Y. Kaihara*, A. Katayama**, K. Ono**,
M. Kurose**, K. Toma**, H. Amano***,
H. Nikawa**, K. Kozai****

Hiroshima University, Hiroshima, Japan
**Department of Paediatric Dentistry, Hiroshima University Hospital*
***Department of Oral Biology and Engineering,*
Division of Oral Health Sciences,
Graduate School of Biomedical Sciences
****Maxillofacial Functional Development Division of Oral Health Sciences, Graduate School of Biomedical Sciences*
*****Department of Paediatric Dentistry,*
Division of Cervico-Gnathostomatology
Graduate School of Biomedical Sciences

e-mail: kaihara@hiroshima-u.ac.jp

Comparative analyses of paediatric dental measurements using plaster and three-dimensional digital models

ABSTRACT

Aim *Recent advances in three-dimensional imaging have led to an increased interest in the application of computer-models in paediatric dentistry. However, in evidence-based paediatric dentistry the accuracy of new methods must be validated before they are introduced to clinical practice. We aimed to compare the accuracy of measurements of digital models obtained using a non-contact 3D measuring system, with direct measurements made on plaster models (gold standard) from children.*
Materials and methods *Twelve pairs of plaster models were obtained from children with deciduous dentition; tooth size, arch width, and arch length were examined. The same parts on each cast were measured twice with at least a 2-week interval between measurements with each method by four examiners. Linear mixed-effects model analyses were performed for comparison of values from the 2 different measurement methods.*
Results *The average difference between the 2 methods in measured values, derived from the final model, was <0.2 mm. Random effect of examiners was always the smallest component of variance, and frequently negligible. Statistics: Intraclass correlation coefficients were typically >90%.*
Conclusion *These results suggest that primary dentition analysis of digital models has a high accuracy level, comparable to that of direct measurement of plaster models by digital calipers.*

Keywords *Accuracy; Non-contact 3D laser scanner; Study model analysis.*

Introduction

Guidance of eruption and development of dentition and occlusion are important paediatric dental interventions. Early diagnosis and successful treatment during childhood contribute not only to the prevention of the development of malocclusions but also to the promotion of the normal development of dental arches and jaw function. In order to predict occlusal development, it is necessary to observe patient dentition and occlusion from the primary dentition period. Study model analysis is an essential component of the assessment of occlusion, diagnosis, treatment planning, and evaluation. Traditionally, the gold standard for obtaining diagnostic measurements involves the use of plaster models. The most important measurements in the evaluation of paediatric patients' dental arches are linear measurements of tooth size, arch width, and arch length. However, in recent years, more accurate digital images of study models (digital models) have been obtained as a result of significant advances in the application of non-contact three-dimensional (3D) measuring systems [Abizadeh et al., 2012; Asquith et al., 2007; Bell et al., 2003; Bootvong et al., 2010; Dalstra and Melsen, 2009; Fleming et al., 2011; Leifert et al., 2009; Okunami et al., 2007; Park et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Rosseto et al., 2009; Santoro et al., 2003; Veenema et al., 2009; Watanabe-Kanno et al., 2009; Zilberman et al., 2003]. Digital models have gained attention for their use in virtual model analysis and treatment simulation [Abizadeh et al., 2012; Asquith et al., 2007; Bell et al., 2003; Bootvong et al., 2010; Dalstra and Melsen, 2009; Fleming et al., 2011; Leifert et al., 2009; Okunami et al., 2007; Park et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Rosseto et al., 2009; Santoro et al., 2003; Veenema et al., 2009; Watanabe-Kanno et al., 2009; Zilberman et al., 2003]. Furthermore, digital models have the advantages of being easy to access [Abizadeh et al., 2012; Bell et al., 2003; Leifert et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Santoro et al., 2003], easy to consult [Dalstra and Melsen, 2009; Leifert et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Rosseto et al., 2009; Veenema et al., 2009; Zilberman et al., 2003], and have minimal storage space requirements [Abizadeh et al., 2012; Bell et al., 2003; Bootvong et al., 2010; Dalstra and Melsen, 2009; Fleming et al., 2011; Leifert et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Santoro et al., 2003; Zilberman et al., 2003]. In order to ensure that these new measuring methods optimise diagnosis and treatment, it is essential that the accuracy and precision of any new type of 3D measuring system be evaluated prior to being introduced to clinical practice. To this end, it must be demonstrated that a new system estimates measurements on digital models that are comparable to measurements derived by the "gold standard" method, in this case the direct measurement of plaster models. Without evidence of the equivalence or superiority of the new methods to conventional ones, the quality of diagnosis and measurement-based predictions may be compromised. According to previous reports, the differences between linear measurements obtained using calipers on plaster models and those obtained from digital models were less than 0.5 mm and deemed clinically insignificant [Bell et al., 2003; Bootvong et al., 2010; Dalstra and Melsen, 2009; Leifert et al., 2009; Quimby et al., 2004; Redlich et al., 2008; Santoro et al., 2003; Zilberman et al., 2003]. However, there is currently little information about the accuracy of the 3D measuring system; a situation that must be rectified before it is introduced to clinical use. Moreover, whilst evidence validating digital models as an alternative to plaster models is accumulating, the methodological quality of the studies is variable [Fleming et al., 2011].

In order to aid paediatric dentists in deciding whether or not measurements carried out on digital models can replace those performed using the current gold standard method, an analysis of measurements made by multiple examiners using both methods on paediatric patients is called for. Such a protocol should include statistical assessment of: whether there is any systematic bias in the measurements obtained with the new method; whether there is a difference in the variability between the 2 methods; and the reliability of measurements made by different examiners. Thus, in this study, we aimed to assess the accuracy of diagnostic linear measurements of digital models obtained from the non-contact 3D measuring system in comparison with the results obtained from the gold standard method (direct measurements made on plaster models) in paediatric patients.

Materials and methods

The study protocol was approved by the Ethic Committee of Epidemiology, Hiroshima University (N. 329).

Sample

As study samples, we obtained 12 pairs of maxillary and mandibular plaster models from 12 Japanese children (6 boys and 6 girls) with deciduous dentition. The study sample included both subjects with normal occlusion and different types of malocclusion (crowding, maxillary protraction, crossbite). All the teeth in the sample had normal morphology and displayed no fracture, remarkable erosion, caries, or restorations.

Measurement points

The purpose of the measurements taken in this study was to enable space analysis and prediction of occlusal development. The tooth sizes, arch widths, and arch lengths were examined. Tooth size was measured as the maximum mesio-distal crown width in 2 teeth (maxillary deciduous right central incisor and mandibular deciduous right second molar). A total of 6 arch-length or arch-width measurements were made: the width of the cusp of deciduous canine on both sides and in each arch; the width of the lingual cervical line of the second deciduous molars on both sides and in each arch; and the length of the interlabial surfaces of the central deciduous incisors to the left and right distal surfaces of the second deciduous molar crown in each arch.

Measurements

All experiments in this study compared the direct measurements of unmarked plaster models with the measurements of digital models obtained from the same subject. Direct measurements were performed using a digital caliper (Digimatic Caliper: CD-15CPX, Mitutoyo, Kawasaki, Japan) with an accuracy of 0.01 mm. The maxillary and mandibular plaster models were individually scanned using a non-contact 3D laser scanner (RexcanDS, Solutionix, Seoul, Korea). The measurement system of the RexcanDS is based on the principle of phase-shifting optical triangulation (Fig. 1). The scanner provides non-contact high-accuracy inspection up to a quoted resolution of 0.016 mm. The object is scanned with halogen light stripes, and the twin cameras receive the light reflected from the surface of the object. This procedure takes approximately 10 min and the file size of the scanned sample was approximately 30 Mb. The generated 3D model of the dental cast is exported in stereolithography format with the 3D coordinates. RapidForm 2006 software (INUS Technology, Seoul, South Korea) was used to reconstruct the scanned image into a digital model. The tooth sizes, arch width, and arch lengths were measured using a function of the RapidForm 2006 software on a computer display by clicking on the measurement points with a computer mouse. To facilitate ease and accuracy of measurements, the digital images of the model could be rotated or enlarged on the screen as required.

A total of 4 examiners participated in the study and were trained in the use of both the digital caliper and the RapidForm 2006 software. In order to avoid that 1 set of measurements influenced the other set of measurements, the samples were not measured with both methods sequentially. Patients were randomly

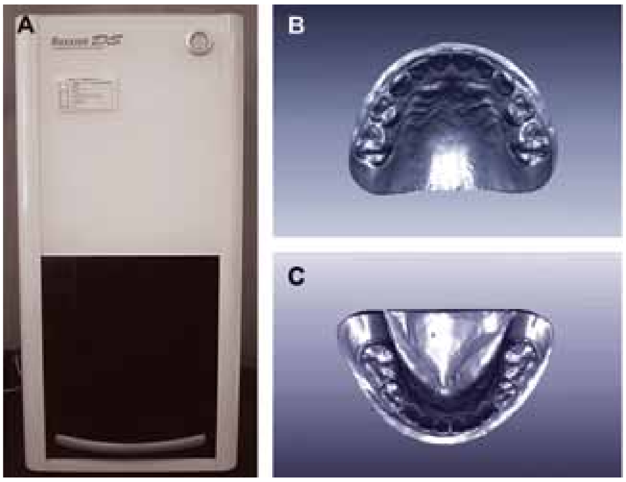

FIG. 1 A Non-contact 3D laser scanner. B, C Panels showing acquired digital images of the maxilla and the mandible.

allocated to 2 of the 4 examiners and the examiner pairs alternated so that all examiners were paired with all of the other examiners over the course of the study. Each examiner used both measurement methods for each patient on 2 occasions separated by at least 2 weeks in order to avoid influence of the first measurements on the second. Thus, for a given dental measurement, there were 4 values from each of the assigned examiners for each patient for each method. In this way, a total of 96 measurements were carried out.

Statistical analysis

A mixed-effects linear regression model was fit to each dataset. The model contained measurement method (calipers versus digital scanning), examination (first versus second) and their interaction as fixed effects. The model contained random effects for patient, examiner and their interaction. Finally, the model contained 2 residual variance estimates, one for each method in order to allow for heteroscedasticity (i.e. unequal variances) between error variances for both methods. Using a likelihood ratio test, it was possible to test whether the 2 residual standard deviations for the methods were statistically significantly different from each other. The initial model was then reduced by means of likelihood-ratio tests to the final model, which might have removed the random effect for patient–by-examiner interaction, the separate residual variances, or both. In addition, if the variance component for the examiner was too numerically small to be estimated reliably by the algorithm, both the examiner and patient–by-examiner variance component were removed from the model. Using the final model, the fixed effects term for method-by-examination interaction was tested by means of the likelihood-ratio test.

Likelihood-ratio tests were used to determine whether the interaction terms (patient × examiner and method × examination) were statistically significant at the 5% level. If not, then these terms were removed from the model. Similarly, the difference in residual variances was tested and removed from the model if it was not statistically significant. The intraclass correlation coefficient (ICC) was computed, according to the following formula: ICC = σ^2 patient/(σ^2 examiner + σ^2 patient + σ^2 patient × examiner + σ^2 residual). In cases where the patient-by-examiner variance was removed from the model, or where both the examiner variance and the patient–by-examiner variance were removed, the corresponding variance terms were taken as zero. When heteroscedastic residual variances were included in the final model, 2 ICC values were computed separately for each of the methods. For each method, a separate Bland–Altman plot was produced by averaging measurements from both examinations for each combination of patient and examiner. From these, difference and average values between the 2 methods were computed for the plots. For equivalence, the null hypothesis was that the acceptable difference between the methods exceeds ±0.2 mm. Post hoc power analysis was performed to determine the statistic power in this study. Two Wald tests were performed, one with hypothesis that the difference was <0.2 mm and the second with hypothesis that the difference was >0.2 mm. The statistical analyses and production of graphs were performed using Stata, Release 11.2 (Stata Corp LP, Texas, USA).

Results

The results include the estimate of the bias (difference in millimeters between the measurements from the 2 methods) obtained from regression model fits, ICC values, and residual standard deviations (if statistically significantly different between methods).

Table 1 shows fixed effects parameters of the final models. The method-by-examination interaction terms in each part were not significant and were not included in the final models. The average difference in measured values between the two methods ("fixed bias"), which are derived from the final model, was always less than 0.2 mm. The difference was statistically significant in a single case: the mandibular arch length, but the difference between methods was negligible (0.16 mm) relative to the measurement itself (arch length or width, or tooth size) which was around 30 mm. In real terms, the difference between methods although statistically significant is only about 0.5% of the magnitude of the measurement. Table 2 shows random effects parameters of final models and ICCs. Random effect of examiners was always the smallest component of variance, and frequently negligible. In 3 cases, there was statistically significant heteroscedasticity (i.e., a difference was seen between methods with regards to their residual variances), but the difference was practically insignificant in comparison to the magnitude of the random effect of the patient. Intraclass correlation coefficient was typically very similar between methods even in cases of statistically significant heteroscedasticity. These were typically greater than 90% and never below the mid-80% range.

Predictions from the model for the fixed effects were used to prepare Bland–Altman plots for inspection of each measurement point dependencies in systematic differences between the 2 measurement methods (Fig. 2). Bland–Altman plots demonstrate substantial

FIXED EFFECTS PARAMETERS						
Type of tooth measurement		Coefficient	SE	P value	95% CI	
					Lower	Upper
Maxillary deciduous incisor	Method	-0.03	0.02	.14	-0.08	0.01
	Examination	0.01	0.02	.68	-0.03	0.04
	Constant	6.73	0.13	<.001	6.48	6.98
Mandibular deciduous molar	Method	0.03	0.04	.46	-0.04	0.10
	Examination	0.01	0.04	.79	-0.06	0.08
	Constant	10.24	0.14	<.001	9.97	10.52
Maxillary intercanine width	Method	0,04	0.05	.50	-0.07	0.14
	Examination	0.04	0.05	.46	-0.06	0.14
	Constant	30.74	0.48	<.001	29.79	31.68
Mandibular intercanine width	Method	0.06	0,04	.12	-0.02	0,15
	Examination	-0.04	0.04	.29	-0.13	0.04
	Constant	23.74	0.49	<.001	22.78	24.70
Maxillary intermolar width	Method	0.00	0.03	.99	-0.06	0.06
	Examination	-0.02	0.03	.50	-0.08	0.04
	Constant	29.45	0.53	<.001	28.41	30.48
Mandibular intermolar width	Method	-0.16	0.04	<.001	-0.24	-0.08
	Examination	0.00	0.04	.92	-0.07	0.07
	Constant	28.04	0.47	<.001	27.12	28.96
Maxillary arch length	Method	0.05	0.05	.35	-0.05	0.16
	Examination	0.11	0.04	<.001	0.04	0.19
	Constant	29.47	0.39	<.001	28.71	30.24
Mandibular arch length	Method	-0.02	0.05	.73	-0.12	0.08
	Examination	0.02	0.05	.72	-0.08	0.12
	Constant	26.66	0.44	<.001	25.80	27.52

Method-by-examiner interaction terms in each part were not statistically significant, and were eliminated from subsequent analysis. SE = standard error; CI = confidence interval.

TAB. 1 Final models of mixed-effects linear regression models.

RANDOM-EFFECTS PARAMETERS						
Type of tooth measurement		Estimate	SE	95% CI		ICC
				Lower	Upper	
Maxillary deciduous incisor	SD (Examiner)	0.02	0.04	0.00	0.76	
	SD (Patient)	0.44	0.09	0.29	0.65	
	SD (Examiner×Patient)	0.05	0.02	0.03	0.11	
	Residual					
	SD (Caliper)	0.06	0.01	0.05	0.08	0.965
	SD (3D)	0.14	0.02	0.11	0.18	0.898
Mandibular deciduous molar	SD (Examiner)	0.06	0.04	0.02	0.20	
	SD (Patient)	0.46	0.10	0.31	0.70	
	SD (Residual)	0.18	0.01	0.15	0.21	0.866
Maxillary intercanine width	SD (Constant)	1.66	0.34	1.11	2.48	
	SD (Residual)	0.26	0.02	0.22	0.30	0.977
Mandibular intercanine width	SD (Constant)	1.69	0.35	1.13	2.52	
	SD (Residual)	0.20	0.02	0.18	0.24	0.986
Maxillary intermolar width	SD (Constant)	1.83	0.37	1.23	2.73	
	SD (Residual)	0.15	0.01	0.13	0.18	0.993
Mandibular intermolar width	SD (Constant)	1.62	0.33	1.09	2.42	
	Residual					
	SD (Caliper)	0.14	0.02	0.11	0.18	0.993
	SD (3D)	0.26	0.03	0.21	0.32	0.976
Maxillary arch length	SD (Examiner)	0.04	0.04	0.01	0.25	
	SD (Patient)	1.35	0.28	0.90	2.01	
	Residual					
	SD (Caliper)	0.35	0.04	0.28	0.43	0.940
	SD (3D)	0.14	0.02	0.11	0.18	0.989
Mandibular arch length	SD (Constant)	1.52	0.31	1.02	2.27	
	SD (Residual)	0.25	0.02	0.22	0.29	0.975

SE = standard error; CI = confidence interval.

TAB. 2 Final models of mixed-effects linear regression models and intraclass correlation coefficients (ICCs) of measurement data.

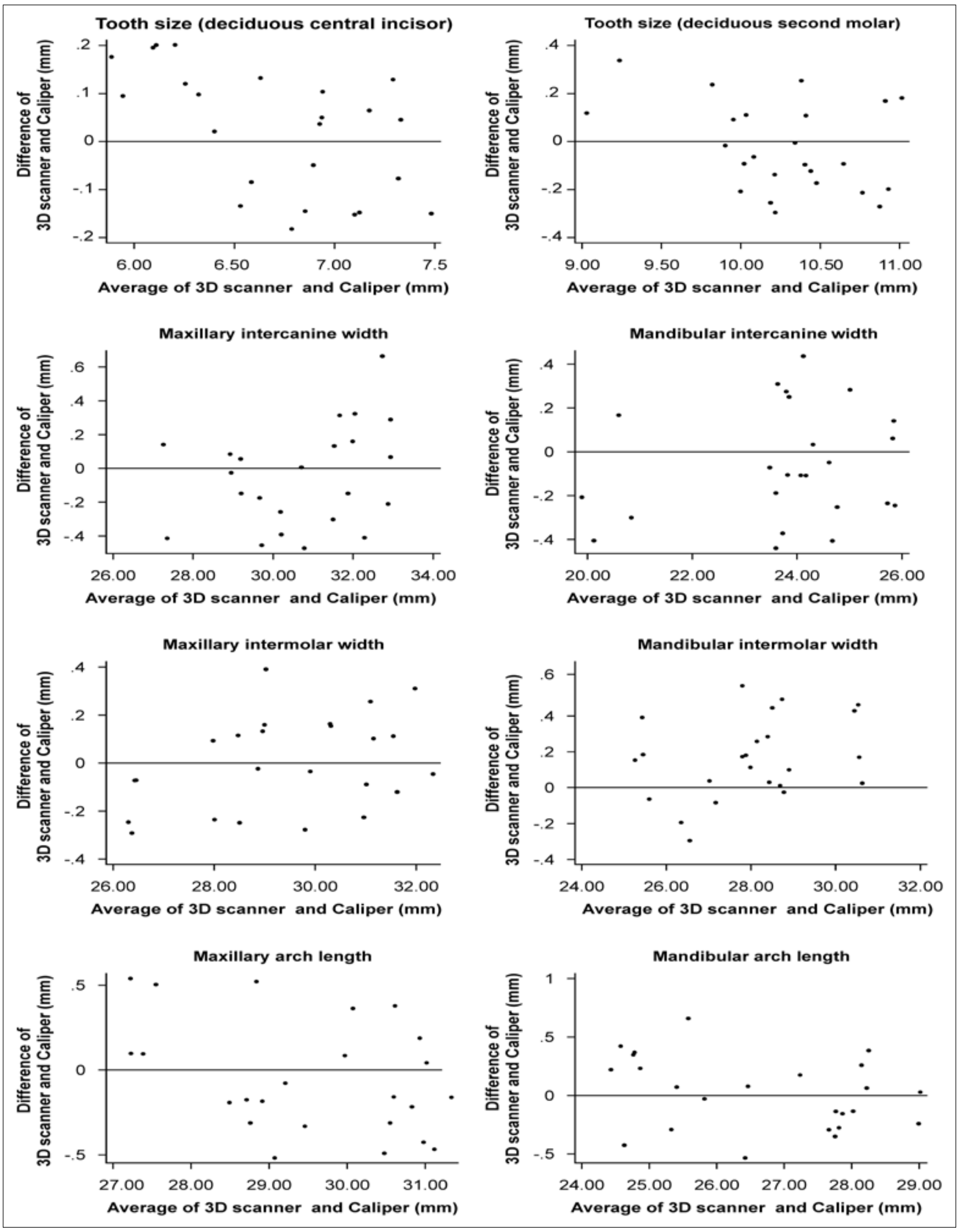

FIG. 2 Bland–Altman plots showing systematic differences between the 2 measurement methods.

dependence of the difference between the 2 methods on the mean value ("proportional bias") in many cases. Measurements made with the calipers tended to be higher than those made with the 3D scanner at the lower end of the measured range; the opposite effect was seen at the higher end of the measured range. Nevertheless, throughout the measured range, the differences between the 2 measurement methods were typically very small in comparison to the magnitude of the measurements (arch length or width, or tooth size).

The post hoc test analysis revealed that the statistical power (in the case of maximum standard deviation in Table 2) was 99%. We rejected the null hypothesis to accept the alternative hypothesis that the difference between the two methods is less than ±0.2 mm.

Discussion

This study aimed to investigate the potential usefulness of a new non-contact 3D measuring system for the measurement of paediatric dentition models by comparison with the current "gold standard" method – making measurements of plaster models. We found no clinically significant difference in accuracy between the direct measurement of plaster models with calipers and the measurement of 3D digital models using the computer software. The post hoc power analysis revealed that the sample size in this study was sufficient. These findings suggest that the non-contact 3D measuring system may be suitable for use in the clinical setting. Bell et al. [2003] reported that the average difference between liner measurements of dental casts and 3D images was 0.27 mm, while Redlich et al. [2008] reported that the difference in space analyses between the caliper and cross-section plane measurements was very small (0.38–0.74 mm). Santoro et al. [2003] demonstrated that tooth width and overbite measurements made on plaster and digital models showed statistically significant differences (range: 0.16–0.49 mm). Quimby et al. [2004] reported similar results in all parameters except overbite and overjet. Asquith et al. [2007] reported differences in mean tooth size, intercanine width, and intermolar width ranging from -0.62 to 0.19 mm. Finally, although Leifert et al. [2009] found that differences between maxillary arch length calculations were small (<0.5 mm), they concluded that the differences were clinically significant. In contrast, the difference between the means of the 2 measurement methods in this research is similar to or smaller than values reported previously, and appears to show no clinically significant difference.

Although 3 measurement points demonstrated statistically significant heteroscedasticity between the residual variances of the 2 methods, the ICCs were greater than 90 and never below the mid-80% range, suggesting good agreement between the methods and achieving high reliability. With regard to differences in variability between the methods, previous reports have described increased [Dalstra and Melsen, 2009] and decreased [Quimby et al., 2004; Zilberman, et al., 2003] variability in the 3D measurements compared to those in plaster models. Despite these contrasting findings, all studies have recognised the usefulness of the 3D scanner [Quimby et al., 2004; Santoro et al., 2003; Zilberman, et al., 2003]. Bootvong et al. [2010] reported that ICCs for tooth, intercanine and intermolar width were greater than 0.70 and concluded that there was substantial to excellent agreement between assessment of tooth dimensions and arch relationships between plaster and virtual models. Our findings are in agreement with these reports, demonstrating the reliability of measurements with different examiners and the potential usefulness of the 3D scanner. Though not clinically relevant, the variation in the data sets may represent challenges in view control or landmark identification performed on a computer screen [Abizadeh et al., 2012; Asquith et al., 2007; Fleming et al., 2011; Leifert et al, 2009; Park et al., 2009; Zilberman et al., 2003]. This may be related to the inter-examiner variability that was observed in this study. Despite the potential for improvement in these aspects, our results show, nonetheless, that the analysis of dentition using a 3D scanner is comparable in terms of accuracy to the "gold standard" method of directly measuring a plaster models using digital caliper.

In addition to the challenge of identifying certain landmarks on the computer screen, this study highlights other areas for further investigation. For example, the reliability of measurements of angle, area, and volume should be examined. It will also be important to verify the performance of the 3D scanning method at sites such as undercuts or curved surfaces, where measurements using 3D scanners are presumed to be less accurate. Clarification of these remaining issues would confirm the reliability and potential usefulness of 3D measurements of dentition models and could lead to the adoption of this simple and convenient method in paediatric clinics and academic research centres. Moreover, our results could help address issues in other similar models such as the superimposing of anatomic structures on 3D-Cone-Beam CT images or facial images, and could also help in treatment simulation/planning situations.

The results of this study suggest that the accuracy of analysing the dentition of children using a 3D scanner is comparable to that of the "gold standard" method that employs plaster models and digital calipers. The 3D scanning system has many advantages over manual measurement methods. It is hoped that these results lead towards a more widespread adoption of this technology in paediatric dentistry.

References

- Abizadeh N, Moles DR, O'Neill J, Noar JH. Digital versus plaster study models: how accurate and reproducible are they? J Orthod 2012;39:151–159.
- Asquith J, Gillgrass T, Mossey P. Three-dimensional imaging of orthodontic models: a pilot study. Eur J Orthod 2007;29:517–522.
- Bell A, Ayoub AF, Siebert P. Assessment of the accuracy of a three-dimensional imaging system for archiving dental study models. J Orthod 2003;30:219–223.
- Bootvong K, Liu Z, McGrath C, Hägg U, Wong RW, Bendeus M, Yeung S. Virtual model analysis as an alternative approach to plaster model analysis: reliability and validity. Eur J Orthod 2010;32:589–595.
- Dalstra M, Melsen B. From alginate impressions to digital virtual models: accuracy and reproducibility. J Orthod 2009;36:36–41.
- Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review. Orthod Craniofac Res 2011;14:1–16.
- Leifert MF, Leifert MM, Efstratiadis SS, Cangialosi TJ. Comparison of space analysis evaluations with digital models and plaster dental casts. Am J Orthod Dentofacial Orthop 2009;136:16.e1–e4.
- Okunami TR, Kusnoto B, BeGole E, Evans CA, Sadowsky C, Fadavi S. Assessing the American Board of Orthodontics objective grading system: digital vs plaster dental casts. Am J Orthod Dentofacial Orthop 2007;131:51–56.
- Park K, Jung DW, Kim JY. Three-dimensional space changes after premature loss of a maxillary primary first molar. Int J Paediatr Dent 2009;19:383–389.
- Quimby ML, Vig KW, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. Angle Orthod 2004;74:298–303.
- Redlich M, Weinstock T, Abed Y, Schneor R, Holdstein Y, Fischer A. A new system for scanning, measuring and analyzing dental casts based on a 3D holographic sensor. Orthod Craniofac Res 2008;11:90–95.
- Rosseto MC, Palma FM, Ferreira RI, Pinzan A, Vellini-Ferreira F. Comparative study of dental arch width in plaster models, photocopies and digitized images. Braz Oral Res 2009;23:190–195.
- Santoro M, Galkin S, Teredesai M, Nicolay OF, Cangialosi TJ. Comparison of measurements made on digital and plaster models. Am J Orthod Dentofacial Orthop 2003;124:101–105.
- Veenema AC, Katsaros C, Boxum SC, Bronkhorst EM, Kuijpers-Jagtman AM. Index of complexity, outcome and need scored on plaster and digital models. Eur J Orthod 2009;31:281–286.
- Watanabe-Kanno GA, Abrão J, Miasiro Junior H, Hiroshi Sánchez-Ayala A, Lagravère MO. Reproducibility, reliability and validity of measurements obtained from Cecile3 digital models. Braz Oral Res 2009;23:288–295.
- Zilberman O, Huggare JA, Parikakis KA. Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models. Angle Orthod 2003;73:301–306.