Association Between Crowding Estimation and Extraction Recommendations in Orthodontics

Kaitlyn Tom

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Dentistry

University of Washington
2023

Committee:
Anne-Marie Bollen
Lloyd Mancl
Heather Woloshyn
Roozbeh Khosravi

Program Authorized to Offer Degree:
School of Dentistry – Department of Orthodontics

University of Washington

Abstract

Association Between Crowding Estimation and Extraction Recommendations in Orthodontics

Kaitlyn Tom, DDS

Chair of the Supervisory Committee:
Dr. Anne-Marie Bollen
Orthodontics Department

Little is known about how precisely orthodontists in the United States (U.S.) assess crowding in their own practices or at what range of crowding they decide to recommend extractions. The purpose of this study is to assess the relationship between estimated crowding in Class I cases and the recommendation for extraction treatment by orthodontists in the U.S. Secondary aims are to evaluate the accuracy and precision of clinicians' estimations compared to objective manual measurements and determine if clinician sex, region, experience, method of crowding assessment, and extraction prevalence, play a role in extraction decision-making.

An electronic survey was prepared using four Class I patients with anterior crowding selected from the University of Washington Orthodontics Clinic. The survey was sent to approximately 10,400 subjects through Facebook and the American Association of Orthodontists (AAO). In addition to the patient cases, questions about the clinicians were included.

From the 297 responses received, it was observed that most clinicians recommended extraction once estimated crowding was classified as severe and reached 8-10mm in either the

maxilla or the mandible. Clinicians' estimations varied widely but on average, were precise within approximately 2mm of the objective measurements. There was a tendency to overestimate crowding, especially by clinicians practicing in the Northeast. Clinicians who reported that they routinely measured crowding or had recommended extractions to more than 10% of their patients were 1.2-2 and 1.4-1.6 times more likely, respectively, to recommend extraction in the cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank:

My mentor, Dr. Bollen, for her guidance throughout this project.

Committee members, Dr. Woloshyn and Dr. Khosravi for their advice.

Dr. Mancl for his help with statistical analysis.

UW Orthodontics graduate students for their support and contribution of patient cases.

UW Alumni Association for their financial support.

INTRODUCTION:

Dental crowding has been identified as the main reason that people seek orthodontic treatment (Gilmore *et al.*, 1984, Gosney, 1986) and may be defined as the rotation, ectopic eruption, or impaction of teeth that occurs when the space required for ideal alignment exceeds the available space in the dental arch (Van der Linden, 1974). This malocclusion is more prevalent in the anterior dentition and affects the mandible more than the maxilla. Mandibular crowding alone, has been found to impact 63% of the population and tends to worsen with age (Proffit *et al.*, 2012).

Crowding can be resolved in several ways, depending on the severity or amount present. Non-extraction treatment using methods such as mesiodistal reduction of teeth or expansion of the dental arch is typically recommended for mild crowding, while extractions may be considered as crowding becomes severe (Proffit *et al.*, 2012). Several studies have examined variation in crowding estimations as well as treatment decisions between clinicians and have found considerable differences (Baelum *et al.*, 2012, Brown *et al.*, 1977, Lee *et al.*, 1999, Naish *et al.*, 2016, Ribarevski *et al.*, 1996, Saghafi *et al.*, 2017, Wallis *et al.*, 2014, Weintraub *et al.*, 1989). There are a number of factors that might contribute to these inconsistencies, such as differences in training and types of diagnostic tools or records utilized (Bader *et al*., 1995). One of the biggest contributors, however, is thought to be the lack of clearly defined standards for classifying orthodontic problems and making treatment decisions (Luke *et al*., 1998). Though the American Association of Orthodontists (AAO) has released clinical practice guidelines to aid in diagnosis and treatment planning, they are very generalized. Overall, literature involving the diagnosis of orthodontic problems is limited. For the purposes of adhering to a similar standard of care when treating orthodontic patients, discrepancies in crowding assessment and treatment

decisions presents an issue. Currently, there are few evidence-based clinical guidelines that identify a specific range for crowding at which extraction treatment should be considered. The only known parameters are those provided by Proffit, who suggests either extraction or non-extraction treatment when crowding is 5-9mm and extraction treatment once crowding reaches 10mm.

While the extraction decision is multifactorial and depends on a number of other factors including facial profile, incisor inclination, periodontal health, and curve of Spee (Ali *et al.,* 2018, Baumrind *et al.*, 1996, Saghafi *et al*., 2017), the identification of a range for crowding at which extractions are generally considered can help to guide orthodontists in making this decision especially when there is still uncertainty after considering all other factors. The extraction of permanent teeth is irreversible and is considered by patients to be one of the highest anxiety-provoking dental procedures (Oosterink *et al*., 2008). Further, the decision between extraction and non-extraction treatment can significantly impact the health of the periodontium. For these reasons, it is especially important for the practitioner to carefully consider all factors before making this recommendation.

Currently, there is little known about how precisely orthodontists in the United States (U.S.) assess crowding in their own practices or at what range of crowding they decide to recommend extraction. The purpose of this study is to assess the relationship between estimated crowding and the recommendation for extraction by orthodontists in the United States. Secondary aims are to evaluate the accuracy and precision of clinicians' estimations of crowding compared to objective manual measurements and determine if clinicians' sex, region, experience, method of crowding assessment and extraction prevalence play a role in extraction recommendation.

MATERIALS & METHODS:

Selection of patient cases

This study was approved by the University of Washington (UW) Institutional Review Board (IRB). Cases were selected from the UW Orthodontic Department based on the following inclusion criteria: full permanent dentition, Class I molar classification, normal incisor inclination, presence of maxillary and or mandibular crowding, lip competency, complete records available (facial and intraoral photographs, intraoral scan, panoramic and lateral cephalometric radiographs). Cases were excluded if they met the following criteria: missing teeth other than third molars, worn or fractured dentition, restorations that make up 50% or more of the crown structure, defective or grossly contoured restorations, root canal treatment, ankylosis, moderate-severe decay, periodontal disease, moderate-severe gingival recession, and other pathology. Of the ten cases that were screened, four met the inclusion criteria and were selected for the survey.

Crowding Analysis

Maxillary and mandibular resin models were printed for each case with a SprintRay® three- dimensional printer. One orthodontic graduate student measured crowding (mm) on each model by subtracting the required arch perimeter from the available arch perimeter (Moyers, 1973). The primary method used to measure available arch perimeter in this study was linear segmentation. For comparison, available arch perimeter was also measured with brass wire. For the linear segment method, the sum of five linear segments comprised the available arch perimeter: mesial of the first molars to the mesial of the first premolars, mesial of the first premolars to the mesial of the lateral incisors, and from the mesial of the lateral incisor to the mesial of the contralateral lateral incisor (adapted from Lundstrom, 1955, Moorrees *et al.*, 1954).

For the brass wire method, a continuous strand of brass wire was adapted around the arch mesial to the first molars over the buccal cusp tips of the premolars and along the ideal positions for the incisal edges of the anterior teeth (Nance, 1947, Carey, 1952). The required arch perimeter was calculated by adding together the widths of all teeth mesial to the first molars. Four weeks after the initial measurements, the same student repeated crowding analysis on each model in a random order. The decision to measure crowding on physical rather than digital models was based on variation in study findings, with some reporting digital measurements to be larger than manual measurements, and others finding them to be smaller (Adobes *et al.*, 2022, Naidu *et al.*, 2013, Santoro *et al*., 2003, Shailendran *et al*., 2021). The assessor was blinded to all records and patient identifiers during the measurement process. The average of the crowding measurements taken from the two time points four weeks apart (timepoint 1 (T1) and timepoint 2 (T2)) was used as the measured crowding value for each case. All landmarks on the lateral cephalometric radiographs were traced by the same orthodontic resident and reviewed by the same faculty member.

Sample

In order to be eligible to participate in this study, an orthodontist or orthodontic resident/graduate student needed to be practicing in the U.S or in training at an accredited U.S orthodontic residency/graduate program.

Electronic survey

A survey was created with SurveyMonkey® and sent to all members participating in the AAO Partners in Research Program as well as members of the Orthodontic Pearls Facebook® group. The consent and recruitment form included with the survey specified that the survey was for clinicians practicing in the United States.

Survey questions were multiple choice and open-ended (Appendix I). The first part of the survey included questions about the clinicians' demographics: sex, region, years of experience, extraction prevalence, and preferred method for crowding assessment. The second part of the survey included images of the records for each case (intra- and extraoral photographs, lateral cephalometric and panoramic radiographs, and scanned models) along with follow-up questions about crowding assessment and preferred extraction or non-extraction treatment. For crowding assessment, a numerical amount was chosen using a sliding scale set at a minimum of 0mm and maximum of 20mm. Crowding was further classified as either none, mild, moderate, or severe. Depending on the clinicians' treatment decision, follow up questions were asked specifying their extraction pattern or method(s) to alleviate crowding if they chose not to extract. Patient cases were randomized in the order that they were presented.

Outcomes

Accuracy is the average difference between the clinician's crowding estimation and the objective measurement. A negative value indicates underestimation by the clinician while a positive value indicates overestimation. Precision, or measurement error, is the average absolute difference between the clinician's crowding estimation and the objective measurement. A lower absolute difference indicates that the clinician's estimation is closer to the objective measurement and more precise.

Statistical Analysis

Based on a 2016 survey of AAO members regarding orthodontists' decisions for extraction, the expected sample size was 200-300 practitioners, which was determined have 80% power with a two-sided .05 significance level to detect a 40% to 50% increase in the odds (odds ratio of 1.4 to 1.5) for recommending extraction for each 1mm increase in crowding if the rate of

recommending extraction is 40-60% for crowding around 5mm (Hsieh *et al.*, 1998). Intra-examiner error in crowding measurements was calculated with Dahlberg's formula. Descriptive summaries were run to determine general trends in the data. Paired t-test and linear regression using the generalized estimating equations (GEE) method to account for the repeated measures were used to compare the accuracy and precision between the cases and arches. Two-sample t-test, one-way ANOVA, non-parametric Fisher's exact test, and Pearson's chi-squared test determined the association between practitioners' characteristics and accuracy as well as extraction recommendation. The Holm's method was used to adjust the statistical significance (p-value) for post-hoc multiple testing. The relationship between crowding estimation and extraction recommendation was examined using scatter plots and locally weighted smoothing (LOESS).

RESULTS:

Clinicians' Demographics and Survey Completion

The survey was distributed to 1,904 subjects through the AAO Partners in Research Program and approximately 8,500 subjects through the Facebook® Orthodontic Pearls group for a total of 10,404 subjects. Of the 373 clinicians who responded to initial questions about their background, 266 completed questions for all cases, 31 completed questions for 1-3 cases, and 76 did not complete any case questions (Table I). A breakdown of cases completed by the 31 subjects who completed questions for 1-3 cases is shown in Table II. 179 of those who completed case questions came from Facebook® and 118 from the AAO. Cases A and B both received 280 responses while Cases C and D received 281 responses. As the Orthodontic Pearls Facebook® group is international, a true response rate for this group could not be determined as it is unknown how many total members within the group practiced in the U.S. Survey response rate

for the AAO group was 7.2%. Overall, experience of clinicians varied, with 23.3% having practiced for less than 5 years, 28.2% between 5-15 years, and 38.9% for more than 15 years (Table III). The majority of responders from the AAO group had more than 15 years of experience (60.1%) while those who responded from Facebook® were more evenly distributed by experience (Appendix II, III). Most respondents were male (63%) and practiced in the Western region (30%). Visual estimation was the primary method of crowding assessment (81.8%) with the majority of the clinicians (57%) reporting they have recommended extractions to 10-25% of their patients (Table III). There was no association between years practiced or clinician sex and reported frequency of extractions (Table IV).

Table I. Number of Clinicians Completing All Cases, 1-3 Cases, and No Cases

Number of cases completed	N = 373
All cases	266 (71.3%)
1-3 cases	31 (8.3%)
No cases	76 (20.4%)

Table II. Number of Clinicians Completing 1-3 Cases

1-3 cases completed	N = 31
Case A only	4 (12.9%)
Case B only	6 (19.3%)
Cases A and B	5 (16.1%)
Cases C and D	8 (25.8%)
Cases B, C, and D	3 (9.6%)
Cases A, C, and D	5 (16.1%)

Table III. Clinicians Demographics: Years Practiced, Sex, Region, Extraction Frequency, and Crowding Assessment Method

Characteristic	Overall, N = 373
Years practiced	
0	36 (9.7%)
<5	87 (23.3%)
5-15	105 (28.2%)
>15	145 (38.9%)
Sex	
Female	134 (35.9%)
Male	235 (63.0%)
Prefer not to say	4 (1.1%)
AAO regional group	
Midwest	74 (19.8%)
Northeast	66 (17.7%)
West	112 (30.0%)
South	104 (27.9%)
Other	17 (4.6%)
Recommend extraction	
<10%	90 (24.2%)
10-25%	212 (57.0%)
25-75%	70 (18.8%)
Crowding assessment	
Visual estimation only	305 (81.8%)
Measurement on study models	43 (11.5%)
Visual estimation and other methods	25 (6.7%)

Table IV. Association Between Clinician Sex, Years Practiced, and Reported Extraction Frequency

	Recommend extraction			
	<10%	10-25%	25-75%	p-value[1]
Years practiced				0.307
<5	24 (19.7%)	72 (59.0%)	26 (21.3%)	
5-15	27 (25.7%)	64 (61.0%)	14 (13.3%)	
>15	39 (26.9%)	76 (52.4%)	30 (20.7%)	
Sex				0.972
Female	34 (25.4%)	75 (56.0%)	25 (18.7%)	
Male	55 (23.5%)	134 (57.3%)	45 (19.2%)	

Crowding Measurement Error:

Dahlberg's error between the student assessor's T1 and T2 crowding measurements (Table V) was small and revealed an error of 0.27mm and 0.34mm for the linear segments and brass wire methods respectively. On average, the crowding measurements were higher using linear segments than with the brass wire method (mean difference of 0.39-1.37mm, $p<0.005$).

Table V. Student Crowding Measurements

Case	*Mx T1 (mm)	*Mx T2 (mm)	*Mx Mean (mm)	*Mn T1 (mm)	*Mn T2 (mm)	*Mn Mean (mm)
Straight segment						
A	3.9	3.6	3.75	4.1	4.3	4.2
B	7.1	7.4	7.25	10.5	10	10.25
C	3.7	3.3	3.5	8.2	8.2	8.2
D	9	9.1	9.05	7.7	7	7.35
Brass wire						
A	2.7	3.1	2.9	3.3	3.8	3.55
B	5.9	6.7	6.3	8.6	8	8.3
C	3.9	3.3	3.6	7.2	7.3	7.25
D	7.8	7.9	7.85	6.8	6.7	6.75

*Mx=maxillary, Mn=mandibular, T1=timepoint 1, T2-timepoint 2

Crowding Assessment and Treatment Recommendation:

Crowding estimation, severity, and treatment preferences for each case are listed in Tables VI, IX, and X. There was wide variation in estimation for all cases, ranging from 0mm to 15mm for the least crowded arch (Tables VII and VIII). Of the four cases, Case A was perceived to have the least crowding with 5.3mm in the maxilla and 5.1mm in the mandible. Crowding was classified as moderate in both arches with most clinicians preferring non-extraction treatment (95%). Cases B and D were found to have the most crowding (moderate-severe) with 8.4-9.4mm in the maxilla and 8-9.1mm in the mandible. The preferred treatment for these cases was with extractions (77.2-84.3%). Crowding in Case C was estimated to be more than in Case A but less than in Cases B and D. Most clinicians favored non-extraction for this case (72.8%).

For the more crowded cases (B and D), the favored extraction pattern was all first premolars. For Cases A and C, the few who chose extractions showed similar preference for all

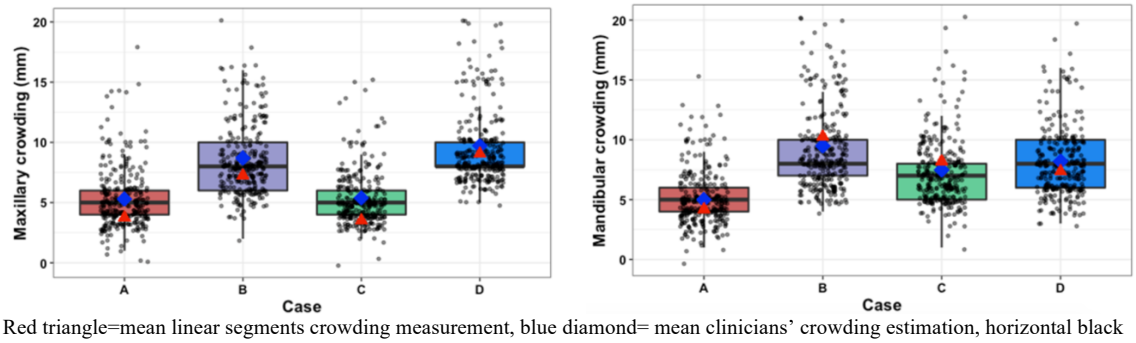
first and second premolars for Case A and for all first and second premolars, as well as one lower incisor for Case C (Appendices IV-VIII). As crowding increased, the preference switched from all second premolars or a lower incisor to first premolars. Examining all cases together, there was a trend to recommend extractions as crowding estimation and severity increased (Table XI). Once crowding reached 8-10mm in either the maxilla or the mandible, most clinicians were recommending extractions (Figure I). Looking at cases individually, a similar crowding threshold for extraction recommendation existed for just the mandible in Cases B and D. When non-extraction treatment was chosen, the majority chose to create space through a combination of expansion, incisor proclination, and interproximal reduction (17.8%). The methods selected for each case are displayed in Appendix IX.

Table VI. Clinician Crowding Estimations

	Case			
Characteristic	**A**	**B**	**C**	**D**
Maxillary crowding (mm)				
Mean (SD)	5.3 (2.6)	8.4 (2.9)	5.3 (2.3)	9.4 (3.1)
Median (IQR)	5.0 (4.0, 6.0)	8.0 (6.0, 10.0)	5.0 (4.0, 6.0)	8.0 (8.0, 10.0)
Range	0.0, 18.0	2.0, 20.0	0.0, 15.0	5.0, 20.0
Mandibular crowding (mm)				
Mean (SD)	5.1 (2.3)	9.1 (3.4)	7.3 (2.8)	8.0 (2.9)
Median (IQR)	5.0 (4.0, 6.0)	8.0 (7.0, 10.0)	7.0 (5.0, 8.0)	8.0 (6.0, 10.0)
Range	0.0, 15.0	4.0, 20.0	1.0, 20.0	3.0, 20.0

Table VII. Maxillary Crowding Estimations vs. Measurements

Table VIII. Mandibular Crowding Estimations vs. Measurements



Red triangle=mean linear segments crowding measurement, blue diamond= mean clinicians' crowding estimation, horizontal black line=median clinicians' crowding estimation

Table IX. Clinician Crowding Severity Classification

	Case			
Characteristic	**A**	**B**	**C**	**D**
Maxillary severity				
None	2 (0.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Mild	101 (36.1%)	6 (2.1%)	119 (42.2%)	0 (0.0%)
Moderate	164 (58.6%)	149 (53.2%)	160 (56.7%)	65 (23.2%)
Severe	13 (4.6%)	125 (44.6%)	3 (1.1%)	215 (76.8%)
Mandibular severity				
None	3 (1.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Mild	111 (39.6%)	0 (0.0%)	24 (8.5%)	15 (5.4%)
Moderate	155 (55.4%)	114 (40.7%)	186 (65.7%)	137 (48.9%)
Severe	11 (3.9%)	166 (59.3%)	73 (25.8%)	128 (45.7%)

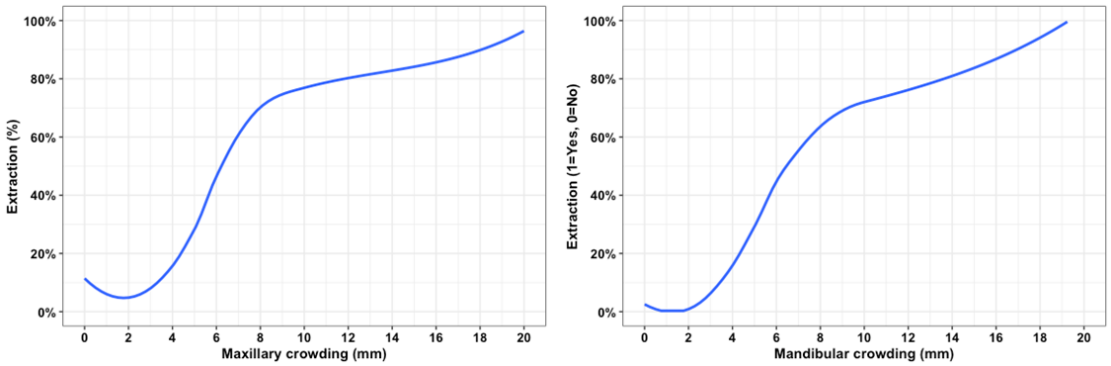
Table X. Clinician Treatment Recommendation

	Case			
Characteristic	**A**	**B**	**C**	**D**
Treatment				
Extractions	14 (5.0%)	217 (77.2%)	77 (27.2%)	236 (84.3%)
Non-extraction	266 (95.0%)	64 (22.8%)	206 (72.8%)	44 (15.7%)

Table XI. Treatment Recommendation by Crowding Severity for All Cases

Characteristic	**Extractions (%)**	**Non-extraction (%)**
Maxillary severity		
None/Mild	9.6%	90.4%
Moderate	39.6%	60.4%
Severe	86.2%	13.8%
Mandibular severity		
None/Mild	3.3%	96.7%
Moderate	38.9%	61.1%
Severe	81.5%	18.5%

Figure I. Crowding Estimation and Case Extraction Recommendation



Blue line= loess smooth

Clinician Accuracy and Precision

Clinicians overestimated crowding in the maxillary arches of all cases and in the mandibular arches of two cases (Table XII). The precision of their estimations was similar for all cases in the maxilla (2.0-2.2mm, p=0.72) and more variable in the mandible (1.7-3.0mm, p<.001). Clinicians were the most precise in Case A, which had the least crowding, and the least precise in Case B, which had the most crowding (Table XIII).

Table XII. Average Difference Between Clinician Crowding Estimations and Measured Values (Accuracy)

	Case			
Characteristic	A	B	C	D
Mx crowding (mm):				
Mean (SD)	1.5 (2.6)	1.2 (2.9)	1.8 (2.3)	0.4 (3.1)
Median (IQR)	1.3 (0.3, 2.2)	0.8 (-1.3, 2.8)	1.5 (0.5, 2.5)	-1.1 (-1.1, 0.9)
Range	-3.8, 14.2	-5.2, 12.8	-3.5, 11.5	-4.1, 10.9
Mn crowding (mm):				
Mean (SD)	0.9 (2.3)	-1.1 (3.4)	-0.9 (2.8)	0.7 (2.9)
Median (IQR)	0.8 (-0.2, 1.8)	-2.2 (-3.2, -0.3)	-1.2 (-3.2, -0.2)	0.7 (-1.3, 2.7)
Range	-4.2, 10.8	-6.2, 9.8	-7.2, 11.8	-4.3, 12.7

Maxillary (Mx) p-value <0.001; p < 0.001 D vs A,C, p = 0.03 B vs C, p = 0.005 B vs D
Mandibular (Mn) p-value < 0.001; p<0.001 A vs B,C, p<0.001 D vs B,C
Positive value=overestimation, negative value=underestimation

Table XIII. Average Absolute Difference Between Clinician Crowding Estimations and Measured Values (Precision)

	Case			
Characteristic	A	B	C	D
Mx crowding (mm):				
Mean (SD)	2.0 (2.2)	2.2 (2.2)	2.0 (2.1)	2.1 (2.2)
Median (IQR)	1.3 (0.8, 2.2)	1.3 (0.8, 2.8)	1.5 (0.5, 2.5)	1.1 (0.9, 2.1)
Range	0.3, 14.2	0.3, 12.8	0.5, 11.5	0.1, 10.9
Mn crowding (mm):				
Mean (SD)	1.7 (1.8)	3.0 (1.9)	2.4 (1.8)	2.2 (2.0)
Median (IQR)	1.2 (0.2, 2.2)	2.8 (1.3, 4.2)	2.2 (1.2, 3.2)	1.7 (0.7, 2.7)
Range	0.2, 10.8	0.3, 9.8	0.2, 11.8	0.3, 12.7

Maxillary (Mx) p-value = 0.72
Mandibular (Mn) p-value < 0.001; p<0.001 A vs B and C, p=0.01 A vs D, p<0.001 B vs C and D

Clinician Demographics and Precision

The associations between clinicians' demographics and precision are displayed in Appendix X. Clinicians that have practiced for less than five years were 0.8mm more precise in their estimations for the maxillary arch of Case C than those with more than 15 years of experience. No differences were detected between experience groups for Cases A, B or D. Regarding extraction prevalence, those who reported having recommended extractions to the fewest patients (<10%) overestimated and were less precise in Cases A and D compared to those who have recommended extractions to 10-25% of patients. Compared to the 10-25% group, maxillary estimations by clinicians from the <10% extraction group were 1mm higher in the maxilla and 0.8mm higher in the mandible for Case A, and 1.1mm higher in the mandible for Case D. By region, clinicians practicing in the Northeast overestimated crowding by the greatest amount in all cases and six of the eight arches (1.2-1.5mm). Precision was similar between sexes and higher for those who reported routinely measuring crowding in cases A and D. In Case A, routine measurers were 0.6 and 0.1mm more precise than visual estimators in the maxilla and mandible, respectively. In Case D, routine measurers were 0.5mm more precise in the mandibular arch.

Clinician Demographics and Extraction Recommendation

The correlations between clinician demographics and extraction recommendation for each case are listed in Appendix XI. Subjects with >15 years of experience were more likely to extract in three of the four cases, with significant differences for Case C. Those who reported recommending extractions to <10% of patients were 1.4-1.6 times less likely to recommend extractions in all cases, with significant differences for Cases B, C, and D. Routine measurers of

crowding on the other hand, were 1.2-2 times more likely to extract in Cases B, C, and D. There was no association between clinician sex or region and case extraction recommendation.

DISCUSSION:

This study aimed to assess the relationship between crowding estimation and extraction decision in Class I crowded cases by clinicians practicing in the United States. Secondarily, the associations between clinician background and estimation precision as well as extraction recommendation were examined. Similar to a 2017 study (Saghafi *et al.*, 2017), the response rate from the AAO was low (7.2%) and most respondents were male with greater than 15 years of experience. Further, it was found that the majority of clinicians reported that they routinely estimated crowding (81.8%), which was expected (Wallis *et al.*, 2014). These clinicians also stated that they have recommended extractions to 10-25% of their patients. Compared to subjects from Facebook®, a greater number of AAO clinicians had been practicing for more than 15 years and were more likely to complete questions for all cases even though fewer total responses were received. While the use of social media increased visibility of the survey to more age groups, survey completion rate was lower and should be considered for future orthodontic surveys.

Comparing crowding measurements using linear segments and brass wire, the values with brass wire were 0.39-1.37 mm lower. While this difference may not be clinically significant, it is consistent with past findings that linear segments underestimate the space available and rounded arch forms overestimate the available space (Adkins *et al.*, 1990, Park *et al.*, 2020). In contrast to Proffit *et al.*, who reports linear segments to be more reliable, our findings indicate both methods have similar reliability (0.27mm vs. 0.34mm difference). Clinicians' average estimations were precise within 1.7-3.0mm of the objective measurements, with vast ranges reported for each case. Crowding estimations for the least crowded arch ranged from 0-18mm, with 9 subjects

reporting there to be 20mm of crowding in at least one arch of one case. Given the small number of clinicians who estimated 20mm, these estimations were not excluded from analyses as their omission would have little impact on the results. The broad range of estimations highlights marked differences in criteria for evaluating crowding and identifies a need for greater standardization of diagnostic guidelines. Wide estimations have also been found in other studies (Brown *et al.*, 1977, Naish *et al.*, 2016, Wallis *et al.*, 2014). Similar to what was reported in Naish *et al.*, 2016, clinicians generally tended to overestimate crowding and were the least precise in the most crowded case. In addition to crowding severity, distribution of crowding may have also affected precision. In the maxillary arch of Case D and the mandibular arch of Case B, estimated crowding was essentially the same but clinicians' estimations were closer to the measured value for Case D. As all of the maxillary crowding in Case D was located around a single blocked out canine tooth, it is likely that the location of the crowding made it easier to approximate. However, a 1mm difference in precision is small and unlikely to be clinically relevant as the majority still agreed in treatment recommendation for the least precisely assessed cases. Similar to the Wallis *et al.* study, the influence of clinicians' years of experience on precision was minimal and did not seem to affect treatment decisions. By region, Northeastern clinicians consistently overestimated crowding in all cases and 6 of 8 arches. This finding could be attributed to regional differences in orthodontic training and diagnosis, but further study is needed. Unexpectedly, clinicians who routinely measured crowding were not more precise overall, though large differences would be unlikely due to the similar amounts of crowding between the cases.

Past literature from 1989 to 2017 examining the relation between clinician experience and extraction tendencies has been controversial, with some studies suggesting a connection

(Baelum *et al.*, 2012, Gentry *et al.*, 2009, Saghafi *et al.*, 2017) and other others finding no association (Wallis *et al.*, 2014, Weintraub *et al.*, 1989). Our recent analysis of clinicians' demographic traits shows no association between years of experience and past extraction prevalence. Of the four cases though, experience was found to play a role in Case C, where crowding was approximated to be between that of Cases A and B. For this case, subjects who had been practicing for more than 15 years were more likely to recommend extraction. If the crowding in Case C can be considered somewhat borderline, this observation draws similarity to the Saghafi *et al.* study which found that clinicians with more than 15 years of experience extracted more in borderline cases. Regarding past extraction prevalence, clinicians who reported recommending extractions at a higher rate in practice were more likely to recommend extractions in the cases, as expected. Less expectedly, routine measurers of crowding were more likely to recommend extractions in three of the four cases even though their estimations were similar to those by clinicians who routinely estimated. Also, Northeastern clinicians were not more likely to recommend extraction in the cases even though they consistently overestimated crowding.

As anticipated, there was greater preference for extraction treatment as crowding in the cases increased. Analysis of all four cases together revealed that most clinicians recommended extractions of four first premolars once crowding was considered severe and reached 8-10mm in the maxilla or the mandible. This crowding amount is similar to that reported by Proffit, who recommends extractions once crowding reaches 10mm. As the extraction of two 7mm premolar teeth in each arch creates more space than what would be needed to alleviate 8-10mm of crowding, some may question the justification of extracting an additional two teeth in the opposing arch that may have even less crowding. It is possible that once maxillary or mandibular crowding totals the width of a single premolar or canine tooth (7mm), most clinicians are starting

to think about extracting a tooth. However, in order to maintain a symmetrical Class I occlusion, extracting contralateral teeth in both the maxilla and mandible is often required. Additionally, any extraction space left over after crowding is alleviated can usually be addressed by slipping posterior anchorage and moving the molars forward. When looking at cases individually, it was interesting to find that a crowding threshold for extractions was only detected for the mandible and not the maxilla in Cases B and D, even though both arches had similar crowding. This observation is similar to that from the Lee *et al*. study which found amount of lower incisor crowding to play a significant role in extraction decision-making. Given the bony limits of the anterior mandible and greater potential for negative periodontal effects with forward tooth movement (Handelman, 1996, Proffit *et al*., 2012), mandibular crowding should be a major consideration in the extraction decision. More studies are needed though to corroborate this relationship.

Though the majority preferred non-extraction treatment for Case C, which had a deeper overbite, there were several clinicians who recommended a lower incisor extraction. As this case had the greatest crowding difference between the maxilla and mandible, future studies might include similar cases with more severe crowding in one arch and minimal to crowding in the other. This may help to identify a potential crowding threshold for extracting in just one arch.

This study has identified an evidence-based numeric range for crowding at which most clinicians recommend extractions in Class I cases. As the extraction decision is multifactorial, clinicians should comprehensively consider all features of a patient's periodontal health, facial profile, incisor inclination, and curve of Spee, in addition to crowding, when making this decision. While it may be argued that most clinicians do not estimate a numeric value for crowding regularly, this range can still be helpful especially when the treatment decision is still

borderline after considering other factors. Diagnostic set-ups can also be informative, as there were a few clinicians who indicated they performed them regularly when assessing crowding.

There were several limitations to this study, the first being the study design. As clinicians would normally be able to manipulate study models from different angles when assessing crowding, the two-dimensional view of the records may have affected their evaluation. Further, manual measurement of all crowding values by the same student assessor places the study at risk of experimenter bias and error. As the small number of cases included do not accurately represent all absolute values for crowding, future studies might include more cases with wider ranges of crowding. Additionally, the crowding values in the cases were fairly low and similar, and so a 1mm difference that was found to be statistically significant may not be clinically significant. One of the study's biggest limitations, however, was case selection. While it would have been more standardized to eliminate confounding variables by selecting patients with straight profiles and normal overbite, the cases chosen were based on availability of Class I patients in treatment at the UW at the time of screening. Additionally, all patients selected for the study were adolescents, which limits the application of the results to this age group. Lastly, many statistical tests were run in this study, which increases the probability of finding a significant difference due to chance when there may be no difference.

CONCLUSIONS:

The present study suggests that on average, crowding estimations by orthodontists in the United States vary widely and are precise within approximately 2mm of manual measurements. Clinicians practicing in the Northeast tended to overestimate crowding while those who routinely measured crowding or reported recommending extraction to >10% of patients were 1.2-2 times more likely to extract in the cases. Most clinicians recommended extractions once crowding was

considered severe and reached 8-10mm in either the maxilla or the mandible. There is some evidence to suggest that the extraction decision is more strongly correlated to mandibular crowding, but further study is needed.

LITERATURE CITED

Adkins, Michael D, et al. “Arch Perimeter Changes on Rapid Palatal Expansion.” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 97, no. 3, 1990, pp. 194–199.

Adobes Martin, Milagros, et al. “Reliability of Tooth Width Measurements Delivered by the Clin-Check Pro 6.0 Software on Digital Casts: A Cross-Sectional Study.” *International Journal of Environmental Research and Public Health*, vol. 19, no. 6, 2022, p. 3581–, https://doi.org/10.3390/ijerph19063581.

Ali, Batool, et al. “Factors Affecting Treatment Decisions for Class I Malocclusions.” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 154, no. 2, 2018, pp. 234–37, https://doi.org/10.1016/j.ajodo.2017.11.035.

Bader, JD, and Shugars, DA. “Variation, Treatment Outcomes, and Practice Guidelines in Dental Practice.” *Journal of Dental Education*, vol. 59, no. 1, 1995, pp. 61–95.

Baelum, V., et al. “Inter-Examiner Variability in Orthodontic Treatment Decisions for Danish Children with ‘Borderline’ Treatment Need.” *European Journal of Orthodontics*, vol. 34, no. 2, 2012, pp. 250–56, https://doi.org/10.1093/ejo/cjq195.

Baumrind, Sheldon, et al. “The Decision to Extract: Part 1—Interclinician Agreement.” *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 109, no. 3, 1996, pp. 297–309.

Brown, W A, et al. “Treatment Planning from Study Models: an Examiner Variability Study.” *The Angle Orthodontist*, vol. 47, no. 2, 1977, pp. 118–122.

Carey, C. W. “Diagnosis and Case Analysis in Orthodontics.” *American Journal of Orthodontics*, vol. 38, no. 3, 1952, pp. 149–61, https://doi.org/10.1016/0002-9416(52)90105-X.

Gentry SI (2009). Extraction decision-making in class I malocclusions: a survey identifying values for definite extraction and non-extraction therapy. MSD Thesis. Saint Louis University: U.S.

Gilmore, Charles A, and Little, Robert M. "Mandibular Incisor Dimensions and Crowding." *American Journal of Orthodontics* 86, no. 6 (1984): 493-502.

Gosney, Mary B. E. “An Investigation into Some of the Factors Influencing the Desire for Orthodontic Treatment.” *British Journal of Orthodontics*, vol. 13, no. 2, 1986, pp. 87–94.

Handelman, C S. “The Anterior Alveolus: Its Importance in Limiting Orthodontic

Treatment and Its Influence on the Occurrence of Iatrogenic Sequelae." *The Angle Orthodontist*, vol. 66, no. 2, 1996, pp. 95–109; discussion 109–10.

Hsieh, FY, Bloch, DA, and Larsen, MD. A simple method for sample size calculation for linear and logistic regression. Statistics in Medicine. 1998; 17:1623-1634.

Lee, R, et al. "Consistency of Orthodontic Treatment Planning Decisions." *Clinical Orthodontics and Research*, vol. 2, no. 2, 1999, pp. 79–84.

Luke, L S, et al. "Consistency of Patient Classification in Orthodontic Diagnosis and Treatment Planning." *The Angle Orthodontist*, vol. 68, no. 6, 1998, pp. 513–520.

Lundström, Anders. "The Significance of Early Loss of Deciduous Teeth in the Etiology of Malocclusion." *American Journal of Orthodontics*, vol. 41, no. 11, 1955, pp. 819–26, https://doi.org/10.1016/0002-9416(55)90187-1.

Moorrees, C F, and Reed, R B. "Biometrics of Crowding and Spacing of the Teeth in the Mandible." *American Journal of Physical Anthropology*, vol. 12, no. 1, 1954, pp. 77–88.

Moyers, Robert E. *Handbook of Orthodontics for the Student and General Practitioner*. 3d ed., Year Book Medical Publishers, 1973.

Naidu, Devan, and Terrence J. Freer. "Validity, Reliability, and Reproducibility of the iOC Intraoral Scanner: A Comparison of Tooth Widths and Bolton Ratios." *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 144, no. 2, 2013, pp. 304–10, https://doi.org/10.1016/j.ajodo.2013.04.011.

Naish, Hywel, et al. "Does a True Knowledge of Dental Crowding Affect Orthodontic Treatment Decisions?" *European Journal of Orthodontics*, vol. 38, no. 1, 2016, pp. 66–70.

Nance, Hays N. "The Limitations of Orthodontic Treatment. II. Diagnosis and Treatment in the Permanent Dentition." *American Journal of Orthodontics and Oral Surgery*, vol. 33, no. 5, 1947, pp. 253–301, https://doi.org/10.1016/0096-6347(47)90123-3.

Oosterink, F.M.D, et al. "What Are People Afraid of during Dental Treatment? Anxiety-Provoking Capacity of 67 Stimuli Characteristic of the Dental Setting." *European Journal of Oral Sciences*, vol. 116, no. 1, 2008, pp. 44–51.

Park, Seo-Hyun, et al. "Evaluation of the Reliability, Reproducibility and Validity of Digital Orthodontic Measurements Based on Various Digital Models among Young Patients." *Journal of Clinical Medicine*, vol. 9, no. 9, 2020, p. 2728.

Proffit, William R., et al. *Contemporary Orthodontics*. 5th ed., Mosby Elsevier, 2012.

Ribarevski, R., et al. "Consistency of Orthodontic Extraction Decisions." *European*

Journal of Orthodontics, vol. 18, no. 1, 1996, pp. 77–80, https://doi.org/10.1093/ejo/18.1.77.

Saghafi, Niousha, et al. "Influence of Clinicians' Experience and Gender on Extraction Decision in Orthodontics." *The Angle Orthodontist*, vol. 87, no. 5, 2017, pp. 641–650.

Santoro, Margherita, et al. "Comparison of Measurements Made on Digital and Plaster Models." *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 124, no. 1, 2003, pp. 101–05, https://doi.org/10.1016/S0889-5406(03)00152-5.

Shailendran, Arun, et al. "Accuracy and Reliability of Tooth Widths and Bolton Ratios Measured by ClinCheck Pro." *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 161, no. 1, 2022, pp. 65–73, https://doi.org/10.1016/j.ajodo.2020.06.048.

Van der Linden, Frans P.G.M. "Theoretical and Practical Aspects of Crowding in the Human Dentition." *The Journal of the American Dental Association (1939)*, vol. 89, no. 1, 1974, pp. 139–153.

Wallis, Colin, et al. "How Good Are We at Estimating Crowding and How Does It Affect Our Treatment Decisions?" *European Journal of Orthodontics*, vol. 36, no. 4, 2014, pp. 465–470.

Weintraub, Jane A, et al. "The Prevalence of Orthodontic Extractions." *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 96, no. 6, 1989, pp. 462–466.

Appendix I. Survey Questions

Demographic Questions:

1) For how many years have you practiced orthodontics?
    a. 0 (resident or grad student)
    b. <5
    c. 5-15
    d. >15

2) What is your sex?
    a. Male
    b. Female
    c. Other
    d. Prefer not to say

3) To which AAO regional constituent group do you belong?
    a. Great Lakes Association of Orthodontists
    b. Middle Atlantic Society of Orthodontists
    c. Midwestern Society of Orthodontists
    d. Northeastern Society of Orthodontists
    e. Pacific Coast Society of Orthodontists
    f. Rocky Mountain Society of Orthodontists
    g. Southern Association of Orthodontists
    h. Southwestern Society of Orthodontists
    i. Other

4) Over the past 10 years, in what percentage of your patients do you estimate you have recommended an extraction treatment plan?
    a. <10%
    b. 10-25%
    c. 25-50%
    d. 50-75%
    e. >75%

5) How do you routinely assess crowding in your practice?
    a. Visual estimation (contact overlap)
    b. Measurement on study models
    c. Other (please specify)

Case Questions:

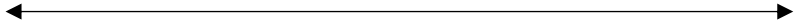
You will be shown 4 cases (A-D), randomized in the order that they are presented. Please use the records below to answer the following questions for Case A. Patients have all permanent dentition present, class I molar relationships, and healthy periodontium.

6) Please rate the severity of maxillary crowding (mm):
*For scale, green arrow next to scanned models=5mm

0 mm 20 mm

7) How would you describe the severity of maxillary crowding?
   a. None
   b. Mild
   c. Moderate
   d. Severe
   e. Not sure

8) Please rate the severity of mandibular crowding (mm):



0 mm 20 mm

9) How would you describe the severity of mandibular crowding?
   a. None
   b. Mild
   c. Moderate
   d. Severe
   e. Not sure

10) Case A is best treated by:
   a. Extractions, other than third molars
   b. Non-extraction

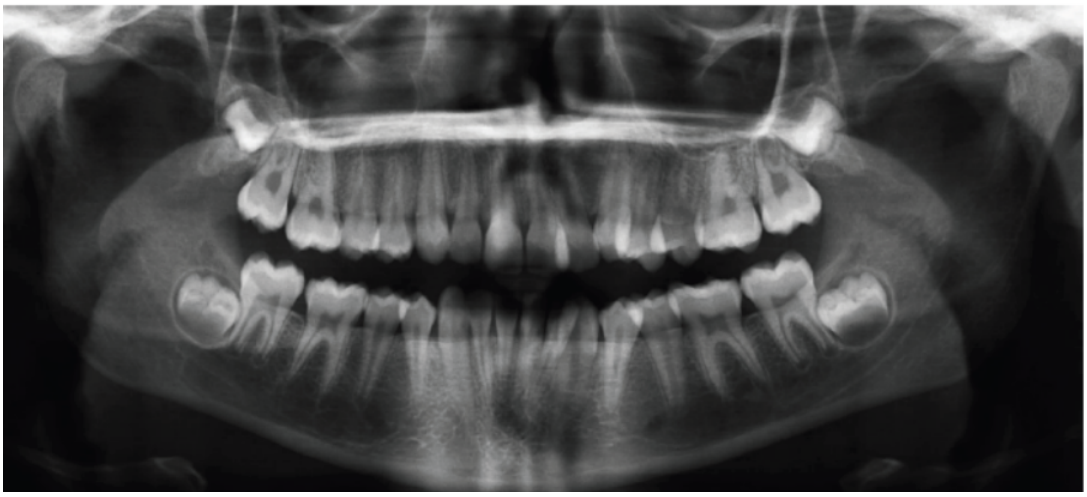
11) Skip this question if you selected "extractions" above. If you selected "non-extraction", how will you create space to level/align the teeth (select all that apply)?
   a. Expansion
   b. Proclination of incisors
   c. Leveling curve of Spee
   d. Leveling curve of Wilson
   e. Distalization of molars
   f. Interproximal reduction
   g. Other (please specify)

12) Skip this question if you selected "non-extraction". If you selected "extractions", please specify the extraction pattern you prefer (select all that apply)?
   a. UR4 (#5)
   b. UR5 (#4)
   c. UL4 (#12)
   d. UL5 (#13)
   e. LL4 (#21)
   f. LL5 (#20)
   g. LR4 (#28)

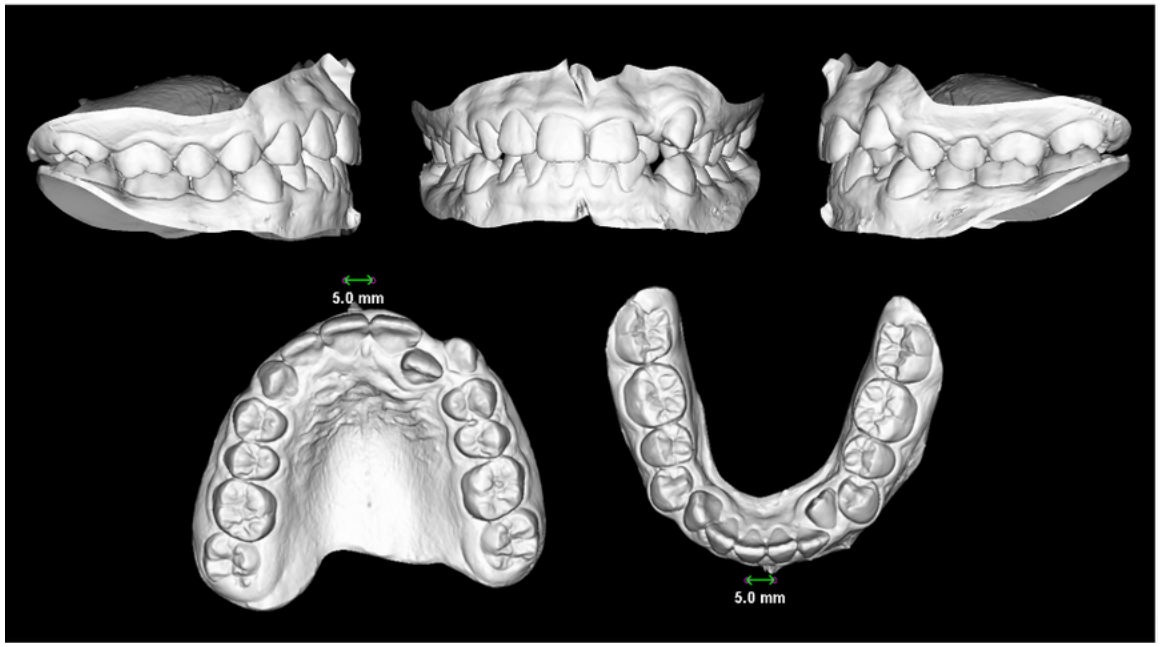
h. LR5 (#29)
i. One lower incisor
j. Other (please specify)

Case Records:

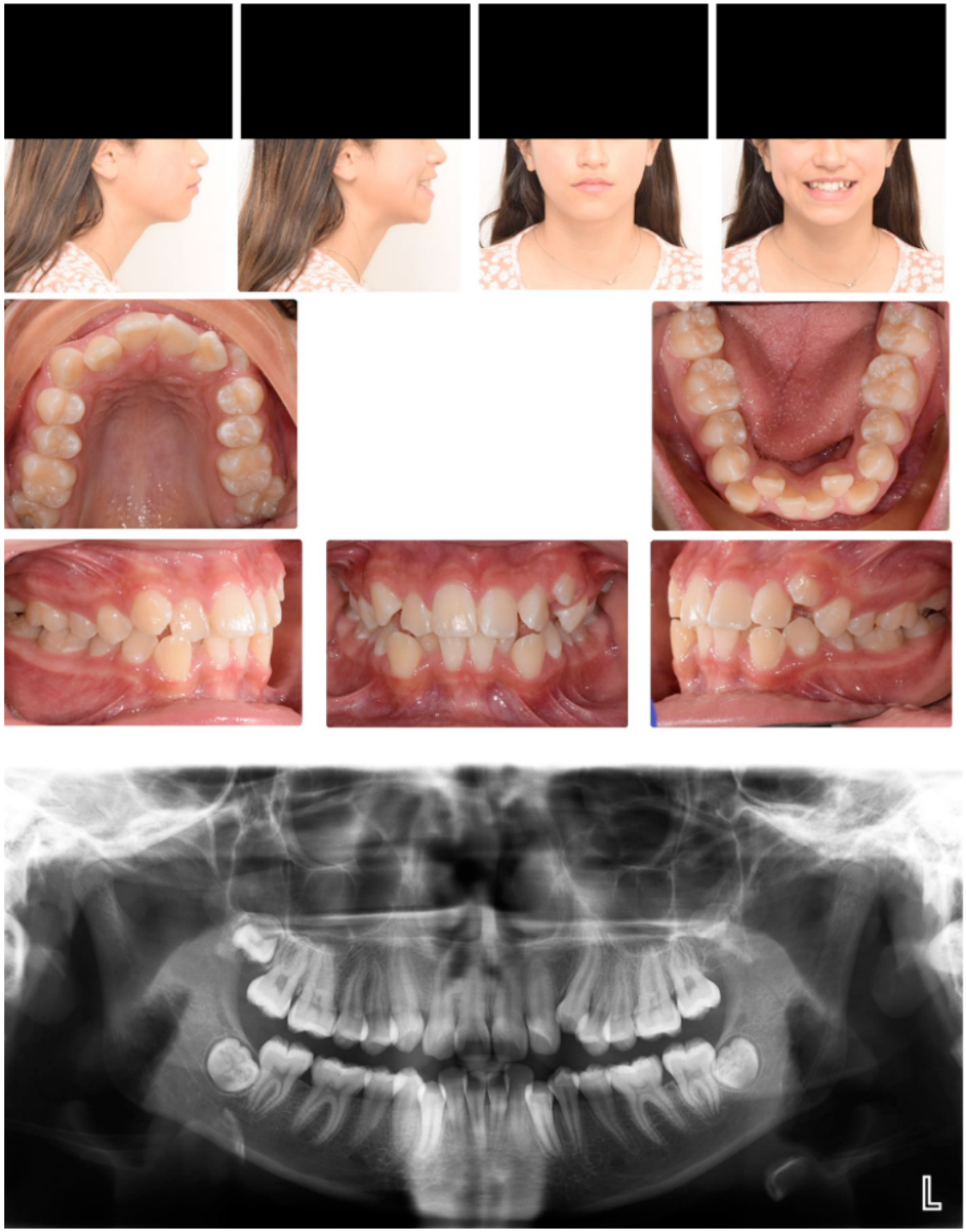
Case A



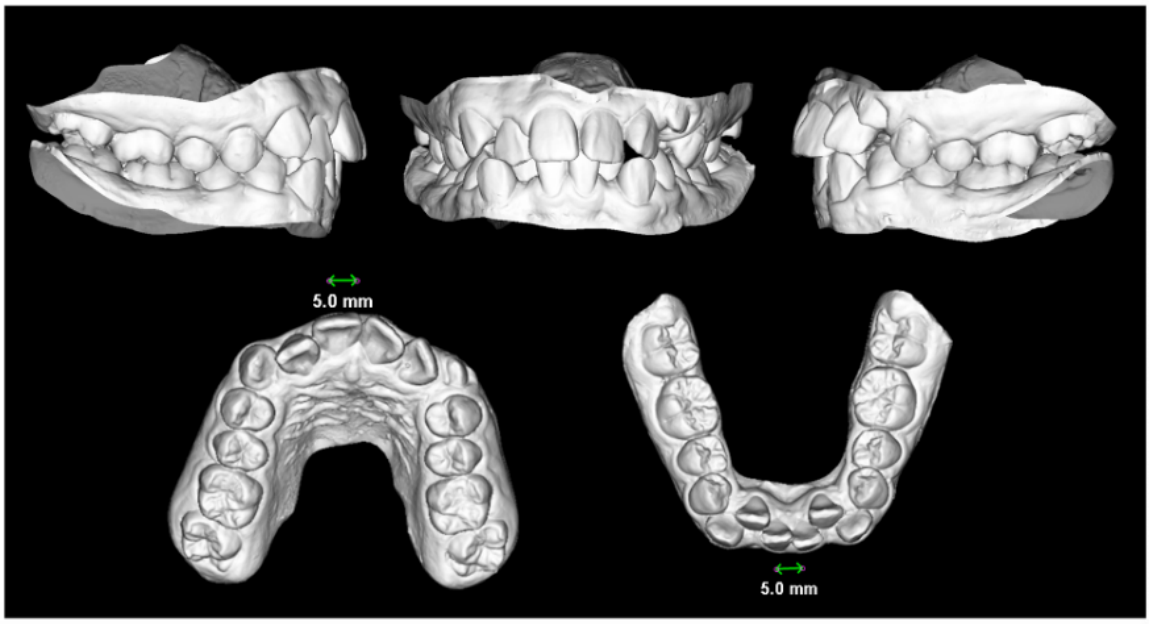
Group/Measurement	Value	Norm	Std Dev	Dev Norm
Maxilla to Cranium				
SNA (°)	81.1	82.0	3.5	-0.3
A - N-Perp (mm)	-1.1	0.0	3.1	-0.3
Mandible to Cranium				
SNB (°)	80.4	80.9	3.4	-0.2
Pog - N-Perp (mm)	-1.3	-4.0	5.3	0.5
A-P relationship				
ANB (°)	0.7	1.6	1.5	-0.6
Wits Appraisal (mm)	-2.7	-1.0	1.0	-1.7 *
Harvold (CoPog)-(CoANS)	12.8	20.0	3.0	-2.4 **
Mx Unit Length (Co-ANS)	48.2	90.0	5.0	-8.4 ******
Md Unit Length (Co-Pog)	61.1	113.0	8.0	-6.5 ******
Inter-incisal relationship				
Interincisal Angle (U1-L1) (°)	142.7	130.0	6.0	2.1 **
Overjet (mm)	2.2	2.5	2.5	-0.1
Overbite (mm)	1.4	2.5	2.0	-0.5
Upper incisors				
U1 - NA (°)	20.2	22.8	5.7	-0.4
U1 - NA (mm)	2.2	4.3	2.7	-0.8
U1 - Palatal Plane (°)	111.1	112.0	6.0	-0.1
Lower incisors				
L1 to Mn Plane (IMPA) (°)	88.8	95.0	7.0	-0.9
L1 - NB (°)	16.4	25.3	6.0	-1.5 *
L1 - NB (mm)	0.8	4.0	1.8	-1.8 *
Vertical Measures				
Mandibular Plane - FH (°)	20.3	25.0	4.5	-1.0 *
Mandibular Plane - SN (°)	27.2	33.0	6.0	-1.0 *
Y-Axis -- Downs (SGn-FH) (°)	58.0	60.6	3.4	-0.8
Y-Axis (SGn-SN) (°)	64.8	67.0	5.5	-0.4
Nasal Height (%)	47.2	43.0	100.0	0.0
Lower Face Height (ANS-Gn) (mm)	32.1	65.0	4.5	-7.3 ******
Palatal plane - FH (°)	-3.0	-1.8	3.9	-0.3
Occlusal Plane - FH (°)	13.2	9.2	5.0	0.8
Chin Button				
Pog - NB (mm)	1.4	1.7	1.7	-0.2
Holdaway Ratio (%)	1.7	1.0	0.5	1.3 *
Facial Pattern				
Upper Lip to E-Plane (mm)	-1.5	-3.8	2.0	1.1 *
Lower Lip to E-Plane (mm)	-1.0	-2.0	2.0	0.5
Zero Meridian (FH-soft tissue NPo) (°)	91.0	91.0	7.0	+0.0
Facial Angle (FH-NPo) (°)	88.6	87.5	3.0	0.4
A-B to Facial Plane (°)	-2.6	-5.2	3.0	0.9
Angle of Convexity (N-A-Pog) (°)	-1.5	7.6	3.0	-3.1 ***



Case B

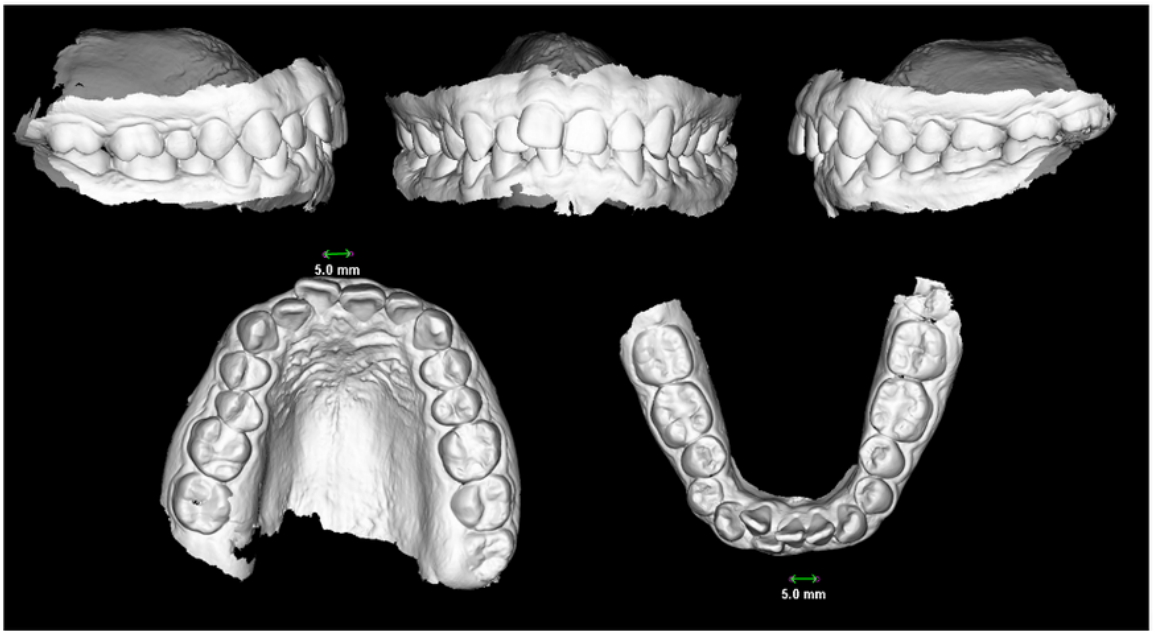


Group/Measurement	Value	Norm	Std Dev	Dev Norm
Maxilla to Cranium				
SNA (°)	81.2	82.0	3.5	-0.2
A - N-Perp (mm)	0.3	3.0	3.1	-0.9
Mandible to Cranium				
SNB (°)	80.2	80.9	3.4	-0.2
Pog - N-Perp (mm)	-2.9	-4.0	5.3	0.2
A-P relationship				
ANB (°)	1.0	1.6	1.5	-0.4
Wits Appraisal (mm)	-1.4	-1.0	1.0	-0.4
Harvold (CoPog)-(CoANS)	21.7	20.0	3.0	0.6
Mx Unit Length (Co-ANS)	79.7	90.0	5.0	-2.1 **
Md Unit Length (Co-Pog)	101.4	113.0	8.0	-1.5 *
Inter-incisal relationship				
Interincisal Angle (U1-L1) (°)	122.6	124.0	6.0	-0.2
Overjet (mm)	3.6	3.4	2.5	0.1
Overbite (mm)	3.1	2.8	2.0	0.1
Upper incisors				
U1 - NA (°)	28.4	22.8	5.7	1.0 *
U1 - NA (mm)	8.1	4.3	2.7	1.4 *
U1 - Palatal Plane (°)	118.9	112.0	6.0	1.2 *
Lower incisors				
L1 to Mn Plane (IMPA) (°)	94.6	95.0	7.0	-0.1
L1 - NB (°)	28.0	25.3	6.0	0.4
L1 - NB (mm)	6.1	4.0	1.8	1.1 *
Vertical Measures				
Mandibular Plane - FH (°)	24.0	24.5	4.5	-0.1
Mandibular Plane - SN (°)	33.2	33.0	6.0	0.0
Y-Axis -- Downs (SGn-FH) (°)	58.6	62.0	3.4	-1.0 *
Y-Axis (SGn-SN) (°)	67.6	67.0	5.5	0.1
Nasal Height (%)	46.6	43.0	100.0	0.0
Lower Face Height (ANS-Gn) (mm)	61.1	65.0	4.5	-0.9
Palatal plane - FH (°)	-0.2	-1.0	3.7	0.2
Occlusal Plane - FH (°)	4.2	10.0	5.0	-1.2 *
Chin Button				
Pog - NB (mm)	-1.7	2.0	1.7	-2.2 **
Holdaway Ratio (%)	-0.3	0.9	0.5	-2.3 **
Facial Pattern				
Upper Lip to E-Plane (mm)	0.8	-3.3	2.0	2.1 **
Lower Lip to E-Plane (mm)	3.1	-2.0	2.0	2.5 **
Zero Meridian (FH-soft tissue NPo) (°)	91.4	96.1	7.0	-0.7
Facial Angle (FH-NPo) (°)	88.4	90.0	3.0	-0.5
A-B to Facial Plane (°)	-0.5	-6.3	3.0	1.9 *
Angle of Convexity (N-A-Pog) (°)	4.1	7.0	3.0	-1.0 *

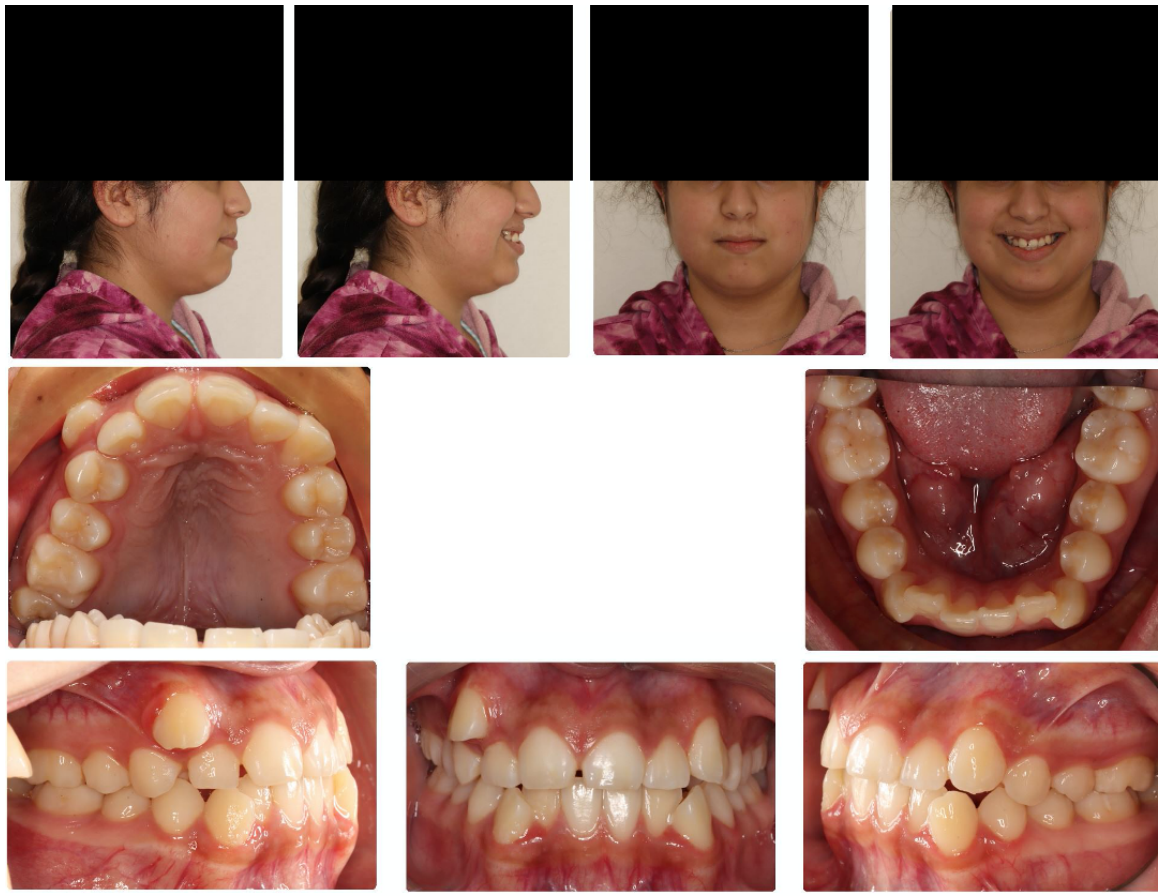


Case C

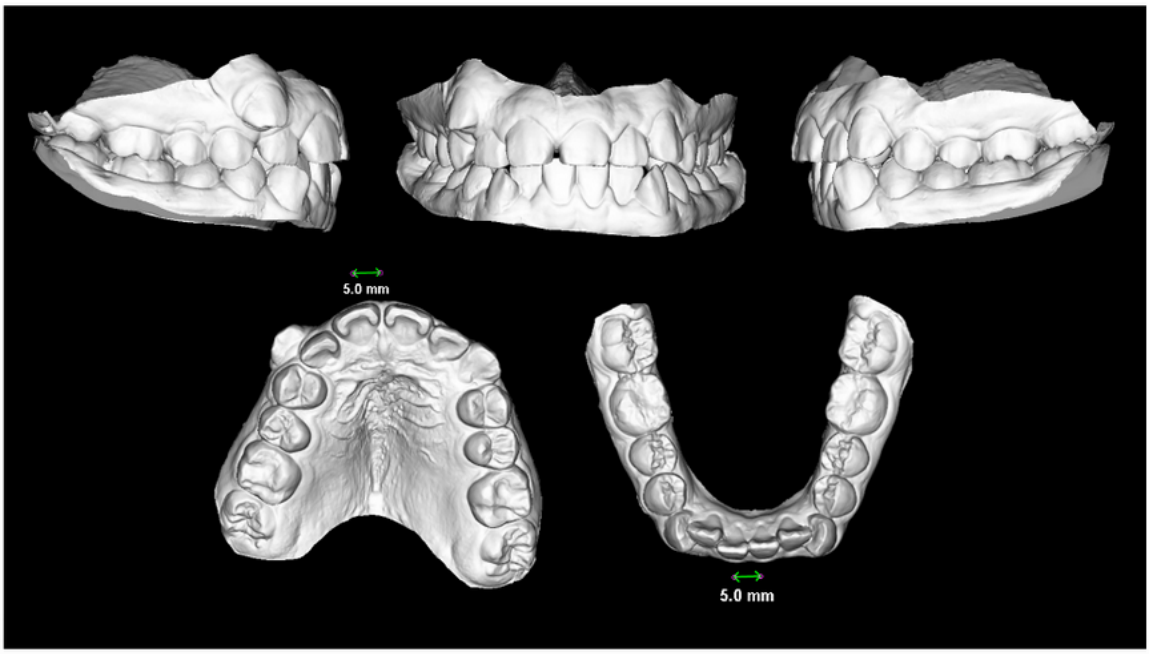
Group/Measurement	Value	Norm	Std Dev	Dev Norm
Maxilla to Cranium				
SNA (°)	80.8	82.0	3.5	-0.3
A - N-Perp (mm)	-2.1	3.4	3.1	-1.8 *
Mandible to Cranium				
SNB (°)	78.6	80.9	3.4	-0.7
Pog - N-Perp (mm)	-8.7	-4.0	5.3	-0.9
A-P relationship				
ANB (°)	2.2	1.6	1.5	0.4
Wits Appraisal (mm)	-0.7	-1.0	1.0	0.3
Harvold (CoPog)-(CoANS)	30.0	20.0	3.0	3.3 ***
Mx Unit Length (Co-ANS)	86.9	90.0	5.0	-0.6
Md Unit Length (Co-Pog)	116.9	113.0	8.0	0.5
Inter-incisal relationship				
Interincisal Angle (U1-L1) (°)	133.9	124.0	6.0	1.7 *
Overjet (mm)	4.3	3.4	2.5	0.4
Overbite (mm)	5.1	2.8	2.0	1.1 *
Upper incisors				
U1 - NA (°)	20.3	22.8	5.7	-0.4
U1 - NA (mm)	8.1	4.3	2.7	1.4 *
U1 - Palatal Plane (°)	110.9	112.0	6.0	-0.2
Lower incisors				
L1 to Mn Plane (IMPA) (°)	94.2	95.0	7.0	-0.1
L1 - NB (°)	23.6	25.3	6.0	-0.3
L1 - NB (mm)	7.2	4.0	1.8	1.8 *
Vertical Measures				
Mandibular Plane - FH (°)	23.9	23.1	4.5	0.2
Mandibular Plane - SN (°)	30.8	33.0	6.0	-0.4
Y-Axis -- Downs (SGn-FH) (°)	63.7	61.6	3.4	0.6
Y-Axis (SGn-SN) (°)	70.6	67.0	5.5	0.7
Nasal Height (%)	40.1	43.0	100.0	-0.0
Lower Face Height (ANS-Gn) (mm)	79.5	60.0	4.5	4.3 ****
Palatal plane - FH (°)	-2.8	-3.2	5.8	0.1
Occlusal Plane - FH (°)	7.4	7.9	5.0	-0.1
Chin Button				
Pog - NB (mm)	0.5	2.9	1.7	-1.4 *
Holdaway Ratio (%)	0.1	0.9	0.5	-1.6 *
Facial Pattern				
Upper Lip to E-Plane (mm)	-0.7	-4.8	2.0	2.0 **
Lower Lip to E-Plane (mm)	1.2	-2.0	2.0	1.6 *
Zero Meridian (FH-soft tissue NPo) (°)	89.3	96.9	7.0	-1.1 *
Facial Angle (FH-NPo) (°)	85.8	91.4	3.0	-1.9 *
A-B to Facial Plane (°)	-2.5	-5.2	3.0	0.9
Angle of Convexity (N-A-Pog) (°)	3.7	4.0	3.0	-0.1



Case D



Group/Measurement	Value	Norm	Std Dev	Dev Norm
Maxilla to Cranium				
SNA (°)	81.1	82.0	3.5	-0.3
A - N-Perp (mm)	-2.8	2.9	3.1	-1.8 *
Mandible to Cranium				
SNB (°)	80.2	80.9	3.4	-0.2
Pog - N-Perp (mm)	-5.8	-4.0	5.3	-0.3
A-P relationship				
ANB (°)	0.8	1.6	1.5	-0.5
Wits Appraisal (mm)	-3.2	-1.0	1.0	-2.2 **
Harvold (CoPog)-(CoANS)	24.1	20.0	3.0	1.4 *
Mx Unit Length (Co-ANS)	80.7	90.0	5.0	-1.9 *
Md Unit Length (Co-Pog)	104.8	113.0	8.0	-1.0 *
Inter-incisal relationship				
Interincisal Angle (U1-L1) (°)	134.1	124.0	6.0	1.7 *
Overjet (mm)	3.0	3.4	2.5	-0.1
Overbite (mm)	1.9	2.8	2.0	-0.5
Upper incisors				
U1 - NA (°)	22.5	22.8	5.7	-0.1
U1 - NA (mm)	6.3	4.8	2.7	0.7
U1 - Palatal Plane (°)	108.4	112.0	6.0	-0.6
Lower incisors				
L1 to Mn Plane (IMPA) (°)	91.3	95.0	7.0	-0.5
L1 - NB (°)	22.6	25.3	6.0	-0.5
L1 - NB (mm)	4.6	4.0	1.8	0.3
Vertical Measures				
Mandibular Plane - FH (°)	25.0	24.8	4.5	0.0
Mandibular Plane - SN (°)	31.0	33.0	6.0	-0.3
Y-Axis -- Downs (SGn-FH) (°)	64.0	62.1	3.4	0.6
Y-Axis (SGn-SN) (°)	70.1	67.0	5.5	0.6
Nasal Height (%)	44.3	43.0	100.0	0.0
Lower Face Height (ANS-Gn) (mm)	66.7	65.0	4.5	0.4
Palatal plane - FH (°)	1.2	-1.5	3.9	0.7
Occlusal Plane - FH (°)	9.2	10.5	5.0	-0.3
Chin Button				
Pog - NB (mm)	1.1	1.8	1.7	-0.5
Holdaway Ratio (%)	0.2	0.9	0.5	-1.2 *
Facial Pattern				
Upper Lip to E-Plane (mm)	-6.0	-3.0	2.0	-1.5 *
Lower Lip to E-Plane (mm)	-2.3	-2.0	2.0	-0.1
Zero Meridian (FH-soft tissue NPo) (°)	91.4	95.9	7.0	-0.7
Facial Angle (FH-NPo) (°)	86.8	89.7	3.0	-0.9
A-B to Facial Plane (°)	-1.7	-6.5	3.0	1.6 *
Angle of Convexity (N-A-Pog) (°)	0.6	7.6	3.0	-2.4 **



Appendix II. AAO Demographics: Years Practiced, Sex, Region, Extraction Frequency, and Crowding Assessment Method

Characteristic	Overall, N = 138
Years practiced	
0	11 (8.0%)
<5	15 (10.9%)
5-15	29 (21.0%)
>15	83 (60.1%)
Sex	
Female	33 (23.9%)
Male	104 (75.4%)
Prefer not to say	1 (0.7%)
AAO regional group	
Midwest	29 (21.0%)
Northeast	24 (17.4%)
West	41 (29.7%)
South	43 (31.2%)
Other	1 (0.7%)
Recommend extraction	
<10%	31 (22.5%)
10-25%	82 (59.4%)
25-75%	25 (18.1%)
Crowding assessment	
Visual estimation only	96 (69.6%)
Measurement on study models	27 (19.6%)
Visual estimation and other methods	15 (10.9%)

Appendix III. Facebook® Demographics: Years Practiced, Sex, Region, Extraction Frequency, and Crowding Assessment Method

Characteristic	Overall, N = 235
Years practiced	
0	25 (10.6%)
<5	72 (30.6%)
5-15	76 (32.3%)
>15	62 (26.4%)
Sex	
Female	101 (43.0%)
Male	131 (55.7%)
Prefer not to say	3 (1.3%)
AAO regional group	
Midwest	45 (19.1%)
Northeast	42 (17.9%)
West	71 (30.2%)
South	61 (26.0%)
Other	16 (6.8%)
Recommend extraction	
<10%	59 (25.2%)
10-25%	130 (55.6%)
25-75%	45 (19.2%)
Unknown	1
Crowding assessment	
Visual estimation only	209 (88.9%)
Measurement on study models	16 (6.8%)
Visual estimation and other methods	10 (4.3%)

Appendix IV. Extraction Preferences

Characteristic	A, N = 297[1]	B, N = 297[1]	C, N = 297[1]	D, N = 297[1]
Extraction pattern				
Lower extractions only	1 (7.1%)	6 (2.8%)	23 (29.5%)	1 (0.4%)
Upper extractions only	1 (7.1%)	0 (0.0%)	0 (0.0%)	1 (0.4%)
Upper and lower extractions	12 (85.7%)	208 (97.2%)	55 (70.5%)	232 (99.1%)

Appendix V. Case A Extraction Patterns

Characteristic	N = 297[1]
Extraction pattern	
One lower incisor	1 (7.1%)
UL4 (#12)	1 (7.1%)
UL5 (#13), LL5 (#20)	1 (7.1%)
UR4 (#5), UL4 (#12), LL4 (#21), LR5 (#29)	1 (7.1%)
UR4 (#5), UL4 (#12), LL4 (#21), LR4 (#28)	5 (35.7%)
UR5 (#4), UL5 (#13), LL5 (#20), LR5 (#29)	5 (35.7%)
Unknown	283

Appendix VI. Case B Extraction Patterns

Characteristic	N = 297[1]
Extraction pattern	
LL4 (#21), LR4 (#28)	1 (0.5%)
One lower incisor	5 (2.3%)
UR4 (#5), LL4 (#21), LR4 (#28)	1 (0.5%)
UR4 (#5), UL4 (#12), LL4 (#21), LR4 (#28)	139 (65.0%)
UR4 (#5), UL4 (#12), LL5 (#20), LR5 (#29)	2 (0.9%)
UR4 (#5), UL4 (#12), Other: LL3 (#22), LR3 (#27)	1 (0.5%)
UR4 (#5), UL5 (#13), LL4 (#21), LR4 (#28)	1 (0.5%)
UR4 (#5), UL5 (#13), LL5 (#20), LR5 (#29)	1 (0.5%)
UR4 (#5), UL5 (#13), One lower incisor	1 (0.5%)
UR5 (#4), UL4 (#12), LL4 (#21), LR4 (#28)	4 (1.9%)
UR5 (#4), UL5 (#13), LL4 (#21), LR4 (#28)	2 (0.9%)
UR5 (#4), UL5 (#13), LL4 (#21), LR4 (#28)	10 (4.7%)
UR5 (#4), UL5 (#13), LL5 (#20), LR5 (#29)	46 (21.5%)
Unknown	83

Appendix VII. Case C Extraction Patterns

Characteristic	N = 297[1]
Extraction pattern	
One lower incisor	20 (25.6%
One lower incisor, Other (Upper IPR)	1 (1.3%)
One lower incisor, Other: IPR upper arch	1 (1.3%)
Other: set up to decide 5s or one lower incisor	1 (1.3%)
UR4 (#5), One lower incisor	1 (1.3%)
UR4 (#5), UL4 (#12), LL4 (#21), LR4 (#28)	23 (29.5%
UR4 (#5), UL4 (#12), LL5 (#20), LR5 (#29)	1 (1.3%)
UR5 (#4), UL5 (#13), LL4 (#21), LR4 (#28)	2 (2.6%)
UR5 (#4), UL5 (#13), LL5 (#20), LR5 (#29)	28 (35.9%
Unknown	219

Appendix VIII. Case D Extraction Patterns

Characteristic	N = 297[1]
Extraction pattern	
One lower incisor	1 (0.4%)
UL4 (#12), LL4 (#21)	1 (0.4%)
UL4 (#12), One lower incisor	2 (0.9%)
UL4 (#12), Other: UR3 (#6), LL3 (#22), LR3 (#27)	1 (0.4%)
UR3 (#6), UL4 (#12), LL4 (#21), LR4 (#28)	1 (0.4%)
UR4 (#5)	1 (0.4%)
UR4 (#5), UL4 (#12)	1 (0.4%)
UR4 (#5), UL4 (#12), LL4 (#21), LR4 (#28)	169 (72.2%)
UR4 (#5), UL4 (#12), LL4 (#21), LR4 (#28).	1 (0.4%)
UR4 (#5), UL4 (#12), LL5 (#20), LR5 (#29)	7 (3.0%)
UR4 (#5), UL5 (#13), LL4 (#21), LR4 (#28)	1 (0.4%)
UR4 (#5), UL5 (#13), LL5 (#20), LR4 (#28)	1 (0.4%)
UR4 (#5), UL5 (#13), LL5 (#20), LR5 (#29)	3 (1.3%)
UR5 (#4), UL4 (#12), LL4 (#21), LR4 (#28)	15 (6.4%)
UR5 (#4), UL4 (#12), LL4 (#21), LR5 (#29)	3 (1.3%)
UR5 (#4), UL4 (#12), LL5 (#20), LR5 (#29)	3 (1.3%)
UR5 (#4), UL5 (#13), LL4 (#21), LR4 (#28)	9 (3.8%)
UR5 (#4), UL5 (#13), LL5 (#20), LR5 (#29)	14 (6.0%)
Unknown	63

Appendix IX. Non-extraction Methods

	Case			
Characteristic	**A**	**B**	**C**	**D**
Expansion	196 (73.7%)	60 (93.8%)	133 (64.6%)	38 (86.4%)
Proclination of incisors	221 (83.1%)	33 (51.6%)	158 (76.7%)	29 (65.9%)
Leveling curve of Spee	75 (28.2%)	18 (28.1%)	67 (32.5%)	10 (22.7%)
Leveling curve of Wilson	86 (32.3%)	30 (46.9%)	62 (30.1%)	14 (31.8%)
Distalization of molars	4 (1.5%)	1 (1.6%)	1 (0.5%)	4 (9.1%)
Interproximal reduction	149 (56.0%)	43 (67.2%)	158 (76.7%)	25 (56.8%)
Other	3 (1.1%)	2 (3.1%)	2 (1.0%)	1 (2.3%)

Appendix X. Association Between Clinician Demographic (Experience, Extraction Prevalence, Crowding Assessment Method, Region, Sex) and Precision

Experience (Years) vs. Precision

Case A

Characteristic	<5 N = 89	5-15 N = 81	>15 N = 127
Mx crowding (mm):			
Mean (SD)	1.9 (1.9)	2.0 (2.3)	2.2 (2.4)
Mn crowding (mm):			
Mean (SD)	1.6 (2.0)	1.5 (1.7)	1.9 (1.8)

Maxillary (Mx) p-value = 0.64
Mandibular (Mx) p-value = 0.40

Case B

Characteristic	<5 N = 89	5-15 N = 81	>15 N = 127
Mx crowding (mm):			
Mean (SD)	2.0 (1.8)	2.1 (2.4)	2.4 (2.4)
Mn crowding (mm):			
Mean (SD)	2.9 (1.9)	2.7 (1.9)	3.2 (2.0)

Maxillary (Mx) p-value = 0.31
Mandibular (Mx) p-value = 0.20

Case C

Characteristic	<5 N = 89	5-15 N = 81	>15 N = 127
Mx crowding (mm):			
Mean (SD)	1.6 (1.6)	1.9 (2.1)	2.4 (2.3)
Mn crowding (mm):			
Mean (SD)	2.3 (1.3)	2.2 (1.7)	2.5 (2.1)

Maxillary (Mx) p-value = 0.01; pairwise testing, p = 0.008 <5 vs >15
Mandibular (Mx) p-value = 0.53

Case D

Characteristic	<5 N = 89	5-15 N = 81	>15 N = 127
Mx crowding (mm):			
Mean (SD)	1.9 (2.3)	1.9 (1.9)	2.4 (2.4)
Mn crowding (mm):			
Mean (SD)	1.8 (1.9)	2.1 (1.9)	2.4 (2.2)

Maxillary (Mx) p-value = 0.18
Mandibular (Mx) p-value = 0.13

Extraction Prevalence vs. Precision

Case A

Characteristic	<10% N = 68	10-25% N = 172	25-75% N = 57
Mx crowding (mm):			
Mean (SD)	2.7 (2.7)	1.7 (1.8)	2.2 (2.7)
Mn crowding (mm):			
Mean (SD)	2.3 (2.3)	1.5 (1.6)	1.8 (1.9)

Maxillary (Mx) p-value = 0.02; pairwise testing, p = 0.02 <10 vs 10-25%
Mandibular (Mn) p-value = 0.03; pairwise testing, p = 0.04 <10 vs 10-25%

Case B

Characteristic	<10% N = 68	10-25% N = 172	25-75% N = 57
Mx crowding (mm):			
Mean (SD)	2.4 (2.2)	2.0 (2.0)	2.6 (2.7)
Mn crowding (mm):			
Mean (SD)	3.0 (1.8)	3.1 (1.9)	2.9 (2.2)

Maxillary (Mx) p-value = 0.15
Mandibular (Mn) p-value = 0.84

Case C

Characteristic	<10% N = 68	10-25% N = 172	25-75% N = 57
Mx crowding (mm):			
Mean (SD)	2.4 (2.2)	1.8 (1.9)	2.1 (2.4)
Mn crowding (mm):			
Mean (SD)	2.2 (1.9)	2.3 (1.6)	2.6 (2.1)

Maxillary (Mx) p-value = 0.20
Mandibular (Mn) p-value = 0.57

Case D

Characteristic	<10% N = 68	10-25% N = 172	25-75% N = 57
Mx crowding (mm):			
Mean (SD)	2.2 (2.1)	1.9 (1.9)	2.6 (3.0)
Mn crowding (mm):			
Mean (SD)	2.9 (2.0)	1.8 (1.7)	2.5 (2.6)

Maxillary (Mx) p-value = 0.29
Mandibular (Mn) p-value = 0.003; pairwise testing, p = 0.0005 <10 vs 10-25%

Crowding Assessment Method vs. Precision

Case A

Characteristic	Measurement N = 34	Visual estimation only N = 239
Mx crowding (mm):		
Mean (SD)	1.5 (1.2)	2.1 (2.3)
Mn crowding (mm):		
Mean (SD)	1.6 (1.3)	1.7 (1.9)

Maxillary (Mx) p-value = 0.06; pairwise testing, p = 0.02
Mandibular (Mn) p-value = 0.82; pairwise testing, p = 0.04

Case B

Characteristic	Measurement N = 34	Visual estimation only N = 239
Mx crowding (mm):		
Mean (SD)	2.2 (2.2)	2.1 (2.2)
Mn crowding (mm):		
Mean (SD)	2.9 (2.0)	3.0 (1.9)

Maxillary (Mx) p-value = 0.5
Mandibular (Mn) p-value = 0.87

Case C

Characteristic	Measurement N = 34	Visual estimation only N = 239
Mx crowding (mm):		
Mean (SD)	2.0 (1.7)	2.0 (2.1)
Mn crowding (mm):		
Mean (SD)	2.2 (1.3)	2.4 (1.7)

Maxillary (Mx) p-value = 0.98
Mandibular (Mn) p-value = 0.68

Case D

Characteristic	Measurement N = 34	Visual estimation only N = 239
Mx crowding (mm):		
Mean (SD)	1.8 (1.6)	2.1 (2.3)
Mn crowding (mm):		
Mean (SD)	1.8 (1.8)	2.3 (2.1)

Maxillary (Mx) p-value = 0.57
Mandibular (Mn) p-value = 0.20; pairwise testing, p = 0.0005

Region vs. Precision

Case A

Characteristic	Midwest N = 57	Northeast N = 55	West N = 92	South N = 83	Other N = 10
Mx crowding (mm):					
Mean (SD)	1.8 (1.9)	3.0 (2.9)	2.1 (2.4)	1.5 (1.6)	1.8 (1.9)
Mn crowding (mm):					
Mean (SD)	1.3 (1.4)	2.5 (2.3)	1.9 (2.0)	1.3 (1.3)	1.3 (2.1)

Maxillary (Mx) p-value = 0.04; pairwise testing, p = 0.03 Northeast vs South
Mandibular (Mn) p-value = 0.005; pairwise testing, p = 0.01
Northeast vs Midwest, p = 0.007 Northeast vs South

Case B

Characteristic	Midwest N = 57	Northeast N = 55	West N = 92	South N = 83	Other N = 10
Mx crowding (mm):					
Mean (SD)	2.4 (2.5)	2.9 (2.6)	2.2 (2.2)	1.6 (1.5)	2.4 (3.2)
Mn crowding (mm):					
Mean (SD)	3.2 (2.1)	3.2 (2.0)	3.0 (2.0)	2.8 (1.8)	3.2 (2.1)

Maxillary (Mx) p-value = 0.03; pairwise testing, p = 0.03 Northeast vs South
Mandibular (Mn) p-value = 0.68

Case C

Characteristic	Midwest N = 57	Northeast N = 55	West N = 92	South N = 83	Other N = 10
Mx crowding (mm):					
Mean (SD)	1.9 (1.8)	2.7 (2.6)	2.1 (2.3)	1.5 (1.3)	2.5 (3.1)
Mn crowding (mm):					
Mean (SD)	2.3 (1.8)	2.4 (2.1)	2.5 (1.9)	2.1 (1.4)	2.6 (1.5)

Maxillary (Mx) p-value = 0.02; pairwise testing, p = 0.04 Northeast vs South
Mandibular (Mn) p-value = 0.60

Case D

Characteristic	Midwest N = 57	Northeast N = 55	West N = 92	South N = 83	Other N = 10
Mx crowding (mm):					
Mean (SD)	1.6 (1.5)	3.1 (3.1)	2.4 (2.3)	1.6 (1.6)	1.7 (2.1)
Mn crowding (mm):					
Mean (SD)	1.7 (1.3)	3.2 (2.9)	2.2 (2.0)	1.7 (1.3)	2.7 (2.6)

Maxillary (Mx) p-value = 0.0008; pairwise testing, p = 0.04
Northeast vs Midwest, p = 0.02 Northeast vs South
Mandibular (Mn) p-value = 0.008; pairwise testing, p = 0.01
Northeast vs Midwest and South

Clinician Sex vs. Precision

Case A

Characteristic	Female N = 99	Male N = 194
Mx crowding (mm):		
Mean (SD)	2.3 (2.8)	1.9 (1.9)
Mn crowding (mm):		
Mean (SD)	1.8 (2.0)	1.7 (1.8)

Maxillary (Mx) p-value = 0.25
Mandibular (Mn) p-value = 0.70

Case B

Characteristic	Female N = 99	Male N = 194
Mx crowding (mm):		
Mean (SD)	2.1 (2.2)	2.2 (2.2)
Mn crowding (mm):		
Mean (SD)	2.9 (2.1)	3.1 (1.9)

Maxillary (Mx) p-value = 0.70
Mandibular (Mn) p-value = 0.55

Case C

Characteristic	Female N = 99	Male N = 194
Mx crowding (mm):		
Mean (SD)	2.1 (2.1)	2.0 (2.1)
Mn crowding (mm):		
Mean (SD)	2.2 (1.7)	2.5 (1.8)

Maxillary (Mx) p-value = 0.64
Mandibular (Mn) p-value = 0.24

Case D

Characteristic	Female N = 99	Male N = 194
Mx crowding (mm):		
Mean (SD)	2.2 (2.4)	2.1 (2.1)
Mn crowding (mm):		
Mean (SD)	2.1 (1.9)	2.2 (2.1)

Maxillary (Mx) p-value = 0.62
Mandibular (Mn) p-value = 0.53

Appendix XI. Association Between Clinician Demographic (Experience, Extraction Prevalence, and Crowding Assessment Method) and Extraction Recommendation

Experience (Years) vs. Extraction Recommendation

Case A

	<5	5-15	>15
Case recommendation			
Extractions	3.0 (3.7%)	2.0 (2.6%)	9.0 (7.4%)
Non-extraction	79.0 (96.3%)	75.0 (97.4%)	112.0 (92.6%)

p = 0.30

Case B

<5	5-15	>15
66.0 (79.5%)	51.0 (68.9%)	100.0 (80.6%)
17.0 (20.5%)	23.0 (31.1%)	24.0 (19.4%)

p = 0.14

Case C

	<5	5-15	>15
Case recommendation			
Extractions	12.0 (14.3%)	17.0 (22.4%)	48.0 (39.0%)
Non-extraction	72.0 (85.7%)	59.0 (77.6%)	75.0 (61.0%)

p = 0.002

Case D

<5	5-15	>15
71.0 (84.5%)	65.0 (86.7%)	100.0 (82.6%)
13.0 (15.5%)	10.0 (13.3%)	21.0 (17.4%)

p = 0.78

Extraction Prevalence vs. Extraction Recommendation

Case A

	<10%	10-25%	25-75%
Case recommendation			
Extractions	1 (1.6%)	6 (3.7%)	7 (12.7%)
Non-extraction	62 (98.4%)	156 (96.3%)	48 (87.3%)

p-value = 0.02; pairwise testing, p >0.05 all comparisons <10,10-25,25-75%

Case B

<10%	10-25%	25-75%
33 (52.4%)	140 (85.9%)	44 (80.0%)
30 (47.6%)	23 (14.1%)	11 (20.0%)

p-value <0.0001; pairwise testing, p = 0.00 <10 vs 10-25%, p = 0.004 <10 vs 25-75%

Case C

	<10%	10-25%	25-75%
Case recommendation			
Extractions	12 (18.8%)	42 (25.8%)	23 (41.1%)
Non-extraction	52 (81.2%)	121 (74.2%)	33 (58.9%)

p-value = 0.02; pairwise testing, p = 0.03 <10 vs 25-75%

Case D

<10%	10-25%	25-75%
41 (64.1%)	145 (90.1%)	50 (90.9%)
23 (35.9%)	16 (9.9%)	5 (9.1%)

p-value <0.000; pairwise testing, p = 0.00 <10 vs 10-25%, p = 0.001 <10 vs 25-75%

Crowding Assessment Method vs. Extraction Recommendation

Case A

	Measurement on models	Visual estimation
Case recommendation		
Extractions	3.0 (10.0%)	10.0 (4.4%)
Non-extraction	27.0 (90.0%)	216.0 (95.6%)

p-value = 0.37

Case B

Measurement on models	Visual estimation
30.0 (96.8%)	166.0 (73.5%)
1.0 (3.2%)	60.0 (26.5%)

p-value = 0.003

Case C

	Measurement on models	Visual estimation
Case recommendation		
Extractions	16.0 (48.5%)	55.0 (24.2%)
Non-extraction	17.0 (51.5%)	172.0 (75.8%)

p-value = 0.02

Case D

Measurement on models	Visual estimation
31.0 (100.0%)	186.0 (82.7%)
0.0 (0.0%)	39.0 (17.3%)

p-value = 0.01

Region vs. Extraction Recommendation

Case A

	Midwest	Northeast	West	South	Other
Case recommendation					
Extractions	2.0 (3.8%)	4.0 (8.0%)	3.0 (3.4%)	4.0 (5.0%)	1.0 (11.1%)
Non-extraction	51.0 (96.2%)	46.0 (92.0%)	85.0 (96.6%)	76.0 (95.0%)	8.0 (88.9%)

p-value = 0.51

Case B

	Midwest	Northeast	West	South	Other
Case recommendation					
Extractions	44.0 (81.5%)	36.0 (70.6%)	67.0 (76.1%)	61.0 (77.2%)	9.0 (100.0%)
Non-extraction	10.0 (18.5%)	15.0 (29.4%)	21.0 (23.9%)	18.0 (22.8%)	0.0 (0.0%)

p-value = 0.36

Case C

	Midwest	Northeast	West	South	Other
Case recommendation					
Extractions	17.0 (30.9%)	14.0 (27.5%)	30.0 (34.1%)	13.0 (16.2%)	3.0 (33.3%)
Non-extraction	38.0 (69.1%)	37.0 (72.5%)	58.0 (65.9%)	67.0 (83.8%)	6.0 (66.7%)

p-value = 0.09

Case D

	Midwest	Northeast	West	South	Other
Case recommendation					
Extractions	47.0 (87.0%)	45.0 (88.2%)	67.0 (77.9%)	68.0 (85.0%)	9.0 (100.0%)
Non-extraction	7.0 (13.0%)	6.0 (11.8%)	19.0 (22.1%)	12.0 (15.0%)	0.0 (0.0%)

p-value = 0.34

Clinician Sex vs Extraction Recommendation

Case A

	Female	Male
Case recommendation		
Extractions	2.0 (2.2%)	12.0 (6.5%)
Non-extraction	89.0 (97.8%)	173.0 (93.5%)

p-value = 0.15

Case B

	Female	Male
Case recommendation		
Extractions	67.0 (72.0%)	147.0 (79.9%)
Non-extraction	26.0 (28.0%)	37.0 (20.1%)

p-value = 0.17

Case C

	Female	Male
Case recommendation		
Extractions	24.0 (26.1%)	51.0 (27.3%)
Non-extraction	68.0 (73.9%)	136.0 (72.7%)

p-value = 0.89

Case D

	Female	Male
Case recommendation		
Extractions	74.0 (80.4%)	158.0 (85.9%)
Non-extraction	18.0 (19.6%)	26.0 (14.1%)

p-value = 0.30