sciendo

DOI: 10.2478/aoj-2009-0014

Validity, reliability and reproducibility of three methods used to measure tooth widths for Bolton analyses

Devan Naidu,* Justin Scott,† Desmond Ong* and Christopher T.C. Ho*

School of Dentistry, The University of Queensland, Brisbane,* and The Queensland Institute of Medical Research, Herston,† Queensland, Australia

Background: Vernier calipers have traditionally been used as the 'gold-standard' for tooth width measurements. New digital methods may prove to be as valid, reliable and reproducible as caliper measurements.
Aims: To determine the validity, reliability and reproducibility of mesio-distal crown measurements made with calipers, using the DigiModel method and on digital photographs, and to determine the validity, reliability and reproducibility when the measurements are employed in Bolton tooth size analyses.
Methods: Twenty-five consecutive study models were used. The maximum mesio-distal crown widths of 12 teeth in each arch (first molar to first molar) were measured with digital calipers, using the DigiModel software (OrthoProof, Caulfield South, Victoria, Australia) and on standardised digital photographs of the models by eight examiners. One examiner measured all models three times using the three measurement methods, two examiners measured all models once using the three measurement methods and five examiners measured all models once using the calipers only. Validity for the widths of individual teeth, the mean tooth width and the Bolton ratios were assessed using a general linear model two-way analysis of variance and pair-wise comparisons between the two digital methods and the calipers. Reliability was assessed with Cronbach's alpha and reproducibility with the intra-class correlation coefficient.
Results: There were statistically significant differences between the tooth widths measured with DigiModel and the calipers ($p < 0.001$) and with the calipers and digital photographs ($p < 0.001$). However these discrepancies were judged not to be clinically significant. The Cronbach's alpha scores for all methods were classified as 'excellent'. The intra-class correlation coefficient values exceeded 0.75 for the digital calipers, 0.79 for DigiModel and 0.54 for the digital photographs.
Conclusions: The validities of DigiModel and digital photographs for tooth width measurements and Bolton analyses are clinically acceptable. The reliability of both methods is also excellent. However, DigiModel software provided more accurate Bolton's ratios and demonstrated greater reproducibility than the digital photographs.
(Aust Orthod J 2009; 25: 97–103)

Received for publication: November 2008
Accepted: February 2009

Devan Naidu: devannaidu1@gmail.com
Justin Scott: Justin.Scott@climatechange.qld.gov.au
Desmond Ong: d.ong@uq.edu.au
Christopher T.C. Ho: c.ho@uq.edu.au

Introduction

An important goal in comprehensive orthodontic treatment is to achieve an optimal final occlusion with an ideal overjet, overbite and buccal segment relationships. Factors vital to the success of this goal are the existence of well-proportioned maxillary and mandibular tooth sizes. In 1958 Bolton used cases with excellent occlusion and proposed desirable tooth size ratios for the anterior teeth and the anterior and posterior teeth, excluding the second and third molars.[1] Traditionally, the maximum mesio-distal widths of the crowns of the teeth, measured with calipers, have been used in these analyses and caliper measurements are widely regarded as the 'gold standard'.

© Australian Society of Orthodontists Inc. 2009

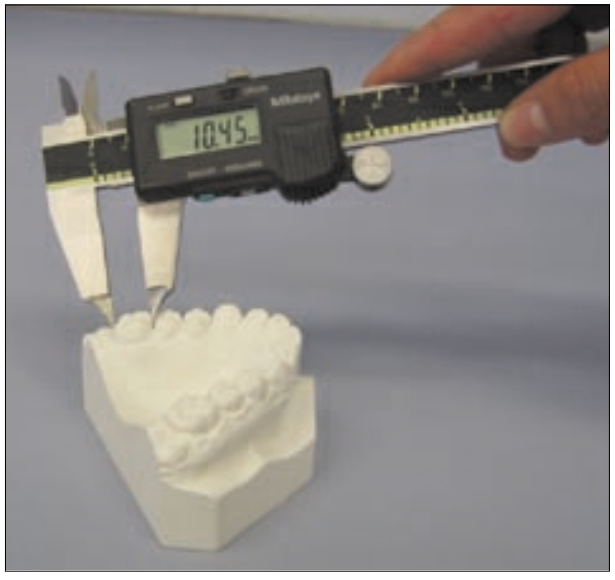

Figure 1. Measurement with Mitutoyo 6"/150 mm Digital Calipers.

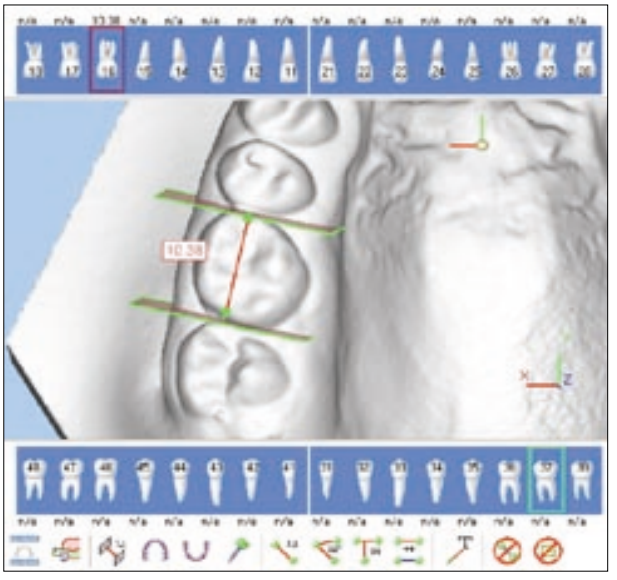

Figure 2. Measurement with DigiModel.

Recently, digital methods and 3-D virtual study models have been introduced to the profession. Some of the new digital methods scan alginate impressions and do not require plaster study models. Digital photographs of plaster study models offer an intermediate step that may suit many clinicians because the technology is readily available, and many clinicians are familiar with the basic equipment. However, no published studies have evaluated the validity, reliability and reproducibility of mesio-distal crown measurements made on 3-D virtual study models from a recently developed method that uses computerised tomography to scan the alginate impressions, and digital photographs.

The aims of this study were to determine the validity, reliability and reproducibility of mesio-distal crown measurements made with calipers, using the DigiModel method and on digital photographs, and to determine the validity, reliability and reproducibility of these measurements in Bolton tooth size analyses.

Materials and methods

Twenty-five consecutive sets of intact pretreatment study models with 12 fully erupted permanent teeth in each arch (first molar to first molar) were used. Study models were excluded if a tooth had a large restoration, an obvious crown deficiency due to caries or fracture, an impression and/or casting artifact or if the digital models had a crown artifact.

The maximum mesio-distal widths of the crowns were measured directly with digital calipers, using the DigiModel software and on digital photographs. The examiner was blinded to the identity of the models by assigning new random numbers for each of the three measurement methods. The maximum mesio-distal width of a tooth was defined as the distance between the anatomical contact areas when the teeth were correctly aligned. In addition, the measurements were made parallel to the occlusal surface and labial/buccal surfaces. All mesio-distal widths were measured to the nearest 0.01 mm and the anterior Bolton's ratio (ABR) and the overall Bolton's ratio (OBR) were calculated using data from each method.

For the first method, a digital caliper (Mitutoyo Corporation, Tokyo, Japan) was used to measure the mesio-distal widths directly (Figure 1). The tips of the calipers were sharpened to permit accurate placement in the interproximal embrasures and the caliper was linked via a RS-232C interface cable to a personal computer. The light source was standardised for all direct measurements. The second method of measurement used the DigiModel software (OrthoProof, Caulfield South, Victoria, Australia). Separate alginate impressions of patients were taken and sent to the OrthoProof laboratory where they were scanned with computerised tomography (CT) and converted into digital models. This data was stored on CDs and returned to the clinic. The digital models were then



Figure 3. A Canon EOS 350D digital camera mounted on a specialised work station.

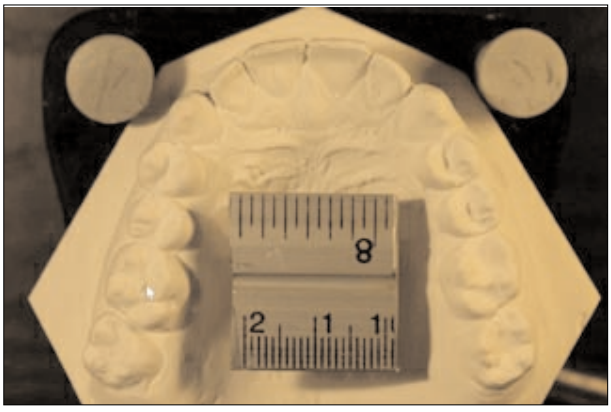

Figure 4. Measurement with digital photos.

uploaded onto a computer and opened with the DigiModel software. Tooth widths were measured using the DigiModel 'Measurement tool' and the magnification, rotation and shifting features were used as required (Figure 2). The third method of measurement involved measurement of the mesio-distal widths on standardised digital photographs. A Canon EOS 350D digital camera with a 100 mm EF Macro Lens 1:2.8 was used. The camera was mounted on a purpose-built stand and adjusted with the aid of two spirit levels attached to the lens so that the sensor plane was horizontal (Figure 3). The study model was then placed on a surveyor model holder and a tile with two spirit levels placed on the model, contacting the incisors and first permanent molars. The surveyor table was then adjusted until the tile was horizontal, which averaged out the curves of Spee and Monson, thus equalising image distortion of the teeth and ensuring that the sensor and occlusal planes were parallel. The images were calibrated with a millimetre ruler attached to the study model at the level of the contact areas. An independent ring light was attached to the camera and three tungsten lamps positioned around each study cast to prevent shadows falling on the teeth. The digital photographs were transferred to a computer and the mesio-distal widths measured with the image measurement software (Image Measurement, School of Dentistry, University of Queensland, Brisbane, Australia) (Figure 4).

Eight independent examiners (A, B, C, D, E, F, G, and H) participated in the study. Examiner A measured all models three times using the three measurement methods. Replications were separated by a period of two weeks. Examiners B and C measured all models once using the three measurement methods and Examiners D, E, F, G, and H measured all models once using the calipers only.

Validity is defined as the extent to which a measurement obtained represents the object of interest.[2] The term 'accuracy' is often used to describe validity.[2] In the present study, validity was considered to be the extent to which measurements from the two digital methods agreed with the caliper measurements. Measurements made by the calipers were taken as the 'true' values or the 'gold standard'. Thus, accuracy was judged as the closeness of the digital values to the caliper measurements. Reliability was considered as the extent to which the measurements were repeatable under identical conditions,[3] and reproducibility was the closeness of agreement between independent results obtained with the same method on identical test material, but under different conditions.[4] It refers to the ability of a measurement technique to be accurately reproduced by another examiner.[4]

Statistical analysis

The data were analysed using the statistical software packages SAS version 9.1 and SPSS version 15.0. Measurements from Examiner A were used to investigate validity and reliability. Validity was assessed using a general linear model (GLM) two-way analysis of variance (ANOVA) and pair-wise comparisons (Tukey's test) between the two digital methods and the calipers.[5,6] This was done for each individual

Table I. Inter-examiner reproducibility.

	Examiners A,B,C,D,E,F,G,H		Examiners A,B,C			
	Digital caliper		DigiModel		Digital photograph	
	ICC	95% CI	ICC	95% CI	ICC	95% CI
Tooth widths	0.96	0.93 – 0.98	0.98	0.96 – 0.99	0.85	0.73 – 0.93
OBR	0.76	0.64 – 0.87	0.79	0.64 – 0.89	0.59	0.36 – 0.77
ABR	0.75	0.63 – 0.86	0.84	0.72 – 0.92	0.54	0.31 – 0.74

ICC, Intra-class correlation coefficient

tooth width, the average tooth width, the OBR and the ABR. A significance level of 0.05 was selected. Reliability was assessed with Cronbach's alpha.[7] A Cronbach's alpha score was derived to examine the association between the three replicate measurements. Measurements from Examiners A, B, and C were used to investigate the reproducibility of the DigiModel and digital photographic measurements. Measurements from all eight examiners were used to investigate the reproducibility of the caliper measurements. Reproducibility was tested with the intra-class correlation coefficient (ICC).[8]

Results

Digital calipers

Cronbach's alpha scores for the digital calipers were: mesio-distal widths 0.99; OBR 0.98; ABR 0.99. The data were in close agreement when the caliper measurements for the eight examiners were compared. The ICC values for the digital calipers exceeded 0.75 (Table I).

DigiModel vs digital calipers

There was a statistically significant difference between the mean tooth widths obtained with DigiModel software and the calipers ($p < 0.001$). Tooth widths measured with the DigiModel software were, on average, 0.08 mm less than those measured with the calipers (95% CI: 0.05, 0.12 mm) (Table II). There were no statistically significant differences between the mean Bolton's ratios obtained with the DigiModel method and the calipers (OBR: $p = 0.75$; ABR: $p = 0.28$). The OBRs obtained with DigiModel were, on average, 0.18 per cent greater than those obtained with the calipers (95% CI: –0.42, 0.78%), and the ABRs were, on average, 0.43 per cent less than those obtained with the calipers (95% CI: –1.11, 0.25%) (Table III). The Cronbach's alpha scores for the DigiModel software were 0.99 for tooth widths, and 0.97 for the OBR and ABR. Inter-examiner measurements with the software were highly correlated with ICC values exceeding 0.79 (Table I).

Digital photographs vs digital calipers

When measurements taken from the digital photographs were compared with the caliper measurements, a statistically significant difference was detected between the mean mesio-distal widths ($p < 0.001$). Tooth width measurements from the digital photographs were, on average, 0.07 mm larger than those from the calipers (95% CI: 0.03, 0.10 mm) (Table II). Highly significant differences were evident for measurements of the lower incisor teeth ($p < 0.001$). There was no statistically significant difference between the mean OBRs from measurements taken from the digital photographs and the calipers ($p = 0.06$), but there was a statistically significant difference between the mean ABRs ($p = 0.02$). The OBRs and ABRs calculated from data obtained from the digital photographs were, on average, 0.59 per cent (95% CI: –0.01, 1.19%) and 0.82 per cent (95% CI: 0.14, 1.50%) greater, respectively, than the OBRs and ABRs obtained using the caliper measurements (Table III). The Cronbach's alpha scores for the digital photographs were 0.99 for measuring tooth widths, 0.99 for the OBR and 0.98 for the ABR. The ICC values exceeded 0.54 (Table I).

Discussion

This study is the first to evaluate the validity, reliability and reproducibility of crown width measurements

Table II. Comparison of measurements with digital caliper, DigiModel and on digital photographs, by Examiner A.

Tooth	Digital caliper	DigiModel			Digital photograph		
	Mean	Difference	95% CI	*p*	Difference	95% CI	*p*
11	8.82	0.08	-0.01 - 0.16	0.074	-0.12	-0.21 - -0.04	**0.004**
12	7.05	0.16	0.07 - 0.24	**<0.001**	-0.20	-0.28 - -0.11	**<0.001**
13	7.93	0.10	0.03 - 0.17	**0.003**	-0.07	-0.14 - 0.00	**0.039**
14	7.29	0.03	-0.05 - 0.11	0.577	-0.09	-0.17 - -0.01	**0.018**
15	6.96	0.03	-0.02 - 0.09	0.334	-0.08	-0.13 - -0.02	**0.008**
16	10.53	0.15	0.05 - 0.25	**0.002**	0.17	0.07 - .27	**<0.001**
21	8.84	0.10	0.01 - 0.19	**0.022**	-0.08	-0.18 - 0.01	0.071
22	6.96	0.11	0.03 - 0.19	**0.005**	-0.24	-0.32 - -0.16	**<0.001**
23	7.91	0.05	-0.02 - 0.13	0.220	-0.07	-0.15 - 0.01	0.101
24	7.29	0.05	-0.02 - 0.11	0.228	0.01	-0.06 - 0.08	0.906
25	7.00	0.08	0.02 - 0.15	**0.013**	0.01	-0.06 - 0.08	0.919
26	10.43	0.17	0.04 - 0.30	**0.008**	0.22	0.10 - 0.35	**<0.001**
31	5.56	0.11	0.04 - 0.19	**0.001**	-0.17	-0.24 - -0.10	**<0.001**
32	6.08	0.11	0.04 - 0.17	**0.001**	-0.17	-0.24 - -0.10	**<0.001**
33	6.90	0.11	0.02 - 0.21	**0.015**	-0.16	-0.25 - -0.07	**<0.001**
34	7.36	0.02	-0.06 - 0.10	0.824	-0.11	-0.19 - -0.03	**0.004**
35	7.41	0.00	-0.08 - 0.08	0.997	-0.10	-0.18 - -0.02	**0.012**
36	11.22	0.06	-0.02 - 0.15	0.195	0.04	-0.05 - 0.12	0.562
41	5.56	0.15	0.08 - 0.23	**<0.001**	-0.17	-0.25 - -0.10	**<0.001**
42	6.04	0.12	0.05 - 0.19	**0.001**	-0.13	-0.21 - -0.06	**<0.001**
43	6.86	0.07	-0.01 - 0.15	0.126	-0.20	-0.28 - -0.12	**<0.001**
44	7.39	0.12	0.03 - 0.21	**0.008**	0.06	-0.03 - 0.15	0.203
45	7.40	0.02	-0.06 - 0.09	0.862	0.01	-0.06 - 0.09	0.900
46	11.11	-0.04	-0.14 - 0.06	0.654	0.04	-0.07 - 0.14	0.660

ANOVA pair-wise comparisons (Tukey's test)
Significant values in bold

made with the DigiModel software, although there are several reports of other software programmes available for 3-D digital study models.[5,9–12] Compared to the caliper measurements, DigiModel tended to underestimate the mesio-distal widths of the teeth by an average of 0.08 mm, overestimate the OBR by an average of 0.18 per cent and underestimate the ABR by an average of 0.43 per cent. These over- and underestimates equated to an overall tooth size discrepancy of 0.18 mm and an anterior tooth size discrepancy of 0.24 mm, which we consider are not clinically significant. According to Shellhart et al., Bolton's ratios calculated with data obtained with needle point dividers may vary by as much as ±2.2 mm, thus emphasising the relative insignificance of the discrepancies we found with the DigiModel software.[13]

When the measurements were repeated using the DigiModel software, the replicates were strongly correlated: the high Cronbach's alpha scores we obtained indicate excellent reliability. The data for reproducibility were also very promising. Roberts and Richmond suggested that an ICC value below 0.4 constitutes 'poor' reliability, between 0.4 and 0.75 is 'fair to good', and above 0.75 represents 'excellent' reliability.[3] Thus, our finding of ICC scores exceeding 0.79 for DigiModel indicates 'excellent' reproducibility. Although the DigiModel method has the same shortcomings as other methods of producing digital models the differences are extremely small and are unlikely to be of clinical significance.[5,9,12]

The validity of tooth width measurements and Bolton analyses made on the digital photographs

Table III. Comparison of the Bolton's ratios using the three measurement methods, by Examiner A.

	Digital caliper		DigiModel			Digital photograph		
	Mean	95% CI	Mean	95% CI	*p*	Mean	95% CI	*p*
OBR	91.66	91.30 - 92.01	91.84	91.48 - 92.19	0.75	92.25	91.89 - 92.60	0.06
ABR	77.95	77.55 - 78.35	77.52	77.12 - 77.92	0.28	78.77	78.37 - 79.17	**0.02**

ANOVA pair-wise comparisons (Tukey's test)
Significant value in bold

were also clinically acceptable. Measurements on the digital photographs were slightly larger (Mean difference: 0.07 mm) than the caliper measurements and the OBR and ABR were overestimated by 0.59 per cent and 0.82 per cent, respectively. The overestimates of the tooth widths contributed 0.44 mm to the OBR and 0.57 mm to the ABR, which are unlikely to be of clinical significance. Although the measurements and Bolton ratios made on the digital photographs had excellent reliability and good reproducibility (ICC scores exceeding 0.54) they were slightly less reproducible than the same measurements made with the DigiModel software.

A variety of reasons could be responsible for the slight inaccuracies in tooth width measurements obtained with the two new methods. For DigiModel, the most likely explanation is shrinkage of the alginate impressions during transportation with the result that the digital models and the crowns were slightly smaller.[14] Furthermore, the occlusal anatomy and interproximal contacts were not well-defined on all digital models, which increased the potential for measurement errors. On the digital photographs, the contact areas of tipped teeth were often difficult to identify because the photographs were 2-D representations of the 3-D study models. This may explain the inaccurate measurements of the lower incisors, which were often tipped. Despite these limitations, the errors with both DigiModel and the digital photographs were small and not clinically significant.

Conclusion

The validities of DigiModel and digital photographs for tooth width measurements and Bolton analyses are clinically acceptable. The reliability of both methods is also excellent. However, DigiModel provided more accurate Bolton's ratios and demonstrated greater reproducibility than the digital photographs.

Corresponding author

Associate Professor Christopher T.C. Ho
School of Dentistry
The University of Queensland
200 Turbot Street
Brisbane Qld 4000
Australia
Tel: +61 7 3365 8084
Email: c.ho@uq.edu.au

References

1. Bolton WA. Disharmony in tooth size and its relation to the analysis and treatment of malocclusion. Angle Orthod 1958; 28:113–30.
2. Houston WJ. The analysis of errors in orthodontic measurements. Am J Orthod 1983;83:382–90.
3. Roberts CT, Richmond S. The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. Br J Orthod 1997;24:139–47.
4. McNaught AD, Wilkinson A. IUPAC Compendium of chemical terminology, 2nd edn. Oxford: Blackwell Science, 1997. URL: 'http://www.iupac.org'. Accessed October 2007.
5. Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW. Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements. Am J Orthod Dentofacial Orthop 2006;129:794–803.
6. Cody PC, Smith JK. Applied statistics and the SAS programming language. Pearson Prentice-Hall: New Jersey, 2006:348–9.
7. Field A. Discovering statistics using SPSS. SAGE Publications: London, 2005:667–76.
8. Hamer RM. Intracc:sas. Calculate reliabilities from intraclass correlations. URL: 'http://www.psych.yorku.ca/lab/sas/intracc.htm'. Accessed October 2007.
9. 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;124:101–5.
10. Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR, Jr. A comparison of 3 computerized Bolton tooth-size analyses with a commonly used method. Angle Orthod 2001;71: 351–7.

11. 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;73:301–6.
12. 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.
13. Shellhart WC, Lange DW, Kluemper GT, Hicks EP, Kaplan AL. Reliability of the Bolton tooth-size analysis when applied to crowded dentitions. Angle Orthod 1995;65: 327–34.
14. Coleman RM, Hembree JH, Weber FN. Dimensional stability of irreversible hydrocolloid impression material. Am J Orthod 1979;75:438–46.