r*	*P**	T1-T2	SD	*p*	*r*	*P**	T1-T2	SD	*p*	*r*	*P**
0.06	0.16	ns	0.96	***	0.02	0.11	ns	0.92	***	−0.01	0.13	ns	0.93	***
0.04	0.13	ns	0.96	***	0.05	0.12	ns	0.97	***	0.00	0.13	ns	0.97	***
0.02	0.10	ns	0.99	***	0.01	0.08	ns	0.99	***	−0.05	0.14	ns	0.99	***
−0.05	0.16	ns	0.99	***	−0.06	0.11	ns	0.99	***	−0.01	0.07	ns	1.00	***
0.02	0.07	ns	0.99	***	−0.02	0.07	ns	0.99	***	−0.01	0.06	ns	1.00	***
0.02	0.06	ns	1.00	***	0.11	0.20	ns	1.00	***	0.02	0.31	ns	0.99	***
0.06	0.17	ns	0.97	***	0.06	0.17	ns	0.98	***	−0.04	0.10	ns	0.99	***
0.06	0.13	ns	0.99	***	0.02	0.10	ns	0.99	***	0.03	0.09	ns	0.99	***
−0.02	0.18	ns	0.99	***	0.04	0.16	ns	0.99	***	0.08	0.27	ns	0.98	***
0.01	0.19	ns	0.99	***	0.00	0.20	ns	0.99	***	−0.03	0.10	ns	1.00	***
0.01	0.12	ns	0.99	***	0.04	0.15	ns	0.99	***	0.06	0.75	ns	0.83	***
−0.06	0.24	ns	0.98	***	−0.06	0.22	ns	0.96	***	−0.02	0.14	ns	0.98	***
−0.07	0.25	ns	1.00	***	−0.03	0.12	ns	1.00	***	−0.04	0.15	ns	1.00	***
0.02	0.31	ns	1.00	***	−0.07	0.29	ns	1.00	***	0.07	0.22	ns	1.00	***

measurements (T2). The Student *t* test for independent samples was used to compare the mean values of interexaminer measurements, and Pearson correlation coefficients (*r* value) were obtained. Intraclass correlation coefficients (ICC) were calculated to determine intraexaminer repeatability. A reliability coefficient is generally regarded as acceptable when greater than 0.7. Statistical significance was set at $p < .05$.

RESULTS

For the error estimation, we firstly analyzed intraexaminer correlations. When the mean differences of the same examiner between the first (T1) and second measurements (T2) were compared, there were no statistically significant differences for the 3 software programs. High correlations for the repeated measurements were found (range 0.83–1.00 for both examiners), and the intraexaminer correlations were significant for each examiner in terms of mean differences obtained from outcomes of the 3 computer analysis systems (range 0.83–1.00 for both examiners) ($p<.001$) (Table 1).

The mean values of interexaminer measurements and the correlations of 3 analysis programs are shown in Table 2. The lowest Pearson *r* value for Dolphin was 0.836, for Quick Ceph 0.760, and for Vistadent 0.650, which was the same angular measurement (S ant-n-ss) (Figure 1) for all tested computer programs ($p<.001$). The highest Pearson *r* value for Dolphin was 1.000 in lateral cephalograms (go-me distance), for Vistadent 1.000 in PA radiographs (gol-m distance), and for Quick Ceph 0.999 in lateral films (go-me distance) and also in PA radiographs (gol-gor distance) (Figures 1 and 2).

To estimate the intraexaminer repeatability, mean values of 3 software systems were compared and ICC was calculated for each examiner (Table 3). The lowest and highest correlation coefficients were 0.76 and 1.00 for the first investigator and 0.71 and 1.00 for the second one. The highest value of ICC was mandibular body length (go-me) in the lateral cephalograms for both examiners ($p<.001$).

Table 2. The mean values of interexaminer measurements and the correlations of three analysis programs[a]

	Dolphin							Vistadent	
	Examiner 1		Examiner 2			Correlation		Examiner 1	
Variables	**Mean**	**SD**	**Mean**	**SD**	***p***	***r***	***P****	**Mean**	**SD**
Lateral cephalogram									
S ant-n-ss, degrees	84.41	0.41	84.42	0.41	ns	0.836	***	84.30	0.41
S ant-n-sm, degrees	88.80	0.54	88.74	0.52	ns	0.903	***	88.68	0.63
go-gn/s ant–n, degrees	20.29	2.41	20.30	2.52	ns	0.992	***	20.59	2.50
ans-pns/go-gn, degrees	17.42	2.01	17.38	2.13	ns	0.984	***	17.76	1.92
S ant-n, mm	65.45	1.53	65.44	1.56	ns	0.990	***	66.22	1.62
go-me, mm	78.31	7.52	78.33	7.55	ns	1.000	***	79.06	7.44
ans-me, mm	64.41	0.68	64.42	0.81	ns	0.944	***	65.38	1.18
n-me, mm	114.84	1.17	114.86	1.24	ns	0.988	***	116.32	1.47
PA cephalogram									
fmtl-fmtr, mm	95.26	2.57	95.20	2.66	ns	0.995	***	93.29	1.16
gol-gor, mm	116.06	1.95	116.02	1.87	ns	0.991	***	112.67	1.90
gol-m-gor, degrees	64.29	1.50	64.32	1.47	ns	0.992	***	60.55	15.05
fmtl-m-fmtr, degrees	49.42	1.59	49.53	1.53	ns	0.977	***	49.93	0.74
gol-m, mm	62.21	12.43	61.82	12.40	ns	0.988	***	60.60	11.83
gor-m, mm	58.41	13.31	58.68	13.19	ns	0.991	***	57.08	12.85

[a] *p* independent *t* test significance; ns indicates not significant; *r*, Pearson correlation; *P**, Pearson correlation significance; PA, posteroanterior.
* $p < .05$; ** $p < .01$; *** $p < .001$. Note that all *p* values are not significant and all *P** are ***.

DISCUSSION

Measurement and tracing errors in cephalometry may depend on 3 main factors: (1) improper landmark identification, in other words, determination of the wrong anatomic or cephalometric points causing miscalculations; (2) incorrect positioning of the head leading to projection errors; and (3) tracing and measuring errors including mechanical deficiency influencing measurement accuracy.[11,12]

The most common technique for landmark identification is the manual technique. An acetate sheet is placed over the cephalometric radiograph, and landmarks are marked.[7] Measurements of the distances and angles between cephalometric landmarks are recorded with a ruler and protractor. Another technique is computer aided; the landmarks are located manually and the computer system completes the analysis. This study examined whether the orthodontic software programs, Dolphin, Quick Ceph, and Vistadent, could analyze digital cephalograms precisely enough. Therefore, we planned to minimize the operator errors of landmark identification by gluing the steel balls to the skull.

The most prominent cause of tracing errors is ambiguity in landmark identification, which requires skills greatly dependent on the examiner's experience.[19] Gliddon *et al.*[20] found significant differences in location of cephalometric landmarks between trained and untrained operators. It is obvious that sufficient knowledge of the medium in digital cephalometric systems reduces errors and improves the reliability of landmark identification.[2] Therefore, we planned to ease landmark identification by attaching marker steel balls as markers on selected points on the dry skull. Two digitizing examiners were then instructed to use the center of the images of the steel balls. Two examiners in our study also had at least 3 years of orthodontic experience and training. On the other hand, reproducible head position between the X-ray source and film is required for standardization.[13] In this study, cephalometric digital images were taken with substantially stable positioning of the skull; thus, projection errors

Table 2. Extended.

Vistadent					Quick Ceph						
Examiner 2			Correlation		Examiner 1		Examiner 2			Correlation	
Mean	SD	*p*	*r*	*P**	Mean	SD	Mean	SD	*p*	*r*	*P**
84.34	0.26	ns	0.650	***	85.70	0.34	85.68	0.35	ns	0.760	***
88.66	0.56	ns	0.779	***	91.40	0.53	91.38	0.53	ns	0.920	***
20.33	2.31	ns	0.988	***	20.02	2.36	20.11	2.39	ns	0.995	***
17.54	2.06	ns	0.971	***	17.16	2.14	16.04	4.16	ns	0.989	***
66.20	1.61	ns	0.946	***	64.41	1.49	60.44	16.01	ns	0.996	***
79.14	7.45	ns	0.999	***	76.98	7.39	77.01	7.45	ns	0.999	***
65.31	1.00	ns	0.835	***	63.08	0.75	63.20	0.77	ns	0.885	***
116.27	1.70	ns	0.966	***	112.46	1.22	112.57	1.35	ns	0.962	***
93.11	1.57	ns	0.947	***	95.07	1.53	94.92	1.50	ns	0.992	***
112.80	1.68	ns	0.963	***	113.40	5.12	113.50	5.13	ns	0.999	***
64.38	1.37	ns	0.942	***	64.32	1.47	63.67	1.32	ns	0.899	***
49.92	0.88	ns	0.896	***	50.22	0.68	50.14	0.69	ns	0.977	***
60.64	11.73	ns	1.000	***	62.48	11.26	62.32	11.49	ns	0.998	***
56.80	12.34	ns	0.985	***	57.39	11.89	56.76	11.46	ns	0.976	***

were minimized. Furthermore, we aimed to evaluate tracing and measuring errors, including mechanical deficiencies of different image analysis programs, with the help of a computer.

Several important factors should be taken into account in landmark selection for reliability studies. Certain cephalometric landmarks are more reliable in either the horizontal or vertical plane. For example, rotation on the vertical axis affects the horizontal measurements, not the vertical measurements.[13] Regarding interobserver errors, statistically significant differences were found at Po, Ar, ANS, and UM (mesiobuccal cusp of upper first molar) points.[2] In these 2 studies,[2,13] the authors supported that the principal factor appeared to be the features of the landmark itself. Malkoc *et al.*[13] evaluated linear and angular measurement changes due to head rotations. In their study, a researcher manually traced and measured the radiographs, and they concluded that more reliable estimations were vertical linear values for lateral films and angular values for PA radiographs in order to minimize the projection errors. In the present study, we also measured both angular and linear values on the digital images, investigating whether or not there were any differences among the measurements of 3 different software programs. Thus, we thought that the selected points and measurements would be sufficient for the reliability testing.

Two types of cephalometric analysis software programs are commercially available: completely automated programs without any assistance from the operator, and semiautomatic programs with hand-operated landmark determination.[21,22] In the latter technique, after manual location of the landmarks, the computer system performs the desired cephalometric analysis. Errors resulting from drawing and measuring with a ruler and a protractor can be eliminated by these computer programs.[23] Performing analysis with the help of a computer was reported not to cause more measurement error than manual marking and drawing, provided that cephalometric points are manually marked.[24] Although manually identifying landmarks on screen-displayed digital radiographs for cephalometric measurement may still be a better method, Chen *et al.*[2] indicated

Table 3. Estimation of intraexaminer repeatability by intraclass correlation coefficient (ICC)[a]

	Examiner 1								Examiner 2							
	Dolphin		Vistadent		Quick Ceph				Dolphin		Vistadent		Quick Ceph			
Variables	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**ICC**	***P+***	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**ICC**	***P+***
Lateral cephalogram																
S ant-n-ss, degrees	84.40	0.41	84.30	0.41	84.19	0.66	0.76	*	84.42	0.41	84.34	0.26	84.41	0.70	0.71	**
S ant-n-sm, degrees	88.80	0.54	88.68	0.63	87.93	0.75	0.76	*	88.74	0.52	88.66	0.56	87.83	0.60	0.73	**
go-gn/s ant–n, degrees	20.29	2.41	20.59	2.50	20.02	2.36	0.99	***	20.30	2.52	20.33	2.31	20.11	2.39	0.99	***
ans-pns/go-gn, degrees	17.42	2.01	17.76	1.92	17.12	2.14	0.98	***	17.38	2.13	17.54	2.06	17.24	2.14	0.99	***
S ant-n, mm	65.45	1.53	66.22	1.62	64.41	1.49	0.98	***	65.44	1.56	66.20	1.61	64.44	1.54	0.99	***
go-me, mm	78.31	7.52	79.06	7.44	76.98	7.39	1.00	***	78.33	7.55	79.14	7.45	77.01	7.45	1.00	***
ans-me, mm	64.41	0.68	65.38	1.18	63.08	0.75	0.86	***	64.42	0.81	65.31	1.00	63.20	0.77	0.94	***
n-me, mm	114.84	1.17	116.32	1.47	112.46	1.22	0.95	***	114.86	1.24	116.27	1.70	112.54	1.34	0.95	***
PA cephalogram																
fmtl-fmtr, mm	95.26	2.57	93.29	1.16	95.07	1.53	0.84	***	95.20	2.66	93.11	1.57	94.92	1.50	0.86	***
gol-gor, mm	116.06	1.95	112.67	1.90	113.4	5.12	0.80	***	116.02	1.87	112.80	1.68	113.50	5.13	0.77	***
gol-m-gor, degrees	64.29	1.50	64.42	1.14	64.32	1.47	0.97	***	64.32	1.47	64.38	1.37	63.67	1.32	0.96	***
fmtl-m-fmtr, degrees	49.42	1.59	49.93	0.74	50.22	0.68	0.85	***	49.53	1.53	49.92	0.88	50.14	0.69	0.88	***
gol-m, mm	62.21	12.43	60.60	11.83	62.48	11.26	0.98	***	61.82	12.40	60.64	11.73	62.32	11.49	0.98	***
gor-m, mm	58.41	13.31	57.08	12.85	57.39	11.89	0.99	***	58.68	13.19	56.80	12.34	56.76	11.46	0.98	***

[a] *P+* indicates ICC significance.
* $p < .05$; ** $p < .01$; *** $p < .001$.

that there were statistically significant differences in identifying landmarks between conventional cephalometric radiographs and their digitized counterparts. In the present study, cephalometric points were manually marked with a mouse-controlled cursor on digital images on screen, and all calculations were automatically performed by the software.

For all 3 techniques, there were no statistically significant differences between the first and second measurements for both examiners. In computerized programs, all measurements were dependent upon only 1 localization point, such as nasion[7] or menton,[21] and the differences may have been caused by difficulty in identification at this point. Erkan *et al*.[18] compared the manual cephalometric tracing with 4 different computerized tracing programs. They concluded that outcomes obtained with the computerized cephalometric analysis were reliable, and both the traditional method and computer-aided systems have the capacity to execute comparable measurements, as similar to the findings of Çavdar *et al*.[14] However, numerous investigations have demonstrated that incompatibility in landmark identification, such as nasion, menton, porion, gonion, mandibular incisor apex, and posterior nasal spine, is the major source of mistakes in routine cephalometry.[25,26] Similarly, results of a more recent study indicated that Na-Pog measurement showed the lowest interexaminer correlation between both manual and Dolphin tracings and the lowest intraexaminer correlation between these 2 techniques for both examiners.[7] In the present study, the lowest Pearson correlation between interexaminer measurements was found to be the same angular measurement (S ant-n-ss) (Figure 1) for all tested computer programs. Since this value is associated with nasion, the finding supported the previous investigations.[25,26] Furthermore, Uysal *et al*.[7] demonstrated the consistency in the parameters related to gonion and did not agree with the findings of those authors.[26] Our results also revealed that the highest Pearson correlation of interexaminer measurements was related to gonion for 3 tested software programs in agreement with the study by Uysal *et al*.[7]

According to our study, interexaminer correlations among the 3 analyses showed that a high consistency was present among the tested computer-generated measurements. On the contrary, Baumrind and Frantz[11] stated that significant interexaminer errors were present in landmark marking, and they proposed retracing as the solution. However, Arponen *et al*.[19] objected to this suggestion because the effect of intraexaminer error would remain unresolved due to repeated identification of landmarks. Their findings indicated that the intraexaminer errors were similar, not smaller than the interexaminer error.

Uysal *et al*.[7] reported that although interexaminer reproducibility of landmarks was not ideal, measurement errors were quite similar. According to their study, intraexaminer repeatability of landmarks both with the conventional and computerized techniques was highly correlated. Additionally, observer errors for tracing and digitizing were analyzed using both the classic method of tracing by hand and a computerized method.[27] In the results of their study, for both manual and digital measurements, intraobserver and interobserver consistency demonstrated a high correlation. In the present study, ICC values obtained from 3 softwares demonstrated high intraexaminer repeatability in terms of measurement accuracy for both examiners. This finding confirmed earlier studies.[27] We assumed that rotation effects to the landmarks were the same level for both examiners. Therefore, the measurements obtained from the same radiograph would be similar between 2 examiners if the software programs generated accurate measurements. Thus, the present results showed that the high correlations among the 3 measurements might result from the high reliability of the evaluated cephalometric programs.

CONCLUSIONS

Our findings suggest that:

- High correlations for repeated measurements were found and mean differences obtained from outcomes of the 3 computer analysis systems were statistically significant for each examiner.
- Interexaminer correlations of 3 analyses showed high consistency.
- Intraclass correlation coefficients demonstrated high intraexaminer repeatability for both examiners. These results revealed that there was a strong consistency among the 3 tested techniques.
- No statistically significant differences were found among the measurements obtained from the 3 different computer-assisted analysis programs.
- The tested analysis programs may be accepted as reliable for clinical use.

ACKNOWLEDGEMENT

We thank Dr. Cemil Çolak, associate professor in the Department of Biostatistics, İnönü University, for his skillful statistical assistance.

REFERENCES

1. Kolokitha OE. Validity of a manual soft tissue profile prediction method following mandibular setback osteotomy. *Eur J Dent*. 2007;1:202–211.
2. Chen YJ, Chen SK, Chang HF, Chen KC. Comparison of landmark identification in traditional versus computer-aided digital cephalometry. *Angle Orthod*. 2000;70:387–392.
3. Athanasiou AE. Orthodontic cephalometry. *Am J Orthod*. 1997;111:458–459.
4. Gotfredsen E, Kragskov J, Wenzel A. Development of a system for craniofacial analysis from monitor-displayed digital images. *Dentomaxillofac Radiol*. 1999;28:123–126.
5. Malkoc S, Demir A, Uysal T, Canbuldu N. Angular photogrammetric analysis of the soft tissue facial profile of Turkish adults. *Eur J Orthod*. 2009;31:174–179.
6. Gungor AY, Turkkahraman H, Yilmaz HH, Yariktas M. Cephalometric comparison of obstructive sleep apnea patients and healthy controls. *Eur J Dent*. 2013;7:48–54.
7. Uysal T, Baysal A, Yagci A. Evaluation of speed, repeatability, and reproducibility of digital radiography with manual versus computer-assisted cephalometric analyses. *Eur J Orthod*. 2009;31:523–528.
8. Leonardi RM, Giordano D, Maiorana F, Greco M. Accuracy of cephalometric landmarks on monitor-displayed radiographs with and without image emboss enhancement. *Eur J Orthod*. 2010;32:242–247.
9. Na KC, Yoon YJ, Kim KW. A study on the errors in the cephalometric measurements. *Korean J Orthod*. 1998;28:75–83.
10. Yoon YJ, Kim KS, Hwang MS, Kim HJ, Choi EH, Kim KW. Effect of head rotation on lateral cephalometric radiographs. *Angle Orthod*. 2001;71:396–403.
11. Baumrind S, Frantz RC. The reliability of head film measurements: 1. Landmark identification. *Am J Orthod*. 1971;60:111–127.
12. Baumrind S, Frantz RC. The reliability of head film measurements. 2. Conventional angular and linear measures. *Am J Orthod*. 1971;60:505–517.
13. Malkoc S, Sari Z, Usumez S, Koyuturk AE. The effect of head rotation on cephalometric radiographs. *Eur J Orthod*. 2005;27:315–321.
14. Çavdar K, Ciger S, Öz, AZ. A comparison of conventional

and computerized cephalometric methods. *Clinical Dentistry and Research*. 2011;35:33–40.

15. Nalcaci R, Ozturk F, Sokucu O. A comparison of two-dimensional radiography and three-dimensional computed tomography in angular cephalometric measurements. *Dentomaxillofac Radiol*. 2010;39:100–106.
16. Tulunoglu O, Esenlik E, Gulsen A, Tulunoglu I. A comparison of three-dimensional and two-dimensional cephalometric evaluations of children with cleft lip and palate. *Eur J Dent*. 2011;5:451–458.
17. Baskin H, Cisneros GJ. A comparison of two computer cephalometric programs. *J Clin Orthod*. 1997;31:231–233.
18. Erkan M, Gurel HG, Nur M, Demirel B. Reliability of four different computerized cephalometric analysis programs. *Eur J Orthod*. 2012;34:318–321.
19. Arponen H, Elf H, Evälahti M, Waltimo-Sirén J. Reliability of cranial base measurements on lateral skull radiographs. *Orthod Craniofac Res*. 2008;11:201–210.
20. Gliddon MJ, Xia JJ, Gateno J, Wong HT, Lasky RE, *et al*. The accuracy of cephalometric tracing superimposition. *J Oral Maxillofac Surg*. 2006;64:194–202.
21. Sommer T, Ciesielski R, Erbersdobler J, Orthuber W, Fischer-Brandies H. Precision of cephalometric analysis via fully and semiautomatic evaluation of digital lateral cephalographs. *Dentomaxillofac Radiol*. 2009;38:401–406.
22. Vucinic P, Trpovski Z, Scepan I. Automatic landmarking of cephalograms using active appearance models. *Eur J Orthod*. 2010;32:233–241.
23. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*. 2000;118:535–540.
24. Gravely J, Benzies P. The clinical significance of tracing error in cephalometry. *J Orthod*. 1974;1:95–101.
25. Broadbent BH. A new X-ray technique and its application to orthodontia. *Angle Orthod*. 1981;51:93–114.
26. Lagravere MO, Low C, Flores-Mir C, Chung R, Carey JP, *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.
27. Sayinsu K, Isik F, Trakyali G, Arun T. An evaluation of the errors in cephalometric measurements on scanned cephalometric images and conventional tracings. *Eur J Orthod*. 2007;29:105–108.