RESEARCH

Influence of a programme of professional calibration in the variability of landmark identification using cone beam computed tomography-synthesized and conventional radiographic cephalograms

EL Delamare^{*1}, GS Liedke¹, MB Vizzotto¹, HLD da Silveira¹, JLD Ribeiro² and HED Silveira¹

¹Department of Surgery and Orthopedics, Faculty of Dentistry, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil; ²Department of Production and Transport, Faculty of Production Engineering, Federal University of Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

Objectives: The validity of any measurement obtained through a cephalogram largely depends on the reproducibility of the cephalometric landmarks. The purpose of this study is to evaluate the influence of a programme of professional calibration (PPC) on the variability of landmark identification comparing conventional radiographs and cone beam CT (CBCT)-synthesized cephalograms.

Methods: 5 graduate students in oral radiology identified 20 cephalometric landmarks from cephalograms generated from conventional radiographs (RADs), Ray-Sum CBCT-synthesized cephalograms (CBTs) and half-skull CBT (HSTs) from 10 patients. After a period of reinforcement on instruction and calibration with inter- and intraexaminer assessment of reproducibility (intraclass coefficient correlation scores > 0.75) for RADs, CBTs and HSTs obtained from 5 different patients, observers were asked to repeat the analysis of the first 10 patients under the same circumstances. Values in millimetres represented each landmark in a table of Cartesian co-ordinates (x- and y-axes).

Results: ANOVA showed significant reduction in variability levels after the PPC, and there were no differences among the methods of image acquisition. Repeated measures ANOVA indicated that the PPC accounted for reduction in variability levels in 14 of 20 landmarks.

Conclusions: The results suggest that a PPC has more influence than the type of image acquisition on variability of landmark identification based on two-dimensional cephalometric analysis. Cephalograms obtained from RAD or CBCT can be considered equivalent for clinical and experimental applications.

Dentomaxillofacial Radiology (2010) **39**, 414–423. doi: 10.1259/dmfr/31924248

Keywords: cephalometric landmark; error; cone beam computed tomography

Introduction

Cephalometric analysis is widely accepted as a diagnostic standard in orthodontic treatment planning and final assessment of the results. Therefore, obtaining adequate samples for analysis and interpretation of this analysis, as well as a comprehensive understanding of

its limitations, are deemed essential in satisfying its objectives.^{1–4}

Several methods have been developed for obtaining and processing cephalograms, since conventional cephalometric radiographs are the original and most accepted source of analysis.^{1,4–9} Using conventional films, identification of landmarks and tracing cephalograms can be manually performed directly over the radiographs^{3,4} or using a software-assisted approach after digitization of the images.^{1,5–8,10,11} Digital radiographs have been recently introduced to clinical practice and have also supplied a reliable source of

^{*}Correspondence to: Eduardo Luiz Delamare, Department of Surgery and Orthopedics, Ground Floor, Faculty of Dentistry, 2492, Ramiro Barcelos Street, Federal University of Rio Grande do Sul, Porto Alegre 90035-003, Rio Grande do Sul, Brazil; E-mail: edudelamare@gmail.com

Received 6 July 2009; revised 5 October 2009; accepted 19 October 2009

diagnostic information for cephalograms under software-assisted production.^{7,9} Some studies, with promising results, have been published suggesting the use of three-dimensional (3D) and multiplanar reconstructions obtained from CT and cone beam CT (CBCT) as a consistent source of images for cephalometric analysis.^{1,12–18} Acknowledging the fact that the analysis of cephalograms is a human-dependent and time-consuming process, some studies have proposed entirely computer-based systems for automatic identification of cephalometric landmarks with positive perspectives.^{19–21} Nonetheless, contemporary clinical practice still carries out most of its cephalometric analysis using human-based approaches and, owing to its subjective nature, it is generally recognized that error is part of the process that must always be considered.²

Several reports confirm that variability of landmark identification follows characteristic patterns and is directly associated with measurement inaccuracies, which vary independently from the chosen imaging method.^{3,8,22,23} During the assessment of treatment outcomes, most cephalometric evaluations report small changes due to therapy, and it is important that these are properly estimated. Therefore, measurement errors produced as a consequence of inappropriate landmark identification and minor changes induced by treatment can overlap, producing a negative influence in the assessment of treatment results.²

The validity of any measurement obtained by a cephalogram largely depends on the reproducibility of the cephalometric landmarks.^{8,24–26} Factors such as the quality of the radiographs, the conditions under which they are measured, and the care and skill of the operator influence the magnitude of identification error.^{3,22} Once these factors are controlled, a limited number of variables may account for the random errors observed in previous studies. In a study by da Silveira and Silveira,²⁷ 40 lateral cephalometric radiographs were selected and sent at different times to 3 different clinics for cephalometric analyses. Of the 32 factors studied, the reproducibility of results was satisfactory for only 4 factors: position of the maxilla relative to the anterior cranial base, inclination of occlusal plane relative to the anterior cranial base, position of lower incisor relative to the nasion–pogonion line and soft-tissue profile of face ($P < 0.05$).²⁷ Since it has been established that there is a direct association between poor reproducibility scores and landmark identification error, these findings might indicate that variables related to individual perception, such as the lack of professional programmes of continuing education, could have an influence in the magnitude of the pattern of landmark identification error.

A limited number of aspects could be involved in the potential minimization of variability of landmark identification, two of these being the implementation of modern types of image acquisition, such as CBCT, and the influence of professional calibration programmes (PPC). Thus, the purpose of this study is

to evaluate the influence of a PPC on the variability of landmark identification comparing conventional radiographs and CBCT-synthesized cephalograms.

Materials and methods

For this comparative interventional study, the first 15 subjects who had the given indications for a cephalometric radiograph and a CBCT scan were included after presenting for treatment. The study was approved by the Institutional Review Board of the Federal University of Rio Grande do Sul, Brazil.

Cephalometric radiographs were taken according to the following radiographic settings: 70 kV, 10 mA and 0.6 s. All radiographs were scanned into JPEG digital format using a Astra 2400S scanner (UMAX, Dallas, TX) and resolution of 300 dpi.

The CBCT images were obtained from an i-CAT scanner (Imaging Sciences International, Hatfield, PA). The volumetric data were generated in the following mode: 129 kVp, 4.7 mA and 40 s, with a resolution of 0.4 voxels. With i-Cat Vision Software (Imaging Sciences International), a Ray-Sum-synthesized^{10,28} cephalometric image was constructed from the 3D CBCT scan by right lateral radiographic projection of the entire volume. A second synthesized cephalometric image was constructed under the same standards after tracing a division on the midsagittal plane and excluding all the data from the right side of the volume. In order to control tipping of the plane, a grid provided by the i-Cat Vision Software preview screen ensured that, using axial, coronal and sagittal views, the midsagittal plane of the model was oriented vertically, the transporionic line was oriented horizontally and the Frankfort horizontal plane was oriented horizontally.¹¹ Before excluding half of the image, an axial inspection using the multiplanar visualization screen ensured that the following landmarks located along the midsagittal aspect were included in the section: anterior nasal spine (Ans), basion (Ba), nasion (N) and posterior nasal spine (Pns). All images were exported to JPEG digital format (Figure 1).

Using a millimetric scale for digital images, a standardized adjustment of the proportions was performed using Radiocef software (RadioMemory, Belo Horizonte, Brazil) in order to ensure that all distances detected on digital images during landmark identification were the same distances found in the real structures. For conventional cephalograms a ruler positioned along the midsagittal plane provided reference to the software. For CBCT images, the size of the field of view (FOV) was the informed reference.

Five dentists from a graduate programme in oral and maxillofacial radiology with at least 2 years of experience with cephalometric radiology analysed the images. The study was divided into three stages. During a pre-PPC stage, all examiners received 30 images from 10 different patients. The group of images enclosed

Figure 1 Types of images used for landmark identification: (a) conventional radiograph; (b) cone beam CT (CBCT)-synthesized cephalometric image; (c) half-skull CBCT-synthesized cephalometric image

three types of examination from each patient: a conventional radiograph (RAD), a CBCT-synthesized cephalometric image (CBT) and a half-skull CBCT-synthesized cephalometric image (HST). The examiners were asked to identify 20 anatomical landmarks (Table 1) for each examination using the Radiocef software (RadioMemory) on a 19 inch liquid crystal display screen with resolution of 1280×1024 pixels. A table of randomly generated numbers was used for the identification of the images and all observations were performed in random order with the purpose of minimizing learning effects. The same cephalometric software generated a table of co-ordinates in the x- and y-axes from all cephalograms, which was subsequently exported to Microsoft Excel® 2007 for Windows (Redmont, WA). The value of each co-ordinate was represented in millimetres (mm).

The PPC consisted of a lecture on cephalometric analysis presented by a PhD graduate in oral radiology

and specialist in orthodontics and practical discussion sessions on cephalometric landmark identification, followed by an assessment of reproducibility. The lecture consisted of:

Part 1 (4 h)

- Sectional and volumetric anatomy of the skull.
- One-by-one explanation of cephalometric landmarks on dry human skulls.
- Virtual demonstration of proper cephalometric landmark identification using computer-assisted approaches.

Part 2 (4 h)

- Introduction to a virtual learning object for radiographic cephalometry.²⁹
- Introduction to software systems for cephalometric calibration.²⁶

Table 1 Landmarks analysed in the study

A point	A	Deepest point of the curve of the maxilla, between anterior nasal spine and the dental alveolus
Anterior nasal spine	Ans	The tip of the anterior nasal spine.
Articulare	Ar	Posterior border of the neck of the condyle
B point	B	Most posterior point in the concavity along the anterior border of symphysis
Basion	Ba	Most inferior posterior point of the occipital bone at the anterior margin of the occipital foramen
Condylion	Co	Most posterior superior point of the condyle
Gnathion	Gn	Midpoint between the most anterior and inferior point on the bony chin
Gonion	Go	Most convex point where the posterior inferior curve of the ramus meet
Mandibular central incisor root apex	LIR	Root apex of the mandibular central incisor
Mandibular central incisor tip	LIT	Incisal tip of the mandibular central incisor
Menton	Me	Most inferior point of the symphysis
Nasion	N	Intersection of the internasal suture with the nasofrontal suture in midsagittal plane
Orbitale	Or	Lowest point of the roof of orbit; most inferior point of the external border of the orbital cavity
Pogonion	Pog	Most anterior point on the midsagittal symphysis
Porion	Po	Highest point of the ear canal; most superior point of the external auditory meatus
Posterior nasal spine	Pns	Tip of the posterior nasal spine
Pterygomaxillary	Ptm	Most posterior superior point of the pterygomaxillary fissure
Sella	S	Centre of the pituitary fossa of the sphenoid bone
Upper central incisor root apex	UIR	Root apex of the maxillary central incisor
Upper central incisor tip	UIT	Incisal tip of the maxillary central incisor

Previous to the evaluation of reproducibility, all observers had used the Radiocef software for two problem-based practical sessions (of 4 h each). During these sessions, observers liberally debated over highly complex cephalometric cases while directly assisted by the professor and peers. At the end of this stage, all observers had to reach consensus on the identification of landmarks from the presented cases.

The evaluation of intra- and interexaminer reproducibility was conducted, with all examiners receiving 15 examinations from 5 different patients. They were asked to identify the same anatomical landmarks used in the pre-PPC stage under exactly the same circumstances. After a period of 15 days, the examiners were solicited to repeat the analysis of these 15 images. The values from the table of co-ordinates on both occasions were exported to SPSS v.15 for Windows and the intraclass coefficient correlation (ICC) was calculated. In order to proceed to the next stage, ICC values of at least 0.75 for intra- and interexaminer reproducibility for the *x*- and *y*-axes in all landmarks had to be achieved by all examiners (data not shown).

The post-PPC stage consisted of a repetition of the pre-PPC stage. Examiners analysed the same 30 images from the first 10 patients under exactly the same circumstances.

With the intention of analysing the pattern of variability before and after the PPC, an adjustment of the co-ordinates (*Ac*) in the *x*- and *y*-axes had to be performed, given that three types of cephalogram, RAD, CBT and HST, had been generated from the same patient (*Pt*). Thus, to ensure that the error for each one of the observations (*Ob*) performed by all examiners was calculated under the same reference values, the following equation (using patient 1 as an example) was used for all patients:

$$Ac_x = Ob_x + \text{mean}(\Sigma Ob_x)Pt_1^* - \text{mean}(\Sigma Ob_x \text{ RAD or CBT or HST})Pt_1^{**} \quad (1)$$

where $*n = 600$ and $**n = 200$.

This preliminary adjustment ensured that different images obtained from the same patient had their initial co-ordinates (0,0) positioned at the exact same location.

Estimates of error were subsequently calculated for each observation and the corresponding corrected co-ordinates. To establish reliable reference values for this purpose, the median of each landmark was used, given that medians are not significantly influenced by atypical observations. The error from each observation was then calculated after subtracting the value of the adjusted co-ordinate from the median of each landmark, according to the type of cephalogram (RAD, CBT and HST) and patient, as demonstrated by the following equation:

$$\text{Error} = \text{adjusted co-ordinate} - \text{median value of observations of the corresponding landmark} \quad (2)$$

The same procedure was applied to all patients. Analyses performed in this study were based on the

standard deviation of the error values. All errors three times higher than the standard deviation of the landmark were considered atypical and were not included in the final estimations of standard deviation.

Statistical analysis was carried out with SPSS® Version 15.0 for windows and Microsoft Excel® 2007 for Windows. Differences among groups have been assessed by ANOVA and repeated measures ANOVA.

Results

Table 2 shows univariate ANOVA for the variability of landmark identification in the *x*- and *y*-axes. This table has been generated using as response variable the standard deviation from the observed error calculated for each one of the observations performed by all examiners after analysing the examinations from 10 patients. All interactions involving the variable landmark and the highest levels of interaction (involving all variables) were grouped to make up the error term.

It is known that the standard deviation is a variable that does not follow a normal distribution; nevertheless, considering the sample sizes used for the calculation of each standard deviation, it is consistent to affirm that the corresponding distribution is similar to a normal model. Additionally, the *F*-test used in the ANOVA is robust for violations of the hypothesis of normality.³⁰ Therefore, this ANOVA has been performed using the proposed variable (standard deviation) for the generation of the output.

Data analysis of the *F*-test and *P*-values indicate that factors "PPC" and "Landmark" were consistently responsible for differences in the variability of identification, whereas factor "Observer" was determined only for differences in the *x*-axis. The method of acquisition did not account for significant differences, indicating that the variations in landmark identification are similar for conventional radiographs and CBCT images.

Using repeated measures ANOVA, methods of acquisition and the influence of the PPC were analysed to identify differences within landmarks individually. The results are consistent with the previous analysis and are presented in Table 3 along with the estimated marginal averages for each case.

This analysis shows that only a limited number of landmarks did not present lower estimates of variability after the period of PPC. Po (porion), Ar (articulare), Pog (pogonion), Me (menton), Ans, L1T (mandibular central incisor tip) and U1T (upper central incisor tip) did not present significant improvements on both co-ordinates, whereas Go (gonion) and B (B point) did not differ on the *x*-axis, and Or (orbitale), S (sella), Ptm (pterygomaxillary), Pns and L1R (mandibular central incisor root apex) did not show improvement in the *y*-axis.

Considering the differences in types of image acquisition, a variation of 0.14 mm within radiographs and tomography has been the only discrepancy detected

Table 2 ANOVA for variability of landmark identification in *x*- and *y*-co-ordinates

Source	Sum of squares	df	Mean square	F ratio	P-value
<i>x</i>-axis					
<i>Main effects</i>					
A: Method	0.413	2	0.206	2.073	0.127
B: Landmark	80.211	19	4.222	42.393	0.000
C: Observer	2.023	4	0.506	5.080	0.001
D: PPC	2.082	1	2.082	20.902	0.000
<i>Interactions</i>					
AC	1.094	8	0.137	1.373	0.205
AD	0.453	2	0.226	2.274	0.104
CD	0.553	4	0.138	1.388	0.237
ACD	0.182	8	0.023	0.229	0.986
Residual	54.871	551	0.100		
Total (corrected)	141.881	599			
<i>y</i>-axis					
<i>Main effects</i>					
A: Method	0.303	2	0.151	1.369	0.255
B: Landmark	85.414	19	4.495	40.694	0.000
C: Observer	0.908	4	0.227	2.054	0.086
D: PPC	2.141	1	2.141	19.380	0.000
<i>Interactions</i>					
AC	1.615	8	0.202	1.827	0.070
AD	0.168	2	0.084	0.761	0.468
CD	1.030	4	0.257	2.330	0.055
ACD	0.527	8	0.066	0.596	0.782
Residual	60.869	551	0.110		
Total (corrected)	152.973	599			

All *F* ratios are based on the residual mean square error
df, degrees of freedom; PPC, programme of professional calibration

as significant on the *x*-axis of landmark S. All other landmarks did not differ significantly.

The number of atypical observations (error values higher than 3 standard deviations for the landmark) was also reduced by the PPC. Before the PPC, for *x*- and *y*-axes, 40 and 64 atypical observations were detected, respectively; these figures were reduced to 17 and 29, respectively, after the PPC. A χ^2 analysis of these findings and estimations of associated probability was performed and indicated that the PPC was responsible for a significant reduction in these atypical errors.

Since all methods of image acquisition were considered equivalent, the average variability for each landmark after the PPC is presented in Table 4. Given that the distinct values noticed in this analysis suggest different levels of difficulty, an estimate of sample size derived from our group of 150 observations was calculated for each landmark with the aim of identifying the number of repeated observations required for reassessment of variability scores on both co-ordinates, therefore avoiding redundant observations and unnecessary repetitions. Table 4 also presents a scale of difficulty based on the results of sample size estimation. Scores from 0 to 5 derived from the averages of both axes were attributed to each landmark.

Discussion

Several studies have proposed diverse designs to evaluate variability of landmark identification,^{3,4,6,7,23,31} and it is generally accepted that each landmark exhibits a

non-circular envelope pattern of variability and difficulty for the observer.^{8,24,32} Consequently, all traced linear and angular measurements will be subject to expected variations, and the extent will be directly related to the magnitude of the variability of the landmark.^{2,23}

Some theories have been proposed in order to explain the nature of landmark variability, which fall into two groups: structure related and image related. Among the former, some of the most cited are blurred images caused by bilateral structures,^{4,5,8,33} shape and position of the structure,^{3,4,8,32} effect of surrounding structures and curved anatomical boundaries.^{3,4,8,32} Image-related causes of variability, such as spatial and contrast resolution of the display device,^{24,32} background luminance level,^{6,24} luminance range of the display system,²⁴ brightness uniformity,^{6,24} extraneous light in the reading room,²⁴ displayed field size,²⁴ viewing distance,²⁴ image motion and monitor flickering,²⁴ signal-to-noise ratio of the displayed image,^{8,24} magnification functions²⁴ and user interface²⁴ have been cited and considered as important sources of error-inducing factors. Controlling these technical factors had to be achieved in order to judge new imaging methods similar or superior to the original hand-trace-based cephalogram.^{4-7,11,30,32,33} This study has verified the effect of a standardized period of training on structure-related causes of variability and the influence of new types of image acquisition on image-related causes.

da Silveira and Silveira²⁷ suggest that it is advisable for professionals involved in cephalometric analysis to undergo periodic calibration in order to minimize the

Table 3 Variability of landmark identification before and after the programme of professional calibration (PPC) and according to different types of examination (in mm)

			Mean σ	SE	P-value		Mean σ	SE	P-value
N	x	Before	0.47	0.08	0.02	HST	0.44	0.08	0.56
		After	0.30	0.03		RAD	0.32	0.08	
						CBT	0.39	0.08	
	y	Before	0.89	0.14	0.01	HST	0.99	0.23	0.51
		After	0.68	0.13		RAD	0.75	0.23	
						CBT	0.61	0.23	
Or	x	Before	1.52	0.16	0.02	HST	1.39	0.25	0.32
		After	1.26	0.14		RAD	1.13	0.25	
						CBT	1.67	0.25	
	y	Before	1.07	0.14	0.94	HST	0.92	0.17	0.47
		After	1.08	0.11		RAD	1.07	0.17	
						CBT	1.22	0.17	
S	x	Before	0.32	0.02	0.02	HST	0.40	0.03	0.00
		After	0.29	0.02		RAD	0.26	0.03	
						CBT	0.25	0.03	
	y	Before	0.35	0.02	0.34	HST	0.39	0.04	0.26
		After	0.39	0.03		RAD	0.32	0.04	
						CBT	0.41	0.04	
Po	x	Before	1.22	0.11	0.06	HST	1.09	0.12	0.70
		After	0.96	0.09		RAD	1.01	0.12	
						CBT	1.16	0.12	
	y	Before	1.00	0.17	0.06	HST	0.72	0.22	0.65
		After	0.77	0.10		RAD	1.00	0.22	
						CBT	0.95	0.22	
Ba	x	Before	1.13	0.13	0.03	HST	0.88	0.16	0.72
		After	0.85	0.09		RAD	1.02	0.16	
						CBT	1.06	0.16	
	y	Before	1.22	0.12	0.04	HST	1.07	0.16	0.97
		After	0.90	0.12		RAD	1.08	0.16	
						CBT	1.03	0.16	
Co	x	Before	1.55	0.14	0.00	HST	1.18	0.20	0.91
		After	0.95	0.12		RAD	1.29	0.20	
						CBT	1.29	0.20	
	y	Before	1.48	0.16	0.04	HST	1.49	0.23	0.49
		After	1.17	0.14		RAD	1.11	0.23	
						CBT	1.39	0.23	
Ar	x	Before	0.60	0.08	0.09	HST	0.48	0.10	0.50
		After	0.46	0.06		RAD	0.62	0.10	
						CBT	0.48	0.10	
	y	Before	0.62	0.08	0.63	HST	0.53	0.10	0.60
		After	0.58	0.06		RAD	0.61	0.10	
						CBT	0.66	0.10	
Go	x	Before	1.29	0.08	0.12	HST	1.18	0.11	0.80
		After	1.10	0.10		RAD	1.15	0.11	
						CBT	1.25	0.11	
	y	Before	1.30	0.10	0.02	HST	1.02	0.14	0.33
		After	1.06	0.09		RAD	1.32	0.14	
						CBT	1.20	0.14	
Me	x	Before	0.64	0.05	0.11	HST	0.63	0.07	0.89
		After	0.55	0.05		RAD	0.59	0.07	
						CBT	0.58	0.07	
	y	Before	0.21	0.02	0.47	HST	0.21	0.03	0.76
		After	0.23	0.02		RAD	0.24	0.03	
						CBT	0.22	0.03	
Pog	x	Before	0.25	0.03	0.50	HST	0.25	0.03	0.90
		After	0.23	0.02		RAD	0.23	0.03	
						CBT	0.23	0.03	
	y	Before	0.97	0.07	0.45	HST	0.94	0.10	0.42
		After	0.89	0.08		RAD	0.84	0.10	
						CBT	1.02	0.10	
Gn	x	Before	0.30	0.03	0.03	HST	0.26	0.03	0.91
		After	0.24	0.02		RAD	0.28	0.03	
						CBT	0.27	0.03	
	y	Before	0.29	0.07	0.03	HST	0.25	0.03	0.99
		After	0.21	0.02		RAD	0.25	0.03	
						CBT	0.25	0.03	

Table 3 Continued.

			Mean σ	SE	P-value		Mean σ	SE	P-value
B	x	Before	0.30	0.03	0.15	HST	0.29	0.02	0.48
		After	0.24	0.02		RAD	0.26	0.02	
						CBT	0.26	0.02	
	y	Before	1.40	0.11	0.04	HST	1.23	0.16	0.62
		After	1.19	0.10		RAD	1.43	0.16	
						CBT	1.24	0.16	
A	x	Before	1.20	0.08	0.00	HST	0.74	0.14	0.34
		After	0.62	0.12		RAD	0.94	0.14	
						CBT	1.04	0.14	
	y	Before	1.37	0.11	0.02	HST	1.12	0.12	0.35
		After	1.10	0.06		RAD	1.37	0.12	
						CBT	1.20	0.12	
Ans	x	Before	1.28	0.15	0.54	HST	1.17	0.16	0.22
		After	1.38	0.09		RAD	1.26	0.16	
						CBT	1.56	0.16	
	y	Before	0.57	0.08	0.82	HST	0.54	0.12	0.45
		After	0.56	0.07		RAD	0.47	0.12	
						CBT	0.68	0.12	
Pns	x	Before	1.63	0.14	0.02	HST	1.41	0.16	0.83
		After	1.26	0.10		RAD	1.40	0.16	
						CBT	1.52	0.16	
	y	Before	0.52	0.05	0.39	HST	0.50	0.08	0.98
		After	0.46	0.06		RAD	0.49	0.08	
						CBT	0.48	0.08	
Ptm	x	Before	1.25	0.13	0.02	HST	1.26	0.18	0.56
		After	1.02	0.09		RAD	1.17	0.18	
						CBT	0.99	0.18	
	y	Before	1.25	0.17	0.17	HST	1.55	0.24	0.66
		After	1.51	0.17		RAD	1.37	0.24	
						CBT	1.23	0.24	
L1R	x	Before	1.96	0.13	0.00	HST	1.57	0.19	0.09
		After	1.57	0.12		RAD	2.12	0.19	
						CBT	1.60	0.19	
	y	Before	1.75	0.09	0.31	HST	1.64	0.13	0.73
		After	1.63	0.10		RAD	1.78	0.13	
						CBT	1.66	0.13	
L1T	x	Before	0.25	0.02	0.09	HST	0.25	0.03	0.38
		After	0.21	0.02		RAD	0.20	0.03	
						CBT	0.24	0.03	
	y	Before	0.28	0.02	0.56	HST	0.31	0.04	0.75
		After	0.30	0.03		RAD	0.27	0.04	
						CBT	0.29	0.04	
U1T	x	Before	0.30	0.04	0.25	HST	0.31	0.04	0.52
		After	0.26	0.02		RAD	0.24	0.04	
						CBT	0.29	0.04	
	y	Before	0.28	0.04	0.30	HST	0.30	0.04	0.09
		After	0.24	0.02		RAD	0.19	0.04	
						CBT	0.30	0.04	
U1R	x	Before	1.51	0.11	0.00	HST	1.13	0.17	0.25
		After	1.21	0.11		RAD	1.52	0.17	
						CBT	1.45	0.17	
	y	Before	1.67	0.10	0.04	HST	1.44	0.17	0.65
		After	1.37	0.13		RAD	1.64	0.17	
						CBT	1.47	0.17	

CBT, CBCT-synthesized cephalogram; HST, half-skull CBT; RAD, conventional radiograph; SE standard error
Refer to Table 1 for explanation of landmarks

lack of reproducibility. The present study aimed to observe whether PPC would result in lower variability scores; the results (Table 2) were consistent with the suggestions made by the aforementioned authors. The development of this programme took into consideration the fact that, after the initial learning stages of radiographic cephalometry, professionals are hardly ever confronted with a revision of its basic concepts and are very unlikely to be presented with reassessments of

reproducibility. Therefore, the errors detected during clinical practice could be caused by modifications of individual perception acquired through time. Since it has been shown that computer and virtual instruments of learning can be effective in radiographic cephalometry teaching,²⁹ the intention of this study was to create an objective programme based not only on instruction by a professor, but also on problem-based individual and peer-oriented practices.

Table 4 Estimated marginal means of variability for the *x* and *y*-axes (in mm), estimated sample of observations for reassessments of qualified examiners* and difficulty ranking for each landmark

Landmark	<i>x</i>		<i>y</i>		Overall min. n*	Difficulty ranking (0–5)
	Mean σ	n*	Mean σ	n*		
LIR	1.57	28	1.63	31	31	5
UIR	1.21	17	1.37	22	22	3
Or	1.26	19	1.08	14	19	3
Ptm	1.02	13	1.25	18	18	3
Go	1.10	14	1.06	13	14	2
Co	0.95	11	1.17	16	16	2
Ans	1.38	22	0.56	4	22	2
Pns	1.26	19	0.46	3	19	2
A	0.62	5	1.10	14	14	2
Ba	0.85	9	0.90	10	10	2
Po	0.96	11	0.77	8	11	2
B	0.24	2	1.19	17	17	2
Pog	0.23	2	0.89	10	10	1
Ar	0.46	3	0.58	5	5	1
N	0.30	2	0.68	6	6	1
Me	0.55	4	0.23	2	4	1
S	0.29	2	0.39	3	3	0
LIT	0.21	1	0.30	2	2	0
UIT	0.26	2	0.24	2	2	0
Gn	0.24	2	0.21	1	2	0

*Average of *n* observations presenting standard error < 0.3 mm
Refer to Table 1 for explanation of landmarks

Table 3 shows that, of the seven landmarks which did not exhibit significant improvement with the PPC on both co-ordinates, four (with difficulty rankings of 1 or less, Table 4) presented scores of variability lower than 0.65 mm before the PPC, which indicates that these landmarks naturally do not impose high levels of difficulty and, therefore, are not subject to major errors of interpretation. The same logic could be applied to a lack of significant improvement in the *y*-axis of landmark S and the *x*-axis of landmarks B and Pog. In contrast, the lack of reduction in variability levels on landmarks Ans and Po on both axes, LIR, Ptm, Pns, Pog and Or on the *y*-axis and Go on the *x*-axis indicates that the difficulty imposed by these landmarks is mostly related to the complexity in visualizing the structure itself and, therefore, these inflexible score ranges could indicate that the PPC may not be capable of overcoming this condition within these landmarks. For all other landmarks and corresponding co-ordinates, significant improvements were observed, as shown in Table 3.

Since there was strict control of confounding variables, once the three types of examination studied were subjected to the exact same conditions of analysis, it can be suggested that CBT, HST (which were based on the Ray-Sum type of voxel visualization) and RAD are equivalent in terms of cephalometric investigation (Table 2). CBCT provides other opportunities of image rendering, such as maximum intensity projections (MIPs)¹⁰ and 3D, but the effect of these methods on the variability of landmark identification is still unknown.

Although some studies suggested that higher variability scores of bilateral landmarks could have been explained by the fact that the images from these are not as clear as unilateral landmarks,^{4,8,9} the findings from

this study show that cephalograms constructed from half-skull projections did not significantly benefit from the removed superimposition. Although the scores of Po, Ba, Ar and Go have been generally lower for HST, it has been observed that the challenge imposed for the identification of these landmarks could not have been explained by a blurring effect of the overlapping contralateral structure. Position and shape of the structure are most likely responsible for these findings, as for all other remaining landmarks.

As a result of this particular difficulty imposed by each landmark, special orientation towards the identification of the most challenging structures should be observed during undergraduate programmes as well as graduate training and research should be carried out to develop a special approach for landmarks which are inflexible to the effect of further training. Reports show that software-assisted methods of calibration are reliable and well-accepted alternatives in terms of evaluating and reducing errors associated with landmark identification for professionals and students.^{26,29} In order to assist this process, this study divided landmarks in categories according to an ordinal scale of difficulty (Table 4). A scale of difficulty was also proposed in a meta-analysis conducted by Trpkova *et al*,³ who found similar results.

It is well established that the variability of landmark identification is a fundamentally responsible factor for errors in interpretation of cephalometric analysis^{3,24,27,32,33} and it is a limitation which will constantly be detected whenever human-based systems of identification over two-dimensional (2D) cephalograms are reviewed. Therefore, since it is acknowledged that human-based approaches will probably never be capable of exactly reproducing gold standards, agreement

over the minimum acceptable levels of variability could be taken as the references themselves. If this factor is not evaluated and controlled, cephalometric measurements are subject to random variations that could lead to unpredictable results.²⁷ Since the variability of landmarks shows distinct patterns and extents of distribution for each structure, some studies have tried to determine the influence of controlled patterns of variability over cephalometric measurements; it has been observed that particular landmarks might inflict considerable ranges of variation in angles and linear distances, even when highly trained professionals are being assessed.^{2,9} Nevertheless, when it comes to assessments of reproducibility of 2D-based cephalograms, the variability in landmark identification has not always been considered;^{10–12,17,32} however, a number of reports do consider this factor.^{2,15,22,23,25} Neglecting this aspect may generate studies in which the number of repeated observations for each landmark might not be adequate to assess the extent of this limitation (even when reasonable ICC scores are reached) and, consequently, results judged as correct and reproducible are subject to the possibility of being the product of an opportune sequence of random approximates. In contrast, the same logic could be applied to results considered as inadequate, even though they have been generated from measurements derived from landmarks with inflexible wide ranges of variation. Understanding the importance and exact consequence of these levels of variability of cephalometric angles and linear distances is, therefore, essential, since it will lead to better interpretation of results, with estimations of value ranges being oriented by landmark complexity and, consequently, more reliable. Unfortunately, only a few studies have had this as an objective.^{2,23}

In view of this limitation and considering that, in this study, the achieved interexaminer ICC values were above 0.75 for all landmarks, Table 4 shows the expected number of repeated observations indicated for reassessment of variability scores on both coordinates, ensuring standard errors below or equal to 0.3 mm. This study used 150 observations for each landmark (considering scores from all observers); however, it has been noticed that, after a certain number of repetitions, variability scores originating from these standard deviations did not differ significantly. The intention of calculating a sample size derived from our group of observations was to facilitate future studies and training activities that aim to evaluate acceptable variability scores without the need for redundant

observations and unnecessary repetitions. The difficulty ranking used in Table 4 shows that 7 out of 8 coordinates from landmarks scoring “0” were below 0.3 mm. Therefore, a standard error of 0.3 mm was chosen as a cut-off point for this estimation, as this value is close to the maximum error expected to be detected on the least challenging landmarks and, consequently, a possible level of variability intrinsic to limitations of human-based observations.

These findings, therefore, suggest a preliminary orientation for sample size criteria of reproducibility studies based on landmark complexity. Nonetheless, further studies are required in order to assess the combined influence and extent of landmark identification errors on the generation of angular and linear measurements.

This factor should also be considered in the development of software intended for automatic landmark identification, which has been studied by several authors.^{19–21} As happens with human-based approaches, computer-based systems must also be assessed in terms of variability in order to gauge reliable margins of error for each measurement consistent with the expected difficulty imposed by each landmark.

The uncertainty related to identification of landmarks on 2D examinations is one of the factors which have led studies to propose models of cephalometric analysis based on 3D data from CT and CBCT.^{18,34} Although an agreement over the most suitable type of analysis has not yet been reached,^{35,36} promising results concerning the variability of landmark identification and, consequently, angular and linear measurements have been observed.¹³

Despite the method of image acquisition used, the results presented in this study indicate that professional training must be prioritized in order to minimize interpretation errors associated with cephalometric analysis. The influence of periodic professional assessment seems to be more important for improving accuracy than new methods of image acquisition.

In conclusion this study suggests that a PPC can significantly reduce the variability of cephalometric landmark identification. A considerable difference in the extent of variability has been observed within landmarks, and a difficulty ranking derived from these levels has been proposed. Differences detected among cephalograms obtained from conventional radiographs and two types of CBCT reconstruction were not statistically significant; thus, these methods can be considered as equivalent for clinical and experimental applications.

References

1. Greiner M, Greiner A, Hirschfelder U. Variance of landmarks in digital evaluations: comparison between CT-based and conventional digital lateral cephalometric radiographs. *J Orofac Orthop* 2007; **68**: 290–298.
2. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. *Eur J Orthod* 2001; **23**: 569–578.
3. Trpkova B, Major P, Prasad N, Nebbe B. Cephalometric landmarks identification and reproducibility: a meta analysis. *Am J Orthod Dentofacial Orthop* 1997; **112**: 165–170.
4. 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.

5. Turner PJ, Weerakone S. An evaluation of the reproducibility of landmark identification using scanned cephalometric images. *J Orthod* 2001; **28**: 221–229.
6. Schulze RK, Gloede MB, Doll GM. Landmark identification on direct digital versus film-based cephalometric radiographs: a human skull study. *Am J Orthod Dentofacial Orthop* 2002; **122**: 635–642.
7. 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**: 301–306.
8. Lou L, Lagravere MO, Compton S, Major PW, Flores-Mir C. Accuracy of measurements and reliability of landmark identification with computed tomography (CT) techniques in the maxillofacial area: a systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2007; **104**: 402–411.
9. Moshiri M, Scarfe WC, Hilgers ML, Scheetz JP, Silveira AM, Farman AG. Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. *Am J Orthod Dentofacial Orthop* 2007; **132**: 550–560.
10. Cattaneo PM, Bloch CB, Calmar D, Hjortshoj M, Melsen B. Comparison between conventional and cone-beam computed tomography-generated cephalograms. *Am J Orthod Dentofacial Orthop* 2008; **134**: 798–802.
11. Kumar V, Ludlow J, Soares Cevidanes LH, Mol A. In vivo comparison of conventional and cone beam CT synthesized cephalograms. *Angle Orthod* 2008; **78**: 873–879.
12. van Vlijmen OJ, Berge SJ, Swennen GR, Bronkhorst EM, Katsaros C, Kuijpers-Jagtman AM. Comparison of cephalometric radiographs obtained from cone-beam computed tomography scans and conventional radiographs. *J Oral Maxillofac Surg* 2009; **67**: 92–97.
13. de Oliveira AE, Cevidanes LH, Phillips C, Motta A, Burke B, Tyndall D. Observer reliability of three-dimensional cephalometric landmark identification on cone-beam computerized tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2009; **107**: 256–265.
14. Silva MA, Wolf U, Heinicke F, Bumann A, Visser H, Hirsch E. Cone-beam computed tomography for routine orthodontic treatment planning: a radiation dose evaluation. *Am J Orthod Dentofacial Orthop* 2008; **133**: 640 e641–645.
15. Periago DR, Scarfe WC, Moshiri M, Scheetz JP, Silveira AM, Farman AG. Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. *Angle Orthod* 2008; **78**: 387–395.
16. Kragkov J, Bosch C, Gyldensted C, Sindet-Pedersen S. Comparison of the reliability of craniofacial anatomic landmarks based on cephalometric radiographs and three-dimensional CT scans. *Cleft Palate Craniofac J* 1997; **34**: 111–116.
17. Chidiac JJ, Shofer FS, Al-Kutoub A, Laster LL, Ghafari J. Comparison of CT scanograms and cephalometric radiographs in craniofacial imaging. *Orthod Craniofac Res* 2002; **5**: 104–113.
18. Swennen GR, Schutyser F. Three-dimensional cephalometry: spiral multi-slice vs cone-beam computed tomography. *Am J Orthod Dentofacial Orthop* 2006; **130**: 410–416.
19. El-Fegh I, Galhood M, Sid-Ahmed M, Ahmadi M. Automated 2-D cephalometric analysis of X-ray by image registration approach based on least square approximator. *Conf Proc IEEE Eng Med Biol Soc* 2008; **2008**: 3949–3952.
20. Rueda S, Alcaniz M. An approach for the automatic cephalometric landmark detection using mathematical morphology and active appearance models. *Med Image Comput Comput Assist Interv* 2006; **9**: 159–166.
21. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop* 2000; **118**: 535–540.
22. Perillo M, Beideman R, Shofer F, Jacobsson-Hunt U, Higgins-Barber K, Laster L, et al. Effect of landmark identification on cephalometric measurements: guidelines for cephalometric analyses. *Clin Orthod Res* 2000; **3**: 29–36.
23. Chen YJ, Chen SK, Yao JC, Chang HF. The effects of differences in landmark identification on the cephalometric measurements in traditional versus digitized cephalometry. *Angle Orthod* 2004; **74**: 155–161.
24. Yu SH, Nahm DS, Baek SH. Reliability of landmark identification on monitor-displayed lateral cephalometric images. *Am J Orthod Dentofacial Orthop* 2008; **133**: e790–796; discussion e791.
25. Arponen H, Elf H, Evalahti M, Waltimo-Siren J. Reliability of cranial base measurements on lateral skull radiographs. *Orthod Craniofac Res* 2008; **11**: 201–210.
26. Silveira HL, Silveira HE, Dalla-Bona RR, Abdala DD, Bertoldi RF, von Wangenheim A. Software system for calibrating examiners in cephalometric point identification. *Am J Orthod Dentofacial Orthop* 2009; **135**: 400–405.
27. da Silveira HL, Silveira HE. Reproducibility of cephalometric measurements made by three radiology clinics. *Angle Orthod* 2006; **76**: 394–399.
28. Farman AG, Scarfe WC. Development of imaging selection criteria and procedures should precede cephalometric assessment with cone-beam computed tomography. *Am J Orthod Dentofacial Orthop* 2006; **130**: 257–265.
29. Silveira HLD, Gomes MJ, Dalla-bona RR, Silveira HED. Evaluation of the radiographic cephalometry learning process by a learning virtual object. *Am J Orthod Dentofacial Orthop* 2009; **136**: 134–138.
30. Zeller RA YY. 22 Power: establishing the optimum sample size. *Handbook of statistics*. Elsevier, Oxford 2007, pp 656–678.
31. Athanasiou AE, Miethke R, Van Der Meij AJ. Random errors in localization of landmarks in postero-anterior cephalograms. *Br J Orthod* 1999; **26**: 273–284.
32. Sayinsu K, Isik F, Trakyalı 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.
33. Roden-Johnson D, English J, Gallerano R. Comparison of hand-traced and computerized cephalograms: landmark identification, measurement, and superimposition accuracy. *Am J Orthod Dentofacial Orthop* 2008; **133**: 556–564.
34. Moreira CR, Sales MA, Lopes PM, Cavalcanti MG. Assessment of linear and angular measurements on three-dimensional cone-beam computed tomographic images. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2009; **108**: 430–436.
35. Swennen GR, Schutyser F, Barth EL, De Groeve P, De Mey A. A new method of 3-D cephalometry Part I: the anatomic Cartesian 3-D reference system. *J Craniofac Surg* 2006; **17**: 314–325.
36. Lagravere MO, Major PW. Proposed reference point for 3-dimensional cephalometric analysis with cone-beam computerized tomography. *Am J Orthod Dentofacial Orthop* 2005; **128**: 657–660.