

University of Alberta

The Validity, Reliability and Time Requirement of
Study Model Analysis using
Cone-Beam Computed Tomography generated
Virtual Study Models

by

Nghe Sieu Luu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
Medical Sciences – Orthodontics

©Nghe Sieu Luu
Fall 2011
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Dedication

To mom and dad,

I am truly grateful for the tireless efforts and magnanimous sacrifices made to ensure every opportunity for success in my life. Your unfaltering love and guidance has enriched my upbringing while your impressive work ethic has inspired and encouraged me to persevere through decades of schooling. As the work herein is symbolic of my highest academic achievement thus far, it represents a mere portion of the fruits of my labor in life as yet. I pledge to never forget where I have come from, continue to make the best of everything I'm given, and I patiently look forward to the day when I can finally give back.

Your son,

劉超毅

Abstract

Objectives

To investigate the validity, reliability and time spent on performing a full orthodontic study model analysis (SMA) on Cone Beam Computed Tomography (CBCT)-generated dental models (Anatomodels) compared with conventional Plaster models and a subset of Extracted Premolars.

Methods

Timed SMA was performed on thirty retrospectively selected patient records. Five evaluators participated in the interrater reliability study and one evaluator for the intrarater reliability and validity studies. Agreement was assessed by ICC and crosstabulations while mean differences were investigated using paired-sample *t*-tests and repeated measures ANOVA.

Results

For all three modalities studied—Anatomodels, Plaster and Extracted Premolars—intrarater reliability was excellent, interrater reliability was moderate to excellent, validity was poor to moderate, and performing SMA on Anatomodels took twice as long as on Plaster.

Conclusions

SMA using CBCT-generated study models was reliable but not always valid and required more time to perform when compared with Plaster models.

Acknowledgements

Marie-Alice Mandich and Neelambar Kaipatur, thanks for your support and friendship over the past three years. Marie-Alice, it's been incredible learning with you, laughing with you, cooking with you, and enjoying life with you. Your energy is refreshing and I love that your heart is always in the right place. Neelambar, thanks for helping me with the transition to Edmonton. You're a reliable friend and a great confidant.

Jean-Marc Retrouvey and Liliya Nikolcheva, thanks for partaking in the article selection process of the systematic review. It was time consuming but hopefully a worthwhile process for all. Donald Taylor and Jean-Marc Retrouvey, thank you both for your mentorship in dental school and for fostering my interest in orthodontics; this thesis would not even exist if it weren't for you both.

To my thesis committee, thank you for ensuring the integrity of my work. Paul Major and Jason Carey, my supervisors, thanks for standing by me as I did what no graduate student ever wants to do—start a new thesis project half way into the program. Carlos Flores-Mir, thanks for giving me the foundation to conduct systematic reviews, it was indispensable for this thesis. Giseon Heo, thanks for teaching me the statistical tools to properly conduct and understand quality scientific research. Tarek El-Bialy, thanks for your encouraging support throughout this research endeavor, and Pierre Boulanger, thanks for your valuable feedback as an external examiner.

Many others deserve acknowledgement for going above and beyond the call of duty. Manuel Lagraverre, along with Carlos, Tarek and Marie-Alice, thank you for participating in the reliability studies. Susan Helwig, your whimsical spirit made my days in the clinic enjoyable and your assistance with coordinating this project was outstanding. To all of the support staff and fellow residents, too many to name individually, thanks for being instrumental at every step of my graduate training.

Last but not least, to my dear family—my father, Quang Luu; my mother, Mai Luu; and my brother, Brian Luu—through peaks and valleys, thank you ever so kindly for constantly being there for me.

Table of Contents

Chapter 1. Introduction	1
1.1 Background	1
1.2 Literature Review	2
1.2.1 Study Model Analysis	2
1.2.2 Virtual Study Models	3
1.2.3 Cone Beam Computed Tomography-generated virtual study models	3
1.2.4 Timed Study Model Analysis	6
1.3 Statement of the Problem and Rationale for Inquiry	7
1.4 Research Objectives	8
1.4.1 Reliability	8
1.4.2 Validity	9
1.4.3 Time	9
1.5 Hypotheses	9
1.5.1 Reliability	9
1.5.2 Validity	11
1.5.3 Time	11
1.6 References	12
Chapter 2. Linear Measurements using Virtual Study Models: A Systematic Review	15
2.1 Introduction	15
2.2 Methods	17
2.2.1 Search strategy	17
2.2.2 Selection of articles	17

2.2.2.1	Screening of articles from electronic databases	18
2.2.2.2	Assessment of entire articles from electronic databases	18
2.2.2.3	Screening of selected references from hand searches	18
2.2.2.4	Assessment of entire articles from hand searches	19
2.2.3	Data Analysis	19
2.2.3.1	Grouping by virtual study model acquisition type	21
2.2.3.2	Grouping by 2-landmark and >2-landmark measurement approaches	21
2.3	Results	22
2.3.1	Reliability of repeated measures for commonly reported linear measurements	24
2.3.2	Validity grouped by acquisition type and measurement approaches	27
2.4	Discussion	29
2.4.1	Reliability	32
2.4.2	Validity	32
2.4.3	Influence of acquisition type on reliability and validity	34
2.4.4	Influence of the number of landmarks in a measurement on validity and reliability	34
2.5	Conclusion	35
2.6	References	36
Chapter 3. Study Model Analysis using CBCT-generated Virtual Study Models		40
3.1	Introduction	40
3.2	Methods and Materials	45
3.2.1	Sample size calculation and selection	47
3.2.2	Parameters used in a Study Model Analysis (SMA)	49
3.2.3	Modalities of assessment	51
3.2.3.1	Conventional plaster study models	51

3.2.3.2	Virtual models generated from Cone Beam Computed Tomography scans	52
3.2.3.3	Extracted premolars from matched samples	52
3.2.4	Experimental Design	53
3.2.5	Data Collection	55
3.2.5.1	Measurements on conventional plaster study models	56
3.2.5.2	Digital caliper measurements on matched-sample of extracted premolars	56
3.2.5.3	Software measurements from CBCT-generated study models	57
3.2.5.4	Timed measurements using computer operating system time clock	59
3.2.6	Statistical Analyses	59
3.2.6.1	Intra-rater Reliability within each modality of assessment	63
3.2.6.2	Inter-rater Reliability within each modality of assessment	64
3.2.6.3	Validity between each modality of assessment	65
3.2.6.4	Timed study model analysis for each modality of assessment	66
3.3	Results	67
3.3.1	Sample characteristics	67
3.3.2	Intra-rater Reliability	67
3.3.3	Inter-rater Reliability	72
3.3.4	Validity	77
3.3.5	Time	80
3.4	Discussion	82
3.4.1	Influence of the number of landmarks in a measurement on validity and reliability	83
3.4.2	Reliability	84
3.4.3	Validity	87
3.4.4	Time	89

3.4.5	Limitations and possible sources of error	92
3.4.6	Transfer of knowledge to clinical practice	98
3.5	Conclusion	101
3.6	Sources of Funding	102
3.7	References	103
Chapter 4. General Discussion and Conclusions		108
<hr/>		
4.1	Synthesis	108
4.2	Limitations	109
4.4	Future Research	111
Appendices		114
<hr/>		
Appendix 2-1. Search strategy and related search terms.		114
Appendix 2-2. Summary of results from electronic databases, as of May 16, 2010, after adapted search strings were applied.		114
Appendix 2-3. Sample calculation using A) pooled data from systematic review, for B) weighted mean difference for the parameter OB in laser-acquired models.		115
Appendix 2-4. Raw data from selected articles of the systematic review.		115
Appendix 3-1. Raw data from this study used for statistical analysis in the software SPSS.		118
Appendix 3-2. Histograms for the differences in measurements of Anatomodels minus Plaster.		132
Appendix 3-3. Histograms for the differences in measurements of Anatomodels minus extracted premolars.		133
Appendix 3-4. Histograms for the differences in measurements of plaster minus extracted premolars.		133
Appendix 3-5. Sample characteristics for Gender and Mean Age for orthodontic records across three modalities: Anatomodels, Plaster, and Extracted Premolars.		133
Appendix 3-6. Intra-rater, Anatomodels: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).		134
Appendix 3-7. Intra-rater, Plaster: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).		135

Appendix 3-8. Inter-rater, Anatomodels: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red). 136

Appendix 3-9. Inter-rater, Plaster: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red). 137

Appendix 3-10. Validity, Anatomodels vs Plaster: nominal parameter crosstabulations of paired assessments with a summary of overall concordant pairs (green) and discordant pairs (red). 138

List of Tables

Table 2-1. Intra-rater, plaster study models: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.	25
Table 2-2. Intra-rater, Laser-acquired virtual models: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.	26
Table 2-3. Validity, Laser-acquired vs plaster: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.	28
Table 2-4. Validity, CBCT-acquired vs plaster: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.	29
Table 3-1. Sample size calculations to detect differences of 0.5 mm or 2.0 mm, comparing projected sample sizes for 1 variable, and Bonferroni adjustments for 13 variables as well as 36 variables.	48
Table 3-2. Intra-rater, Anatomodels: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.	70
Table 3-3. Intra-rater, Plaster: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.	71
Table 3-4. Intra-rater, Extracted Premolars: ICC and repeated measures ANOVA mean differences shown for mesiodistal width measurements of each extracted premolar.	72
Table 3-5. Inter-rater, Anatomodels: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.	75
Table 3-6. Inter-rater, Plaster: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.	76
Table 3-7. Inter-rater, Extracted Premolars: ICC and repeated measures ANOVA mean differences shown for mesiodistal width measurements of each extracted premolar.	77

Table 3-8. Validity, Anatomodels vs Plaster: ICC and paired-sample mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.	79
Table 3-9. Validity, Anatomodels vs Extracted Premolar: ICC and paired-sample mean differences for each premolar.	80
Table 3-10. Validity, Plaster vs Extracted Premolar: ICC and paired-sample mean differences for each premolar.	80
Table 3-11. Time required measuring all parameters in a study model analysis during: A. Reliability studies within one evaluator for Anatomodels and Plaster separately, and between five evaluators for Anatomodels and Plaster separately; B. Validity study comparing Anatomodels to Plaster.	81
Table 3-12. Comparison of estimated resources, time, and related costs for Plaster in-house versus Anatomodels via CBCT scan in-house or outsourced.	92
Table 3-13. Caution is advised for the identified 2-landmark Linear parameters due to agreement that was poor ($ICC < 0.600$), or mean differences that were both statistically significant ($p\text{-value} < 0.05$) and clinically large (> 0.5 mm).	98
Table 3-14. Caution is advised for the identified > 2 -landmark linear parameters due to poor agreement ($ICC < 0.600$), or mean differences that were both statistically significant ($p\text{-value} < 0.05$) and clinically large (> 2.0 mm), and for the identified categorical parameters due to potentially high discordances. A summary from the time studies is also provided.	99

List of Figures

- Figure 1-1. Images from InVivoDental software: (A) A patient's CBCT volumetric scan, (B) Study models permitting visualization of roots and unerupted teeth, (C) Occlusal view of mandibular dental arch. 1
- Figure 2-1. Flow chart of the selection process. 23
- Figure 3-1. Precision using voxels: A, The dimensions of a voxel showing the length of the diagonal; B, The precision of measuring the distance between two adjacent voxels may, at worst, be 0.52 mm, so values lower than this amount should be interpreted with caution. 46
- Figure 3-2. Study Flow Chart. 53
- Figure 3-3. Study Model Analysis form. 55
- Figure 3-4. Step-by-step pictorial instructions for performing a single tooth width measurement with Anatomodels in the software InVivo 5.0 build 229. 58
- Figure 3-5. The accuracy of selecting voxels, outlined in blue, for the boundary of an object which follows a path (orange line) through points A, B and C. Selecting point A (green circle) is perfectly accurate since the orange line goes through the center of the voxel. But, in attempting to select points B and C, we are forced to select a neighboring voxel which centers at point B' and C' (yellow circles), respectively. Since the diagonal of a voxel with 0.3 mm sides is 0.52 mm, Point B' has as much as 0.26 mm error from the true Point B. Taking into account the error for point C', one can note that the accuracy of selecting two voxels can have a total error of much as about 0.5 mm. 94
- Figure 3-6. Artifacts in Anatomodels: A, due to metal streak artifacts; B, demonstrating disappearing surfaces. 96

Chapter 1. Introduction

1.1 Background

Study model analysis (SMA) is important for accurate diagnosis and treatment planning in dentistry. In performing a SMA, common diagnostic parameters are measured on dental models. Conventionally, dental study models are obtained by taking impressions of a patient's upper and lower dentition, which is often an uncomfortable intraoral procedure. A novel method involving Cone Beam Computed Tomography (CBCT) scans and the InVivoDental software (Anatomage, San Jose, CA) offers an alternative to obtaining study models, as illustrated in Figure 1-1, from which SMA can be performed without taking impressions. A full SMA using CBCT scans has not yet been reported in the literature.

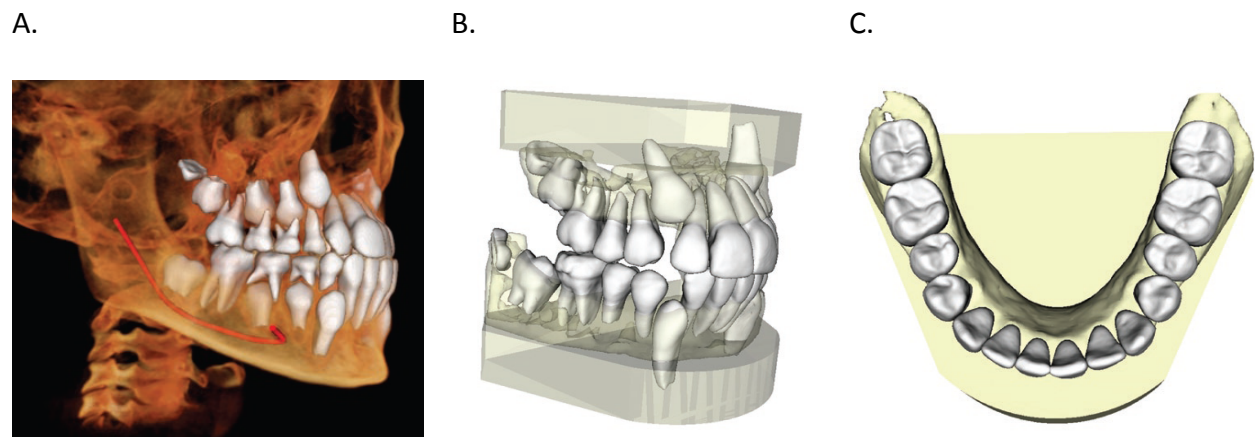


Figure 1-1. Images from InVivoDental software: (A) A patient's CBCT volumetric scan, (B) Study models permitting visualization of roots and unerupted teeth, (C) Occlusal view of mandibular dental arch.

1.2 Literature Review

Much research has been conducted on select parameters of a SMA on virtual models compared with the gold standard, plaster models. Virtual models can be generated in several ways, more recently through CBCT scans, and their utility in clinical practice should be explored. High quality research methodologies involving investigations of both quantitative (i.e. linear) and qualitative (i.e. categorical) parameters of a SMA should include assessments of reliability, validity and time. Specifically, reliability refers to the consistency with which a measurement can be made, validity refers to the ability to truly measure what is intended, and time refers to how long it takes to perform such measurements. Such inquiries are important among the many considerations prior to making decisions to implement new techniques in clinical practice.

1.2.1 Study Model Analysis

Study model analysis traditionally utilizes measurement calipers to evaluate linear distances between certain landmarks on dental study models. Commonly reported quantitative parameters include overjet, overbite, intermolar width, intercanine width, mesiodistal tooth widths and arch perimeter¹. But, a full study model analysis also includes qualitative parameters which can be assessed by visual inspection, such as molar and canine Angle classification, arch symmetry, size and shape. Since SMA has customarily been performed on

conventional dental study models, measurements obtained from plaster dental casts can be considered the gold standard.

1.2.2 Virtual Study Models

Virtual study models have many advantages² and a number of studies³⁻⁶ that compared them with plaster concluded that the differences in diagnostic measurements are not clinically significant. Virtual study models can be generated by various approaches including laser scanning, holographic scanning, stereophotogrammetry capture, or CBCT scanning. A systematic review on the reliability and validity of virtual models compared to plaster will be presented in Chapter 2. A similar systematic review by Fleming *et al.* (2011)⁷ reported that virtual study models offer a high degree of validity when compared to direct measurement on plaster models with differences likely to be clinically acceptable; however, they did not consider reliability measures. On the other hand, CBCT-generated study models, though virtual, have not been fully investigated for the reliability, validity and time requirements for diagnostic measurements in a full SMA; this will be reported in Chapter 3.

1.2.3 Cone Beam Computed Tomography-generated virtual study models

CBCT is an increasingly popular radiographic technique able to produce theoretically undistorted 3D images⁸ of the dentofacial complex from which exact measurements may be

performed. Among its many applications, CBCT is particularly useful in oral surgery, implantology and orthodontics⁹.

A novel method¹⁰ utilizing CBCT scans of the oral region and the InVivoDental software (Anatmage, San Jose, CA) offers an alternative to obtaining study models from which SMA can be performed without taking impressions. After uploading a CBCT dataset, Anatmage will return a new CBCT dataset with teeth digitally segmented. Subsequently, SMA can be performed on the CBCT-generated virtual study models, Anatomodels (Anatmage, San Jose, CA), using the InVivoDental software.

In early investigations, Lagravere et al. (2008)¹¹ compared measurements on CBCT images to a coordinate measuring machine and found that the accuracy of linear and angular measurements from titanium markers on a synthetic mandible were within 1 mm and 1 degree, respectively. Agreement as measured by Intraclass Correlation Coefficient was near perfect when assessed across each of the three dimensions (x, y and z axes).

Linear measurements based on cephalometric landmarks of CBCT reconstructions were found by Periago et al. (2010)¹² to be $1.13 \pm 1.47\%$ smaller than anatomic landmarks on human skulls and these differences were found to be statistically significant but clinically acceptable. Mean percentage measurement error on CBCT was $2.31 \pm 2.11\%$ which was higher than repeated measurements on skulls of $0.63 \pm 0.51\%$. However, the study used dry skulls and the authors were unable to simulate soft tissue effects of attenuation on image quality and admitted that

“the dimensional accuracy of 3D measurements would be somewhat less on patient derived data.” Distances such as Sella-Nasion or Gonion-Menton were investigated but differences in measurements involving the teeth, as would be performed in a model analysis, were not assessed.

Later, Ganguly *et al.* (2011)¹³ confirmed that linear measurements of bone height on CBCT in the presence of soft tissue was underestimated by on average 0.31 ± 0.61 mm compared to direct measurements on the same six cadaver heads. There were no statistically significant differences between repeated measurements. Although the sample size was small, these authors concluded that CBCT-based linear measurements were sufficiently accurate for clinical use. Still, measurements involving teeth were not assessed.

Today, virtual models with digitally segmented teeth are available and, thus far, have been validated in two studies^{14,15} using select linear parameters of a full SMA. Kau *et al.* (2010)¹⁴ compared Anatomodels to virtual study models (OrthoCAD, Cadent, Fairview, NJ) and found no statistical significance for the mean differences of 0.79 ± 2.33 mm for maxillary Little’s Index, 0.14 ± 1.39 mm for mandibular Little’s Index, 0.03 ± 1.31 mm for overjet, and -0.20 ± 1.67 mm for overbite. Furthermore, these differences might not be considered clinically important.

Tarazona *et al.* (2011)¹⁵ conducted a broader study of linear parameters on Anatomodels compared to two dimensional scans from the occlusal perspectives of plaster models. They found no clinical differences based on mean differences of no more than 1% for mesiodistal

tooth sizes, maxillary and mandibular intercanine and intermolar widths, as well as arch lengths (i.e. maxillary and mandibular arch perimeter). No justification was provided for their chosen level of clinical significance.

1.2.4 Timed Study Model Analysis

In deciding whether to implement a new technology in clinical practice, it is worthwhile to consider how much time and resources are required to utilize the technology. At this point, however, it is unclear if the process of obtaining measurements using the InVivoDental software is time consuming compared with conventional plaster models. To our knowledge, no study has compared the time efficiency of a full SMA on plaster casts to virtual models. There were a few reports¹⁵⁻¹⁸ of time required to perform what could be considered as only portions of a full SMA.

Tomassetti *et al.* (2001)¹⁶ studied a sample of 22 patients and reported average times performing Bolton analyses of 8 minutes and 4 seconds using plaster and 5 minutes and 16 seconds using OrthoCAD virtual models. These Bolton calculations imply the measurement of all mesiodistal tooth widths from first molar to first molar in both arches.

In agreement with these findings, Mullen *et al.* (2007)¹⁷ found that Bolton analyses, which is commonly assessed in SMA, when performed on 30 plaster models was on average 1 minute and 4 seconds slower than corresponding virtual models (eModels, GeoDigm, Chanhassen,

MN). They explained that the longer times using plaster models could be due to the extra steps of having to write down the measurements for each tooth, whereas with eModels, the measurements were automatically calculated at the click of a button.

Reporting the opposite trend, however, in a study of 32 plaster study models and corresponding eModels, Horton *et al.* (2010)¹⁸ compared mesiodistal dimensions from maxillary and mandibular first molar to first molar. The average time to measure the plaster study models was 4 minutes and 15 seconds while measurements that involved freely rotating the digital models on-screen took on average 7 minutes and 1 second.

On the other hand, Tarazona *et al.* (2011)¹⁵ found in their study of 27 patients that the average time to perform linear measurements on both arches using Anatomodels was 3 minutes 8 seconds compared with static scans of plaster models which took 4 minutes 56 seconds. It is unclear how measurements on virtual models were performed so quickly compared with previously mentioned studies.

1.3 Statement of the Problem and Rationale for Inquiry

Dental study models, whether virtual or made of conventional plaster, are typically obtained by taking alginate impressions of the maxillary and mandibular teeth. This is frequently an uncomfortable intraoral procedure that demands effective behavioral management¹⁹.

Furthermore, impression-taking is a resource intensive process, requiring chair time, staff time, lab time and associated costs.

On the other hand, at a current list price of about \$70 USD, Anatomage will electronically produce a set of Anatomodels which is a new CBCT dataset with teeth segmented by their proprietary process. If the measurements from the CBCT study models are found to be valid, reliable, and time-efficient, it may represent an acceptable alternative for the purposes of model analyses.

1.4 Research Objectives

The main research objectives of this thesis are studies on reliability, validity and time of study model analysis on Anatomodels compared with matched samples of plaster models and a matched subset of extracted premolars.

1.4.1 Reliability

The first objective was to assess the reliability of study model analysis performed using Anatomodels compared to matched samples of conventional plaster dental study models, as well as to selected matched samples of extracted premolars.

1.4.2 Validity

The next objective was to assess the validity of study model analysis performed using Anatomodels compared to matched samples of conventional plaster dental study models, as well as to selected matched samples of extracted premolars.

1.4.3 Time

The final objective was to assess the time efficiency of study model analysis performed using Anatomodels compared with conventional plaster dental study models.

1.5 Hypotheses

The following research hypotheses regarding measures of mean differences will be investigated for the studies on reliability, validity and time. Similar hypotheses can be stated regarding measures of agreement for the studies on reliability and validity.

1.5.1 Reliability

Intra-rater reliability was investigated through one evaluator on the basis of the null hypotheses for each parameter of a SMA across the following three modalities:

H₀: Using Anatomodels, there is no difference between the mean measurements among the five repeated trials.

H₀: Using plaster study models, there is no difference between the mean measurements among the five repeated trials.

H₀: Using extracted premolars, there is no difference between the mean measurements among the five repeated trials.

Inter-rater reliability was investigated through five evaluators on the basis of the null hypothesis for each parameter of a SMA across the following three modalities:

H₀: Using Anatomodels, there is no difference between the mean measurements among the individual trials of five evaluators.

H₀: Using plaster study models, there is no difference between the mean measurements among the individual trials of five evaluators.

H₀: Using extracted premolars, there is no difference between the mean measurements among the individual trials of five evaluators.

1.5.2 Validity

The validity of measurements was investigated through one evaluator with the null hypotheses across the following matched pairs of groups:

H_0 : There is no difference between the mean measurements on Anatomodels and corresponding plaster study models.

H_0 : There is no difference between the mesiodistal dimensions of premolars on Anatomodels and corresponding extracted premolars.

H_0 : There is no difference between the mesiodistal dimensions of premolars on plaster study models and corresponding extracted premolars.

1.5.3 Time

The speed of a full study model analysis was investigated with the following null hypothesis:

H_0 : Within a single evaluator, there is no difference between the time required to perform measurements on Anatomodels and corresponding plaster study models.

H₀: Within a single evaluator, there is no difference between the times of repeated measurements on Anatomodels.

H₀: Within a single evaluator, there is no difference between the times of repeated measurements on plaster study models.

H₀: Between multiple evaluators, there is no difference between the times of repeated measurements on Anatomodels.

H₀: Between multiple evaluators, there is no difference between the times of repeated measurements on plaster study models.

1.6 References

1. Kahl-Nieke B, Fischbach H, Schwarze CW. Treatment and postretention changes in dental arch width dimensions--a long-term evaluation of influencing cofactors. *Am J Orthod Dentofacial Orthop* 1996;109:368-378.
2. Joffe L. OrthoCAD: digital models for a digital era. *J Orthod* 2004;31:344-347.
3. 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 e11-14; discussion 16.

4. 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.
5. 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.
6. 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.
7. Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review. *Orthodontics & Craniofacial Research* 2011;14:1-16.
8. Scarfe WC, Farman AG, Sukovic P. Clinical applications of cone-beam computed tomography in dental practice. *J Can Dent Assoc* 2006;72:75-80.
9. Hilgers ML, Scarfe WC, Scheetz JP, Farman AG. Accuracy of linear temporomandibular joint measurements with cone beam computed tomography and digital cephalometric radiography. *Am J Orthod Dentofacial Orthop* 2005;128:803-811.
10. Chenin DL, Chenin DA, Chenin ST, Choi J. Dynamic Cone-Beam Computed Tomography in Orthodontic Treatment. *J Clin Orthod* 2009;43:507-512.
11. Lagravere MO, Carey J, Toogood RW, Major PW. Three-dimensional accuracy of measurements made with software on cone-beam computed tomography images. *Am J Orthod Dentofacial Orthop* 2008;134:112-116.

12. Periago DR, Scarfe WC, Moshiri M, Scheetz JP, Silveira AM, Farman AG. Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. *Angle Orthod* 2008;78:387-395.
13. Ganguly R, Ruprecht A, Vincent S, Hellstein J, Timmons S, Qian F. Accuracy of linear measurement in the Galileos cone beam computed tomography under simulated clinical conditions. *Dento-Maxillo-Facial Radiology* 2011;40:299-305.
14. Kau CH, Littlefield J, Rainy N, Nguyen JT, Creed B. Evaluation of CBCT digital models and traditional models using the Little's Index. *Angle Orthod* 2010;80:435-439.
15. Tarazona B, Llamas JM, Cibrian R, Gandia JL, Paredes V. A comparison between dental measurements taken from CBCT models and those taken from a Digital Method. *European Journal of Orthodontics* 2011.
16. Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR, Jr. A comparison of 3 computerized Bolton tooth-size analyses with a commonly used method. *Angle Orthodontist* 2001;71:351-357.
17. Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2007;132:346-352.
18. Horton HM, Miller JR, Gaillard PR, Larson BE. Technique comparison for efficient orthodontic tooth measurements using digital models. *Angle Orthod* 2009;80:254-261.
19. Kaakko T, Horn MT, Weinstein P, Kaufman E, Leggott P, Coldwell SE. The influence of sequence of impressions on children's anxiety and discomfort. *Pediatr Dent* 2003;25:357-364.

Chapter 2. Linear Measurements using Virtual Study Models: A Systematic Review

2.1 Introduction

A key process in diagnosis and treatment planning in dentistry is the study model analysis (SMA). In performing a SMA, common diagnostic parameters¹ are measured on dental models, such as overjet, intermolar width, and arch perimeter. Such linear measurements might further be classified as those that involve two landmarks (2-landmark measures), and those that involve more than two landmarks (>2-landmark measures).

Conventionally, SMA is performed on plaster dental casts using measurement calipers. As such, measurements from plaster study models can be considered the gold standard.

In recent decades, three-dimensional (3D) virtual study models have made headway into dentistry, spearheaded by proposed advantages² such as no physical storage space requirements, simple measuring and storing of data, storage and integration into digital records, chairside retrieval and viewing, and transferability.

The available literature on three-dimensional virtual dental study models has largely focused on those acquired by laser³⁻¹⁷ (Laser-acquired), while others have investigated holographic

scanning¹⁸, stereophotogrammetry capture¹⁹ and more recently, by cone-beam computed tomography (CBCT)²⁰⁻²² (CBCT-acquired).

Both validity and reliability are important measures. Reliability refers to the consistency with which a measurement can be made and validity refers to the ability to truly measure what is intended. It is our opinion that demonstrated reliability in repeated measurements within virtual models and plaster separately are necessary before interpreting validity between the two modalities.

Numerous studies have investigated the validity and reliability of linear measurements made on plaster versus virtual study models, but a systematic review has not been performed to collectively summarize their conclusions. To our knowledge, the only systematic review on virtual study models by Fleming *et al.* (2011)²³ summarized assessments of validity but not reliability.

The aims of this study were to perform a systematic review of the literature to assess the validity and reliability of linear measurements using virtual versus plaster dental study models, grouping our analysis by virtual model acquisition type and the number of landmarks used in a given measurement.

2.2 Methods

The following research methodology was employed for this systematic review.

2.2.1 Search strategy

The PICO²⁴ search strategy (Appendix 2-1) was adopted for this study and the resulting search string was tailored for PubMed (from 1966 to May 16, 2010) and adapted with no limits for the following online databases: OVID Medline, OVID – All EBM Reviews, and Lilacs (Appendix 2-2).

2.2.2 Selection of articles

Eligibility of selected articles was determined in four phases. Selection of articles at each stage was performed by three researchers. Discrepancies were discussed and final selections were agreed upon by majority vote. All non-English papers selected at each stage were appropriately translated.

2.2.2.1 Screening of articles from electronic databases

In Phase I of the selection process, from the electronic database results, the titles and abstracts were screened with the following selection criteria:

- Main focus was on the assessment of linear measurements in 3D virtual models of the human dentition.

2.2.2.2 Assessment of entire articles from electronic databases

In Phase II of the selection process, the whole article from those selected in Phase I were retrieved where possible and the following selection criteria were applied:

- Validity and reliability measures provided
- Gold standard measurements taken from plaster casts
- Minimum sample size of 10

2.2.2.3 Screening of selected references from hand searches

In Phase III of the selection process, the reference lists from the selected articles in Phase II were screened with the same selection criteria as Phase I:

- Main focus was on the assessment of linear measurements in 3D virtual models of the human dentition.

2.2.2.4 Assessment of entire articles from hand searches

In Phase IV of the selection process, the retrievable articles from Phase III were assessed with the same selection criteria as Phase II:

- Validity and reliability measures provided
- Gold standard measurements taken from plaster casts
- Minimum sample size of 10

2.2.3 Data Analysis

In this systematic review, the important measures were reliability and validity. Reliability refers to the consistency with which a measurement can be made and this was assessed by reports of mean difference, agreement (Intraclass correlation coefficient, ICC) and correlation (Pearson's correlation coefficient, PCC) of repeated measures using virtual and plaster models. Validity refers to the ability to truly measure what is intended and this was also assessed using measures of mean difference, agreement (ICC) and correlation (PCC) between virtual and plaster models.

Relevant data was tabulated in a spreadsheet using Excel 2007 (Microsoft, Redmond, WA). For both validity and reliability, the data was weighted by sample size and analyzed by descriptive statistics. An example of a calculation for weighted mean difference is provided in Appendix 2-

3. A minimum sample size of 10 ensured that studies with good methodology—ones that measured both reliability and validity—were not excluded. Furthermore, weighted means allowed us to pool the results from studies that had relatively lower sample sizes. Conversely, weighted means allowed those studies with higher sample sizes to contribute more to the findings of this systematic review. In the calculation of weighted mean differences, as an example, individual mean differences multiplied by their respective sample sizes, as reported in the study, were added together and then divided by the total sum of the associated sample sizes. Weighted ICC and weighted PCC were calculated in a similar manner.

Of the selected articles, inter-rater reliability^{3,10,17,21} was uncommonly reported, so only intra-rater reliability^{4,5,7,10,11,15,19-22} in terms of mean differences, ICC and PCC were tabulated. Other reported measures of reliability^{6,8,9,12-14,16,18} such as standard deviations, random error, or statements confirming tests of repeated measurements, were also accepted but not summarized. Furthermore, because reliability is always within a single modality (i.e. within plaster models or virtual models alone), weighted mean differences were calculated by first converting reported differences into absolute values.

The parameters summarized in this systematic review were, by inspection, the most commonly reported of the selected articles. The parameters that could not be categorized under one of the commonly reported linear parameters, but were nonetheless reported in the literature, were noted but not summarized in this paper.

In this systematic review, we set clinically relevant thresholds for mean differences for 2-landmark linear measurements at 0.5 mm and for >2-landmark linear measurements at 2.0 mm. Although largely unsubstantiated by the literature, three^{4,7,10} of the selected articles specified clinically significant mean differences and their thresholds were in line with ours. Asquith *et al.* (2007)⁴ suggested clinically significant differences of 0.5 mm for tooth width measurements or 5% for larger measurements. Goonewardene *et al.* (2008)⁷ argued that variations of 1-2 mm in crowding measurements could influence extraction versus non-extraction treatment plans. Mullen *et al.* (2007)¹⁰ proposed that less than 1.5 mm of tooth structure discrepancy per arch could be clinically insignificant.

2.2.3.1 Grouping by virtual study model acquisition type

Data for all virtual study models were grouped to investigate any differences between virtual model acquisition types.

2.2.3.2 Grouping by 2-landmark and >2-landmark measurement approaches

The collected data was also grouped to investigate differences between 2-landmark and >2-landmark linear measurements.

2.3 Results

A flow chart of the selection process is illustrated in Figure 2-1. The search strategy (Appendix 2-2) revealed 278 potential articles from electronic databases after duplicates were removed. From the list of 278 potential articles, three reviewers identified 59 retrievable articles by majority vote based on the titles and abstracts, and subsequently selected 20 after reading the entire articles. From these 20 articles, 238 unique references were identified from which 62 retrievable articles were screened, but ultimately, no additional articles were selected from the hand-searches. After specifically excluding three of the originally included articles, a final total of 17 articles were selected for this review. The data collected from the selected articles is compiled in Appendix 2-4.

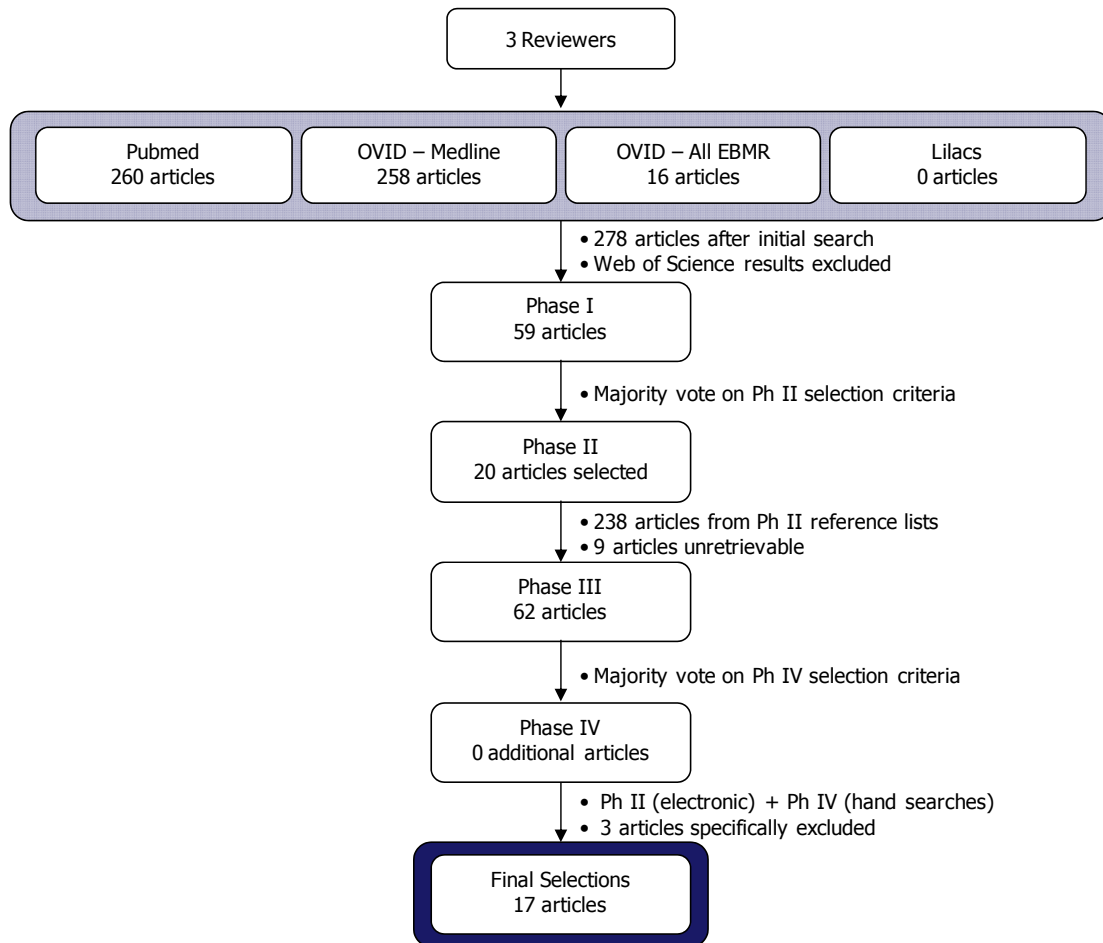


Figure 2-1. Flow chart of the selection process.

Three articles that were ultimately excluded^{11,18,19} had initially satisfied the selection criteria at each phase. However, since our intention was to pool relevant data, the information reported in the three excluded articles was unsuitable in the context of this systematic review and needed support by further independent studies. One study assessed virtual models of neonatal cleft palate¹¹ patients without any erupted teeth. Another study investigated virtual models acquired by holographic scanning¹⁸, but the paper was published two decades ago. Similarly, the study on models acquired by stereophotogrammetry¹⁹ has not been revisited for almost a decade.

2.3.1 Reliability of repeated measures for commonly reported linear measurements

The intra-rater reliability of repeated measures for plaster study models and laser-acquired virtual study models are presented in Table 2-1 and Table 2-2, respectively. Although the intra-rater reliability data for CBCT-acquired models will not be presented in a table due to insufficient comparative data, ICC values from two studies^{21,22} were above 0.80 and PCC values from the third study²⁰ were well above 0.90 which suggested good agreement and excellent correlation of repeated measures.

Intra-rater reliability for both plaster (Table 2-1) and laser-acquired (Table 2-2) study models were reported for all of the common 2-landmark and >2-landmark measurements. All weighted mean differences were under 0.5 for the 2-landmark parameters and under 1.5 mm for the >2-landmark parameters. For repeated measurements in plaster, ICC values were around 0.85 for all 2-landmark parameters and above 0.98 for crowding; similarly PCC values were above 0.91 for 2-landmark parameters and above 0.96 for arch perimeter. For repeated measurements in laser-acquired models, ICC values were near 0.99.

Table 2-1. Intra-rater, plaster study models: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.

Parameter	Absolute Difference		Agreement		Correlation	
	N	Mean (mm)	N	ICC	N	PCC
<i>Plaster, Linear measurements, 2 landmarks</i>						
Overjet	114	0.18	15	0.852	-	-
Overbite	104	0.15	15	0.852	-	-
Tooth 1-1	90	0.02	15	0.852	-	-
Tooth 1-2	80	0.03	15	0.852	-	-
Tooth 1-3	80	0.02	15	0.852	34	0.933
Tooth 1-4	80	0.04	15	0.852	-	-
Tooth 1-5	80	0.04	15	0.852	-	-
Tooth 1-6	80	0.05	15	0.852	-	-
Tooth 2-1	80	0.02	15	0.852	34	0.944
Tooth 2-2	80	0.07	15	0.852	-	-
Tooth 2-3	80	0.01	15	0.852	-	-
Tooth 2-4	80	0.01	15	0.852	-	-
Tooth 2-5	80	0.04	15	0.852	-	-
Tooth 2-6	90	0.00	15	0.852	-	-
Tooth 3-1	80	0.03	15	0.852	-	-
Tooth 3-2	80	0.04	15	0.852	-	-
Tooth 3-3	80	0.03	15	0.852	-	-
Tooth 3-4	90	0.04	15	0.852	-	-
Tooth 3-5	80	0.05	15	0.852	-	-
Tooth 3-6	80	0.07	15	0.852	-	-
Tooth 4-1	80	0.01	15	0.852	-	-
Tooth 4-2	80	0.00	15	0.852	-	-
Tooth 4-3	80	0.03	15	0.852	-	-
Tooth 4-4	80	0.00	15	0.852	-	-
Tooth 4-5	80	0.05	15	0.852	34	0.913
Tooth 4-6	80	0.06	15	0.852	34	0.999
Mx_IMW	90	0.18	15	0.852	-	-
Mx_ICW	80	0.19	15	0.852	-	-
Mn_IMW	80	0.13	15	0.852	-	-
Mn_ICW	90	0.04	15	0.852	-	-
<i>Plaster, Linear measurements, >2 landmarks</i>						
Mx_Perim	24	0.51	-	-	34	0.999
Mx_Crowd	80	0.67	50	0.991	-	-
Mn_Perim	24	0.48	-	-	34	0.961
Mn_Crowd	80	0.19	50	0.979	-	-
Bolton6	24	0.32	-	-	-	-
Bolton12	24	0.58	-	-	-	-

Abbreviations: ICC, Intraclass Correlation Coefficient; PCC, Pearson's Correlation Coefficient, Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

Table 2-2. Intra-rater, Laser-acquired virtual models: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.

Parameter	Absolute Difference		Agreement		Correlation	
	N	Mean (mm)	N	ICC	N	PCC
<i>Laser-acquired, Linear measurements, 2 landmarks</i>						
Overjet	114	0.13	-	-	-	-
Overbite	104	0.09	-	-	-	-
Tooth 1-1	90	0.07	-	-	-	-
Tooth 1-2	80	0.06	-	-	-	-
Tooth 1-3	80	0.00	-	-	-	-
Tooth 1-4	80	0.04	-	-	-	-
Tooth 1-5	80	0.00	-	-	-	-
Tooth 1-6	80	0.08	-	-	-	-
Tooth 2-1	80	0.08	-	-	-	-
Tooth 2-2	80	0.07	-	-	-	-
Tooth 2-3	80	0.03	-	-	-	-
Tooth 2-4	80	0.02	-	-	-	-
Tooth 2-5	80	0.02	-	-	-	-
Tooth 2-6	90	0.07	-	-	-	-
Tooth 3-1	80	0.07	-	-	-	-
Tooth 3-2	80	0.03	-	-	-	-
Tooth 3-3	80	0.06	-	-	-	-
Tooth 3-4	90	0.03	-	-	-	-
Tooth 3-5	80	0.02	-	-	-	-
Tooth 3-6	80	0.04	-	-	-	-
Tooth 4-1	80	0.04	-	-	-	-
Tooth 4-2	80	0.11	-	-	-	-
Tooth 4-3	80	0.04	-	-	-	-
Tooth 4-4	80	0.01	-	-	-	-
Tooth 4-5	80	0.10	-	-	-	-
Tooth 4-6	80	0.07	-	-	-	-
Mx_IMW	90	0.13	-	-	-	-
Mx_ICW	80	0.07	-	-	-	-
Mn_IMW	80	0.36	-	-	-	-
Mn_ICW	90	0.03	-	-	-	-
<i>Laser-acquired, Linear measurements, >2 landmarks</i>						
Mx_Perim	24	1.13	-	-	-	-
Mx_Crowd	80	0.13	50	0.987	-	-
Mn_Perim	24	1.07	-	-	-	-
Mn_Crowd	80	0.06	50	0.986	-	-
Bolton6	24	0.69	-	-	-	-
Bolton12	24	1.08	-	-	-	-

Abbreviations: ICC, Intraclass Correlation Coefficient; PCC, Pearson's Correlation Coefficient, Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

2.3.2 Validity grouped by acquisition type and measurement approaches

The validity of commonly reported linear parameters subgrouped by 2-landmark and >2-landmark measurements between plaster and specific acquisition types, Laser-acquired or CBCT-acquired, are presented in Table 2-3 and Table 2-4, respectively.

For laser-acquired study models (Table 2-3), the mean differences compared with plaster study models were well below 0.5 mm for 2-landmark measures, and less than 1 mm for >2-landmark measures. The majority of parameters were reported in terms of ICC with weighted values that tended to be above 0.90.

The virtual study models acquired by CBCT scanning (Table 2-4) had mean differences compared with plaster study models of below 0.5 mm for 2-landmark measures. None of the articles included in this systematic review reported mean differences for >2-landmark measures. Although none of the articles reported ICC values, weighted PCC values from one study²⁰ ranged from 0.62 to 0.99.

Table 2-3. Validity, Laser-acquired vs plaster: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.

Parameter	Difference ¹		Agreement		Correlation	
	N	Mean (mm)	N	ICC	N	PCC
<i>Laser-acquired vs. Plaster, Linear measurements, 2 landmarks</i>						
Overjet	204	-0.06	80	0.967	-	-
Overbite	194	-0.19	80	0.913	-	-
Tooth 1-1	140	-0.02	80	0.911	-	-
Tooth 1-2	130	-0.04	80	0.968	-	-
Tooth 1-3	130	0.00	80	0.900	-	-
Tooth 1-4	130	-0.02	80	0.908	-	-
Tooth 1-5	130	-0.02	80	0.882	-	-
Tooth 1-6	130	-0.01	80	0.942	-	-
Tooth 2-1	130	-0.04	80	0.945	-	-
Tooth 2-2	130	-0.05	80	0.963	-	-
Tooth 2-3	130	0.00	80	0.984	-	-
Tooth 2-4	130	-0.01	80	0.948	-	-
Tooth 2-5	130	-0.02	80	0.966	-	-
Tooth 2-6	140	-0.05	80	0.896	-	-
Tooth 3-1	100	-0.07	80	0.907	-	-
Tooth 3-2	100	-0.05	80	0.891	-	-
Tooth 3-3	100	-0.03	80	0.914	-	-
Tooth 3-4	110	-0.05	80	0.918	-	-
Tooth 3-5	100	-0.03	80	0.939	-	-
Tooth 3-6	100	-0.11	80	0.917	-	-
Tooth 4-1	100	-0.08	80	0.901	-	-
Tooth 4-2	100	-0.05	80	0.908	-	-
Tooth 4-3	100	-0.06	80	0.906	-	-
Tooth 4-4	100	-0.05	80	0.972	-	-
Tooth 4-5	100	-0.02	80	0.963	-	-
Tooth 4-6	100	-0.07	80	0.918	-	-
Mx_IMW	160	0.13	101	0.943	-	-
Mx_ICW	130	0.07	101	0.927	-	-
Mn_IMW	150	0.18	80	0.988	-	-
Mn_ICW	140	0.08	80	0.983	-	-
<i>Laser-acquired vs. Plaster, Linear measurements, >2 landmarks</i>						
Mx_Perim	74	0.58	-	-	-	-
Mx_Crowd	155	-0.09	80	0.984	-	-
Mn_Perim	94	0.83	-	-	-	-
Mn_Crowd	155	0.43	80	0.966	-	-
Bolton6	24	-0.04	-	-	-	-
Bolton12	24	-0.38	-	-	-	-

¹ Negative mean difference when measurements from Plaster are larger
Abbreviations: ICC, Intraclass Correlation Coefficient; PCC, Pearson's Correlation Coefficient,
Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Intercanine Width;
Perim, Arch Perimeter; Crowd, crowding if negative;
Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

Table 2-4. Validity, CBCT-acquired vs plaster: mean difference, agreement and correlation values weighted by sample size shown for most commonly reported parameters, grouped by 2-landmark and >2-landmark linear measurements.

Parameter	Difference		Agreement		Correlation	
	N	Mean (mm)	N	ICC	N	PCC
<i>CBCT-acquired vs. Plaster, Linear measurements, 2 landmarks</i>						
Overjet	15	-0.31	-	-	-	-
Overbite	15	-0.21	-	-	-	-
Tooth 1-1	40	-0.10	-	-	34	0.878
Tooth 1-2	40	-0.16	-	-	34	0.898
Tooth 1-3	40	-0.10	-	-	34	0.846
Tooth 1-4	40	-0.06	-	-	34	0.773
Tooth 1-5	40	-0.09	-	-	34	0.699
Tooth 1-6	40	-0.17	-	-	34	0.746
Tooth 2-1	40	-0.13	-	-	34	0.828
Tooth 2-2	40	-0.10	-	-	34	0.812
Tooth 2-3	40	-0.10	-	-	34	0.822
Tooth 2-4	40	-0.10	-	-	34	0.806
Tooth 2-5	40	-0.12	-	-	34	0.712
Tooth 2-6	40	-0.18	-	-	34	0.882
Tooth 3-1	40	-0.12	-	-	34	0.704
Tooth 3-2	40	-0.14	-	-	34	0.854
Tooth 3-3	40	-0.12	-	-	34	0.786
Tooth 3-4	40	-0.08	-	-	34	0.725
Tooth 3-5	40	-0.08	-	-	34	0.836
Tooth 3-6	40	-0.09	-	-	34	0.838
Tooth 4-1	40	-0.15	-	-	34	0.617
Tooth 4-2	40	-0.15	-	-	34	0.827
Tooth 4-3	40	-0.12	-	-	34	0.723
Tooth 4-4	40	-0.15	-	-	34	0.894
Tooth 4-5	40	-0.09	-	-	34	0.885
Tooth 4-6	40	-0.12	-	-	34	0.850
Mx_IMW	15	-0.16	-	-	34	0.995
Mx_ICW	15	-0.12	-	-	34	0.987
Mn_IMW	15	-0.12	-	-	34	0.988
Mn_ICW	15	-0.14	-	-	34	0.980
<i>CBCT-acquired vs. Plaster, Linear measurements, >2 landmarks</i>						
Mx_Perim	-	-	-	-	34	0.996
Mx_Crowd	-	-	-	-	-	-
Mn_Perim	-	-	-	-	34	0.979
Mn_Crowd	-	-	-	-	-	-
Bolton6	-	-	-	-	-	-
Bolton12	-	-	-	-	-	-

¹ Negative mean difference when measurements from Plaster are larger
Abbreviations: ICC, Intraclass Correlation Coefficient; PCC, Pearson's Correlation Coefficient,
Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Intercanine Width;
Perim, Arch Perimeter; Crowd, crowding if negative;
Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

2.4 Discussion

The systematic review sought to investigate the validity and reliability of virtual study models compared with plaster, grouping our analysis by acquisition type and the number of landmarks used in a given measurement. The compiled data demonstrated the high validity and reliability of a number of 2-landmark and >2-landmark measurements, particularly from laser-acquired virtual study models.

This systematic review and the one by Fleming *et al.* (2011)²³ selected 17 articles each. However, slight differences in our selection criteria resulted in our studies selecting only nine articles^{7-10,12,13,15,17,22} in common. We chose to focus on quantitative linear measurements only; therefore, of the articles that Fleming chose to include, we had rejected because they focused on PAR²⁵, ABO²⁶⁻²⁸ or ICON²⁹ scores, which are qualitative ordinal measures. We also rejected an article³⁰ that Fleming accepted because we found no reports on reliability of repeated measurements. Of the articles that Fleming chose to exclude, we chose to accept two studies that used artificial occlusal set-ups^{3,16} since they are assessments of linear measurements nonetheless, and another study that placed marking points on the casts in black pen⁴ since those points did not affect the parameters that we chose to summarize. Finally, our search strategy selected an additional five relevant articles^{5,6,14,20,21} as of May 2010 that were not mentioned by Fleming's systematic review—three^{6,14,21} of which were published by the time their search was conducted in January of 2010.

Virtual study models acquired by laser scanning represented 14 out of the 17 selected articles, while those acquired by CBCT scanning were reported in the remaining 3. The number of good quality studies on laser-acquired study models is remarkable but emerging approaches using CBCT show promise. However, two^{21,22} of the selected studies using CBCT still required impressions, so errors may be replicated³ as the process goes from the mouth to alginate impressions and finally to virtual models.

The reliability and validity of newer approaches that generate virtual study models from direct CBCT scans of the patient's mouth³¹ compared with the gold standard plaster models have yet to be reported. Given the high reliability and validity of virtual models acquired by laser scanning, one might consider laser-acquired models as clinically acceptable substitutes for plaster models. Based on this premise, the differences in Little's Index, overjet and overbite were shown by Kau *et al.* (2010)³² to be both statistically and clinically insignificant.

By inspection, the most commonly reported 2-landmark linear parameters were overjet, overbite, maxillary and mandibular mesiodistal tooth sizes from first molar to first molars, inclusive, as well as maxillary and mandibular intermolar and intercanine widths. The commonly reported >2-landmark linear parameters were maxillary and mandibular arch perimeter and crowding, as well as Bolton anterior and Bolton overall discrepancies. This list highlights the linear parameters most comparable with existing literature; therefore, future studies on study model analyses should consider investigating these parameters.

A full study model analysis should also involve categorical parameters, such as Angle's classification, but good quality studies incorporating these were infrequently reported. It would be worthwhile, then, for future studies to investigate the reliability and validity of the linear parameters listed above in addition to categorical parameters.

2.4.1 Reliability

Intra-rater reliability of repeated measures on plaster study models as well as virtual study models for 2-landmark measures showed clinically insignificant mean differences at the 0.5 mm threshold while both agreement and correlation were good to excellent for the parameters that were reported. For >2-landmark measures, mean differences were below the 2 mm threshold indicating clinically insignificant differences in repeated measures as well as excellent agreement and correlation. Intra-rater reliability, then, was good to excellent for virtual study models and the same can be said for plaster as the differences in repeated measurements of both 2-landmark and >2-landmark linear parameters were judged to be clinically insignificant.

2.4.2 Validity

The validity of virtual compared to plaster study models for all 2-landmark and >2-landmark linear parameters showed clinically insignificant mean differences. This agrees with the findings of Fleming *et al.* (2011)²³ who reported that virtual models offer a high degree of validity when compared to direct measurement on plaster models. Compared to plaster, for 2-

landmark parameters, there was excellent agreement using laser-acquired models, while correlation using CBCT-acquired models ranged from poor to excellent. In contrast, Fleming did not summarize agreement in terms of ICC or PCC values.

Overjet, overbite, and all tooth width measurements from first molar to first molar using laser-acquired study models were clinically insignificant compared with plaster, but the negative weighted mean differences suggested a tendency towards larger measurements on plaster models. Intermolar and intercanine distances on laser-acquired models, however, had a tendency towards smaller measurements on plaster, but again, the weighted mean differences were clinically insignificant. Similarly, differences in arch perimeter, crowding and Bolton measurements were clinically insignificant. Agreement for all 2-landmark measures and arch crowding were excellent.

Compared with the compiled data from articles on laser-acquired study models, which had combined sample sizes that ranged from 100 to 204 per parameter, the data on CBCT-acquired study models had relatively smaller sample sizes that ranged from 15 to 40. As observed with laser-acquired study models, the weighted mean differences were all negative indicating a tendency towards larger measurements on plaster, but this finding had no clinical relevance. Correlation of CBCT-acquired study models compared with plaster was poor for mesiodistal measurements of teeth 1-5 and 4-1, moderate for teeth 1-4, 1-6, 2-5, 3-1, 3-3, 3-4, 4-3, and good or better for all remaining 2-landmark and arch perimeter measures. There was no obvious explanation for this variation in correlation.

2.4.3 Influence of acquisition type on reliability and validity

There were no perceived differences in intra-rater reliability and validity across the various acquisition types. The variation in correlation for 2-landmark measures from CBCT-acquired models was the only inconsistent finding, but further independent studies are required to confirm this. Aside from this possibly anomalous finding, overall, the mean differences were clinically insignificant and the correlation and agreement were good to excellent. These findings were consistent across laser-acquired and CBCT-acquired virtual models compared with plaster.

2.4.4 Influence of the number of landmarks in a measurement on validity and reliability

In magnitude, there was a tendency for the reliability and validity of 2-landmark measures to have smaller mean differences than >2-landmark measures, regardless of acquisition type. For example, for the 2-landmark parameters, repeated tooth width measurements in plaster showed less than 0.1 mm absolute difference while overjet, overbite, intermolar and intercanine distances had double the absolute differences but less than 0.2 mm. For >2-landmark parameters, differences in arch perimeter, crowding and Bolton discrepancies ranged higher than 0.2 mm, up to 0.7 mm. Although these findings were not clinically significant, this pattern for increasing absolute difference relative to the number of landmarks could be detected by inspection for repeated measurements in laser-acquired models as well.

2.5 Conclusion

1. The intra-rater reliability was high for 2-landmark and >2-landmark linear measurements performed on laser-acquired models or CBCT-acquired models and similar to measurements on plaster models.
2. The validity was high for 2-landmark and >2-landmark linear measurements comparing laser-acquired models or CBCT-acquired models to plaster study models and the weighted mean differences were clinically insignificant.
3. Agreement of measurements was excellent with less variability than correlation.
4. Acquisition type had no perceived influences on reliability and validity.
5. >2-landmark measures tended to have higher mean differences than 2-landmark measures.
6. Virtual study models are clinically acceptable compared with plaster study models in regards to intra-rater reliability and validity of selected linear measurements.

2.6 References

1. Kahl-Nieke B, Fischbach H, Schwarze CW. Treatment and postretention changes in dental arch width dimensions--a long-term evaluation of influencing cofactors. *Am J Orthod Dentofacial Orthop* 1996;109:368-378.
2. Joffe L. OrthoCAD: digital models for a digital era. *J Orthod* 2004;31:344-347.
3. Alcan T, Ceylanoglu C, Baysal B. The relationship between digital model accuracy and time-dependent deformation of alginate impressions. *Angle Orthodontist* 2009;79:30-36.
4. Asquith J, Gillgrass T, Mossey P. Three-dimensional imaging of orthodontic models: a pilot study. *European Journal of Orthodontics* 2007;29:517-522.
5. Bootvong K, Liu Z, McGrath C, Hagg U, Wong RW, Bendeus M et al. Virtual model analysis as an alternative approach to plaster model analysis: reliability and validity. *European Journal of Orthodontics* 2010.
6. Cha BK, Choi JI, Jost-Brinkmann PG, Jeong YM. Applications of three-dimensionally scanned models in orthodontics. *International Journal of Computerized Dentistry* 2007;10:41-52.
7. Goonewardene RW, Goonewardene MS, Razza JM, Murray K. Accuracy and validity of space analysis and irregularity index measurements using digital models. *Australian Orthodontic Journal* 2008;24:83-90.
8. Horton HM, Miller JR, Gaillard PR, Larson BE. Technique comparison for efficient orthodontic tooth measurements using digital models. *Angle Orthod* 2009;80:254-261.
9. Keating AP, Knox J, Bibb R, Zhurov AI. A comparison of plaster, digital and reconstructed study model accuracy. *Journal of Orthodontics* 2008;35:191-201; discussion 175.

10. Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2007;132:346-352.
11. Oosterkamp BC, van der Meer WJ, Rutenfrans M, Dijkstra PU. Reliability of linear measurements on a virtual bilateral cleft lip and palate model. *Cleft Palate Craniofac J* 2006;43:519-523.
12. Quimby ML, Vig KWL, Rashid RG, Firestone AR. The accuracy and reliability of measurements made on computer-based digital models. *Angle Orthodontist* 2004;74:298-303.
13. Santoro M, Galkin S, Teredesai M, Nicolay OF, Cangialosi TJ. Comparison of measurements made on digital and plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2003;124:101-105.
14. Sjogren AP, Lindgren JE, Huggare JA. Orthodontic Study Cast Analysis-Reproducibility of Recordings and Agreement Between Conventional and 3D Virtual Measurements. *J Digit Imaging* 2009.
15. 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. *American Journal of Orthodontics & Dentofacial Orthopedics* 2006;129:794-803.
16. 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.

17. 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.
18. Miras D, Sander FG. [The accuracy of holograms compared to other model measurements]. *Fortschritte der Kieferorthopadie* 1993;54:203-217.
19. Bell A, Ayoub AF, Siebert P. Assessment of the accuracy of a three-dimensional imaging system for archiving dental study models. *Journal of Orthodontics* 2003;30:219-223.
20. El-Zanaty HM, El-Beialy AR, Abou El-Ezz AM, Attia KH, El-Bialy AR, Mostafa YA. Three-dimensional dental measurements: An alternative to plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2010;137:259-265.
21. Naidu D, Scott J, Ong D, Ho CT. Validity, reliability and reproducibility of three methods used to measure tooth widths for bolton analyses. *Australian Orthodontic Journal* 2009;25:97-103.
22. Watanabe-Kanno GA, Abrao J, Miasiro Junior H, Sanchez-Ayala A, Lagravere MO. Reproducibility, reliability and validity of measurements obtained from Cere3 digital models. *Pesquisa Odontologica Brasileira = Brazilian Oral Research* 2009;23:288-295.
23. Fleming PS, Marinho V, Johal A. Orthodontic measurements on digital study models compared with plaster models: a systematic review. *Orthodontics & Craniofacial Research* 2011;14:1-16.
24. Sutherland SE. Evidence-based dentistry: Part I. Getting started. *J Can Dent Assoc* 2001;67:204-206.

25. Mayers M, Firestone AR, Rashid R, Vig KW. Comparison of peer assessment rating (PAR) index scores of plaster and computer-based digital models. *Am J Orthod Dentofacial Orthop* 2005;128:431-434.
26. Costalos PA, Sarraf K, Cangialosi TJ, Efstratiadis S. Evaluation of the accuracy of digital model analysis for the American Board of Orthodontics objective grading system for dental casts. *American Journal of Orthodontics & Dentofacial Orthopedics* 2005;128:624-629.
27. 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. *American Journal of Orthodontics & Dentofacial Orthopedics* 2007;131:51-56.
28. Hildebrand JC, Palomo JM, Palomo L, Sivik M, Hans M. Evaluation of a software program for applying the American Board of Orthodontics objective grading system to digital casts. *Am J Orthod Dentofacial Orthop* 2008;133:283-289.
29. Veenema AC, Katsaros C, Boxum SC, Bronkhorst EM, Kuijpers-Jagtman AM. Index of Complexity, Outcome and Need scored on plaster and digital models. *European Journal of Orthodontics* 2009;31:281-286.
30. 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. *Orthodontics & Craniofacial Research* 2008;11:90-95.
31. Macchi A, Carrafiello G, Cacciafesta V, Norcini A. Three-dimensional digital modeling and setup. *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;129:605-610.
32. Kau CH, Littlefield J, Rainy N, Nguyen JT, Creed B. Evaluation of CBCT digital models and traditional models using the Little's Index. *Angle Orthod* 2010;80:435-439.

Chapter 3. Study Model Analysis using CBCT-generated Virtual Study Models

3.1 Introduction

Study model analysis (SMA) is an important process for accurate diagnosis and treatment planning in dentistry. In performing a SMA, common diagnostic parameters are measured on dental study models¹. Conventionally this is done using plaster models, but the current trend is moving toward using virtual models.

In evaluating SMA on new modalities, the important considerations are reliability and validity. Reliability refers to the consistency with which a measurement can be made and validity refers to the ability to truly measure what is intended. Both can be assessed using mean difference and agreement (Intraclass correlation coefficient, ICC) between virtual models and plaster models.

Quantitative parameters, usually linear distances, are an important component of a full SMA, but qualitative (categorical) parameters can also be assessed, such as molar and canine Angle's classification, arch symmetry, size and shape. Since SMA has customarily been performed on plaster models, such measurements can be considered the gold standard. The true gold standard, however, exists inside the patient's mouth (direct teeth and/or dental arch measurements), but due to access, it may not be possible to obtain accurate measurements on

live teeth. In certain circumstances, however, when treatment plans call for extractions, those teeth can be kept thereafter for direct measurement.

Conventionally, dental study models are obtained by taking impressions of a patient's upper and lower dentition, which is often an uncomfortable intraoral procedure. Furthermore, impression-taking can be a resource-intensive process requiring chair time, staff time, lab time and material costs. From the impressions, it is then possible to produce physical plaster study models as well as three-dimensional (3D) virtual study models. Although these study models can be diagnostic representations of crowns, occlusal anatomy and their interrelationships²⁻⁵, they cannot show the relationship of the roots and other anatomic structures.⁶

One of the current diagnostic trends in dentistry is Cone Beam Computed Tomography (CBCT), which is a theoretically undistorted⁷ radiographic approach to visualizing anatomy in 3D. From such scans, it was recognized that one could manually segment the teeth individually, including the roots, digitally using computer software.⁶ In a less painstaking process that does not require manual segmentation by the end user, Anatomage (San Jose, CA) can create Anatomodels, virtual models generated from CBCT scans with teeth already segmented by their proprietary process which can later be individually manipulated and measured in 3D on a computer screen. The approach of using CBCT offers an intriguing alternative to obtaining study models from which SMA can be performed without taking impressions.

Lagravere *et al.* (2008)⁸ compared measurements on CBCT images to a coordinate measuring machine, and found that the accuracy of linear and angular measurements from titanium markers on a synthetic mandible were within 1 mm and 1 degree, respectively, and agreement as measured by ICC was near perfect when assessed across each of the three dimensional axes (x, y and z).

Linear measurements based on cephalometric landmarks of CBCT reconstructions were found by Periago *et al.* (2010)⁹ to be $1.13 \pm 1.47\%$ smaller than anatomic landmarks human skulls and these differences were found to be statistically significant but clinically acceptable. However, the study used dry skulls and the authors were unable to simulate soft tissue effects of attenuation on image quality and admitted that “the dimensional accuracy of 3D measurements would be somewhat less on patient derived data.” Distances such as Sella-Nasion or Gonion-Menton were investigated but differences in measurements involving the teeth, as would be performed in a model analysis, were not assessed.

Later, Ganguly *et al.* (2011)¹⁰ confirmed that linear measurements of bone height on CBCT in the presence of soft tissue was on average 0.31 ± 0.61 mm smaller compared to direct measurements on the same six cadaver heads. Although the sample size was small, these authors concluded that CBCT-based linear measurements were sufficiently accurate for clinical use. Still, measurements involving teeth were not assessed.

The use of CBCT serial slices to perform linear measurements of tooth length was investigated in a thesis study by Rosenblatt (2010)¹¹. A total sample of 26 subjects previously treatment planned for premolar extractions had CBCT scans taken with an iCAT (Imaging Sciences International, Hatfield, PA) set to a voxel resolution of 0.25 mm. From these patients, 48 extracted premolars were collected. Measurements of tooth lengths on the CBCT slices underestimated on average 1.6 mm (p-value < 0.001) and 95% CI (1.1,2.0) the true length as measured directly on extracted premolars.

A full SMA using CBCT-generated virtual study models has not yet been reported in the literature, but a few parameters have been previously validated against laser-acquired study models¹². In their study, Kau *et al.* (2010)¹² took a sample of 30 subjects and used virtual study models (OrthoCAD, Cadent, Fairview, NJ) as their gold standard to compare with Anatomodels generated from CBCT scans taken with a Galileos (Sirona, Charlotte, NC) cone beam scanner at a voxel resolution of 0.125 mm. In OrthoCAD models and Anatomodels, the mean maxillary Little's Index scores, a relative measure of crowding, were found to be 9.65 mm and 8.87 mm, respectively; similarly, for the mandibular teeth, the mean scores were 6.41 mm and 6.27 mm, respectively. Again, comparing OrthoCAD models to Anatomodels, the mean overjet measurements were 2.29 mm and 2.26 mm, respectively, while mean overbite measurements were 2.29 and 2.26 mm. By way of paired t-test, no statistical significance could be demonstrated for all comparable measurements.

Tarazona *et al.* (2011)¹³ conducted a related study comparing Anatomodels to static two dimensional images of the occlusal perspective of plaster models. The following linear parameters were investigated on 27 subjects: mesiodistal tooth sizes, maxillary and mandibular intercanine and intermolar widths, as well as arch perimeters. Statistically significant differences at a 0.05 alpha level were found for mean dimensions of teeth 1-4, 2-6, 3-4, 4-5, mandibular intercanine width and arch perimeter. However, with mean differences less than 0.5 mm or up to 1%, no clinical significance could be stated. Pearson's correlation for parameters grouped by mesiodistal tooth sizes, or intercanine widths, intermolar widths and arch perimeters were above 0.99. Intra-rater reliability over three trials was found to be acceptable but inter-rater reliability was not assessed.

Despite the opportunity to utilize study models acquired via CBCT, a practical consideration before implementing this in practice is the time required to perform a full SMA. Tomassetti *et al.* (2001)¹⁴ reported average times performing Bolton analyses of 8 minutes and 4 seconds using plaster and 5 minutes and 16 seconds using virtual models. In agreement with these findings, Mullen *et al.* (2007)¹⁵ found that Bolton analyses on average 1 minute and 4 seconds slower on plaster than corresponding virtual models. Similarly, Tarazona *et al.* (2011)¹³ found that the average time to perform linear measurements on both arches using static scans of plaster models took 4 minutes 56 seconds compared with Anatomodels which took 3 minutes 8 seconds. On the other hand, Horton *et al.* (2010)¹⁶ found that average time to measure the plaster study models was 4 minutes and 15 seconds while measurements that involved freely rotating the virtual models on-screen took on average 7 minutes and 1 second.

The purpose of this study was to investigate the reliability, validity and time requirements of quantitative and qualitative measurements in a full SMA using Anatomodels compared with plaster dental study models as well as a subset of extracted premolars.

3.2 **Methods and Materials**

Approval for this study was granted by the University of Alberta, Health Research Ethics Board, Biomedical panel under the study ID Pro00010202. The subset of patients from this study who had extracted premolars originated from another study¹¹ that was granted approval under the ID Pro00002248.

In this study, we could not expect to identify differences of less than 0.5 mm since the resolution of some of the reconstructed CBCT scans were at 0.3x0.3x0.3 mm voxels while the subset of patients with extracted premolars that were a part of the other study¹¹ was scanned at 0.25x0.25x0.25 mm voxels. This is because any linear measurement involves the identification of two points; thus, the smallest measurable distance is between the centers of two adjacent voxels. At worst, the neighboring voxels will not be bordering one another such that a common side is shared, but rather, diagonally adjacent to each other such that a vertex on each voxel touch. The distance from the center of one voxel to the vertex and then to the center of the adjacent voxel is equivalent to the length of the diagonal of a voxel. For a cube of side x , the length of the diagonal is equal to $x\sqrt{3}$. In this study, we would expect the poorest

resolution to come from the voxels with 0.3 mm sides because it is larger than the voxels with 0.25 mm sides. Therefore, the precision of measurements with voxels of 0.3 mm sides is $0.3\sqrt{3}$ or 0.52 mm (Figure 3-1). Measurements smaller than 0.5 mm, then, should be interpreted with caution. With larger measurements such as arch perimeter, larger variability can be expected, and from a clinical standpoint, if the mean differences were greater than 2.0 mm, then the measurement would have little diagnostic value. For these reasons, we assumed thresholds for clinically relevant mean differences for 2-landmark linear measurements of 0.5 mm and for >2-landmark linear measurements of 2.0 mm.

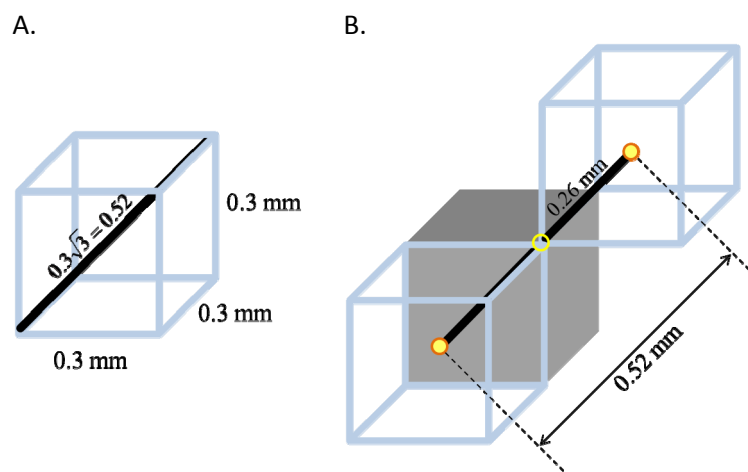


Figure 3-1. Precision using voxels: A, The dimensions of a voxel showing the length of the diagonal; B, The precision of measuring the distance between two adjacent voxels may, at worst, be 0.52 mm, so values lower than this amount should be interpreted with caution.

Differences of 0.5 mm for tooth widths and 5% for larger measurements were determined to be clinically significant by Asquith *et al.* (2007)¹⁷. Furthermore, Goonewardene *et al.* (2008)¹⁸ argued that extraction versus non-extraction treatment plans could be influenced by variations of 1-2 mm in crowding measurements. But at less than 1.5 mm of tooth structure discrepancy

in an arch, Mullen *et al.* (2007)¹⁵ decided that this could be clinically insignificant. Our proposed thresholds for clinical relevance of 0.5 mm for 2-landmark linear measurements and 2.0 mm for >2-landmark linear measurements, then, would be in line with these authors.

3.2.1 Sample size calculation and selection

Based on the data from a previous study comparing virtual models to plaster⁵ we took a standard deviation, σ , of 0.58 and set a statistical power, $1-\beta$, of 0.9 to detect a difference, δ , between Anatomodels and plaster of 0.5 mm at a significance level, α , of 0.05. The sample size was calculated applying 0.5 mm as the mean difference using the following equation specified by Rosner (2010)¹⁹:

$$n = \frac{\sigma^2(z_{1-\beta} + z_{1-\alpha/2})^2}{\delta^2} \quad (1)$$

Mathematically, one can note from equation 1 that the sample size requirement would be increased as a function of the following: greater sample variance (increased σ^2), reduction in significance level (decreased α), larger power (increased $1-\beta$), or smaller desired detectable difference (decreased δ).

The relevant scenarios and their effect on the sample size calculation for detectable differences of 0.5 mm as well as 2.0 mm are compared in Table 3-1. When only one variable is considered

at the $\alpha = 0.05$ level, the minimum sample size would be 15. However, we originally sought to investigate 13 variables in a split mouth study approach, so a Bonferroni-corrected alpha, $\alpha = 0.05/13$, was applied from which a minimum sample size of 24 was derived. When multiple t tests are performed over 13 variables, the Bonferroni correction was necessary to minimize the possibility of falsely declaring statistical differences by chance alone. It was ultimately decided that many more variables need to be considered for a full SMA, so a larger sample size requirement was expected. The Bonferroni-corrected alpha for a total of 36 linear variables was calculated as $\alpha = 0.05/36$, but this only mildly increased the minimum sample size to 27. A target sample of around 30 patients was decided as this size was consistent with similar studies^{15,16,20-26} investigating the validity and reliability of three-dimensional study models.

Table 3-1. Sample size calculations to detect differences of 0.5 mm or 2.0 mm, comparing projected sample sizes for 1 variable, and Bonferroni adjustments for 13 variables as well as 36 variables.

	σ	α	$z_{1-\alpha/2}$	β	$z_{1-\beta}$	Minimum Sample Size	
						$\delta = 0.5 \text{ mm}$	$\delta = 2.0 \text{ mm}$
No adjustments, 1 variable	0.58	0.05	1.9600	0.1	1.2816	15	1
Bonferroni-adjusted, 13 variables	0.58	0.004	2.8905	0.1	1.2816	24	2
Bonferroni-adjusted, 36 variables	0.58	0.001	3.1970	0.1	1.2816	27	2

This study ultimately investigated a retrospective sample of 30 consecutive patients chosen from the University of Alberta graduate orthodontic clinic between February 2007 to November 2009. The inclusion criteria were patients with fully erupted permanent dentition whose diagnostic records included good quality plaster study models and CBCT scans. Patients were excluded if the plaster models contained obvious chips or bubbles or the CBCT scans showed evidence of movement artifacts.

As this was a retrospective study, orthodontic treatment planning was independently completed and 11 patients were prescribed premolar extraction therapy. A total of twenty-two extracted premolars used in a separate study¹¹ were available for direct assessment of mesiodistal widths in this study. The teeth were inspected to ensure no obvious cracks or chips and then preserved in ethanol.

3.2.2 Parameters used in a Study Model Analysis (SMA)

Commonly used qualitative and quantitative parameters were included in the full SMA used for this study. The qualitative parameters were assessed by viewing perpendicular to the buccal surfaces of the teeth in question for Angle classifications, and from the occlusal view for arch form assessments. Bolton ratios were converted to millimeter differences with positive amounts corresponding to mandibular tooth mass excess (or maxillary deficiency) and negative amounts corresponding to maxillary tooth mass excess (or mandibular deficiency). The quantitative parameters were all linear measurements, which were further grouped by those requiring only two landmarks (2-landmarks), and those requiring more than two landmarks (>2-landmarks). Note that the only parameters that apply to the extracted premolar group were the mesiodistal widths of teeth 14, 15, 24 and 25.

The following ten qualitative (categorical) parameters and their levels were used:

- i. Right molar Angle classification (I, II, or III)
- ii. Right canine Angle classification (I, II, or III)
- iii. Left molar Angle classification (I, II, or III)
- iv. Left canine Angle classification (I, II, or III)
- v. Maxillary Arch Symmetry (symmetric or asymmetric)
- vi. Maxillary Arch Size (narrow, average, or expanded)
- vii. Maxillary Arch Shape (U-shaped, V-shaped, tapered, or squared)
- viii. Mandibular Arch Symmetry (symmetric or asymmetric)
- ix. Mandibular Arch Size (narrow, average, or expanded)
- x. Mandibular Arch Shape (U-shaped, V-shaped, tapered, or squared)

Thirty quantitative 2-landmark parameters included:

- i. Overjet
- ii. Overbite
- iii. Maxillary Intermolar Width
- iv. Maxillary Inter canine Width
- v. Mandibular Intermolar Width
- vi. Mandibular Inter canine Width
- vii. Twelve (12) Maxillary mesiodistal tooth sizes
16, 15, 14, 13, 12, 11, 21, 22, 23, 24, 25, 26
- viii. Twelve (12) Mandibular mesiodistal tooth sizes
46, 45, 44, 43, 42, 41, 31, 32, 33, 34, 35, 36

Six quantitative >2-landmark parameters included:

- i. Maxillary arch perimeter (four segments, mesial to first molars)
- ii. Maxillary arch crowding (mesial to first molars)
- iii. Mandibular arch perimeter (four segments, mesial to first molars)
- iv. Mandibular arch crowding (mesial to first molars)
- v. Bolton 6 – anterior ratio
- vi. Bolton 12 – overall ratio

Additionally, the time at the start of the SMA and the time at the end were recorded so that the actual time used to perform the full SMA could be calculated.

3.2.3 Modalities of assessment

Three modalities were compared in this study: Plaster dental study models, CBCT-generated study models, and matched samples of extracted premolars.

3.2.3.1 Conventional plaster study models

During the records-taking process, each patient received alginate impressions and immediately afterwards, the models were poured up and trimmed by an in-house lab technician at the University of Alberta as per the manufacturer's instruction. By inspection, the plaster models

must have been of good quality with no obvious chips or bubbles, otherwise, the patient was excluded from the study.

3.2.3.2 Virtual models generated from Cone Beam Computed Tomography scans

CBCT scans for the subjects with extracted premolars were taken with the 12-bit iCAT (Imaging Sciences International, Hatfield, PA) set to a 40 second scan, 120 kVp, 47 mAs, to allow image reconstruction into DICOM format at 0.25x0.25x0.25 mm voxels. The rest of the CBCT scans using the same iCAT machine were prescribed at 120 kVp, 24 mAs and voxel sizes of 0.30x0.30x0.30 mm. The DICOM datasets were uploaded to Anatomage (San Jose, CA) and processed into Anatomodels, the company's product name for CBCT-generated study models. Aside from the customary initial orthodontic records fee, patients were not assessed additional fees for the CBCT scan or the Anatomodels.

3.2.3.3 Extracted premolars from matched samples

A subset of 11 patients from the total 30 in this study underwent premolar extractions as a part of their orthodontic treatment. Consequently, matched samples of twenty-two extracted premolars were available from which direct mesiodistal width measurements were made.

3.2.4 Experimental Design

A flowchart for the research plan is illustrated in Figure 3-2. The gold standard for SMA was using plaster study model, but for a subset of the sample, we had extracted premolars from which to compare true mesiodistal width dimensions. Extracted teeth, if measured correctly, represent an ideal gold standard as distortion from impression taking and pouring is avoided. Our experimental comparison was the CBCT-generated study model, or in other words, the Anatomodel. Our experimental design had three arms of study across all three modalities: intra-rater reliability, inter-rater reliability, and validity. In addition, the time spent to perform each SMA on plaster and Anatomodels was tracked.

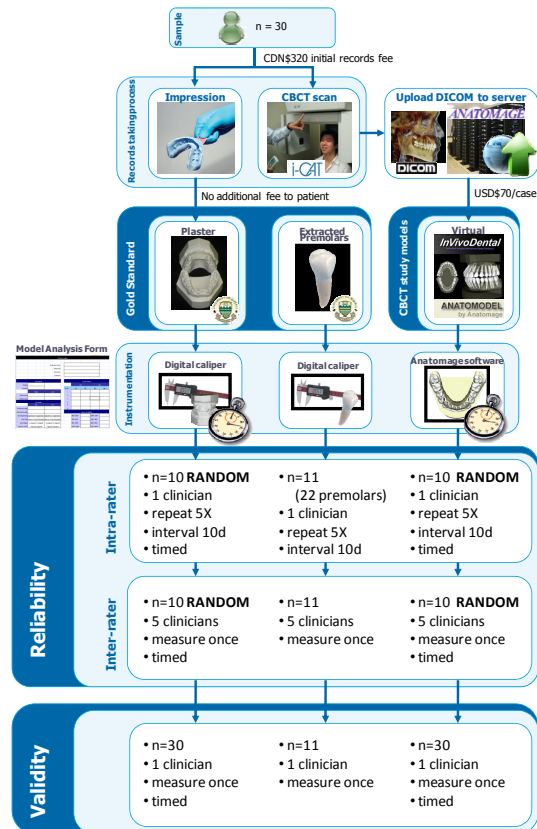


Figure 3-2. Study Flow Chart.

Intra-rater reliability was assessed over five trials as performed by one senior orthodontic resident. Ten subjects were randomly chosen from the subset of eleven subjects who had extracted premolars so that useful comparisons across the three modalities could be made. For both plaster and Anatomodels, timed SMA was repeated for 10 subjects five times at intervals of ten days apart, with assessments limited to five unique cases per day in random order to minimize bias due to fatigue. Similarly, twenty-two extracted premolars were measured in random order from the subset of 11 subjects, repeated five times at ten day intervals.

Inter-rater reliability was assessed across five evaluators: one senior orthodontic resident, and orthodontists of 0.5, 1, 16, and 23 years of clinical experience. Data collected from the last trial in the intra-rater reliability study was used since the most efficient ways to use the software, learned from the preceding four trials, were then taught to the participants of the inter-rater reliability study. Timed SMA was performed for both plaster and Anatomodels from the 10 subjects used in the intra-rater reliability study in random order, and then twenty-two extracted premolars were measured in random order from the subset of 11 subjects.

For the validity studies, the same senior orthodontic resident performed timed SMA on 30 subjects in random order on Anatomodels and then plaster, limited to only five cases per day. Twenty-two extracted premolars were measured in random order from the subset of 11 subjects.

The time to perform the full SMA in each trial of the studies on intra-rater reliability, inter-rater reliability and validity was recorded by the principle investigator.

3.2.5 Data Collection

The recording of data from the SMA was manually written into copies of the form shown in Figure 3-3, and then tabulated in a spreadsheet using Excel 2007 (Microsoft, Redmond, WA). Categorical measures were coded into numbers and arch perimeter in plaster was entered as the sum of four separate segments.

Model Analysis				
		Evaluator name:	<input type="text"/>	
		Patient ID:	<input type="text"/>	
		Start time:	<input type="text"/>	
		End time:	<input type="text"/>	
Articulated				
Overjet	<input type="text"/>	<input type="text"/>		
Overbite	<input type="text"/>	<input type="text"/>		
	Right	Left		
Molar class	<input type="text"/>	<input type="text"/>		
Canine class	<input type="text"/>	<input type="text"/>		
Occlusals				
	Mx	Mn		
Intermolar width	<input type="text"/>	<input type="text"/>		
Intercanine width	<input type="text"/>	<input type="text"/>		
Arch Symmetry	Symmetric / Asymmetric	Symmetric / Asymmetric		
Arch Size	Narrow / Avg / Expanded	Narrow / Avg / Expanded		
Arch Shape <i>(please circle)</i>	U-shaped / V-shaped/ Tapered / Squared	U-shaped / V-shaped/ Tapered / Squared		
Tooth Widths				
	Mx		Mn	
Tooth	Q1	Q2	Q3	Q4
1	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
2	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
3	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
4	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
5	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
6	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Space Available				
16M - 13M	<input type="text"/>	36M - 33M	<input type="text"/>	
13M - 11M	<input type="text"/>	33M - 31M	<input type="text"/>	
11M - 23M	<input type="text"/>	31M - 43M	<input type="text"/>	
23M - 26M	<input type="text"/>	43M - 46M	<input type="text"/>	

Figure 3-3. Study Model Analysis form.

3.2.5.1 Measurements on conventional plaster study models

Linear measurements on plaster study models were performed using the same digital caliper (Model IP67, Mitutoyo Canada, Mississauga, ON) for all evaluators. The tips were ground to a fine point and each time the caliper was used, it was calibrated by approximating the tips and pressing the “origin” button. The product specifications stated a resolution of 0.01 mm and an accuracy of ± 0.02 mm. Measurements of overbite and overjet were taken with a periodontal probe to the nearest 0.5 mm. The desired landmarks that correspond to each parameter of the SMA were reviewed with evaluators before they performed their measurements. For example, overjet was measured from the incisal edge of the most prominent upper incisor to the labial surface of the corresponding lower incisors.

3.2.5.2 Digital caliper measurements on matched-sample of extracted premolars

Linear measurements on extracted premolars were performed using the same digital caliper as that used for the plaster study models. After each use, the tips were disinfected using Caviwipes (Metrex, Orange, CA). Desired landmarks for mesiodistal widths of premolars were at the expected contact points at the height of contours on the mesial and distal surfaces of the premolars. This was reviewed before each evaluator performed their measurements. Though not a timed process, the extracted premolars were quickly measured and all returned to their vials of ethanol within three to five minutes.

3.2.5.3 Software measurements from CBCT-generated study models


Anatomodels were viewed using the software InVivo 5.0 build 229 (Anatomage, San Jose, CA) and linear measurements were shown onscreen to the nearest 0.01 mm. Arch perimeter measurements were conducted using the multiple-points measurement tool which automatically calculated the sum of the four segments and returned a single value.

All evaluators were given a five-minute tutorial on how to perform the measurements using the software and had a chance to practice on a sample Anatomodel not included in this study.

Figure 3-4 illustrates the procedure used to measure the mesiodistal tooth dimension for one tooth in Anatomodels.

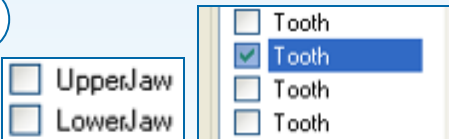
Like with the plaster models, the desired landmarks for each parameter of the SMA on Anatomodels were reviewed before evaluators performed their measurements. During the study, the primary investigator sat nearby to answer software-related questions as they arose.

1



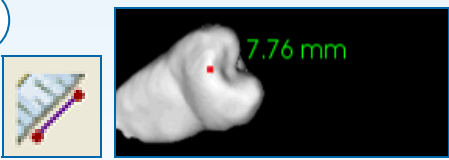
- Load the Anatomodel

2



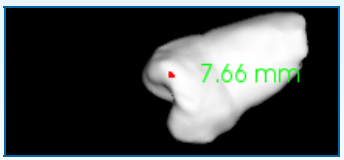
- Click to hide "UpperJaw"
- Click to hide "LowerJaw"
- Click the appropriate but often difficult to find checkbox to unhide the target tooth

3



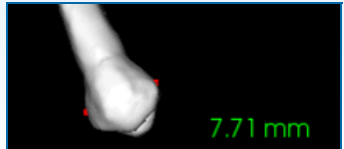
- Click the "Distance measurement" tool
- Manipulate the view to the distal aspect
- Click to place the first point on the distal

4



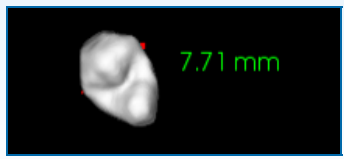
- Manipulate the view to the mesial aspect
- Click to place the final point on the mesial

5



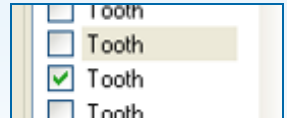
- Confirm landmark accuracy

6



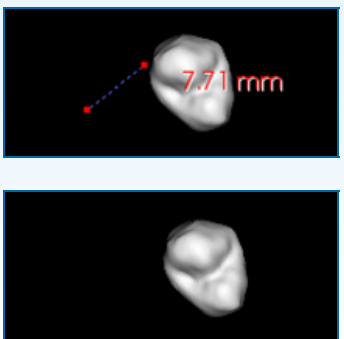
- Make note of the measurement value

7



- Click the appropriate checkboxes to unhide the next tooth and hide the tooth we just measured

8



- Click the measurement and press the delete button to keep the screen uncluttered
- Repeat from Step 3 for each additional tooth

Figure 3-4. Step-by-step pictorial instructions for performing a single tooth width measurement with Anatomodels in the software InVivo 5.0 build 229.

3.2.5.4 Timed measurements using computer operating system time clock

The time required, in minutes, to perform all of the intended measurements in a SMA was calculated by recording the start and finish times and then taking the difference. The times were read and announced by the primary investigator from the time clock of a computer operating system (Windows 7, Microsoft, Redmond, WA) and subsequently recorded by the evaluator.

3.2.6 Statistical Analyses

All of the tabulated data was transferred to a statistical software package (SPSS version 16, IBM, Armonk, NY) for analysis. The data in both reliability and validity studies was examined from different perspectives: 1) in terms of the pattern of responses, as in agreement; 2) in terms of the magnitude of measures, as in the mean differences, and 3) in terms of the extent of evidence for that difference, as in statistical significance. For the reliability and validity studies, agreement of measurements was assessed by way of intraclass correlation coefficient (ICC) and cross tabulations. Mean difference of measurements was investigated by way of paired *t* tests in the validity studies and repeated measures ANOVA in the reliability studies. Statistical significance was interpreted from the *p*-values; this is contrasted by clinical significance as defined in Section 3.2, which is a subjective assessment of both the degree of agreement and the magnitude of mean differences.

The agreement among multiple groups of raters or measurements is best estimated with ICC. Because ICC evaluates the true variance in multiple groups of ratings among all sources of variance, it is better than measures such as Pearson's correlation which can only measure the strength of the linear relationship between two variables. ICC values can range between 0, corresponding to no agreement, to 1, corresponding to perfect agreement. In this study, we will consider all ICC values above 0.8 as excellent, above 0.7 to be good, above 0.6 to be moderate, and below 0.6 to be poor. Because we had set high diagnostic expectations for our study, the proposed ICC scale is more demanding than the ranges suggested by Rosner (2010)¹⁹ who considered ICC values above 0.75 as excellent, between 0.4 to 0.75 as fair to good, and below 0.4 to be poor agreement.

In a complete assessment of clinical relevance, both high agreement and low mean differences are desirable; any other combination should be accepted with caution. Hypothetically, the validity of a given parameter on Anatomodels compared with plaster may be unacceptable despite the mean difference being close to zero (i.e. a magnitude judged to be clinically insignificant) because the ICC could also be close to zero (i.e. interpreted as poor agreement). The low mean difference could be explained by the fact that the average of the measurements across all subjects on Anatomodels for a given parameter is about the same as the average on plaster; but when the recorded values are inspected sequentially, poor agreement results from the fact that the measurements for each subject from Anatomodels do not match well with the corresponding values recorded from plaster. Such a combination of low mean difference but poor agreement may render the measurement untrustworthy from a clinical stand point.

Alternatively, the reliability across five trials of measurements for a given parameter may be unacceptable because it shows a high mean difference between one pair of trials (i.e. a magnitude judged to be clinically significant), despite a high ICC (i.e. interpreted as high agreement). This is because repeated measures ANOVA tests are used to screen for the presence of differences between at least two means; where this difference comes from can only be determined from pairwise comparisons of two trials at a time. If, for instance, the mean of the first trial is much lower than the remaining four, then the worst difference from all of the pairwise comparisons can cross the clinically significant threshold. On the other hand, ICC can be calculated across all five trials simultaneously, so four of the five trials may have excellent agreement, offsetting the moderate agreement from the first trial. Indeed, if such a scenario occurred and the first trial was eliminated from the analysis, the reliability could become acceptable.

In assessments of statistical significance, the p-value measures to what extent the data is consistent with the null hypothesis. A scientific statement can then be made without any consideration for the magnitude of the difference (i.e. clinical significance) for which it corresponds to. A p-value below the statistical significance level, in this case, $\alpha = 0.05$, is evidence against the null hypothesis and larger p-values represents insufficient evidence against the null hypothesis. In the paired-sample *t* tests, interpretation of the p-value translates to the extent the data demonstrates that a difference, regardless of how much difference, exists between the two groups compared. Alternatively, in the repeated measures

ANOVA, interpretation of the p-value tells us the strength of evidence that the means for at least one pair of within-subject factors differ, regardless of how much difference exists.

Statistical significance is an objective uniform decision-making criterion in hypothesis testing¹⁹ and should not be dismissed in lieu of clinical significance, which tends to rely on subjective impressions. Without evidence for a statistical difference, the mean difference, which is used to determine clinical significance, may be misleading since the 95% confidence interval may include zero (interpreted as no difference).

For repeated measures ANOVA tests, when the p-value is less than 0.05, this should be interpreted as evidence that a mean difference exists for one or more pairs out of the 10 possible two-group combinations. The p-value for repeated measures ANOVA relates to the tests on all five trials or evaluators simultaneously and should not be interpreted as between any two groups specifically. On the other hand, the p-value for validity and time within one evaluator should be interpreted as between the two groups specified in the tables.

To recapitulate, the results for the reliability studies will be presented in tables with ICC values and their 95% confidence intervals (CI) as well as the repeated measures ANOVA tests showing the worst and best differences for two groups chosen from five and then p-values for the ANOVA test result overall for each parameter. The results of the validity studies are presented in tables with ICC values and 95% CI, mean differences and 95% CI for paired-sample *t* tests and the p-values for the differences in each parameter.

As a note regarding the cross tabulations, it is statistically desirable to have at least 5 counts in any given cell; otherwise, as in this study, we could only report tendencies for the nominal parameters.

3.2.6.1 Intra-rater Reliability within each modality of assessment

Intra-rater reliability was assessed in terms of agreement and mean difference for 10 patients with one evaluator over five trials for the three modalities of interest separately: Anatomodels, plaster and extracted premolars.

Agreement for quantitative parameters was assessed by way of ICC. A two-way mixed model was used since the evaluator was not randomly selected, but fixed. We were interested in seeing if there were identical patterns of scores as opposed to similar patterns of scores, so we further set the test to absolute agreement. Individual ratings were the analysis of interest, so we read the ICC values from the results of single measure reliability.

Agreement of qualitative parameters was assessed by way of concordances in the cross tabulation of two trials randomly chosen from five. A concordant pair is when a pair of observations from two different trials is the same; if the observations were different from one trial to another, then this pair of observations is said to be discordant. The combination of comparisons of groups of two from five trials can be calculated as 5 choose 2, which equates to 10 combinations for each of the ten nominal parameters. In order to simplify the analysis,

cross-tabulations were performed on two trials randomly chosen out of five for each parameter. Agreement was reported as a percentage of the number of observed concordant pairs over the total number of possible concordant pairs across all ten nominal parameters.

To quantify the mean difference between the repeated measurements, a repeated measures ANOVA was performed using Trial (1, 2, 3, 4, 5) as the within-subjects factor and the output organized by groups based on the parameter. Bonferroni comparisons were applied to the main effects and the pairwise mean difference for each parameter was summarized in terms of the best and worst mean differences. The reported p-values were read from the tests of within-subjects effects taking into account Mauchley's Test of Sphericity.

3.2.6.2 Inter-rater Reliability within each modality of assessment

Inter-rater reliability was assessed similarly in terms of agreement and mean difference with five evaluators for 10 patients using Anatomodels, plaster and extracted premolars, separately. The fifth trial from the intra-rater reliability data was reused as the data for one of the evaluators in this inter-rater reliability analysis.

Agreement for quantitative parameters was assessed by way of ICC set at two-way mixed model and absolute agreement. The ICC values were read from the results of single measure reliability.

Agreement of qualitative parameters was assessed by way of concordances in the cross tabulation of two evaluators randomly chosen from five. Agreement was reported as a percentage of the number of observed concordant pairs over the total number of possible concordant pairs across all ten nominal parameters.

To quantify the mean difference between the repeated measurements, a repeated measures ANOVA was performed using Evaluator (NL, MM, CF, ML, TE) as the within-subjects factor and the output organized by groups based on the Parameter. Bonferroni comparisons were applied to the main effects and the pairwise mean difference for each parameter was summarized in terms of the best and worst mean differences. The reported p-values were read from the tests of within-subjects effects taking into account Mauchley's Test of Sphericity.

3.2.6.3 Validity between each modality of assessment

Validity was assessed in terms of agreement and mean difference for all 30 patients between pairwise combinations of Anatomodels, plaster and extracted premolars, separately.

Agreement for quantitative parameters was assessed by way of ICC set at two-way mixed model and absolute agreement. The ICC values were read from the results of single measure reliability.

Agreement of qualitative parameters was assessed by way of concordances in the cross tabulation of Anatomodels and plaster. Agreement was reported as a percentage of the number of observed concordant pairs over the total number of possible concordant pairs across all ten nominal parameters.

To quantify the mean difference between pairwise combinations of the three modalities, paired samples *t* tests were employed.

3.2.6.4 Timed study model analysis for each modality of assessment

The time to measure all of the parameters in a SMA was calculated for the Anatomodel and plaster groups during the intra-rater reliability, inter-rater reliability and validity studies. During the validity study, when comparing Anatomodels and plaster over all 30 cases, the mean difference was derived from a paired-samples T-test. During the reliability studies over 10 cases, repeated measures ANOVA was used separately for Anatomodels and plaster to expose the worst and best differences between the means of two time measurements out of five trials or evaluators.

3.3 Results

The raw data used in the statistical analyses are in Appendix 3-1. Histograms demonstrating normal distribution of differences in measurements for Anatomodels compared with Plaster, Anatomodels compared with extracted premolars, and plaster compared with extracted premolars, are presented in Appendix 3-2, Appendix 3-3, and Appendix 3-4, respectively. Model assumptions were checked and satisfied prior to performing the statistical tests. The following are the findings from this study.

3.3.1 Sample characteristics

The sample characteristics are summarized in Appendix 3-5 showing Gender and Mean Age for orthodontic records across the three modalities. Moreover, fifteen CBCT scans were taken with incompletely occluded arches, so it was not possible to assess Angle classification, overjet and overbite for these patients.

3.3.2 Intra-rater Reliability

A summary of the results from ICC and repeated measures ANOVA tests are presented for Anatomodels in Table 3-2, plaster in Table 3-3 and extracted premolars in Table 3-4.

Intra-rater reliability for 10 Anatomodels was excellent for most parameters across all five trials of measurements (Table 3-2). All parameters had ICC values above 0.8, so there was excellent agreement across the five trials; in support of this finding were the short 95% confidence intervals and this indicates low variability in agreement for all parameters. At the 0.05 significance level, there was strong statistical evidence (p -value <0.05) to show differences in the means of at least one pair of trials for overbite, tooth 1-6 and maxillary intermolar width; however, at worst, the differences were around 0.5 mm or less. For >2 -landmark linear measures, statistically, there was strong evidence (p -value <0.05) for a difference in the means of at least one pair of trials in maxillary arch perimeter and maxillary crowding, but the worst mean differences were less than 2 mm. When comparing the best mean differences between pairs of trials, no parameter exceeded 0.13 mm, while the worst mean differences between repeated measures were around 0.5 mm or less for 2-landmark measures and less than 2 mm for >2 landmark measures.

Intra-rater reliability for 10 plaster models was also excellent for most parameters across five trials of measurements (Table 3-3). The majority of parameters had ICC values well above 0.8, so there was excellent agreement across the five trials; in support of this finding were the tight 95% confidence intervals which indicate low variability in agreement for all parameters.

Although the 95% confidence intervals varied from poor to excellent across five trials for tooth 2-6, maxillary arch perimeter and maxillary crowding, the ICC values were above 0.7 and, therefore, agreement was acceptably moderate. There was evidence (p -value <0.05) to show differences in the means of at least one pair of trials for repeated measurements of teeth 1-1,

1-5, 2-2 and 3-2, but none of these parameters had mean differences greater than 0.5 mm. All >2-landmark parameters had mean differences that were less than 1.5 mm.

Intra-rater reliability for extracted premolars was excellent (Table 3-4). The sample size for each group of premolars was low but they were particularly low for teeth 1-5 and 2-5, so interpretations will only be attempted for teeth 1-4 and 2-4. All ICC values were close to 1.0 and worst mean differences were 0.02 mm with p-values greater than 0.05.

Cross tabulations for intra-rater reliability are presented for Anatomodels in Appendix 3-6 and plaster in Appendix 3-7. Based on the 96% concordant pairs in the cross tabulations (Appendix 3-6) of selected trial comparisons, there tended to be good intra-rater agreement for the qualitative parameters using Anatomodels. In other words, out of all trials chosen for comparison, paired observations were the same 96% of the time. Two instances of discordant pairs in Anatomodels arose from assessments of maxillary symmetry, and one instance from assessments of mandibular symmetry as well as maxillary arch shape. Furthermore, 96% concordant pairs suggest excellent intra-rater agreement among the qualitative parameters using plaster (Appendix 3-7). One instance each of discordant pairs arose from assessments of left canine classification, maxillary shape, maxillary size, and mandibular symmetry.

Table 3-2. Intra-rater, Anatomodels: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.

Parameter	N	Intra-rater Reliability		Mean Differences (mm)			
		ICC	95% CI	Worst	Best	p-value	
<i>Anatomodels, Linear measurements, 2 landmarks</i>							
Overjet	10	0.905	(0.788,0.971)	0.43	0.02	0.113	
Overbite	10	0.947	(0.871,0.985)	0.53	0.01	0.016	*
Tooth 1-1	10	0.871	(0.723,0.960)	0.10	0.01	0.728	
Tooth 1-2	10	0.975	(0.940,0.993)	0.03	0.00	0.987	
Tooth 1-3	10	0.916	(0.813,0.975)	0.10	0.01	0.412	
Tooth 1-4	10	0.919	(0.818,0.976)	0.06	0.00	0.869	
Tooth 1-5	10	0.927	(0.835,0.978)	0.09	0.00	0.432	
Tooth 1-6	10	0.913	(0.799,0.974)	0.28	0.05	0.015	*
Tooth 2-1	10	0.962	(0.911,0.989)	0.12	0.00	0.337	
Tooth 2-2	10	0.965	(0.917,0.990)	0.12	0.00	0.334	
Tooth 2-3	10	0.920	(0.820,0.976)	0.09	0.00	0.288	
Tooth 2-4	10	0.915	(0.809,0.975)	0.06	0.00	0.614	
Tooth 2-5	10	0.898	(0.777,0.969)	0.10	0.00	0.344	
Tooth 2-6	10	0.863	(0.711,0.958)	0.19	0.01	0.310	
Tooth 3-1	10	0.962	(0.909,0.989)	0.04	0.00	0.957	
Tooth 3-2	10	0.905	(0.790,0.972)	0.06	0.01	0.647	
Tooth 3-3	10	0.876	(0.734,0.962)	0.12	0.01	0.568	
Tooth 3-4	10	0.813	(0.623,0.940)	0.13	0.01	0.213	
Tooth 3-5	10	0.894	(0.767,0.968)	0.09	0.01	0.563	
Tooth 3-6	10	0.919	(0.819,0.976)	0.08	0.00	0.726	
Tooth 4-1	10	0.945	(0.873,0.984)	0.07	0.01	0.569	
Tooth 4-2	10	0.957	(0.900,0.987)	0.06	0.00	0.618	
Tooth 4-2a	1 ^a	-	-	0.56	0.01	-	
Tooth 4-3	10	0.866	(0.716,0.959)	0.11	0.00	0.482	
Tooth 4-4	10	0.867	(0.717,0.959)	0.17	0.00	0.136	
Tooth 4-5	10	0.977	(0.944,0.993)	0.07	0.01	0.212	
Tooth 4-6	10	0.902	(0.784,0.970)	0.15	0.00	0.335	
Mx_IMW	10	0.984	(0.959,0.995)	0.52	0.09	0.010	*
Mx_ICW	10	0.934	(0.849,0.980)	0.31	0.03	0.482	
Mn_IMW	10	0.965	(0.918,0.990)	0.27	0.00	0.639	
Mn_ICW	10	0.968	(0.924,0.991)	0.32	0.02	0.198	
<i>Anatomodels, Linear measurements, >2 landmarks</i>							
Mx_Perim	10	0.929	(0.802,0.980)	1.71	0.04	<0.001	*
Mx_Crowd	10	0.889	(0.746,0.967)	1.47	0.03	0.029	*
Mn_Perim	10	0.934	(0.850,0.981)	0.82	0.09	0.193	
Mn_Crowd	10	0.920	(0.820,0.976)	1.10	0.01	0.093	
Bolton6	10	0.936	(0.853,0.981)	0.30	0.01	0.575	
Bolton12	10	0.887	(0.775,0.966)	0.78	0.13	0.165	

^a Cannot be computed because the sum of caseweights is less than or equal 1.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval;

*, p-value <0.05; Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width;

ICW, Inter canine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

Table 3-3. Intra-rater, Plaster: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.

Parameter	N	Intra-rater Reliability		Mean Differences (mm)		
		ICC	95% CI	Worst	Best	p-value
<i>Plaster models, Linear measurements, 2 landmarks</i>						
Overjet	10	0.926	(0.832,0.978)	0.20	0.05	0.420
Overbite	10	0.935	(0.852,0.981)	0.20	0.00	0.723
Tooth 1-1	10	0.980	(0.946,0.995)	0.13	0.01	0.001 *
Tooth 1-2	10	0.989	(0.973,0.997)	0.07	0.00	0.358
Tooth 1-3	10	0.940	(0.862,0.982)	0.08	0.00	0.203
Tooth 1-4	10	0.946	(0.875,0.984)	0.05	0.01	0.688
Tooth 1-5	10	0.949	(0.878,0.985)	0.12	0.02	0.028 *
Tooth 1-6	10	0.931	(0.841,0.980)	0.13	0.01	0.054
Tooth 2-1	10	0.990	(0.976,0.997)	0.05	0.00	0.459
Tooth 2-2	10	0.980	(0.949,0.994)	0.14	0.03	0.027 *
Tooth 2-3	10	0.936	(0.855,0.981)	0.08	0.00	0.224
Tooth 2-4	10	0.830	(0.651,0.946)	0.15	0.01	0.274
Tooth 2-5	10	0.960	(0.906,0.988)	0.09	0.00	0.252
Tooth 2-6	10	0.784	(0.573,0.930)	0.08	0.00	0.860
Tooth 3-1	10	0.971	(0.932,0.992)	0.05	0.00	0.503
Tooth 3-2	10	0.970	(0.925,0.991)	0.09	0.01	0.012 *
Tooth 3-3	10	0.942	(0.867,0.983)	0.10	0.00	0.120
Tooth 3-4	10	0.901	(0.781,0.970)	0.04	0.00	0.889
Tooth 3-5	10	0.847	(0.677,0.952)	0.05	0.00	0.813
Tooth 3-6	10	0.937	(0.852,0.982)	0.21	0.02	0.089
Tooth 4-1	10	0.902	(0.782,0.970)	0.07	0.00	0.600
Tooth 4-2	10	0.941	(0.864,0.982)	0.06	0.00	0.420
Tooth 4-2a	1 ^a	-	-	0.09	0.01	-
Tooth 4-3	10	0.952	(0.887,0.986)	0.12	0.01	0.062
Tooth 4-4	10	0.945	(0.873,0.984)	0.05	0.00	0.689
Tooth 4-5	10	0.901	(0.782,0.970)	0.14	0.00	0.118
Tooth 4-6	10	0.964	(0.916,0.990)	0.10	0.01	0.274
Mx_IMW	10	0.992	(0.981,0.998)	0.18	0.04	0.648
Mx_ICW	10	0.985	(0.963,0.996)	0.11	0.01	0.843
Mn_IMW	10	0.977	(0.945,0.993)	0.13	0.00	0.886
Mn_ICW	10	0.978	(0.947,0.994)	0.15	0.01	0.750
<i>Plaster models, Linear measurements, >2 landmarks</i>						
Mx_Perim	10	0.794	(0.593,0.934)	1.33	0.04	0.358
Mx_Crowd	10	0.735	(0.502,0.934)	1.31	0.16	0.394
Mn_Perim	10	0.927	(0.833,0.978)	0.76	0.02	0.095
Mn_Crowd	10	0.915	(0.799,0.975)	1.27	0.16	0.006 *
Bolton6	10	0.934	(0.848,0.980)	0.18	0.01	0.700
Bolton12	10	0.899	(0.779,0.970)	0.36	0.04	0.623

a. Cannot be computed because the sum of caseweights is less than or equal 1.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval;

*, p-value <0.05; Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width;

ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12,

Bolton millimeter, positive when Mandibular Excess

Table 3-4. Intra-rater, Extracted Premolars: ICC and repeated measures ANOVA mean differences shown for mesiodistal width measurements of each extracted premolar.

Parameter	N	Intra-rater Reliability		Mean Differences (mm)		
		ICC	95% CI	Worst	Best	p-value
<i>Extracted Premolars</i>						
Tooth 14	8	0.998	(0.995,1.000)	0.02	0.00	0.532
Tooth 15	3 ^a					
Tooth 24	9	0.999	(0.997,1.000)	0.02	0.00	0.177
Tooth 25	2 ^a					

a. Test values not reported due to low sample size.
Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval

3.3.3 Inter-rater Reliability

A summary of the results from ICC and repeated measures ANOVA tests are presented for Anatomodels in Table 3-5, plaster in Table 3-6 and extracted premolars in Table 3-7.

Inter-rater reliability for 10 Anatomodels was moderate to excellent for most parameters as measured among five evaluators (Table 3-5). The ICC values for most parameters varied around 0.8, suggesting good to excellent agreement. The measurements of tooth 3-4 and maxillary crowding had ICC values that were below 0.6 with wide 95% confidence intervals, suggesting that these parameters had unacceptably poor and highly variable agreement. There was statistical evidence (p -value < 0.05) to show that differences existed between the means of at least two evaluators for measurements of overbite, teeth 2-3, 2-5, 3-1, 4-1, maxillary intermolar width, mandibular intercanine width, maxillary and mandibular perimeter and crowding, and the mean differences were above the clinical thresholds except for measurements of teeth widths.

Inter-rater reliability for 10 plaster models was moderate to excellent for most parameters as measured among five evaluators (Table 3-6). The ICC values for most parameters varied around 0.8, suggesting predominantly good agreement. The ICC values were below 0.6 and 95% confidence intervals wide for measurements of overjet, tooth 2-3, and mandibular perimeter, suggesting that these parameters had unacceptably poor and highly variable agreement. There was statistical evidence (p -value < 0.05) to show that differences existed between the means of at least two evaluators for measurements of overbite, overjet, tooth 1-6, 2-4, 3-4, 4-4, 4-6, maxillary and mandibular intercanine widths, arch perimeter and crowding; the mean differences exceed clinically relevant thresholds for all of these parameters except for the teeth width measurements and maxillary crowding.

Inter-rater reliability for the extract teeth was high (Table 3-7). Similar to intra-rater reliability, interpretations will only be attempted for teeth 1-4 and 2-4. All ICC values were above 0.9 and worst mean differences were less than 0.2 mm with p -values greater than 0.05.

Cross tabulations for inter-rater reliability are presented for Anatomodels in Appendix 3-8 and plaster in Appendix 3-9. With 81% overall concordant pairs (Appendix 3-8) of selected trial comparisons, there tended to be good inter-rater agreement for the qualitative parameters using Anatomodels. Discordant pairs in Anatomodels arose from assessments of left molar and canine classifications, right canine classification, maxillary arch shape, size and symmetry, as well as mandibular size and symmetry. Of these parameters, only maxillary shape had a potentially excessive number of discordances for five out of a possible ten pairs. Furthermore, 73% overall concordant pairs suggest good inter-rater agreement among the qualitative parameters using plaster (Appendix 3-9). Discordant pairs in plaster were much more frequent than with Anatomodels, affecting all parameters but right molar classification and mandibular symmetry. The parameters that had potentially excessive number of discordances of four or more out of ten pairs included maxillary shape, size and symmetry, as well as mandibular shape and size.

Table 3-5. Inter-rater, Anatomodels: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.

Parameter	N	Inter-rater Reliability		Mean Differences (mm)			
		ICC	95% CI	Worst	Best	p-value	
<i>Anatomodels, Linear measurements, 2 landmarks</i>							
Overjet	10	0.864	(0.710,0.958)	0.21	0.02	0.808	
Overbite	10	0.906	(0.785,0.972)	0.69	0.04	0.021	*
Tooth 1-1	10	0.909	(0.797,0.973)	0.09	0.01	0.765	
Tooth 1-2	10	0.939	(0.857,0.982)	0.24	0.02	0.050	
Tooth 1-3	10	0.771	(0.556,0.925)	0.19	0.01	0.445	
Tooth 1-4	10	0.795	(0.592,0.934)	0.23	0.01	0.145	
Tooth 1-5	10	0.818	(0.631,0.942)	0.21	0.02	0.138	
Tooth 1-6	10	0.867	(0.718,0.959)	0.16	0.01	0.615	
Tooth 2-1	10	0.904	(0.787,0.971)	0.17	0.02	0.356	
Tooth 2-2	10	0.619	(0.351,0.862)	0.38	0.01	0.438	
Tooth 2-3	10	0.771	(0.505,0.928)	0.34	0.03	0.001	*
Tooth 2-4	10	0.734	(0.500,0.910)	0.17	0.01	0.564	
Tooth 2-5	10	0.691	(0.431,0.894)	0.32	0.01	0.005	*
Tooth 2-6	10	0.735	(0.497,0.912)	0.11	0.00	0.964	
Tooth 3-1	10	0.942	(0.866,0.983)	0.15	0.01	0.129	
Tooth 3-2	10	0.684	(0.412,0.892)	0.31	0.01	0.013	*
Tooth 3-3	10	0.850	(0.684,0.954)	0.07	0.01	0.927	
Tooth 3-4	10	0.559	(0.289,0.830)	0.32	0.00	0.080	
Tooth 3-5	10	0.771	(0.556,0.925)	0.21	0.00	0.313	
Tooth 3-6	10	0.815	(0.627,0.941)	0.23	0.01	0.186	
Tooth 4-1	10	0.808	(0.610,0.939)	0.23	0.00	0.035	*
Tooth 4-2	10	0.827	(0.647,0.945)	0.18	0.01	0.309	
Tooth 4-2a	1 ^a	-	-	0.41	0.00	-	
Tooth 4-3	10	0.799	(0.600,0.935)	0.13	0.03	0.610	
Tooth 4-4	10	0.739	(0.508,0.913)	0.17	0.01	0.511	
Tooth 4-5	10	0.818	(0.631,0.942)	0.21	0.03	0.362	
Tooth 4-6	10	0.729	(0.495,0.908)	0.28	0.04	0.215	
Mx_IMW	10	0.837	(0.663,0.949)	0.75	0.05	0.021	*
Mx_ICW	10	0.749	(0.524,0.916)	1.11	0.07	0.178	
Mn_IMW	10	0.860	(0.701,0.957)	1.17	0.02	0.120	
Mn_ICW	10	0.934	(0.832,0.981)	0.72	0.02	0.001	*
<i>Anatomodels, Linear measurements, >2 landmarks</i>							
Mx_Perim	10	0.679	(0.268,0.902)	5.72	0.13	<0.001	*
Mx_Crowd	10	0.566	(0.188,0.849)	5.51	0.87	<0.001	*
Mn_Perim	10	0.776	(0.495,0.931)	3.03	0.08	<0.001	*
Mn_Crowd	10	0.742	(0.481,0.916)	3.00	0.11	<0.001	*
Bolton6	10	0.884	(0.749,0.964)	0.39	0.06	0.437	
Bolton12	10	0.833	(0.654,0.947)	0.54	0.01	0.815	

a. Cannot be computed because the sum of caseweights is less than or equal 1.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval;

*, p-value <0.05; Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width;

ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12,

Bolton millimeter, positive when Mandibular Excess

Table 3-6. Inter-rater, Plaster: ICC and repeated measures ANOVA mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.

Parameter	N	Inter-rater Reliability		Mean Differences (mm)		
		ICC	95% CI	Worst	Best	p-value
<i>Plaster models, Linear measurements, 2 landmarks</i>						
Overjet	10	0.550	(0.249,0.829)	1.35	0.05	<0.001 *
Overbite	10	0.771	(0.460,0.931)	1.65	0.05	<0.001 *
Tooth 1-1	10	0.923	(0.826,0.977)	0.15	0.01	0.195
Tooth 1-2	10	0.639	(0.377,0.870)	0.39	0.01	0.330
Tooth 1-3	10	0.712	(0.470,0.902)	0.13	0.02	0.451
Tooth 1-4	10	0.830	(0.649,0.946)	0.22	0.03	0.078
Tooth 1-5	10	0.856	(0.697,0.955)	0.18	0.00	0.189
Tooth 1-6	10	0.823	(0.624,0.945)	0.34	0.01	0.030 *
Tooth 2-1	10	0.949	(0.883,0.985)	0.11	0.01	0.401
Tooth 2-2	10	0.965	(0.917,0.990)	0.07	0.00	0.837
Tooth 2-3	10	0.587	(0.318,0.845)	0.25	0.00	0.168
Tooth 2-4	10	0.841	(0.657,0.951)	0.22	0.01	0.032 *
Tooth 2-5	10	0.866	(0.715,0.959)	0.14	0.01	0.183
Tooth 2-6	10	0.702	(0.458,0.897)	0.31	0.06	0.125
Tooth 3-1	10	0.842	(0.669,0.951)	0.06	0.00	0.847
Tooth 3-2	10	0.900	(0.780,0.970)	0.13	0.00	0.280
Tooth 3-3	10	0.890	(0.760,0.967)	0.07	0.01	0.762
Tooth 3-4	10	0.874	(0.704,0.963)	0.23	0.02	0.001 *
Tooth 3-5	10	0.843	(0.672,0.951)	0.20	0.00	0.077
Tooth 3-6	10	0.810	(0.597,0.941)	0.45	0.06	0.002 *
Tooth 4-1	10	0.935	(0.851,0.981)	0.10	0.02	0.129
Tooth 4-2	10	0.855	(0.693,0.955)	0.06	0.00	0.764
Tooth 4-2a	1 ^a	-	-	0.54	0.00	-
Tooth 4-3	10	0.888	(0.753,0.966)	0.19	0.03	0.055
Tooth 4-4	10	0.875	(0.719,0.962)	0.22	0.01	0.027 *
Tooth 4-5	10	0.834	(0.659,0.948)	0.18	0.01	0.203
Tooth 4-6	10	0.826	(0.613,0.947)	0.36	0.01	0.001 *
Mx_IMW	10	0.854	(0.683,0.955)	1.93	0.05	0.081
Mx_ICW	10	0.957	(0.893,0.988)	0.52	0.09	0.013 *
Mn_IMW	10	0.939	(0.859,0.982)	0.45	0.04	0.131
Mn_ICW	10	0.905	(0.775,0.972)	0.92	0.06	0.027 *
<i>Plaster models, Linear measurements, >2 landmarks</i>						
Mx_Perim	10	0.838	(0.551,0.955)	3.07	0.01	<0.001 *
Mx_Crowd	10	0.787	(0.548,0.933)	1.94	0.01	<0.001 *
Mn_Perim	10	0.522	(0.195,0.819)	4.66	0.32	<0.001 *
Mn_Crowd	10	0.655	(0.345,0.882)	3.66	0.15	<0.001 *
Bolton6	10	0.721	(0.476,0.906)	0.10	0.01	0.994
Bolton12	10	0.811	(0.620,0.939)	0.75	0.02	0.274

a. Cannot be computed because the sum of caseweights is less than or equal 1.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval;

*, p-value <0.05; Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width;

ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12,

Bolton millimeter, positive when Mandibular Excess

Table 3-7. Inter-rater, Extracted Premolars: ICC and repeated measures ANOVA mean differences shown for mesiodistal width measurements of each extracted premolar.

Parameter	N	Inter-rater Reliability		Mean Differences (mm)		
		ICC	95% CI	Worst	Best	p-value
<i>Extracted Premolars</i>						
Tooth 14	8	0.938	(0.845,0.985)	0.17	0.03	0.275
Tooth 15	3 ^a					
Tooth 24	9	0.913	(0.799,0.976)	0.15	0.02	0.372
Tooth 25	2 ^a					

a. Test values not reported due to low sample size.
Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval

3.3.4 Validity

A summary of the ICC and paired-sample *t* tests are presented for Anatomodels versus plaster in Table 3-8, Anatomodels vs extracted premolars in Table 3-9 and plaster vs extracted premolars in Table 3-10.

The validity of measurements on 30 Anatomodels compared with plaster (Table 3-8) was mostly poor to moderate in terms of agreement but with low mean differences. A number of parameters had ICC values below 0.6 and wide 95% confidence intervals including teeth 1-1, 1-3, 2-3, 2-5, 3-4, 3-5, 3-6, 4-5, and 4-6, maxillary arch perimeter, and Bolton anterior and Bolton overall measurements. There was, however, statistical evidence (p -value <0.05) to show that differences existed between Anatomodels and plaster for the mean measurements of teeth 1-1, 1-2, 1-3, 1-5, 2-1, 2-2, 2-3, 2-4, 2-5, mandibular intermolar width, maxillary and mandibular arch perimeter and crowding, and Bolton anterior and overall measurements; however, only

maxillary arch perimeter had a magnitude of mean difference, 3.38 mm and 95% CI (2.48, 4.28), that exceeded the clinically significant threshold.

Crosstabulations for Anatomodels vs. plaster are presented in Appendix 3-10. Out of a total possible two hundred and fifty two categorical comparisons, 92% were concordant pairs, suggesting excellent agreement across all categorical parameters. Nineteen out of two hundred and fifty two discordant pairs were observed across all ten nominal parameters except right molar and right canine classification, but no single parameter had more than four discordant pairs out of a possible thirty pairs.

Compared to extracted premolars (Table 3-9), Anatomodels had ICC values well above 0.9 and measurements on average up to 0.08 mm larger, while plaster (Table 3-10) had ICC values only slightly above 0.7 with measurements on average up to 0.17 mm smaller. All of the p-values were above 0.05. Again, analysis was only attempted for teeth 1-4 and 2-4 because the sample sizes for these teeth were not too small.

Table 3-8. Validity, Anatomodels vs Plaster: ICC and paired-sample mean differences shown for each parameter, grouped by linear measurements requiring 2 landmarks, and those requiring more than 2 landmarks.

Parameter	N	Agreement		Difference (mm) ^a		
		ICC	95% CI	Mean	95% CI	p-value
<i>Anatomodels vs. Plaster, Linear measurements, 2 landmarks</i>						
Overjet	18	0.927	(0.815,0.972)	0.02	(-0.31,0.35)	0.905
Overbite	18	0.925	(0.808,0.971)	0.27	(-0.17,0.70)	0.219
Tooth 1-1	30	0.558	(0.159,0.781)	0.35	(0.16,0.54)	0.001 *
Tooth 1-2	29	0.772	(0.552,0.889)	0.19	(0.02,0.37)	0.031 *
Tooth 1-3	30	0.532	(0.196,0.752)	0.29	(0.09,0.49)	0.007 *
Tooth 1-4	30	0.749	(0.540,0.872)	0.09	(-0.04,0.22)	0.166
Tooth 1-5	30	0.611	(0.319,0.796)	0.18	(0.02,0.33)	0.026 *
Tooth 1-6	30	0.724	(0.499,0.858)	0.07	(-0.12,0.27)	0.454
Tooth 2-1	30	0.630	(0.108,0.844)	0.47	(0.27,0.67)	<0.001 *
Tooth 2-2	30	0.863	(0.654,0.941)	0.21	(0.08,0.34)	0.003 *
Tooth 2-3	30	0.549	(0.214,0.763)	0.23	(0.07,0.39)	0.007 *
Tooth 2-4	30	0.773	(0.540,0.890)	0.16	(0.03,0.28)	0.014 *
Tooth 2-5	30	0.569	(0.271,0.768)	0.17	(0.00,0.34)	0.046 *
Tooth 2-6	30	0.714	(0.478,0.853)	-0.01	(-0.21,0.19)	0.908
Tooth 3-1	30	0.648	(0.385,0.815)	-0.08	(-0.23,0.07)	0.269
Tooth 3-2	30	0.727	(0.501,0.860)	-0.04	(-0.16,0.09)	0.559
Tooth 3-3	30	0.718	(0.485,0.855)	-0.02	(-0.18,0.14)	0.796
Tooth 3-4	30	0.595	(0.312,0.783)	-0.12	(-0.28,0.03)	0.120
Tooth 3-5	30	0.429	(0.099,0.677)	0.14	(-0.05,0.34)	0.143
Tooth 3-6	30	0.560	(0.251,0.764)	0.04	(-0.22,0.30)	0.772
Tooth 4-1	30	0.682	(0.437,0.834)	-0.08	(-0.20,0.03)	0.159
Tooth 4-2	30	0.704	(0.466,0.847)	-0.05	(-0.18,0.09)	0.500
Tooth 4-2a	1 ^b	-	-			
Tooth 4-3	30	0.612	(0.325,0.795)	0.03	(-0.14,0.20)	0.733
Tooth 4-4	30	0.660	(0.397,0.823)	0.03	(-0.14,0.21)	0.694
Tooth 4-5	30	0.367	(0.027,0.635)	0.14	(-0.05,0.33)	0.147
Tooth 4-6	30	0.518	(0.209,0.736)	0.23	(-0.00,0.47)	0.051
Mx_IMW	30	0.953	(0.900,0.978)	0.17	(-0.19,0.54)	0.339
Mx_ICW	30	0.873	(0.750,0.937)	0.14	(-0.42,0.71)	0.607
Mn_IMW	30	0.949	(0.885,0.976)	0.39	(0.05,0.73)	0.026 *
Mn_ICW	30	0.954	(0.907,0.978)	0.29	(-0.01,0.59)	0.061
<i>Anatomodels vs. Plaster, Linear measurements, >2 landmarks</i>						
Mx_Perim	30	0.536	(-0.092,0.824)	3.38	(2.48,4.28)	<0.001 *
Mx_Crowd	30	0.864	(0.710,0.936)	1.06	(0.18,1.94)	0.020 *
Mn_Perim	30	0.777	(0.371,0.909)	1.71	(0.88,2.54)	<0.001 *
Mn_Crowd	30	0.718	(0.205,0.888)	1.75	(1.00,2.49)	<0.001 *
Bolton6	30	0.506	(-0.097,0.807)	-1.57	(-1.99,-1.16)	<0.001 *
Bolton12	30	0.504	(0.006,0.770)	-1.95	(-2.73,-1.17)	<0.001 *

a. Positive mean difference when measurements from Anatomodel are larger

b. Cannot be computed because the sum of caseweights is less than or equal 1.

Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval; *, p-value <0.05; Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Intercanine Width; Perim, Arch Perimeter; Crowd, crowding if negative; Bolton6/Bolton12, Bolton millimeter, positive when Mandibular Excess

Table 3-9. Validity, Anatomodels vs Extracted Premolar: ICC and paired-sample mean differences for each premolar.

Parameter	N	Agreement		Difference (mm) ^a		
		ICC	95% CI	Mean	95% CI	p-value
<i>Anatomodels vs. Extracted Premolars</i>						
Tooth 14	8	0.963	(0.842,0.992)	0.08	(-0.07,0.22)	0.245
Tooth 15	3 ^b					
Tooth 24	9	0.957	(0.835,0.990)	0.05	(-0.09,0.20)	0.400
Tooth 25	2 ^b					

a. Positive mean difference when measurements from Anatomodels are larger
b. Test values not reported due to low sample size.
Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval

Table 3-10. Validity, Plaster vs Extracted Premolar: ICC and paired-sample mean differences for each premolar.

Parameter	N	Agreement		Difference (mm) ^a		
		ICC	95% CI	Mean	95% CI	p-value
<i>Plaster vs. Extracted Premolars</i>						
Tooth 14	8	0.731	(0.185,0.938)	-0.17	(-0.51,0.17)	0.286
Tooth 15	3 ^b					
Tooth 24	9	0.755	(0.243,0.939)	-0.08	(-0.36,0.20)	0.531
Tooth 25	2 ^b					

a. Positive mean difference when measurements from Plaster are larger
b. Test values not reported due to low sample size.
Abbreviations: ICC, Intraclass Correlation Coefficient; CI, Confidence Interval

3.3.5 Time

The average time, in minutes, required to perform measurements for all parameters in a SMA during the validity and reliability studies are presented in Table 3-11.

Within one evaluator over 10 subjects during the intra-rater reliability study (Table 3-11A), at worst, it took on average an additional 5.91 minutes longer than the best trial to perform a SMA on Anatomodels. The same comparison in plaster revealed on average only 2.34 additional minutes over the best trial. Between 5 evaluators over 10 subjects during the inter-

rater reliability study, at worst, the slowest evaluator may take 8.15 minutes and 2.33 minutes longer than the fastest evaluator for Anatomodels and plaster, respectively.

Across all 30 subjects during the validity study (Table 3-11B), the average time to perform all of the component measurements in the SMA was about 10 minutes using Anatomodels and 6 minutes using plaster. There was convincing evidence to show a statistical difference in the mean time to perform the same SMA in Anatomodels of 3.96 minutes and 95% CI (3.44, 4.48) longer than with plaster.

Table 3-11. Time required measuring all parameters in a study model analysis during: A. Reliability studies within one evaluator for Anatomodels and Plaster separately, and between five evaluators for Anatomodels and Plaster separately; B. Validity study comparing Anatomodels to Plaster.

A.

Modality	N	Time (minutes)		Mean Differences (minutes)		
		Mean	95% CI	Worst	Best	p-value
Within one evaluator						
Anatomodels	10	10.67	(10.35,11.00)	5.91	0.64	<0.001 *
Plaster	10	6.49	(6.19,6.78)	2.34	0.18	<0.001 *
Between five evaluators						
Anatomodels	10	12.52	(11.42,13.62)	8.15	0.08	<0.001 *
Plaster	10	6.22	(5.92,6.51)	2.33	0.09	<0.001 *

Abbreviations: CI, Confidence Interval; *, p-value <0.05

B.

Parameter	N	Mean Time (minutes)		Difference (minutes) ^a		
		Anatomodels	Plaster	Mean	95% CI	P-value
<i>Anatomodels vs. Plaster</i>						
Time	30	9.91	5.96	3.96	(3.44,4.48)	<0.001 *

a. Positive mean difference when time using Anatomodels was longer
Abbreviations: CI, Confidence Interval; *, p-value <0.05

3.4 Discussion

This study sought to investigate the performance of SMA using Anatomodels compared with plaster study models. A comprehensive analysis of validity, intra-rater reliability and inter-rater reliability using ten nominal (categorical) parameters, thirty scale (linear) 2-landmark parameters and six scale (linear) >2-landmark parameters over three modalities was performed.

In our methodology, we defined clinically relevant thresholds of 0.5 mm for 2-landmark linear measurements and 2.0 mm for >2-landmark linear measurements and then applied them to the interpretation of our results. In contrast, very few publications state a level of clinical significance as there is only mild support from the literature^{15,17,18}.

Notably, even though all thirty datasets were uploaded to Anatomage on the same day, only four cases were sent back 1.5 months later. After contacting the company, the majority of the cases were processed by the 2 month mark, four remaining cases by the 4 month mark, and the last case finally arrived 5 months after it was initially submitted. A representative from Anatomage cited busy production lines and miscommunications as reasons for the almost half-year delay. The unfortunate turnaround time is unacceptable in the context of providing timely diagnosis and treatment for patients.

3.4.1 Influence of the number of landmarks in a measurement on validity and reliability

The act of performing measurements has an element of uncertainty and is subject to error.²⁷ Uncertainty can be the result of random or systematic effects. Random effects can influence repeated measures in irregular ways whereas systematic effects influence the results in the same way for each repeated measurement. Error can arise from problems with the measuring instrument, instability of the item being measured, difficulties in the measurement process, improper calibration, lack of operator skill, sampling biases, and environmental factors.²⁷ In using an instrument to perform a single measurement, all of the aforementioned factors come into play, and by extension, summing multiple such measurements can multiply uncertainty and error.

In this study, we found that parameters utilizing only 2 landmarks (i.e. those involving a single measurement) had much lower and often clinically insignificant mean differences compared with parameters requiring more than 2 landmarks (i.e. those involving the calculation of multiple measurements). This finding is consistent with the findings from a systematic review of the literature on linear measurements using virtual study models (Chapter 2).

Interestingly, within the >2-landmark parameters for both the reliability and validity studies, the mean differences in arch crowding, Bolton anterior, and Bolton overall, which use upwards to twenty-four component measurements, were paradoxically better than arch perimeter

which use only four component measurements. When calculating multiple measurements, it is possible that errors from component measurements do not compound in the same direction such that the net effect is similar to a phenomenon known as regression towards the mean²⁸. Essentially, with greater number of components in a given calculation come greater opportunities for variation but also for errors to cancel each other out.

3.4.2 Reliability

The excellent intra-rater reliability for Anatomodels was due to the excellent agreement and clinically insignificant mean differences for linear parameters (Table 3-2), even for those that showed statistical significance. The ICC values were above 0.8 and the pairwise mean differences ranged from 0.00 to 0.56 mm across all 2-landmark parameters and 0.01 to 1.71 mm across all >2-landmark parameters. The moderate to excellent intra-rater reliability for plaster models was due to the excellent agreement and clinically insignificant mean differences for linear parameters (Table 3-3). The ICC values were above 0.7 and mean differences ranged from 0.00 to 0.21 mm across all 2-landmark parameters, and 0.01 to 1.33 mm across all >2-landmark parameters. The near perfect intra-rater reliability for extracted premolars was due to the excellent agreement and clinically insignificant mean differences for mesiodistal width measurements (Table 3-4). The ICC values were almost 1.0 and mean differences ranged from 0.00 to 0.02 mm. In summary, repeated measurements of linear parameters performed by a single evaluator using Anatomodels, plaster models or extracted teeth were consistent over multiple trials. Categorical parameters had excellent agreement in Anatomodels (Appendix 3-6)

as well as in plaster models (Appendix 3-7) owing to 96% concordance of measures. This suggests that repeated assessments of categorical parameters by a single evaluator using either Anatomodels or plaster models were also consistent over multiple trials.

For most linear parameters, inter-rater reliability using Anatomodels (Table 3-5), plaster models (Table 3-6), and extracted premolars (Table 3-7) had moderate to excellent agreement and clinically insignificant mean differences. This suggests that for most parameters, the mean measurements were consistent and acceptable among the individual trials of five evaluators.

Using Anatomodels (Table 3-5), the problematic parameters with low agreement and/or high mean differences were overbite, tooth 3-4, maxillary intermolar width, mandibular intercanine width, maxillary and mandibular arch perimeter, as well as maxillary and mandibular crowding. The parameters with mean differences that crossed the clinically significant thresholds, but also showed no statistically significant differences, while still exhibiting moderate agreement or better, were maxillary intercanine and mandibular intermolar widths. The discrepancy in maxillary intercanine width, for example, can be explained by measurements from one evaluator (NL) being different enough—but not so much that it could be deemed an outlier—from the remaining four evaluators (MM, CF, ML, TE) that it crossed the clinically relevant threshold of greater than 0.5 mm, but the poor agreement from the one evaluator (NL) was offset by the excellent agreement between the remaining four evaluators (MM, CF, ML, TE), giving an overall good agreement rating. Indeed, if the one evaluator (NL) is excluded from the analysis, the worst mean difference drops below 0.5 mm and agreement improves to excellent.

Using plaster (Table 3-6), the problematic parameters were overjet, overbite, tooth 2-3, maxillary and mandibular intercanine widths and perimeter, as well as mandibular crowding. Maxillary intermolar width had mean differences as high as 1.93 mm between two evaluators, but there was no statistical significance because the 95% confidence interval for the difference included zero (we cannot rule out the possibility that there is no difference). Suspect discordances affected maxillary shape in Anatomodels and maxillary shape, size and symmetry, as well as mandibular shape and size in plaster. In a practical sense, these parameters may still be useful for a single clinician, as evidenced by the high intra-rater reliability using either Anatomodels or plaster, but communication of these parameters between different clinicians may be meaningless.

Overall, intra-rater reliability was better than inter-rater reliability: mean differences were smaller, agreement and concordances were higher. Intra-rater reliability for all three modalities had no clinically significant findings. For extracted premolars, the worst mean differences were 0.02 mm and this happens to be the stated accuracy of the instrument used to perform those measurements. Interestingly, although intra-rater reliability is commonly reported in the literature on the linear accuracy of virtual models, it was not reported in the study¹² comparing Anatomodels to OrthoCAD while inter-rater reliability between two observers was merely reported as being adequate by paired t-tests.

Inter-rater reliability for all three modalities had a few parameters that were possibly clinically significant, particularly maxillary arch perimeter and maxillary crowding. It is possible that systematic error due to improper calibration could account for the lower and more variable agreement between five evaluators compared to a single evaluator. The same can be argued for the greater number of parameters with discordances as well as clinically significant mean differences. Moreover, discordances were observed more often in plaster despite the unfamiliar Anatomodels. It is possible that biases related to the extra experience using plaster led to variation between evaluators since this discrepancy was not observed in the intra-rater reliability study.

It should be noted that the discussion of worst mean differences should be kept in the context of the best mean differences which rarely exceeded 0.2 mm and never more than 1 mm for either intra-rater or inter-rater reliability. In other words, although clinically significant differences may be noted in the inter-rater reliability studies, the potential exists for practically no differences in reliability at all perhaps through more practice or training.

3.4.3 Validity

The experimental workflow started with studies of intra-rater reliability and then inter-rater reliability, and based on the encouraging results from the reliability studies, we finally went on to study validity. Ironically, based on the ICC values of the worst case parameters, agreement was worse in the validity studies compared with the reliability studies. If a bias due to fatigue

was present, this might be addressed by increasing the interval between measurements to greater than 10 days or reducing the number of models assessed per day.

There was questionable validity for linear parameters in Anatomodels compared with plaster (Table 3-8) due to mostly poor to moderate agreements despite a majority of low mean differences. In this study, twelve out of thirty-six linear parameters used in a full study model analysis had unacceptable validity based on either poor agreement or clinically significant mean differences. Taken altogether, the parameters in Anatomodels that had unacceptable validity compared with plaster, based on poor agreement or clinically significant mean differences, were: teeth 1-1, 1-3, 2-3, 2-5, 3-4, 3-5, 3-6, 4-5, 4-6, maxillary arch perimeter, Bolton anterior and Bolton overall measurements.

This is in contrast with Tarazona *et al.* (2011)¹³ who found that the mean measurements of teeth 1-4, 2-6, 3-4, 4-5, mandibular intercanine widths, and mandibular arch perimeter had statistical but not clinically significant differences of within 0.5 mm compared with 2D images of plaster models. Our finding is perhaps due to inconsistent landmark identification between conventional plaster and the unfamiliar digital counterpart in Anatomodels. Most of the nominal parameters had evidence of at least one pair of discordance and this might be explained by the lack of standardized definitions before the study began.

Similar to the findings by Kau *et al.* (2010)¹² who reported no statistical significance for the respective mean differences in overbite and overjet of 0.03 mm and -0.20 mm larger in

Anatomodels, this study found no evidence for a statistical difference in the mean overbite and overjet measurements of 0.27 mm and 0.02 mm larger in Anatomodels, respectively.

No discordances in the ten nominal parameters for all 30 subjects (Appendix 3-10) were cause for concern. Furthermore, there was higher validity of Anatomodels to extracted premolars (Table 3-9) than plaster models to extracted premolars (Table 3-10), but both comparisons showed moderate to excellent agreement and clinically insignificant mean differences. Validity with respect to the extracted premolars needs to be interpreted with caution due to the small sample size, but is arguable that agreement with extracted premolars using plaster was inferior to Anatomodels. If this is true, it is likely because the anatomical contact points are accessible in both the extracted premolars and Anatomodels, but not usually on plaster. Based on the trend of high agreement and low mean differences, it is probable that compared with the true gold standard of extracted premolars, measurements on both Anatomodels and plaster were valid.

3.4.4 Time

The time data in Table 3-11 suggested that the validity study had faster times than the intra-rater reliability study which showed faster times than the inter-rater reliability study. It is likely that five trials for Anatomodels and plaster in the intra-rater reliability study afforded more practice to become more efficient in the subsequent validity study. However, for the

evaluators who participated in the inter-rater reliability study, none had previous experience using Anatomodels, so the slower times to perform SMA was not surprising.

Comparing modalities, SMA using plaster had faster times than with Anatomodels by about four minutes. The extra four minutes to use Anatomodels may also be considered clinically significant to some clinicians. In plaster, the mean differences in time spent performing SMA were just slightly faster between different evaluators than within one person. Operator skill could have contribute to this finding since all of the evaluators were more experienced at performing measurements on the traditional modality of plaster than the principal investigator.

The length of time to use Anatomodels was prolonged by the tedious process of performing a single tooth width measurement with the software, as outlined in the Materials and Methods section. In plaster, the process is to simply place one tip of the digital caliper on the distal aspect, the other tip on the mesial aspect, make note of the measurement value, and then move on to the next tooth.

In the end, this study found that performing SMA on Anatomodels can take about four minutes longer than using plaster. This performance is similar to a study by Horton *et al.* (2009)¹⁶ who reported that mesiodistal measurements on virtual models took about three minutes longer than using plaster. In contrast, earlier studies reported an opposite trend of about one to three minutes faster using virtual models^{14,15} compared with plaster. Whereas our study showed mean times as fast as about 10 minutes using Anatomodels, Tarazona *et al.* (2011)¹³ were able

to perform most of their linear measurements on both arches in about 3 minutes. It is not likely that assessments of ten categorical parameters and a few more linear parameters in our study took up the extra 7 to 10 minutes. Our study performed the measurements by dynamically rotating the models, showing and hiding teeth to reveal the interproximal contact areas; it is not clear if Tarazona *et al.* (2011)¹³ approached the measurements in the same way.

A discussion of the extra time for SMA on Anatomodels should be considered in the context of the total time and costs involved compared with plaster. A thorough analysis on the resources, time, and related costs involved is beyond the scope of this paper but an unofficial comparison might be as presented in Table 3-12. The resources that need to be considered for traditional in-house records include both time and costs for panoramic and cephalometric radiographs, clinic chair time, lab time, sterilization, materials and overhead, and finally the time to perform SMA on plaster models. The comparable resources for CBCT-generated digital models, assuming they are out-sourced, involve practically no time from the practice but possibly only the related costs for the referral to the imaging center, which may or may not have the cost included for a radiologist report and for the Anatomodels, and then there is the time spent to perform SMA on Anatomodels.

Table 3-12. Comparison of estimated resources, time, and related costs for Plaster in-house versus Anatomodels via CBCT scan in-house or outsourced.

Resource	Estimated time	Related costs*
<i>Traditional records with impressions in-house</i>		
Radiographs: Pan + Ceph	10 min	\$ 170.00
Clinic	30 min	240.00
Lab	60 min	100.00
Sterilization	50 min	30.00
Study model analysis on Plaster	11 min	5.00
	2 hr 41 min	\$ 546.00
<i>CBCT scan in-house, models outsourced</i>		
Radiographs: CBCT imaging in-house	10 min	\$ 200.00
Radiologist report: screening for pathology	Outsourced	125.00
Anatomodels	Outsourced	70.00
Study model analysis on Anatomodels	11 min	6.00
	21 min	\$ 401.00
<i>CBCT scan and models outsourced</i>		
CBCT imaging center referral	Outsourced	\$ 450.00
Radiologist report (may be included)	Outsourced	-
Anatomodels (may be included)	Outsourced	-
Study model analysis on Anatomodels	11 min	6.00
	11 min	\$ 456.00

* Estimated Assistant/Technician salary, Materials and Overhead costs

3.4.5 Limitations and possible sources of error

Although Anatomodels may appear more cost effective overall, it should be noted that CBCT-generated study models alone should not justify taking the imaging. However, if the imaging is being acquired for other diagnostic purposes, then secondary model analysis does not expose the patient to any additional radiation.

Since Anatomodels are produced via a proprietary process, there is an underlying assumption that when teeth are segmented from CBCT scans, it is done correctly along true anatomic

contours. Any differences arising from this segmentation process, then, will contribute to systematic error. The process of segmentation to define the boundary of an object involves complicated schemes that apply transfer functions to take into account the values of neighboring voxels.²⁹ Such processes can select the voxels that lie on the surface between two different materials, such as tooth and bone. For adjacent teeth, however, segmenting the contact area that consists of two of the same materials, enamel and enamel, likely requires volume cut outs, slicing or peeling. The error of this process as it relates to segmenting human teeth has not been fully studied.

One should be cautioned not to attempt to extend the findings from this study to other potential measurements with Anatomodels not investigated, such as root length or angular measurements. The measurements in this study involved landmarks on enamel only. Further studies are required for measurements that rely on the segmentation of other materials in the body that have different expected densities, such as dentin, bone, cartilage, and soft tissues.

In the absence of complicating factors³⁰ such as partial volume average, noise, artifacts and threshold settings, it is theoretically possible to define a single point by selecting only one voxel. Additional voxels may help to identify the single voxel of interest but they are not necessary in the act of selecting a single voxel. When defining the true boundary of an object, at best, the line for this boundary will cross directly through the center of a voxel. But when attempting to select a boundary that truly goes between voxels, one is forced to select the center of one of the surrounding voxels. At worst, then, the accuracy for the selection of a single voxel of 0.3

mm sides will be unavoidably off by the equivalent of half the diagonal of the voxel, or 0.26 mm. Given that a 2-landmark measurement will require the selection of two voxels, then, we would expect errors in accuracy to be as much as two half-diagonal distances, or around 0.5 mm, as illustrated in Figure 3-5.

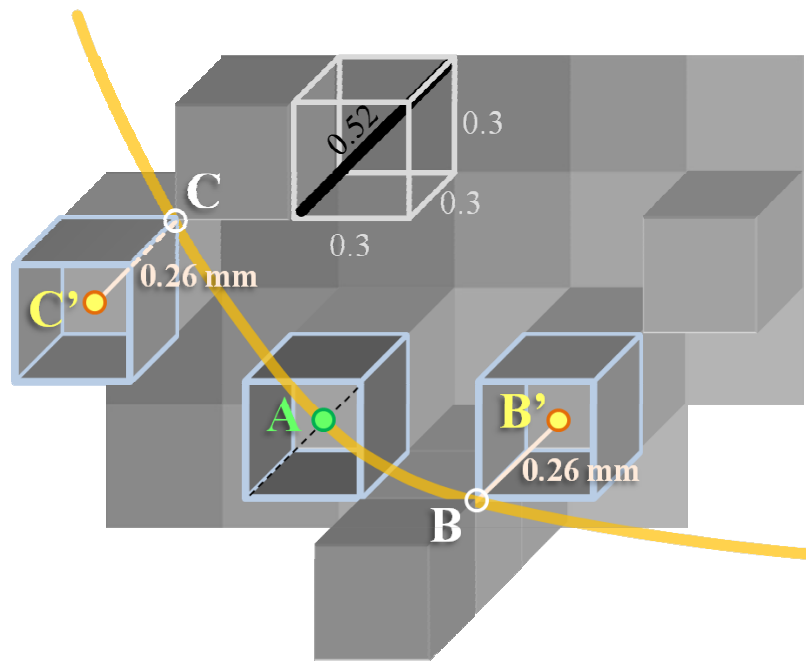


Figure 3-5. The accuracy of selecting voxels, outlined in blue, for the boundary of an object which follows a path (orange line) through points A, B and C. Selecting point A (green circle) is perfectly accurate since the orange line goes through the center of the voxel. But, in attempting to select points B and C, we are forced to select a neighboring voxel which centers at point B' and C' (yellow circles), respectively. Since the diagonal of a voxel with 0.3 mm sides is 0.52 mm, Point B' has as much as 0.26 mm error from the true Point B. Taking into account the error for point C', one can note that the accuracy of selecting two voxels can have a total error of much as about 0.5 mm.

The accuracy of different segmentation protocols on CBCT-acquired surface models of mandibles in cadaver heads with intact soft tissue was recently reported by Fourie *et al* (2011)³¹. In their study, CBCT scans of seven fresh-frozen cadaver heads using a KaVo 3D machine (KaVo Dental GmbH, Bismarckring, Germany) were segmented for the mandibles

commercially by an experienced technician and in-house by a clinician in oral maxillofacial surgery, and then compared with subsequently macerated mandibles acquired by laser scanning. Compared with the laser-acquired mandibles, the mean differences in linear measurements were 0.33 ± 0.43 mm larger by commercial segmentation, which was more accurate than 0.76 ± 0.39 mm larger by doctor segmentation.

The tendency toward positive mean differences of Anatomodels compared with extracted premolars suggests that conservative segmentation of voxel datasets in CBCT-generated virtual models occurred resulting in larger than expected measurements on Anatomodels. On the other hand, the negative differences in average mesiodistal measurements of plaster compared with extracted premolars suggest dimensional changes in plaster such that measurements were systematically smaller than in reality. This may be the result of imbibition of water³² causing the alginate impression material to expand, thus resulting in a slightly smaller than expected stone cast. Again, these statements should be interpreted with caution due to the relatively small sample of extracted premolars.

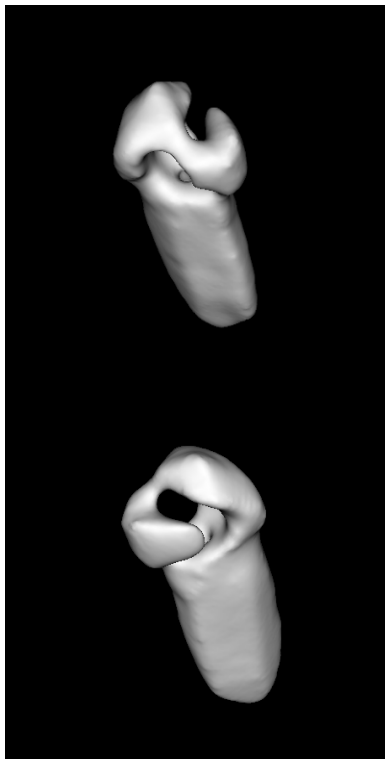
A few Anatomodels had defects due to possible patient movement or streak artifacts.

Radiographically, dental fillings are strongly attenuating objects which cause metal streak artifacts that are seen in reconstructed images as dark streaks in the direction of highest attenuation³³ (Figure 3-6A). The result on Anatomodels is an incomplete reconstruction of a segmented tooth and such missing surfaces will challenge the veracity of measurements.

Oftentimes, though, there is still enough reconstructed tooth structure to make an approximate measurement.

Sporadically, parts of the tooth would disappear upon manipulation onscreen and this behavior can be reproduced on other computers. Since the measurement points are placed only on the volume that is visible, sometimes they end up being placed inside the pulp or on the inner surface of the tooth (Figure 3-6A). “Shaking” the tooth usually causes the surfaces to reappear, but it is an annoying occurrence nonetheless.

A.



B.

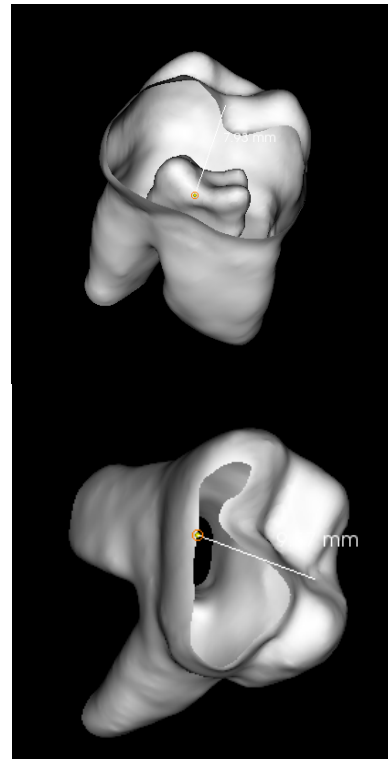


Figure 3-6. Artifacts in Anatomodels: A, due to metal streak artifacts; B, demonstrating disappearing surfaces.

It has been recognized that standardized definitions for the intended parameters on study models can improve reliability and validity.³⁴ Although much care was taken to train all evaluators, it is possible that the personal biases due to what the evaluators are normally accustomed to diagnosing occasionally may have overrode attempts to utilize common landmarks for this study. Couple this problem with an unfamiliar and sometimes finicky modality and it comes as little surprise that agreement between Anatomodels and plaster was poor to moderate for many parameters. Fortunately, the mean differences in measurements were generally clinically insignificant.

The InVivo5 software has room for improvement. The numerous steps to gather data for a full SMA using Anatomodels certainly is a hindrance and perhaps keyboard shortcuts could speed up the process. The segmented teeth could be identified by Anatomage according to Fédération Dentaire Internationale (FDI) notation³⁵, Universal number system³⁶, or Zsigmondy/Palmer notations^{37,38} to permit easier identification from the list of available objects in the Anatomodel. It would be worthwhile to have the ability to store, show and hide measurements presented in a table format in future builds of the software. When placing points for a measurement, there was no undo option, and when multiple points were placed within close proximity of one another, an option to open a sub-menu to select a specific point from a list would be helpful. It would be useful to have cross-sectional views from the models, not the radiographic slices, to easily measure overbite and overjet as well as evaluate interdigitation.

3.4.6 Transfer of knowledge to clinical practice

A summary of the interpretations from this study are presented in Table 3-13 and Table 3-14.

Table 3-13. Caution is advised for the identified 2-landmark Linear parameters due to agreement that was poor (ICC<0.600), or mean differences that were both statistically significant (p-value<0.05) and clinically large (>0.5 mm).

Variables Investigated	Intrarater Reliability		Interrater Reliability		Validity
	Anatomodels	Plaster	Anatomodels	Plaster	Anatomodels vs Plaster
<i>Linear Measurements, 2 landmarks</i>					
Overjet				Overjet	
Overbite	Overbite		Overbite	Overbite	
Tooth 1-1					Tooth 1-1
Tooth 1-2					
Tooth 1-3					Tooth 1-3
Tooth 1-4					
Tooth 1-5					
Tooth 1-6					
Tooth 2-1					
Tooth 2-2					
Tooth 2-3				Tooth 2-3	Tooth 2-3
Tooth 2-4					
Tooth 2-5					Tooth 2-5
Tooth 2-6					
Tooth 3-1					
Tooth 3-2					
Tooth 3-3					
Tooth 3-4			Tooth 3-4		Tooth 3-4
Tooth 3-5					Tooth 3-5
Tooth 3-6					Tooth 3-6
Tooth 4-1					
Tooth 4-2					
Tooth 4-2a					
Tooth 4-3					
Tooth 4-4					
Tooth 4-5					Tooth 4-5
Tooth 4-6					Tooth 4-6
Mx_IMW	Mx_IMW		Mx_IMW		
Mx_ICW				Mx_ICW	
Mn_IMW					
Mn_ICW			Mn_ICW	Mn_ICW	

Abbreviations: Mx_, Maxillary; Mn_, Mandibular; IMW, Intermolar Width; ICW, Inter canine Width

Table 3-14. Caution is advised for the identified >2-landmark linear parameters due to poor agreement (ICC<0.600), or mean differences that were both statistically significant (p-value<0.05) and clinically large (>2.0mm), and for the identified categorical parameters due to potentially high discordances. A summary from the time studies is also provided.

Variables Investigated	Intrarater Reliability		Interrater Reliability		Validity
	Anatomodels	Plaster	Anatomodels	Plaster	Anatomodels vs Plaster
<i>Linear Measurements, >2 landmarks</i>					
Mx_Perim			Mx_Perim	Mx_Perim	Mx_Perim
Mx_Crowd			Mx_Crowd		
Mn_Perim			Mn_Perim	Mn_Perim	
Mn_Crowd			Mn_Crowd	Mn_Crowd	
Bolton6					Bolton6
Bolton12					Bolton12
<i>Categorical parameters</i>					
R_Molar					
R_Canine					
L_Molar					
L_Canine					
Mx_Shape			Mx_Shape	Mx_Shape	
Mx_Size				Mx_Size	
Mx_Symm				Mx_Symm	
Mn_Shape				Mn_Shape	
Mn_Size				Mn_Size	
Mn_Symm					
<i>Time</i>					
Mean (minutes)	10.67	6.49	12.52	6.22	9.91, 5.96
Mean Difference	5.91	2.34	8.15	2.33	3.96

Abbreviations: Mx_, Maxillary; Mn_, Mandibular; Perim, Arch Perimeter; Crowd, Crowding; Bolton6/Bolton12, Bolton millimeter; R_, Right; L_, Left; Symm, Symmetry

Linear and categorical measurements on Anatomodels were just as reliable as on plaster when performed by one clinician.

Between different clinicians, most linear and categorical measurements were reliable. But due to poor agreement and/or unacceptable mean differences, a number of parameters may have different relevance when communicated between different clinicians. In Anatomodels, the

parameters that should be communicated with caution between clinicians were overbite, tooth 3-4, maxillary intermolar width, mandibular intercanine width, maxillary and mandibular arch perimeter, maxillary and mandibular crowding, as well as maxillary shape; using plaster, the problematic parameters were overjet, overbite, tooth 2-3, maxillary and mandibular intercanine widths and perimeter, mandibular crowding, maxillary shape, size and symmetry, as well as mandibular shape and size.

The validity of Anatomodels compared with plaster was unacceptably poor to moderate for many parameters, including teeth 1-1, 1-3, 2-3, 2-5, 3-4, 3-5, 3-6, 4-5, 4-6, maxillary arch perimeter, Bolton anterior and Bolton overall measurements. The validity of Anatomodels or plaster for the mesiodistal measurements of teeth 1-4 and 2-4 were high.

Finally, performing a study model analysis on Anatomodels will take, on average, four minutes longer than on plaster.

The turnaround time to process Anatomodels can be unpredictable and improvements to the interface could be made to allow more efficient measuring before this could be considered a viable option in the workflow of a typical private practice.

Future research will be required to elucidate if other potential uses of CBCT-generated study models, such as tooth set ups and the possibility to fabricate intraoral appliances, can be utilized in practice.

3.5 Conclusion

In performing study model analysis for all three modalities investigated—Anatomodels, plaster and extracted premolars— the following conclusions could be made:

1. Intra-rater reliability was excellent.
 - a. Repeated measurements of 2-landmark and >2-landmark parameters as performed by a single evaluator using Anatomodels, plaster models, or extracted premolars, were consistent over multiple repeated trials owing to the moderate to excellent agreement and clinically insignificant mean differences.
 - b. There was excellent agreement of nominal (categorical) parameters.
2. Inter-rater reliability was moderate to excellent for most parameters.
 - a. Measurements of most linear parameters using Anatomodels, plaster, or extracted premolars were consistent and acceptable between the individual trials of five evaluators owing to moderate to good agreement and clinically insignificant mean differences.
 - b. Suspect discordances affected only one out of ten nominal parameters in Anatomodels, whereas discordances in plaster affected half of the nominal parameters.

3. Validity was poor to moderate for many parameters.
 - a. Between Anatomodels and corresponding plaster study models, unacceptable differences in agreement or mean measurements could be demonstrated for 13 out of 36 linear parameters.
 - b. Compared with extracted premolars, the validity of mesiodistal measurements on Anatomodels or plaster were high.
 - c. High concordances for the ten nominal parameters were present.
4. Time spent on Anatomodels can be almost twice as long as that on plaster.
 - a. Across all subjects, study model analyses on Anatomodels took, on average, 10 minutes while plaster took 6 minutes.

3.6 Sources of Funding

The author would like to thank and acknowledge The Fund for Dentistry at the University of Alberta and the Alpha Omega Foundation for their grants in support of this research initiative.

3.7 References

1. Kahl-Nieke B, Fischbach H, Schwarze CW. Treatment and postretention changes in dental arch width dimensions--a long-term evaluation of influencing cofactors. *Am J Orthod Dentofacial Orthop* 1996;109:368-378.
2. 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 e11-14; discussion 16.
3. 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.
4. 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.
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. Macchi A, Carrafiello G, Cacciafesta V, Norcini A. Three-dimensional digital modeling and setup. *American Journal of Orthodontics and Dentofacial Orthopedics* 2006;129:605-610.
7. Scarfe WC, Farman AG, Sukovic P. Clinical applications of cone-beam computed tomography in dental practice. *J Can Dent Assoc* 2006;72:75-80.
8. Lagravere MO, Carey J, Toogood RW, Major PW. Three-dimensional accuracy of measurements made with software on cone-beam computed tomography images. *Am J Orthod Dentofacial Orthop* 2008;134:112-116.

9. Periago DR, Scarfe WC, Moshiri M, Scheetz JP, Silveira AM, Farman AG. Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. *Angle Orthod* 2008;78:387-395.
10. Ganguly R, Ruprecht A, Vincent S, Hellstein J, Timmons S, Qian F. Accuracy of linear measurement in the Galileos cone beam computed tomography under simulated clinical conditions. *Dento-Maxillo-Facial Radiology* 2011;40:299-305.
11. Rosenblatt MR. Tooth Length Measurement Accuracy and Reliability with Cone-Beam CT and Panoramic Radiography Orthodontics. Edmonton, AB: University of Alberta; 2010.
12. Kau CH, Littlefield J, Rainy N, Nguyen JT, Creed B. Evaluation of CBCT digital models and traditional models using the Little's Index. *Angle Orthod* 2010;80:435-439.
13. Tarazona B, Llamas JM, Cibrian R, Gandia JL, Paredes V. A comparison between dental measurements taken from CBCT models and those taken from a Digital Method. *European Journal of Orthodontics* 2011.
14. Tomassetti JJ, Taloumis LJ, Denny JM, Fischer JR, Jr. A comparison of 3 computerized Bolton tooth-size analyses with a commonly used method. *Angle Orthodontist* 2001;71:351-357.
15. Mullen SR, Martin CA, Ngan P, Gladwin M. Accuracy of space analysis with emodels and plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2007;132:346-352.
16. Horton HM, Miller JR, Gaillard PR, Larson BE. Technique comparison for efficient orthodontic tooth measurements using digital models. *Angle Orthod* 2009;80:254-261.
17. Asquith J, Gillgrass T, Mossey P. Three-dimensional imaging of orthodontic models: a pilot study. *European Journal of Orthodontics* 2007;29:517-522.

18. Goonewardene RW, Goonewardene MS, Razza JM, Murray K. Accuracy and validity of space analysis and irregularity index measurements using digital models. *Australian Orthodontic Journal* 2008;24:83-90.
19. Rosner B. *Fundamentals of biostatistics*. Boston: Brooks/Cole, Cengage Learning; 2010.
20. Cha BK, Choi JI, Jost-Brinkmann PG, Jeong YM. Applications of three-dimensionally scanned models in orthodontics. *International Journal of Computerized Dentistry* 2007;10:41-52.
21. El-Zanaty HM, El-Beialy AR, Abou El-Ezz AM, Attia KH, El-Bialy AR, Mostafa YA. Three-dimensional dental measurements: An alternative to plaster models. *American Journal of Orthodontics & Dentofacial Orthopedics* 2010;137:259-265.
22. Keating AP, Knox J, Bibb R, Zhurov AI. A comparison of plaster, digital and reconstructed study model accuracy. *Journal of Orthodontics* 2008;35:191-201; discussion 175.
23. Miras D, Sander FG. [The accuracy of holograms compared to other model measurements]. *Fortschritte der Kieferorthopadie* 1993;54:203-217.
24. Naidu D, Scott J, Ong D, Ho CT. Validity, reliability and reproducibility of three methods used to measure tooth widths for bolton analyses. *Australian Orthodontic Journal* 2009;25:97-103.
25. 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. *American Journal of Orthodontics & Dentofacial Orthopedics* 2006;129:794-803.
26. 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.

27. Bell S. A Beginner's Guide to Uncertainty of Measurement. In: Laboratory NP, editor. Teddington, Middlesex: Crown; 1999.
28. Stigler SM. Regression towards the mean, historically considered. *Stat Methods Med Res* 1997;6:103-114.
29. Halazonetis DJ. From 2-dimensional cephalograms to 3-dimensional computed tomography scans. *Am J Orthod Dentofacial Orthop* 2005;127:627-637.
30. Molen AD. Considerations in the use of cone-beam computed tomography for buccal bone measurements. *Am J Orthod Dentofacial Orthop* 2010;137:S130-135.
31. Fourie Z, Damstra J, Schepers RH, Gerrits PO, Ren Y. Segmentation process significantly influences the accuracy of 3D surface models derived from cone beam computed tomography. *Eur J Radiol* 2011.
32. Nicholls JI. The measurement of distortion: theoretical considerations. *Journal of Prosthetic Dentistry* 1977;37:578-586.
33. De Man B, Nuyts J, Dupont P, Marchal G, Suetens P. Metal Streak Artifacts in X-ray Computed Tomography: A Simulation Study. *IEEE Trans Nucl Sci* 1999;46:691-696.
34. Pair JW, Luke L, White S, Atchinson K, Englehart R, Brennan R. Variability of study cast assessment among orthodontists. *American Journal of Orthodontics and Dentofacial Orthopedics* 2001;120:629-638.
35. Keiser-Nielsen S. Federation Dentaire Internationale. Two-Digit System of designating teeth. *Int Dent J* 1971;21:104-106.

36. Parreidt J. Zählung der Zähne und Benennung der verschiedenen Zahnsorten. In: Felix A, editor. Zahnärztliche Mitteilungen aus der chirurgischen Universitätspoliklinik zu Leipzig.

Germany; 1882. p. 10-15.

37. Zsigmondy A. Grundzüge einer praktischen Methode zur raschen und genauen Vormerkung der zahnärztlichen Beobachtungen und Operationen. Dtsch Vjschr Zahnk 1861;1:209-211.

38. Palmer C. Proceedings of the tenth annual meeting of the American Dental Association.

Nashville, TN: Dent Cosmos; 1870: p. 522-523.

Chapter 4. General Discussion and Conclusions

4.1 Synthesis

The act of obtaining diagnostic measurements from dental study models is a process referred to as study model analysis (SMA). Traditionally, SMA is performed using calipers on dental study models made of plaster stone but technological advances have made possible the virtualization of this process. Further advances have created opportunities to extract virtual study models from volumetric radiographic scans via Cone Beam Computed Tomography (CBCT) of the oral region, in addition to providing the anatomical information necessary for adequate diagnosis and treatment planning.

The first objective of this thesis (Chapter 2) was to perform a systematic review of the literature on virtual study models. From this review, commonly reported linear parameters were identified for inclusion in a full SMA, and a distinction was made between 2-landmark and >2-landmark linear parameters. Though not commonly reported, we acknowledged that a full SMA should also include qualitative parameters, and investigations should include tests of inter-rater and intra-rater reliability, validity, as well as comparisons of the time used to perform such measurements.

The second part of the thesis (Chapter 3) tested the reliability, validity and time efficiency of performing a full SMA on CBCT-generated virtual models (Anatomodels) compared with plaster

study models. The investigation had an inimitable opportunity to further confirm measurements with the *true* gold standard, extracted premolars for a subset of the sample. We were also fortunate to have multiple clinicians—well-trained in the use of traditional plaster study models but minimally trained in the use of Anatomodels—participate in the inter-rater reliability study to substantiate the findings from the principal investigator.

4.2 Limitations

Despite the encouraging findings from the systematic review (Chapter 2), there was paucity of scientific information on qualitative diagnostic parameters which is an essential component of a full SMA. Although intra-rater reliability was often reported, inter-rater reliability was not. The time required to perform a full SMA, whether in plaster or virtual study models, was almost never reported. At best, most virtual study models in use today, from a visible tooth structure point of view, reveal only half the relevant information that a CBCT-generated study can potentially provide, such as the relationship of the roots and other anatomic structures. The linear accuracy of CBCT-generated study models has been investigated for a few parameters but other important parameters, in addition to the time efficiency of performing such measurements, remain to be reported.

In conducting the study on Anatomodels (Chapter 3), a number of limitations were already discussed. Graphical artifacts and software interface issues hindered the performance of a full SMA. There was an unacceptable—almost half-year—turnaround time to have Anatomodels

processed by Anatomage. By far the greatest limitation was the poor to moderate validity for many parameters on Anatomodels compared with plaster. These limitations need to be considered before Anatomodels can be deemed a viable option in the workflow of a typical private orthodontic practice.

4.3 Findings and Conclusions

In the systematic review of the literature on virtual models (Chapter 2), four types of acquisition were identified: holographic scanning, stereophotogrammetry capture, laser scanning or Cone Beam Computed Tomography (CBCT) scanning. Only articles on laser-acquired and CBCT -acquired were ultimately included. Neither reliability nor validity had perceivable influences by acquisition type. >2-landmark measures tended to have higher mean differences than 2-landmark measures. Agreement, as measured by ICC, was just as high with less variability than correlation, as measured by PCC. Overall, it could be said that the validity and reliability of linear measurements performed on virtual dental study models were high and comparable to similar measurements on plaster. Therefore, virtual study models could be considered clinically acceptable compared with plaster study models in regards to the intra-rater reliability and validity of linear measurements.

Based on the findings of this study (Chapter 3), linear and categorical measurements on Anatomodels were just as reliable as on plaster when performed by one clinician. Between different clinicians, most linear and categorical measurements were reliable. But due to poor

agreement and/or unacceptable mean differences, a number of parameters may have different relevance when communicated between different clinicians. In Anatomodels, the parameters that should be communicated with caution between clinicians were overbite, tooth 3-4, maxillary intermolar width, mandibular intercanine width, maxillary and mandibular arch perimeter, maxillary and mandibular crowding, as well as maxillary shape; using plaster, the problematic parameters were overjet, overbite, tooth 2-3, maxillary and mandibular intercanine widths and perimeter, mandibular crowding, maxillary shape, size and symmetry, as well as mandibular shape and size. The validity of Anatomodels compared with plaster was unacceptably poor to moderate for many parameters, including teeth 1-1, 1-3, 2-3, 2-5, 3-4, 3-5, 3-6, 4-5, 4-6, maxillary arch perimeter, Bolton anterior and Bolton overall measurements. The validity of Anatomodels or plaster for the mesiodistal measurements of teeth 1-4 and 2-4 were high. Finally, performing a study model analysis on Anatomodels will take, on average, four minutes longer than on plaster.

4.4 **Future Research**

Much opportunity exists to extend the findings from this thesis to future research. It is possible that alternative software exist to perform SMA on CBCT-generated study models. For example, the SureSmile system (OraMetrix, Richardson, TX) also has a proprietary process in which teeth are individually segmented from CBCT scans with certain landmarks placed by their technicians and subsequent linear measurements already provided. Qualitative measurements, however, are not provided, so a full SMA is still not possible without some extra work. It should be noted

that the SureSmile system is meant to be used as a finishing technique, so CBCT scans are generally taken during orthodontic treatment, not necessarily before treatment.

One of the unique software features of InVivo5 (Anatomage, San Jose, CA) not investigated in this study is the ability to freely move and rotate teeth onscreen and therefore simulate and preview tooth set ups. The focus of a future study could be on tooth set ups in Anatomodels. At the moment, however, segmented teeth can be approximated so close that they will overlap without warning. An option to prevent such “collisions” would be useful. In cases of missing teeth, it would be practical to have tooth templates with customizable buccolingual and mesiodistal dimensions.

The occlusal anatomy in Anatomodels lacked the level of detail that plaster can provide. Furthermore, it is unclear if the curvatures of the buccal and lingual surfaces are accurate enough to allow fabrication of intraoral appliances. Further research is needed to determine if well-fitting lab appliances can be fabricated from Anatomodels.

Depending on how the dental assistant was trained at the University clinic, some patients had CBCT scans that allowed reconstruction at 0.30x0.30x0.30 voxels while other scans offered higher resolution at 0.25x0.25x0.25 mm voxels. The influence of the voxel size on validity and reliability was not investigated.

Finally, one should not attempt to extend the findings from this study to other potential measurements with Anatomodels not investigated, such as root length or angular measurements. The measurements in this study involved landmarks on enamel only. Further studies are required for measurements that rely on the segmentation of other materials in the body that have different expected densities, such as dentin, bone, cartilage, and soft tissues.

Appendices

Appendix 2-1. Search strategy and related search terms.

PICO	Patient	Intervention	Comparison	Outcome
Question	For dental patients...	...are linear measurements...	... using digital study models compared with conventional plaster study modelsvalid and reliable?
Search terms	orthod* dental dentistry	caliper calliper measur* assess*	"study models" "study casts" "plaster models" "plaster casts" "digital models" "virtual models" orthocad emodel geodigm digimodel orthoproof	reproduc* reliab* valid* accur*

Appendix 2-2. Summary of results from electronic databases, as of May 16, 2010, after adapted search strings were applied.

Database	Search string	Results
Pubmed	(orthod* OR dental OR dentistry) AND (caliper OR calliper OR measur* OR assess*) AND ("study models" OR "study casts" OR "plaster models" OR "plaster casts" OR "digital models" OR "virtual models" OR orthocad OR emodel OR geodigm OR digimodel OR orthoproof) AND (reproduc* OR reliab* OR valid* OR accur*)	260
OVID Medline	((orthod* or dental or dentistry) and (caliper or calliper or measur* or assess*) and ("study models" or "study casts" or "plaster models" or "plaster casts" or "digital models" or "virtual models" or orthocad or emodel or geodigm or digimodel or orthoproof) and (reproduc* or reliab* or valid* or accur*)).af.	258
OVID All EBMR	((orthod* or dental or dentistry) and (caliper or calliper or measur* or assess*) and ("study models" or "study casts" or "plaster models" or "plaster casts" or "digital models" or "virtual models" or orthocad or emodel or geodigm or digimodel or orthoproof) and (reproduc* or reliab* or valid* or accur*)).af.	16
LILACS	(orthod? OR dental OR dentistry) AND (caliper OR calliper OR measure? OR assess?) AND ("study models" OR "study casts" OR "plaster models" OR "plaster casts" OR "digital models" OR "virtual models" OR orthocad OR emodel OR geodigm OR digimodel OR orthoproof) AND (reproduce? OR reliab? OR valid? OR accur?)	0

Appendix 2-3. Sample calculation using A) pooled data from systematic review, for B) weighted mean difference for the parameter OB in laser-acquired models.

A.

Study	Year	Title	Company	Acquisition	n	Parameter	Validity		
							Value	Units	Type
Watanabe-Kanno GA,	2009	Reproducibility, reliability and	Bibliocast	CBCT	15	OB	-0.21	mm	Mean difference
Bootvong K, Liu Z,	2010	Virtual model analysis as an	Cadent	Laser	80	OB	0.16	mm	Mean difference
Quimby ML, Vig KWL,	2004	The accuracy and reliability of	Cadent	Laser	50	OB	-0.66	mm	Mean difference
Santoro M, Galkin S,	2003	Comparison of measurements made	Cadent	Laser	20	OB	-0.4901	mm	Mean difference
Sjogren AP, Lindgren JE,	2009	Orthodontic Study Cast Analysis-	Ortolab	Laser	20	OB	0	mm	Mean difference
Stevens DR, Flores-Mir C,	2006	Validity, reliability, and	GeoDigm	Laser	24	OB	-0.3	mm	Mean difference

B.

Weighted mean for Laser-acquired parameter “OB”:

$$= \frac{80(0.16) + 50(-0.66) + 20(-0.49) + 20(0) + 24(-0.3)}{80 + 50 + 20 + 20 + 24}$$

$$= \frac{-37.2}{194}$$

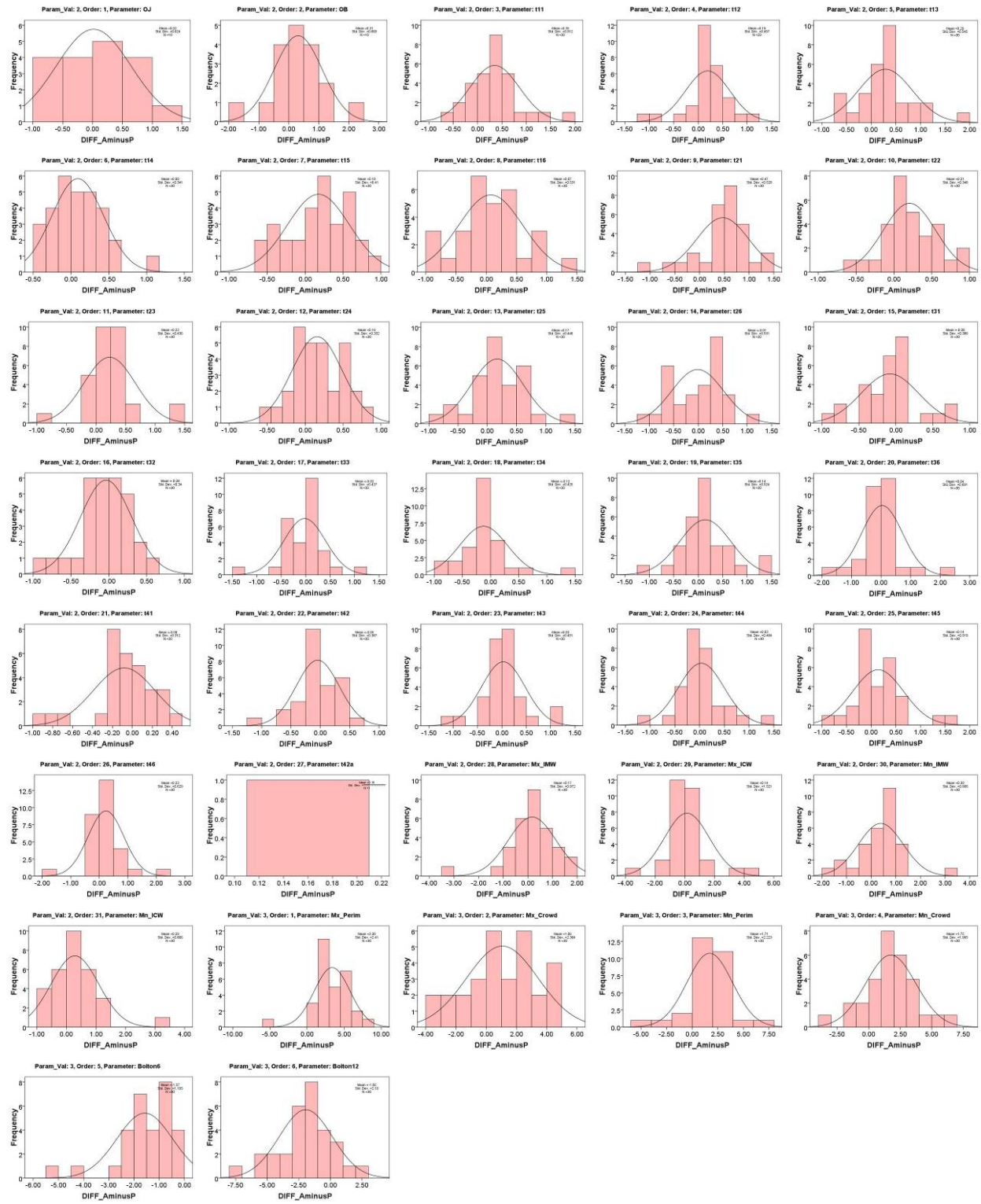
$$= -0.19 \text{ mm}$$

Appendix 2-4. Raw data from selected articles of the systematic review.

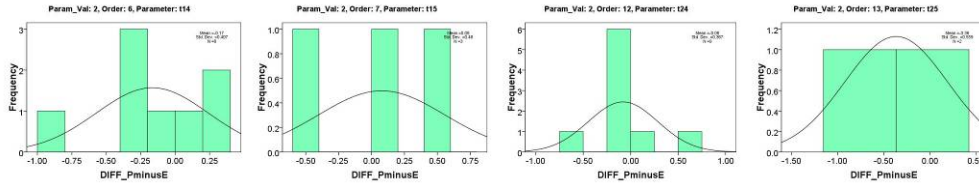
Study	Year	Title	Company	Acquisition type	n	Parameter	Reliability				Validity													
							Plaster models		3D model		Correlation Coefficient (R)		Mean difference (mm)											
							Absolute	Units	Type	Absolute	Units	Type	Value	Units	Value	Units								
Aican T, Ceylanoglu C, Baysal B	2009	The relationship between digital model accuracy and time-dependent deformation of alginate impressions	3Shape	Laser	21	Mx IMW	0.773	ICC	Inter-rater	0.74	ICC	Inter-rater	0.786	ICC	0.381	mm	Absolute difference							
							0.721	ICC	Inter-rater	0.786	ICC	Inter-rater	0.774	ICC	0.123	mm	Absolute difference							
							0.82	ICC	Inter-rater	0.847	ICC	Inter-rater	0.75	ICC	0.097	mm	Absolute difference							
							0.815	ICC	Inter-rater	0.752	ICC	Inter-rater	0.744	ICC	0.055	mm	Absolute difference							
Asquith J, Gillgrass T, Mossey P	2007	Three-dimensional imaging of orthodontic models: a pilot study	Arius3D	Laser	10	11 MDW	0.11	mm	Intra-rater	0.27	mm	Intra-rater			0.16	mm	Mean difference							
							0.01	mm	Intra-rater	0.15	mm	Intra-rater			-0.19	mm	Mean difference							
							0.01	mm	Intra-rater	0.11	mm	Intra-rater			-0.38	mm	Mean difference							
							0.12	mm	Intra-rater	0.08	mm	Intra-rater			0.1	mm	Mean difference							
							0.12	mm	Intra-rater	0.11	mm	Intra-rater			-0.11	mm	Mean difference							
							0.09	mm	Intra-rater	0.02	mm	Intra-rater			-0.05	mm	Mean difference							
							0.06	mm	Intra-rater	0.59	mm	Intra-rater			-0.62	mm	Mean difference							
							0.11	mm	Intra-rater	1.4	mm	Intra-rater			-4.78	mm	Mean difference							
							0.5	mm	Intra-rater	0.37	mm	Intra-rater			-0.07	mm	Mean difference							
							0.05	mm	Intra-rater	0.32	mm	Intra-rater			-0.37	mm	Mean difference							
							0.04	mm	Intra-rater	0.14	mm	Intra-rater			-0.39	mm	Mean difference							
							Bootvong K, Liu Z, McGrath C, Hagg U, Wong RW, Bendeus M et al	2010	Virtual model analysis as an alternative approach to plaster model analysis: reliability and validity	Cadent	Laser	80	16 MDW	0.05	mm	Intra-rater	0.08	mm	Intra-rater	0.942	ICC	0.02	mm	Mean difference
														0.04	mm	Intra-rater	0	mm	Intra-rater	0.882	ICC	0.01	mm	Mean difference
0.04	mm	Intra-rater	0.04	mm	Intra-rater	0.908								ICC	0.02	mm	Mean difference							
0.02	mm	Intra-rater	0	mm	Intra-rater	0.9								ICC	0.05	mm	Mean difference							
0.03	mm	Intra-rater	0.06	mm	Intra-rater	0.968								ICC	0.01	mm	Mean difference							
0.01	mm	Intra-rater	0.05	mm	Intra-rater	0.911								ICC	-0.02	mm	Mean difference							
0.02	mm	Intra-rater	0.08	mm	Intra-rater	0.945								ICC	-0.01	mm	Mean difference							
0.07	mm	Intra-rater	0.07	mm	Intra-rater	0.963								ICC	-0.01	mm	Mean difference							
0.01	mm	Intra-rater	0.03	mm	Intra-rater	0.984								ICC	0.02	mm	Mean difference							
0.01	mm	Intra-rater	0.02	mm	Intra-rater	0.948								ICC	0.03	mm	Mean difference							
0.04	mm	Intra-rater	0.02	mm	Intra-rater	0.966								ICC	0.03	mm	Mean difference							
0	mm	Intra-rater	0.07	mm	Intra-rater	0.896								ICC	-0.03	mm	Mean difference							
0.07	mm	Intra-rater	0.04	mm	Intra-rater	0.917								ICC	-0.05	mm	Mean difference							
0.05	mm	Intra-rater	0.02	mm	Intra-rater	0.939								ICC	0.03	mm	Mean difference							
0.04	mm	Intra-rater	0.02	mm	Intra-rater	0.918								ICC	0.04	mm	Mean difference							
0.03	mm	Intra-rater	0.06	mm	Intra-rater	0.914								ICC	0.02	mm	Mean difference							
0.04	mm	Intra-rater	0.03	mm	Intra-rater	0.891								ICC	0.03	mm	Mean difference							
0.03	mm	Intra-rater	0.07	mm	Intra-rater	0.907								ICC	-0.02	mm	Mean difference							
0.01	mm	Intra-rater	0.04	mm	Intra-rater	0.901								ICC	-0.03	mm	Mean difference							
0	mm	Intra-rater	0.11	mm	Intra-rater	0.908								ICC	0.03	mm	Mean difference							
0.03	mm	Intra-rater	0.04	mm	Intra-rater	0.906								ICC	0	mm	Mean difference							
0	mm	Intra-rater	0.01	mm	Intra-rater	0.972								ICC	0.02	mm	Mean difference							
0.05	mm	Intra-rater	0.1	mm	Intra-rater	0.963								ICC	0.03	mm	Mean difference							
0.06	mm	Intra-rater	0.07	mm	Intra-rater	0.918								ICC	-0.01	mm	Mean difference							
0.19	mm	Intra-rater	0.07	mm	Intra-rater	0.967								ICC	-0.03	mm	Mean difference							
0.03	mm	Intra-rater	0.03	mm	Intra-rater	0.983								ICC	0.05	mm	Mean difference							
0.2	mm	Intra-rater	0.07	mm	Intra-rater	0.984								ICC	-0.01	mm	Mean difference							
0.13	mm	Intra-rater	0.36	mm	Intra-rater	0.988								ICC	-0.1	mm	Mean difference							
0.05	mm	Intra-rater	0.06	mm	Intra-rater	0.967								ICC	0.19	mm	Mean difference							
0.05	mm	Intra-rater	0.06	mm	Intra-rater	0.913								ICC	0.16	mm	Mean difference							
0.67	mm	Intra-rater	0.13	mm	Intra-rater	0.984								ICC	0.07	mm	Mean difference							
0.19	mm	Intra-rater	0.06	mm	Intra-rater	0.966								ICC	0.15	mm	Mean difference							
0.2	mm	Intra-rater	0.05	mm	Intra-rater	0.903								ICC	0.12	mm	Mean difference							
Cha BK, Choi JJ, Jost-Brinkmann PG, Jeong YM	2007	Applications of three-dimensionally scanned models in orthodontics	NUS	Laser	30	13 MDW	0.1	mm	SD	0.1	mm	SD			0.11	mm	Mean difference							
							0.1	mm	SD	0.1	mm	SD			0.03	mm	Mean difference							
							0	mm	SD	0.1	mm	SD			0.04	mm	Mean difference							
							0.1	mm	SD	0.1	mm	SD			0.04	mm	Mean difference							

Study	Year	Title	Company	Acquisition type	n	Parameter	Reliability						Validity				
							Plaster models			3D model			Correlation Coefficient (R)		Mean difference (mm)		
							Absolute	Units	Type	Absolute	Units	Type	Value	Units	Value	Units	Type
				Laser	20	14 MDW	Tested	PCC		Tested	PCC			-0.2816	mm	Mean difference	
				Laser	20	13 MDW	Tested	PCC		Tested	PCC			-0.2224	mm	Mean difference	
				Laser	20	12 MDW	Tested	PCC		Tested	PCC			-0.3164	mm	Mean difference	
				Laser	20	11 MDW	Tested	PCC		Tested	PCC			-0.2605	mm	Mean difference	
				Laser	20	21 MDW	Tested	PCC		Tested	PCC			-0.2395	mm	Mean difference	
				Laser	20	22 MDW	Tested	PCC		Tested	PCC			-0.2763	mm	Mean difference	
				Laser	20	23 MDW	Tested	PCC		Tested	PCC			-0.2375	mm	Mean difference	
				Laser	20	24 MDW	Tested	PCC		Tested	PCC			-0.2836	mm	Mean difference	
				Laser	20	25 MDW	Tested	PCC		Tested	PCC			-0.2375	mm	Mean difference	
				Laser	20	26 MDW	Tested	PCC		Tested	PCC			-0.1632	mm	Mean difference	
				Laser	20	36 MDW	Tested	PCC		Tested	PCC			-0.3605	mm	Mean difference	
				Laser	20	35 MDW	Tested	PCC		Tested	PCC			-0.2862	mm	Mean difference	
				Laser	20	34 MDW	Tested	PCC		Tested	PCC			-0.325	mm	Mean difference	
				Laser	20	33 MDW	Tested	PCC		Tested	PCC			-0.2447	mm	Mean difference	
				Laser	20	32 MDW	Tested	PCC		Tested	PCC			-0.3842	mm	Mean difference	
				Laser	20	31 MDW	Tested	PCC		Tested	PCC			-0.2816	mm	Mean difference	
				Laser	20	41 MDW	Tested	PCC		Tested	PCC			-0.2605	mm	Mean difference	
				Laser	20	42 MDW	Tested	PCC		Tested	PCC			-0.3592	mm	Mean difference	
				Laser	20	43 MDW	Tested	PCC		Tested	PCC			-0.2888	mm	Mean difference	
				Laser	20	44 MDW	Tested	PCC		Tested	PCC			-0.3066	mm	Mean difference	
				Laser	20	45 MDW	Tested	PCC		Tested	PCC			-0.2138	mm	Mean difference	
				Laser	20	46 MDW	Tested	PCC		Tested	PCC			-0.3053	mm	Mean difference	
				Laser	20	OB	Tested	PCC		Tested	PCC			-0.4901	mm	Mean difference	
				Laser	20	OJ	Tested	PCC		Tested	PCC			-0.0099	mm	Mean difference	
Sjogren AP, Lindgren JE, Huggare JA	2009	Orthodontic Study Cast Analysis- Reproducibility of Recordings and Agreement Between Conventional and 3D Virtual Measurements	Ortolab	Laser	20	Mx IMW	0.26	mm	SD	0.31	mm	SD		0.1	mm	Mean difference	
				Laser	20	Mn IMW	0.41	mm	SD	0.18	mm	SD		0.2	mm	Mean difference	
				Laser	20	Mx arch circumference	0.37	mm	SD	0.44	mm	SD		-0.9	mm	Mean difference	
				Laser	20	Mn arch perimeter	0.43	mm	SD	0.25	mm	SD		-0.7	mm	Mean difference	
				Laser	20	OJ	0.19	mm	SD	0.12	mm	SD		-0.2	mm	Mean difference	
				Laser	20	OB	0.18	mm	SD	0.1	mm	SD		0	mm	Mean difference	
Stevens DR, Flores-Mir C, Nebbe B, Raboud DW, Heo G, Major PW	2006	Validity, reliability, and reproducibility of plaster vs digital study models: comparison of peer assessment rating and Bolton analysis and their constituent measurements	GeoDigm	Laser	24	Bolton 6	0.32	mm	Intra-rater	0.69	mm	Intra-rater		-0.04	mm	Mean difference	
				Laser	24	Bolton 12	0.58	mm	Intra-rater	1.08	mm	Intra-rater		-0.38	mm	Mean difference	
				Laser	24	Mx length 3-3	0.31	mm	Intra-rater	0.58	mm	Intra-rater		-0.59	mm	Mean difference	
				Laser	24	Mn length 3-3	0.21	mm	Intra-rater	0.62	mm	Intra-rater		-0.4	mm	Mean difference	
				Laser	24	Mx arch perimeter	0.51	mm	Intra-rater	1.13	mm	Intra-rater		-0.2	mm	Mean difference	
				Laser	24	Mn arch perimeter	0.48	mm	Intra-rater	1.07	mm	Intra-rater		0.2	mm	Mean difference	
				Laser	24	OJ	0.49	mm	Intra-rater	0.25	mm	Intra-rater		0.01	mm	Mean difference	
				Laser	24	OB	0.47	mm	Intra-rater	0.2	mm	Intra-rater		-0.3	mm	Mean difference	
				Laser	24	Centerline	0.33	mm	Intra-rater	0.14	mm	Intra-rater		-0.1	mm	Mean difference	
Watanabe-Kanno GA, Abrao J, Miasiro Junior H, Sanchez-Ayala A, Lagravere MO	2009	Reproducibility, reliability and validity of measurements obtained from Ceelec3 digital models	Bilibocast	CBCT	15	36 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.14	mm	Mean difference	
				CBCT	15	35 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	34 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.19	mm	Mean difference	
				CBCT	15	33 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.13	mm	Mean difference	
				CBCT	15	32 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	31 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.13	mm	Mean difference	
				CBCT	15	41 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.15	mm	Mean difference	
				CBCT	15	42 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	43 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.21	mm	Mean difference	
				CBCT	15	44 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.19	mm	Mean difference	
				CBCT	15	45 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.21	mm	Mean difference	
				CBCT	15	46 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.39	mm	Mean difference	
				CBCT	15	16 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	15 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	14 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.11	mm	Mean difference	
				CBCT	15	13 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.11	mm	Mean difference	
				CBCT	15	12 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.15	mm	Mean difference	
				CBCT	15	11 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.14	mm	Mean difference	
				CBCT	15	21 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.18	mm	Mean difference	
				CBCT	15	22 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.07	mm	Mean difference	
				CBCT	15	23 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.19	mm	Mean difference	
				CBCT	15	24 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.19	mm	Mean difference	
				CBCT	15	25 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.19	mm	Mean difference	
				CBCT	15	26 MDW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.2	mm	Mean difference	
				CBCT	15	Mx ICW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.12	mm	Mean difference	
				CBCT	15	Mx IPW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.11	mm	Mean difference	
				CBCT	15	Mx IMW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.16	mm	Mean difference	
				CBCT	15	Mn ICW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.14	mm	Mean difference	
				CBCT	15	Mn IPW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.13	mm	Mean difference	
				CBCT	15	Mn IMW	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.12	mm	Mean difference	
				CBCT	15	OB	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.21	mm	Mean difference	
				CBCT	15	OJ	0.852	ICC	Intra-rater	0.824	ICC	Intra-rater		-0.31	mm	Mean difference	
Zilberman O, Huggare JA, Parikakis KA	2003	Evaluation of the validity of tooth size and arch width measurements using conventional and three-dimensional virtual orthodontic models	Cadent	Laser	20	Incisors	0.0549	mm	Random Error	0.0742	mm	Random Error	0.975	PCC			
				Laser	20	Canines	0.0635	mm	Random Error	0.0725	mm	Random Error	0.827	PCC			
				Laser	20	Premolars	0.0598	mm	Random Error	0.0775	mm	Random Error	0.763	PCC			
				Laser	20	Molars	0.0783	mm	Random Error	0.1173	mm	Random Error	0.849	PCC			
				Laser	20	Arch widths	0.1815	mm	Random Error	0.2031	mm	Random Error	0.998	PCC			

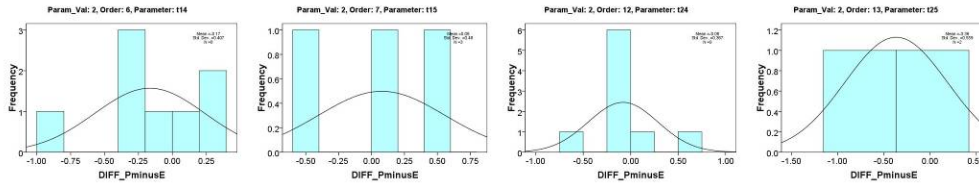
Appendix 3-2. Histograms for the differences in measurements of Anatomodels minus Plaster.



Appendix 3-3. Histograms for the differences in measurements of Anatomodels minus extracted premolars.



Appendix 3-4. Histograms for the differences in measurements of plaster minus extracted premolars.



Appendix 3-5. Sample characteristics for Gender and Mean Age for orthodontic records across three modalities: Anatomodels, Plaster, and Extracted Premolars.

	Anatomodels	Plaster	Extracted Premolars
Gender (% of total)			
Male	18 (60%)		6 (55%)
Female	12 (40%)		5 (45%)
Mean Age in Years (St.dev.)			
Impressions for Plaster	16.5 (5.5)		14.8 (2.5)
CBCT scan	16.8 (5.4)		15.3 (2.4)

Appendix 3-6. Intra-rater, Anatomodels: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).

Anatomodels, Intra-rater

<p><u>Right Molar</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td style="background-color: #d9ead3;">8</td><td>0</td><td>0</td><td>8</td></tr> <tr><td>II</td><td>0</td><td style="background-color: #d9ead3;">2</td><td>0</td><td>2</td></tr> <tr><td>III</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 3 vs 5</p>		I	II	III		I	8	0	0	8	II	0	2	0	2	III	0	0	0	0		8	2	0	10	<p><u>Left Molar</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td style="background-color: #d9ead3;">8</td><td>0</td><td>0</td><td>8</td></tr> <tr><td>II</td><td>0</td><td style="background-color: #d9ead3;">2</td><td>0</td><td>2</td></tr> <tr><td>III</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 2 vs 1</p>		I	II	III		I	8	0	0	8	II	0	2	0	2	III	0	0	0	0		8	2	0	10	<p><u>Right Canine</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td style="background-color: #d9ead3;">7</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>II</td><td>0</td><td style="background-color: #d9ead3;">3</td><td>0</td><td>3</td></tr> <tr><td>III</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>7</td><td>3</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 2 vs 3</p>		I	II	III		I	7	0	0	7	II	0	3	0	3	III	0	0	0	0		7	3	0	10	<p><u>Left Canine</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td style="background-color: #d9ead3;">7</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>II</td><td>0</td><td style="background-color: #d9ead3;">3</td><td>0</td><td>3</td></tr> <tr><td>III</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>7</td><td>3</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 4 vs 3</p>		I	II	III		I	7	0	0	7	II	0	3	0	3	III	0	0	0	0		7	3	0	10
	I	II	III																																																																																																				
I	8	0	0	8																																																																																																			
II	0	2	0	2																																																																																																			
III	0	0	0	0																																																																																																			
	8	2	0	10																																																																																																			
	I	II	III																																																																																																				
I	8	0	0	8																																																																																																			
II	0	2	0	2																																																																																																			
III	0	0	0	0																																																																																																			
	8	2	0	10																																																																																																			
	I	II	III																																																																																																				
I	7	0	0	7																																																																																																			
II	0	3	0	3																																																																																																			
III	0	0	0	0																																																																																																			
	7	3	0	10																																																																																																			
	I	II	III																																																																																																				
I	7	0	0	7																																																																																																			
II	0	3	0	3																																																																																																			
III	0	0	0	0																																																																																																			
	7	3	0	10																																																																																																			

<p><u>Maxillary Shape</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>U</td><td>V</td><td>Tp</td><td>Sq</td><td></td></tr> <tr><td>U</td><td style="background-color: #d9ead3;">9</td><td>0</td><td>0</td><td style="background-color: #f2dede;">1</td><td>10</td></tr> <tr><td>V</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Tp</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td></tr> <tr><td>Sq</td><td>0</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>9</td><td>0</td><td>0</td><td>1</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 5 vs 2</p>		U	V	Tp	Sq		U	9	0	0	1	10	V	0	0	0	0	0	Tp	0	0	0	0	0	Sq	0	0	0	0	0		9	0	0	1	10	<p><u>Maxillary Size</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Na</td><td>Av</td><td>Ex</td><td></td></tr> <tr><td>Na</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Av</td><td>0</td><td style="background-color: #d9ead3;">10</td><td>0</td><td>10</td></tr> <tr><td>Ex</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>0</td><td>10</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 3 vs 4</p>		Na	Av	Ex		Na	0	0	0	0	Av	0	10	0	10	Ex	0	0	0	0		0	10	0	10	<p><u>Maxillary Symmetry</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Sy</td><td>As</td><td></td></tr> <tr><td>Sy</td><td style="background-color: #d9ead3;">7</td><td style="background-color: #f2dede;">1</td><td>8</td></tr> <tr><td>As</td><td style="background-color: #f2dede;">1</td><td style="background-color: #d9ead3;">1</td><td>2</td></tr> <tr><td></td><td>8</td><td>2</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 3 vs 5</p>		Sy	As		Sy	7	1	8	As	1	1	2		8	2	10
	U	V	Tp	Sq																																																																											
U	9	0	0	1	10																																																																										
V	0	0	0	0	0																																																																										
Tp	0	0	0	0	0																																																																										
Sq	0	0	0	0	0																																																																										
	9	0	0	1	10																																																																										
	Na	Av	Ex																																																																												
Na	0	0	0	0																																																																											
Av	0	10	0	10																																																																											
Ex	0	0	0	0																																																																											
	0	10	0	10																																																																											
	Sy	As																																																																													
Sy	7	1	8																																																																												
As	1	1	2																																																																												
	8	2	10																																																																												

<p><u>Mandibular Shape</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>U</td><td>V</td><td>Tp</td><td>Sq</td><td></td></tr> <tr><td>U</td><td style="background-color: #d9ead3;">7</td><td>0</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>V</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Tp</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td></tr> <tr><td>Sq</td><td>0</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">3</td><td>3</td></tr> <tr><td></td><td>7</td><td>0</td><td>0</td><td>3</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 1 vs 4</p>		U	V	Tp	Sq		U	7	0	0	0	7	V	0	0	0	0	0	Tp	0	0	0	0	0	Sq	0	0	0	3	3		7	0	0	3	10	<p><u>Mandibular Size</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Na</td><td>Av</td><td>Ex</td><td></td></tr> <tr><td>Na</td><td style="background-color: #d9ead3;">0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Av</td><td>0</td><td style="background-color: #d9ead3;">10</td><td>0</td><td>10</td></tr> <tr><td>Ex</td><td>0</td><td>0</td><td style="background-color: #d9ead3;">0</td><td>0</td></tr> <tr><td></td><td>0</td><td>10</td><td>0</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 2 vs 5</p>		Na	Av	Ex		Na	0	0	0	0	Av	0	10	0	10	Ex	0	0	0	0		0	10	0	10	<p><u>Mandibular Symmetry</u></p> <table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Sy</td><td>As</td><td></td></tr> <tr><td>Sy</td><td style="background-color: #d9ead3;">8</td><td>0</td><td>8</td></tr> <tr><td>As</td><td style="background-color: #f2dede;">1</td><td style="background-color: #d9ead3;">1</td><td>2</td></tr> <tr><td></td><td>9</td><td>1</td><td>10</td></tr> </table> <p style="text-align: center;">Trials 3 vs 2</p>		Sy	As		Sy	8	0	8	As	1	1	2		9	1	10
	U	V	Tp	Sq																																																																											
U	7	0	0	0	7																																																																										
V	0	0	0	0	0																																																																										
Tp	0	0	0	0	0																																																																										
Sq	0	0	0	3	3																																																																										
	7	0	0	3	10																																																																										
	Na	Av	Ex																																																																												
Na	0	0	0	0																																																																											
Av	0	10	0	10																																																																											
Ex	0	0	0	0																																																																											
	0	10	0	10																																																																											
	Sy	As																																																																													
Sy	8	0	8																																																																												
As	1	1	2																																																																												
	9	1	10																																																																												

Overall Pairs	Count (%)
Concordant	96 (96%)
Discordant	4 (4%)
Total	100 (100%)

Abbreviations: U, U-shaped; V, V-shaped; Tp, Tapered; Sq, Squared; Na, Narrow; Av, Average; Ex, Expanded; Sy, Symmetric; As, Asymmetric

Appendix 3-7. Intra-rater, Plaster: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).

Plaster, Intra-rater

Right Molar

	I	II	III	
I	8	0	0	8
II	0	2	0	2
III	0	0	0	0
	8	2	0	10

Trials 4 vs 3

Left Molar

	I	II	III	
I	8	0	0	8
II	0	2	0	2
III	0	0	0	0
	8	2	0	10

Trials 5 vs 2

Right Canine

	I	II	III	
I	7	0	0	7
II	0	3	0	3
III	0	0	0	0
	7	3	0	10

Trials 1 vs 3

Left Canine

	I	II	III	
I	7	0	0	7
II	1	2	0	3
III	0	0	0	0
	8	2	0	10

Trials 4 vs 3

Maxillary Shape

	U	V	Tp	Sq	
U	8	0	0	0	8
V	1	1	0	0	2
Tp	0	0	0	0	0
Sq	0	0	0	0	0
	9	1	0	0	10

Trials 3 vs 5

Maxillary Size

	Na	Av	Ex	
Na	0	1	0	1
Av	0	9	0	9
Ex	0	0	0	0
	0	10	0	10

Trials 4 vs 3

Maxillary Symmetry

	Sy	As	
Sy	10	0	10
As	0	0	0
	10	0	10

Trials 2 vs 1

Mandibular Shape

	U	V	Tp	Sq	
U	6	0	0	0	6
V	0	0	0	0	0
Tp	0	0	0	0	0
Sq	0	0	0	4	4
	6	0	0	4	10

Trials 3 vs 2

Mandibular Size

	Na	Av	Ex	
Na	0	0	0	0
Av	0	10	0	10
Ex	0	0	0	0
	0	10	0	10

Trials 3 vs 1

Mandibular Symmetry

	Sy	As	
Sy	9	1	10
As	0	0	0
	9	1	10

Trials 5 vs 3

Overall Pairs	Count (%)
Concordant	96 (96%)
Discordant	4 (4%)
Total	100 (100%)

Abbreviations: U, U-shaped; V, V-shaped; Tp, Tapered; Sq, Squared; Na, Narrow; Av, Average; Ex, Expanded; Sy, Symmetric; As, Asymmetric

Appendix 3-8. Inter-rater, Anatomodels: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).

Anatomodels, Interrater

Right Molar	Left Molar	Right Canine	Left Canine																																																																																																				
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td>8</td><td>0</td><td>0</td><td>8</td></tr> <tr><td>II</td><td>0</td><td>2</td><td>0</td><td>2</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p>NL vs MM</p>		I	II	III		I	8	0	0	8	II	0	2	0	2	III	0	0	0	0		8	2	0	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td>7</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>II</td><td>1</td><td>2</td><td>0</td><td>3</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p>TE vs ML</p>		I	II	III		I	7	0	0	7	II	1	2	0	3	III	0	0	0	0		8	2	0	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td>7</td><td>1</td><td>0</td><td>8</td></tr> <tr><td>II</td><td>0</td><td>2</td><td>0</td><td>2</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>7</td><td>3</td><td>0</td><td>10</td></tr> </table> <p>MM vs TE</p>		I	II	III		I	7	1	0	8	II	0	2	0	2	III	0	0	0	0		7	3	0	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>I</td><td>II</td><td>III</td><td></td></tr> <tr><td>I</td><td>5</td><td>0</td><td>0</td><td>5</td></tr> <tr><td>II</td><td>2</td><td>3</td><td>0</td><td>5</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>7</td><td>3</td><td>0</td><td>10</td></tr> </table> <p>CF vs NL</p>		I	II	III		I	5	0	0	5	II	2	3	0	5	III	0	0	0	0		7	3	0	10
	I	II	III																																																																																																				
I	8	0	0	8																																																																																																			
II	0	2	0	2																																																																																																			
III	0	0	0	0																																																																																																			
	8	2	0	10																																																																																																			
	I	II	III																																																																																																				
I	7	0	0	7																																																																																																			
II	1	2	0	3																																																																																																			
III	0	0	0	0																																																																																																			
	8	2	0	10																																																																																																			
	I	II	III																																																																																																				
I	7	1	0	8																																																																																																			
II	0	2	0	2																																																																																																			
III	0	0	0	0																																																																																																			
	7	3	0	10																																																																																																			
	I	II	III																																																																																																				
I	5	0	0	5																																																																																																			
II	2	3	0	5																																																																																																			
III	0	0	0	0																																																																																																			
	7	3	0	10																																																																																																			

Maxillary Shape	Maxillary Size	Maxillary Symmetry																																																																													
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>U</td><td>V</td><td>Tp</td><td>Sq</td><td></td></tr> <tr><td>U</td><td>4</td><td>0</td><td>0</td><td>0</td><td>4</td></tr> <tr><td>V</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Tp</td><td>4</td><td>0</td><td>0</td><td>0</td><td>4</td></tr> <tr><td>Sq</td><td>1</td><td>0</td><td>0</td><td>1</td><td>2</td></tr> <tr><td></td><td>9</td><td>0</td><td>0</td><td>1</td><td>10</td></tr> </table> <p>TE vs ML</p>		U	V	Tp	Sq		U	4	0	0	0	4	V	0	0	0	0	0	Tp	4	0	0	0	4	Sq	1	0	0	1	2		9	0	0	1	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Na</td><td>Av</td><td>Ex</td><td></td></tr> <tr><td>Na</td><td>0</td><td>2</td><td>0</td><td>2</td></tr> <tr><td>Av</td><td>0</td><td>8</td><td>0</td><td>8</td></tr> <tr><td>Ex</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>0</td><td>10</td><td>0</td><td>10</td></tr> </table> <p>MM vs ML</p>		Na	Av	Ex		Na	0	2	0	2	Av	0	8	0	8	Ex	0	0	0	0		0	10	0	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Sy</td><td>As</td><td></td></tr> <tr><td>Sy</td><td>8</td><td>0</td><td>8</td></tr> <tr><td>As</td><td>2</td><td>0</td><td>0</td></tr> <tr><td></td><td>10</td><td>0</td><td>10</td></tr> </table> <p>CF vs ML</p>		Sy	As		Sy	8	0	8	As	2	0	0		10	0	10
	U	V	Tp	Sq																																																																											
U	4	0	0	0	4																																																																										
V	0	0	0	0	0																																																																										
Tp	4	0	0	0	4																																																																										
Sq	1	0	0	1	2																																																																										
	9	0	0	1	10																																																																										
	Na	Av	Ex																																																																												
Na	0	2	0	2																																																																											
Av	0	8	0	8																																																																											
Ex	0	0	0	0																																																																											
	0	10	0	10																																																																											
	Sy	As																																																																													
Sy	8	0	8																																																																												
As	2	0	0																																																																												
	10	0	10																																																																												

Mandibular Shape	Mandibular Size	Mandibular Symmetry																																																																													
<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>U</td><td>V</td><td>Tp</td><td>Sq</td><td></td></tr> <tr><td>U</td><td>7</td><td>0</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>V</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Tp</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Sq</td><td>0</td><td>0</td><td>0</td><td>3</td><td>3</td></tr> <tr><td></td><td>7</td><td>0</td><td>0</td><td>3</td><td>10</td></tr> </table> <p>ML vs NL</p>		U	V	Tp	Sq		U	7	0	0	0	7	V	0	0	0	0	0	Tp	0	0	0	0	0	Sq	0	0	0	3	3		7	0	0	3	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Na</td><td>Av</td><td>Ex</td><td></td></tr> <tr><td>Na</td><td>1</td><td>1</td><td>0</td><td>2</td></tr> <tr><td>Av</td><td>2</td><td>6</td><td>0</td><td>8</td></tr> <tr><td>Ex</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td></td><td>3</td><td>7</td><td>0</td><td>10</td></tr> </table> <p>CF vs MM</p>		Na	Av	Ex		Na	1	1	0	2	Av	2	6	0	8	Ex	0	0	0	0		3	7	0	10	<table border="1" style="border-collapse: collapse; text-align: center;"> <tr><td></td><td>Sy</td><td>As</td><td></td></tr> <tr><td>Sy</td><td>6</td><td>0</td><td>6</td></tr> <tr><td>As</td><td>3</td><td>1</td><td>4</td></tr> <tr><td></td><td>9</td><td>1</td><td>10</td></tr> </table> <p>TE vs NL</p>		Sy	As		Sy	6	0	6	As	3	1	4		9	1	10
	U	V	Tp	Sq																																																																											
U	7	0	0	0	7																																																																										
V	0	0	0	0	0																																																																										
Tp	0	0	0	0	0																																																																										
Sq	0	0	0	3	3																																																																										
	7	0	0	3	10																																																																										
	Na	Av	Ex																																																																												
Na	1	1	0	2																																																																											
Av	2	6	0	8																																																																											
Ex	0	0	0	0																																																																											
	3	7	0	10																																																																											
	Sy	As																																																																													
Sy	6	0	6																																																																												
As	3	1	4																																																																												
	9	1	10																																																																												

Overall Pairs	Count (%)
Concordant	81 (81%)
Discordant	19 (19%)
Total	100 (100%)

Abbreviations: U, U-shaped; V, V-shaped; Tp, Tapered; Sq, Squared; Na, Narrow; Av, Average; Ex, Expanded; Sy, Symmetric; As, Asymmetric

Appendix 3-9. Inter-rater, Plaster: nominal parameter crosstabulations of paired trials randomly chosen from five with a summary of overall concordant pairs (green) and discordant pairs (red).

Plaster, Interrater

<p><u>Right Molar</u></p> <table border="1"> <tr><td>I</td><td>8</td></tr> <tr><td>II</td><td>2</td></tr> <tr><td>III</td><td>0</td></tr> <tr><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p>TE vs CF</p>	I	8	II	2	III	0	8	2	0	10	<p><u>Left Molar</u></p> <table border="1"> <tr><td>I</td><td>7</td><td>1</td><td>0</td><td>8</td></tr> <tr><td>II</td><td>0</td><td>2</td><td>0</td><td>2</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>7</td><td>3</td><td>0</td><td>10</td></tr> </table> <p>ML vs TE</p>	I	7	1	0	8	II	0	2	0	2	III	0	0	0	0	7	3	0	10	<p><u>Right Canine</u></p> <table border="1"> <tr><td>I</td><td>6</td><td>0</td><td>1</td><td>7</td></tr> <tr><td>II</td><td>0</td><td>3</td><td>0</td><td>3</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>6</td><td>3</td><td>1</td><td>10</td></tr> </table> <p>NL vs MM</p>	I	6	0	1	7	II	0	3	0	3	III	0	0	0	0	6	3	1	10	<p><u>Left Canine</u></p> <table border="1"> <tr><td>I</td><td>7</td><td>0</td><td>0</td><td>7</td></tr> <tr><td>II</td><td>1</td><td>2</td><td>0</td><td>3</td></tr> <tr><td>III</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>8</td><td>2</td><td>0</td><td>10</td></tr> </table> <p>NL vs TE</p>	I	7	0	0	7	II	1	2	0	3	III	0	0	0	0	8	2	0	10
I	8																																																																					
II	2																																																																					
III	0																																																																					
8	2	0	10																																																																			
I	7	1	0	8																																																																		
II	0	2	0	2																																																																		
III	0	0	0	0																																																																		
7	3	0	10																																																																			
I	6	0	1	7																																																																		
II	0	3	0	3																																																																		
III	0	0	0	0																																																																		
6	3	1	10																																																																			
I	7	0	0	7																																																																		
II	1	2	0	3																																																																		
III	0	0	0	0																																																																		
8	2	0	10																																																																			

<p><u>Maxillary Shape</u></p> <table border="1"> <tr><td>U</td><td>6</td><td>2</td><td>0</td><td>1</td><td>9</td></tr> <tr><td>V</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>Tp</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Sq</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>7</td><td>2</td><td>0</td><td>1</td><td>10</td></tr> </table> <p>NL vs MM</p>	U	6	2	0	1	9	V	1	0	0	0	1	Tp	0	0	0	0	0	Sq	0	0	0	0	0	7	2	0	1	10	<p><u>Maxillary Size</u></p> <table border="1"> <tr><td>Na</td><td>1</td><td>3</td><td>0</td><td>4</td></tr> <tr><td>Av</td><td>0</td><td>3</td><td>0</td><td>3</td></tr> <tr><td>Ex</td><td>0</td><td>3</td><td>0</td><td>3</td></tr> <tr><td>1</td><td>9</td><td>0</td><td>10</td></tr> </table> <p>TE vs ML</p>	Na	1	3	0	4	Av	0	3	0	3	Ex	0	3	0	3	1	9	0	10	<p><u>Maxillary Symmetry</u></p> <table border="1"> <tr><td>Sy</td><td>5</td><td>2</td><td>7</td></tr> <tr><td>As</td><td>3</td><td>0</td><td>3</td></tr> <tr><td>8</td><td>2</td><td>10</td></tr> </table> <p>MM vs TE</p>	Sy	5	2	7	As	3	0	3	8	2	10
U	6	2	0	1	9																																																								
V	1	0	0	0	1																																																								
Tp	0	0	0	0	0																																																								
Sq	0	0	0	0	0																																																								
7	2	0	1	10																																																									
Na	1	3	0	4																																																									
Av	0	3	0	3																																																									
Ex	0	3	0	3																																																									
1	9	0	10																																																										
Sy	5	2	7																																																										
As	3	0	3																																																										
8	2	10																																																											

<p><u>Mandibular Shape</u></p> <table border="1"> <tr><td>U</td><td>4</td><td>1</td><td>1</td><td>1</td><td>7</td></tr> <tr><td>V</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Tp</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>Sq</td><td>1</td><td>1</td><td>0</td><td>1</td><td>3</td></tr> <tr><td>5</td><td>2</td><td>1</td><td>2</td><td>10</td></tr> </table> <p>ML vs MM</p>	U	4	1	1	1	7	V	0	0	0	0	0	Tp	0	0	0	0	0	Sq	1	1	0	1	3	5	2	1	2	10	<p><u>Mandibular Size</u></p> <table border="1"> <tr><td>Na</td><td>0</td><td>4</td><td>0</td><td>4</td></tr> <tr><td>Av</td><td>0</td><td>6</td><td>0</td><td>6</td></tr> <tr><td>Ex</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>10</td><td>0</td><td>10</td></tr> </table> <p>TE vs CF</p>	Na	0	4	0	4	Av	0	6	0	6	Ex	0	0	0	0	0	10	0	10	<p><u>Mandibular Symmetry</u></p> <table border="1"> <tr><td>Sy</td><td>10</td><td>0</td><td>10</td></tr> <tr><td>As</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>10</td><td>0</td><td>10</td></tr> </table> <p>CF vs NL</p>	Sy	10	0	10	As	0	0	0	10	0	10
U	4	1	1	1	7																																																								
V	0	0	0	0	0																																																								
Tp	0	0	0	0	0																																																								
Sq	1	1	0	1	3																																																								
5	2	1	2	10																																																									
Na	0	4	0	4																																																									
Av	0	6	0	6																																																									
Ex	0	0	0	0																																																									
0	10	0	10																																																										
Sy	10	0	10																																																										
As	0	0	0																																																										
10	0	10																																																											

Overall Pairs	Count (%)
Concordant	73 (73%)
Discordant	27 (27%)
Total	100 (100%)

Abbreviations: U, U-shaped; V, V-shaped; Tp, Tapered; Sq, Squared; Na, Narrow; Av, Average; Ex, Expanded; Sy, Symmetric; As, Asymmetric

Appendix 3-10. Validity, Anatomomodels vs Plaster: nominal parameter crosstabulations of paired assessments with a summary of overall concordant pairs (green) and discordant pairs (red).

Anatomomodels vs Plaster

Right Molar

	I	II	III	
I	14	0	0	14
II	0	3	0	3
III	0	0	1	1
	14	3	1	18

Left Molar

	I	II	III	
I	13	0	0	13
II	0	3	0	3
III	1	0	1	2
	14	3	1	18

Right Canine

	I	II	III	
I	12	0	0	12
II	0	5	0	5
III	0	0	1	1
	12	5	1	18

Left Canine

	I	II	III	
I	11	1	0	12
II	0	4	0	4
III	2	0	0	2
	13	5	0	18

Maxillary Shape

	U	V	Tp	Sq	
U	26	0	0	1	27
V	0	1	0	0	1
Tp	0	0	0	0	0
Sq	0	0	0	2	2
	26	1	0	3	30

Maxillary Size

	Na	Av	Ex	
Na	0	1	0	1
Av	0	28	0	28
Ex	0	1	0	1
	0	30	0	30

Maxillary Symmetry

	Sy	As	
Sy	26	2	28
As	1	1	2
	27	3	30

Mandibular Shape

	U	V	Tp	Sq	
U	22	0	0	3	25
V	0	0	0	0	0
Tp	0	0	0	0	0
Sq	1	0	0	4	5
	23	0	0	7	30

Mandibular Size

	Na	Av	Ex	
Na	0	1	0	1
Av	0	29	0	29
Ex	0	0	0	0
	0	30	0	30

Mandibular Symmetry

	Sy	As	
Sy	25	0	25
As	4	1	5
	29	1	30

Overall Pairs	Count (%)
Concordant	233 (92%)
Discordant	19 (8%)
Total	252 (100%)

Abbreviations: U, U-shaped; V, V-shaped; Tp, Tapered; Sq, Squared; Na, Narrow; Av, Average; Ex, Expanded; Sy, Symmetric; As, Asymmetric