ORIGINAL ARTICLE

Validity, reliability, and reproducibility of plaster vs digital study models: Comparison of peer assessment rating and Bolton analysis and their constituent measurements

Daron R. Stevens,[a] Carlos Flores-Mir,[b] Brian Nebbe,[c] Donald W. Raboud,[d] Giseon Heo,[e] and Paul W. Major[f]
Nampa, Idaho, and Edmonton, Alberta, Canada

Introduction: The objective of this validation study was to compare standard plaster models (the current gold standard for cast measurements) with their digital counterparts made with emodel software (version 6.0, GeoDigm, Chanhassen, Minn) for the analysis of tooth sizes and occlusal relationships—specifically the Bolton analysis and the peer assessment rating (PAR) index and their components. **Methods:** Dental casts were poured from 24 subjects with 8 malocclusion types grouped according to American Board of Orthodontics categories. Measurements were made with a digital caliper to the nearest 0.01 mm from plaster models and with the software from the digital models. A paired samples *t* test was used to compare reliability and validity of measurements between plaster and digital methods. **Results:** Reproducibility of digital models via the concordance correlation coefficient was excellent in most cases and good in some. Although statistically significant differences in some measurements were found for the reliability and validity of the digital models via the average mean of the absolute differences of repeated measurements, none was clinically significant. Grouping of the measurements according to the 8 American Board of Orthodontics categories produced no significant difference (Kruskal-Wallis test). No measurement associated with Bolton analysis or PAR index made on plaster vs digital models showed a clinically significant difference. The PAR analysis and its constituent measurements were not significantly different clinically between plaster and emodel media. **Conclusions:** Preliminary results did not indicate that digital models would cause an orthodontist to make a different diagnosis of malocclusion compared with plaster models; digital models are not a compromised choice for treatment planning or diagnosis. (Am J Orthod Dentofacial Orthop 2006;129:794-803)

Orthodontics, like many other areas of life, is going digital. Many orthodontists are joining other health professionals in using paperless patient information systems that include virtual chart notes and health histories, and digital photographs and radiographs. However, a major obstacle for orthodontists is the necessity of plaster study models of a patient's dentition for treatment planning.

In late 1999, OrthoCAD (Cadent, Carlstadt, NJ) developed and released to market virtual digital dental casts. Then, in early 2001, emodels (GeoDigm, Chanhassen, Minn) came to market. The technology of digital study models allows an orthodontist to send a patient's alginate impression or existing plaster study model to 1 of these companies for processing into a virtual 3-dimensional (3D) computerized image. This image is then available to the orthodontist for downloading from the company's web-site within 5 days. Software from the imaging companies allows the orthodontist to view the image and manage it in a virtual 3D environment.

Replacement of plaster orthodontic models with these new virtual counterparts can benefit orthodontics in the following areas: (1) efficiency of having patient records instantly accessible on the computer screen vs retrieving plaster models from a storage area; (2) saving money on the monthly cost of storage space needed for the thousands of traditional plaster models an orthodon-

[a]Private practice, Nampa, Idaho.
[b]Clinical associate professor, Orthodontic Graduate Program, Department of Dentistry, University of Alberta, Edmonton, Alberta, Canada.
[c]Clinical assistant professor, Orthodontic Graduate Program, Department of Dentistry, University of Alberta, Edmonton, Alberta, Canada.
[d]Associate professor, Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada.
[e]Assistant professor, Statistics, Orthodontic Graduate Program, University of Alberta, Edmonton, Alberta, Canada.
[f]Professor, Director of Orthodontic Graduate Program, University of Alberta, Edmonton, Alberta, Canada.
Supported by 3M Unitek and GeoDigm Corporation.
Reprint requests to: Dr Paul W. Major, Faculty of Medicine and Dentistry, Room 4051B, Dentistry/Pharmacy Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2N8; e-mail, major@ualberta.ca.
Submitted, May 2004; revised and accepted, August 2004.
0889-5406/$32.00

doi:10.1016/j.ajodo.2004.08.023

tist accumulates in his or her career; (3) accuracy, efficiency, and ease of measurement of tooth and arch sizes and dental crowding; (4) accurate and simple diagnostic setups of various extraction patterns; (5) the ability to send virtual images anywhere in the world for instant referral or consultation as needed or for internet study clubs; and (6) objective rather than subjective model grading analysis for American Board of Orthodontics (ABO) certification.[1]

A review of the literature did not identify any studies that tested the clinical applicability of emodels vs plaster; however, previous studies showed that the dimensional accuracy of laser surface scanned digital models is within about 0.05 mm.[2-5] Several studies tested the accuracy of OrthoCAD vs plaster, but to date there have been no comparisons between emodels and plaster.[6-10]

The makers of emodel and OrthoCAD guard their secret proprietary methods of model fabrication. Although their products appear similar on the computer screen, the 2 companies have fundamental differences in their laser surface-scanning techniques. Emodel scans the surface of a complete plaster model, whereas OrthoCAD uses a "destructive scanning" process that takes many scans of a model in thin slices. This is repeated until the entire model has been shaved and scanned. Thus, the result is a typical OrthoCAD file of 3000 kilobytes (3 megabytes) because the internal aspects of the model are scanned and recorded—even though the internal information is completely unusable. An emodel file is about 800 kilobytes because of surface scanning only. Emodel uses software to "slice through" the image, whereas OrthoCAD actually slices through the model and images it. From a storage and data transfer standpoint, the smaller file size of the emodel is an advantage.

The purpose of this study was to compare the current gold-standard plaster model with the digital counterpart of emodel for the analysis of tooth sizes and occlusal relationships—specifically the Bolton analysis and the peer assessment rating (PAR) index and their components.

MATERIAL AND METHODS

Approval to conduct this study was obtained from the University of Alberta Health Research Ethics Board. The study sample comprised pretreatment diagnostic study models of 24 randomly selected subjects, subdivided with 3 subjects in each of 8 malocclusion categories. Statistics from a pilot study estimated that the total sample size of the main project should be 10 to 13 subjects with $\alpha = 0.05$ and power of 95%. However, because we wanted the subjects to represent the entire range of malocclusions treated by orthodontists, a sample size of 24 was chosen to allow adequate representation in each category.

The sample was selected from the initial patient records at the University of Alberta Orthodontic Clinic. Two hundred twenty-five records were categorized into the 8 groups, and 3 patients were selected by a random number generator from each group. Each subject was in the permanent dentition from first molar to first molar without orthodontic appliances. The 8 categories were based on the following criteria, which closely resemble the various categories in the Case Report Category Specifications of the ABO Information for Candidates Phase III clinical examination publication[11]: (1) Class I malocclusion with $<$4 mm of crowding in both arches (nonextraction); (2) Class I malocclusion with 4.1 to 8 mm of crowding in at least 1 arch (borderline extraction); (3) Class I malocclusion with $>$8.1 mm of crowding in at least 1 arch (extraction necessary); (4) Class II Division 1 malocclusion ($<$4 mm of crowding): FMA $\geq$30° or SNa G0-Gn angle $\geq$37°; (5) Class II Division 1 malocclusion ($>$5 mm of crowding): mandibular arch length discrepancy that requires extraction of permanent teeth in at least the mandibular arch; (6) anteroposterior skeletal discrepancy: Class II malocclusion with (ANB angle $\geq$6°) or Class III malocclusion with ANB angle not less than $-2°$; (7) deep overbite malocclusion: overbite $\geq$100%, retroclined maxillary central incisors, and FMA $\leq$22° or SNa G0-Gn angle $\leq$29°, no specifications for crowding; (8) transverse discrepancy: posterior crossbite malocclusion that requires complete appliance treatment; before treatment, at least 1 posterior quadrant must be in complete lingual or buccal crossbite.

The diagnostic study models of the subjects in this study were duplicated by taking alginate impressions of the models and pouring the impressions in plaster as soon as possible. No positive or negative bubbles in the plaster or digital models were present. The bite was recorded by hand with a wax wafer. When the duplicate models were properly trimmed, they were sent via overnight courier to GeoDigm for scanning into digital models. The same plaster models were returned via overnight courier so that they could be used for the direct-measurement part of the experiment. By using the same plaster cast for direct measurements and digital replications, distortion or variation among alginate impressions was avoided. Any variance in measurements between plaster and emodels would then be attributable to operator variation or an inherent distortion in the emodel image.

Three examiners working independently—a senior orthodontic resident as the primary examiner (D.R.S.)

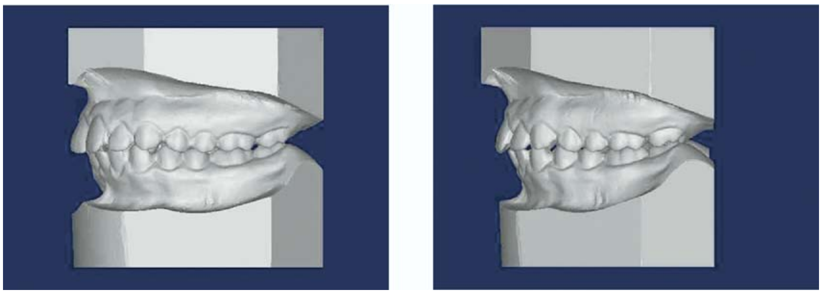

Fig. Emodel molar relationship rotated on z-axis. **A**, Patient with relatively solid Class I molar relationship when shown in emodel's preset left-buccal view. **B**, Same model appears to be in end-to-end molar relationship. Special care must be taken when assessing digital model for molar and canine relationships because slight rotation around z-axis can affect diagnosis of molar and canine relationships (still images from emodel software).

and 2 licensed orthodontists (B.N. and C.F-M.) with a minimum 5 years of orthodontic experience as secondary examiners—recorded the measurements on the plaster and digital models (each plaster and digital model was measured 3 times by the primary examiner and once by each secondary examiner). Measurements included tooth size from first molar to first molar in both arches, used for the Bolton analysis, and all measurements necessary for the PAR index. Rather than solely using a traditional incremental approximation for the PAR measurements (0-1 mm, 1-2 mm, 2-3 mm), actual measurements in millimeters along with approximations were recorded for contact displacements greater than 1 mm, overjet, overbite, deviation from ideal posterior interdigitation, and midline deviation on both casts. Maximum mesiodistal width was recorded for each tooth based on the expected contact point if the teeth were properly aligned. Overbite was measured as the greatest amount of maximum vertical overlap between a maxillary central and a mandibular central incisor. Overjet was also measured from the labial surface of the most anterior mandibular central incisor to the labial surface of the most anterior maxillary central incisor despite differences in labial inclinations of the maxillary incisors.

All plaster measurements were made with an electronic digital caliper (Sealey, Bury St Edmunds, United Kingdom) to the nearest 0.01 mm. Tooth size was measured on the digital models with the analysis tools from emodels (software version 6.0), to the nearest 0.01 mm. All teeth were measured from the direct occlusal view (by using the emodels' Auto Center feature) for consistency of object measurement between plaster and emodels. A 15-in LCD laptop screen with resolution of 1280 × 1024 pixels and 0.26-mm diagonal dot pitch with 32-bit color was used.

For severely malpositioned anterior teeth, the images were rotated on screen, and the measurements were made from the occlusal view to provide better visibility. For ease and accuracy of measurements, the images were enlarged on screen as needed by using the magnifying feature. Overjet and overbite were also measured with the analysis tools. Displacements in the posterior and midlines were measured with the "grid: 1 mm crosshair" feature. Posterior displacement from ideal interdigitation was assessed with a view perpendicular to the posterior quadrant because a slight rotation of the model quickly changed the operator's perception (Fig).

Reliability was considered as the extent to which a measurement was repeatable under identical conditions for the new diagnostic test (emodels) and the gold standard (plaster).[12] Validity was considered as the extent to which the new diagnostic test (emodels) measured against the gold standard (plaster).[12] A reproducibility index, called the concordance correlation coefficient (CCC), was introduced by Lin.[13] It evaluates the agreement between 2 readings (from the same sample) by measuring the variation for the 45° line through the origin (the concordance line).[13,14] CCC contains the measurements of accuracy and precision.

Statistical analysis

All measurements were recorded in a Microsoft Excel 2000 spreadsheet (Microsoft, Redmond, Wash) and analyzed with SPSS version 11.5 (SPSS, Chicago, Ill). Individual variables were graphed and assessed for normal distribution. Also, variables were evaluated to see whether they fulfilled normality (Shapiro-Wilks test) and homogeneity of variances (Levene test) criteria.

Reproducibility of measurement for intraexaminer and interexaminer measurements were tested with the CCC.[13-16] Also, a paired-samples *t* test was used to compare reliability between plaster and emodels for both

Table I. Intraexaminer and interexaminer* CCC to evaluate reproducibility of measurements between plaster and emodel

	Intraexaminer						*Interexaminer*					
	Plaster			*emodel*			*Plaster*			*emodel*		
		CI			*CI*			*CI*			*CI*	
Measurement	*CCC*	*Lower*	*Upper*	*CCC*	*Lower*	*Upper*	*CCC*	*Lower*	*Upper*	*CCC*	*Lower*	*Upper*
Bolton 6 (mm)	0.976	0.940	0.991	0.916	0.814	0.963	0.918	0.822	0.963	0.874	0.752	0.938
Bolton 12 (mm)	0.945	0.861	0.979	0.885	0.742	0.951	0.863	0.710	0.938	0.875	0.735	0.944
6 max length (mm)	0.992	0.989	0.995	0.972	0.939	0.987	0.965	0.926	0.983	0.963	0.909	0.985
6 man length (mm)	0.993	0.990	0.995	0.944	0.873	0.976	0.955	0.909	0.978	0.913	0.830	0.956
12 max length (mm)	0.993	0.981	0.997	0.963	0.929	0.980	0.948	0.912	0.969	0.949	0.912	0.971
12 man length (mm)	0.993	0.986	0.997	0.971	0.941	0.986	0.957	0.930	0.974	0.954	0.923	0.973
PAR (raw score)	0.943	0.881	0.973	0.916	0.827	0.960	0.920	0.817	0.966	0.897	0.778	0.954
PAR US (score)	0.939	0.870	0.972	0.868	0.723	0.939	0.876	0.750	0.941	0.731	0.489	0.868
PAR UK (score)	0.964	0.922	0.984	0.872	0.734	0.941	0.911	0.814	0.958	0.763	0.538	0.886
Overjet (mm)	0.980	0.956	0.991	0.990	0.978	0.996	0.887	0.779	0.944	0.878	0.770	0.937
Overbite (mm)	0.891	0.774	0.950	0.984	0.963	0.993	0.797	0.604	0.902	0.957	0.895	0.983
Overbite quantitative (mm)	0.871	0.719	0.943	0.823	0.633	0.920	0.856	0.703	0.934	0.887	0.780	0.944
Total contact displacement (mm)	0.859	0.708	0.935	0.803	0.614	0.905	0.803	0.591	0.911	0.756	0.533	0.881
Post contact displacement (mm)	0.798	0.595	0.905	0.808	0.619	0.909	0.740	0.510	0.871	0.776	0.572	0.889
Ant contact displacement (mm)	0.902	0.795	0.954	0.853	0.701	0.931	0.807	0.604	0.912	0.761	0.520	0.890
Total contact displacement (n)	0.746	0.539	0.868	0.813	0.625	0.912	0.796	0.559	0.913	0.699	0.470	0.840
Post contact displacement (n)	0.618	0.331	0.800	0.698	0.439	0.850	0.699	0.458	0.844	0.604	0.358	0.772
Ant contact displacement (n)	0.873	0.747	0.939	0.903	0.783	0.958	0.822	0.606	0.925	0.816	0.603	0.921
Right interdig quantitative (mm)	0.649	0.349	0.828	0.785	0.591	0.893	0.561	0.230	0.775	0.589	0.302	0.779
Left interdig quantitative (mm)	0.728	0.465	0.873	0.776	0.571	0.890	0.569	0.235	0.782	0.665	0.385	0.832
Right interdig (mm from "good")	0.826	0.639	0.921	0.830	0.687	0.912	0.427	0.144	0.646	0.739	0.490	0.876
Left interdig (mm from "good")	0.795	0.573	0.908	0.853	0.700	0.931	0.541	0.313	0.710	0.756	0.517	0.886
Centerline (mm)	0.822	0.647	0.915	0.971	0.929	0.988	0.671	0.430	0.822	0.728	0.466	0.872
Centerline quantitative (mm)	0.938	0.845	0.976	0.591	0.261	0.797	0.782	0.557	0.900	0.573	0.224	0.792
Post x-bite severity (mm)	0.991	0.982	0.996	0.988	0.973	0.994	0.975	0.946	0.988	0.976	0.948	0.989
Ant x-bite severity (mm)	0.962	0.916	0.983	0.618	0.351	0.791	0.954	0.909	0.977	0.632	0.352	0.809

Values closest to 1.00 are most reproducible.
CI, Confidence interval; *Max,* maxillary; *man,* mandibular; *post,* posterior; *ant,* anterior; *interdig,* interdigitation; *x-bite,* crossbite.
*Time trial #2 of primary examiner randomly selected for comparison.

intraexaminer and interexaminer measurements. The validity of emodel measurements was assessed with a paired-samples *t* test to compare the mean emodel and plaster model measurements for the 5 trials of the 24 subjects. A nonparametric Kruskal-Wallis test was used to detect differences among the 8 malocclusion groups.

RESULTS

Reproducibility of digital models via CCC

Intraexaminer and interexaminer (Table I) reproducibility of the digital models with the CCC was generally high; the average intraexaminer CCCs were 0.923 (range, 0.618-0.993) for plaster and 0.882 (range, 0.591−0.990) for emodels; the interexaminer CCCs were 0.851 for plaster (range, 0.427-0.975) and 0.835 (range, 0.573-0.976) for emodels. In 22 and 23 of the 26 CCCs, the values exceeded 0.750 for intraexaminer plaster and emodels, respectively. In 19 and 17 of the 26 CCCs, the measurements exceeded 0.750 for interexaminer plaster and emodels, respectively. No CCC value was lower than 0.400. The highest CCC values for intraexaminer measurements were for maxillary 12 lengths and mandibular 6 and 12 lengths for plaster (CCC = 0.993) and overjet measurements on emodels (CCC = 0.990). The lowest CCC values for intraexaminer measurements were for the number of posterior contact displacements counted for plaster (CCC = 0.618) and quantitative centerline measurements on emodels (CCC = 0.591). The highest CCC values for interexaminer measurements were for posterior crossbite assessment for plaster (CCC = 0.975) and emodels (CCC = 0.976). The lowest CCC values for interexaminer measurements were for the right quadrant measurement from "good interdigitation" for plaster (CCC = 0.427) and quantitative centerline measurements on emodels (CCC = 0.573).

Table II. Intraexaminer: average mean of absolute difference of repeated measurements

	Descriptives				*Descriptives*				*Paired samples* t *test*	
	Plaster ave (diff)				*emodel ave (diff)*				*Diff (±) plaster (−) emodel*	
Measurement	*Mean*	*SD*	*Range (mn/mx)*		*Mean*	*SD*	*Range (mn/mx)*		*Diff (+ value = plaster is larger)*	P *value*
Bolton 6 (mm)	0.32	0.17	0.04	0.78	0.69	0.32	0.23	1.31	−0.38	.000
Bolton 12 (mm)	0.58	0.37	0.10	1.37	1.08	0.51	0.39	1.99	−0.50	.001
6 max length (mm)	0.31	0.16	0.04	0.62	0.58	0.38	0.09	1.47	−0.27	.001
6 man length (mm)	0.21	0.12	0.04	0.57	0.62	0.38	0.05	1.41	−0.41	.000
12 max length (mm)	0.51	0.28	0.03	1.27	1.13	0.67	0.11	2.27	−0.62	.000
12 man length (mm)	0.48	0.27	0.07	1.19	1.07	0.53	0.23	2.10	−0.59	.000
PAR (raw score)	2.42	1.58	0.67	6.00	2.92	1.47	0.67	6.67	−0.50	.307
PAR US (score)	2.83	2.04	0.00	6.67	3.72	2.97	0.00	12.00	−0.89	.201
PAR UK (score)	2.69	2.03	0.67	6.67	4.56	4.00	0.67	14.67	−1.86	.047
Overjet (mm)	0.49	0.31	0.08	1.29	0.25	0.29	0.02	1.32	0.24	.000
Overbite (mm)	0.47	0.57	0.06	2.85	0.20	0.21	0.03	0.98	0.27	.012
Overbite quantitative (mm)	0.17	0.41	0.00	1.33	0.28	0.48	0.00	2.00	−0.11	.444
Total contact displacement (mm)	3.07	1.95	0.95	8.61	3.42	3.14	0.49	12.89	−0.35	.524
Post contact displacement (mm)	2.36	1.27	0.75	6.46	2.16	1.61	0.42	6.23	0.20	.611
Ant contact displacement (mm)	1.33	1.14	0.15	4.85	1.80	1.56	0.26	7.67	−0.47	.046
Total contact displacement (n)	1.64	0.88	0.67	4.00	1.53	0.77	0.00	3.33	0.11	.597
Post contact displacement (n)	1.50	0.82	0.00	3.33	1.11	0.70	0.00	2.67	0.39	.027
Ant contact displacement (n)	0.61	0.59	0.00	2.00	0.69	0.42	0.00	1.33	−0.08	.479
Right interdig quantitative (mm)	0.33	0.39	0.00	1.33	0.22	0.32	0.00	0.67	0.11	.162
Left interdig quantitative (mm)	0.22	0.38	0.00	1.33	0.19	0.31	0.00	0.67	0.03	.788
Right interdig (mm from "good")	0.38	0.45	0.00	1.45	0.43	0.41	0.00	1.67	−0.05	.608
Left interdig (mm from "good")	0.37	0.50	0.00	2.19	0.40	0.43	0.00	1.33	−0.04	.791
Centerline (mm)	0.33	0.57	0.00	2.89	0.14	0.17	0.00	0.33	0.19	.090
Centerline quantitative (mm)	0.06	0.19	0.00	0.67	0.19	0.37	0.00	1.33	−0.14	.057
Post x-bite severity (mm)	0.06	0.19	0.00	0.67	0.08	0.23	0.00	0.67	−0.03	.575
Ant x-bite severity (mm)	0.08	0.23	0.00	0.67	0.33	0.71	0.00	2.67	−0.25	.083

Max, Maxillary; *man,* mandibular; *post,* posterior; *ant,* anterior; *interdig,* interdigitation; *x-bite,* crossbite; *mn,* minimum; *mx,* maximum; *ave,* average; *diff,* difference.

Reliability of digital models via the average mean of absolute differences of repeated measurements

Most of the mean differences in tooth-size measurements were not statistically significant ($P > .0021 = .05/24$). The individual mean tooth measurement differences between measurement values and standard deviations for repeated intraexaminations and interexaminations were significant at $\alpha = 0.05$ (individual results not shown). However, it was necessary to adjust the significance level to control overall probability of making an incorrect decision, because 24 teeth were compared simultaneously. Other significance level α's were adjusted according to the group of measurements. The average intraexaminer differences of tooth-width measurements were 0.10 mm (range, 0.05-0.18 mm) for plaster and 0.19 mm (range, 0.14-0.24 mm) for emodels. The average interexaminer differences of tooth-width measurements were 0.17 mm (range, 0.12-0.24 mm) for plaster and 0.22 mm (range, 0.17-0.33 mm) for emodels.

Intraexaminer differences in repeated tooth-width measurements between plaster and emodels resulted in mostly statistically insignificant *P* values ($P < .0021 = .05/24$) except tooth numbers 14, 11, 22, 35, 45, 44, 42, and 41, all of which were premolars and incisors.

Paired-samples *t* tests showed that intraexaminer (Table II) Bolton 6 ratios between plaster and emodels differed by an average of −0.38 mm ($P < .001$) and Bolton 12 ratios differed by −0.50 mm ($P = .001$); the average difference was slightly larger in the digital format. The difference between repeated plaster and digital models' sums of lengths for maxillary 6, mandibular 6, maxillary 12, and mandibular 12 dentition lengths were statistically significant (*P* values ranging from <.000 to .001). Interexaminer (Table III) Bolton 6 ratios differed by an average of −0.36 mm ($P = .001$) and Bolton 12 ratios differed by −0.22 mm ($P = .185$).

A paired-samples *t* test for intraexaminer (Table II) results showed a statistically significant difference in reliability of overjet measurement between the plaster

Table III. Interexaminer*: average mean of absolute difference of repeated measurements

	Descriptives				Descriptives				Paired samples t test	
	Plaster ave (diff)				emodel ave (diff)				Diff (±) plaster (−) emodel	
Measurement	Mean	SD	Range (mn/mx)		Mean	SD	Range (mn/mx)		Diff (+ value = plaster is larger)	P value
Bolton 6 (mm)	0.49	0.27	0.14	1.21	0.86	0.42	0.29	1.67	−0.36	.001
Bolton 12 (mm)	0.85	0.48	0.15	1.95	1.06	0.59	0.30	2.49	−0.22	.185
6 max length (mm)	0.64	0.33	0.09	1.41	0.68	0.40	0.12	1.67	−0.04	.700
6 man length (mm)	0.60	0.21	0.23	1.13	0.80	0.39	0.13	1.60	−0.21	.047
12 max length (mm)	1.38	0.63	0.53	3.08	1.41	0.64	0.33	2.79	−0.03	.872
12 man length (mm)	1.25	0.61	0.35	3.57	1.37	0.56	0.25	2.31	−0.12	.477
PAR (raw score)	2.92	1.80	0.67	7.33	2.97	2.05	0.00	8.00	−0.06	.917
PAR US (score)	4.14	2.17	0.67	8.00	5.64	4.51	0.67	14.67	−1.50	.082
PAR UK (score)	4.33	2.73	0.67	10.67	6.36	5.37	0.67	17.33	−2.03	.045
Overjet (mm)	0.85	0.82	0.15	3.13	0.93	0.98	0.03	4.01	−0.08	.676
Overbite (mm)	0.78	0.71	0.03	3.35	0.38	0.27	0.05	1.12	0.40	.014
Overbite quantitative (mm)	0.25	0.33	0.00	0.67	0.25	0.33	0.00	0.67	0.00	1.000
Total contact displacement (mm)	4.05	2.34	0.37	9.61	4.30	3.22	0.69	14.87	−0.25	.764
Post contact displacement (mm)	2.76	1.24	0.51	4.87	2.59	1.49	0.59	6.74	0.17	.682
Ant contact displacement (mm)	2.28	1.63	0.39	7.21	2.37	2.13	0.52	8.13	−0.10	.865
Total contact displacement (n)	1.53	0.77	0.00	3.33	1.89	1.12	0.67	5.33	−0.36	.158
Post contact displacement (n)	1.33	0.59	0.67	2.67	1.36	0.77	0.00	2.67	−0.03	.892
Ant contact displacement (n)	0.78	0.70	0.00	2.67	0.81	0.81	0.00	3.33	−0.03	.896
Right interdig quantitative (mm)	0.39	0.44	0.00	1.33	0.31	0.34	0.00	0.67	0.08	.450
Left interdig quantitative (mm)	0.42	0.33	0.00	0.67	0.25	0.33	0.00	0.67	0.17	.083
Right interdig (mm from "good")	0.85	1.03	0.00	3.63	0.60	0.41	0.00	1.33	0.26	.180
Left interdig (mm from "good")	0.77	0.95	0.00	3.65	0.53	0.52	0.00	1.67	0.25	.224
Centerline (mm)	0.51	0.71	0.00	3.23	0.44	0.55	0.00	2.67	0.07	.645
Centerline quantitative (mm)	0.19	0.31	0.00	0.67	0.25	0.38	0.00	1.33	−0.06	.539
Post x-bite severity (mm)	0.14	0.34	0.00	1.33	0.11	0.32	0.00	1.33	0.03	.328
Ant x-bite severity (mm)	0.11	0.25	0.00	0.67	0.44	0.73	0.00	2.67	−0.33	.020

Max, Maxillary; *man,* mandibular; *post,* posterior; *ant,* anterior; *interdig,* interdigitation; *x-bite,* crossbite; *mn,* minimum; *mx,* maximum; *ave,* average; *diff,* difference.
*Time trial #2 of primary examiner randomly selected for comparison.

and digital models (0.24 mm, P <.0001). Overbite differences were also statistically significant, with 20 of the 24 digital overbite measurements smaller than plaster; the mean difference was 0.27 mm (P =.012). Because the PAR index scores ranged from a mean difference of raw PAR score = 0.50 (range, 0.00-4.67), United States weighted PAR score = 0.89 (range, 0.00-7.33), and United Kingdom weighted PAR score = 1.86 (range, 0.00-2), the ranges are taken from our raw data set not reported in this article but included here for comparison capability. Interexaminer (Table III) results were similar.

Validity of digital models

Mean plaster and emodels tooth-size measurement differences for the 5 time trials combined (the difference between the measurements means for plaster and emodels) were 0.01 to 0.21 mm. The greatest mean difference was found for the maxillary left central incisor (0.21 mm). One half of the tooth-size measurements were statistically significant, with plaster measurement means almost equally both larger and smaller than emodel measurement means (individual results not shown).

Bolton 6 and Bolton 12 ratios between plaster and emodels were not significantly different (P = .790 and P = .084, respectively). The difference between repeated plaster and digital models' sums of lengths for the maxillary and mandibular 6 teeth (0.59 and 0.40 mm) were statistically significant (P <.0125 [.05/4]) with P values of <.000 and .004, respectively. Maxillary and mandibular 12 dentition length differences (both 0.20 mm) were statistically insignificant (P > .0125 [.05/4]) with P values of .226 and .256, respectively (Table IV).

A paired-samples t test did not identify statistically significant differences (P >.0167 [.05/3]) for PAR measurements. Overjet differences of mean measure-

Table IV. Measurement means for 5 time trials per category

	Descriptives				Descriptives				Paired-samples t test		Ave of absolute differences (Pl−Em)	
	Plaster mean				emodel mean				Diff (±) plaster (−) emodel			
Measurement	Mean	SD	Range (mn/mx)		Mean	SD	Range (mn/mx)		Diff (+ value = plaster is larger)	P value	Mean	SD
Bolton 6 (mm)	−0.51	1.80	−6.41	2.79	−0.55	2.00	−7.39	2.55	0.04	.790	0.60	0.38
Bolton 12 (mm)	−0.37	2.20	−6.22	2.92	−0.75	2.64	−7.99	3.34	0.38	.084	0.92	0.58
6 max length (mm)	46.15	3.22	39.01	51.73	45.56	3.19	38.39	50.26	0.59	.000	0.69	0.52
6 man length (mm)	36.23	2.32	27.96	39.28	35.84	2.34	28.16	38.77	0.40	.004	0.61	0.40
12 max length (mm)	94.78	5.33	79.95	105.11	94.58	5.25	80.89	105.00	0.20	.226	0.69	0.43
12 man length (mm)	86.96	5.17	70.80	94.24	87.16	5.44	70.92	95.34	−0.20	.256	0.65	0.55
PAR (raw score)	25.08	9.30	10.80	52.20	25.91	8.79	10.20	48.00	−0.83	.128	2.11	1.62
PAR US (score)	27.03	10.39	12.00	56.20	26.98	9.77	10.40	54.20	0.04	.941	2.14	1.66
PAR UK (score)	30.84	13.52	13.60	70.80	30.77	12.60	14.20	69.20	0.08	.906	2.43	1.82
Overjet (mm)	4.90	2.97	1.82	13.33	4.91	2.98	2.06	13.70	−0.01	.884	0.33	0.21
Overbite (mm)	3.96	1.75	0.40	6.54	3.67	1.82	0.13	6.68	0.30	.001	0.38	0.27
Overbite quantitative (mm)	1.23	0.92	0.00	3.00	1.31	1.00	0.00	3.00	−0.08	.144	0.18	0.21
Total contact displacement (mm)	20.99	7.47	6.20	34.97	23.70	7.81	5.89	40.28	−2.71	.003	3.70	3.05
Post contact displacement (mm)	10.21	4.21	3.33	19.14	11.10	4.59	2.84	21.17	−0.89	.108	2.11	1.70
Ant contact displacement (mm)	10.78	4.53	2.87	19.72	12.61	4.84	3.05	24.09	−1.83	.000	1.98	1.85
Total contact displacement (n)	10.71	2.67	4.80	16.80	11.57	2.86	3.80	15.60	−0.86	.006	1.31	0.95
Post contact displacement (n)	5.29	1.78	2.00	9.20	5.61	1.70	1.60	8.80	−0.32	.189	1.00	0.61
Ant contact displacement (n)	5.42	1.84	2.40	9.80	5.96	2.02	2.20	9.80	−0.54	.000	0.68	0.42
Right interdig quantitative (mm)	0.98	0.59	0.00	2.00	1.01	0.58	0.00	2.00	−0.03	.682	0.28	0.27
Left interdig quantitative (mm)	0.78	0.58	0.00	1.80	0.84	0.57	0.00	2.00	−0.06	.475	0.29	0.26
Right interdig (mm from "good")	1.64	1.05	0.00	3.35	1.61	1.09	0.00	3.30	0.03	.782	0.34	0.31
Left interdig (mm from "good")	1.28	1.12	0.00	3.29	1.38	1.11	0.00	3.30	−0.10	.411	0.45	0.38
Centerline (mm)	1.32	1.10	0.00	3.45	1.23	1.04	0.00	3.90	0.10	.300	0.34	0.28
Centerline quantitative (mm)	0.61	0.63	0.00	2.00	0.45	0.46	0.00	1.40	0.16	.007	0.21	0.22
Post x-bite severity (mm)	0.74	1.84	0.00	6.00	0.75	1.86	0.00	6.00	−0.01	.747	0.04	0.12
Ant x-bite severity (mm)	0.67	1.09	0.00	3.20	0.63	0.98	0.00	3.00	0.03	0.590	0.15	0.26

Max, Maxillary; *man,* mandibular; *post,* posterior; *ant,* anterior; *interdig,* interdigitation; *x-bite,* crossbite; *mn,* minimum; *mx,* maximum; *ave,* average; *diff,* difference; *Pl,* plaster; *Em,* emodel.

ments were not statistically significant; however, overbite differences of mean measurements were statistically significant ($P < .025$ [.05/2]) with a P value of .001 and a difference of 0.30 mm, with plaster measurements being larger. Digital model measurements for anterior crowding (total contact displacement in millimeters) were statistically significantly larger (−2.71 mm, $P = .003$) (Table IV).

A nonparametric Kruskal-Wallis test for all measurements showed no statistically significant difference among the 8 malocclusion groups for any measurement (Table V).

DISCUSSION

Reproducibility of digital model measurements

The CCC[13-16] showed that all 50 intraexaminer measurements had excellent reproducibility for both plaster and emodels except 7 measurements (4 plaster and 3 emodels), which were good. The CCC was used because the intraclass correlation coefficient (ICC) measures reliability under the model of equal marginal distributions; however, when the marginal distributions are not equal (inaccurate), the ICC captures the deviations and considers those unreliable. Furthermore, the commonly used Pearson correlation coefficient provides only an expression of the linearity between values for repeated measures if the data are plotted while ignoring the inaccuracy component. In contrast, the CCC can segregate inaccuracy from unreliability.[16]

To generalize an acceptable level of orthodontic reliability, Roberts and Richmond[12] suggested that an ICC value for R below 0.4 is poor reliability, between 0.4 and 0.75 is fair to good, and above 0.75 is excellent. Lin et al[16] stated that the CCC, ICC, and Pearson correlation coefficient depend largely on the analytical range and the intrasample variation and can be compared as long as they have similar clinical interpretations. Thus, we applied the ranges of Roberts and

Table V. Nonparametric Kruskal-Wallis test *P* values to detect differences among 8 groups (with differences of repeated measurement means for all 5 time trials per category)

		Range		
Measurement	*Mean*	*Mn*	*Mx*	P *values*
Bolton 6 (mm)	0.04	−1.15	1.42	.725
Bolton 12 (mm)	0.38	−1.77	2.38	.210
6 max length (mm)	0.59	−0.40	2.00	.504
6 man length (mm)	0.40	−0.90	1.49	.217
12 max length (mm)	0.20	−1.05	1.72	.191
12 man length (mm)	−0.20	−2.11	1.47	.049
PAR (raw score)	−0.83	−5.20	4.20	.448
PAR US (score)	0.04	−3.60	8.00	.126
PAR UK (score)	0.08	−5.60	6.00	.275
Overjet (mm)	−0.01	−0.84	0.53	.521
Overbite (mm)	0.30	−0.44	1.02	.797
Overbite quantitative (mm)	−0.08	−0.80	0.40	.591
Total contact displacement (mm)	−2.71	−11.62	5.00	.678
Post contact displacement (mm)	−0.89	−5.89	5.17	.755
Ant contact displacement (mm)	−1.83	−5.72	1.04	.517
Total contact displacement (n)	−0.86	−3.60	2.20	.690
Post contact displacement (n)	−0.32	−2.40	2.20	.535
Ant contact displacement (n)	−0.54	−1.40	0.80	.694
Right interdig quantitative (mm)	−0.03	−0.60	1.00	.265
Left interdig quantitative (mm)	−0.06	−1.00	0.60	.142
Right interdig (mm from "good")	0.03	−0.85	1.01	.075
Left interdig (mm from "good")	−0.10	−1.20	1.24	.174
Centerline (mm)	0.10	−0.80	0.94	.900
Centerline quantitative (mm)	0.16	−0.20	0.80	.456
Post x-bite severity (mm)	−0.01	−0.40	0.40	.386
Ant x-bite severity (mm)	0.03	−0.80	0.80	.066

Max, Maxillary; *man,* mandibular; *post,* posterior; *ant,* anterior; *interdig,* interdigitation; *x–bite,* crossbite; *mn,* minimum; *mx,* maximum.

Richmond.[12] The 5 sets of measurements made by the 3 independent examiners were found to be statistically correlated, both for the plaster and digital models, via the CCC, indicating good to mostly excellent intraexaminer and interexaminer reliability for both media.

In this study, we did not detect a clinically significant difference in reliability between plaster and emodel tooth-size or PAR index measurements. Among the intraexaminer tooth–size measurement differences between plaster and emodels, two thirds resulted in statistically insignificant *P* values; however, all measurement differences were clinically insignificant (range, 0.01-0.16 mm). Interexaminer results showed that no tooth-size measurement difference was statistically or clinically significant (mean measurement differences ranged from 0.00 to 0.09 mm).

Bolton 6 measurements fell within a clinically insignificant mean difference of intraexaminer = −0.38 and interexaminer = −0.36, again showing that the digital measurement was less reliable; this was to be expected with individual tooth widths. There were no other consistent correlations of this type among the other measurements in this study.

There appears to be no clinically relevant difference in reliability between digital and plaster models. Although the ranges of the United States and United Kingdom PAR weighted scores are higher, this is due to the multiplication factor placed on various aspects of each score. The important fact is that the raw PAR score varies only by 0.5 of a point (not clinically significant) on average, with a range of less than 6 points for either plaster or emodels. Interexaminer results were similar.

Aside from tooth-size and related Bolton analysis results, no correlations among the other measurements were consistently higher or lower for plaster and emodels.

Validity of digital model measurements

No difference of mean measurement was clinically significant for Bolton 6 and 12 (0.04 and 0.38 mm, respectively), any arch length (range, 0.20-0.59 mm), PAR score (range, 0.04-0.83), overjet (0.01 mm), or overbite (0.30 mm). Emodels are clinically valid for study-model measurement of Bolton, PAR, and their constituents.

Santoro et al[8] reported OrthoCAD digital tooth-width measurements that were always smaller. Our study did not identify a consistent tooth-width bias with digital models, but the range of differences was similar to the results of Santoro et al in that they were clinically insignificant.

Possible reasons for the differences between this study (emodels) and that of Santoro et al[8] (OrthoCAD) include: (1) OrthoCAD and emodels might have different distortions in the fabrication process; (2) there was a minimum number of repeated time trials in their study (1 time trial by each of 2 examiners—negating the possibility of intraexaminer analysis); (3) potential operator differences when clicking the mouse pointer on tooth locations for width; and (4) differences in model details between OrthoCAD and emodel software around tooth borders or contact points to see an accurate point of measurement.

A nonparametric Kruskal-Wallis test for all measurements showed no statistically or clinically significant differences among the 8 groups.

Any difference between plaster and digital models in this study cannot be attributed to alginate impression distortion because all digital models were made from the same plaster casts used in this study. If one assumes that the digital model is accurate (a near-exact replica of a plaster model) in size, the most likely explanation for the difference is that digital models result in more valid measurements than plaster because there is no physical barrier of the caliper dictating placement of measurement points; however, this also allows one to click the mouse pointer either within or on the outside surface of the teeth. As long as a careful measuring point is selected on the computer screen, it would be reasonable to believe that digital measurements are more valid than those made by calipers on plaster.

Another contributing factor to the difference between plaster and emodel measurement might be the operator's learning curve in precisely measuring with the computer mouse on the screen. Once this was learned, it was easier to measure on the computer screen.

Slight differences in the measurements for overjet and overbite are most likely attributed to the intimate contact points achieved on the digital image compared with the bulky digital caliper on a plaster cast, but the differences had no clinical significance (<0.30-mm difference).

Quimby et al[10] used "accuracy" as a synonym for validity in a comparison between computer-based digital models and a dentoform model (gold standard). According to Colton,[17] "accuracy" encompasses both unbiasedness (the tendency to arrive at the true or correct value) and precision (the spread of a series of observations). Thus, medical data can be unbiased but imprecise, and vice versa. The in-vitro approach of Quimby et al[10] of using a dentoform to evaluate accuracy was useful. Plaster models were considered in our in-vivo study as the gold standard to evaluate only validity because it is ethically and clinically impossible to obtain the exact tooth measurements (in which case accuracy could also be evaluated) to be used as the gold standard.

According to Quimby et al,[10] the validity and the reliability of digital models (OrthoCad) were clinically acceptable. Our results support their findings with a different digital model system (emodels).

Limitations of emodels

Digital models present several unique challenges compared with plaster models. Because the 3D computer image is displayed on a 2D screen, the greatest challenge was observing crossbites. Especially in the posterior, teeth can falsely appear in crossbite on the screen. Or they will seem to have a positive overjet in the posterior segment when they really do not. This phenomenon seems to depend on the amount of zoom and rotation; however, we found that a standard preset view of the buccal or anterior segment was correct after rotating the model in various positions for verification. If there is any question about crossbite, it can easily be checked with the vertical or horizontal cross section function, but this is somewhat time-consuming and frustrating.

Details for midlines, occlusal anatomy, and wear facets are not as clear on the digital emodel. However, for a deep overbite, it is easier and more reliable to check the midline and measure overbite and overjet by using the software's cross-sectioning tool.

It is more difficult to quantify the precise interdigitation of a digital model than it is with plaster. Maybe this will improve with future software releases, but the pictures on the screen seem to show more openbites than their plaster counterparts. This could be a function of zoom. More zoom exaggerates things that would not normally be noticeable in plaster. This phenomenon might be good or bad—but it should be acknowledged so the viewer can learn to separate a clinically significant problem from an insignificant but magnified problem. It is possible that emodels will allow clinicians to see imperfections and improvee other aspects of a patient's malocclusion; this in turn will raise the bar in the clinical outcomes of patient treatment.

The practitioner should be aware of the potential for distortion in the shipment of alginate impressions when using emodels as the company suggests. Coleman et al[18] found significant dimensional changes between dental models poured within 1 hour of the hydrocolloid impression compared with pouring 24 hours later. Obviously, this would be translated into the digital image. Further studies are necessary to determine this important aspect of distortion with alginate impressions, which could negate some or all of our results. Disinfection protocols could also alter the impressions after some time.[19-21]

The average age of the examiners in this study was 33.3 (range, 31-38). The learning curve for the product was steep and short-lived. As with most computer-related activities, those more familiar with computers in general will probably experience a shorter learning curve with emodels and be able to achieve precise measurements without complications.

CONCLUSIONS

1. No measurement associated with Bolton analysis or PAR index made on plaster vs digital models showed a clinically significant difference.
2. Digital models are a clinically acceptable replacement for plaster casts for the routine measurements made in most orthodontic practices.
3. Because the PAR analysis and its constituent measurements are not significantly different clinically between plaster and emodel media, and preliminary

results gave no indication that digital models would cause an orthodontist to make a different diagnosis of malocclusion than with plaster models, digital models are not a compromised choice for treatment planning and diagnosis.

REFERENCES

1. Harrell WE Jr, Hatcher DC, Bolt RL. In search of anatomic truth: 3-dimensional digital modeling and the future of orthodontics. Am J Orthod Dentofacial Orthop 2002;122:325-30.
2. Kuroda T, Motohashi N, Tominaga R, Iwata K. Three-dimensional dental cast analyzing system using laser scanning. Am J Orthod Dentofacial Orthop 1996;110:365-9.
3. Motohashi N, Kuroda T. A 3D computer-aided design system applied to diagnosis and treatment planning in orthodontics and orthognathic surgery. Eur J Orthod 1999;21:263-74.
4. Sohmura T, Kojima T, Wakabayashi K, Takahashi J. Use of an ultrahigh-speed laser scanner for constructing three-dimensional shapes of dentition and occlusion. J Prosthet Dent 2000;84:345-52.
5. Kusnoto B, Evans CA. Reliability of a 3D surface laser scanner for orthodontic applications. Am J Orthod Dentofacial Orthop 2002;122:342-8.
6. 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.
7. Garino F, Garino GB. Comparison of dental arch measurements between stone and digital casts. World J Orthod 2002;3:250-4.
8. 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.
9. 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.
10. 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.
11. American Board of Orthodontics. Information for Candidates "Orange Book" (May 2001). Available at: http://www.americanboardortho.com/professionals/road_to_cert/common/info/#. Accessed December 15, 2003.
12. 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.
13. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45:255-68.
14. Lin LI. Assay validation using the concordance correlation coefficient. Biometrics 1992;48:599-604.
15. Lin L, Torbeck LD. Coefficient of accuracy and concordance correlation coefficient: new statistics for methods comparison. PDA J Pharm Sci Technol 1998;52:55-9.
16. Lin L, Hedayat A, Sinha B, Yang M. Statistical methods in assessing agreement: models, issues, and tools. J Am Stat Assoc 2002;97:257-70.
17. Colton T. Statistics in medicine. Boston: Little, Brown and Company; 1974.
18. Coleman RM, Hembree JH Jr, Weber FN. Dimensional stability of irreversible hydrocolloid impression material. Am J Orthod 1979;75:438-46.
19. Johnson GH, Chellis KD, Gordon GE, Lepe X. Dimensional stability and detail reproduction of irreversible hydrocolloid and elastomeric impressions disinfected by immersion. J Prosthet Dent 1998;79:446-53.
20. Boden J, Likeman P, Clark R. Some effects of disinfecting solutions on the properties of alginate impression material and dental stone. Eur J Prosthodont Restor Dent 2001;9:131-5.
21. Taylor RL, Wright PS, Maryan C. Disinfection procedures: their effect on the dimensional accuracy and surface quality of irreversible hydrocolloid impression materials and gypsum casts. Dent Mater 2002;18:103-10.