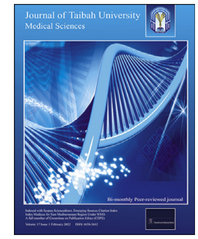
Journal of Taibah University Medical Sciences (2022) 17(1), 57–66

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com



Original Article

The accuracy and reliability of WebCeph for cephalometric analysis

Yassir A. Yassir, PhD [a,*], Aya R. Salman, B.D.S. [b] and Sarah A. Nabbat, MSc [b]

[a] *Orthodontic Department, College of Dentistry, University of Baghdad, Iraq*
[b] *Department of Orthodontics, Ministry of Health, Baghdad, Iraq*

Received 5 May 2021; revised 27 July 2021; accepted 4 August 2021; Available online 22 September 2021

الملخص

أهداف البحث: تقارن هذه الدراسة دقة وموثوقية "ويبسيف" (برنامج قائم على الويب لتحليل قياس الرأس) مع برنامج الكمبيوتر أوتوكاد.

طرق البحث: تم تحليل عينة من مخططات رأسية جانبية رقمية لـ ٥٠ مريضا لتقويم الأسنان قبل المعالجة باستخدام برنامج "ويبسيف" وبرنامج أوتوكاد (كمقياس قياسي). في كل مخطط رأسي، تم وضع علامة على ١٧ معلما و١١ قياسا وتنفيذها كمعلمات للهيكل العظمي والأسنان والأنسجة الرخوة. استخدمنا ست زوايا وخمس قياسات خطية. كما تم استخدام اختبار تي المقترن لتقييم التحيز المنهجي. واستخدام تحليل الارتباط داخل الطبقة واختطاط بلاند-ألتمان مع تحليل الانحدار الخطي لتقييم التوافق بين الطريقتين.

النتائج: كانت هناك موثوقية كافية للقياسات باستخدام "ويبسيف" وأوتوكاد. أظهر اختبار تي المقترن فروق ذات دلالة إحصائية لخمس زوايا واثنين من القياسات الخطية. كما أظهر اختبار الارتباط داخل الطبقة بين "ويبسيف" وأوتوكاد توافقا جيدا جدا إلى ممتاز لجميع القياسات باستثناء القاطعة السفلية لزاوية مستوى الفك السفلي. وأظهر اختطاط بلاند-ألتمان حدا مقبولا نسبيا للاتفاق لثلاث زواويا واثنين من القياسات الخطية فقط، وكشف تحليل الانحدار الخطي عن تحيز نسبي كبير بين الطريقتين لأربع زوايا وخط تجميل الشفة العليا. وتم تحسين التحيز المنهجي ومستوى الاتفاق باستخدام "ويبسيف" شبه التلقائي.

الاستنتاجات: هناك مشاكل مختلفة متأصلة في "ويبسيف" التلقائي مثل ضعف تحديد المعالم / تتبع الأنسجة الرخوة وعدم اتساق القياسات. يمكن لـ "ويبسيف" شبه التلقائي التغلب على بعض قيود "ويبسيف" التلقائي، الذي يجب استخدامه لتحليل قياس الرأس بقدر كبير من الحذر.

الكلمات المفتاحية: تحليل قياس الرأس؛ تقويم الأسنان؛ برامج قياس الرأس؛ ويبسيف؛ اختبار الارتباط داخل الطبقة

* Corresponding address: Orthodontic Department, College of Dentistry, University of Baghdad, Iraq.
E-mail: yassirkyassir@gmail.com (Y.A. Yassir)

Peer review under responsibility of Taibah University.

Production and hosting by Elsevier

Abstract

Objective: This study compares the accuracy and reliability of WebCeph (web-based program for cephalometric analysis) with the AutoCAD computer software.

Materials and methods: A sample of pretreatment digital lateral cephalograms of 50 orthodontic patients was analysed with WebCeph and AutoCAD software (as a standard measure). On each cephalogram, 17 landmarks and 11 measurements were marked and performed as skeletal, dental, and soft–tissue parameters. We used six angular and five linear measurements. A paired t-test was used to assess the systematic bias. The intraclass correlation coefficient (ICC) and Bland–Altman plot with linear regression analysis were used to assess the agreement between the two methods.

Results: There was adequate reproducibility for the measurements with both WebCeph and AutoCAD. The paired t-test showed statistically significant differences for five angular and two linear measurements ($P < 0.05$). The ICC test between WebCeph and AutoCAD revealed very good to excellent agreement for all measurements, except for the lower incisor to mandibular plane angle. The Bland–Altman plot visually showed a relatively acceptable limit of agreement for three angular and two linear measurements only, and the linear regression analysis revealed a significant proportional bias between the two methods for four angles and the upper lip-Esthetic line (U Lip-E Line). The systematic bias and level of agreement improved with the use of the semi-automatic WebCeph.

Conclusions: Different problems, such as poor landmark identification/soft tissue tracing and inconsistency of measurements, are inherent to the automatic WebCeph. The semi-automatic WebCeph can overcome some limitations of the automatic WebCeph; however, it should be

1658-3612 https://doi.org/10.1016/j.jtumed.2021.08.010

used for cephalometric analysis with a great deal of caution.

Keywords: Cephalometric analysis; Cephalometric software; Intraclass correlation; Orthodontic; WebCeph

Introduction

Cephalometric analysis aids in evaluating dentofacial proportions, identifying the anatomic basis of malocclusion, and analysing growth and treatment-related changes.[1] During orthodontic/orthognathic treatment planning, cephalometric analysis is considered an essential diagnostic method, particularly when a skeletal discrepancy exists.[2,3]

Recent improvements in computer and software technology have led to the introduction of computer-aided cephalometric analysis in addition to the conventional (manual) method[4] which includes direct measurement of cephalometric angles with a protractor using an overlay of acetate tracing paper secured over the radiograph after identifying the landmarks and drawing the lines with pencil and ruler, with or without the use of a view box.[5]

Computerised software and smartphone applications can automatically identify the landmarks and complete the measurements once the digital radiograph is imported using artificial intelligence (AI) technology (e.g. WebCeph web program). Alternatively, the operator can manually identify the landmarks and then have the measurements automatically calculated (e.g. OneCeph application, WebCeph web program) or give orders to calculate specific measurements determined by the operator (e.g. AutoCAD software). The latter operations are considered semi-automated cephalometric analysis.

The use of AI is prevalent in numerous aspects of daily life, and AI-based algorithms are now widely used in technology. Given these recent developments in computing, such AI algorithms can be used for simple and complicated tasks and, consequently, show promise for various health care fields.[6] Despite the variety of applicable techniques for automated identification of cephalometric landmarks, it remains uncertain whether these methods are able to detect cephalometric landmarks within a clinically acceptable range.[7] A systematic review by **Leonardi** et al.[8] revealed little scientific evidence to support the use of automatic landmark identification when compared to manual tracing due to the greater number of errors. Recently, various studies[6,7,9,10] have shown that AI is as accurate as human examiners in landmark identification and could be a feasible option for repeated identification of multiple cephalometric landmarks. However, a systematic review by **Hung** et al.[11] concluded that it is necessary to confirm the reliability and applicability of AI models before they are introduced to clinical practice.

WebCeph is an AI-based orthodontic and orthognathic online platform that is recently gaining popularity due to its many desirable options that may simplify orthodontic treatment planning and acquisition of patient records. These include automatic cephalometric tracing, cephalometric analysis, visual treatment simulation, automatic superimposition, image archiving, and a photo gallery. Additionally, WebCeph allows for manual landmark editing with automatic calculation of measurements.

The AutoCAD software can be used to determine linear and angular measurements as well as other measurements such as surface area of any digital image. While this software was not originally designed for cephalometric analysis, it permits magnification adjustment and easy handling of any digital image to simplify landmark identification and line drawing. It has shown adequate reliability in landmark identification and angular and linear measurement when compared to the Viewbox 3.1.1 software.[12] Additionally, it has shown reliability and accuracy in linear and angular measurement in other medical fields when compared to manual measurement with a goniometer[13,14] or other software packages, such as ImageTool® and Adobe Photoshop®.[15]

Computer software must be accurate and reliable.[16] Because software is now widely available and regularly used in orthodontic practice, it is necessary to assess their accuracy to identify the most appropriate type to use.[4,17]

Therefore, the AutoCAD software (manual landmark identification) and WebCeph web program (both automatic and semi-automatic cephalometric tracing) were selected to be compared in this study with the aim of evaluating their accuracy and reliability (especially between AutoCAD and the automatic tracing of WebCeph). The null hypothesis stated that there would be no significant difference in the cephalometric tracing performed by WebCeph and AutoCAD.

Materials and Methods

Sample

Fifty cephalometric radiographs were selected for this study from a list of pre-orthodontic patients' records, using a systematic randomisation method. These were divided into gender and stratified according to different skeletal patterns. All the radiographs were taken by the same machine (Planmeca Promax®), using a standard method whereby the patients were positioned in the cephalostat with the Frankfort plane parallel to the floor, the sagittal plane at a right angle to the path of the X-ray, and the teeth in centric occlusion. To minimise random errors, the following exclusion criteria were applied: patients with impacted, un-erupted, or missing incisors; craniofacial deformity; and radiographs with poor quality.

The current study was carried out in compliance with the Declaration of Helsinki, and all applicable data protection laws were respected.

Method

Each lateral cephalometric radiograph digital image was initially assigned a specific identifier and then downloaded and saved to a computer (Dell, Latitude E5440) before being imported to the WebCeph and AutoCAD software. Magnification correction was undertaken based on a known distance of 10 mm between two fixed points on the cephalostat rod in the radiograph. Landmark identification was performed twice on the digital images with both options of WebCeph (automatically and semi-automatically). For the AutoCAD, landmarks identified manually using the cursor and then the calculation of each measurement was completed by giving orders to measure angles and lines. Finally, each analysed image was stored individually. All the measurements were carried out by the same investigator (ARS) on a daily basis of 10 radiographs/day using WebCeph first. Then, after an interval of two weeks, the same radiographs were re-measured using AutoCAD software.

To achieve adequate calibration, the investigator (ARS) who performed the measurements received training from an expert orthodontist (YAY). Furthermore, 10 radiographs were measured with AutoCAD by both investigators (ARS and YAY) to test for inter-examiner reliability and by the same investigator on two different occasions with an interval of four weeks to assess intra-examiner reliability.

Seventeen landmarks were marked on each cephalogram, and 11 measurements were made, indicating skeletal, dental, and soft tissue parameters, including six angular and five linear measurements. Bilateral structures were averaged to make a single landmark.[5,18]

The following measurements were used in the study: SNA (), SNB (), ANB (), SN-MP (), U1-MaxP (), L1-MP (), N–Me (mm), U1-NPog (mm), L1-NPog (mm), Upper lip Lip-E Line (mm), Lower Lip-E Line (mm).

Statistical analysis

The statistical analyses were performed using the Statistical Package for Social Sciences for Windows, version 25.0 (SPSS Inc., Chicago, Illinois, USA). Inter- and intra-examiner reliability was assessed using the intraclass correlation coefficient (ICC) for all the measurements. Ten radiographs were re-measured after a four-week interval with AutoCAD and semi-automatic WebCeph, whereas no reliability test was required for the automatic WebCeph, as the measurements were completed automatically. The minimum, maximum, mean, mean difference, and standard deviation were used to describe the data. The Shapiro–Wilk test was used to inspect the normality of the data distribution. A paired t-test was used to assess the systematic bias between the two methods (in the case of abnormally distributed data, the Wilcoxon signed-rank test was used instead). Intraclass correlation and a Bland–Altman plot (limit of agreement) with linear regression analysis were used to assess the agreement between the two methods of measurements. The significance level was set as $P < 0.05$, and a 95% confidence interval was estimated for the outcomes in the study groups. The ICC values were set as follows:[19]

ICC <0.5: Poor reliability/agreement
ICC between 0.5 and 0.75: Moderate reliability/agreement
ICC between 0.75 and 0.9: Good reliability/agreement
ICC >0.90: Excellent reliability/agreement

A clinically relevant difference was determined when the difference in the angular and linear measurements was greater than 2° or 2 mm, respectively.[20–22]

Results

The total analysed cephalograms were 50 for each group (25 female, 25 male). The distribution of malocclusion was as follows: 30 skeletal Class I, 10 skeletal Class II, and 10 skeletal Class III. Figure 1 shows the measurements used in the currect study.

Intra-examiner reliability

The ICC revealed that the inter-examiner reliability was excellent (0.972–0.999) and the intra-examiner reliability very good to excellent (0.843–0.999) for the measurements by the AutoCAD software. For the semi-automatic WebCeph, retesting the same radiographs twice resulted in an ICC that ranged from 0.756 to 0.980.

Normality

The assumption of normality was violated for seven variables with the automatic WebCeph, two variables with the semi-automatic WebCeph, and three variables with AutoCAD ($P < 0.05$), as revealed by the Shapiro–Wilk test. This was confirmed by inspecting the histograms and Q–Q plots of these variables.

Descriptive statistics

The descriptive statistics showed that there were no missing data. The maximum differences (compared with AutoCAD) for the angular measurements were 8.42° and 7.42° (L1-MP angle) for the semi-automatic and automatic WebCeph, respectively, whereas for the linear measurements, the difference was 2.66 mm for N–Me. With the exception of the L1-MP angle, all the measurements with the semi-automatic WebCeph were closer to those of AutoCAD when compared with the automatic WebCeph (Tables 1 and 2).

Systematic bias

It was decided to use a paired t-test backed up by the Wilcoxon signed-rank test, as some variables were not

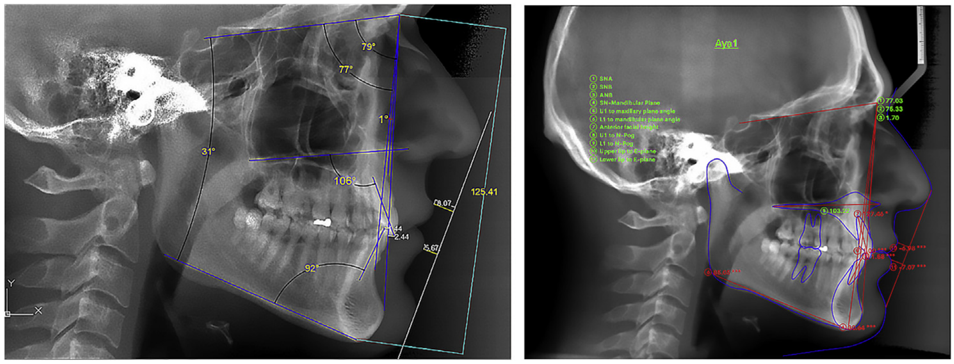


Figure 1: Landmarks and measurements used in this study (Above: AutoCAD, Below: WebCeph).

normally distributed. The paired t-test showed that seven variables, namely SNB, ANB, SN-MP, U1-MaxP, L1-MP, N–Me, and upper lip-E line, significantly differed between the automatic WebCeph and AutoCAD ($P < 0.05$) (Table 2). The Wilcoxon signed-rank test revealed similar results to that of the paired t-test, except for the lower lip-E line, which was statistically significant in the Wilcoxon signed-rank test ($P = 0.04$). The differences between the semi-automatic WebCeph and AutoCAD were significant for five variables (SN-MP, L1-MP, N–Me, L1-NPog, and upper lip-E line). The Wilcoxon signed-rank test confirmed these results (Table 2).

Agreement

The ICC test revealed no agreement between the automatic and semi-automatic WebCeph and AutoCAD for the L1-MP angle, whereas the other measurements showed good to excellent agreement between the automatic WebCeph and AutoCAD (0.768 and 0.910) and excellent agreement

Table 1: Descriptive Statistics for the measurements used in this study.

Variables	N	Minimum	Maximum	Mean	Std. Deviation
SNA Web Auto (°)	50	72	88	82.29	3.01
SNA Web Semi (°)	50	73	91	82.07	3.50
SNA AutoCAD (°)	50	72	92	82.14	3.71
SNB Web Auto (°)	50	71	83	78.32	2.84
SNB Web Semi (°)	50	72	86	79.58	3.19
SNB AutoCAD (°)	50	71	86	79.60	3.44
ANB Web Auto (°)	50	−1	9	3.95	2.17
ANB Web Semi (°)	50	−4	9	2.49	2.36
ANB AutoCAD (°)	50	−4	8	2.50	2.31
SN-MP Web Auto (°)	50	28	46	34.03	4.47
SN-MP Web Semi (°)	50	22	45	32.55	4.60
SN-MP AutoCAD (°)	50	22	45	33.04	4.73
U1-MaxP Web Auto (°)	50	102	125	112.48	4.99
U1-MaxP Web Semi (°)	50	102	134	114.65	6.41
U1-MaxP AutoCAD (°)	50	99	135	114.40	7.31
L1-MP Web Auto (°)	50	73	100	86.86	4.90
L1-MP Web Semi (°)	50	73	104	85.86	6.22
L1-MP AutoCAD (°)	50	76	107	94.28	6.12
N–Me Web Auto (mm)	50	107.94	140.02	118.89	7.43
N–Me Web Semi (mm)	50	104.89	135.45	116.89	7.15
N–Me AutoCAD (mm)	50	106.36	133.85	116.24	7.39
U1-NPog Web Auto (mm)	50	1.58	19.21	7.22	4.08
U1-NPog Web Semi (mm)	50	0.88	17.74	7.06	4.05
U1-NPog AutoCAD (mm)	50	0.73	17.63	7.11	3.97
L1-NPog Web Auto (mm)	50	0.19	14.09	4.10	2.79
L1-NPog Web Semi (mm)	50	0.30	14.50	3.79	2.87
L1-NPog AutoCAD (mm)	50	0.40	14.75	4.11	2.90
U Lip-E Line Web Auto (mm)	50	−9.10	13.25	−3.28	3.39
U Lip-E Line Web Semi (mm)	50	−10.61	0.30	−4.23	2.50
U Lip-E Line AutoCAD (mm)	50	−10.09	2.72	−4.73	2.63
L Lip-E Line Web Auto (mm)	50	−7.50	8.65	−1.93	3.27
L Lip-E Line Web Semi (mm)	50	−6.95	6.00	−1.49	2.80
L Lip-E Line AutoCAD (mm)	50	−6.73	7.54	−1.69	2.94

between the semi-automatic WebCeph and AutoCAD (ICC: 0.961−0.991) (Table 3). Figures 2 and 3 show the Bland−Altman plot for the measurement means and differences between the automatic WebCeph and AutoCAD. The error size was not within the acceptable range for most of the variables, as some points deviated from the limit of agreement and were not that close to the mean difference line. Additionally, a trend of a greater number of points above and below the mean difference was also noted. The only variables that visually showed a relatively acceptable limit of agreement were the SNA angle, ANB angle, SN-MP angle, N−Me, and L1-NPog. To confirm this, a linear regression analysis (Table 4) revealed a significant proportional bias between the two methods, i.e. significant differences for the SNA, SNB, U1MaxP, L1-MP angles, and upper lip-E line.

Discussion

The lateral cephalometric radiograph is an essential record for the diagnosis of anteroposterior and vertical discrepancies and the evaluation of the relationship between soft tissue and dental structures.[23] Therefore, the method used for cephalometric analysis must be accurate, safe, and highly reproducible.[16]

This study was primarily designed to evaluate the accuracy and reliability of WebCeph for cephalometric analysis in comparison to the AutoCAD software. AutoCAD had been proved in previous studies[12−15] to produce adequate linear and angular measurements and was therefore used in the current study as a standard measure. WebCeph has been recently introduced as an AI-based orthodontic and orthognathic online platform that can automatically perform cephalometric analysis and assist in arranging patients' records according to a uniform standard. However, its performance has not yet been tested, and it is necessary to verify its accuracy before it can be approved for clinical use.

The results of the current study revealed several areas in which the two programs did not concur, especially in terms of systematic bias. Therefore, we can reject the null hypothesis stating that there would be no significant difference in the cephalometric analysis performed by WebCeph and AutoCAD.

Both programs achieved adequate reliability when tested at different intervals. As inter-examiner errors have been found to be more frequent than intra-examiner errors,[24] all the measurements were undertaken by a single operator to achieve standardisation and avoid possible errors between operators.

Different sagittal and vertical skeletal patterns were included in this study to ensure the presence of all the possible variations in anteroposterior and vertical jaw relationships that could be faced during cephalometric tracing and analysis. Although in some instances the standard deviation was found to be slightly greater with the AutoCAD results, testing the normality revealed that the heterogeneity was greater with WebCeph. This may be explained by the inconsistency of the analysis undertaken with WebCeph with wide range of variation of most of the measurements when comapred to AutoACD.

Table 2: Paired t test assessing the systematic bias between the WebCeph and AutoCAD (d.f. = 49).

Paired Variables	Mean Difference	95% Confidence Interval of the Difference		t	P
		Lower	Upper		
SNA Web Auto - SNA AutoCAD	0.15	−0.59	0.89	0.41	0.684
SNA Web Semi - SNA AutoCAD	−0.07	−0.35	0.20	−0.52	0.603
SNB Web Auto - SNB AutoCAD	−1.29	−1.90	−0.67	−4.18	0.000***
SNB Web Semi - SNB AutoCAD	−0.02	−0.24	0.21	−0.14	0.892
ANB Web Auto - ANB AutoCAD	1.45	1.10	1.80	8.29	0.000***
ANB Web Semi - ANB AutoCAD	−0.02	−0.19	0.16	−0.17	0.865
SN-MP Web Auto - SN-MP AutoCAD	0.99	0.13	1.86	2.31	0.025*
SN-MP Web Semi - SN-MP AutoCAD	−0.49	−0.74	−0.24	−3.90	0.000***
U1-MaxP Web Auto - U1-MaxP AutoCAD	−1.92	−3.19	−0.64	−3.01	0.004**
U1-MaxP Web Semi - U1-MaxP AutoCAD	0.17	−0.60	0.94	0.45	0.653
L1-MP Web Auto - L1-MP AutoCAD	−7.42	−10.42	−4.41	−4.96	0.000***
L1-MP Web Semi - L1-MP AutoCAD	−8.42	−11.88	−4.96	−4.89	0.000***
N−Me Web Auto - N−Me AutoCAD	2.66	1.65	3.66	5.31	0.000***
N−Me Web Semi - N−Me AutoCAD	0.65	0.15	1.15	2.61	0.012*
U1-NPog Web Auto - U1-NPog AutoCAD	0.11	−0.30	0.52	0.55	0.588
U1-NPog Web Semi - U1-NPog AutoCAD	−0.06	−0.35	0.24	−0.38	0.705
L1-NPog Web Auto - L1-NPog AutoCAD	−0.00	−0.31	0.30	−0.02	0.985
L1-NPog Web Semi - L1-NPog AutoCAD	−0.31	−0.44	−0.19	−5.07	0.000***
U Lip-E Line Web Auto - U Lip-E Line AutoCAD	1.46	0.78	2.13	4.31	0.000***
U Lip-E Line Web Semi - U Lip-E Line AutoCAD	0.50	0.24	0.76	3.88	0.000***
L Lip-E Line Web Auto - L Lip-E Line AutoCAD	−0.24	−0.91	0.43	−0.72	0.475
L Lip-E Line Web Semi - L Lip-E Line AutoCAD	0.19	−0.08	0.47	1.43	0.159

* Significant.
** Highly significant.
*** Very highly significant.

Table 3: Intraclass correlation coefficient for the measurements between the WebCeph and AutoCAD.

Method	Variables	Intraclass Correlation	95% Confidence Interval	
			Lower Bound	Upper Bound
Web Auto/AutoCAD	SNA (°)	0.828	0.696	0.902
Web Semi/AutoCAD		0.982	0.968	0.990
Web Auto/AutoCAD	SNB (°)	0.828	0.597	0.916
Web Semi/AutoCAD		0.986	0.975	0.992
Web Auto/AutoCAD	ANB (°)	0.825	0.049	0.942
Web Semi/AutoCAD		0.982	0.969	0.990
Web Auto/AutoCAD	SN-MP (°)	0.868	0.761	0.926
Web Semi/AutoCAD		0.988	0.971	0.995
Web Auto/AutoCAD	U1-MaxP (°)	0.832	0.678	0.909
Web Semi/AutoCAD		0.961	0.931	0.978
Web Auto/AutoCAD	L1-MP (°)	−1.570	−2.116	0.102
Web Semi/AutoCAD		−1.997	−2.182	−2.142
Web Auto/AutoCAD	N−Me (mm)	0.910	0.681	0.963
Web Semi/AutoCAD		0.983	0.968	0.991
Web Auto/AutoCAD	U1-NPog (mm)	0.968	0.943	0.982
Web Semi/AutoCAD		0.984	0.971	0.991
Web Auto/AutoCAD	L1-NPog (mm)	0.964	0.936	0.980
Web Semi/AutoCAD		0.991	0.967	0.996
Web Auto/AutoCAD	U Lip-E Line (mm)	0.768	0.472	0.885
Web Semi/AutoCAD		0.959	0.899	0.980
Web Auto/AutoCAD	L Lip-E Line (mm)	0.833	0.706	0.905
Web Semi/AutoCAD		0.971	0.948	0.983

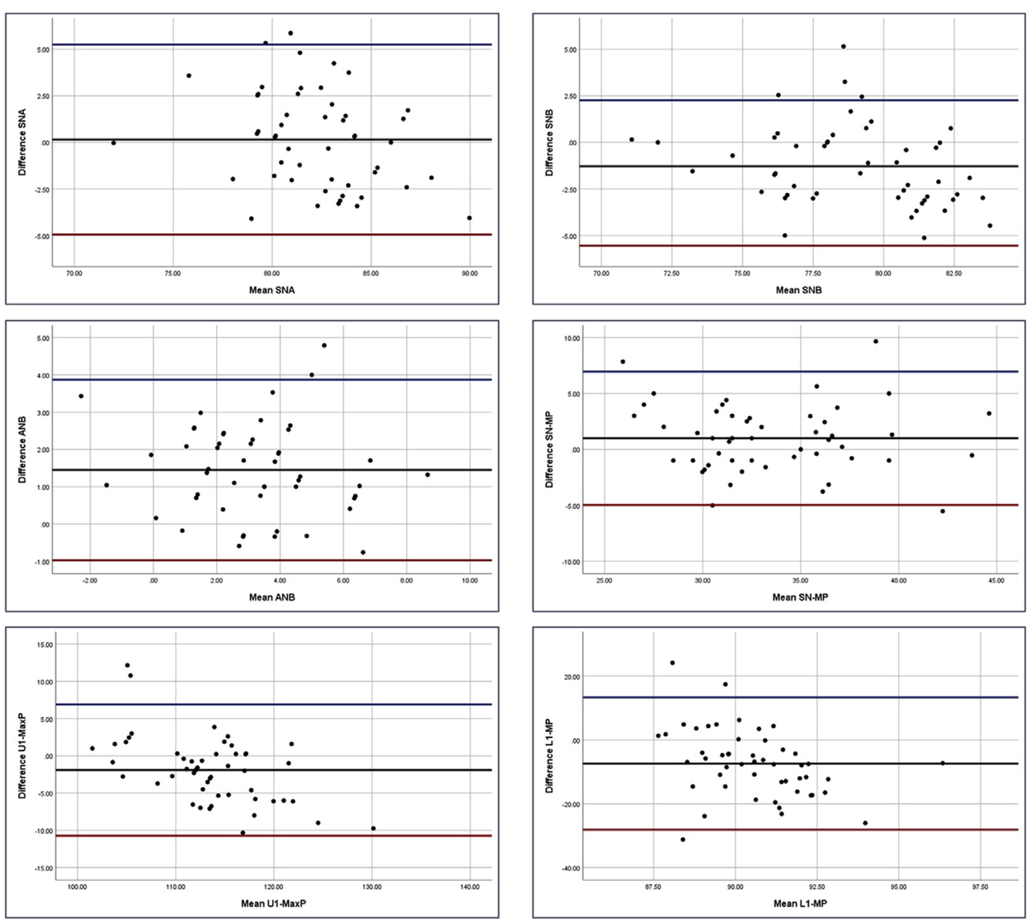


Figure 2: Bland−Altman plot for the angular measurement means and differences of the automatic WebCeph and AutoCAD.

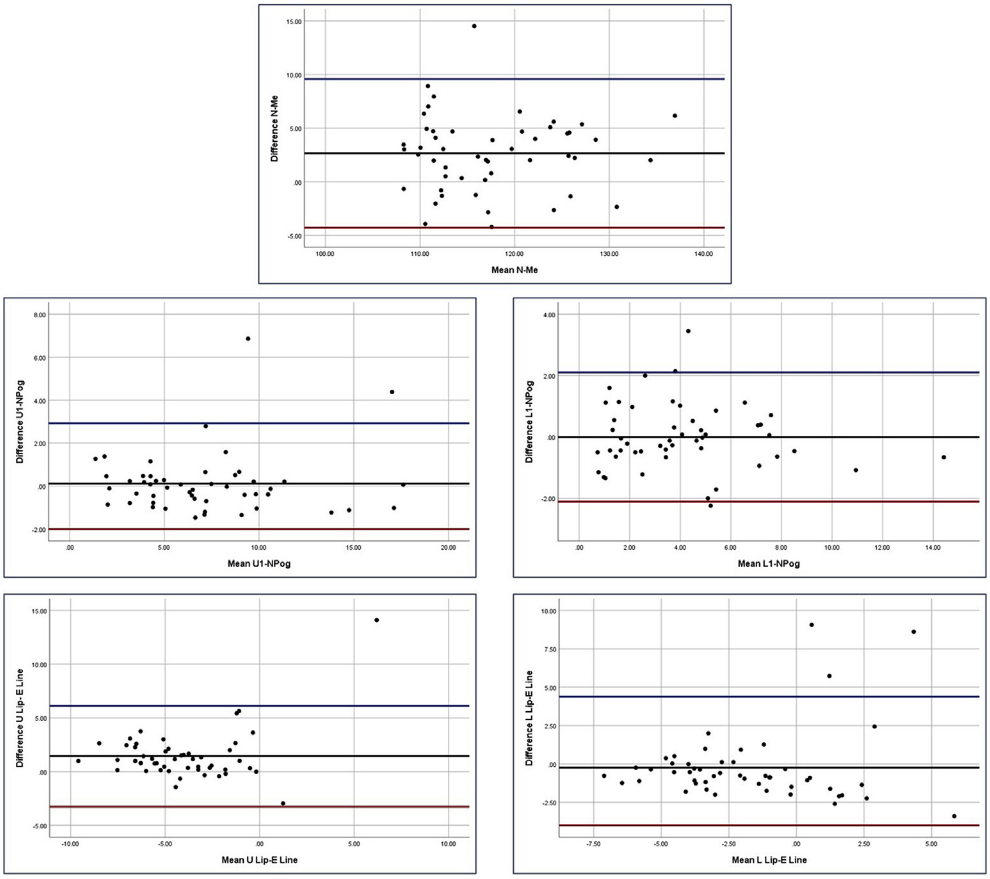


Figure 3: Bland–Altman plot for the linear measurement means and differences of the automatic WebCeph and AutoCAD.

Table 4: Linear regression analysis assessing the proportional bias between the automatic WebCeph and AutoCAD.

Variable Means	Unstandardized Coefficients		t	Sig.
	B	Std. Error		
SNA (°)	−0.24	0.12	−2.11	0.040*
SNB (°)	−0.21	0.10	−2.08	0.042*
ANB (°)	−0.07	0.08	−0.79	0.432
SN-MP (°)	−0.06	0.10	−0.63	0.530
U1-MaxP (°)	−0.42	0.09	−4.49	0.000***
L1-MP (°)	−2.40	0.84	−2.85	0.006**
N−Me (mm)	0.01	0.07	0.08	0.933
U1-NPog (mm)	0.03	0.05	0.55	0.587
L1-NPog (mm)	−0.04	0.06	−0.69	0.494
U Lip-E Line (mm)	0.29	0.12	2.52	0.015*
L Lip-E Line (mm)	0.13	0.12	1.08	0.288

* Significant.
** Highly significant.
*** Very highly significant.

Regarding the systematic bias between the automatic WebCeph and AutoCAD, two variables (L1-MP and N−Me) showed significant clinical and statistical differences greater than 2° for the angular measurement and 2 mm for the linear measurement. Six variables showed a significant statistical difference that was slightly below the assigned level of clinical difference: lower lip-E line (0.24 mm), SN-MP (0.99°), SNB (1.29°), ANB (1.45°), U1-MaxP (1.92°), and upper lip-E line (1.46 mm). The upper and lower incisor angles were significantly decreased in WebCeph, and it was found by checking the data individually that the two methods produced differences that were often greater than 20°, owing to miscalculations by WebCeph. However, the systematic bias was reduced when using the semi-automatic WebCeph, and the differences, except for that of L1-MP, were of no clinical importance. AutoCAD software was

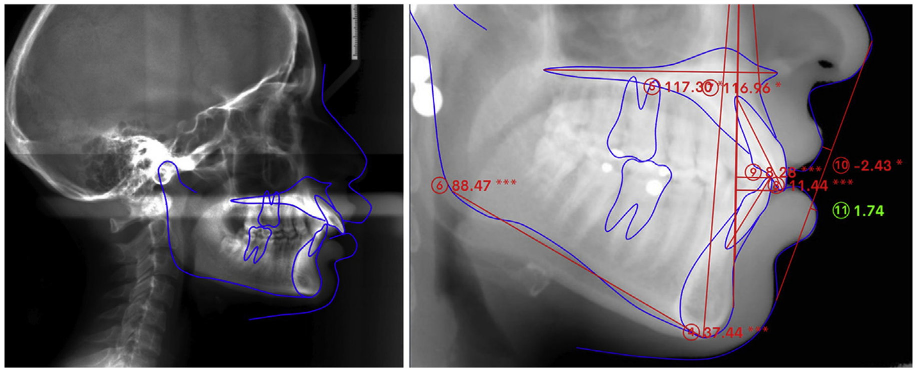


Figure 4: Two examples of poor definition of mandibular plane and soft tissue tracing with WebCeph

chosen for this study for purposes of comparison because it combines the features of manual and digital tracing. Unlike the automatic WebCeph, landmarks can be initially manually identified in AutoCAD, and then the cephalometric analysis can be completed by giving orders to calculate the measurements. The use of computer programs has been reportedly shown to reduce errors that may result from the manual drawing of lines and measuring with a ruler and protractor in the conventional method.[25] The current results agreed with the findings of the systematic review by **Leonardi** et al.[8] which concluded there was a lack of scientific evidence to support automatic landmark identification, as more errors resulted from it when compared with manual identification. However, in this study, even after using the semi-automatic option in WebCeph with manual landmark identification, about half of the results still showed statistically significant differences due to errors in calculating the measurements. On the other hand, most of these were of no clinical importance. Therefore, the use of the semi-automatic option may be considered a step towards enhancing the WebCeph outcomes.

Although the ICC revealed that the major discrepancy in the agreement between the automatic WebCeph and AutoCAD was only with the L1-MP angle, which did not show any level of agreement, the limit of agreement and regression analysis exposed different scenarios. The limit of agreement showed a visual problem with the Bland−Altman plot,[26] especially for SNB, U1-MaxP, L1-MP, U1-NPog, upper lip-E line, and lower lip-E line. Additionally, there was a significant proportional bias with the SNA, SNB, U1-MaxP, L1-MP, upper lip-E line variables. Several studies[6,18,24,27−30] have found adequate reliability between computer programs and manual methods, and the differences in these studies were slightly smaller than those in the current study. Besides, the individual behaviour of WebCeph with its limited consistency during the analysis could reduce its reliability as a substitute for other pre-tested programs. The semi-automatic WebCeph showed greater improvement in agreement with AutoCAD, but the problem of miscalculating the L1-MP remained. This could be attributed to the discrepancy in the mandibular plane identification.

To summarise, most of the discrepancies were found in the angular measurements. The present findings indicate that tracing with the automatic WebCeph led to obvious problems related to accuracy, such as inaccuracy in landmark identification (this was frequently observed where the points were identified either outside the bone or in the wrong location), inaccuracy of soft tissue outline tracing (where the tracing line was clearly drawn away from the soft tissue outline), and inadequate identification of the average of bilateral points (Figure 4). These are all vital issues that can directly influence the outcome of the analysis and were identified in all the radiographs measured in this study. The landmark identification problem can be overcome by using the semi-automatic WebCeph, but this did not completely solve the calculation errors.

AI algorithms have been shown to be capable of evaluating cephalometric images with high accuracy by different investigations,[6,7,9,10] but the WebCeph algorithm seems to be unable to do so and must be revised before it should be used clinically. On the other hand, several advantages were also noted: WebCeph is freely available in different languages; can be used with computers or smartphones (iOS and Android); is user-friendly, easily applied, and provides explanatory demonstrations; calibrates images and corrects magnification more quickly and easily compared to AutoCAD; allows for manual correction to digital tracing; provides a template for documenting patients' photos and radiographs; allows for the addition of new definitions for any unavailable landmark and the development of a custom analysis; and enables all the results to be easily exported and transferred for communication with others.

The main limitation of this study is that the reliability was assessed by only two investigators.

Conclusions and recommendations

Different problems, such as poor landmark identification/soft tissue tracing and inconsistency of measurements, are inherent to the automatic WebCeph. The use of the semi-automatic WebCeph can overcome some of the limitations of the automatic WebCeph. Therefore, unless it is developed further, WebCeph should only be used for cephalometric analysis with a great deal of caution accompanied by visual checks by a clinician.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not for-profit sectors.

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

The authors confirm that this study was prepared in accordance with COPE roles and regulations. Given the nature of the study, an IRB review was not required.

Authors' contributions

YAY conceived and designed the study, conducted research, provided research materials, and collected and organised data and was responsible for project administration, resources, software, supervision, validation, visualisation, and writing of the original draft of the article. **ARS** performed the data curation, formal analysis, and investigation and provided logistical support. **SAN** was responsible for project administration, supervision, and writing of the final draft of the article. All authors critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

References

1. Proffit WR, Fields HW, Larson BE, Sarver DM. *Contemporary orthodontics*. 6th ed. Philadelphia, PA: Elsevier; 2019.
2. Baumrind S, Miller DM. Computer-aided head film analysis: the University of California San Francisco method. **Am J Orthod 1980**; 78(1): 41–65. https://doi.org/10.1016/0002-9416(80)90039-1.
3. Forsyth DB, Shaw WC, Richmond S, Roberts CT. Digital imaging of cephalometric radiographs, Part 2: image quality. **Angle Orthod 1996**; 66(1): 43–50. https://doi.org/10.1043/0003-3219(1996)066<0043:DIOCRP>2.3.CO;2.
4. 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(1): 341–346. https://doi.org/10.13005/bpj/1645.
5. Jacobson A, Jacobson RL. *Radiographic cephalometry from basic to 3-D imaging*. 2nd ed. Quintessence Publishing co, Inc.; 2006.
6. Kunz F, Stellzig-Eisenhauer A, Zeman F, Boldt J. Evaluation of a fully automated cephalometric analysis using a customized convolutional neural network. **J Orofac Orthop 2020**; 81(1): 52–68. https://doi.org/10.1007/s00056-019-00203-8.
7. Arık SÖ, Ibragimov B, Xing L. Fully automated quantitative cephalometry using convolutional neural networks. **J Med Imaging 2017**; 4(1):014501. https://doi.org/10.1117/1.JMI.4.1.014501.
8. Leonardi R, Giordano D, Maiorana F, Spampinato C. Automatic cephalometric analysis. **Angle Orthod 2008**; 78(1): 145–151. https://doi.org/10.2319/120506-491.1.
9. Hwang H, Park J, Moon J, Yu Y, Kim H, Her S, et al. Automated identification of cephalometric landmarks: Part 2- Might it be better than human? **Angle Orthod 2020**; 90(1): 69–76. https://doi.org/10.2319/022019-129.1.
10. Kim H, Shim E, Park J, Kim Y, Lee U, Kim Y. Web-based fully automated cephalometric analysis by deep learning. **Comput Methods Programs Biomed 2020**; 194:105513. https://doi.org/10.1016/j.cmpb.2020.105513.
11. Hung K, Montalvao C, Tanaka R, Kawai T, Bornstein MM. The use and performance of artificial intelligence applications in dental and maxillofacial radiology: a systematic review. **Dentomaxillofac Radiol 2020**; 49(1):20190107. https://doi.org/10.1259/dmfr.20190107.
12. Nahidh M, Al-Jarad AF, Aziz ZA. The reliability of AutoCAD program in cephalometric analysis in comparison with pre-programmed cephalometric analysis software. **Iraqi Dent J 2012**; 34(1): 35–40. https://doi.org/10.26477/idj.v34i1.172.
13. Piqué-Vidal C, Maled-García I, Arabi-Moreno J, Vila J. Radiographic angles in hallux valgus: differences between measurements made manually and with a computerized program. **Foot Ankle Int 2006**; 27(3): 175–180. https://doi.org/10.1177/107110070602700304.
14. Nasseri N, Hadian MR, Bagheri H, Talebian S, Olyaei G. Reliability and accuracy of joint position sense measurement in laboratory and clinic; utilising a new system. **Acta Med Iran 2007**; 45(5): 395–404.
15. Quieregatto PR, Hochman B, Furtado F, Machado AFP, Neto MS, Ferreira LM. Image analysis software versus direct anthropometry for breast measurements. **Acta Cir Bras 2014**; 29(10): 688–695. https://doi.org/10.1590/s0102-86502014001600010.
16. Celik E, Polat-Ozsoy O, Toygar Memikoglu TU. Comparison of cephalometric measurements with digital versus conventional cephalometric analysis. **Eur J Orthod 2009**; 31(3): 241–246. https://doi.org/10.1093/ejo/cjn105.
17. Polat-Ozsoy O, Gokcelik A, Toygar Memikoglu TU. Differences in cephalometric measurements: a comparison of digital versus hand-tracing methods. **Eur J Orthod 2009**; 31(3): 254–259. https://doi.org/10.1093/ejo/cjn121.
18. Goracci C, Ferrari M. Reproducibility of measurements in tablet-assisted, PC-aided, and manual cephalometric analysis. **Angle Orthod 2014**; 84(3): 437–442. https://doi.org/10.2319/061513-451.1.
19. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. New Jersey: Prentice Hall; 2000.
20. 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(2): 155–161. https://doi.org/10.1043/0003-3219(2004)074<0155:TEODIL>2.0.CO;2.
21. Akhare PJ, Dagab AM, Alle RS, Shenoyd U, Garla V. Comparison of landmark identification and linear and angular measurements in conventional and digital cephalometry. **Int J Comput Dent 2013**; 16(3): 241–254.
22. Livas C, Dellib K, Spijkervetc FKL, Vissinkd A, Dijkstrae PU. Concurrent validity and reliability of cephalometric analysis using smartphone apps and computer software. **Angle Orthod 2019**; 89(6): 889–896. https://doi.org/10.2319/021919-124.1.
23. Aksakallı S, Yılancı H, Görükmez E, Ramoğlu SI. Reliability assessment of orthodontic apps for cephalometrics. **Turk J Orthod 2016**; 29(4): 98–102. https://doi.org/10.5152/TurkJOrthod.2016.1618.
24. Erkan M, Gurel HG, Nur M, Demirel B. Reliability of four different computerized cephalometric analysis programs. **Eur J Orthod 2011**; 34(3): 318–321. https://doi.org/10.1093/ejo/cjr008.
25. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. **Am J Orthod Dentofacial Orthop 2000**; 118(5): 535–540. https://doi.org/10.1067/mod.2000.110168.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. **Lancet 1986**; 1(8476): 307–310.
27. Al-Nasseri NA. The accuracy of computerised cephalometric analysis compared to conventional manual method. **Mustansiria Dent J 2005**; 2(1): 7–14.

28. Uthman AT, Al-Sahaf NH. Validity and reliability of cephalometric measurements in traditional versus digitized cephalometry. **Mustansiria Dent J 2006**; 3(3): 233—238.
29. de Araújo Guedes P, de Souza JEN, Tuji FM, Nery EM. A comparative study of manual vs. computerized cephalometric analysis. **Dental Press J Orthod 2010**; 15(2): 44—51. https://doi.org/10.1590/S2176-94512010000200007.
30. de Abreu DP, Freitas KMS, Nomura S, Valarelli FP, Cançado RH. Comparison among manual and computerized cephalometrics using the softwares dolphin imaging and dentofacial planner. **Dent Oral Craniofac Res 2016**; 2(6): 1—5. https://doi.org/10.15761/DOCR.1000186.

How to cite this article: Yassir YA, Salman AR, Nabbat SA. The accuracy and reliability of WebCeph for cephalometric analysis. J Taibah Univ Med Sc 2022;17(1):57—66.