Original Research

Smartphone and computer cephalometric analysis: A trueness and precision study

Mónica Amorim[1],* , Fernanda Diamantino[2] , Rui Pereira[1] , Luís Jardim[1]

[1] Departamento de Ortodontia, Faculdade de Medicina Dentária, Universidade de Lisboa, Lisbon, Portugal.
[2] Centro de Estatística e Aplicações (CEAUL), Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal.

A R T I C L E   I N F O

A B S T R A C T

Objectives: Performing digital cephalometric analysis on apps is a convenient feature of handheld devices. The objective of this study was to assess the accuracy of the OneCeph cephalometric tracing app on a smartphone and computer.

Methods: 34 lateral cephalograms were traced in two sessions using three methods: OneCeph on a smartphone, OneCeph on a computer, and NemoCeph on a computer as the reference. For trueness analysis, the measurements were compared between each test method and the reference. For precision analysis, the measurements were compared between sessions.

Results: Regarding trueness analysis, significant differences were found between OneCeph and NemoCeph for ANB (smartphone, -0.3±0.68; computer, -0.3±0.52), OL/SN (smartphone, -1.1±2.68; computer -2.0±2.98), and GoGn/SN (smartphone, -0.5±1.27; computer, -0.8±1.56). OneCeph reproducibility was high for all cephalometric variables except OL/SN, both on the smartphone and the computer (ICC [95%CI]: 0.888 [0.773-0.944] and 0.842 [0.583-0.931], respectively). Repeatability was high for all cephalometric variables except OL/SN and UINA on the smartphone (ICC [95%CI]: 0.889 [0.730-0.950] and 0.831 [0.687-0.912], respectively). Tracing time was significantly higher for both OneCeph methods.

Conclusions: OneCeph demonstrated adequate accuracy and efficiency on both interfaces. Clinical judgment is advised when interpreting the measurement output of handheld devices, which may provide a higher frequency of gross landmark identification errors. (Rev Port Estomatol Med Dent Cir Maxilofac. 2022;63(4):204-212)

* Corresponding author.
   E-mail address: monica.amorim92@yahoo.com (Mónica Amorim).

Análise cefalométrica em smartphone e computador: Um estudo de justeza e precisão

R E S U M O

Palavras-chave:
Exatidão
Apps
Cefalometria
Telemóvel

Objetivos: A realização de análise cefalométrica digital em *apps* pode ser uma funcionalidade conveniente dos dispositivos portáteis. O presente estudo teve como objetivo avaliar a exatidão da aplicação OneCeph em telemóvel e computador.

Métodos: 34 teleradiografias foram traçadas e retraçadas com três métodos: OneCeph em telemóvel, OneCeph em computador e com NemoCeph em computador como referência. Na análise de exatidão as medições foram comparadas entre os métodos. Na análise de precisão as medições foram comparadas entre repetições.

Resultados: Na análise de exatidão foram encontradas diferenças significativas entre os métodos OneCeph e NemoCeph para ANB (-0,3±0,68 em telemóvel; -0,3±0,52 em computador), OL/SN (-1,1±2,68; -2,0±2,98), GoGn/SN (-0,5±1,27 e -0,8±1,56). A reprodutibilidade de OneCeph foi elevada para todas as variáveis exceto OL/SN em telemóvel e computador (ICC [95%CI]: 0.888 [0,773-0,944] e 0.842 [0,583-0,931], respetivamente). A repetibilidade foi elevada para todas as variáveis exceto OL/SN e UINA em telemóvel (0,889 [0,730-0,950]). O tempo de traçado foi significativamente superior nos métodos OneCeph.

Conclusões: OneCeph demonstrou exatidão e eficiência adequadas em ambas as interfaces. Aconselha-se discernimento clínico aquando da interpretação dos resultados obtidos por dispositivos portáteis, que podem apresentar maior frequência de erros de marcação grosseiros. (Rev Port Estomatol Med Dent Cir Maxilofac. 2022;63(4):204-212)

Introduction

Cephalometric analysis has been playing for long a fundamental role in the diagnosis, treatment planning, and evaluation of dental and skeletal changes in orthodontic patients. Since the 1980s, computer-guided cephalometric analysis has been of special interest, reflected by the increasing number and quality of related scientific publications.[1] In addition to its space- and time-saving qualities, this procedure is user-friendly, eliminates the measurement error inherent in an operator using a ruler or protractor, and has demonstrated suitable accuracy.[2-6]

Meanwhile, handheld devices have also shown applicability in the medical field, namely in diagnosis, monitoring, and communication with patients, usually through the use of apps – downloadable small-scale software applications.[7,8] Likewise, apps with applicability in orthodontics are now available[9,10] and may be of special interest to clinicians working in various clinical settings.

OneCeph (NXS, Telangana, India) is a free-of-charge cephalometric analysis app for Android operating systems. This software allows users to perform, save, and export measurements for 15 pre-programmed and commonly used cephalometric analysis sets and, in addition, customize a "favorite" analysis using the available variables.

The term accuracy was defined by ISO 5725-1 as describing the closeness between a test result and the accepted reference value, which involves a combination of random and bias components. The bias component is quantified by the method's trueness – the closeness of agreement between the average value from multiple observations and an accepted reference value. The random component is quantified by the method's precision considering its repeatability (same test conditions) and reproducibility (different test conditions, e.g., different operators or equipment). Published reports on the accuracy of cephalometric tracing apps are scarce and include conditions that vary with the software version, display device, screen size, use of a landmark identification instrument, and reference tracing method.[11-14]

A recent paper[14] published accuracy results for cephalometric analysis with OneCeph, reporting a single-examiner experiment on a Samsung Galaxy S8 smartphone with a 5.8" screen (2960x1440). The lowest repeatability estimates (intra-class correlation coefficients, ICC) were registered for lower incisor protrusion, mandibular plane angle, upper incisor protrusion, and lower incisor inclination (ICC [95% confidence interval]: 0.647 [0.451, 0.784], 0.658 [0.466, 0.791], 0.701 [0.529, 0.818], and 0.867 [0.778, 0.922], respectively). On the other hand, high repeatability (ICC>0.9) was registered for SNA, SNB, ANB, upper incisor inclination, and interincisal angle. The authors also reported high reproducibility for all cephalometric variables when comparing the OneCeph method with the chosen reference standard on a 13.5" (2256x1504) computer screen. In terms of trueness, the authors reported statistically significant differences – although hardly clinically significant – for the mandibular plane angle (1.5° mean increase), upper incisor protrusion (-0.4 mm), lower incisor protrusion (-0.5 mm), and inter-incisal angle (+1.2°).

Using larger display screens or a stylus has been suggested as potentially enhancing app-based cephalometric analysis.[14] Following previously published findings, the main purpose of this study was to assess the trueness and precision of OneCeph on a smartphone screen using a stylus pen and on a computer screen using an optical mouse, having NemoCeph set as the reference standard. Additionally, the method's efficiency was assessed by comparing the tracing time between techniques.

Material and Methods

This study evaluated pretreatment lateral cephalograms performed at (anonymized) using an Orthoralix 9200 2D Digital Pan Ceph (Gendex, Hatfield, EUA) between October 2016 and July 2018. Selection criteria included digital radiographs with a 2429x2121 resolution showing permanent dentition in maximum intercuspation. Images were excluded from the sample if the machine lateral poles did not overlap or any relevant structures had been cropped out. No modifications were made to the images, and a single examiner performed the cephalometric analysis with maximum screen brightness on the laptop and smartphone. The cephalograms were plotted using a single method randomly selected per day, and this process was repeated a week later for the second method and two weeks later for the third method (T1). Thus, 6 weeks were needed to retrace all cephalograms with each method (T2). A total of 18 anatomical landmarks (Figure 1) and 13 cephalometric variables were selected according to Steiner's analysis[15] and tested with each method on a training set of 12 lateral cephalograms. The (anonymized) Ethics Review Board approved this study.

The cephalometric analysis technique performed at the orthodontic graduate clinic was selected as the reference method (NemoCeph software, 2017, Nemotec, Madrid, Spain). This analysis was performed on a laptop computer (Lenovo™ ideapad 320) with a 14" HD (1366x768) display and an optical mouse (Qilive CP-1485). OneCeph (beta 7) was used for cephalometric analysis on a smartphone (Samsung A5, Samsung Telecommunications, Suwon, South Korea) with a 5.2" (1920x1280) display. A capacitive stylus touch pen with a 2-mm diameter tip and a 6.8-mm diameter transparent disk was used for landmark identification. OneCeph was also used on the computer through a screen-sharing software (VysorPro 2.1.7, Google Commerce, Seattle, Washington). In this case, the computer setup was the same as for the computer tracing method.

A literature search was conducted on European cephalometric norms, and the standard deviation values of relevant variables were registered. The minimum sample size (n=34) was determined for a 0.5 effect size, 80% power, and 5% alpha (G*Power 3.1.9.2) to detect a clinically significant difference of 3.50° – half the highest standard deviation value registered for U1/NA.[16] Statistical analysis was performed using SPSS 25 software (IBM, Armonk, NY, USA). Precision was explored by computing point and 95% confidence interval estimates for the inter-session (repeatability) and inter-method (reproducibility) ICCs by a two-way mixed effects model, single measures, and
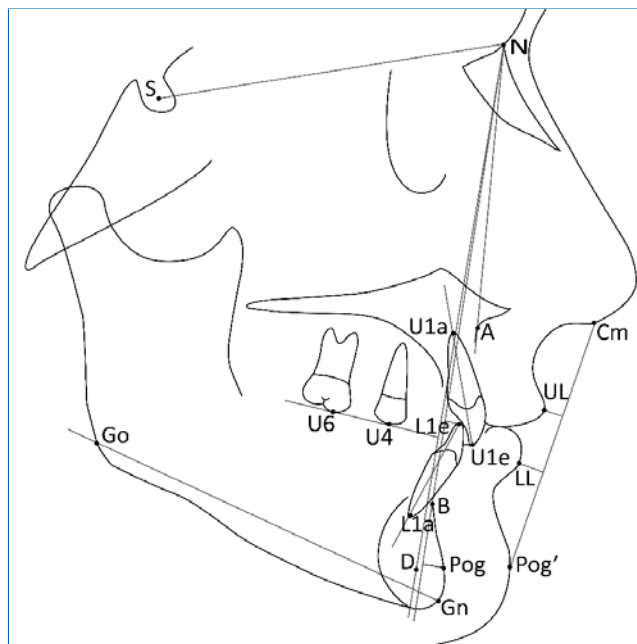
Figure 1. Description of the landmarks and measurements used in the study. Sella (S): Center of sella turcica radiographic outline. Nasion (N): Most anterior midline point on the frontonasal suture. Subspinale (Point A): Deepest midline point on the anterior outline of the maxilla. Supramentale (Point B): Deepest midline point on the anterior outline of the mandible. Point D: Center of the radiographic outline of the mandibular symphysis. Upper molar point (U6): Most superior and anterior cusp tip of the upper first permanent molar. Upper premolar point (U4): Most superior and anterior cusp tip of the upper first permanent premolar. Gonion (Go): Most posterior and inferior point on the angle of the mandible. Gnathion (Gn): Midpoint between the most anterior and inferior points on the mandibular symphysis. Upper incisor apex (U1a): Most apical point of the most prominent maxillary central incisor. Upper incisor edge (U1e): Most incisal point of the most prominent maxillary central incisor. Lower incisor apex (L1a): Most apical point of the most prominent mandibular central incisor. Lower incisor edge (L1e): Most incisal point of the most prominent mandibular central incisor. Collumela (Cm): The midpoint between Subnasale and the nasal tip. Soft tissue Pogonion (Pog´): Most anterior midline point of the soft tissue chin. Labrale superius (UL): Most anterior midline point of the upper lip. Labrale inferius (LL): Most anterior midline point of the lower lip. Pogonion (Pog): Most anterior midline point of the mandibular symphysis. SNA(°): the angle between Sella (S), Nasion (N), and A point; SNB(°): the angle between S, N, and B point; ANB(°): the angle between A, N, and B points; SND(°): the angle between S, N, and D points; UINA(°): the angle between the axis of the upper incisor (U1e-U1a) and NA line; LINB(°): the angle between the axis of the lower incisor (L1e-L1a) and NB line; GoGn/SN(°): the angle between the mandibular line (Go-Gn) and SN line; OL/SN(°): the angle between the occlusal line (U6-U4) and SN line; UINA (mm): the distance between the upper incisor edge and NA line; LINB (mm): the distance between the lower incisor edge and NB line; PogNB (mm): the distance between Pog and NB line; UL (mm): the distance between *labrale superius* and S line (Pog'-Columella, Cm); LL (mm): the distance between *labrale inferius* and S line.

absolute agreement. Kolmogorov-Smirnov tests were applied to verify normality before assessing trueness with paired Student's *t*-tests or Wilcoxon signed rank tests accordingly. The significance level was set at 0.05.

Results

Table 1 describes the means and standard deviations for cephalometric measurements and tracing time by session and method. Table 2 shows the trueness analysis results. Statistically significant differences were found between both test methods and NemoCeph for variables ANB (smartphone, -0.3±0.68; computer, -0.3±0.52), occlusal plane angle (smartphone, -1.1±2.68; computer, -2.0±2.98), and mandibular plane angle (smartphone, -0.5±1.27; computer, -0.8±1.56). Additionally, the lower lip measurement showed a statistically significant increase with smartphone tracing (0.3±0.57), while SNB (0.5±1.19) and SND (0.5±1.08) showed a significant increase in the computer OneCeph method. Figure 2 shows that the OneCeph tracings on the smartphone and computer produced, overall, a similar count of outlier observations, except for the inclination of the upper incisor, which only registered a relevant number of positive outliers for the smartphone analysis.

Statistically significant differences regarding tracing time were found for OneCeph both on the smartphone and computer (p<0.001). Tracing time was significantly longer and more variable on the phone (18.6±14.96), with an average increase of 29% compared to NemoCeph. Although OneCeph

tracings on the computer had a shorter tracing time, they still showed a statistically significant mean increase of 12% (8.0±8.61) compared to the reference method.

In terms of reproducibility (Table 3), the occlusal plane registered the lowest ICC estimates for both the smartphone (ICC [95%CI]: 0.888 [0.773, 0.944]) and the computer (ICC [95%CI]: 0.842 [0.583, 0.931]) OneCeph cephalometric analyses, with all other cephalometric measurements registering values above 90%. As for the method's repeatability (Table 4), intraclass correlation coefficients revealed OL/SN° as the least repeatable variable for NemoCeph (ICC:0.916; 95%CI [0.839, 0.957]) and for OneCeph on the computer (ICC:0.929; 95%CI [0.860, 0.965]). On the smartphone, OL/SN also registered low repeatability (ICC:0.889; 95%CI [0.739, 0.950]), only second to the linear variable UINA (ICC:0.831; 95%CI [0.687, 0.912]). Graphically, it seems that NemoCeph presented less T2-T1 variability, considering the lowest frequency of outlier observations (Figure 3). The linear variables presented similar variability between methods. Additionally, the outlier analysis revealed that the highest T2-T1 values were obtained with OneCeph on the smartphone for variables including the A-point. These discrepant observations were considered relevant for the present study and, as such, were included in the analysis.

Finally, when all cephalometric measurements were considered (Figure 4), OneCeph on the computer registered a higher frequency of inter-session differences below 1 degree/mm compared to the reference (77.8% vs. 76.0%). On the other hand, smartphone analysis produced the highest inter-session differences, with 3.2% of measurements corresponding to differences above 5 degrees/mm, compared to only 0.7% of cases

Table 1. Descriptive statistics for tracing time and cephalometric measurements by method and session. SD, standard deviation; min, minimum; max, maximum.

	NemoCeph				OneCeph				OneCeph on computer			
	T1		T2		T1		T2		T1		T2	
	Mean (SD)	[min; max]	Mean (SD)	[min; max]	Mean (SD)	[min; max]	Mean (SD)	[min; max]	Mean (SD)	[min; max]	Mean (SD)	[min; max]
Time (s)	64.1 (8.56)	[51.0; 85.0]	55.1 (4.12)	[48.0; 63.0]	82.7 (14.99)	[60.0; 121.0]	71.9 (7.20)	[59.0; 94.0]	72.1 (8.88)	[63.0; 103.0]	72.9 (6.94)	[60.0; 91.0]
SNA (°)	80.9 (3.55)	[74.9; 90.6]	81.3 (3.83)	[75.2; 92.8]	80.9 (3.83)	[74.5; 92.6]	80.7 (4.19)	[73.6; 93.6]	81.1 (4.22)	[74.0; 92.6]	80.8 (3.83)	[73.4; 91.4]
SNB (°)	77.9 (4.53)	[70.6; 88.8]	78.1 (4.67)	[69.9; 89.4]	78.2 (4.7)	[70.0; 89.2]	78.1 (4.73)	[70.2; 88.8]	78.4 (5.08)	[70.1; 94.1]	77.9 (4.60)	[70.0; 88.9]
ANB (°)	3.0 (3.28)	[-4.0; 9.8]	3.2 (3.36)	[-4.3; 9.3]	2.7 (3.43)	[-4.3; 9.4]	2.6 (3.12)	[-3.9; 7.6]	2.7 (3.22)	[-4.1; 9.7]	2.8 (3.26)	[-4.7; 9.4]
SND (°)	75.6 (4.59)	[68.0; 86.4]	75.8 (4.72)	[68.1; 86.6]	75.7 (4.74)	[68.1; 86.6]	75.8 (4.78)	[68.7; 86.2]	76.0 (5.06)	[68.3; 91.3]	75.6 (4.71)	[67.7; 86.5]
PogNB (mm)	1.5 (2.17)	[-1.8; 5.6]	1.6 (1.99)	[-2.5; 5.6]	1.5 (2.05)	[-1.9; 5.6]	1.5 (2.14)	[-2.3; 5.4]	1.5 (2.10)	[-2.3; 5.3]	1.6 (2.05)	[-2.3; 5.7]
OL/SN (°)	19.1 (5.57)	[6.5; 29.0]	18.9 (5.66)	[7.0; 31.8]	18.0 (6.39)	[2.2; 34.8]	16.5 (6.05)	[5.1; 27.3]	17.1 (6.64)	[5.2; 32.8]	17.9 (6.23)	[3.4; 33.8]
GoGn/SN (°)	35.4 (6.68)	[22.0; 51.9]	35.3 (6.51)	[22.9; 50.6]	34.9 (7.06)	[20.3; 52.7]	35.1 (6.91)	[19.9; 50.9]	34.6 (7.48)	[20.7; 52.0]	34.8 (7.03)	[21.2; 51.8]
UINA (°)	25.9 (7.86)	[-2.6; 37.9]	25.3 (7.92)	[-3.1; 37.5]	25.8 (7.81)	[-1.0; 37.3]	25.7 (7.42)	[-0.6; 37.8]	26.0 (7.64)	[-2.9; 38.3]	25.8 (7.45)	[-2.0; 39.1]
UINA (mm)	5.2 (2.60)	[-3.2; 10.1]	4.9 (2.80)	[-3.7; 9.6]	5.5 (3.02)	[-3.4; 10.6]	5.6 (3.30)	[-3.0; 16.1]	5.3 (2.76)	[-3.2; 10.7]	5.3 (2.64)	[-3.0; 10.3]
LINB (°)	26.3 (8.92)	[4.5; 42.5]	26.1 (9.05)	[5.2; 43.2]	26.7 (8.97)	[4.2; 44.0]	25.7 (9.03)	[4.0; 43.8]	26.5 (8.87)	[3.9; 44.1]	26.1 (8.81)	[2.3; 42.4]
LINB (mm)	5.2 (2.98)	[-2.1; 11.8]	5.2 (3.24)	[-2.4; 12.5]	5.2 (3.09)	[-2.8; 12.3]	5.1 (3.03)	[-2.2; 12]	5.1 (3.09)	[-2.7; 12.1]	5.3 (3.01)	[-2.0; 12.1]
UL (mm)	0.2 (2.76)	[-5.9; 6.2]	-0.1 (2.67)	[-6.2; 5.5]	0.3 (2.85)	[-7.0; 6.6]	0.1 (2.96)	[-7.1; 6.6]	0.0 (2.63)	[-5.7; 5.7]	0.0 (2.84)	[-6.5; 5.8]
LL (mm)	1.6 (2.82)	[-5.0; 8.7]	1.4 (2.74)	[-4.9; 7.7]	1.9 (3.00)	[-5.6; 8.3]	1.8 (3.01)	[-5.3; 8.7]	1.8 (2.86)	[-4.5; 8.6]	1.6 (2.89)	[-5.3; 8.6]

Table 2. Time and trueness analysis – descriptive statistics for method bias and comparison results obtained with paired Student's t-tests/Wilcoxon tests on T1 data. SD, standard deviation.

|	OneCeph – NemoCeph		OneCeph on computer – NemoCeph	
	Mean (SD)	**p**	**Mean (SD)**	**p**
Time (s)	18.6 (14.96)	<0.001	8.0 (8.61)	<0.001
SNA (°)	0.0 (1.11)	0.819	0.2 (1.36)	0.425
SNB (°)	0.3 (0.84)	0.079	0.5 (1.19)	0.027
ANB (°)	-0.3 (0.68)	0.010	-0.3 (0.52)	0.002
SND (°)	0.2 (0.83)	0.239	0.5 (1.08)	0.017
PogNB (mm)	0.0 (0.46)	0.882	0.0 (0.32)	0.425
OL/SN (°)	-1.1 (2.68)	0.025	-2.0 (2.98)	0.001
GoGn/SN (°)	-0.5 (1.27)	0.027	-0.8 (1.56)	0.004
UINA (°)	0.0 (3.16)	0.979	0.2 (2.07)	0.795
UINA (mm)	0.4 (1.05)	0.050	0.2 (0.86)	0.255
LINB (°)	0.4 (2.27)	0.351	0.2 (2.23)	0.676
LINB (mm)	0.0 (0.39)	0.762	-0.1 (0.36)	0.055
UL (mm)	0.1 (0.56)	0.292	-0.2 (0.69)	0.102
LL (mm)	0.3 (0.57)	0.008	0.1 (0.50)	0.189

Table 4. Precision analysis: intra-class correlation coefficients regarding method repeatability (T1 vs T2). ICC, intra-class correlation coefficient; 95% CI, 95% confidence interval.

	Method		
	NemoCeph	**OneCeph**	**OneCeph on computer**
	ICC [95% CI]	**ICC [95% CI]**	**ICC [95% CI]**
SNA (°)	0.971 [0.930, 0.986]	0.951 [0.905, 0.975]	0.945 [0.891, 0.972]
SNB (°)	0.985 [0.971, 0.993]	0.992 [0.984, 0.996]	0.962 [0.922, 0.981]
ANB (°)	0.993 [0.982, 0.997]	0.932 [0.868, 0.965]	0.987 [0.975, 0.994]
SND (°)	0.985 [0.969, 0.993]	0.993 [0.986, 0.997]	0.969 [0.936, 0.985]
PogNB (mm)	0.981 [0.963, 0.991]	0.970 [0.940, 0.985]	0.974 [0.948, 0.987]
OL/SN (°)	0.916 [0.839, 0.957]	0.889 [0.730, 0.950]	0.929 [0.860, 0.965]
GoGn/SN (°)	0.990 [0.981, 0.995]	0.989 [0.978, 0.994]	0.979 [0.959, 0.989]
UINA (°)	0.957 [0.915, 0.978]	0.930 [0.864, 0.964]	0.973 [0.946, 0.986]
UINA (mm)	0.973 [0.941, 0.987]	0.831 [0.687, 0.912]	0.962 [0.925, 0.981]
LINB (°)	0.981 [0.963, 0.991]	0.940 [0.882, 0.970]	0.965 [0.931, 0.982]
LINB (mm)	0.990 [0.979, 0.995]	0.985 [0.971, 0.993]	0.986 [0.971, 0.993]
UL (mm)	0.973 [0.944, 0.987]	0.977 [0.954, 0.989]	0.971 [0.943, 0.985]
LL (mm)	0.986 [0.963, 0.994]	0.988 [0.974, 0.994]	0.982 [0.964, 0.991]

Table 3. Precision analysis: intra-class correlation coefficients regarding method reproducibility on T1 measurements. ICC, intra-class correlation coefficient; 95% CI, 95% confidence interval.

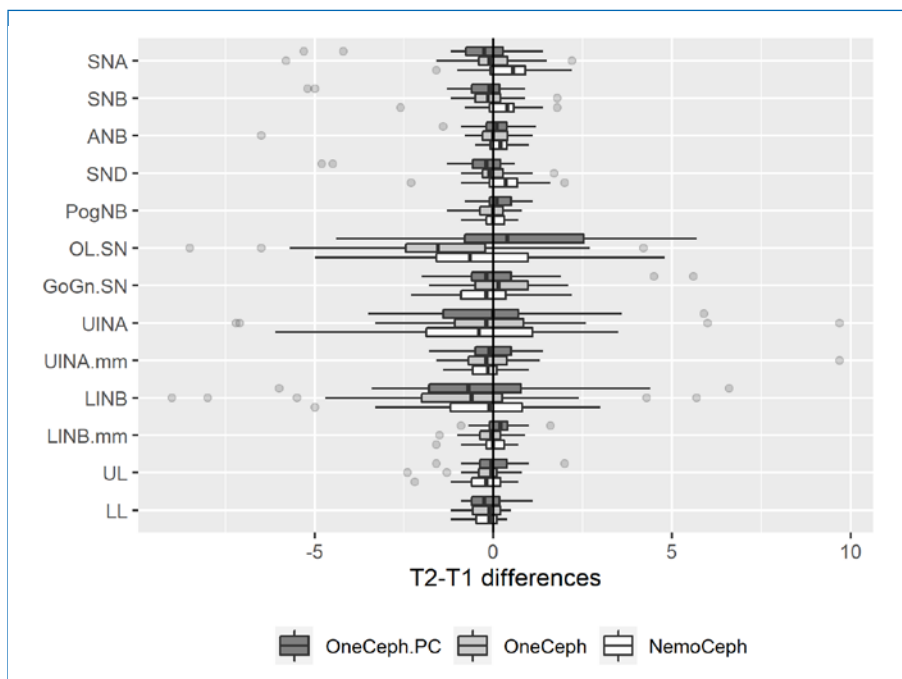|	Method	
	OneCeph vs. NemoCeph	**OneCeph on computer vs. NemoCeph**
	ICC [95% CI]	**ICC [95% CI]**
SNA (°)	0.956 [0.913, 0.978]	0.940 [0.884, 0.969]
SNB (°)	0.982 [0.964, 0.991]	0.966 [0.927, 0.983]
ANB (°)	0.976 [0.943, 0.989]	0.984 [0.954, 0.993]
SND (°)	0.984 [0.968, 0.992]	0.971 [0.936, 0.986]
PogNB (mm)	0.977 [0.955, 0.989]	0.989 [0.978, 0.994]
OL/SN (°)	0.888 [0.773, 0.944]	0.842 [0.583, 0.931]
GoGn/SN (°)	0.981 [0.959, 0.991]	0.970 [0.924, 0.987]
UINA (°)	0.921 [0.847, 0.960]	0.965 [0.931, 0.982]
UINA (mm)	0.925 [0.850, 0.962]	0.947 [0.897, 0.973]
LINB (°)	0.968 [0.937, 0.984]	0.969 [0.939, 0.984]
LINB (mm)	0.992 [0.984, 0.996]	0.992 [0.984, 0.996]
UL (mm)	0.980 [0.960, 0.990]	0.965 [0.931, 0.983]
LL (mm)	0.977 [0.944, 0.989]	0.984 [0.969, 0.992]

with NemoCeph.

Discussion

This study aimed to assess the accuracy of a cephalometric analysis app used in two input setups: a smartphone and a computer. Overall, both methods showed clinically acceptable results considering bias and random-error analysis compared to the reference method, as well as acceptable repeatability results. Regarding trueness analysis, although statistically significant differences were registered, these were not considered clinically significant, considering the magnitude of each cephalometric variable. Livas,[14] who compared OneCeph smartphone tracing with a computer program, also reported statistically significant differences regarding the method's bias, namely for the mandibular plane inclination, upper incisor protrusion, lower incisor protrusion, and interincisal angle. The present study only found significant bias in the smartphone method associated with the occlusal and mandibular plane inclinations. Additionally, it registered lower mean differences for these variables in smartphone tracings, which may be explained by methodological differences, such as using a stylus, as explained below.

Regarding the ICC estimates, while some papers define acceptability above 0.75 or 0.8,[17] others generally follow the classification cited by Koo & Li,[18] where values between 0.75 and 0.9 indicate good reliability and values greater than 0.90 indicate excellent reliability. However, whether an ICC value is good enough should depend on the intended use of the method,[19] and what may be considered acceptable in sociological and behavioral research may not be sufficient in medical re-
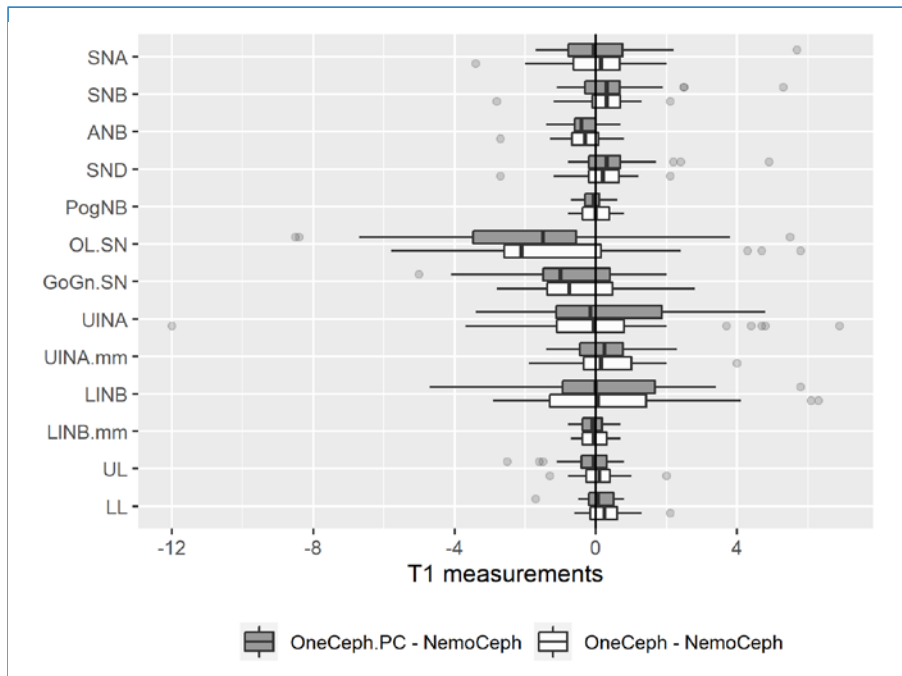
Figure 2. Trueness analysis – clustered boxplots for method bias by cephalometric variable, on T1 data.

Figure 3. Precision analysis – clustered boxplots of T2-T1 differences by cephalometric variable and method.

search,[20] especially considering intra-rater repeatability on cephalometric measurements. Therefore, this study considered a more conservative threshold. Following previous criteria used in sports science and medicine, the authors considered ICC point estimates 'high' over 0.9, 'moderate' between 0.8 and 0.9, and 'insufficient' below 0.8.[21] OneCeph methods

displayed high reproducibility for all cephalometric measurements except OL/SN, where point estimates decreased both on smartphone and computer OneCeph tracings, despite remaining above the 0.8 threshold. These results are in line with the study by Livas,[14] although the occlusal plane was not included in their analysis. The comparison of ICC estimates
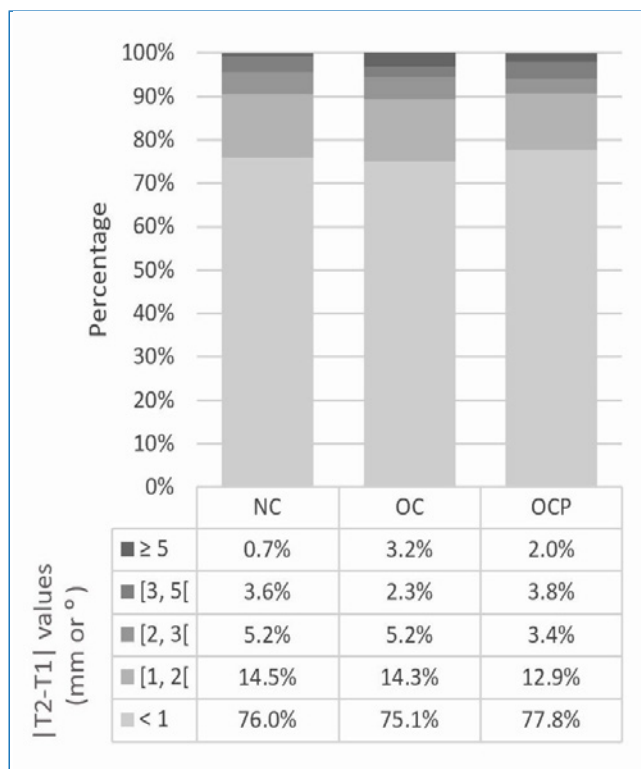
Figure 4. Stacked bar chart of absolute T2-T1 differences by method. NC, NemoCeph; O, OneCeph; OPC, OneCeph on computer.

|T2-T1| values (mm or °)	NC	OC	OCP
≥ 5	0.7%	3.2%	2.0%
[3, 5[	3.6%	2.3%	3.8%
[2, 3[	5.2%	5.2%	3.4%
[1, 2[	14.5%	14.3%	12.9%
< 1	76.0%	75.1%	77.8%

across test methods by cephalometric variables showed no tendency to favor either setup.

Regarding repeatability, high ICC estimates were registered for all variables measured with NemoCeph and OneCeph on the computer. The smartphone tracings, however, showed moderate repeatability for UINA (mm). The authors confirmed that by removing an outlier observation from the data, this ICC estimate increased to 0.964, which agrees with previously published results.[14] The outlier analysis for repeatability found the highest differences with OneCeph on a smartphone for variables including the A-point. Upon further investigation, a gross identification mistake performed during smartphone tracing was detected for this landmark. During landmark identification on the smartphone, because the cephalogram is magnified and not completely visible on screen, the user has to drag the image around to locate the next landmark; this may have resulted in the outlier observation if the correct structure remained cropped out of the screen. This finding was included in the data analysis as it may be related to the tracing technique itself and, therefore, serve to alert readers to a possible disadvantage of using handheld devices for cephalometric analysis.

Overall, the effect of image magnification on the accuracy of digital cephalometric analysis has not been described in the literature, possibly due to the inherent difficulty in quantifying it and because it most likely depends on several factors such as screen size, operator, cephalometric landmark, and an eventual interaction between these. Besides stating whether mag-

nification was performed during analysis, authors should also state the cephalogram image resolution, which, when provided with the screen resolution, indicates if more information would be visible upon magnification. For instance, the present sample involved cephalograms with a 2429x2121 resolution, traced either on a 1920x1280 smartphone or a 1366x768 computer screen; with this setup, not all image data is visible if either screen shows the entire cephalogram. Besides, cephalometric software usually includes horizontal and/or vertical toolbars that absorb screen space. Nevertheless, the amount of clinically relevant information in each pixel also depends on machine-related factors, some of which may be improved by the operator.[22] Thus, tracing time could be improved by using larger display screens for cephalometric analysis so that magnification would not be necessary.

However, the present version of NemoCeph software includes both automatic magnification and image repositioning so that the examiner does not have to drag the image around. This feature showed a clear positive impact on shortening tracing time and may enhance cephalometric measurements' repeatability based on the lower frequency discrepancies above 5 degrees/mm. The operator's previous familiarity with NemoCeph software might have also contributed to these results.

Another factor that may have interfered with the OneCeph methods' accuracy, and has not yet been described in the literature, is the type of pointer used for landmark location. The app displays an opaque circle shape to pinpoint the landmark after its identification and allows the user to make immediate corrections before accepting its position and continuing the analysis. The size of this marker may not remain constant in different magnifications, and during the analysis, the marker may easily approach the equivalent of 2 mm in diameter on the cephalogram scale. Although this has little to do with the examiner's incorrect location of the point when the landmark is first identified, it can affect error detection.

For landmark identification on the smartphone, a capacitive stylus was chosen due to its low cost and compatibility with touch screens, deeming it suitable for most smartphones and tablets.[23] The tip design was selected to favor accuracy, as an alternative to capacitive styli with larger soft rubber tips. Other options, including finer, ball-point pen-like tips, are limited to compatibility with specific models of handheld devices, meaning the input framework was evaluated and modified to allow for active pen input.[24] Despite the selection of the thinnest compatible tip, its 2-mm diameter may have incorporated errors that may have remained unidentified within the circle marker. This rationale may explain why tracing in OneCeph through the computer resulted in slightly better repeatability results since the identification was performed with the tip of the cursor. However, the use of screen-sharing software, even in its highest quality settings, may affect the image resolution or quality, not detailed by the manufacturers; this may have somehow affected tracing accuracy in this study.

Although the general results of the OneCeph app for cephalometric analysis in orthodontic practice seem favorable, cephalometric values should always be subject to clinical judgment alongside the radiograph and be regarded solely as an auxiliary tool for patient diagnosis.

The design chosen for the study involved treating cephalometric measurements directly as dependent variables, which favors more clinically relevant results. It has been shown that each landmark presents a non-random distribution of identification errors,[25] and this was not further explored in this study. However, an additional step in which the landmarks would have to be manually re-identified through software for coordinate pair attribution would likely imply additional operator error. A study design involving more than one examiner and more than two tracing sessions could have produced more relevant results, although impacting its feasibility.

Conclusions

OneCeph demonstrated adequate accuracy and efficiency both on smartphone and computer interfaces. Although its use in cephalometric analysis seems suitable in orthodontic clinical practice, clinical judgment is critical when interpreting the measurement output, especially if landmark identification is performed on handheld devices, which may result in a higher frequency of gross landmark identification errors.

Acknowledgments

Conflict of interest

The authors have no conflicts of interest to declare.

Ethical disclosures

Protection of human and animal subjects. The authors declare that no experiments were performed on humans or animals for this study.

Confidentiality of data. The authors declare that they have followed their work center protocols on access to patient data and for its publication.

Right to privacy and informed consent. The authors declare that no patient data appear in this article.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Mónica Amorim: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing. **Fernanda Diamantino:** Formal analysis, Supervision, Writing – review and editing. **Rui Pereira:** Conceptualization, Methodology, Supervision, Writing – review and editing. **Luís Jardim:** Conceptualization, Formal analysis, Methodology, Project administration, Resources, Supervision, Validation, Writing – review and editing.

ORCID

Mónica Amorim (iD) 0000-0001-9127-5955
Fernanda Diamantino (iD) 0000-0002-9217-167X
Rui Pereira (iD) 0000-0002-6544-5786
Luís Jardim (iD) 0000-0002-8869-3299

REFERENCES

1. Iwanicki J, Mikulewicz M, Zietek M. Computer technologies in dentofacial orthopeadics – historical review. Dent Med Probl. 2008;45:349-53.
2. Celik E, Polat-Ozsoy O, Toygar Memikoglu TU. Comparison of cephalometric measurements with digital versus conventional cephalometric analysis. Eur J Orthod. 2009;31:241-6.
3. Erkan M, Gurel HG, Nur M, Demirel B. Reliability of four different computerized cephalometric analysis programs. Eur J Orthod. 2012;34:318-21.
4. Paixão MB, Sobral MC, Vogel CJ, Araujo TMd. Comparative study between manual and digital cephalometric tracing using Dolphin Imaging software with lateral radiographs. Dent Press J Orthod. 2010;15:123-30.
5. Tikku T, Khanna R, Maurya R, Srivastava K, Bhushan R. Comparative evaluation of cephalometric measurements of monitor-displayed images by Nemoceph software and its hard copy by manual tracing. J Oral Biol Craniofac Res. 2014;4:35-41.
6. 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;31:523-8.
7. Choi JS, Yi B, Park JH, et al. The uses of the smartphone for doctors: an empirical study from Samsung Medical Center. Healthc Inform Re. 2011;17:131-8.
8. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res. 2012;14:e128.
9. Baheti MJ, Toshniwal N. Orthodontic apps at fingertips. Progress in orthodontics. 2014;15:36.
10. Singh P. Orthodontic apps for smartphones. J Orthod. 2013;40:249-55.
11. Aksakalli S, Yilanci H, Gorukmez E, Ramoglu SI. Reliability assessment of orthodontic apps for cephalometrics. Turk J Orthod. 2016;29:98-102.
12. Sayar G, Kilinc DD. Manual tracing versus smartphone application (app) tracing: a comparative study. Acta Odontol Scand. 2017;75:588-94.
13. Shettigar P, Shetty S, Naik RD, Basavaraddi SM, Patil AK. A comparative evaluation of reliability of an android-based app and computerized cephalometric tracing program for orthodontic cephalometric analysis. Biomed Pharmacol J. 2019;12:341-6.
14. Livas C, Delli K, Spijkervet FK, Vissink A, Dijkstra PU. Concurrent validity and reliability of cephalometric analysis using smartphone apps and computer software. Angle Orthod. 2019;89:889-96.
15. Steiner CC. Cephalometrics for you and me. Am J Orthod. 1953;39:729-55.
16. Thilander B, Persson M, Adolfsson U. Roentgen–cephalometric standards for a Swedish population. A

longitudinal study between the ages of 5 and 31 years. Eur J Orthod. 2005;27:370-89.

17. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420-8.

18. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155-63.

19. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation – a discussion and demonstration of basic features. PloS ONE. 2019;14:e0219854.

20. Vaz S, Falkmer T, Passmore AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test–retest reliability. PloS ONE. 2013;8:e73990.

21. Vincent WJ, Weir JP. Statistics in kinesiology. 4th ed. Champaign, IL: Human Kinetics; 2012.

22. Uffmann M, Schaefer-Prokop C. Digital radiography: The balance between image quality and required radiation dose. Eur J Radiol. 2009;72:202-8.

23. Wilcher D. Arduino Electronics Blueprints. Birmingham, UK: Packt Publishing Ltd; 2015.

24. Hughes AS. Active pen input and the android input framework. San Luis Obispo, CA: California Polytechnic State University; Master of Science in Electrical Engineering; 2011.

25. Baumrind S, Frantz RC. The reliability of head film measurements: 1. Landmark identification. Am J Orthod. 1971;60:111-27.