Review

Accuracy and reliability of Keynote for tracing and analyzing cephalometric radiographs

Ali K. Hamad[a]; Ferdinand M. Machibya[b]; Matilda M. Mlangwa[c]; David N. Ngassapa[d]

ABSTRACT

Objectives: To evaluate the reliability and accuracy of Keynote for tracing and analyzing cephalograms in comparison to Quick Ceph Studio.

Materials and Methods: This was a cross-sectional study, which utilized the lateral cephalometric digital images (radiographs) from 49 patients. The study site was the Dental Radiology unit in the School of Dentistry of the Muhimbili University of Health and Allied Sciences (MUHAS), in Dar es Salaam, Tanzania. Cephalograms were imported to Quick Ceph Studio and then to Keynote for analysis. Minimum, maximum, mean, standard deviation, and mean difference were used to describe the data. Agreement between the two techniques was assessed by the Bland-Altman plot, linear regression, and interexaminer reliability tests. A level of significance was considered at $P < .05$, and a 95% CI was estimated for the outcomes in the study groups.

Results: The majority of the mean values obtained from Quick Ceph were greater ($P < .05$) than those obtained from Keynote. According to Bland-Altman plots, all measurements were within the limit of agreement except for only five linear variables. The interexaminer reliability test showed no agreement between the two instruments for all linear parameters except for the LAFH: TAFH, whereas all angular measurements revealed good to excellent agreement (ICC: 0.75 to 0.97) between the methods.

Conclusions: The measurements obtained with the Keynote software were found to be clinically reliable since the limits did not exceed the maximum acceptable difference between the methods. The two software instruments were considered to be in agreement and can be used interchangeably. (*Angle Orthod*. 2025;00:000–000.)

KEY WORDS: Keynote; Quick Ceph; Orthodontics; Cephalometric; Reliability; Tracing

INTRODUCTION

Currently, the analysis of cephalometric radiographs is commonly performed by a computer-assisted method that may involve either manual or auto-identification of cephalometric landmark points on the monitor.[1] Previous literature reported that, as long as the landmark points are identified manually, a computerized cephalometric analysis does not induce more measurement error than the traditional tracing method.[2,3] Additionally, the use of computers is expected to minimize any error caused by operator fatigue and provide effective evaluation with a high rate of reproducibility.[4] The commonly used preprogrammed cephalometric analysis software includes Quick Ceph, Dolphin Imaging, Nemoceph, and Vistadent.[5] However, availability and affordability of these commercially available software remains questionable.[6] This is due to the fact that some of them seem to be very expensive for many clinicians.

Keynote is a free software program that can be a cost-effective way to perform analysis of cephalograms. Notably, Keynote is a program developed by Apple for presenting visual data. Additionally, it was reported that Keynote is an alternative for freely available software that can be more cost-effective for performing cephalometric analysis.[7] Therefore, the

[a] Assistant Lecturer, Department of Anatomy, School of Biomedical Sciences, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[b] Senior Lecturer and Dean, School of Dentistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[c] Senior Lecturer, Department of Orthodontics, Paedodontics and Community Dentistry, School of Dentistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

[d] Professor, Department of Anatomy, School of Biomedical Sciences, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania.

Corresponding author: Dr Ali K. Hamad, Department of Anatomy, School of Biomedical Sciences, Muhimbili University of Health and Allied Sciences, United Nations Rd, Dar es Salaam, Tanzania
(e-mail: habibkham@yahoo.com)

null hypothesis of this study was that there would be no difference between the analysis performed by Quick Ceph and that performed by Keynote. However, in clinical orthodontics, effectiveness of the presented cephalometric tracing software needs to be assessed for its accuracy to allow clinicians to select appropriate methods and tools for analysis.[8] For these reasons, this study aimed to evaluate accuracy and reliability of Keynote software for tracing and analyzing cephalometric radiographs compared to Quick Ceph Studio.

MATERIALS AND METHODS

This cross-sectional study was approved (MUHAS-REC-05-2023-1654) by the Research and Publications Committee of the Muhimbili University Senate. The digital images of lateral cephalograms of 49 patients (24 females and 25 males) were obtained from the Dental Radiology unit of the School of Dentistry (MUHAS). The images were taken using cone beam computed tomography (CBCT) (X-VIEW 3D Pan Ceph, Trident S.r.l, Italy) according to the standard regulation of radiation[9] of Tanzania. The cephalograms were selected using a systematic randomization method. They were then categorized based on gender. To reduce random error, the following exclusion criteria were used: craniofacial abnormality, missing incisors, presence of impacted or unerupted teeth, and a radiograph with poor quality. The inclusion criteria were: lateral cephalogram of patients with no crowding, presence of all teeth (third molars may or may not be present), and no history of orthodontic treatment. The cephalograms were imported to Quick Ceph Studio, (Version 5.2.6, Quick Ceph Systems, Inc., FL, USA) and then to Keynote for macOS, (version 14.1, Apple Inc., USA) for analysis (Figure 1). The magnification corrections for each option were initially undertaken based on the distance of 10 mm between two fixed points on the Cephalostat rod in the Cephalogram. Thirteen anatomical landmarks were selected and manually identified using a cursor, followed by calculations of 10 angular and 6 linear measurements using both software applications. All measurements were taken by a single operator daily, with a maximum of 10 cephalograms per day (using Quick Ceph first). After an interval of a minimum of 1 week, the same images were remeasured using the Keynote application.

The Kolmogorov-Smirnov test was used to review normality of the data distribution. Intra- and inter-examiner reliability for each measurement was assessed using the intraclass correlation coefficient (ICC). The minimum, maximum, mean, standard deviation, and mean difference were used to describe the data. A systematic bias between Quick Ceph and Keynote software was assessed by paired t-test. A Bland-Altman plot,[10] interexaminer reliability, and linear regression tests were applied for assessment of the agreement between the two techniques of measurement. Clinically relevant differences were considered when the difference was greater than 2° or 2 mm for the angular and linear measurements, respectively.[11] A statistical significance level of $P < .05$ was considered and a 95% confidence interval was estimated for the outcomes in the study group. Data analysis was conducted using RStudio Desktop for macOS 12+, (Posit Software, Boston, Mass, USA).

RESULTS

There were 49 cephalograms analyzed including 24 from males and 25 from females. Based on the skeletal pattern distribution, there were 29 Class I, 12 Class II, and 8 Class III skeletal patterns. The ICC demonstrated that the intra-examiner reliability was very good to excellent (0.86 to 0.99) and moderate to excellent (0.73 to 0.99) for the measurements by Quick Ceph and Keynote software, respectively (Table 1). The maximum differences were 2.56° (interincisal angle) and 15.52 mm (TAFH) for the angular and linear measurements, respectively (Table 2). It was also noted that the majority of the mean values obtained from Quick Ceph were greater than those obtained from Keynote. However, a paired t-test showed that there was a significant difference ($P < .05$) in all parameters except for only five angular variables, namely the Saddle Angle, SNA, SNB, ANB, and the FMIA when compared between the two methods (Table 3). Additionally, linear regression analysis revealed a significant ($P < .05$) proportional bias between the two methods.

According to the Bland-Altman plots (Figures 2 and 3), all measurements were within the limit of agreement as the bias analysis was close to zero except for five linear variables: anterior cranial base, posterior cranial base, ramus height, LAFH, and TAFH. The mean differences of these mentioned parameters drifted away from zero, indicating a systematic bias between the two methods. To confirm this, the interexaminer reliability test (Table 4) showed no agreement between the two instruments for all linear parameters except for LAFH: TAFH, whereas all angular measurements revealed good to excellent agreement between the two approaches (ICC: 0.75 to 0.97). In addition, trends for a greater number of data points randomly distributed around the mean difference lines were also observed, suggesting good agreement (Figures 2 and 3).

Based on the clinically relevant difference, the error size for all measurements was within the acceptable range, and the limits for most variables did not exceed the maximum acceptable difference (2° and 2 mm, for angular and linear measurements, respectively) between
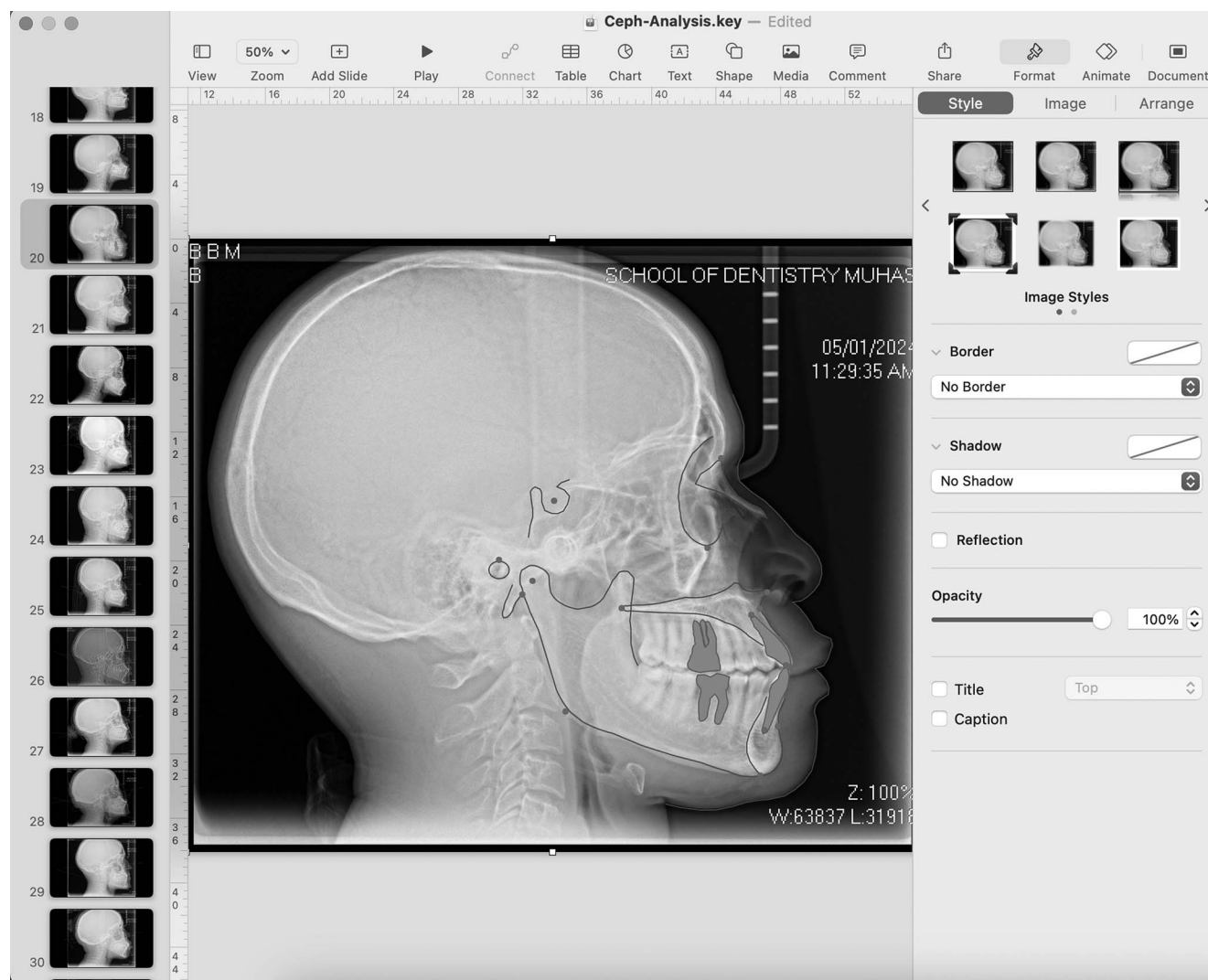
Figure 1. Cephalometric analysis using Keynote software.

the methods. All measurements that revealed a significant difference in the paired *t*-test were also within the limit of agreement.

DISCUSSION

The use of computerized cephalometric analysis techniques has been reportedly shown to minimize errors resulting from manual drawing of lines and measuring with a ruler and protractor in the conventional method.[2] The current study aimed to evaluate the reliability and accuracy of Keynote software for tracing and analyzing cephalograms in comparison to Quick Ceph software. Quick Ceph was used in the current study as a standard tool because it had been shown in previous studies[12,13] to produce adequate angular and linear measurements. Keynote software, on the other hand, has recently been introduced as pre-installed software and is freely available on all

Apple computers as a cost-effective alternative method for digital cephalometric analysis.[7] This was the first study that analyzed its performance and verified its accuracy for clinical use.

The study involved several sagittal and vertical skeletal patterns, ensuring the existence of any potential variation in the vertical and anteroposterior relationship of the jaws that could be faced when performing cephalometric analysis. Each software application has achieved sufficient reliability when tested at different intervals. As intra-examiner errors are less frequent than interexaminer,[3] the measurements were undertaken by the same investigator to avoid any possible error between operators and to achieve a required standard.

To analyze agreement between the methods, various studies have used the Pearson correlation coefficient (measuring the association instead of agreement); however, this statistical technique can

Table 1. Intraclass Correlation Coefficient (ICC) for Quick Ceph and Keynote Measurements

| | Quick Ceph | | | Keynote | | |
| | | 95% CI | | | 95% CI | |
Variables	ICC	Lower	Upper	ICC	Lower	Upper
Angular measurement						
Saddle angle	0.96	0.75	0.99	0.85	0.52	0.96
Articular angle	0.97	0.83	0.99	0.93	0.76	0.98
Gonion angle	0.86	0.55	0.96	0.75	0.31	0.93
SNA	0.93	0.77	0.97	0.79	0.38	0.94
SNB	0.96	0.87	0.98	0.94	0.78	0.98
ANB	0.94	0.80	0.98	0.98	0.92	0.98
FMA	0.95	0.83	0.98	0.73	0.26	0.93
FMIA	0.98	0.91	0.98	0.94	0.79	0.98
IMPA	0.97	0.89	0.99	0.99	0.96	1.00
Interincisal	0.99	0.97	1.00	0.97	0.95	1.00
Linear measurement						
Anterior cranial base	0.86	0.56	0.96	0.85	0.37	0.96
Posterior cranial base	0.98	0.93	0.99	0.95	0.80	0.99
Ramus height	0.98	0.94	1.00	0.87	0.45	0.97
LAFH	0.99	0.95	1.00	0.92	0.72	0.98
TAFH	0.98	0.61	1.00	0.90	0.67	0.97
LAFH: TAFH	0.89	0.59	0.97	0.81	0.39	0.95

Table 3. Paired *t*-test Assessing Systematic Bias Between Quick Ceph and Keynote

| | | 95% CI of the Difference | | | |
Variable	Mean Difference	Lower	Upper	t	P Value
Angular measurement					
Saddle angle	0.06	−0.98	1.09	0.11	.915
Articular angle	2.32	1.05	3.59	3.66	<.001
Gonion angle	−2.35	−3.32	−1.38	−4.89	<.001
SNA	0.04	−0.32	0.39	0.23	.819
SNB	−0.03	−0.29	0.24	−0.21	.832
ANB	0.11	−0.18	0.40	0.74	.461
FMA	−1.61	−2.27	−0.95	−4.88	<.001
FMIA	−0.68	−1.77	0.41	−1.26	.214
IMPA	2.36	1.14	3.58	3.89	<.001
Interincisal	2.56	1.13	3.99	3.59	<.001
Linear measurements					
Anterior cranial base	8.47	7.85	9.08	27.84	<.001
Posterior cranial base	4.27	3.11	5.43	7.39	<.001
Ramus height	13.19	11.99	14.41	4.07	<.001
LAFH	11.19	10.49	11.89	32.22	<.001
TAFH	15.52	14.47	16.58	29.59	<.001
LAFH: TAFH	1.70	1.38	2.02	10.75	<.001

be misleading and inappropriate.[14] Therefore, the dataset was analyzed by applying a graphical technique with an appropriate use of regression to determine 95% limits of agreement and confidence intervals as well as to quantify the disagreement between two measurement techniques.[15,16] According to the Bland-Altman plots in the present study, all measurements were in acceptable agreement. Even the variables that revealed a significant difference in the paired *t*-test (Table 3) were also within acceptable limits (Figures 2 and 3) since the decision on what was acceptable

agreement was a predetermined clinical judgment. Thus, although there was a significant difference in some of the parameters and wide limits of agreement in the Bland-Altman plots between the instruments, clinically, the analysis can be carried out with both software programs.

The interchangeability of Keynote and Quick Ceph cannot be generalized across all parameters. Although angular measurements such as ANB fell within clinically acceptable ranges of agreement, particularly those involving facial height or cranial base dimensions demonstrated

Table 2. Descriptive Statistics for All Measurements Used in the Study[a]

| | | Quick Ceph | | | Keynote | | |
Variable	n	Mini.	Maxi.	Mean ± SD	Mini.	Maxi.	Mean ± SD
Angular measurement							
Saddle angle	49	107.2	140.2	123.28 ± 6.53	108.0	134.0	123.22 ± 6.44
Articular angle	49	130.1	168.8	150.81 ± 8.37	134.0	168.0	148.49 ± 7.87
Gonion angle	49	104.3	133.1	117.59 ± 6.76	107.0	139.0	119.94 ± 6.79
SNA	49	79.6	96.8	87.51 ± 4.02	78.0	97.0	87.47 ± 4.13
SNB	49	74.9	91.4	83.09 ± 4.03	75.0	91.0	83.12 ± 4.17
ANB	49	−1.0	8.7	4.41 ± 2.49	−1.0	11.0	4.31 ± 2.38
FMA	49	11.8	33.6	23.53 ± 5.05	12.0	34.0	25.14 ± 4.58
FMIA	49	39.7	79.3	55.42 ± 7.43	40.0	72.0	56.10 ± 6.92
IMPA	49	83.4	116.4	101.06 ± 6.97	84.0	112.0	98.69 ± 6.39
Interincisal	49	100.1	136.4	116.44 ± 9.39	96.0	133.0	113.88 ± 9.50
Linear measurement							
Anterior cranial base	49	66.0	81.5	72.28 ± 3.38	57.0	72.0	63.82 ± 3.14
Posterior cranial base	49	26.6	45.2	37.21 ± 4.15	24.0	52.0	32.94 ± 4.75
Ramus height	49	34.2	61.4	50.46 ± 6.07	28.0	49.0	37.27 ± 4.68
LAFH	49	63.3	89.7	75.97 ± 6.12	53.0	78.0	64.78 ± 5.36
TAFH	49	110.2	145.3	127.24 ± 7.29	7.41	97.0	111.71 ± 6.96
LAFH: TAFH	49	55.8	65.1	59.64 ± 2.26	54.2	63.9	57.94 ± 2.40

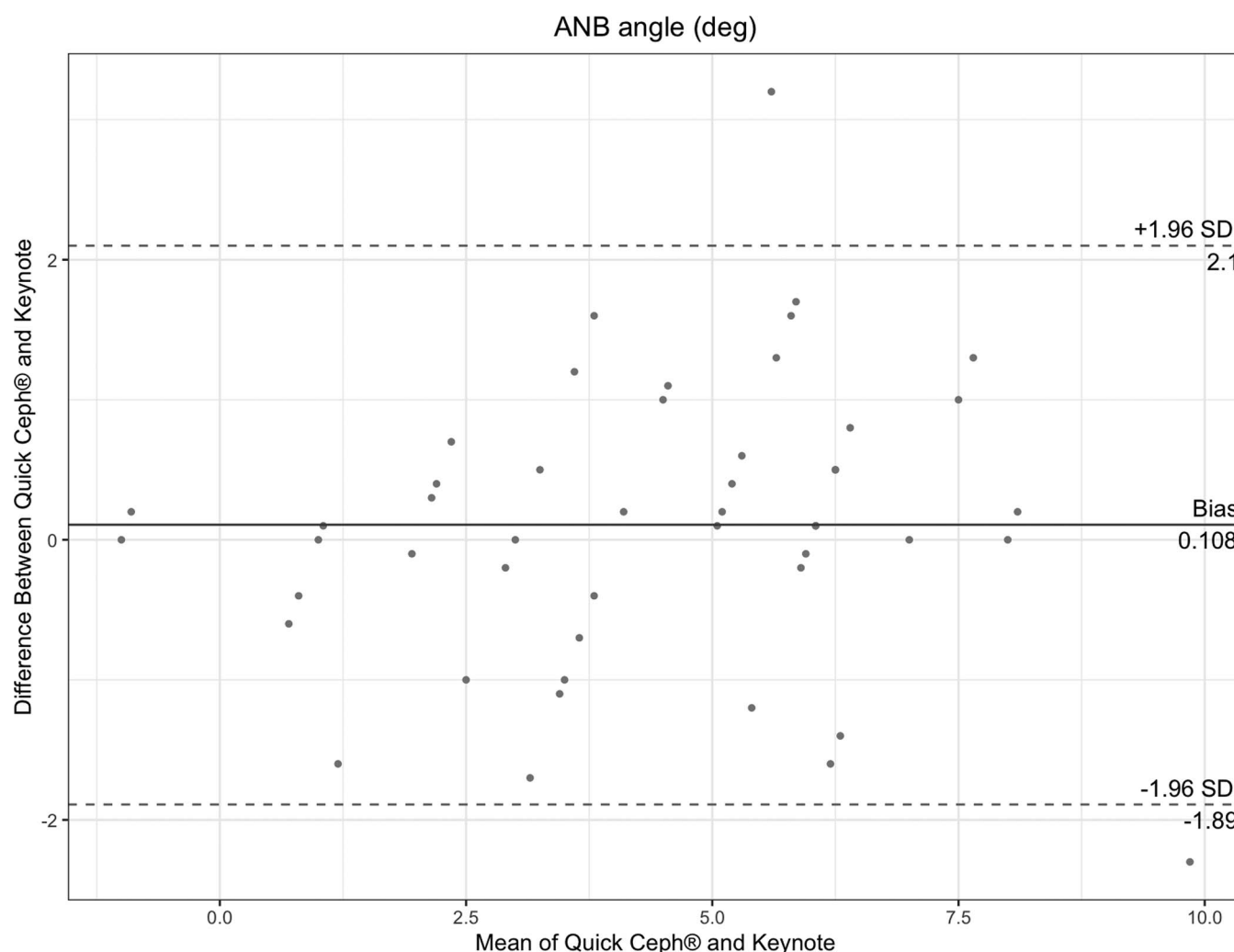[a] Mini indicates minimum value; Maxi, maximum value; SD, standard deviation.

Figure 2. Bland-Altman plots presenting angular variables in each Quick Ceph and Keynote method.

proportional bias that limits their direct comparability. Clinicians should exercise caution and validate critical measurements manually when using Keynote for precision-sensitive applications. The present findings concurred with previous literature which concluded that the differences between measurements derived from the cephalometric radiographs with the two different digitized methods were statistically significant but clinically acceptable.[3] Additionally, previous literature reported that the findings obtained from the same patient may vary as a function of the cephalometric measurement approach used.[17]

Although the ICC (interexaminer reliability) showed that the major discrepancies in agreement between Quick Ceph and Keynote software were only with the linear measurements, especially for posterior cranial base, ramus height, anterior cranial base, LAFH, and TAFH, which showed poor level of agreement, a linear regression analysis showed a different scenario. The

regression exhibited a significant proportional bias between the two techniques for both angular and linear variables. The lower reliability of some measurements could have been due to difficulty in the identification of some landmark points. For instance, some studies indicated that identification of Porion, Orbitale, Articulare, and Gonion points on lateral cephalograms can be challenging,[18,19] and any measurement based on the Frankfort horizontal plane might be erroneous;[20] this justifies the lower reliability of the linear measurements in the present study. Similar findings for linear values were reported by Kumar et al.[6] using Nemoceph and Foxit PDF Reader, and Celik et al.[8] using Vistadent software vs the Jiffy orthodontic evaluation program.

The accuracy of software-based cephalometric analysis has been extensively evaluated in previous studies, aiming to improve efficiency and precision compared to traditional manual methods.[4] Automated
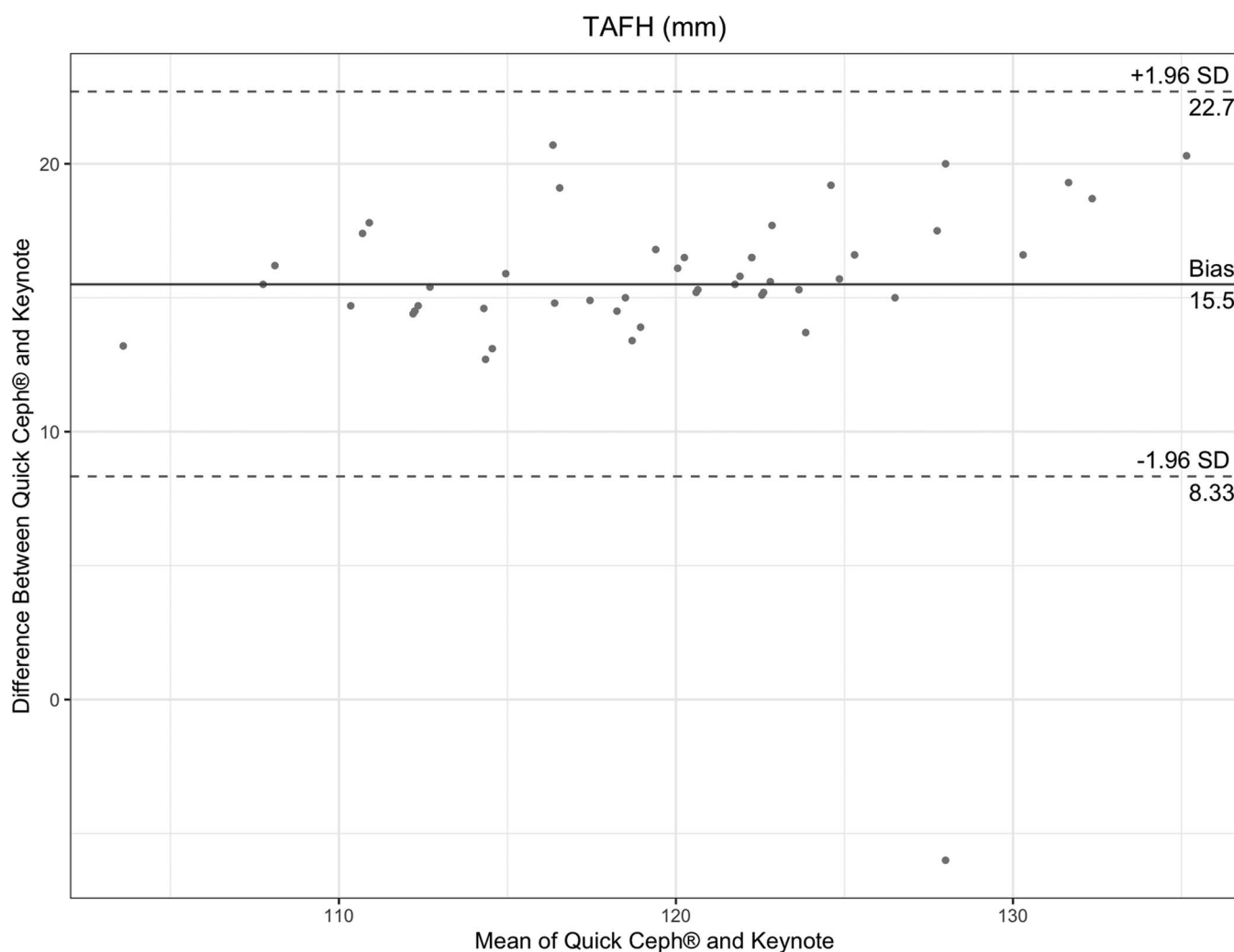
Figure 3. Bland-Altman plots for the linear variables in each Quick Ceph and Keynote method.

and semiautomated software tools, including the one assessed in this study, have demonstrated comparable reliability to manual methods for most parameters,[21] particularly angular measurements such as ANB. However, discrepancies are often observed in linear measurements, likely due to variances in landmark identification.[11] These findings were in agreement with those of the current study, indicating that, while Keynote is a cost-effective alternative, its accuracy must be carefully evaluated, particularly for parameters requiring high precision.

Modern cephalometric software tools, while showing significant proportional biases in some angular and linear parameters, typically provide measurements within clinically acceptable ranges. This makes them reliable alternatives to manual methods for routine orthodontic and surgical assessments. However, clinicians should be cautious of systematic bias in specific linear measurements, particularly in cases requiring high precision, such as craniofacial anomaly assessments or detailed growth

monitoring.[22] Literature suggests that such biases may arise from differences in how software tools and manual methods interpret landmarks, particularly in complex or ambiguous anatomical regions.[23] For example, cranial base measurements are sensitive to errors in landmark identification due to overlapping structures, whereas facial height parameters rely on clear differentiation of anatomical boundaries, which software tools may not always accurately identify.[24] Understanding the acceptable range of error for each parameter is vital, emphasizing the need for software tools to be calibrated and validated against manual methods with clinically defined tolerances for meaningful interpretation.[25]

Although manual cephalometric tracing is often considered the gold standard due to the ability of experienced clinicians to adjust for individual anatomical variations or radiographic artifacts,[26] it remains time-consuming and susceptible to inter- and intra-operator variability.[6,24] Studies have shown that most software tools provide reliable measurements for both linear

Table 4. Interexaminer Reliability for the Measurements Between Quick Ceph and Keynote

Variable	ICC	95% Confidence Interval		P Value
		Lower Bound	Upper Bound	
Angular measurement				
Saddle angle	0.85	0.75	0.91	****
Articular angle	0.82	0.64	0.91	****
Gonion angle	0.83	0.55	0.92	****
SNA	0.95	0.92	0.97	****
SNB	0.97	0.96	0.99	****
ANB	0.91	0.85	0.95	****
FMA	0.84	0.57	0.93	****
FMIA	0.86	0.76	0.92	****
IMPA	0.75	0.51	0.87	****
Interincisal	0.83	0.66	0.91	****
Linear measurement				****
Anterior cranial base	0.18	−0.03	0.53	****
Posterior cranial base	0.41	−0.05	0.69	****
Ramus height	0.18	−0.04	0.51	****
LAFH	0.32	−0.02	0.70	****
TAFH	0.26	−0.03	0.64	****
LAFH: TAFH	0.70	−0.06	0.90	****

**** $P = .0001$.

and angular parameters, though their accuracy may vary depending on the clarity of landmarks and the precision of the algorithms.[27] For instance, angular measurements such as SNA and SNB show high consistency due to reliance on well-defined craniofacial landmarks. Conversely, discrepancies in parameters like the Saddle angle and FMIA highlight the susceptibility of certain variables to software-dependent errors, likely arising from differences in landmark identification or scaling.[28]

Overall, these findings underscore the importance of understanding the limitations of cephalometric software tools and their alignment with clinical needs. Although these tools can significantly enhance efficiency and consistency, ensuring their accuracy and addressing systematic biases in critical measurements are essential for safe and effective clinical application.

Limitations

The reliability of the methods was evaluated by a single investigator. To mitigate this limitation, additional reliability assessments using multiple investigators are planned in future studies. This would help ensure that the results are not influenced by individual bias and provide a more robust evaluation of the methods.

CONCLUSIONS

• The findings of the present study exposed some areas where the two software applications were

inconsistent in the analysis, particularly in terms of linear measurements and systematic bias. Consequently, the null hypothesis stating that there is no statistically significant difference between the analysis carried out by Quick Ceph and Keynote can be rejected.
• Measurements obtained with Keynote software used in the current study were shown to be clinically reliable.
• Since the limits did not exceed the maximum acceptable difference between methods, the two software programs are considered to be in agreement and can be used interchangeably.

REFERENCES

1. Jeon S, Lee K. Comparison of cephalometric measurements between conventional and automatic cephalometric analysis using convolutional neural network. *Prog Orthod.* 2021;22:14.
2. Liu JK, Chen YT, Cheng KS. Accuracy of computerized automatic identification of cephalometric landmarks. *Am J Orthod Dentofacial Orthop*. 2000;118(5):535–540.
3. Erkan M, Gurel HG, Nur M, Demirel B. Reliability of four different computerized cephalometric analysis programs. *Eur J Orthod.* 2012;34(3):318–321.
4. Sangroula P, Sardana HK, Kharbanda OP, Duggal R. Comparison of reliability and validity of posteroanterior cephalometric measurements obtained from AutoCEPH© and dolphin® cephalometric software programs with manual tracing. *J Indian Orthod Soc*. 2018;52:106–114.
5. Paul PL, Tania SD, Rathore S, Missier S, Shaga B. Comparison of accuracy and reliability of automated tracing Android app with conventional and semiautomated computer aided tracing software for cephalometric analysis – a cross-sectional study. *Int J Orthod Rehabil*. 2022;13(4):39–51.
6. Kumar M, Kumari S, Shetty P, Kumar R, Shetty P. Comparative evaluation of Nemoceph and Foxit PDF Reader for Steiner's cephalometric analysis. *J Contemp Dent Pract.* 2019;20(9):1051–1055.
7. Shahrul A. Technique for cephalometric analysis using Keynote. *J Indian Orthod Soc*. 2022;56(3):299–301.
8. 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.
9. Abu-Tayyem H, Alshamsi A, Quadri M. Soft tissue cephalometric norms in Emirati population: a cross-sectional study. *J Multidiscip Healthc*. 2021;14:2863–2869.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307–310.
11. Livas C, Delli K, Spijkervet F, Vissink A, Dijkstra P. Concurrent validity and reliability of cephalometric analysis using smartphone apps and computer software. *Angle Orthod.* 2019;89(6):889–896.
12. Oz AZ, Akcan CA, Ciger S. Evaluation of the soft tissue treatment simulation module of a computerized cephalometric program. *Eur J Dent*. 2014;8:229–233.
13. Takahashi K, Shimamura Y, Tachiki C, Nishii Y, Hagiwara M. Cephalometric landmark detection without X-rays combining coordinate regression and heatmap regression. *Sci Rep*. 2023;13:20011.

14. Grilo LM, Grilo HL. Comparison of clinical data based on limits of agreement. *Biom Letters*. 2012;49(1):45–56.

15. Bland J, Altman D. Agreement between methods of measurement with multiple observations per individual. *J Biopharm Stat.* 2007;17:571–582.

16. Giavarina D. Understanding Bland Altman analysis. *Biochem Med*. 2015;25(2):141–151.

17. Gómez-Medina IP, Aguilar-Pérez DA, Escoffié-Ramírez M, Herrera-Atoche JR, Aguilar-Pérez FJ. Evaluation of diagnostic agreement among cephalometric measurements for determining incisor position and inclination. *Int J Morphol*. 2020;38(5):1386–1391.

18. Santoro M, Jarjoura K, Cangialosi TJ. Accuracy of digital and analogue cephalometric measurements assessed with the sandwich technique. *Am J Orthod Dentofacial Orthop*. 2006;129(3):345–351.

19. Ganna PS, Shetty SK, Yethadka MK, Ansari A. An evaluation of the errors in cephalometric measurements on scanned lateral cephalometric images using computerize program and conventional tracing. *J Indian Orthod Soc*. 2014;48(4):388–392.

20. Khosravani S, Esmaeili S, Mohammadi NM, Eslamian L, Motamedian SR. Inter and intra-rater reliability of lateral cephalometric analysis using 2D Dolphin Imaging Software. *Journal of Dental School*; 2020;38(4):148–152.

21. Subramanian AK, Chen Y, Almalki A, Sivamurthy G, Kafle D. Cephalometric analysis in orthodontics using artificial intelligence-a comprehensive review. *Biomed Res Int.* 2022; 16:1880113.

22. Gribel BF, Gribel MN, Frazão DC, McNamara JA Jr, Manzi FR. Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans. *Angle Orthod*. 2011;81(1):26–35.

23. Narkhede S, Rao P, Sawant V, et al. Digital versus manual tracing in cephalometric analysis: a systematic review and meta-analysis. *J Pers Med*. 2024;14(6):566.

24. Smołka P, Nelke K, Struzik N, et al. Discrepancies in cephalometric analysis results between orthodontists and radiologists and artificial intelligence: a systematic review. *Appl Sci.* 2024;14(12):4972.

25. Kumar R, Choudhary RK, Archana, et al. Calibration of medical devices: method and impact on operation quality. *Int Pharm Sci*. 2023;16(1):128.

26. Gorla LFO, Dos Santos JC, Carvalho PHA, Hochuli-Vieira E, Gabrielli MAC. Accuracy of manual and virtual predictive tracings in patients submitted to orthognathic surgery. *J Craniofac Surg.* 2023;34(4):1165–1169.

27. Narkhede S, Rao P, Sawant V, et al. Digital versus manual tracing in cephalometric analysis: a systematic review and meta-analysis. *J Pers Med.* 2024;14(6):566.

28. Turner PJ, Weerakone S. An evaluation of the reproducibility of landmark identification using scanned cephalometric images. *J Orthod*. 2001;28(3):221–229.