Original article

Reliability and validity of measurements on digital study models and plaster models

Ralph Philip Reusch*,†, Wieland Heuer**,***,†, Meike Stiesch**, Daniela Wenzel**** and Marc Philipp Dittmer*****

Departments of *Orthodontics and **Prosthetic Dentistry and Biomedical Materials Science, Hannover Medical School, ***Private Practice, Detmold, ****Department of Biometry and *****Centre for Dental, Oral and Maxillofacial Medicine, Hannover Medical School, Germany, *****Private Practice, Sarstedt

Correspondence to: Marc Philipp Dittmer, Edith-Weyde-Str. 7, 31157 Sarstedt, Germany. E-mail: marc@drdittmer.de

† These authors contributed equally to the success of this study.

Summary

Objective: To compare manual plaster cast and digitized model analysis for accuracy and efficiency.

Material and methods: Nineteen plaster models of orthodontic patients in permanent dentition were analyzed by two calibrated examiners. Analyses were performed with a diagnostic calliper and computer-assisted analysis after digitization of the plaster models. The reliability and efficiency of different examiners and methods were compared statistically using a mixed model.

Results: Statistically significant differences were found for comparisons of all 28 teeth ($P < 0.001$), mandibular intermolar width (IMW, $P = 0.0453$), and overjet ($P < 0.001$ to $P = 0.0329$). Single-tooth measurements tended to have larger values when measured manually and the SD was between 0.06 and 1.33 mm. Digital analyses gave significantly higher values for mandibular IMW and overjet. Less time was needed for digital measurements.

Conclusion: Clinical significance of the differences between the methods compared did not appear significant. 3D laser-scanned plaster model analysis appears to be an adequate, reliable, and time saving alternative to analogue model analysis using a calliper.

Introduction

Impression taking and plaster model production are established methods in dentistry and plaster models serve as basis for documentation and diagnosis. In orthodontics, model analysis is used routinely and is a key factor to treatment planning and review of orthodontic progress. In recent years, digitalization techniques in dentistry have advanced and methods to digitalize plaster models into 3D virtual models have been established. In 1999, Align Technology Inc. (San Jose, California, USA) introduced OrthoCad, which is a digital model service, based on a proprietary scanning process of plaster models (1). Three years later, GeoDigm Corp. (Falcon Heights, Minnesota, USA) introduced ‘emodels’, a digitalization service for plaster models using non-destructive laser scanning (1). Several studies have been performed to compare digital model analyses with the so-called ‘gold standard’, which means the manual measurement of plaster models with a calliper. Measurements with OrthoCad models (2–4) or emodels (5–7) seem to be generally as precise and reliable as measurements made with plaster models. However, before 3D data are available, plaster models or the original impressions have to be sent to the provider and this is not cost efficient. For this reason, desktop laser scanners were developed to transfer the scanning process into practice or in-house laboratory. The reliability of these desktop scanners, e.g. those of 3Shape (Copenhagen, Denmark), has been evaluated by various authors, who found that digital models obtained by 3Shape scanners are reliable for measurements and the evaluation of dental arch relationships (8–10). However, the analytical software (Orthoanalyzer; 3Shape) of this system has not been tested yet. Besides the general advantages of virtual models, e.g. less storage space, lower costs, instant accessibility, and transfer anywhere in the world for instant referral or consultation (11), Orthoanalyzer software offers a workflow for a step-by-step analysis
of the models, followed by a tool for a diagnostic or therapeutic set-up. This set-up can be divided into substeps according to the amount of tooth movement. The resultant virtual models can then be plotted via additive methods or milled via subtractive manufacturing methods and then be used for the production of orthodontic auxiliaries or appliances.

Because of the advantages and the development of fast and accurate intraoral scanners, it is foreseeable that virtual models will replace plaster models in the medium term. Accordingly, the software for virtual model analysis will become more important and should deliver results that are at least as reliable and valid as those from plaster model analysis. Hence, the aim of this study was to compare traditional manual model analysis with an orthodontic caliper with the virtual model analysis of digitalized plaster models. Inter- and intra-individual reproducibility was also evaluated. The following hypothesis was formulated: digital model analysis is as reliable and measurements obtained are equally reliable compared to manual model analysis.

Materials and methods
Pre-treatment plaster model sets of 20 patients were randomly selected from patients seeking orthodontic treatment in the Hannover Medical School postgraduate program ‘Master of Science in Lingual Orthodontics’. All plaster models were made in the same laboratory, with similar time intervals from impression taking to model production. The criteria for inclusion were a fully erupted permanent dentition, similar shape of the models—with a base parallel to the occlusal plane—and full integrity of the model: cracked or damaged models were not used. Furthermore, models with bonded retainers, appliances, attachments, or prostheses were excluded. After applying these criteria, 19 models remained in the pool of cases. The patients’ ages ranged between 15 and 47 years; 9 male and 10 female volunteers had Class I (11 patients), Class II (6 patients), and Class III malocclusions (2 patients). Eleven of those 19 patients had also slight crowding in the anterior region.

Using G-Power (G-Power 3.1.9.2, Franz Faul, University of Kiel, Kiel, Germany), power and sample sizes were calculated. Power calculation for the digital and manual value groups of single-tooth measurements revealed that a sample size of 7 would have a power of 95 per cent to detect difference in means of 2 mm (e.g. a first condition mean [\(\mu_1\)] of 9.92 mm [\(\pm 0.13\) mm] and a second condition mean [\(\mu_2\)] of 7.92 mm [\(\pm 1.33\) mm]). Furthermore, power calculation for the single-tooth values of the inter-examiner groups revealed that a sample size of 18 would have a power of 80 per cent to detect difference in means of 0.19 mm (e.g. a first condition mean [\(\mu_1\)] of 5.79 mm [\(\pm 0.19\) mm] and a second condition mean [\(\mu_2\)] of 5.6 mm [\(\pm 0.2\) mm]).

Each model set was anonymized with a number and then digitalized using a 3D laser scanning device with two cameras with 5.0 megapixels each (D800; 3Shapedental, Copenhagen, Denmark). The accuracy of this scanner is given by the manufacturer as 15 microns. After separate scanning of the upper and lower jaw plaster model, a bite scan was performed in the wax-bite position, in order to ensure identical bites in the virtual and plaster model sets. The output format used for calculating the model was an open standard STL file of about 22–25 MB for each set of plaster models. The collected data were saved as STL files and imported to OrthoAnalyzer software (Build: 1.2.2.2; 3Shapedental) for further editing. To establish identical preconditions, the virtual model sets were oriented in the OrthoAnalyzer to the occlusal plane before measurements were performed. In-process magnification or zooming in the desired model area was allowed, in order to give maximal resolution.

Calibration of both examiners was achieved by collaboration in two sample cases. Measurements were directly compared and discussed until final definition. After the two examiners had been trained and calibrated, a standardized workflow was established for the manual and digital measurements to assure that the quality and time courses were similar for the two types of measurement.

Each model set was measured 3 times at intervals of 3 days by the 2 examiners. The following parameters were included:

1. Tooth widths were measured for teeth 17–27 and 37–47. The marks were the contact points between neighbouring teeth or the greatest mesiodistal diameter.
2. Upper jaw transversal width between palatal cusps of first molars (Max. IMW, i.e. intermolar width) and of canines (Max. ICW, i.e. intercanine width).
3. Lower jaw transversal width between central fossae of first molars (Mand. IMW) and of the canine cusps (Mand. ICW).
4. Overjet, measured from the mesiobuccal surface of the lower incisor to the mesiobuccal surface of the upper incisor parallel to the occlusal plane. The maximum distance detected was used (Overjet).
5. Overbite, measured from incisor edge to incisor edge in a 90° angle to the occlusal plane. The maximum distance detected was used (Overbite); see also Figures 1 and 2.
6. Midline discrepancy was measured from the contact point of the upper central incisors to the contact point of the lower central incisors (Midline Discrepancy).
7. Duration of time to take above measurements (Time Needed).

Distances were measured in millimetres (mm) and time was documented in seconds. Manual model analysis was performed with a diagnostic calliper (‘Zürcher Model’, Smile Dental®, D, article number: 06-0075) to an accuracy of one tenth (1/10) of a mm. No magnification was used for manual model analysis. All data were recorded in a standardized document to guarantee similar workflows. Model number and trial were documented. Digital values were saved to an accuracy of one hundredth (1/100) of a mm, as this is a standard feature of the software. No post-measurement rounding was operated.

Documented data were captured in a Microsoft Excel™ 2007 file and processed for descriptive statistics. Further statistical analysis with the aid of a mixed model was performed by the procedure implemented in SAS™ 9.3.

![Figure 1. Typical screenshot of a virtual model during analysis of the overbite.](https://academic.oup.com/ejo/article/38/1/22/2599883)
The null hypothesis was defined as no difference between the manual and digital methods and between evaluators. Standard deviation, mean, and coefficient of variation were calculated for each patient, examiner, and method, yielding four different sample groups for each patient. Intra-examiner and inter-examiner data were evaluated for the different methods. Means and standard deviation data were presented in scatter plot graphs to depict inter- and intra-examiner correlation.

$P$ values were calculated for each measured parameter. $P < 0.05$ was taken as statistically significant, corresponding to rejection of the null hypothesis. $P$ values were presented in a separate table, in order to emphasize statistically significant parameters.

### Results

#### Comparison of tooth widths

The overall comparison between the manual and digital methods shows that some results differed significantly ($P < 0.001$, see Table 1). There was a significant difference in inter-examiner variation ($P < 0.001$). There was a significant effect of the examiners in the manual method ($P < 0.001$), but not in the digital method ($P = 0.228$). The mean values for tooth width from the manual method were larger than the values from the digital method (0.1–2.5 mm). Examiner 2 gave larger values than Examiner 1 for both the digital and manual methods. The standard deviations of the width measurements were similar for the two methods. The SE of all combinations ranged from 0.0163 to 0.023 mm. With the digital method, both examiners found remarkably low values with tooth 27, which were almost 20 per cent lower than with the manual method, thus leading to a high overall standard deviation.

#### Comparison of transversal measurements (Max. IMW, Max. ICW, Mand. IMW, and Mand. ICW)

The transverse measurements were not significantly influenced by examiner or method, with the exception of a single value: Examiner 1 found significantly higher values for the mandibular IMW with manual measurements than with digital ones ($P = 0.0453$). The SE for the combined measurements was 0.1824 mm.

#### Comparison of overjet

The overall comparison found statistically significant differences between the examiners ($P < 0.001$) and method of measurement ($P = 0.0002$) (see Table 1). There were striking differences between the two examiners with respect to the mean values in the manual measurements (3.68 versus 5.11 mm). Furthermore, there were statistically significant differences between the digital results for examiner one and his manual results ($P < 0.001$), as well as to both the manual ($P < 0.001$) and digital ($P < 0.001$) results for Examiner 2. There were also significant differences between the digital results for Examiner 2 and the manual results for Examiner 1 ($P = 0.0329$).

### Comparison of overbite values

For overbite, there were no statistically significant differences with respect to method of measurement or examiner (Table 1).

#### Midline

For midline, there were no statistically significant differences with respect to method of measurement or examiner (Table 1).

#### Comparison of time needed for measurements

Time needed for analyses was statistically significant influenced by the method ($P < 0.0001$). Digital trials were 75–92 seconds faster than manual trials. A minimum time needed for digital measurements was 4:29 minutes, the maximum time was 10:54 minutes, whereas manual measurements took between 6:24 and 9:56 minutes. Thus, the bandwidth of time needed to complete a trial was lower for the manual method. Comparing the two examiners, no significance was detected ($P = 0.2370$). Both examiners showed the same mean value of 6:17 minutes for digital measurements.

### Discussion

The trend to virtual models in orthodontics is clear and different software programs are available to perform virtual model analyses. Several studies have compared different model analysis software programs and the conventional manual analysis (10, 12–14). In these studies, scanner and software were not provided by a single manufacturer, which makes the workflow from scanning to analysis more demanding. A simple workflow is available with the 3Shape system, which offers an intraoral scanner or extraoral laser scanning boxes as well as an analysis tool with the OrthoAnalyzer software. To the best of the authors’ knowledge, no study has evaluated the OrthoAnalyzer software yet. In the present study, this software was compared for the first time with the manual analyses with a caliper. The number of samples needed to give conclusive results was determined by a review of current literature and proofed by a power analysis. Watanabe et al. (15), Leifert et al. (16), Zilberman et al. (17), and Asquith et al. (18) selected between 10 and 25 models. According to the results of these studies, 19 models were calculated as an adequate number of samples.

Various options are available to obtain the data set for a full dental arch. The direct intraoral scanning method is already available and seems to have sufficient reliability for diagnostic purposes (19, 20) but is still not routinely used in dental practice. Furthermore, scanning conventional impressions directly by either computed tomography or with lasers has the advantage that it eliminates plaster models but has the potential disadvantage that correct bite registration can be challenging to achieve. The third and most common method for digitalization of dental arches is the scanning of plaster models by various scanning techniques (e.g. laser scanning, structured-light scanning) (21, 22). With a scan in the wax-bite position, a high correlation can be assumed between the plaster model set and the virtual model set. In the current study, a laser scanning method was therefore applied.
Table 1. Values, lower and upper limit of the 95% confidence interval for all combinations: comparison of all 28 teeth (Comp. all 28 teeth), maxillary intermolar width (Comp. Max. ICW), mandibular intermolar width (Comp. Mand. ICW), mandibular intercanine width (Comp. Mand. ICW), and midline discrepancy (Comp. Mand. Inter.). 

| Method of measurement | Examiner 1 versus Examiner 2 | Examiner 1 versus Man. | Examiner 2 versus Man. | Examiner 1 versus Dig. | Examiner 2 versus Dig. 
|-----------------------|-----------------------------|------------------------|------------------------|------------------------|------------------------
|                       | Lower                        | Upper                  | Lower                  | Upper                  | Lower                  | Upper                  |
| Comp. all 28 teeth    | -0.0001                      | 0.0001                 | -0.0001                | 0.0001                 | -0.0001                | 0.0001                 |
| Comp. Max. ICW        | -0.0001                      | 0.0001                 | -0.0001                | 0.0001                 | -0.0001                | 0.0001                 |
| Comp. Mand. ICW       | -0.0001                      | 0.0001                 | -0.0001                | 0.0001                 | -0.0001                | 0.0001                 |
| Comp. Mand. Inter.    | -0.0001                      | 0.0001                 | -0.0001                | 0.0001                 | -0.0001                | 0.0001                 |
| Dig., digital         | -0.0001                      | 0.0001                 | -0.0001                | 0.0001                 | -0.0001                | 0.0001                 |

As model analysis is a routine in an orthodontic practice and commonly performed with plaster casts, we deliberately chose a diagnostic analogue calliper (Zürcher Model) to conduct our measurements. This is less accurate than a digital calliper, although these differences are not clinically significant and diagnostic analogue callipers are widely used. Nevertheless, the advantages offered by the software included standard features like zoom, turn, and tilt of the models as well as the accuracy grade (1/100 mm).

Defining the correct landmark is a decisive issue when looking at reliability and intra- and inter-examiner validity. Standard deviations in molar regions increased significantly, but only for the digital method. This might be caused by difficulties in identifying tooth margins. Problems in setting correct landmarks on digital models have been reported by other authors (5, 23–26). Hunter and Priest (26) also found significantly larger values for posterior teeth but reported a reduction in variation as the technique was refined. Dalstra and Melsen (23) concur with these findings and suggest that there is a learning curve, especially for digital measurements, that would lead to minor deviations as the examiner becomes familiar with the technique. Although explicit emphasis was put on calibration of both examiners, there is evidence that Examiner 1 was not able to reproduce landmarks as precisely during digital trials as Examiner 2 or as the manual trials. Another possible explanation for variations and large standard deviations is the technique of digitalization. Undercuts of models, especially in approximate regions, can lead to scan shadows and thus to inaccuracy (15).

Single-tooth measurements are claimed to be highly precise and reliable. As our data show, measurements with an analogue calliper have better reliability than digital measurements. Analogue data reveal a smaller standard deviation for teeth in the posterior area (lowest SD 0.06 mm at tooth 34 and 42), whereas the front teeth show larger standard deviations (0.25 mm at tooth 12). In contrast, digital measurements are highly reproducible in the front area, up to the premolar region (lowest SD 0.1 mm at tooth 44), but show less accuracy in the molar region (highest SD 1.3 mm at tooth 27). Tooth 27 showed very high standard deviations throughout using the digital method, possibly indicating an inherent error within the software or in the handling of the software by the examiners. Data of manual measurements show the highest standard error for maxillary ICW (SE 2.73–3.89 mm), although this does not mean statistical significance. Moreover, digital measurements of mandibular ICW and especially overjet were comparatively inaccurate and tended to be larger than manual results. Asquith and McIntyre (8) defined systematic errors of more than 0.5 mm for single-tooth measurements and overjet or measurements with more than 5 per cent discrepancy from reference marks as clinically unacceptable. The results of the present study suggest that except for the variation on 27, the software delivered reliable results and should be able to replace plaster cast analysis. Nevertheless, validity of digital results are more critical than manual measurements as the quality of the digitalization process and interpretation of anatomic structures and predefined contact points lead to higher standard deviations. Furthermore, there is a risk of making type 1 error when several measurements are made with both methods.

Time taken to measure digital models was significantly lower than the manual method (P ≤ 0.0001). This is a decisive criterion for choosing a procedure for daily routine in an orthodontic practice. The potential time saving can nearly be 2 minutes per model.

While the difference in time to analyze models digitally or manually was significant, there was little inter-examiner variation, particularly for digital models: both examiners took 6:17 minutes, whereas...
analogue measurements took 7:32 minutes for Examiner 1 and 7:49 minutes for Examiner 2.

Conclusion

3D laser-scanned plaster model analysis appears to be an adequate and reliable alternative to the conventional method of model analysis with an analogue calliper and it appears to be more efficient. In spite of hard and software bias in determining the correct landmark, digital model analysis should be accurate enough for treatment planning. Discrepancies in individual tooth diameters and linear measurements were not clinically significant for most values.

Software advantages, such as embedded calculation of required data, can be used to optimize efficiency in orthodontic workflows.

Further emphasis should be placed on improving efficiency and efficacy of digital analysis; this should help practitioners to use digital techniques in an optimal manner during their daily work.

Acknowledgements

The plaster models used in the present study were provided by the Department of Orthodontics, Hannover Medical School, (Head: Prof. Schwestka-Polly), whose support is gratefully acknowledged. The Orthoanalyzer Software was made available by 3Shape, whose support is also gratefully acknowledged.

References