

# **Facemask Therapy: Outcomes and Assessing the Need for Subsequent Treatment**

Lauren E. Hagel

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

Master of Science in Dentistry

University of Washington

2022

Committee:

Burcu Bayirli

Geoffrey Greenlee

Lawrence Garetto

Program Authorized to Offer Degree:

School of Dentistry – Department of Orthodontics

University of Washington

**Abstract**

Facemask Therapy: Outcomes and Assessing the Need for Subsequent Treatment

Lauren E. Hagel, DDS

Burcu Bayirli

Clinical Associate Professor - Department of Orthodontics

**Objective:** The purpose of this study was to evaluate a cohort of patients who had early treatment with a facemask to determine initial outcomes of the treatment, evaluate post-treatment relapse, assess the long term need for subsequent treatment, and identify patient characteristics that may be indicative of future treatment need.

**Methods:** Patients treated with a facemask were identified at the University of Washington School of Dentistry Graduate orthodontics clinic and two Seattle-based private orthodontic practices. Treated patients were frequency matched to historic untreated controls. Cephalometric changes between different time points were compared among the two groups to determine which changes can be attributed to growth versus treatment. For the treated group, the need for orthognathic surgery was assessed by three orthodontists, independently. Based on their recommendations, a logistic regression model was developed to identify variables associated with surgical need both when the patient is in the permanent dentition and prior to initiation of facemask therapy.

**Results:** The following significant mean changes were seen over the course of facemask therapy in the treated group: SNA increased by  $0.9^\circ$ , ANB increased by  $1.0^\circ$ , mandibular plane increased by  $1.0^\circ$ , overjet increased by 3.3 mm, IMPA decreased by  $1.9^\circ$ , interincisal angle decreased by  $4.0^\circ$ , and U1-NA increased by  $3.7^\circ$ . The change in SNB, Wits Appraisal, and LFH were not statistically significant. Over an average of 3.1 years following interceptive treatment, there was a statistically significant decrease in overjet by 1.2 mm, ANB by  $1.4^\circ$  and an increase in SNB by  $2.1^\circ$ . All other values did not show significant change. Even after early treatment with a facemask therapy, orthognathic surgery as the ideal subsequent treatment was recommended for 41% of patients. Decreased overjet, proclination of the maxillary incisors, retroclination of the lower incisors, and increased lower facial height were associated with surgical recommendations. The lower incisor position prior to interceptive treatment is also associated with the future need for surgery.

**Conclusions:** Although there was significant skeletal change over the course of interceptive treatment, similar change was also seen in untreated controls. Thus, the positive change in overjet as a result of facemask therapy is attributed more to dental change than skeletal. The skeletal change we see during the interceptive treatment is most likely due to growth and not facemask therapy. Following treatment, the decrease in overjet can be attributed to continued mandibular growth. The position of the maxillary and mandibular incisors are important characteristics to take into consideration when determining the probability of facemask therapy preventing the future need for orthognathic surgery.

©Copyright 2022  
Lauren E. Hagel, DDS

## TABLE OF CONTENTS

INTRODUCTION: .....	1
MATERIALS & METHODS .....	4
RESULTS .....	10
DISCUSSION .....	22
CONCLUSIONS .....	30
LITERATURE CITED .....	31
APPENDIX.....	35

## **Acknowledgements**

I would like to thank my research committee for their guidance and support. This project would not be the same without your contributions and dedication.

- Dr. Burcu Bayirli
- Dr. Geoff Greenlee
- Dr. Lawrence Garetto

I would also like to thank Dr. Lloyd Mancel for conducting all of the statistical analysis. Drs. Heather Woloshyn and Reid Winkler for contributing patients from their practices to this study and Drs. Greg Huang, Anne-Marie Bollen, and Bobby Cohanin for participating in the panel evaluation of patient records.

To my family, friends & mentors - thank you for supporting me through so many years of school. I would not be here without all of you!

## **INTRODUCTION:**

Epidemiologic studies show that between 5 to 15% of the United States population are affected by Class III malocclusions. Asian populations are reported to have a Class III prevalence of 4-19% (Alexander *et al.* 2009), while Caucasian prevalence is 0.2-12% (Zhang *et al.*, 2015; Woon *et al.*, 2011; Zionic Alexander *et al.* 2009). An underlying skeletal discrepancy between the maxilla and mandible is often the cause of the Class III relationship. 75% of Class III patients have either a retrognathic maxilla or a combination of maxillary retrognathism and mandibular prognathism (Zhang *et al.*, 2015). Early growth modification for patients with Class III malocclusion often includes maxillary protraction therapy with a reverse-pull headgear (also known as a facemask).

The short-term treatment effects of facemask therapy are well understood and documented in orthodontic literature through observational studies, randomized clinical trials, systematic reviews, meta-analyses, and a Cochrane review. There is a consensus amongst authors that Class III malocclusions can be successfully corrected with facemask therapy during early treatment. This is demonstrated through beneficial changes in molar classification, overjet, skeletal position of the maxilla, position of the incisors, and soft tissue profile (Ngan *et al.*, 1996, Turley *et al.*, 1998, Kapust *et al.*, 1998, Kim *et al.*, 1999, Vaughn *et al.*, 2005, Seehra *et al.*, 2011 Jager *et al.* 2001, Foersch *et al.*, 2015, Zhang *et al.*, 2015).

There are very few studies that compare treatment effects of the facemask to untreated controls. A study published by Sung *et al.* (1998) and meta-analysis published by Jager *et al.*, (2001) used Class I controls to demonstrate changes attributed to facemask therapy. These results may be challenging to interpret, because there are significant differences in growth patterns between Class I patients and Class III patients (Alexander *et al.*, 2009). Patients with a Class III

skeletal pattern generally exhibit maxillary growth deficiency and mandibular growth excess, unlike their Class I counterparts (Zere *et al.*, 2018) and these skeletal components tend to worsen with growth (Zionic Alexander *et al.*, 2009). In 2010, Mandall *et al.* conducted a randomized control trial that evaluated short term effects of the facemask appliance compared to an untreated Class III sample. Their study demonstrated a 1.4° increase in SNA, 2.1° increase in ANB, and 70% of patients achieving positive overjet over the course of treatment. They concluded that facemask therapy results in favorable skeletal and dental effects in the short term compared to untreated controls.

Even if facemask therapy is successful in the short term, it is important for practitioners to understand the long-term outcomes for their patients. Class III orthodontic patients have significant relapse potential due to unpredictable and potentially unfavorable growth of the maxilla and mandible (Mandall *et al.*, 2010, 2012). This puts patients treated early at risk for outgrowing the therapeutic correction achieved by the facemask. Thus, long term treatment stability and later orthodontic treatment needs are challenging to predict. For many patients, subsequent orthodontic treatment (including orthognathic surgery) may still be necessary to correct the skeletal discrepancy and underlying malocclusion. The definition of surgical need differed amongst studies but overall, current orthodontic literature reports that between 66-75% of early treatment facemask patients do *not* require future orthognathic surgery (Hagg *et al.*, 2003, Bacetti *et al.*, 2004, Wells *et al.*, 2006, Masucci *et al.*, 2011, Mandall *et al.*, 2016).

Having a better understanding of factors contributing to improved long-term stability after facemask treatment allows practitioners to give families a prediction for success. Previous studies have looked for possible skeletal predictors of long-term failure after facemask treatment (Bacetti *et al.*, 2004, Wells *et al.*, 2006, Masucci *et al.*, 2011). These studies reported

increased mandibular ramus length, acute cranial base angle, increased inclination of mandibular plane, decreased posterior vertical height, decreased mandibular height, increased gonial angle, downward inclination of the mandibular plane to Frankfort horizontal, a mesial molar relationship, and decrease in overjet as predictors for treatment failure. None of these studies, however, included dental cephalometric variables in their regression models as possible predictors for failure.

For young patients presenting with a Class III malocclusion, the determination of risk factors affecting long-term success are central to determining if early treatment will be worthwhile. Several factors must be considered, including treatment effects (both beneficial and detrimental), costs (direct and indirect), and the probability that early treatment will prevent future, possibly more invasive treatment. There is evidence to suggest that early treatment can decrease the severity of the malocclusion from medically necessary to elective (Jolley *et al.*, 2010) and creates less financial burden (Bresnahan *et al.*, 2010). However, if facemask therapy is not successful and subsequent orthodontic treatment with orthognathic surgery is necessary, there will be significant additional treatment costs.

The goals of this study were to assess the short-term post treatment outcomes of facemask therapy during early orthodontic treatment, the long-term stability of the treatment into the full permanent dentition and the need for subsequent orthodontic treatment in a cohort of patients treated with early limited orthodontics for Class III malocclusion. Predictors associated with the need for orthognathic surgery were explored for both patients in the permanent dentition as well as prior to facemask therapy.

It is our hope that this study will provide clinicians guidance in recommending early treatment to developing Class III patients by better understanding which patients are likely to

have successful and sustained correction and which patients will not. We hope that a better understanding of long-term treatment outcomes and the identification of possible predictors for surgical need will help guide treatment recommendations at an early age to maximize patient benefit and minimize patient harm.

## **MATERIALS & METHODS:**

Orthodontic patient databases at the University of Washington and two private practices were searched for all consecutively treated patients who underwent facemask therapy as part of an orthodontic treatment plan from January 2009 to July 2021. A total of 83 patients were identified; 42 were treated at the University of Washington School of Dentistry Graduate Orthodontics clinic and the remaining 40 patients were treated in a private practice setting.

The inclusion criteria were (1) intra-oral and extra-oral diagnostic photos and/or cephalometric radiographs available prior to facemask therapy (T1) and at the conclusion of facemask therapy (T2). The exclusion criteria were (1) illegible or non-diagnostic radiographs; (2) patients with craniofacial syndromes; and (3) patients treated concurrently with growth altering medications.

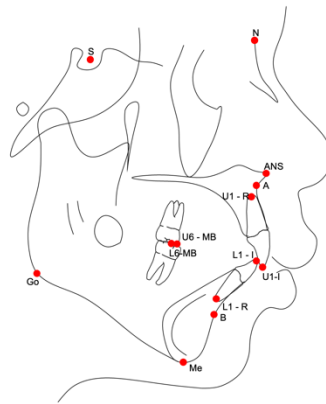
For each patient included, available orthodontic records (intra-oral/extra-oral photos and cephalometric radiographs) were collected at two time points: prior to facemask therapy (T1) and after facemask therapy (T2). If available, records at a third time point (patients in full permanent dentition and at least one-year post-facemask therapy) were also collected (T3). Of the 83 included patients, 52 patients had records (photos, cephalometric radiograph, or both) available at T3.

Each chart was reviewed for demographic and relative clinical information. Demographic data included age, gender, and ethnicity. The presence of a functional shift, the number of months the facemask was worn, and an evaluation of compliance were also recorded. Compliance, evaluated as adherence to practitioners' instructions as described in the chart notes, was divided into three categories: compliant, sub-optimal, or non-compliant. Patients were categorized as compliant if the chart notes reflected that the appliance was worn as directed, sub-optimal if there was any report in the chart of non-compliance, and non-compliant if records indicated appliance was worn minimally or not at all. In addition to facemask therapy, most of the patients also received expansion and dental alignment with limited fixed appliances. Retention protocols following early treatment generally included maxillary Hawley retainers. If any treatment was conducted in the mandible, either mandibular Hawley retainers or a lower lingual holding arch was placed.

Extra-oral and intra-oral photographs were evaluated for each patient at each time point available by three evaluators simultaneously and a consensus diagnosis was reached (LH, BB, GG). Evaluators were blinded to the age of the patient and timepoint. The following variables were determined: Angle classification, overjet, patient's profile pattern, and growth status. Overjet was categorically defined as positive, zero, mildly negative, or severely negative. Profile was defined as concave (upper jaw behind chin), convex (upper jaw prominent relative to chin), or orthognathic (straight line from the bridge of the nose to the base of the upper lip to the chin) (Proffit *et al.*, 2013). Based on facial photo evaluation, growth status was also evaluated, and patients were described as pre-pubertal, circum-pubertal, or post-pubertal.

Untreated controls were selected from patients participating in the Bolton-Brush, Burlington Growth, Denver Growth, Fels Longitudinal, Forsyth Twin, Iowa Growth, Michigan

Growth, and Oregon Growth studies. A search was conducted for all “Class III” patients in their data base. Controls were frequency matched at T1 to the treated patients based on gender and age within two years. T2 and T3 cephalometric radiographs for each selected patient were chosen based on age within two years. After matching for age and gender, each control subject was also matched on ANB, Wits, SNA, and overjet to within one standard deviation of the mean value of the treated subject.



**Figure 1: Identified Cephalometric Landmarks;** S, sella; N, nasion; ANS, anterior nasal spine; A, A-point; U1-R, U1 root apex; U1-I, U1 incisal edge; L1-I, L1 incisal edge; L1-R, L1 root apex; B, B-point; Me, menton; U6-MB, U6 mesiobuccal cusp; L6-MB, L1 mesiobuccal cusp; Go, gonion

Cephalometric radiographs were de-identified and randomized for all patients (treated and controls) and timepoints. The radiographs were imported into Dolphin Imaging Software (Patterson Dental, St. Paul, Mn) and a custom cephalometric analysis was performed.

Magnification of the cephalometric radiographs were accounted for in the software. Radiographs were traced by two investigators (LEH and BB) independently and resulting values compared. If the difference between two values from each individual investigator differed beyond the standard deviation for that variable, the points were re-identified together to reach a consensus. Reported values are average values between the two investigators.

Cephalometric landmarks identified are shown in Figure 1. From these points, the following variables were measured and reported: overjet (mm), ANB (degree) SNA (degree), SNB (degree), Wits appraisal (mm), IMPA (degree), U1-L1 (degree), U1-NA (degree), SN-MP (degree), % lower facial height ( $(ANS-Me)/(N-ANS+ANS-Me)$  %). Mean cephalometric values and standard deviations are reported for each time point as well as mean change from T1 to T2, T2 to T3, and T1 to T3.

For each cephalometric radiograph, the cervical vertebrae maturation (CVM) stage was determined through consensus discussion between three investigators (LEH, BB, GMG). Determination was based on the method outlined by McNamara and Franchi (McNamara *et al.*, 2018). Investigators were blinded to patient age, dentition, and time point. After determination of CVM stage, patients were assigned to one of three categories: pre-pubertal (CVM stage 1 or 2), circum-pubertal (CVM stage 3 or 4), or post-pubertal (CVM stage 5 or 6).

A panel evaluation was used to identify which patients would benefit from a treatment plan including orthognathic surgery at T3. Three experienced orthodontists analyzed patient records (age, photographs, cephalometric radiographs) of 49 patients at T3. Panelists were asked to respond to a simple survey asking:

1. Would you treat this patient now, comprehensively?
  - 1) Yes
  - 2) No
2. Independent of whether you would treat now or wait, does the ideal treatment plan for this patient include orthognathic surgery?
  - 1) Definitely not
  - 2) Likely not

- 3) Likely yes
- 4) Definitely yes

### *Statistical Analysis*

#### *Sample Size Calculation*

The primary clinical outcomes of interest for the comparisons between the treated sample and control group were overjet, ANB, Wits appraisal, and U1-NA (degrees). A 2mm change in overjet, 3° change in ANB, 4mm change in the Wits appraisal, and 3° change in U1-NA were deemed clinically significant based on previous work (Mandall *et al.*, 2012). A sample size of 30 in each group (treated and control) has at least 80% power to detect a difference in means of 2 mm for overjet, 3° for ANB, and 4 mm for Wits appraisal assuming a standard deviation between 0.67 and 0.75 using a two-group t-test with a 0.05 significance level and frequency matching of treated patients to control patients. A total sample size of 30-40 patients per group (treated and control) was needed for the comparison portion of the study.

#### *Statistical Methods*

Descriptive statistics were used to report both categorical and continuous variables at each time point (if applicable). The average change for the cephalometric variables between timepoints and associated 95% confidence intervals were reported. A paired t-test was used to detect if the change between time points were significantly different.

Although control patients were frequency matched to treated patients based on age and years between time point, there were some differences in the cephalometric variables, age at T1, and length of time between time points. Analysis of covariance (ANCOVA) was used to compare treated and control patients to adjust for these differences. The difference between control and treated patients was adjusted for age at the specific time point for direct comparisons. The

difference between control and treated patients was adjusted for the cephalometric measure at the initial time point for the change, age, and time between the specific time points when evaluating change data.

To determine surgical needs and identify pre- and post-treatment variables that may be associated with the need for future surgery, first the evaluators' responses to the survey were evaluated and consensus determined regarding a patient's need for surgery. Surgical need was determined by majority consensus. Using the binary decision on the need for surgery, univariate and multivariate logistic regression models were developed to determine predictive values at T1, T2, and T3 as well as the change from T1 to T2 that predict the future need for orthognathic surgery. For each variable, odds ratios, 95% confidence intervals, and p-values are reported. The odds ratios are reported for one standard deviation increase (except for the change from T1 to T2), because clinicians typically rely on deviations from a cephalometric norms.

#### *Reliability Studies*

Inter- and intra-rater reliability was assessed by re-analyzing 10 cephalometric radiographs and patient photos two weeks apart. To compare the two measurements, the mean and standard deviation was computed for each set of measurements. The mean of the differences and 95% confidence interval for the mean difference, the intraclass correlation coefficient (ICC), 95% confidence interval for the ICC, Dahlberg's error, and the minimum and maximum for the absolute value of the difference between the two measurements were calculated and reported in Appendices 1-4. For the categorical measures, a cross tabulation of the two measurements was created and the agreement was summarized by the percent agreement and kappa statistic. Inter- and intra-reliability data is reported in appendix 5 for cephalometric and non-cephalometric variables.

## RESULTS:

Descriptive statistics (gender, race/ethnicity, and age at each timepoint) for the treated sample are reported in Tables 1 and 2.

Table 1: Descriptive Statistics for Treated Group

	T1	T2	T3
<b>Characteristic</b>	<b>N = 82</b>	<b>N = 76</b>	<b>N = 52</b>
<b>Gender</b>			
Female	39 (47.6%)	36 (46.8%)	26 (50.0%)
Male	43 (52.4%)	41 (53.2%)	26 (50.0%)
<b>Ethnicity</b>			
African American	6 (7.3%)	6 (7.8%)	3 (5.8%)
Asian	26 (31.7%)	25 (32.5%)	13 (25.0%)
Caucasian	37 (45.1%)	36 (46.8%)	26 (50.0%)
Hispanic	5 (6.1%)	4 (5.2%)	4 (7.7%)
Middle Eastern	4 (4.9%)	4 (5.2%)	2 (3.8%)
Mixed	4 (4.9%)	2 (2.6%)	4 (7.7%)

Table 2: Mean Age for Treated Subjects

	<b>All</b>		<b>Male</b>		<b>Female</b>	
	<b>N</b>	<b>Age (SD)<sup>1</sup></b>	<b>N</b>	<b>Age (SD)<sup>1</sup></b>	<b>N</b>	<b>Age (SD)<sup>1</sup></b>
<b>T1</b>	82	9.3 (1.5)	43	9.7 (1.5)	39	8.8 (1.4)
<b>T2</b>	76	11.0 (1.5)	40	11.5 (1.5)	36	10.5 (1.4)
<b>T3</b>	52	14.2 (1.9)	26	15 (2.1)	26	13.3 (1.4)

<sup>1</sup>Age reported in years

The average length of time between T1-T2 (treatment time) was 1.7 years (SD 0.8), T2-T3 was 3.3 years (1.6 years), and T1-T3 was 4.9 years (SD 1.6 years).

CVM stage, Angle Classification, overjet evaluation, profile pattern, and clinical assessment of growth status are reported in Appendix 6. ‘Unknown’ indicates inadequate records for that time point. Some patients had both photographs and cephalometric radiographs, and some had only photographs or a cephalometric radiograph.

We attempted to find other patient variables indicative of skeletal age. These included cervical vertebrae maturation determination and assessment maturity from photographs.

Photographic maturity was based on the presence of secondary sex characteristics and changes attributed to puberty visible on the facial photographs. Reliability of these methods were low; these values are not reported and analyzed in the manuscript but can be found in appendix 3.

#### *Other Clinical and Treated Related Data Reported for Treated Sample*

The presence or absence of a functional shift and patient compliance was determined from chart review. 46 treated patients (56%) had a functional shift (anterior, lateral, or both) at T1. Of the patients with a shift, 41 (73%) had an anterior shift, 2 (4%) had a combination of anterior and lateral shifts, and 3(5%) had a lateral shift. The shifts ranged from 0.5 mm to 3 mm.

54 (66%) of the 82 treated patients were deemed as compliant, 25 (30%) had sub-optimal compliance, and 3 (4%) were non-compliant. The mean overall time the facemask was worn was 12 months (SD=7.1) with a range of 2 months to 47 months. Compliant patients had an average of 9.9 months (SD=4.5) of facemask wear with a range of 2 to 21 months. The non-compliant patients had a mean of 9 months (SD=1.7) of facemask wear with a range of 7–10 months. The suboptimal group had an average of 16.8 months (SD=9.5) of facemask wear with a range of 7-47 months of wear.

Early treatment for many of these patients also included rapid palatal expansion and upper and lower fixed appliances. We did not report on specific treatment regimens for individual patients.

#### *Cephalometric Measures at Each Timepoint for Treated Sample*

A summary of average cephalometric values and 95% confidence intervals at each timepoint for treated patients are reported in Table 3. Cephalometric norms reported in the orthodontic literature and their standard deviations are also listed for comparison. At all-time

points, our treated sample was statistically different from the reported norms for all bolded variables reported in Table 3. At T1, all characteristics, except SNB, were statistically different for our treated group compared to cephalometric norms. This could be attributed to the sample being more midface deficient compared to norms. The included patients were likely still growing; so, it is expected that SNB will continue to increase. At T2, overjet and U1-L1 angle were normalized and no longer significantly different than the norm. At T3, overjet decreased and remained statistically significant. The average value for U1-L1 angle remained consistent with the normal values. SNA also increased from T2 to T3 and at T3 was not statistically different than the norms. Average SNB values were not statistically different than cephalometric norms at any time points.

Table 3: T1, T2, and T3 Mean Cephalometric Values of Sample Compared to Cephalometric Norms

Variable	Norms	T1 (N=82)		T2 (N=77)		T3 (N=52)	
	Value SD)	Mean (SD)	95% CI	Mean (SD)	95% CI	Mean (SD)	95% CI
Overjet (mm)	2.5 (2.5)	<b>-0.9 (2.01)</b>	-1.3, -0.4	2.5 (1.75)	2.1, 2.9	<b>1.2(2.00)</b>	0.7, 1.8
ANB (degree)	1.6 (1.5)	<b>-0.6 (2.03)</b>	-0.9, -0.1	<b>0.4 (2.32)</b>	-0.1, 1.0	<b>-0.8 (2.67)</b>	-1.5, 0.0
SNA (degree)	82 (3.5)	<b>79.9 (4.01)</b>	79.0, 80.8	<b>80.6 (4.35)</b>	79.6, 81.5	81.3 (4.61)	80.1, 82.6
SNB (degree)	80.9 (3.4)	80.4 (3.98)	79.5, 81.3	80.1 (3.88)	79.2, 81.0	82.1 (4.82)	80.7, 83.4
Wits (mm)	-1 (1)	<b>-5.0 (2.85)</b>	-5.6, -4.4	<b>-4.4 (3.00)</b>	-5.0, -3.7	<b>-5.7 (3.60)</b>	-6.7, -4.6
IMPA (degree)	95 (7)	<b>86.4 (6.91)</b>	84.9, 87.9	<b>84.3 (7.38)</b>	82.6, 86.0	<b>84.4 (7.55)</b>	82.3, 86.5
U1-L1 (degree)	131 (7)	<b>134.4 (11.07)</b>	132.0, 136.9	130.8 (9.87)	128.5, 133.0	131.1 (9.19)	128.5, 133.6
U1-NA (degree)	22.8 (5.7)	<b>25.2 (7.38)</b>	23.5, 26.8	<b>28.9 (7.05)</b>	27.3, 30.5	<b>29.3 (6.02)</b>	27.6, 31.0
MP(degree)	33 (6)	<b>34.6 (5.86)</b>	33.3, 35.8	<b>35.6 (5.84)</b>	34.3, 36.9	<b>33.9 (6.48)</b>	32.1, 35.7
LFH (%)	57 (0.035)	<b>55.5 (2.73)</b>	54.9, 56.1	<b>55.4 (2.96)</b>	54.7, 56.1	<b>56.1 (2.21)</b>	55.5, 56.7

<sup>1</sup>CI = Confidence interval

<sup>2</sup>Bold values represent statistically significant difference at p=0.05

### *Cephalometric Changes for Treated Sample*

Mean cephalometric values for treated sample were compared between T1 and T2, T2 and T3 as well as T1 and T3 to determine any significant changes. Mean changes over the three different time point comparisons, 95% confidence intervals, and P-values are reported in Table 4.

From time point T1-T2, there was significant change in overjet, ANB, SNA, IMPA, U1-L1, U1-NA, and mandibular plane. ANB increased by 1.0°, SNA increased by 0.9°, and

mandibular plane increased by 0.9°. Dentally, overjet increased by 3.3 mm, IMPA decreased by 1.9°, U1-L1 decreased by 4.0°, and U1-NA increased by 3.7°. The change in SNB, Wits Appraisal, and LFH were not statistically significant.

From T2-T3, changes in ANB (decreased by 1.4°), SNA (increased by 0.8°), SNB (increased by 2.1°), and mandibular plane (decreased by 1.6°) were all statistically significant. Dentally, overjet significantly decreased by 1.2 mm, but all other variables were not statistically different.

From T1-T3, there was significant change in overjet (increased by 2mm) IMPA (decreased by 1.7°), U1-L1 (decreased by 3.4°), and U1-NA (increased by 3.9°). Skeletally, there was an overall increase in SNA by 1.3° and SNB by 1.5°. ANB, Wits appraisal, mandibular plane angle and LFH did not change significantly.

Table 4: Mean Cephalometric Changes between Time Points for Treated Sample

Variable	T1-T2			T2-T3			T1-T3		
	T2-T1 <sup>1</sup>	95% CI <sup>2</sup>	P-Value <sup>3,4</sup>	T3-T2 <sup>1</sup>	95% CI <sup>2</sup>	P-Value <sup>3,4</sup>	T3-T1 <sup>1</sup>	95% CI <sup>2</sup>	P-Value <sup>3,4</sup>
Overjet (mm)	3.3 (2.5)	2.8, 3.9	<b>&lt;0.001</b>	-1.2 (2.2)	-1.9, -0.6	<b>&lt;0.001</b>	2.0 (2.7)	1.2, 2.7	<b>&lt;0.001</b>
ANB (degree)	1.0 (1.9)	0.6, 1.4	<b>&lt;0.001</b>	-1.4 (2.3)	-2.0, -0.7	<b>&lt;0.001</b>	-0.2 (2.4)	-0.9, 0.4	0.47
SNA (degree)	0.9 (2.4)	0.3, 1.4	<b>0.002</b>	0.8 (1.8)	0.3, 1.3	<b>0.002</b>	1.3 (2.9)	0.5, 2.1	<b>0.002</b>
SNB (degree)	-0.1 (2.3)	-0.6, 0.4	0.68	2.1 (2.1)	1.5, 2.7	<b>&lt;0.001</b>	1.5 (3.1)	0.6, 2.4	<b>&lt;0.001</b>
Wit (mm)	0.6 (3.2)	-0.1, 1.3	0.10	-1.4 (3.2)	-2.4, -0.5	<b>0.002</b>	-0.9 (3.5)	-1.8, 0.1	0.058
IMPA (degree)	-1.9 (4.8)	-3.0, -0.8	<b>0.001</b>	0.0 (4.3)	-1.3, 1.3	0.46	-1.7 (4.7)	-3.0, -0.4	<b>0.004</b>
U1-L1 (degree)	-4.0 (8.9)	-6.0, -1.9	<b>&lt;0.001</b>	0.1 (7.7)	-2.2, 2.3	0.71	-3.4 (9.4)	-6.0, -0.8	<b>0.029</b>
U1-NA (degree)	3.7 (7.2)	2.1, 5.4	<b>&lt;0.001</b>	1.0 (5.8)	-0.7, 2.8	0.27	3.9 (7.2)	1.9, 5.9	<b>&lt;0.001</b>
MP (degree)	0.9 (3.2)	0.2, 1.6	<b>0.016</b>	-1.6 (2.3)	-2.3, -1.0	<b>&lt;0.001</b>	-0.3 (3.7)	-1.4, 0.7	0.71
LFH (%)	0.0 (2.0)	-0.5, 0.4	0.88	0.0 (2.0)	-0.5, 0.4	0.21	0.0 (2.0)	-0.5, 0.4	0.46

<sup>1</sup>Mean (SD), <sup>2</sup>CI = Confidence Interval

<sup>3</sup>Paired t-test between mean variable values

<sup>4</sup>Bold values represent statistically significant results

### *Comparison of Treated Sample and Control Patients*

Treated (n=82) and control patients (n=38) were matched by age within 2 years at T1, T2, and T3. Table 5 reports the mean age and p-values. The two samples are not statistically different in age at the three time points. CVM was evaluated for both the treated and control groups at each timepoint and compared using a Fisher's Exact test. There was no significant difference in

CVM stage for the two groups at all three time points (Table 6). Additionally, the length of time between T1-T2, T2-3, and T1-T3 was determined for the treated and control groups and compared using a Welch Two Sample t-test. There was also no statistical difference in time elapsed between records points for the treated and control groups.

Table 5: Difference in age amongst treated and control groups at T1, T2, and T3

	Age at T1 <sup>1</sup>	Age at T2 <sup>1</sup>	Age at T3 <sup>1</sup>
<b>Treated</b> (N=82)	9.3 (1.5)	11.0 (1.5)	14.1 (1.9)
<b>Control</b> (N=38)	9.0 (1.3)	11.1 (1.4)	14.5 (1.4)
	p-value <sup>2</sup> : 0.42	p-value <sup>2</sup> : 0.78	p-value <sup>2</sup> : 0.28

<sup>1</sup>Age reported in years (Standard Deviation)  
<sup>2</sup>Welch Two Sample t-test

Table 6: Difference in CVM amongst treated and control groups at T1, T2, and T3

	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>p-value<sup>4</sup></b>
<b>CVM at T1</b>			0.12
1,2	41 (51.2%)	26 (68.4%)	
3,4	38 (47.5%)	11 (28.9%)	
5,6	1 (1.2%)	1 (2.6%)	
<b>CVM at T2</b>			0.57
1,2	26 (35.6%)	16 (42.1%)	
3,4	40 (54.8%)	17 (44.7%)	
5,6	7 (9.6%)	5 (13.2%)	
<b>CVM at T3</b>			0.26
1,2	1 (1.9%)	0 (0.0%)	
3,4	28 (53.8%)	13 (39.4%)	
5,6	23 (44.2%)	20 (60.6%)	

<sup>1</sup>N=82, <sup>2</sup>N=38, <sup>3</sup>n (%)  
<sup>4</sup>Fisher's exact test

All cephalometric variable values were compared at T1 for the treated and control groups. Table 7 reports the means and standard deviations for the treated and controls at T1. The means were compared using a Welch Two Sample t-test. Overjet, ANB, SNB and Wits appraisal were significantly different at T1 for the two groups, with the treated group displaying more severe Class III facies.

Table 7: Comparison of Treated and Control Mean Cephalometric Values at T1

<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>p-value<sup>4,5</sup></b>
Overjet (mm)	-0.9 (2.0)	0.1 (1.0)	<b>&lt;0.001</b>
ANB (degree)	-0.5 (2.0)	1.4 (2.3)	<b>&lt;0.001</b>
SNA (degree)	79.9 (4.0)	79.4 (2.5)	0.43
SNB (degree)	80.4 (4.0)	78	<b>&lt;0.001</b>
Wits (mm)	-5.0 (2.8)	-1.5 (1.7)	<b>&lt;0.001</b>
IMPA (degree)	86.4 (6.9)	87.6 (5.7)	0.34
U1-L1 (degree)	134.4 (11.1)	134.0 (9.9)	0.85
U1-NA (degree)	25.1 (7.4)	22.6	0.053
MP (degree)	34.6 (5.9)	36.3 (5.1)	0.094
LFH (%)	55.5 (2.7)	55.9 (2.2)	0.39

<sup>1</sup>N=82, <sup>2</sup>N=38

<sup>3</sup>Mean (SD)

<sup>4</sup>Welch Two Sample t-test

<sup>5</sup>Bold indicates statistical significance

To account for these differences, age and cephalometric variables were adjusted at T1 using analysis of covariance (ANCOVA). The adjusted differences at each individual timepoint are reported in Table 8. The differences between the control and treated patients at the individual timepoints are adjusted for age at the respective time point (i.e., for T1 data, the adjusted difference between control and treated patients is adjusted for age at T1, at T2 it is adjusted for age at T2, at T3 it is adjusted for age at T3). After adjustment, ANB, Wits Appraisal, overjet, and SNB were still statistically different at T1 between the two groups. At T2, overjet, ANB, SNB, IMPA, and U1-NA were statistically different. At T3, ANB, SNB, Wits Appraisal, and U1-NA were statistically different.

Table 8: Adjusted Comparison of Treated and Control Mean Cephalometric Values at T1, T2, & T3

<b>T1</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	-0.9 (2.0)	0.1 (1.0)	-1	-1.7, -0.4	<b>0.003</b>
ANB (degree)	-0.5 (2.0)	1.4 (2.3)	-1.9	-2.7, -1.0	<b>&lt;0.001</b>
SNA (degree)	79.9 (4.0)	79.4 (2.5)	0.5	-0.1, 1.9	0.51
SNB (degree)	80.4 (4.0)	78 (3.1)	2.4	0.1, 3.9	<b>0.001</b>
Wits (mm)	-5.0 (2.8)	-1.5 (1.7)	-3.4	-4.3, -2.4	<b>&lt;0.001</b>
IMPA (degree)	86.4 (6.9)	87.6 (5.7)	-0.9	-3.4, 1.6	0.46
U1-L1 (degree)	134.4 (11.1)	134.0 (9.9)	0.6	-3.6, 4.7	0.79
U1-NA (degree)	25.1 (7.4)	22.6 (6.1)	2.3	-0.4, 5.0	0.091
MP (degree)	34.6 (5.9)	36.3 (5.1)	-2	-4.1, 0.2	0.073
LFH (%)	55.5 (2.7)	55.9 (2.2)	-0.4	-1.4, 0.6	0.43
<b>T2</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	2.5 (1.8)	0.5 (0.9)	2.0	1.4, 2.6	<b>&lt;0.001</b>
ANB (degree)	0.4 (2.3)	1.4 (2.3)	-0.9	-1.8, 0.0	<b>0.050</b>
SNA (degree)	80.5 (4.4)	79.7 (3.8)	0.9	-0.8, 2.5	0.30
SNB (degree)	80.1(3.9)	78.3 (3.6)	1.8	0.3, 3.3	<b>0.017</b>
Wits (mm)	-4.3 (3.0)	-1.6 (1.9)	-2.8	-3.8, -1.7	<b>&lt;0.001</b>
IMPA (degree)	84.3 (7.4)	87.3 (5.5)	-3.0	-5.6, -0.3	<b>0.027</b>
U1-L1 (degree)	130.8 (9.9)	132.5 (8.4)	-1.8	-5.5, 2.0	0.35
U1-NA (degree)	28.9 (7.1)	24.4 (6.3)	4.5	1.8, 7.2	<b>0.001</b>
MP (degree)	35.6 (5.8)	36.1 (5.2)	-0.5	-2.7, 1.7	0.64
LFH (%)	55.4 (3.0)	55.7 (2.7)	-0.1	-1.2, 0.9	0.83
<b>T3</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	1.2 (2.0)	0.5 (0.9)	0.7	-0.04, 1.4	0.066
ANB (degree)	-0.8 (2.7)	0.6 (2.2)	-1.5	-2.6, 0.4	<b>0.007</b>
SNA (degree)	81.3 (4.6)	80.3 (2.8)	1.1	-0.7, 2.9	0.23
SNB (degree)	82.1 (4.8)	79.7 (3.3)	2.5	0.7, 4.4	<b>0.009</b>
Wits (mm)	-5.7 (3.6)	-1.7 (2.0)	-4.0	-5.4, -2.7	<b>&lt;0.001</b>
IMPA (degree)	84.4 (7.5)	86.2 (6.5)	-1.9	-5.1, 1.3	0.23
U1-L1 (degree)	131.1 (9.2)	133.1 (8.9)	-2.1	-6.2, 1.9	0.30
U1-NA (degree)	29.3 (6.0)	25.4 (6.1)	4.1	1.5, 6.8	<b>0.003</b>
MP (degree)	33.9 (6.5)	34.9 (5.4)	-1.1	-3.8, 1.6	0.43
LFH (%)	56.1 (2.2)	55.8 (2.6)	0.3	-0.8, 1.3	0.61

<sup>1</sup>N=82 <sup>2</sup>N=38

<sup>3</sup>Mean (SD)

<sup>4</sup>ANCOVA, <sup>5</sup>CI = Confidence Interval

<sup>6</sup>Bold indicates statistical significance

The change in values from T1-T2, T2-T3, and T1-T3 for each variable was compared between the treated and control samples. The values, adjusted change, and p-values are reported in Table 9.

For changes between T1 and T2, the difference between the control and treated patients was adjusted for the cephalometric measure at T1, age at T1, and the time from T1 to T2. The difference in change from T1-T2 between the two groups were limited to dental variables: overjet, IMPA, and U1-NA. All other change in variables were not statistically significant.

For changes between T2 and T3, the difference between the control and treated patients was adjusted for the cephalometric measure at T2, age at T2 and the time from T2 to T3. The only significant variable change was Wits appraisal, with ANB approaching statistical significance (p-value = 0.058). No other change was statistically significant.

For changes between T1 and T3, the difference between the control and treated patients was adjusted for the cephalometric measure at T1, age at T1, and time from T1 to T3. Change in overjet, Wits appraisal, and U1-NA all showed statistically significant differences between the two groups.

Table 9: Adjusted Comparison of Treated and Control Mean Change in Cephalometric Values from T1-T2, T2-T3, and T1-T3

<b>T1-T2 Changes</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	3.3 (2.5)	0.4 (0.8)	2.2	1.6, 2.8	<b>&lt;0.001</b>
ANB (degree)	1.0 (1.9)	0.0 (1.4)	0.54	-0.2, 1.3	0.15
SNA (degree)	0.9 (2.4)	0.3 (2.9)	0.83	-0.2, 1.8	0.11
SNB (degree)	-0.1 (2.3)	0.3 (2.6)	0.11	-0.8, 1.0	0.81
Wits (mm)	0.6 (3.2)	-0.1 (0.9)	-1.1	-2.2, 0.1	0.069
IMPA (degree)	-1.9 (4.8)	-0.3 (3.3)	-1.7	-3.4, 0.0	<b>0.048</b>
U1-L1 (degree)	-4.0 (8.9)	-1.6 (5.5)	-2.5	-5.2, 0.2	0.074
U1-NA (degree)	3.7 (7.2)	1.7 (4.7)	3.2	0.9, 5.5	<b>0.007</b>
MP (degree)	0.9 (3.2)	-0.3 (2.7)	0.77	-0.4, 1.9	0.18
LFH (%)	0.0 (2.0)	-0.2 (1.9)	0.31	-0.3, 1.0	0.34
<b>T2-T3 Changes</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	-1.2 (2.2)	-0.1 (0.6)	-0.12	-0.9, 0.7	0.77
ANB (degree)	-1.4 (2.3)	-0.9 (1.8)	-0.88	-1.8, 0.0	0.058
SNA (degree)	0.8 (1.8)	0.4 (2.7)	0.46	-0.4, 1.3	0.30
SNB (degree)	2.1 (2.1)	1.4 (2.1)	0.80	-0.1, 1.7	0.089
Wits (mm)	-1.4 (3.2)	-0.3 (0.9)	-2.0	-3.3, -0.7	<b>0.003</b>
IMPA (degree)	0.0 (4.3)	-0.8 (4.0)	0.38	-1.6, 2.4	0.71
U1-L1 (degree)	0.1 (7.7)	0.4 (6.2)	-0.70	-3.8, 2.4	0.65
U1-NA (degree)	1.0 (5.8)	1.5 (4.9)	1.5	-0.7, 3.7	0.17
MP (degree)	-1.6 (2.3)	-1.4 (2.4)	-0.22	-1.2, 0.8	0.66
LFH (%)	0.2 (1.5)	0.0 (1.4)	0.22	-0.4, 0.9	0.50
<b>T1-T3 Changes</b>					
<b>Variable</b>	<b>Treated<sup>1,3</sup></b>	<b>Control<sup>2,3</sup></b>	<b>Adjusted Difference<sup>4</sup></b>	<b>95% CI<sup>4,5</sup></b>	<b>p-value<sup>4,6</sup></b>
Overjet (mm)	2.0 (2.7)	0.2 (1.1)	0.94	0.25, 1.6	<b>0.008</b>
ANB (degree)	-0.2 (2.4)	-0.9 (1.6)	-0.01	-1.0, 1.0	0.98
SNA (degree)	1.3 (2.9)	0.8 (2.6)	0.66	-0.6, 1.9	0.28
SNB (degree)	1.5 (3.1)	1.8 (2.7)	0.39	-0.9, 1.7	0.55
Wits (mm)	-0.9 (3.5)	-0.3 (1.0)	-1.9	-3.4, -0.5	<b>0.010</b>
IMPA (degree)	-1.7 (4.7)	-0.9 (4.1)	-0.70	-2.7, 1.3	0.49
U1-L1 (degree)	-3.4 (9.4)	-1.2 (6.7)	-2.3	-5.6, 0.9	0.16
U1-NA (degree)	3.9 (7.2)	2.9 (5.2)	2.9	0.5, 5.2	<b>0.018</b>
MP (degree)	-0.3 (3.7)	-1.7 (3.2)	0.87	-0.7, 2.4	0.27
LFH (%)	0.1 (1.8)	-0.3 (1.7)	0.38	-0.4, 1.2	0.34

<sup>1</sup>N=82 <sup>2</sup>N=38

<sup>3</sup>Mean (SD)

<sup>4</sup>ANCOVA, <sup>5</sup>CI = Confidence Interval

<sup>6</sup>Bold indicates statistical significance

### *Surgical Evaluation of the Treated Sample*

Summaries for the decision for surgery are given below. Six patients without cephalometric measures at T3 were excluded from the analysis. When asked if they would

commit to initiating treatment now, the evaluators were not homogenous in their agreement. Of the 49 patients presented, evaluator 3 committed to treating five patients (10.2%), evaluator 1 committed to 17 patients (34.7%), and evaluator 2 committed to 45 (91.8%) patients.

The survey data was combined for the panelists to allow for analysis of commitment to treat. Unanimous or majority decisions were used for analysis. Panelists felt comfortable committing to starting comprehensive treatment for 18 patients (36.7%). Commitment to treat was not statistically different for males or females (p-value = 0.91).

The indication for surgery as the ideal treatment plan for the treated patients was more homogenous amongst evaluators. Evaluator 1 and 3 indicated surgery for 22 (44.9%) patients and evaluator 2 indicated surgery for 19 (38.8%). Data was combined for all evaluators in the same way as for commitment to treat. Complete or majority consensus among evaluators was summarized as the outcome for binary analysis. When combined, the committee indicated that surgical treatment would be the ideal plan for 21 patients (42.9%).

#### *Cephalometric Comparison between Surgical and Non-surgical Recommendation Groups*

Variables at T1 and T2 as well as the change in mean values from T1-T2 were also compared between the two groups, patients indicated for surgery and those who were not. At T1, the variables that are significantly different between the surgical and non-surgical groups are IMPA, Wits appraisal, and mandibular plane (Table 10). At T2, IMPA, Wits appraisal, and mandibular plane remained significantly different and ANB reached significance. When looking at the change from T1-T2, there were no significant differences between the two groups.

Table 10: Comparison Amongst Mean Cephalometric Values at T1, T2 and Change from T1-T2 for Surgical and Non-Surgical Patients

Variable	T1			T2			T2-T1		
	Yes <sup>2,4</sup>	No <sup>3,4</sup>	p-value <sup>1</sup>	Yes <sup>2,4</sup>	No <sup>3,4</sup>	p-value <sup>1</sup>	Yes <sup>2,4</sup>	No <sup>3,4</sup>	p-value <sup>1</sup>
Overjet (mm)	-0.4 (2.5)	-1.0 (2.0)	0.41	2.3 (1.9)	2.6 (1.3)	0.55	2.7 (2.9)	3.6 (2.1)	0.29
ANB (degree)	-1.1 (1.9)	-0.1 (1.7)	0.063	-0.1 (1.7)	1.1(2.3)	<b>0.041</b>	0.9 (2.0)	1.3 (1.3)	0.52
Wits(mm)	-5.8 (3.0)	-4.1 (2.6)	<b>0.037</b>	-5.4 (2.9)	-3.5 (1.9)	<b>0.012</b>	0.4 (3.1)	0.3 (2.5)	0.95
IMPA (degree)	82.2 (5.3)	88.6 (6.1)	<b>&lt;0.001</b>	79.7 (5.5)	86.7 (7.2)	<b>&lt;0.001</b>	-2.5 (4.8)	-1.6 (4.4)	0.50
U1-NA (degree)	26.2 (7.1)	24.1 (7.5)	0.33	29.0 (6.6)	26.9 (6.6)	0.30	2.8 (7.4)	3.0 (8.2)	0.93
LFH (%)	55.9 (2.0)	56.1 (1.9)	0.83	56.1 (1.9)	55.8 (2.2)	0.63	0.2 (1.8)	-0.2 (1.3)	0.40
SNA (degree)	79.2 (4.6)	81.0 (4.1)	0.17	80.2 (5.2)	80.7 (4.7)	0.75	1.0 (3.1)	0.2 (2.0)	0.32
SNB (degree)	81.0 (4.2)	80.2 (4.9)	0.55	80.4 (4.8)	79.6 (3.7)	0.55	0.1 (2.5)	-1.0 (1.9)	0.083
U1-L1 (degree)	134.2 (11.6)	135.8 (10.3)	0.60	132.8 (10.2)	131.6 (8.5)	0.67	-3.0 (8.9)	-3.6 (10)	0.85
MP (degree)	32.3 (5.3)	36.8 (6.0)	<b>0.009</b>	37.9 (5.5)	33.9 (5.9)	<b>0.027</b>	1.0 (2.5)	1.6 (3.3)	0.53

<sup>1</sup>Welch Two Sample t-test

<sup>2</sup>N=2, <sup>3</sup>N=28

<sup>4</sup>Mean (SD)

#### *Associated Cephalometric Variables with the Need of Surgery*

A univariate logistic regression was used to seek variables predictive for surgery at all three timepoints. Table 11 reports the odds ratio, 95% confidence interval and p-value for the cephalometric variables included in the univariate model. The odds ratio is reported in units of clinical standard deviation for each variable and describes the relationship between the variable and need for surgery if that variable is increased by one unit of standard deviation. At T1, Wits appraisal, IMPA and LFH % were significantly associated with the need for surgery. At T2, Wits, IMPA, LFH% remain associated while ANB and mandibular plane become associated. Finally, at T3, mandibular plane is no longer significantly associated while U1-NA and overjet become associated with the need for surgery.

Table 11: Univariate Logistic Regression Evaluating Variables Associated with the Need for Surgery

Variable	T1 (N=49)			T2 (N=44)			T3 (N=49)		
	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>
Overjet (SD=2.5 mm)	1.3	0.7, 2.6	0.38	0.7	0.3, 1.9	0.54	0.1	0.0, 0.3	<b>0.002</b>
ANB (SD=1.5 degree)	0.7	0.5, 1.0	0.065	0.7	0.5, 1.0	<b>0.051</b>	0.5	0.3, 0.7	<b>&lt;0.001</b>
Wits (SD=1 mm)	0.8	0.6, 1.0	<b>0.040</b>	0.7	0.5, 0.9	<b>0.020</b>	0.5	0.3, 0.7	<b>&lt;0.001</b>
IMPA (SD=7 degree)	0.2	0.1, 0.5	<b>0.003</b>	0.2	0.1, 0.5	<b>0.58</b>	0.3	0.1, 0.6	<b>0.004</b>
U1-NA (SD=5.7 degree)	1.3	0.8, 2.0	0.32	1.3	0.8, 2.4	0.29	2.1	1.2, 4.2	<b>0.024</b>
MP (SD=6 mm)	2.5	1.3, 5.7	0.82	2.1	1.1, 4.8	<b>0.039</b>	1.6	0.9, 3.0	0.10
SNA (SD=3.5 degree)	0.7	0.4, 1.1	0.16	0.9	0.6, 1.4	0.74	0.8	0.5, 1.2	0.29
SNB (SD=3.4 degree)	0.9	0.6, 1.4	0.53	1.2	0.7, 2.0	0.54	1.3	0.9, 2.0	0.25
U1-L1 (SD=7 degree)	1.1	0.8, 1.6	0.60	1.1	0.7, 1.8	0.66	1.2	0.8, 1.8	0.52
LFH % (SD=0.35%)	1.0	0.9, 1.1	<b>0.015</b>	2.1	1.1, 4.8	<b>0.039</b>	1.1	1.0, 1.3	<b>0.027</b>

<sup>1</sup>OR = Odds Ratio, CI = Confidence Interval

<sup>2</sup>Reported in Standard Deviation Units

<sup>3</sup>Bold indicates statistical significance

Due to high correlation among the included variables, a multivariate model was created with the following variables: overjet, ANB, Wits appraisal, IMPA, U1-NA, Mandibular plane, and LFH. The results of this model are reported in Table 12. In the multivariate model, the only variable significantly associated at both T1 and T2 is IMPA. At T3, overjet and LFH % were also become significant.

Table 12: Multivariate Logistic Regression evaluating Variables Associated with the Need for Surgery

Variable	T1 (N=49)			T2 (N=44)			T3 (N=49)		
	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>	OR <sup>1,2</sup>	95% CI <sup>1</sup>	p-value <sup>3</sup>
Overjet (SD=2.5 mm)	0.8	0.2, 2.6	0.73	0.7	0.2, 2.9	0.65	0.0	0.0, 0.3	<b>0.018</b>
ANB (SD=1.5 degree)	0.9	0.5, 1.7	0.73	1.0	0.6, 1.7	0.95	1.3	0.5, 3.4	0.53
Wits (SD=1 mm)	0.9	0.6, 1.3	0.54	0.8	0.5, 1.2	0.31	0.7	0.3, 1.2	0.24
IMPA (SD=7 degree)	0.2	0.0, 0.8	<b>0.032</b>	0.2	0.0, 0.7	<b>0.017</b>	0.1	0.0, 0.9	<b>0.083</b>
U1-NA (SD=5.7 degree)	1.5	0.7, 3.3	0.35	2.2	0.9, 6.0	0.11	3.8	0.8, 3.0	0.11
MP(SD=6 mm)	2.0	0.8, 5.7	0.16	1.3	0.4, 4.2	0.68	1.0	0.2, 3.9	0.92
LFH % (SD=0.35%)	1.0	0.9, 1.2	0.77	1.1	0.9, 1.2	0.49	1.3	1.0, 1.8	<b>0.047</b>

<sup>1</sup>OR = Odds Ratio, CI = Confidence Interval

<sup>2</sup>Reported in Standard Deviation Units

### *Associated Non-Cephalometric Variables with the Need of Surgery*

Non-cephalometric variables (CVM stage, clinical evaluation of growth status, Angle classification, profile pattern, compliance, presence of a functional shift, and ethnicity) evaluated at T1 did not show any significant association with surgical need (Appendix 7).

## **DISCUSSION:**

### *Treatment Outcomes during Early Treatment with a Facemask*

This study showed significant and favorable maxillary skeletal change during treatment as indicated by an increase in SNA and ANB. The mean change in overjet was positive leading to correction of anterior crossbite. There was also significant proclination of the maxillary incisors and retroclination of the lower incisors which contributed to the positive change in overjet and masked the skeletal disharmony of the maxilla and mandible.

Several previous studies, including a randomized controlled trial (Mandal *et al.*, 2010) and a systematic review (Kim *et al.*, 1999) report cephalometric changes following facemask therapy in the short term. Table 13 shows the mean change in variables between the three studies.

Table 13: Comparison of Cephalometric Outcome Variables after Facemask Treatment from the present study, Mandall *et al.*, 2010 and Kim *et al.* 1999

<u>Variable</u>	<u>Present study</u>	<u>Mandall <i>et al.</i></u>	<u>Kim <i>et al.</i></u>
Overjet	3.3 (2.5)	4.4 (2.7)	** <sup>1</sup>
SNA	0.9 (2.4)	1.4 (2.1)	1.7 (1.6)
SNB	-0.1 (2.3) <sup>1</sup>	-0.7 (1.5)	-1.2 (1.2)
ANB	1.0 (1.9)	2.1 (2.3)	2.8 (1.9)
Wits	0.6 (3.2)	** <sup>1</sup>	4.7 (5)
IMPA	-1.9 (4.8)	-4.9 (4.1)	-2.9 (7.2)
U1-L1	-4.0 (8.9)	-0.4 (6.8) <sup>1</sup>	** <sup>1</sup>

<sup>1</sup>Not reported

The randomized controlled trial showed more significant change for all variables except interincisal angle. The differences seen between the studies could be attributed to many factors

including patient ethnicity, age, and compliance. All of the patients treated in the Mandall study were of Caucasian descent.. In Kim's systematic review, studies with either Caucasian or Asian decent were included. Present study included a wider variety of ethnicities, with 31.7% Asian, 45.1% Caucasian and the remainder being African American, Hispanic, Middle Eastern, or a combination of ethnicities. The age range of treated subjects in this study was also older than the other two studies and the treated sample included patients who were sub-optimal or non-compliant in wearing their facemask. The difference in interincisal angle could be attributed more significant proclination of the maxillary incisors compared to Mandall's work.

The treated sample in this study showed an increase in mandibular plane angle during treatment, whereas the control group had a decrease. The mean difference between the two groups did not reach statistical significance, but the trend is similar to other studies that show a downward and backward rotation of the maxilla and subsequent increase in mandibular plane angle (Ngan *et al.*, 1996, Turley *et al.*, 1998, Kapust *et al.*, 1998, Kim *et al.*, 1999, Vaughn *et al.*, 2005, Seehra *et al.*, 2011 Jager *et al.* 2001, Foersch 2015, Zhang 2015). This change is especially important when treating patients with concurrent anterior open bites. Clinicians need to ensure that the force direction of the facemask does not induce even more clockwise rotation of the maxilla and an increase in the mandibular plane angle.

When comparing this treated sample to the control group, the mean differences in all skeletal variables were not significantly different from T1 to T2. This indicates that some of the skeletal change we see from T1 to T2 in the treated group is attributed to normal skeletal growth.

The change in the mean dental variables, however, was statistically different between the treated and control groups. There was more retroclination of the mandibular incisors and proclination of the maxillary incisors for the treated group. Zionic Alexander *et al.* reported that

Class III patients showed some proclination of the maxillary incisors (about 1° per year) and retroclination of the lower incisors compared to Class I patients. They attributed these changes to the Class III growth pattern. Although some of the dental change may be due to growth, there was still more proclination of the maxillary incisors for our treated group than was reported in Zionix Alexander et al.'s growth study. The difference can be attributed to the treatment effects of the facemask. There is significantly more overjet correction for the treated group as well but since there is no significant skeletal difference, this change is attributed to dental compensation and not a skeletal effect of the appliance on the maxilla.

Similar to the previously published literature, we saw a favorable skeletal effects and an increase in overjet for our treatment group. In determining treatment success however, it is imperative to also evaluate the sustainability of this change following treatment.

#### *Factors Contributing to Relapse Following Facemask Treatment*

In this study, the average amount of time between T2 and T3 for our treated group was 3.1 years and the mean age for treated patients at T3 was 14.3 years. When comparing the untreated controls from the T2-T3, there were no significant differences in change in SNA, SNB, and ANB demonstrating similar growth and development patterns. The position of the dentition remained stable for both groups indicating minimal dental relapse for the treated group. This is most likely attributed to retention protocols following treatment.

Focusing on the treated group, both SNA and SNB significantly increased from T2-T3. However, SNB increased more than SNA, resulting in a significant reduction in ANB and overjet. This is consistent with the growth findings reported by Zionix Alexander *et al.* as they

also reported minimal change in SNA but a significant increase in SNB related to continued mandibular growth.

The mean changes in IMPA, interincisal angle, and U1-NA (dental related variables) from T2-T3 remained stable, likely attributed to retention protocols. Even though dental positions were maintained, there was still a reduction in overjet. This change can be attributed to skeletal change as ANB also significantly decreased. The skeletal relapse can be attributed to mandibular growth as SNB increased and SNA remained constant. Mandibular plane also decreased from T2 to T3 which can be attributed to growth and relapse.

#### *Need for Orthognathic Surgery following Facemask Therapy*

The panelists felt it was clinically appropriate to begin further treatment (whether surgical or non-surgical) for only 18 of the 49 patients presented. We might infer that they believed there may still be growth potential for the remaining 31 patients. This is important to consider when evaluating the surgical data, because the recommendation for surgery may change for some patients depending on future growth. If growth continues. The evaluators indicated that for 43% of patients, ideal future orthodontic treatment would involve orthognathic surgery. This is higher than the previous reports in the literature of 33-38% of patients needing surgery (Hagg *et al.*, 2003, Baccetti *et al.*, 2004, Wells *et al.*, 2006, Masucci *et al.*, 2011, & Mandall *et al.*, 2016). This may be explained by differences in ethnicity and age as well as the approach to surgical evaluation.

#### *Predictive Variables Associated with Surgical Need*

We evaluated associated variables at T3 with the need for surgery to determine what might be the driving force behind a provider's recommendation for surgery. Univariate

modeling showed that decreased overjet, ANB, Wits appraisal, and IMPA as well as increased U1-NA and lower facial height were associated with a higher need for surgery. A multivariate model was created to address high correlation between the variables of interest. In the multivariate model, only decreased overjet and increased lower facial height were associated with the need for surgery. A small sample size and highly correlated variables make interpretation difficult. Based on the two models, variables at T3 that are most highly associated with the need for surgery include overjet, IMPA, U1-NA and lower facial height. In addition to being predictive, a 5.7° degree increase in U1-NA increases the need for surgery by 1.45 times. Increase in IMPA and overjet are protective in regard to the need for surgery when interpreting the reported odds ratios. Wits appraisal and ANB are also likely important but are highly correlated with other variables. It is critically important to look at the position and compensation of the incisors when evaluating the need for orthognathic surgery. Increased lower facial height may be indicative of corresponding anterior open bite which also might drive the decision for surgery.

Identifying specific characteristics associated with need for future surgery at an early age would improve prediction of facemask therapy success. Our data showed that the more retroclined or compensated the lower incisors are at a young age, the higher chances the changes that the patient will need future surgery. However, decreased Wits appraisal and increased mandibular plane angle were also associated similar to the values at T3. Due to the high correlations amongst all variables, their unique contribution is difficult to discern. Increased mandibular plane angle may also be associated with a concurrent anterior open bite, which may also be a driving factor for the future need of surgery.

The mean cephalometric changes from T1 to T2 were also evaluated for association with surgical need. None of the treatment changes were significantly associated with future need for surgery. However, there is likely due to insufficient power to accurately discern any relationship.

Overall, we found a significant number of patients treated with a facemask who were judged to require orthognathic surgery post-treatment. We were not able to judge the surgical need for the matched control sample because of inadequate available records.

#### *When Should Facemask Therapy Be Recommended? – Clinical and Ethical Analysis*

Is facemask therapy a benefit and is early treatment “worth it” for a patient? Will facemask treatment prevent future need for orthognathic surgery? According to Mandall *et al.* “evidence would support an orthodontist offering early protraction facemask treatment, because there is always the chance that a patient will respond extremely favorably, whilst cautioning that such change is not guaranteed.” (Mandall *et al.*, 2010) While it may be true that some patients will respond favorably, our results showed that the favorable changes were primarily due to compensation of the incisors and not significant skeletal change.

Facemask therapy may reduce the complexity of the surgical procedures or limit the case to “one jaw” instead of two. This idea has no support in the literature and the claim is solely based on expert opinion. Orthognathic surgery comes with more inherent risks than orthodontic treatment alone but may also provide additional facial esthetic benefits for a patient.

The economic burden of treatment impacts all involved parties. For the patient, this includes the indirect cost of missing school (Bresnahan, *et al.*, 2010). For the family, there may be direct costs of the orthodontic treatment as well as indirect costs of taking time off work, transportation, childcare for other children in the family. This is especially important in low-

income families where work may not allow flexibility of schedules and transportation may be an additional burden (Bresnahan, *et al.*, 2010).

For the patients funded by public insurance, there are several economic factors to consider. In Washington State, anterior crossbite and “skeletal Class III” problems are weighted heavily as medically necessary malocclusions and many patients who could benefit from a facemask are often approved for early care. In many instances, early treatment can reduce the overall severity of the malocclusion to an elective status (Jolley *et al.*, 2010). This may not be true Class III patients who may still need surgery in the future. If an early treatment is not successful in correcting medically necessary malocclusion and the patient requires orthognathic surgery, the system is again responsible for repeat orthodontic treatment as well as surgical costs.

Surgical intervention also carries significant burdens for all of those involved. These include direct costs, many appointments with the orthodontist and surgical team, inherent surgical risks, and a challenging healing period. In some cases, positive overjet and anterior coupling can be achieved through dental compensation. Although the dentition may not be in the most ideal position, surgery and its inherent risks are prevented.

Psychosocial patient benefit is another important factor to consider when deciding on treatment. The literature does support that orthodontic treatment can positively affect a patient’s overall quality of life (King *et al.*, 2011). In their 2010 paper, Mandall *et al.* assessed the psychosocial outcomes and found minimal improvement in self-esteem after treatment compared to untreated controls.

When considering early treatment with facemask therapy, providers need to consider if the overall benefit will outweigh the burdens and if this treatment is in the best interest for each

individual patient. For some patients, early treatment may prevent the future need for surgery. For those who with severe skeletal discrepancy, surgery may be inevitable.

### *Limitations*

Our retrospective cohort study sample size is relatively small, and only powered to detect changes between the treated group and control group. The exploration of variables associated with the need for future surgery was likely underpowered. Our sample was also younger than providers would typically assess when judging the need for surgery, and the prospect of continued growth may have clouded the surgical decision.

There was also likely selection bias in our treated sample. The most severe cases eligible for facemask therapy may not have been included at T1 because providers decided against attempting early facemask therapy. Additionally, the sample available at T3 for surgical evaluation may not have included successful facemask patients who never presented for further treatment. Finally, some of the patients were lost to follow-up after early treatment and also not included in the surgical evaluation.

The selection and matching of control subjects was difficult due to the limited availability of untreated controls similar to the treatment sample and the variable quality of the historical control records. Additionally, inherent differences likely exist between the two groups, including secular changes in growth, ethnic differences, and limited number of available Class III controls.

Many of the variables of interest are highly correlated with each other, making it challenging to determine which ones are contributing most to the predicted need for surgery. Surgical determination in our sample could also be due to other unexamined factors including facial features, or other concurrent malocclusion like anterior open bite.

**CONCLUSIONS:**

1. Facemask therapy induces more dental incisor compensation than skeletal change, resulting in less correction of the midface deficiency and more masking of skeletal disharmony.
2. Relapse following early treatment can be attributed to continued growth of the mandible.
3. Overjet, IMPA, U1-NA and lower facial height appear to drive the decision about subsequent surgical or non-surgical treatment plan.
4. Need for future surgical treatment is associated with increased retroclination of the lower incisors at a young age.

## LITERATURE CITED

- Baccetti T., Tollaro I., (1998) "A retrospective comparison of functional appliance treatment of Class III malocclusion in the deciduous and mixed dentitions," *Eur J Orthod*, 20, pp. 309-317.
- Baccetti T., McGill J. S., Franchi L., McNamara J. A., Tollaro I., (1998) "Skeletal effects of early treatment of Class III malocclusion with maxillary expansion and face-mask therapy," *Am J Orthod Dentofacial Orthop*, 113, pp. 333-343.
- Baccetti T., Franchi L., McNamara J. A. (2000) "Treatment and posttreatment craniofacial changes after rapid maxillary expansion and facemask therapy," *Am J Orthod Dentofacial Orthop*, 118, pp. 404-413.
- Baccetti T., Franchi L., McNamara J. A., (2004) "Cephalometric variables predicting the long term success or failure of combined rapid maxillary expansion and facial mask therapy," *Am J Orthod Dentofacial Orthop*, 126, pp. 16-22.
- Baik H.S., "Clinical results of maxillary protraction in Korean children," *Am J Orthod Dentofacial Orthop*, 108, pp. 583-92.
- Campbell P., (1983) "The dilemma of Class III treatment – Early or Late?," *Angle Orthod*. 53, pp. 176-191.
- Chong Y-H., Ive J., Artun J. (1996) "Changes following the use of protraction headgear for early correction of Class III malocclusion," *Angle Orthod*, 66, pp. 351-362.
- Cozza P., Marino A., Mucedero M. (2004) "An orthopaedic approach to the treatment of Class III malocclusion in the early mixed dentition," *Eur J Orthod*, 26, pp. 191-199.
- Hata S., Itoh T., Nakagawa M., Kamogashi., Ichikawa K., Matsumoto M., *et al.* (1987) "Biomechanical effects of maxillary protraction on the craniofacial complex," *Am J Orthod Dentofacial Orthop*, 91, pp. 583-92.
- Hagg U., Tse A., Bendeus M., Rabie A. B. M., (2003) "Long-term follow-up of early treatment with reverse headgear," *Eur J Orthod*, 25, pp. 95-102.
- Jacobson A., Evans W.B., Preston C.B., (1974) Sadowsky P. L., "Mandibular prognathism," *Am J Orthod*, 66, pp. 140-171.
- Jolley C. J., Huang G. J., Greenlee G.M., Spiekerman C., Kiyak H.A., King G.J., (2010) "Dental effects of interceptive orthodontic treatment in a Medicaid population: Interim results from a randomized clinical trial," *Am J Orthod Dentofacial Orthop*, 137, pp. 324-333.
- Kapust A.J., Sinclair P. M., Turley P.K., (1998) "Cephalometric effects of facemask/expansion therapy in Class III children: a comparison of three age groups," *Am J Orthod Dentofacial Orthop*, 113, pp. 204-12.

- Kim J-H., Viana M. A. G., Graber T. M., Omerza F. F., and Begole E. A., (1999) "The effectiveness of protraction face mask therapy: A meta-analysis," *Am J Orthod Dentofacial Orthop*, 115, pp. 675-85.
- Macdonald K. E., Kapust A. J., Turley P. K., (1999) "Cephalometric changes after the correction of Class III malocclusion with maxillary expansion/facemask therapy," *Am J Orthod Dentofacial Orthop*, 116, pp. 13-24.
- Mandall N.A., DiBiase A., Littlewood S., Nute S., Stivaros N., McDowall R., Shargill I., Worthington H., Cousley R., Dyer F., Mattick., & Doherty B., (2010) "Is early Class III protraction facemask treatment effective? A multicentre randomized, controlled trial: 15-month follow-up," *J. Orthod*, 37, pp 149-161.
- Mandall N.A., Mandall N.A., Cousley R., DiBiase A., Dyer F., Littlewood S., Mattick R., Nute S., Doherty B., Stivaros N., McDowall R., Shargill I., Ahmad A., Walsh T., & Worthington H., (2012) "Is early Class III protraction facemask treatment effective? A multicentre randomized, controlled trial: 3-year follow-up," *J. Orthod*, 39, pp 176-185.
- Mandall N.A., Cousley R., Dibiasa A., Dyer F., Littlewood., Mattick R., Nute S. J., Doherty., Stivaros., McDowall R., Shargill I., Worthington H. V., (2016) "Early Class III protraction facemask treatment reduces the need for orthognathic surgery: multi-centre, two-arm parallel randomized, controlled trial," *J. Orthod*, 43, pp 164-175.
- Masucci C., Franchi L., Defraia E., Mucedero., Cozza P., Baccetti T., (2011) "Stability of rapid maxillary expansion and facemask therapy: A long-term controlled study," *Am J Orthod Dentofacial Orthop*, 140, pp. 493-500.
- McNamara J.A, Franchi L. (2018) "The cervical vertebral maturation method: A user's guide," *Angle Orthod*, 88 (2), pp. 133-143.
- Mellion Z. J., Behrents R. G., Johnston L. E. (2013) "The pattern of facial skeletal growth and its relationship to various common indexes of maturation," *Am J Orthod Dentofacial Orthop*, 143, pp. 845-854.
- Nartallo-Turley P., Turley P. (1998) "Cephalometric effects of combined palatal expansion and facemask therapy on Class III malocclusion," *Angle Orthod*, 68, pp. 217-224.
- Ngan P.W., Hagg U., Yiu C., Merwin D., Wei Shy., (1996) "Treatment response to maxillary expansion and protraction," *Eur J Orthod*, 18, pp. 151-68.
- Ngan P., Moon W., (2015) "Evolution of Class III treatment in orthodontics," *Am J Orthod Dentofacial Orthop*, 148, pp. 22-36.
- Perinetti P., Primozić J., Sharma B., Cioffi I., Contardo L. (2018) "Cervical vetebral maturation method and mandibular growth peak: a longitudinal study of diagnostic reliability," *Eur J Orthod*, 40 (6), pp 666-672.

- Predko-Engel A., Kaminek M., Langova K., Kowalski P., Fudalej P. S. (2015) "Reliability of the cervical vertebrae maturation method," *Bratisl Lek Listy*, 116 (4), pp. 222-226.
- Proffit W.R., Fields W. F., Sarver D.M. (2013) *Contemporary Orthodontics*, 5<sup>th</sup> ed. Mosby, Saint Louis, MO.
- Seehra J., Fleming P.S., Mandall N., DiBiase A.T. (2011) "A comparison of two different techniques for early correction of Class III malocclusion," *Angle Orthod*, 82, pp. 96 – 101.
- Shanker S., Ngan P., Wade D., Beck M. Yiu C., Hagg U, et all (1996) "Cephalometric A-point changes during and after maxillary protraction and expansion," *American Journal of Orthodontics and Dentofacial Orthopedics*, 110, pp 423-430.
- Sung J. S., Baik H. S. (1998) "Assessment of skeletal and dental changes by maxillary protraction," *Am J Orthod Dentofacial Orthop*, 114 pp 492-502.
- Takada K, Petdachai S, Sakuda M (1993) "Changes in dentofacial morphology in skeletal Class III children treated by a modified maxillary protraction headgear and a chin cup: a longitudinal cephalometric appraisal," *European Journal of Orthodontics*, 15, pp 211-221.
- Vaughn A., Mason B., Moon H-B., Turley P. K., (2005) "The effects of maxillary protraction therapy with or without rapid palatal expansion: A prospective, randomized clinical trial," *Am J Orthod Dentofacial Orthop*, 128, pp. 299-309.
- Watkinson S., Harrison J.E., Furness S., Worthington H.V., (2013) "Orthodontic treatment for prominent lower front teeth (Class III malocclusion) in children," *Cochrane Database of Systematic Reviews*, 9, pp
- Wells A. P., Sarver D. M., Proffit W.R. (2006) "Long-term efficacy of reverse pull headgear therapy," *Angle Orthod*, 76, pp. 915-922.
- Westwood P. V., McNamara J. A., Baccetti T., Franchi L., Sarver D. M., (2003) "Long term effects of Class III treatment with rapid maxillary expansion and facemask therapy followed by fixed appliances," *Am J Orthod Dentofacial Orthop*, 123, pp. 306-320.
- Williams S., Anderson C.E., (1986) "The morphology of the potential Class III skeletal pattern in the growing child," *Am J Orthod Dentofacial Orthop*, 89, pp. 302-311.
- Woon S. C., Thiruvengkatachari B., (2011) "Early orthodontic treatment for Class III malocclusion: A systematic review and meta-analysis," *Am J Orthod Dentofacial Orthop*, 151, pp. 28-52.
- Wylie W. L., Johnson E. L. (1952) "Rapid evaluation of facial dysplasia in the vertical plane," *Angle Orthod*, 22 (3), pp. 165-182.

- Zere E., Chaudhari P.K., Sharan J., Dhingra K., Tiwari N., (2018) “Developing Class III malocclusions: Challenges and Solutions,” *Clinical, Cosmetic, and Investigational Dentistry*, 10, pp. 99-116
- Zhang W., Qu H-C., Yu M., Zhang Y., (2015) “The effects of maxillary protraction with or without rapid maxillary expansion and age factors in treating Class III malocclusion: A meta-analysis,” *PloS One*, 10(6):e0130096.
- Zionic Alexandar A.E., McNamara J.A., Franchi L., Baccetti T. (2009) “Semilongitudinal cephalometric study of craniofacial growth in untreated Class III malocclusion,” *Am J Orthod Dentofacial Orthop*, 135 (6), pp. 700.e1-700.e14

## APPENDIX

Appendix 1: Reliability of average of 2 Cephalometric measurements

Measure	1st Mean (SD)	2nd Mean (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Overjet	0.8 (2.7)	0.8 (2.7)	0.0 (0.4) [-0.2, 0.1]	0.99 (0.98, 1.00)	0.3 (0.0, 0.7)
ANB	-0.9 (2.8)	-0.8 (2.8)	-0.1 (1.0) [-0.5, 0.3]	0.94 (0.87, 0.97)	0.7 (0.0, 4.2)
SNA	82.2 (6.1)	82.2 (5.7)	0.0 (1.7) [-0.6, 0.6]	0.96 (0.92, 0.98)	1.1 (0.0, 6.1)
SNB	83.2 (5.5)	83.1 (5.2)	0.1 (0.8) [-0.2, 0.4]	0.99 (0.98, 0.99)	0.6 (0.0, 2.6)
Wits	-5.8 (3.1)	-5.5 (3.2)	-0.4 (1.0) [-0.7, 0.0]	0.95 (0.89, 0.97)	0.7 (0.0, 2.7)
IMPA	84.1 (7.1)	83.9 (6.5)	0.2 (1.7) [-0.5, 0.8]	0.97 (0.93, 0.98)	1.2 (0.0, 4.2)
U1-L1	131.0 (8.0)	131.1 (8.0)	-0.1 (2.3) [-0.9, 0.7]	0.96 (0.92, 0.98)	1.6 (0.0, 5.2)
U1-NA	29.0 (6.9)	29.3 (6.8)	-0.3 (2.1) [-1.0, 0.5]	0.95 (0.91, 0.98)	1.5 (0.0, 4.5)
MP	33.6 (4.6)	33.4 (4.4)	0.2 (0.7) [0.0, 0.5]	0.99 (0.98, 0.99)	0.5 (0.0, 1.9)
LFH	56.4 (2.0)	56.4 (2.1)	0.0 (0.6) [-0.2, 0.3]	0.95 (0.90, 0.98)	0.5 (0.0, 2.0)

Appendix 2: Cephalometric Intra-rater reliability for Evaluator Hagel

Measure	1st Mean (SD)	2nd Mean (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Overjet	0.8 (2.7)	0.7 (2.8)	0.0 (0.4) [-0.1, 0.2]	0.99 (0.98, 1.00)	0.3 (0.0, 1.0)
ANB	-0.9 (2.7)	-0.9 (2.8)	0.0 (1.1) [-0.5, 0.4]	0.92 (0.84, 0.96)	0.8 (0.0, 4.7)
SNA	82.2 (6.3)	82.1 (5.7)	0.0 (2.0) [-0.7, 0.8]	0.94 (0.89, 0.97)	1.4 (0.1, 7.1)
SNB	83.1 (5.6)	83.0 (5.4)	0.1 (1.1) [-0.3, 0.5]	0.98 (0.96, 0.99)	0.8 (0.1, 2.5)
Wits	-5.8 (3.1)	-5.5 (3.2)	-0.3 (1.0) [-0.7, 0.0]	0.94 (0.88, 0.97)	0.8 (0.1, 2.7)
IMPA	83.3 (6.8)	83.8 (6.5)	-0.5 (2.1) [-1.3, 0.3]	0.95 (0.89, 0.97)	1.5 (0.1, 4.7)
U1-L1	131.7 (7.1)	131.5 (8.1)	0.2 (2.9) [-0.9, 1.2]	0.93 (0.86, 0.97)	2.0 (0.0, 6.6)
U1-NA	29.0 (6.4)	29.1 (6.8)	-0.1 (2.7) [-1.1, 0.9]	0.92 (0.84, 0.96)	1.9 (0.3, 5.4)
MP	33.8 (4.8)	33.4 (4.5)	0.4 (1.1) [0.0, 0.8]	0.97 (0.94, 0.99)	0.8 (0.0, 2.8)
LFH	56.3 (1.9)	56.3 (2.2)	0.0 (0.6) [-0.2, 0.2]	0.95 (0.90, 0.98)	0.5 (0.0, 1.6)

Appendix 3: Cephalometric Intra-rater reliability for Evaluator Bayirli

Measure	1st Mean (SD)	2nd Mean (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Overjet	0.8 (2.7)	0.9 (2.7)	-0.1 (0.5) [-0.3, 0.1]	0.98 (0.96, 0.99)	0.4 (0.0, 1.1)
ANB	-0.9 (2.8)	-0.8 (2.8)	-0.2 (1.0) [-0.6, 0.2]	0.93 (0.86, 0.97)	0.7 (0.1, 3.7)
SNA	82.3 (6.0)	82.3 (5.7)	0.0 (1.6) [-0.6, 0.6]	0.96 (0.93, 0.98)	1.1 (0.1, 5.1)
SNB	83.2 (5.4)	83.1 (5.0)	0.1 (1.0) [-0.2, 0.5]	0.98 (0.96, 0.99)	0.7 (0.0, 3.2)
Wits	-5.8 (3.1)	-5.5 (3.3)	-0.4 (1.0) [-0.8, 0.0]	0.94 (0.88, 0.97)	0.8 (0.1, 2.7)
IMPA	84.9 (7.7)	84.0 (6.6)	0.9 (2.3) [0.0, 1.7]	0.94 (0.88, 0.97)	1.7 (0.2, 6.3)
U1-L1	130.3 (9.0)	130.7 (8.1)	-0.4 (3.0) [-1.5, 0.7]	0.94 (0.88, 0.97)	2.1 (0.1, 6.2)
U1-NA	29.1 (7.4)	29.5 (7.0)	-0.4 (2.3) [-1.3, 0.4]	0.95 (0.90, 0.98)	1.6 (0.1, 5.0)
MP	33.4 (4.4)	33.4 (4.3)	0.0 (0.9) [-0.3, 0.3]	0.98 (0.96, 0.99)	0.6 (0.1, 1.9)
LFH	56.5 (2.0)	56.5 (2.1)	0.1 (0.8) [-0.2, 0.4]	0.93 (0.85, 0.96)	0.6 (0.0, 2.4)

Appendix 4: Cephalometric Inter-rater reliability for Hagel and Bayirli

Measure	1st Mean <sup>1</sup> (SD)	2nd Mean <sup>2</sup> (SD)	Diff. (SD) [95% CI]	ICC (95% CI)	D. Error (Min, Max)
Overjet	0.8 (2.7)	0.8 (2.7)	-0.1 (0.5) [-0.3, 0.1]	0.99 (0.97, 0.99)	0.3 (0.0, 1.2)
ANB	-0.9 (2.7)	-0.9 (2.8)	0.0 (0.6) [-0.2, 0.3]	0.97 (0.95, 0.99)	0.4 (0.0, 1.3)
SNA	82.2 (6.3)	82.3 (6.0)	-0.1 (1.1) [-0.5, 0.3]	0.98 (0.97, 0.99)	0.8 (0.0, 2.0)
SNB	83.1 (5.6)	83.2 (5.4)	-0.2 (1.1) [-0.6, 0.2]	0.98 (0.96, 0.99)	0.8 (0.0, 2.7)
Wits	-5.8 (3.1)	-5.8 (3.1)	0.0 (0.4) [-0.1, 0.2]	0.99 (0.98, 1.00)	0.3 (0.0, 0.8)
IMPA	83.3 (6.8)	84.9 (7.7)	-1.6 (2.3) [-2.4, -0.7]	0.93 (0.86, 0.97)	1.9 (0.0, 6.0)
U1-L1	131.7 (7.1)	130.3 (9.0)	1.4 (2.9) [0.3, 2.4]	0.93 (0.85, 0.96)	2.2 (0.0, 5.5)
U1-NA	29.0 (6.4)	29.1 (7.4)	-0.1 (1.9) [-0.8, 0.6]	0.96 (0.93, 0.98)	1.3 (0.0, 3.7)
MP	33.8 (4.8)	33.4 (4.4)	0.4 (1.3) [-0.1, 0.9]	0.95 (0.91, 0.98)	1.0 (0.0, 3.7)
LFH	56.3 (1.9)	56.5 (2.0)	-0.2 (0.5) [-0.4, 0.0]	0.96 (0.92, 0.98)	0.4 (0.0, 0.9)

<sup>1</sup>Hagel  
<sup>2</sup>Bayirli

Appendix 5: Categorical Variable Reliability

Variable	Percent Agreement	Kappa Statistic
CVM	80%	0.65
Angle Classification	85.7%	0.72
Overjet	92.9%	0.86
Profile Pattern	66.7%	0.43
Growth Status	73.3%	0.49

Appendix 6: Clinical Data Summary for Treated Subjects

<b>Variable</b>	<b>T1 N = 82<sup>1</sup></b>	<b>T2 N = 82<sup>1</sup></b>	<b>T3 N = 82<sup>1</sup></b>
<b>Cervical Vertebrae Maturation Stage</b>			
1,2	41 (51%)	26 (36%)	1 (1.9%)
3,4	38 (48%)	40 (55%)	28 (54%)
5,6	1 (1.2%)	7 (9.6%)	23 (44%)
Unknown	2	9	30
<b>Angle Classification</b>			
I	25 (33%)	31 (40%)	17 (29%)
II	6 (8.0%)	9 (12%)	2 (3.4%)
III	44 (59%)	37 (48%)	39 (67%)
Unknown	7	5	24
<b>Overjet</b>			
(-/-)	7 (9.3%)	0 (0%)	3 (5.2%)
(-)	45 (60%)	5 (6.3%)	7 (12%)
(0)	14 (19%)	1 (1.3%)	10 (17%)
(+)	9 (12%)	73 (92%)	38 (66%)
Unknown	7	3	24
<b>Profile Pattern</b>			
Concave	38 (47%)	20 (25%)	22 (37%)
Orthognathic	26 (32%)	39 (49%)	24 (41%)
Convex	17 (21%)	20 (25%)	13 (22%)
Unknown	1	3	23
<b>Clinical Evaluation of Growth Status</b>			
Pre-Pubertal	73 (90%)	49 (62%)	3 (5.1%)
Pubertal	8 (9.9%)	27 (34%)	45 (76%)
Post-Pubertal	0 (0%)	3 (3.8%)	11 (19%)
Unknown	1	3	23

<sup>1</sup>n (%)

## Appendix 7: Need for Surgery by Clinical Variables

Need for surgery			
Characteristic	No	Yes	p-value <sup>1</sup>
<b>CVM</b>			0.58
1,2	16 (55%)	13 (45%)	
3,4	12 (63%)	7 (37%)	
<b>Growth status</b>			>0.99
Pre-pubertal	25 (58%)	18 (42%)	
Pubertal	3 (60%)	2 (40%)	
<b>Angle Classification</b>			0.24
I	8 (73%)	3 (27%)	
II	1 (25%)	3 (75%)	
III	13 (48%)	14 (52%)	
<b>Profile</b>			0.75
Concave	12 (55%)	10 (45%)	
Convex	6 (55%)	5 (45%)	
Orthognathic	10 (67%)	5 (33%)	
<b>Compliance</b>			0.13
Compliant	21 (66%)	11 (34%)	
Non-compliant	0 (0%)	2 (100%)	
Sub-optimal	7 (47%)	8 (53%)	
<b>Shift</b>			0.52
Anterior	14 (67%)	7 (33%)	
Anterior/lateral	1 (100%)	0 (0%)	
Lateral	0 (0%)	1 (100%)	
Anterior and/or lateral	15 (65%)	8 (35%)	
Not reported	2 (50%)	2 (50%)	
<b>Ethnicity</b>			0.37
Asian	9 (69%)	4 (31%)	
Caucasian	14 (58%)	10 (42%)	
Other	5 (42%)	7 (58%)	