Premolar Extraction Cases Treated with Aligners:

Are Contemporary Results Comparable to Fixed Appliances?

Rachel Teitelbaum

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Dentistry

University of Washington

2026

Committee:

Greg Huang

Burcu Bayirli

Sanghee Lee

Sandra Tai

Program Authorized to Offer Degree:

Department of Orthodontics

University of Washington

Abstract

Premolar Extraction Cases Treated with Aligners:

Are Contemporary Results Comparable to Fixed Appliances?

Rachel Teitelbaum

Chair of the Supervisory Committee:

Greg Huang

Department of Orthodontics

Introduction: Historically, it has been challenging to treat premolar extraction cases with aligners to the same occlusal standards as fixed appliances. The purpose of this study was to assess whether premolar extraction cases treated with Invisalign in the past 5 years have more comparable results to cases treated with fixed appliances.

Methods: Matched pairs of premolar extraction cases were analyzed from 3 private practice orthodontists considered to be experts in both the fixed appliances and aligner techniques. Pre- and post-treatment records were assessed for (1) occlusal outcomes, (2) treatment efficiency, (3) cephalometric parameters, and (4) root parallelism in extraction cases treated with Invisalign versus full fixed appliances whose treatment was completed in the past 5 years.

Results: ABO-OGS and PAR scores were comparable for both groups, showing no statistically significant differences. Aligner cases completed treatment 2 months faster, on average, than fixed appliance cases, but this difference was not significantly different. Cephalometric analyses indicated that at the end of treatment, lower incisors had a greater change in retroclination and overbite was deeper in the aligner group. Teeth adjacent to extraction sites in the aligner group had more divergent roots compared to the fixed group.

Conclusion: Occlusal outcomes and treatment efficiency were similar in the aligner and fixed appliances patients treated in the past 5 years. However, incisor angulation, overbite, and root parallelism were not managed as well with aligners.

Acknowledgments

I would like to thank Dr. Greg Huang for his guidance, patience, and mentorship throughout this entire process. His extensive expertise and support were invaluable to the completion of this work.

I am also grateful to Dr. Sanghee Lee for guiding me whenever challenges arose and for helping me remain calm and focused throughout the project. I would like to thank Dr. Burcu Bayirli for sharing her research expertise and providing thoughtful feedback that strengthened this study.

I would like to acknowledge Dr. Sandra Tai for her mentorship, clinical insight, and for contributing cases to this project. Additionally, I extend my gratitude to Dr. Masashi Makino and Dr. Sean Holliday for generously sharing cases and offering their clinical perspectives, which were vital to this work.

Thank you to Dr. Lloyd Mancl for his work on the statistical analysis and interpretation.

Special thanks to Christin Hao (DDS Class of 2026) for her assistance with data collection and organization. I am also thankful to Dr. Rooz Khosravi for his constant encouragement and for his technical and logistical support with model scoring.

Finally, I would like to thank my co-residents Neal, Anissa, Tammy, and Julie, as well as my fiancé Jared, my mom, and my cat, Moses for their unwavering support, encouragement, and for keeping me grounded throughout this journey.

Table of Contents

List of Tables

List of Figures

Introduction

Since the late 1990s, providers have been using the Invisalign clear aligner system to treat orthodontic patients. At its introduction, Invisalign was perceived as a treatment option mainly for adults with mild malocclusions. However, as the technique has matured and developed, aligners are now used in patients of all ages and have been extended to moderate and even severe malocclusions.

In 2015, Align Technology Inc. introduced the G6 protocol, which was designed specifically to provide more predictable outcomes for premolar extraction cases. SmartForce features, such as optimized retraction attachments and optimized anchorage attachments, were introduced to better control root movements while retracting canines and protracting molars. Additionally, intrusive forces were added to incisors to prevent unwanted vertical deepening. SmartStage technology also utilized specific staging of aligners, first moving posterior teeth, then canines, followed, lastly, by retraction of incisors. Align Technology Inc. claimed that the combination of SmartForce features and SmartStage technology, packaged as the G6 protocol, would produce better clinical outcomes for premolar extraction cases (Align Technology, 2014).

Additionally, providers have been modifying their aligner strategies for premolar extraction cases. Whether utilizing the G6 features or implementing their own auxiliary methods, like different staging techniques, use of elastics, or creative attachments, one thing remains the same: both orthodontists and Align Technology Inc. are continuously working to achieve better occlusal outcomes and more efficient treatment.

Several studies have assessed the effectiveness of Invisalign aligners for premolar extraction cases. However, minimal research has been completed in recent years to see how these new advances in technology and the improved strategies of providers affect the outcomes of

Invisalign aligner cases compared to premolar extraction cases treated with fixed appliances. The purpose of this study was to determine if recent advances in aligner techniques and technology result in outcomes similar to fixed appliances in premolar extraction cases. Specifically, the aims were to investigate differences in (1) occlusal outcomes, (2) treatment efficiency, (3) cephalometric differences, and (4) root parallelism in premolar extraction cases treated with Invisalign aligners versus full fixed appliances whose treatment was completed in the past 5 years.

Background

Premolar extraction cases treated with Invisalign aligners have historically been challenging to finish to a high quality. In one of the first studies conducted on the topic in 2008, Baldwin et al. reported on a randomized controlled trial of 24 cases with at least 1 premolar extraction. Subjects were treated with hard or soft aligner material and dental casts and panoramic radiographs were analyzed. Baldwin et. al found that treating premolar extraction cases with clear aligners was most often associated with tipping of adjacent teeth into the extraction site. They also found that these unwanted side effects could be corrected with subsequent fixed appliances (Baldwin, et al., 2008). Additional case reports have found similar side effects. In 2006, researchers found that although clear aligners were successful in aligning teeth and reducing crowding in a premolar extraction case, they could not control root angulation of teeth adjacent to the extraction site. Similarly to Baldwin, they suggested the use of a hybrid technique to correct root position with fixed appliances after completion of aligners (Giancotti, et al., 2006).

Further studies confirming these unwanted side effects were published in more recent years. They also point to unfavorable incisor and canine lingual crown torque, distal tipping of canines, intrusion of first molars, mesial tipping of first and second molars, and extrusion of incisors and canines (Dai et al., 2019; Dai et al., 2021; Song et al., 2023). Dai et al. completed two studies - one on maxillary premolar extraction cases and one with 4 premolar extraction cases. Both studies evaluated actual tooth movement and compared to predicted tooth movement as described by the digital treatment plan. In both studies, differences were found between achieved and predicted tooth movement and unwanted side effects, such as mesial displacement of molars and distal tipping of canines. Researchers also found there were unpredicted movements in maxillary 1st molars, extrusion of central incisors, and lingual inclination of maxillary and mandibular central incisors (Dai et al., 2019; Dai et al., 2021). As recently as 2024, Dai et al., showed there were differences in root position between the predicted and actual movements. The largest differences were with anterior roots in the anterior-posterior dimension (F. Dai et al., 2024).

In 2023, Tang et al. reported that in 60 cases of mandibular premolar extraction treated with clear aligners, the relative anchorage loss of mandibular first molars was 25% in first premolar extraction cases and 40% in second premolar extraction cases. Similar to the results seen in the Dai et al. cases, Tang et al also noted the unpredicted extrusion of central incisors and canines, as well as the centrals having more lingual crown torque and the canines having more distal crown tipping (Tang et al., 2023). Notably, differences existed between maxillary and mandibular space closure with clear aligners. Finite element analysis has shown that although similar unwanted side effects occur in both maxillary and mandibular arches, differences may exist. Mandibular arches had more relative posterior anchorage and anterior retraction than space closure in the maxilla. Researchers also found that the longer the root of an anterior tooth, the less tipping occurred, and the more vertical changes occurred (Liu et al., 2023).

Studies that have compared the differences between premolar extraction cases treated with Invisalign aligners versus fixed appliances also highlight these unwanted side effects, while showing how cases treated with different appliances have significantly different outcomes. Song, et al. compared cephalometric and panoramic skeletal and dental changes between cases treated with Invisalign G5 protocol and fixed appliances with TADs. The authors reported that compared to cases treated with fixed appliances, Invisalign cases showed more tipping of teeth in the extraction space, as well as more posterior teeth mesialization. Additionally, these authors found that cases treated with Invisalign aligners took on average approximately 6 months longer to complete than cases in the fixed group (Song et al., 2023). Chen et al. completed a similar study, but compared cases treated with Invisalign aligners to those treated with Damon Q fixed system. 61 patients were treated with either appliance and extractions were of first upper premolars and lower first or second premolars. Unlike previous studies, this study explicitly stated that Invisalign cases were treated with the newest G6 protocol. The goal of the investigators was to measure differences in incisor retraction by analyzing pre-treatment and post-treatment cephalometric x-rays. Chen, et al. found that not only did the Invisalign group have a longer treatment time, but they also finished with more lingual crown inclination and less incisor retraction than the Damon Q group (Chen, et al., 2022).

Despite the findings above, some authors have reported satisfactory results in premolar extraction cases treated with the Invisalign system, stating that they were able to achieve good alignment, good occlusion, and ideal overbite and overjet (Hönn & Göz, 2006; Womack, 2006). Gaffuri et al. looked at the ABO- OGS scores and ABO standard cephalometric analysis of 24 cases of premolar extraction cases, 12 treated with Invisalign G6 protocol and 12 with fixed. They found that ABO scores post-treatment for both groups were comparable and that both groups had good outcomes. This group pointed to Invisalign's latest features like optimized attachments and the use of inter-arch elastics as the reason for these results (Gaffuri et al., 2020). Additional studies found no significant difference in ABO-OGS, DI, or PAR scores between post-treatment groups in extraction cases treated with Invisalign and fixed (Jaber et al., 2022; Li et al., 2015). Additionally, more recent studies using 3D imaging CBCT have investigated root angulation between extraction cases treated with clear aligners compared to fixed. The cases treated with Invisalign clear aligners showed comparable parallelism for teeth adjacent to the extraction sites. (Al-Gumaei et al., 2025).

Many clinicians have proposed techniques to minimize some of the unwanted side effects with extraction therapy. Cao et al. saw less incisor extrusion and lingual tipping when a small amount of space between anterior teeth was staged into aligner treatment before retraction (Cao et al., 2023). Others have suggested other staging techniques to achieve improved outcomes and minimize unwanted side effects (Samoto & Vlaskalic, 2014). Feng et al. tested an anti-tipping design with distal crown tipping of posterior teeth and mesial crown tipping of canines prior to alignment and space closure. Their findings suggested that this anti-tipping design could be effective in decreasing tipping of adjacent teeth into the extraction site (Feng et al., 2022). Similar overcorrection protocols have been theorized by others to achieve more predictable results (F. F. Dai et al., 2019; Jaber et al., 2022; Song et al., 2023). Routinely, prior to the latest Invisalign innovations, simultaneous and reciprocal staging were used. The "frog staging" technique and the G6 staging have been proposed most recently to help with unwanted side effects. The "frog staging" uses elements of the classic Tweed tip back bend by adding distal crown into the posterior segments while incisors are retracting. The staging also breaks up movements, moving the second premolar and canine first, pausing the movement and then moving posterior molars and incisors. The G6 staging method of retraction first retracts canines while molars move rotationally or buccal-lingually. Similar tweed tip back is added into the posterior segment for anchorage. After a specific number of stages, the canines continue to retract while the molars stop movement(Tai, 2025). Others have proposed the use of specific optimized attachments, power ridges, TADs, or elastics to improve success in extraction patients (Bowman, 2015; Feng et al., 2022; Ren et al., 2022; Yang et al., 2023). Although individual authors have reported good results with many of these techniques, there does not seem to be a clear consensus.

To date, the literature on premolar extraction cases treated with Invisalign aligners presents mixed results. Some studies have reported good success, particularly in more recently years, where others still highlight the complex challenges of using this system for extraction cases. Because the Invisalign technique for extractions has evolved, and because providers have much greater experience treating extraction cases with aligners, we wished to compare extraction cases treated with aligners or fixed appliances within the past 5 years.

Materials and Methods

It was determined by the Human Subjects Division at the University of Washington that this study was exempt from IRB review because all patient information was de-identified prior to transmission to the study team at UW.

A power analysis was performed to determine the minimum sample size required to achieve 80% power at a significant level of α =0.05. Based on mean difference in ABO-OGS score of 6 points with a 6-point standard deviation (effect size = 1) a minimum of 32 matched pairs was determined to be required.

Cases were contributed from three private practice orthodontists who had significant expertise in treating premolar extractions cases with both fixed appliances, as well as the Invisalign technique. The doctors who agreed to participate were given detailed instruction on case selection, data collection, and transfer of records. They first identified premolar extraction patients who were treated with Invisalign aligners. Doctors were instructed to start with the most recently completed case and move backwards until June 1st, 2019, or until 30 patients were identified. This was done to consecutively include aligner cases that had been completed most recently. Inclusion criteria were adult age 18 to 45, Angle Class I or Class II Division 1, availability of initial and final records, treatment completed after June 1st, 2019, and at least two premolars extracted in at least 1 arch. Exclusion criteria included: medical conditions that may have affected compliance or rate of tooth movement, clefts, craniofacial abnormalities, facial syndromes, and switching to fixed appliances to complete treatment. Providers were asked to keep track of any patients that switched from aligners to fixed, to calculate the percent of aligner patients that converted to fixed appliances.

Next, doctors were asked to identify matched fixed appliances patients, based on age (within 5 years), gender, molar relationship, number of extractions, initial crowding or spacing, and use of TADs. Doctors were allowed to contribute up to 20 matched pairs.

Once matched, information for all the subjects was gathered by the office staff, de-identified, and transmitted to the study team. The treatment information included: total treatment time, total number of visits, total number of emergency visits, and use and type of elastics. For aligner cases, we additionally asked for the change regimen, number of refinements, and any specific techniques used (attachments, staging, auxiliary mechanisms). Finally, de-identified initial and final models, panoramic x-rays, and lateral cephalometric were also transmitted.

Initial study models were uploaded to SureSmile software, to objectively measure Discrepancy Index (DI) scores for the pre-treatment models. DI scores for overjet, lateral open bite, occlusal relationship and lingual posterior crossbite were manually corrected to reflect discrepancies between the SureSmile and American Board of Orthodontics rules for scoring the DI. The cephalometric section of the DI was manually scored for each case (Table 1).

Table I. Corrected Discrepancy Index SureSmile Scores

Correction Area	Discrepancy Index Rule as described by the ABO	Correction made in SureSmile
Overjet	"Overjet is a measurement between two antagonistic anterior teeth (lateral or central incisors) comprising the greatest overjet and is measured from the facial surface of the most lingual mandibular tooth to the middle of the incisal edge of the more facially positioned maxillary tooth."	SureSmile included canines when measuring overjet. Manually adjusted for greatest overjet at lateral or central incisor
Lateral open bite	"For each maxillary posterior tooth (from the 1st premolar to the 2rd molar) in an open bite relationship ≥ 0.5 mm from its opposing tooth, measure cusp to cusp"	SureSmile included any opening greater than 0mm away from the opposing tooth. It was also measured from canine posteriorly instead of from first premolar. Scores for lateral open bite were manually adjusted based on ABO description.
Occlusal Relationship	"If the mesiobuccal cusp of the maxillary first molar occludes with the buccal groove of the mandibular first molar or anywhere between the buccal groove and the mesiobuccal or distobuccal	SureSmile scored 1 point for if there was a discrepancy between the mesiobuccally cusp of the maxillary first molar and the buccal groove of the mandibular first molar.

	cusps, no pts are scored. If the mesiobuccal cusp of the maxillary first molar occludes with the mesiobuccal (Class II end-to-end) or distobuccal (Class III end-to-end) cusps of the mandibular first molar, then 2 pts per side are scored.”	
Lingual posterior XB	“For each maxillary posterior tooth where the maxillary buccal cusp is lingual to the buccal cusp tip of the opposing mandibular tooth (from the first premolar to the third molar), 1 pt is scored.”	SureSmile scores 1 point when cusps maxillary buccal cusps were edge to edge with the buccal cusp tip the opposing mandibular tooth.

Final study models were uploaded to SureSmile software to objectively measure ABO-OGS scores for the post-treatment models. ABO-OGS were manually corrected to reflect discrepancies between the SureSmile and American Board Orthodontics-OGS rules. We did not score the root angulation section, as we examined the parallelism of roots more precisely in question #4. Scores were corrected in the following areas (Table 2).

Table II. Corrected ABO-OGS SureSmile Scores

Correction Area	Grading System for Dental Casts as described by the ABO	Correction made in SureSmile
Third molars	“Third molars are not scored unless they substitute for the second molars.”	SureSmile gave points in all categories to third molars. Manually subtracted third molar scores from each category.
Occlusal Contacts	“If the distolingual cusp is short or diminutive, it should	SureSmile gave points for all cusps and did not determine

	not be considered in the evaluation."	which cusps were considered short or diminutive. Manually reviewed occlusal contact scores and determined whether points were given for short or diminutive cusps and adjusted accordingly.
2 Points	"No more than 2 points are scored for any tooth."	SureSmile gave more than 2 points to one tooth in several categories. Scores were manually corrected to only give a maximum of 2 points per each tooth for each category.

Final models were 3D printed and were manually measured using the Peer Assessment Rating (PAR) system by three calibrated PAR raters. One member of the research team measured all 60 models, and 2 senior faculty members measured half each. If scores were within 2 points of each other the average was taken. If scores were greater than 2 points from each other, both the senior member of the research team and the main investigator re-measured until a consensus on the score was reached.

Initial and final cephalometric X-rays were uploaded to Dolphin software. A custom tracing analysis was used to measure changes in mandibular plane angle, upper and lower incisor angulation and AP positioning, vertical molar measurements, molar root angulation, and AP measure of anchorage loss (Figs 1-3).

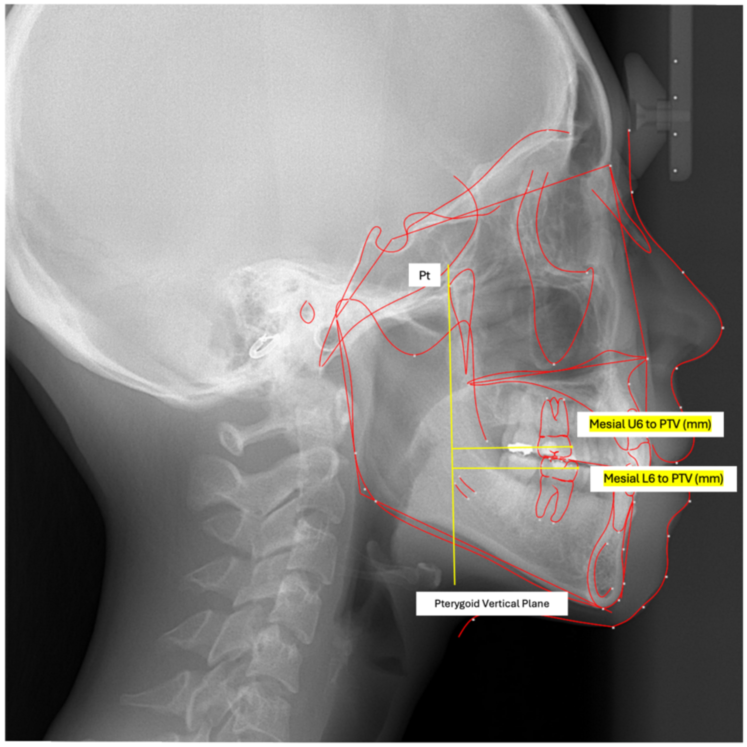

Figure I: Cephalometric X-ray showing how changes in first molar A-P movement were measured by Pterygoid Vertical Plane to mesial of the upper and lower first molars (PTV-U6 and PTV-L6).

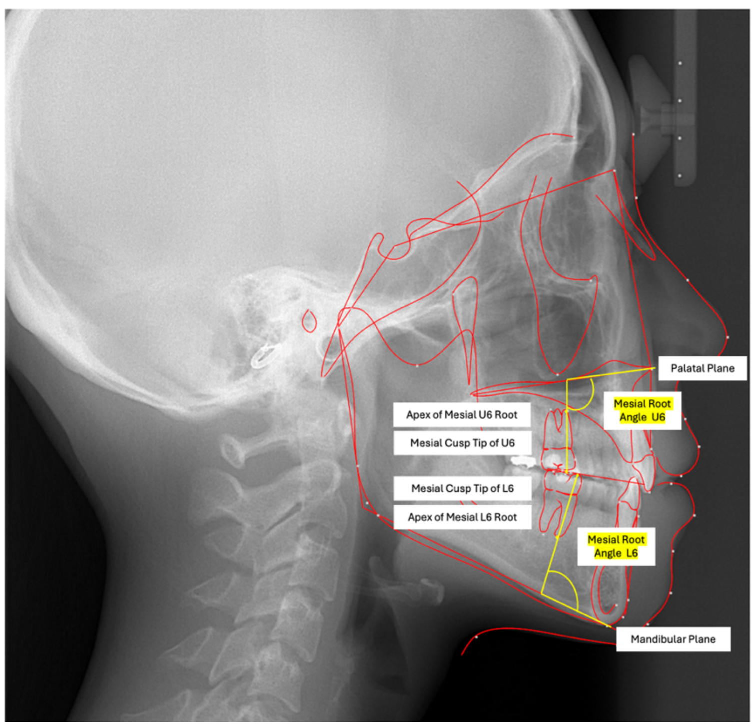

Figure II: Cephalometric X-ray showing how changes in mesial root angulation of first molars was measured. The line created from the apex of the mesial root and the mesial cusp tip was measured to the mandibular plane for the lower first molar and to the palatal plane for the upper first molar.

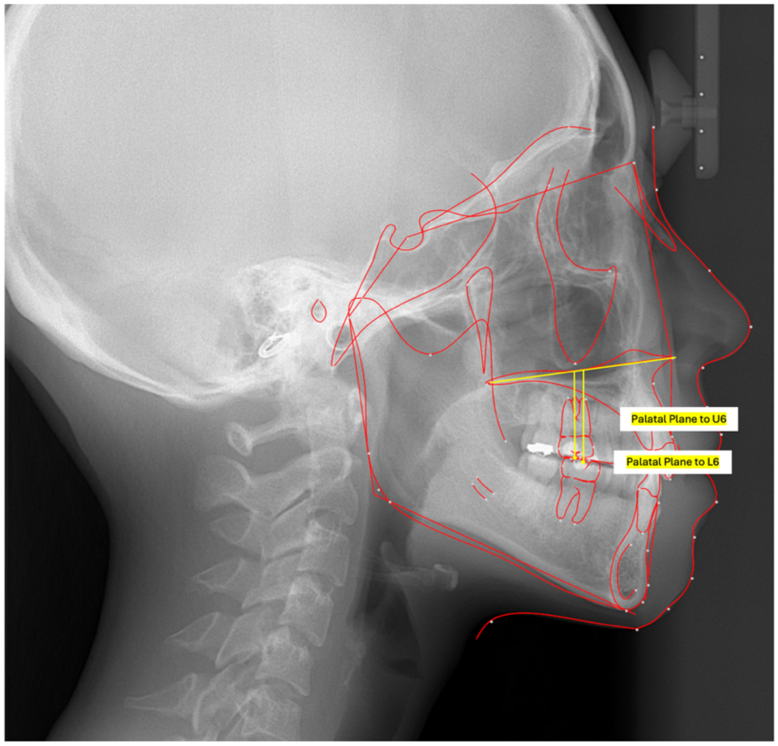

Figure III: Cephalometric X-ray showing how changes in vertical positioning of the first molars was measured. The distance from the palatal plane to the occlusal surface of the upper first molar and to the occlusal surface (mesiobuccal cusp tip?) of the lower first molar.

Finally, root parallelism was measured using the final panoramic X-rays. The angle between the long axis of the two adjacent teeth to the extracted premolar (either canine and second premolar or first premolar and molar) were recorded (Fig 4). When root angles were divergent, we assigned them negative angle values. When root angles were convergent, we assigned them positive angle values. The amount of posterior crowding was estimated and explored as a factor that might impact final root parallelism.

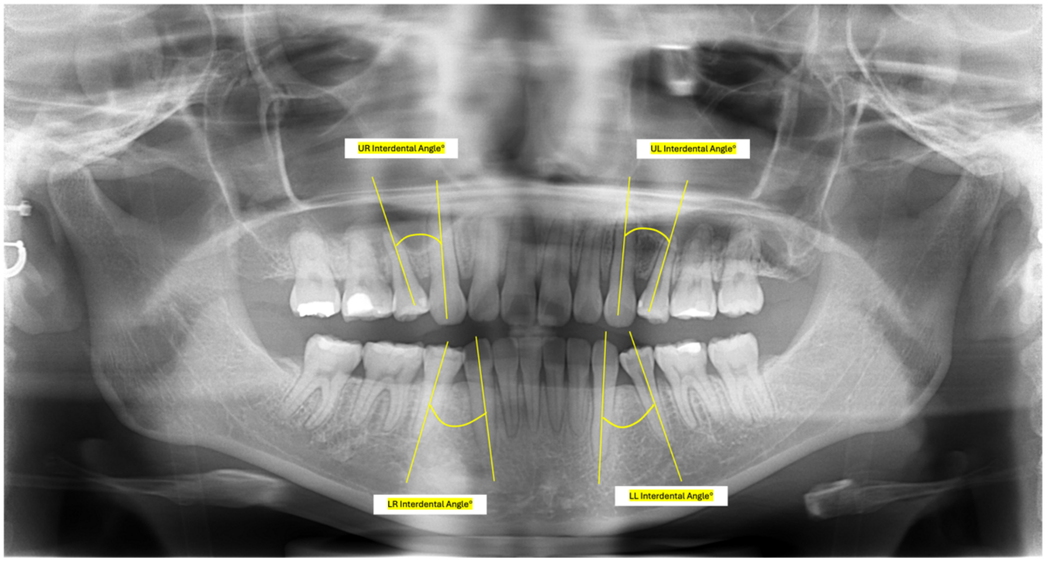

Figure IV : Panoramic X-ray root parallelism was measured. The angle created from the long axis of the teeth adjacent to the extraction site were measured.

Results

Intra-rater Reliability

18 initial and final cephalometric X-rays were randomly selected and traced two weeks apart by the primary investigator (RT) to assess intra-rater reliability. Intra-rater reliability was measured three different ways, using 95% confidence intervals (CI) for the mean differences, the intraclass correlation coefficient (ICC) and 95% CI for the ICC, and Dahlberg's error (Table 3). Intra-reliability was considered good to excellent for all cephalometric landmarks. T1 refers to the pre-treatment and T2 refers to the post-treatment values. 1st refers to timepoint 1 and 2nd refers to the second timepoint.

Table III. Intra-rater Reliability for Cephalometric Tracing

Overbite

Measure	1st Mean (SD)	2nd Mean (SD)	Diff (SD) [95%CI]	ICC (95% CI)	D. Error (Min, Max)
T1	1.1 (3.5)	1.6 (3.6)	-0.5 (0.6) [-1.0, 0.0]	0.98 (0.91, 0.99	0.5 (0.1, 1.1)
T2	3.4 (2.4)	3.7 (2.1)	-0.3 (0.6) [-0.7, 0.2]	0.97 (0.87, 0.99)	0.4 (0.0, 1.1)

T1 & T2	2.2 (3.1)	2.6 (3.1)	-0.4 (0.6) [-0.7, -0.1]	0.98 (0.94, 0.99)	0.5 (0.0, 1.1)
T2-T1	2.3 (5.0)	2.1 (4.9)	0.2 (0.9) [-0.5, 0.9]	0.98 (0.94, 1.00)	0.6 (0.2, 1.7)

6 cast models were PAR scored by 3 calibrated members of the research team at least 1 week apart. Intra-rater reliability was measured using the intraclass correlation coefficient (ICC) and 95% confidence interval (CI) for ICC (Table 4). Intra-reliability was considered fair to excellent for all 3 scorers. Inter-rater reliability was also measured using intraclass correlation coefficient (ICC), 95% CI for the ICC and was fair to excellent for the raters.

Table IV. Intra-rater and Inter-rater Reliability for PAR Scoring

Intra-rater Reliability	**ICC (95% CI)**
Rater A	0.94 (0.67 to 0.99)
Rater B	0.99 (0.96 to 1.0)
Rater C	0.97 (0.83 to 0.99)

Inter-rater Reliability	**ICC (95% CI)**
Rater A vs Rater B	0.94 (0.69 to 0.99)
Raters A vs Rater C	0.98 (0.88 to 0.99

From the three participating doctors, 74 cases (37 pairs) were sent to the research team. After careful review of pre- and post- treatment records, 22 of the 37 matches were determined to be perfect matches on specific teeth that were extracted. An additional 8 pairs were acceptable for inclusion, but not perfectly matched on the exact extraction teeth. For example, in three of these cases, the total number of premolars removed was equivalent, but there were some differences in combinations of 1st versus 2nd premolars. There were also two cases in which canines were removed instead of premolars, and two cases where a lower incisor was extracted instead of a premolar. Seven pairs of patients were deemed ineligible, because the matching was inadequate with respect to the extractions. Most of these involved molar extractions.

Two out of the three participating doctors provided information on the cases that switched to fixed appliances. One doctor reported that 2 out of 22 (9.1%) identified aligner cases switched to fixed appliances. The second doctor reported that 4 out of their 30 identified cases (13.3%) switched to fixed appliances. On average, 11.5% of aligner cases switched to fixed appliances in these two doctors' offices.

Demographics

Of the 30 matched pairs, resulting in a total of 60 patient cases, the average patient age was 25.6 years old. 24 pairs (80%) were female, and 6 pairs (20%) were male. 24 pairs (76.3%) started as Angle Class I, and 7 pairs (23.3%) started Angle Class II (Table 5). Most cases, 22 pairs (73.3%) started with mild to moderate arch length deficiency, and 8 pairs (26.7%) started with severe arch length deficiency. Of the three providers that contributed cases, Provider A treated 15 pairs making up about half of our sample (50%), Provider B contributed 12 pairs (40%), and provider C contributed 3 pairs (10%).

Of the 60 cases, 37 (62%) had an extraction pattern of all first premolars, 6 (10%) had only upper first premolars, 6 (10%) had upper first premolars and asymmetric lower premolars extraction, 4 (7%) had upper first premolars and lower second premolars, and 7 (11%) had some other extraction pattern.

Table V. Patient Characteristics, All Pairs (N=30)

Variable	**Aligner**	**Fixed**	**p-value**
Age, mean (SD)	25.3 (5.0)	25.9 (5.1)	0.511[1]
Gender:			
- **Female**	24 (80.0%	24 (80.0%)	
- **Male**	6 (20.0%)	6 (20.0%)	
Provider:			
- **A**	15 (50%)	15 (50%)	
- **B**	12 (40%)	12 (40%)	
	3 (10%)	3 (10%)	

- **C**			
Molar Relationship: - **Class I** - **Class II**	23 (76.7%) 7 (23.3%)	23 (76.7%) 7 (23.3%)	
Initial Arch Length - **Mild/moderate ALD** - **Severe ALD**	22 (73.3%) 8 (26.7%	22 (73.3%) 8 (23.3%)	

[1]Paired t-test,

Question #1 – Occlusal Outcomes

Total discrepancy index (DI) scores showed no difference in pre-treatment scores for both groups. The average total DI score for the aligner group was 21.2 points and the average score for the fixed group was 21.0 (p-value = 0.930) (Table 6). When restricting our analysis to the 22 perfect matches only, the scores were also not significantly different (p-value =0.695). All subcategories did not show any statistically significant differences in both groups.

Table VI. Initial DI Scores, All Pairs (N=30)

Variable	**Aligner N=30[1]**	**Fixed N=30[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Overjet	2.5 (1.4)	2.7 (1.5)	-0.2	-0.8, 0.5	0.616
Overbite	1.0 (1.2)	1.2 (1.1)	-0.2	-0.8, 0.5	0.616
Anterior XB	1.3 (2.8)	0.8 (1.6)	0.5	-0.6, 1.5	0.356
Anterior open bite	1.6 (2.7)	1.0 (2.7)	0.5	-0.7, 1.8	0.387
Lateral open bite	0.7 (2.1)	0.4 (1.3)	0.3	-0.7, 1.2	0.564
Crowding	2.8 (2.7)	2.1 (2.5)	0.7	-0.3, 1.7	0.173
Occlusal Relationship	1.3 (2.3)	1.8 (2.5)	-0.5	-1.2, 0.2	0.133
Lingual Posterior XB	0.9 (1.0)	1.3 (1.6)	-0.4	-1.2, 0.3	0.236
Cephalometric	9.2 (8.1)	9.7 (7.1)	-0.5	-4.4, 3.4	0.808
Total	21.2 (9.6)	21.0 (10.6)	0.2	-4.4, 4.8	0.930

[1] Mean (SD)
[2]Paired t-test

American Board of Orthodontics-Objective Grading System (ABO-OGS) scores showed no significant differences in scores for post-treatment models in both groups (Table 7). The average ABO-

OGS score for the aligner group was 51.1 and for the fixed group was 51.8 (p-value = 0.794). When restricting our analysis to the 22 perfect matches, differences between both groups was still not significantly different (p-value = 0.503). All subcategories did not show any statistically significant differences in both groups. However, the subcategory that showed the largest absolute difference was occlusal contacts. The aligner group scored on average 14.3 points whereas the fixed group scored 12.6 (p-value = 0.158). It is important to note that these scores do not include the root angulation portion of the ABO-CRE, as we assessed this more precisely in the root parallelism portion of this study.

Based on research published by Scott et al., an adjusted ABO-OGS score was arrived at by a subtracting 18 points from the scores measured by SureSmile (Scott et al., 2019). Average scores by provider were not significantly different for their cases treated in each group (Table 8).

Table VII. Final ABO-OGS Scores, All Pairs (N=30)

Variable	**Aligner N=30**[1]	**Fixed N=30**[1]	**Difference**[2]	**95% CI**[2]	**p-value**[2]
Alignment	11.8 (3.2)	11.6 (3.5)	0.2	-1.4, 1.8	0.834
Marginal ridges	4.5 (2.6)	5.2 (2.6)	-0.6	-2.0, 0.7	0.353
Buccolingual inclination	3.5 (2.4)	3.7 (2.5)	-0.2	-1.5, 1.1	0.719
Overjet	9.4 (4.1)	10.5 (2.9)	-1.2	-3.0, 0.6	0.199
Occlusal contacts	14.3 (4.7)	12.6 (3.3)	1.7	-0.7, 4.0	0.158
Occlusal relationship	7.5 (3.0)	8.1 (3.0)	-0.6	-2.2, 1.0	0.437
Interproximal contacts	0.2(0.7)	0.1 (0.4)	0.1	-0.2, 0.4	0.522
Total	51.1 (10.3)	51.8 (7.8)	-0.7	-6.1, 4.7	0.794
Adj Total (Scott et. al)	33.1	33.8	-0.7		

[1] Mean (SD)
[2]Paired t-test

Table VIII. Final ABO-OGS Adjusted Score Per Provider

Provider	**Average ABO Score Aligner Cases Mean (SD)**	**Average ABO Score Fixed Cases Mean (SD)**	**Differences**	**p-value**[2]
A (N=15,15)	34.4 (9.6)	32 (7.8)	2.4	0.510
B (N=12,12)	30.7 (7.2)	34.4 (8.1)	-3.75	0.351
C (N=3,3)	36.0 (23.6)	40.0 (1.73)	-4.0	0.808

[2]Paired t-test

Peer Assessment Rating (PAR) scores of final models showed no significant difference between the aligner group and the fixed group (Table 9). The average PAR score for the aligner group was 6.1 points and the average PAR score for the fixed group was 6.3 points (p-value = 0.826). Only one category, overbite, showed a significant difference in average points. The aligner group had scored on average 1.2 points for overbite whereas the fixed group scored on average 0.6 points (p-value =0.033). Average scores by provider are presented in Table 10.

Table IX. Final PAR Scores, All Pairs (N=30)

Variable	**Aligner Mean (SD)[1]**	**Fixed Mean (SD)[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Alignment Mx	0.2 (0.6)	0.5 (0.8)	-0.3	-0.7, 0.1	0.127
Alignment Mn	0.2 (0.4)	0.3 (0.5)	-0.1	-0.4, 0.1	0.234
Posterior AP	3.3 (0.8)	3.0 (0.8)	0.3	-0.1, 0.7	0.152
Post. Transverse	0.2 (0.5)	0.4 (0.7)	-0.2	-0.5, 0.2	0.335
Post. Vertical	0.0 (0.1)	0.0 (0.1)	0.0	0.0, 0.1	0.573
Overjet	1.0 (2.3)	1.3 (2.5)	-0.3	-1.4, 0.8	0.586
Overbite	1.2 (1.1)	0.6 (0.9)	0.7	0.1, 1.3	0.033
Centerline	0.0 (0.0)	0.3 (0.9)	-0.3	-0.6, 0.1	0.103
Total PAR Score	6.1 (3.0)	6.3 (4.0)	-0.2	-2.0, 1.6	0.826

[1] Mean (SD)
[2]Paired t-test

Variable	**Aligner Median (IQR)[1]**	**Fixed Median (IQR)[1]**	**p-value[2]**
Alignment Mx	0.0 (0.0, 0.0)	0.0 (0.0, 1.0)	0.072
Alignment Mn	0.0 (0.0, 0.0)	0.0 (0.0, 0.5)	0.252
Posterior AP	3.3 (3.0, 4.0)	3.0 (2.5, 3.5)	0.186
Posterior Transverse	0.0 (0.0, 0.0)	0.0 (0.0, 0.5)	0.386
Post. Vertical	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.773
Overjet	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.706
Overbite	2.0 (0.0, 2.0)	0.0 (0.0, 1.0)	0.035
Centerline	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.174
Total PAR Score	5.5 (4.0, 6.5)	5.0 (3.0, 9.0)	0.722

[1] Mean (SD)
[2]Wilcoxon signed rank test with continuity correction

Table X. Final PAR Score Per Provider

Provider	Average ABO Score Aligner Cases Mean (SD)	Median	Average ABO Score Fixed Cases Mean (SD)	Median
A (N=15,15)	6.1(3.1)	5.5	4.2(2.6)	3.5
B (N=12,12)	5.4(3.1)	4.75	9.2(3.9)	8.75
C (N=3,3)	9.2 (2.8)	9	5.1(3.3)	3.5

Question #2 – Treatment Efficiency

While the aligner group had a 2-month decrease in treatment time (mean = 26.3, SD = 8.3) compared to the fixed group (mean = 28.1, SD = 6.2), this difference was not statistically significant (p-value = 0.346) (Table 11). The fixed group required a significantly greater number of visits compared to the aligner cases (Fixed, mean = 24.0; Aligners, mean = 12.3, p-value < 0.001). The majority (83.3%) of aligner cases did not require any emergency visits, whereas the majority (56.6%) of fixed cases required at least one emergency visit. Most aligner cases and all the fixed cases required some elastic usage, the most popular choice for both groups being a Class II pattern (Aligner, mean = 76.7%; Fixed, mean = 66.7%). The next most used elastic pattern for the aligner group was vertical (43.3%), followed by Class III (23.3%). In contrast, the next most common elastic choice for the fixed cases was a Class III pattern (46.7%) with vertical elastics being the third choice (30.0%). Of the 30 aligner cases, 6 (20.0%) required only one refinement, the majority, 17 (56.7%), required two refinements, and finally, 7 (23.3%) required three or more refinements. More than half of these cases (66.7%) used a 7-day tray changing regimen. The rest of the cases (33.3%) used some combination of 5–14-day changes. Treatment efficiency results did not change significantly when looking at the 22 perfect matches compared to the 30 total adequate matches.

Table XI. Treatment Efficiency Characteristics, All Pairs (N=30)

Variable	**Aligner N=30**[1]	**Fixed N=30**[1]	**Difference**[2]	**95% CI**[2]	**p-value**[2]
Treatment time (months)	26.3 (8.3)	28.1 (6.2)	-1.8	-5.5, 2.0	0.346[2]
Number of visits	12.3 (4.5)	24.0 (8.6)	-11.7	-14.4, -9.0	<0.001[2]
Number of emergency visits					
0	25 (83.3%)	13 (43.3%)			0.029[3]
1	2 (6.7%)	7 (23.3%)			
2+	3 (10.0%)	10 (33.3%)			

[1] Mean (SD)
[2]Paired t-test, [3]McNemar-Bowker test

Question #3 – Cephalometric Differences

When looking at cephalometric changes from initial to final in both groups the biggest difference was in lower incisor angulation (Table 12). On average L1-NB° decreased by 10.6° in the aligner group, whereas in the fixed group L1-NB° decreased by only 2.6° (p<0.001). Likewise, the incisor angle to mandibular plane angle (IMPA) showed similar results – with IMPA in the aligner cases decreasing on average 10.3°, and in the fixed group only decreasing 2.3° (p<0.001). Changes in overbite were also shown to be statically significant, with overbite increasing on average 1.2mm in the aligner groups and decreasing on average 0.2mm in the fixed group (regression p=0.002).

The changes in upper incisor angulation to NA were not significantly different between the aligner group (-14.1°) and the fixed group (-13.6°). Similarly, for upper incisor angulation to palatal plane, the aligner group showed slightly more of a decrease in angulation (-14.9°) compared to the fixed group (-13.9°). When looking at the AP changes in incisor retraction, the lower incisors (L1-NB mm) in the aligner group had greater retraction, on average. The change was -6.0 mm for aligner patients, compared to -4.5 mm for the fixed patients. However, this difference was not statistically significant. The upper incisors (U1-NA mm) retracted a similar amount for the aligner group, -7.8 mm, compared to the fixed group, -7.1 mm. The changes in mandibular plane angle (SN-MP) and relationship of the upper jaw to the lower jaw, (ANB), were not significantly different between the groups.

To assess changes in molar positioning we looked specifically just at the perfectly paired data, in which each pair had identical premolar extractions. Angulation of lower molar mesial root showed a significant difference in changes between the two groups (Table 13). In the aligner group, this changed by -0.1°, on average, indicating minimal change. In the fixed group, this angle increased by 3.5°, on average (p-value =0.007). In both groups, the mesial of the upper and lower first molar, relative to the pterygoid vertical plane, increased, which assessed A-P changes. However, the changes between the groups was not significantly different in the upper molar (2.3 mm aligner group versus 2.7 mm fixed group; p-value = 0.708) or the lower molar (1.0mm aligner group versus 2.3mm fixed group; p-value = 0.230). The difference between upper and lower molars is suggestive of better anchorage (less mesial molar movement) in the mandible compared to the maxilla. Vertical changes in upper and lower first molars were not statistically significant.

We recognized that could be larger changes in lower molar positioning when we looked specifically at the cases where lower second premolars were removed (Table 14). The change in mesial root angulation was significantly different between groups, with the aligner groups showing a greater decrease in angulation of -3.1° whereas the fixed group had a positive change of 2.2° (p-value = 0.013). The changes in AP and vertical movements compared between aligner and fixed groups were not significant. However, on average, the aligner group has had greater mesial movement of molars compared to looking at all the perfect matches together. Indicating an increased loss of posterior anchorage when lower second molars are extracted.

Table XII. Cephalometric Changes, All Pairs (N=30)

Variable	**Aligner Mean (SD)**	**Fixed Mean (SD)**	**Difference**	**95% CI**	**p-value**	**Regression p-value**
ANB, °	-0.4 (2.1)	0.1 (1.2)	-0.6	-1.3, 0.2	0.160	0.228
SN-MP. °	-0.3 (2.7)	0.0 (2.0)	-0.4	-1.8, 1.1	0.613	0.634
Overbite, mm	1.2 (3.7)	-0.2 (3.4)	1.5	-0.4, 3.3	0.110	0.002
U1-NA, °	-14.1 (8.2)	-13.6 (6.7)	-0.5	-4.4, 3.3	0.786	0.316
U1-PP, °	-14.9 (7.8)	-13.9 (6.1)	-1.0	-4.7, 2.7	0.598	0.231
L1-NB, °	-10.6 (6.1)	-2.6 (7.7)	-8.0	-11.1, -4.9	<0.001	<0.001
IMPA, °	-10.4 (6.4)	-2.3 (7.5)	-8.2	-11.1, -5.2	<0.001	<0.001
U1-NA, mm	-7.8 (5.2)	-7.1 (4.8)	-0.7	-2.9, 1.5	0.523	0.350
L1-NB, mm	-6.0 (4.2)	-4.5 (3.4)	-1.6	-3.3, 0.1	0.066	0.077

Table XIII. Molar Cephalometric Changes, Perfect Pairs only (N=22)

Variable	**Aligner Mean (SD)**	**Fixed Mean (SD)**	**Difference**	**95% CI**	**p-value**	**Regression p-value**
PP-U6, mm	-1.7 (5.9)	0.1 (5.1)	-1.8	-6.2, 2.5	0.393	0.583
PP-L6, mm	-1.6 (6.9)	-0.1 (6.6)	-1.5	-6.8, 3.8	0.572	0.861
U6 M root, °	-2.4 (5.9)	0.0 (7.6)	-2.4	-7.1, 2.2	0.287	0.249
L6 M root, °	-0.1 (6.3)	3.5 (5.6)	-3.7	-6.8, -0.5	0.027	0.007
PTV-U6, mm	2.3 (3.7)	2.7 (7.6)	-0.5	-4.2, 3.3	0.794	0.708
PTV-L6, mm	1.0 (5.0)	2.3 (6.9)	-1.3	-4.7, 2.1	0.429	0.230

Table XIV. Second Premolar Cephalometric Changes, L5 Perfect Pairs only (N=4)

Variable	**Aligner Mean (SD)**	**Fixed Mean (SD)**	**Difference**	**95% CI**	**p-value**	**Regression p-value**
PP-L6, mm	0.3 (1.3)	0.5 (1.8)	-0.2	-5.0, 4.6	0.903	0.745
L6 M root, °	-3.1 (6.1)	2.2 (2.5)	-5.3	-15.8, 5.2	0.207	0.013
PTV-L6, mm	1.5 (4.9)	3.1 (1.3)	-1.6	-10.7, 7.6	0.622	0.958

Question #4 – Root Parallelism

When assessing root parallelism for teeth adjacent to the extraction sites, we saw a statistically significant difference. In general, the aligner group showed more divergence of roots, whereas the fixed group showed more parallel, or even convergent, roots (Table 15). The biggest difference was in the upper right quadrant, where the average angle in the aligner group was -6.1°, and the average angle in the fixed group was 5.1° ($p<0.001$). The next biggest difference was in the lower left quadrant where the average angle for the aligner group was -13.4° and the average angle for the fixed group was -3.0° ($p<0.001$). The lower right quadrant had an average angle in the aligner group of -11.8° and -1.8° for the fixed group. Finally, the smallest difference was in the upper left quadrant, where the average angle for the aligners group was -4.6° and the average angle for the fixed group was 1.4° ($p<0.001$). When comparing the maxilla on average to the mandible for the aligner group; the mandible showed more divergence in roots overall -13.1° compared to -5.3° on average in the maxilla. For the fixed group the mandible also showed slightly more divergence compared to the maxilla with the average angle for the mandible being -2.4° and 3.3° for the maxilla. When we looked specifically at the perfect matched groups that just had first premolars extracted, these differences were even larger (appendix Table 12).

Table XV. Root Parallelism, All pairs (w/ premolar extractions N=25-30)

Variable	**Aligner Mean (SD)**[1]	**Fixed Mean (SD)**[1]	**Difference**[2]	**95% CI**[2]	**p-value**[2]
Upper right quadrant	-6.1 (6.1)	5.1 (6.3)	-11.5	-14.5, -8.4	<0.001
Upper left quadrant	-4.6 (5.9)	1.4 (7.4)	-6.2	-9.7, -2.8	<0.001
Maxilla (average)	-5.3 (4.9)	3.3 (6.1)	-8.9	-11.6, -6.1	<0.001
Lower left quadrant	-13.4 (8.2)	-3.0(8.1)	-11.0	-15.6, -6.4	<0.001
Lower right quadrant	-11.8 (9.0)	-1.8 (8.3)	-10.1	-15.5, -4.7	<0.001
Mandible (average)	-13.1 (7.7)	-2.4 (7.2)	-10.7	-15.3, -6.1	<0.001

[1]Mean (SD)
[2]Paired t-test

When assessing changes in root parallelism, we wanted to assess if the amount of space impacted the final angulation. Available space was measured from estimated posterior crowding subtracted from the amount of tooth material removed during the extraction, all measured in mm. Average premolar size as described by Bolton were used (Bolton, 1958, n.d.). When looking at each quadrant separately there was minimal difference between average available space in the aligner group compared to the fixed group.

All quadrants had relatively the same amount of space to close (Table 16). In the upper right, the average space to close in the aligner group was 6.6 mm and in the fixed group was 6.5 mm (p-value = 0.493). In the upper left quadrant, the average space to close in the aligner group was 6.7 mm and in the fixed group was 6.7 mm (p-value =0.754). The total average available space in the upper arch was 13.3 mm in the aligner group and 13.2 mm in the fixed group (p-value=0.760). In the lower, the lower left quadrant had on average 6.1 mm of space to close in the aligner group and 6.3 mm of space to closure in the fixed group (p-value =0.206). In the lower right quadrant the average space to close was 5.9 mm in the aligner group and 6.3 mm in the fixed group (p-value =0.229) The total average available space in the lower arch was 11.9mm in the aligner group and 12.7mm in the fixed group (p-value = 0.090).

Table XVI. Available Space, All pairs (w/ premolar extractions N=25-30)

Variable	**Aligner Mean (SD)[1]**	**Fixed Mean (SD)[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Upper right quadrant	6.6 (0.6)	6.5 (0.8)	0.1	-0.2, 0.5	0.493
Upper left quadrant	6.7 (0.6)	6.7 (0.7)	0.0	-0.4, 0.3	0.754
Maxilla (average)	13.3 (0.9)	13.2 (1.2)	0.1	-0.4, 0.6	0.760
Lower left quadrant	6.1 (1.0)	6.3 (0.9)	-0.3	-0.7, 0.2	0.206
Lower right quadrant	5.9 (1.4)	6.3 (0.8)	-0.4	-1.2, 0.3	0.229

Mandible (average)	11.9 (1.6)	12.7 (1.4)	-0.8	-1.7, 0.1	0.090

[1]Mean (SD)
[2]Paired t-test

In only a few cases was the available space in a quadrant less than 5 mm, and in general there was no impact between the amount of space to close (typically 5mm to 7mm per quadrant) and the degree of root tipping (Fig 5). Even some cases that had less than 5 mm still exhibited poor root parallelism in the aligner group. The scattergrams indicate that on average, aligner patients had roots adjacent to the extraction site that were typically divergent, while in the fixed group, root parallelism was typically better (Fig 6).

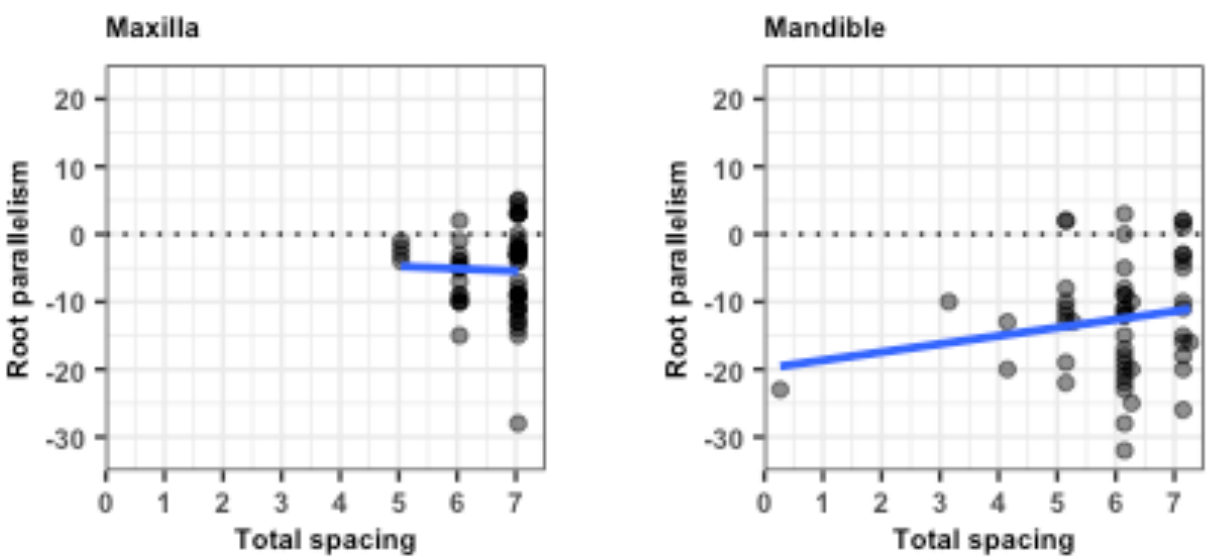

Figure V: Plots showing the relationship between total spacing and root parallelism for the aligner cases.
Maxilla: Spearman rank correlation coefficient = 0.05 (p-value = 0.727).
Mandible: Spearman rank correlation coefficient = 0.12 (p-value = 0.414)

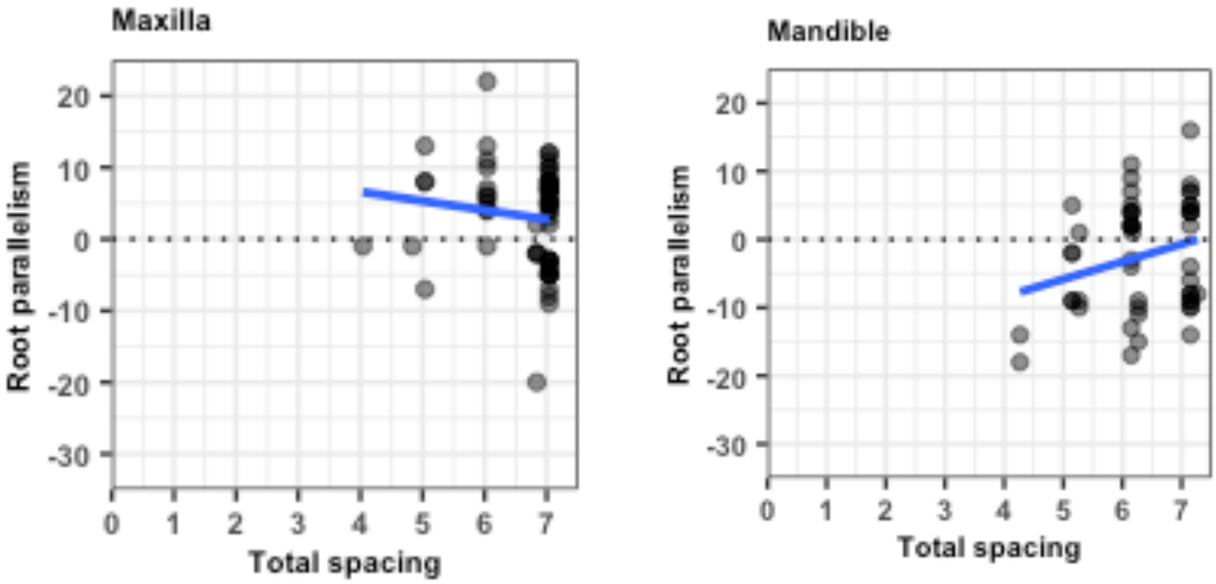

Figure VI: Plots showing the relationship between total spacing and root parallelism for the fixed cases.
Maxilla: Spearman rank correlation coefficient = -0.12 (p-value = 0.365).
Mandible: Spearman rank correlation coefficient = 0.15 (p-value = 0.306).

Discussion

Our study assessed 4 parameters related to premolar extraction cases treated with aligners and fixed appliances: occlusal outcomes, treatment efficiency, cephalometric differences, and root parallelism.

Our results are generally consistent with previous studies that assessed ABO-OGS and PAR scores and found comparable results when comparing cases treated with Invisalign and fixed appliances (Gaffuri et al., 2020; Jaber et al., 2022; Li et al., 2015). One difference that we did find was overbite assessed with the PAR, with the aligner group having greater overbite scores.

When looking at treatment efficiency, our results showed that the aligner group, on average, took 2 months less than their matched fixed cases, but this was not statistically significant. This contrasted with previous studies that looked at treatment time and found on average that the aligner cases took longer to treat, almost 6 months more on average than fixed cases ((Song et al., 2023). Additionally, our results found that the aligners were superior to fixed in having fewer treatment visits and emergency visits. Both the aligner groups and fixed groups relied on elastics to achieve desirable outcomes, with both groups favoring Class II patterns. However, the aligner group used more vertical elastics then the fixed group.

Our results for the cephalometric analysis highlighted some of the still present unwanted side effects in the aligner cases compared to the fixed cases. Aligner cases showed a statistically significant greater decrease in lower incisor inclination. This indicates that aligners do not manage lower incisor angulation well. We also noted that overbite deepened more in aligner cases. This is again consistent with previous research reporting unwanted extrusion of upper and lower incisors during space closure in aligner cases (Liu et al., 2023; Tang et al., 2023) Additionally, greater uprighting of first molar mesial roots was observed in the fixed group. Although, the difference between groups was significant, we did not see the hypothesized tipping of the mesial root in the aligner group that we would have predicted based on the previous research. This may be explained by the fact that majority of the cases were first premolar extraction cases. However, when the analysis was limited to second premolar extraction cases, greater mesial root tipping was observed, and this difference was significant between groups. These unwanted side effects are consistent with previous research (F. F. Dai et al., 2019; F. fan Dai et al., 2021; Song et al., 2023; Tang et al., 2023). Contrasting with some the previous results, we did not find any significant difference in mesialization of the first molars in the aligner group. In fact, the fixed group showed slightly more mesial movement. A possible explanation may be that there was more incisor retraction in the aligner groups. Additionally, providers potentially were able to better control anchorage in with aligner group with elastics or other techniques.

Our results for root parallelism are consistent with prior studies that showed significant divergent root angulations of the teeth adjacent to extraction sites (Baldwin et al., 2008; Giancotti et al., 2006). However, they contrasted with more recent studies that looked at comparing root parallelism using 3D imaging and showed that in some instances aligners had better root parallelism then the fixed group (Al-Gumaei et al., 2025). Our study showed that root parallelism adjacent to extraction sites is still challenging to achieve, particularly in the mandible. This is likely due to the greater bone density in the mandible (Devlin et al., n.d.).

Root parallelism remained divergent for the aligner cases; however, the degrees of divergence appears to be improved compared to earlier studies. The average change in root angulation for the aligner cases in Baldwin's study was 17.3° of divergence, whereas in our study the average root divergence was -5.3° in the upper arch and -13.1° in the lower arch. Although Baldwin et al. measured change from original to final parallelism, and our study only looked at final root parallelism, assuming a relative parallel starting point suggests that the G6 protocol may provide better control of tipping of teeth into the extraction site during space closure. Nevertheless, our findings still show significant differences between those cases treated with aligners compared to those treated with fixed appliances, signifying that this challenge has not been completely solved.

Selection Criteria:

One aspect we wanted to address is how providers select cases to be treated with each treatment modality, i.e. how does the provider decide if a premolar extraction case should be treated with aligners or with fixed. When we asked each of our expert providers, they provided an assortment of different responses. All three providers stated that the patient's autonomy in their desired treatment modality played a large role, with all of them stating that they felt comfortable treating almost all cases with either option. Along with patient related factors, whether they thought a patient would be compliant with aligner wear was emphasized with one provider. All three providers stated that when greater anchorage control is warranted, they would be more likely to consider fixed appliances, especially in cases where second premolars were the teeth to be extracted. One provider explained that further occlusal characteristics are taken into consideration. For deep bite cases or posterior open bite, fixed appliances would be preferred, whereas in a cases with severe crowding or an anterior open bite, aligners would be preferred. All three providers reported that they have completed at least 100 premolar extractions cases with aligners in the past 10 years.

Additionally, we had information on two of the providers on the percentage of cases that needed to switch to fixed appliances during the duration of treatment. When asked, one provider explained one common reasons for needing to switch was inability to control the vertical dimension, which we can see is still a challenge for the aligner cases. Additionally, they stated that loss of tracking due to either rotations or mesially tipped teeth could often lead them to make the switch. We do need to consider that our results exclude patients who may not have been progressing well with aligners.

Strengths

This study had several strengths. Unlike many other studies that may have used a general pool of fixed cases for controls, our study matched cases on gender, age, molar classification, initial crowding or spacing, and the provider. This matched design helped reduce potentially confounding variables. Additionally, we collected data in 2023 and looked back at cases from the completed in the last 5 years, meaning all the aligner cases included were finished after 2018.

This helped ensure that these were cases utilizing the latest in Invisalign technology, like the G6 protocol. Finally, during our analyses, the study team employed blinded measurement.

ABO Scores

We decided to use an objective scoring method for the DI and ABO-OGS, which was SureSmile. ABO-OGS scores were quite similar in all categories, as well as overall. However, we noticed that the scores were higher compared to what we typically see for completed cases. SureSmile is an objective system with a specified algorithm for measuring the scores. Recent studies have shown that the SureSmile automatic scoring system gives significantly greater total scores to models than the same models scored by hand. There was an average of 18 points higher in the SureSmile scores according to a study conducted in 2019 by Scott et. al. Their results showed that SureSmile gave significantly more points in alignment, overjet, and occlusal contacts. Some of their reasons for these large differences in points were the inability to use a measuring gauge directly for alignment, as well as visual and physical interferences that occur when examining a physical model that requires hand articulating (Scott et al., 2019). In a sense the scores may be higher than anticipated because there is no "wiggle" room when doing the scoring, compared to when a human is doing the scoring. For instance, if a human was looking at the model and saw the overjet was only over by 0.1 mm but all the teeth fit well, they may not score this, however, any slight 0.1 mm deviation would result in points in the SureSmile software. It is also possible for the occlusal contacts, to our eye is difficult to discriminate the degree of opening, whereas the SureSmile system may deduct points for even the smallest amount of non-contact. Furthermore, scores may be higher because our sample of cases is representative of routine every day cases treated in private practice. They were not cases that the practitioners were expecting to be evaluated by the ABO-OGS index.

Limitations

There were challenges from the utilization of 2D X-rays for this analysis. X-rays were all taken by three different providers on three different X-ray machines. Some lateral cephalometric X-rays and Panoramic X-rays were constructed from CBCT scans. Distortion and overlapping structures can often cause challenges when selecting landmarks. For example, selecting the landmarks for mesial root angulation was particularly challenging, as it was often unclear and difficult to clearly visualize due to overlapping structures. Additionally, using panoramic X-rays has some potential for distortion. When lower first premolars were extracted, we were measuring the angle of the canine to second premolar. We know that canine root position is not always accurately represented on a panoramic X-rays. Another issue we found when measuring root parallelism angles was the challenges coming from root dilacerations. When a root was dilacerated, it made it considerably more challenging to identify the exact long access of the tooth being measured.

Finally, cases that switched from aligners to fixed appliances were excluded from our sample. As previously discussed, aligner patients who were not progressing well would have been excluded, which may have led to overly optimistic results in the patients that were included in the study.

Clinical Implications:

The results of this study indicate that according to ABO-OGS and PAR standards, as well as treatment efficiency, clear aligners can be considered a good choice for treating premolar extraction cases. When providers have the expertise, experience, and technology to manage these cases, they can expect good occlusal results. However, we also found several areas in which treating cases with clear aligners can still be improved. Providers should continue to be aware of incisor torque control issues, as well as vertical overbite control, when treating these cases with clear aligners. Additionally, providers should be aware of issues with divergent roots and tipping of teeth into the extraction sites, particularly when second premolars are extracted.

In a sense, treating premolar extraction cases with aligners may be comparable to closing space on a light round arch wire. A stiffer or more rigid aligner material could potentially behave more like a stainless-steel wire and therefore be more effective at controlling the challenges we see with root angulation and incisor torque. Ideally, new advancements in aligner material properties and thickness are on the horizon, especially as we venture into printed aligners. However, until such material changes become incorporated into mainstream aligner systems, increased awareness of these limitations may help providers make more informed clinical decisions regarding case selection, incorporation of overcorrections into the digital setup, and the use of auxiliaries to help mitigate these challenges.

Suggestions for Future Investigation:

Examining whether specific attachment designs, particular staging patterns, or quantifying how much overcorrection is needed for optimal results could be interesting to investigate in the future. To answer these questions, prospective designs would be advantageous in order to fully capture the impact of these variables.

Conclusion:

The purpose of this study was to assess whether premolar extraction cases treated with Invisalign in the past 5 years have more comparable results to cases treated with fixed appliances. Specifically, differences in (1) occlusal outcomes, (2) treatment efficiency, (3) cephalometric differences, and (4) root parallelism in premolar extraction cases were investigated. This study found that in terms of overall occlusal outcomes and treatment efficiency, aligner cases completed in the past 5 years were similar to those treated with fixed appliances. However, incisor angulation, overbite, and root parallelism are still challenging to manage with aligners.

References

Al-Gumaei, W. S., Jian, F., Zhang, X., Tang, Y., Yang, Y., Zhang, H., Lai, W., & Long, H. (2025). Three-Dimensional Comparative Analysis of Root Parallelism in First Premolar Extraction Cases: Clear Aligner Versus Fixed Orthodontic Appliance. *Orthodontics and Craniofacial Research*, *28*(2), 336–344. https://doi.org/10.1111/ocr.12874

Align Technology. (2014). *ALIGN TECHNOLOGY ANNOUNCES INVISALIGN G6 CLINICAL INNOVATIONS FOR ORTHODONTIC TREATMENT OF FIRST PREMOLAR EXTRACTIONS*. Align Technology. https://investor.aligntech.com/news-releases/news-release-details/align-technology-announces-invisalign-g6-clinical-innovations

Baldwin, D. K., King, G., Ramsay, D. S., Huang, G., & Bollen, A. M. (2008). Activation time and material stiffness of sequential removable orthodontic appliances. Part 3: Premolar extraction patients. *American Journal of Orthodontics and Dentofacial Orthopedics*, *133*(6), 837–845. https://doi.org/10.1016/j.ajodo.2006.06.025

Bolton, 1958. (n.d.).

Bowman, S. J. (2015). Creative Adjuncts for Clear Aligners Part 3 Extraction and interdisciplinary Treatment. *Journal of Clinical Orthodontics* , *49*(4), 249–262. www.aligntech.com.

Cao, Y., Wang, Z. W., Chen, D., Liu, L., Li, D. X., Li, N., Ying, S. Q., Liu, X., & Jin, F. (2023). The effect of space arrangement between anterior teeth on their retraction with clear aligners in first premolar extraction treatment: a finite element study. *Progress in Orthodontics*, *24*(1). https://doi.org/10.1186/s40510-023-00484-1

Chen, J., Wen, J., Huang, L., Zhang, L., Han, L., & Li, H. (2022). Comparisons of maxillary incisor retraction effects for patients with first premolar extractions between Damon Q and Invisalign®: A retrospective study. *Medicine (United States)*, *101*(40). https://doi.org/10.1097/MD.0000000000030919

Dai, F. F., Xu, T. M., & Shu, G. (2019). Comparison of achieved and predicted tooth movement of maxillary first molars and central incisors: First premolar extraction treatment with Invisalign. *Angle Orthodontist*, *89*(5), 679–687. https://doi.org/10.2319/090418-646.1

Dai, F. fan, Xu, T. min, & Shu, G. (2021). Comparison of achieved and predicted crown movement in adults after 4 first premolar extraction treatment with Invisalign. *American Journal of Orthodontics and Dentofacial Orthopedics*, *160*(6), 805–813. https://doi.org/10.1016/j.ajodo.2020.06.041

Dai, F., Sang, Y., Zeng, J., Wang, H., Pan, Y., Zhao, J., Xu, T., & Shu, G. (2024). How accurate is predicted root movement achieved in four first-premolar extraction cases with Invisalign? *Orthodontics and Craniofacial Research*, *27*(6), 985–995. https://doi.org/10.1111/ocr.12842

Devlin, Volume, & Number. (n.d.). *THE JOURNAL OF PROSTHETIC DENTISTRY MATERIAL AND METHODS*.

Feng, X., Jiang, Y., Zhu, Y., Hu, L., Wang, J., Qi, Y., & Ma, S. (2022). Comparison between the designed and achieved mesiodistal angulation of maxillary canines and posterior teeth and influencing factors: First premolar extraction treatment with clear aligners. *American Journal of Orthodontics and Dentofacial Orthopedics*, *162*(2), e63–e70. https://doi.org/10.1016/j.ajodo.2022.05.006

Gaffuri, F., Cossellu, G., Lanteri, V., Brotto, E., & Farronato, M. (2020). *Comparative Effectiveness of Invisalign and Fixed Appliances in First-Premolar Extraction Cases*. www.jco-online.com

Giancotti, A., Greco, M., & Manpieri, G. (2006). Extraction Treatment using Invisalign Technique. *Progress in Orthodontics* , *7*(1), 32–43.
Hönn, M., & Göz, G. (2006). Invisalign®-Behandlung in einem Prämolarenextraktionsfall. *Journal of Orofacial Orthopedics*, *67*(5), 385–394. https://doi.org/10.1007/s00056-006-0609-6
Jaber, S. T., Hajeer, M. Y., & Burhan, A. S. (2022). The Effectiveness of In-house Clear Aligners and Traditional Fixed Appliances in Achieving Good Occlusion in Complex Orthodontic Cases: A Randomized Control Clinical Trial. *Cureus*. https://doi.org/10.7759/cureus.30147
Li, W., Wang, S., & Zhang, Y. (2015). The effectiveness of the Invisalign appliance in extraction cases using the the ABO model grading system: a multicenter randomized controlled trial. *International Journal of Clinical and Experimental Medicine*, *8*(5), 8276–8282.
Liu, J. qi, Zhu, G. yin, Wang, Y. gan, Zhang, B., Yao, K., & Zhao, Z. he. (2023). Different biomechanical effects of clear aligners in closing maxillary and mandibular extraction spaces: Finite element analysis. *American Journal of Orthodontics and Dentofacial Orthopedics*, *163*(6), 811-824.e2. https://doi.org/10.1016/j.ajodo.2022.07.021
Ren, L., Liu, L., Wu, Z., Shan, D., Pu, L., Gao, Y., Tang, Z., Li, X., Jian, F., Wang, Y., Long, H., & Lai, W. (2022). The predictability of orthodontic tooth movements through clear aligner among first-premolar extraction patients: a multivariate analysis. *Progress in Orthodontics*, *23*(1). https://doi.org/10.1186/s40510-022-00447-y
Samoto, H., & Vlaskalic, V. (2014). A Customized Staging Procedure to Improve the Predictability of Space Closure with Sequential Aligners. *Journal of Clinical Orthodontics*, *48*(6), 359–367. www.aligntech.com.
Scott, J. D., English, J. D., Cozad, B. E., Borders, C. L., Harris, L. M., Moon, A. L., & Kasper, F. K. (2019). Comparison of automated grading of digital orthodontic models and hand grading of 3-dimensionally printed models. *American Journal of Orthodontics and Dentofacial Orthopedics*, *155*(6), 886–890. https://doi.org/10.1016/j.ajodo.2018.11.011
Song, J. H., Lee, J. H., Joo, B. H., Choi, Y. J., Chung, C. J., & Kim, K. H. (2023). Treatment outcome comparison of Invisalign vs fixed appliance treatment in first premolar extraction patients. *American Journal of Orthodontics and Dentofacial Orthopedics*. https://doi.org/10.1016/j.ajodo.2023.10.014
Tai, S. K. (2025). Biomechanical considerations and staging strategies for premolar extraction using clear aligners: a clinical review. In *Seminars in Orthodontics*. W.B. Saunders. https://doi.org/10.1053/j.sodo.2025.08.002
Tang, Z., Chen, W., Mei, L., Abdulghani, E. A., Zhao, Z., & Li, Y. (2023). Relative anchorage loss under reciprocal anchorage in mandibular premolar extraction cases treated with clear aligners. *The Angle Orthodontist*, *93*(4), 375–381. https://doi.org/10.2319/102222-727.1
Womack, W. R. (2006). Case Report: Four-Premolar Extraction Treatment with Invisalign. *Journal of Clinical Orthodontics*, *40*(8), 493–500. www.aligntech.com.
Yang, Y., Yang, R., Liu, L., Zhang, X., Jiang, Q., Fan, Q., Zhang, H., Long, H., & Lai, W. (2023). The effects of aligner anchorage preparation on mandibular first molars during premolar-extraction space closure with clear aligners: A finite element study. *American Journal of Orthodontics and Dentofacial Orthopedics*, *164*(2), 226–238. https://doi.org/10.1016/j.ajodo.2022.12.013

Appendices

Appendix I. Determining Minimum Sample Size

Mean difference in Txt time	Standard Deviation	Alpha	Power	Sample Size needed (Total)
2 months	2	0.05	80%	32
	4			126
	6			282
4 months	2	0.05	80%	8
	4			32
	6			70
6 months	2	0.05	80%	4
	4			14
	6			32
8 months	2	0.05	80%	2
	4			8
	6			18

Mean difference in ABO score	Standard Deviation	Alpha	Power	Sample Size needed (Total)
4 points	2	0.05	80%	8
	4			32
	6			70
6 points	2	0.05	80%	4
	4			14
	6			32
8 points	2	0.05	80%	2
	4			8
	6			18
10 points	2	0.05	80%	2
	4			6
	6			12

Appendix II. Intra-rater Reliability other Cephalometric measurements

U1-PP

Measure	**1st Mean (SD)**	**2nd Mean (SD)**	**Diff (SD) [95%CI]**	**ICC (95% CI)**	**D. Error (Min, Max)**
T1	119.2 (8.3)	118.0 (8.3)	1.1 (2.8) [-1.0, 3.3]	0.94 (0.77, 0.99)	2.0 (0.5, 6.1)
T2	100.9 (5.5)	100.7 (5.7)	0.3 (1.6) [-1.0, 1.5]	0.96 (0.85, 0.99)	1.1 (0.1, 3.5)

T1 & T2	110.0 (11.6)	109.3 (11.3)	0.7 (2.3) [-0.4, 1.8]	0.98 (0.95, 0.99)	1.6 (0.1, 6.1)
T2-T1	-18.3 (8.6)	-17.4 (7.6)	-0.9 (3.0) [-3.2, 1.4]	0.93 (0.76, 0.98)	2.1 (0.4, 5.2)

IMPA

Measure	**1st Mean (SD)**	**2nd Mean (SD)**	**Diff (SD) [95%CI]**	**ICC (95% CI)**	**D. Error (Min, Max)**
T1	101.1 (6.3)	100.9 (6.2)	0.2 (1.8) [-1.2, 1.6]	0.96 (0.85, 0.99)	1.2 (0.2, 2.9)
T2	91.7 (9.0)	91.0 (9.0)	0.6 (1.2) [-0.3, 1.6]	0.99 (0.96, 1.00)	0.9 (0.2, 2.2)
T1 & T2	96.4 (9.0)	96.0 (9.1)	0.4 (1.5) [-0.4, 1.2]	0.99 (0.96, 0.99)	1.1 (0.2, 2.9)
T2-T1	-9.4 (6.4)	-9.9 (5.9)	0.4 (1.7) [-0.9, 1.7]	0.96 (0.86, 0.99)	1.2 (0.2, 4.2)

U1 – NA

Measure	**1st Mean (SD)**	**2nd Mean (SD)**	**Diff (SD) [95%CI]**	**ICC (95% CI)**	**D. Error (Min, Max)**
T1	26.9 (10.4)	26.4 (9.5)	0.6 (3.1) [-1.8, 2.9]	0.96 (0.83, 0.99)	2.1 (0.2, 7.8)
T2	9.7 (5.3)	10.1 (5.7)	-0.3 (1.3) [-1.3, 0.6]	0.98 (0.90, 0.99)	0.9 (0.1, 2.5)
T1 & T2	18.3 (11.9)	18.2 (11.3)	0.1 (2.3) [-1.0, 1.3]	0.98 (0.95, 0.99)	1.6 (0.1, 7.8)
T2-T1	-17.2 (10.6)	-16.3 (8.9)	-0.9 (3.0) [-3.2, 1.4]	0.96 (0.83, 0.99)	2.1 (0.4, 7.3)

L1-NA

Measure	**1st Mean (SD)**	**2nd Mean (SD)**	**Diff (SD) [95%CI]**	**ICC (95% CI)**	**D. Error (Min, Max)**
T1	36.6 (4.1)	36.2 (3.8)	0.3 (1.4) [-0.7, 1.4]	0.94 (0.79, 0.99)	0.9 (0.3, 2.6)
T2	26.4 (6.5)	25.6 (6.8)	0.5 (1.0) [-0.3, 1.3]	0.99 (0.94, 1.00)	0.8 (0.1, 2.2)
T1 & T2	31.5 (7.4)	31.1 (7.5)	0.4 (1.2) [-0.2, 1.0]	0.99 (0.97, 0.99)	0.9 (0.1, 2.6)
T2-T1	-10.1 (6.0)	-10.3 (6.6)	0.2 (1.9) [-1.3, 1.6]	0.96 (0.84, 0.99)	1.3 (0.0, 2.9)

U6 Mesial Root Angulation

Measure	1st Mean (SD)	2nd Mean (SD)	Diff (SD) [95%CI]	ICC (95% CI)	D. Error (Min, Max)
T1	86.6 (5.0)	86.1 (3.8)	0.5 (2.1) [-1.1, 2.1]	0.90 (0.64, 0.98)	1.4 (0.1, 3.9)
T2	86.8 (7.3)	86.8 (5.6)	0.0 (2.6) [-2.0, 2.0]	0.93 (0.74, 0.98)	1.7 (0.1, 4.6)
T1 & T2	86.7 (6.1)	86.4 (4.7)	0.2 (2.3) [-0.9, 1.4]	0.92 (0.79, 0.97)	1.6 (0.1, 4.6)
T2-T1	0.2 (6.1)	0.7 (4.6)	-0.5 (2.7) [-2.6, 1.6]	0.88 (0.59, 0.97)	1.9 (0.4, 5.9)

L6 Mesial Root Angulation

Measure	1st Mean (SD)	2nd Mean (SD)	Diff (SD) [95%CI]	ICC (95% CI)	D. Error (Min, Max)
T1	96.9 (3.6)	97.7 (3.2)	-0.7 (1.9) [-2.2, 0.8]	0.84 (0.46, 0.96)	1.4 (0.2, 3.6)
T2	94.5 (4.9)	96.0 (4.4)	-1.5 (2.1) [-3.2, 0.1]	0.85 (0.51, 0.96)	1.8 (0.1, 4.9)
T1 & T2	95.7 (4.3)	96.8 (3.8)	-1.1 (2.0) [-2.1, -0.1]	0.85 (0.65, 0.94)	1.6 (0.1, 4.9)
T2-T1	-2.4 (4.5)	-1.7 (3.5)	-0.8 (2.4) [-2.6, 1.1]	0.82 (0.43, 0.96)	1.7 (0.1, 4.3)

Appendix III. Patient Characteristics, Perfect Pairs (N=22)

Variable	**Aligner**	**Fixed**	**p-value**
Age, mean (SD)	25.9 (4.5)	27.1 (5.2)	0.177[1]
Gender:			
- **Female**	20 (90.9%)	20 (90.9%)	>0.999[2]
- **Male**	2 (9.1%)	2 (9.1%)	
Provider:			
- **A**	15 (68.2%)	15 (68.2%)	>0.999[2]
- **B**	5 (22.7%)	5 (22.7%)	
- **C**	2 (9.1%)	2 (9.1%)	
Molar Relationship:			
- **Class I**	20 (90.9%)	20 (90.9%)	>0.999[2]
- **Class II**	2 (9.1%)	2 (9.1%)	

[1]Paired t-test; [2]McNemar-Bowker test

Appendix IV. Initial DI Scores, Perfect Pairs (N=22)

Variable	Aligner N=22[1]	Fixed N=22[1]	Difference[2]	95% CI[2]	p-value[2]
Overjet	2.5 (1.3)	2.8 (1.3)	-0.4	-1.0, 0.3	0.277
Overbite	1.0 (1.2)	1.2 (1.1)	-0.2	-1.0, 0.6	0.648
Anterior XB	1.1 (2.4)	0.9 (1.6)	0.3	-1.0, 1.5	0.654
Anterior open bite	1.9 (3.0)	1.1 (3.0)	0.8	-0.8, 2.4	0.322
Lateral open bite	0.9 (2.4)	0.4 (1.3)	0.5	-0.7, 1.8	0.367
Crowding	3.4 (2.8)	2.6 (2.6)	0.8	-0.5, 2.1	0.215
Occlusal Relationship	0.8 (2.1)	1.1 (2.2)	-0.3	-0.9, 0.4	0.378
Lingual Posterior XB	0.8 (1.0)	1.3 (1.8)	-0.5	-1.5, 0.5	0.298
Cephalometric	9.0 (8.2)	11.3 (7.2)	-2.4	-6.9, 2.2	0.296
Total	21.4 (10.0)	22.6 (11.3)	-1.3	-7.0, 4.5	0.650

[1] Mean (SD)
[2]Paired t-test

Appendix V. Final ABO-OGS Scores, Perfect Pairs (N=22)

Variable	Aligner N=22[1]	Fixed N=22[1]	Difference[2]	95% CI[2]	p-value[2]
Alignment	12.5 (2.8)	11.4 (3.4)	1.1	-0.7, 2.8	0.208
Marginal ridges	4.5 (2.4)	4.4 (1.9)	0.0	-1.2, 1.3	0.940
Buccolingual inclination	3.6 (2.5)	3.9 (2.5)	-0.3	-1.8, 1.3	0.717
Overjet	9.7 (4.6)	10.9 (2.9)	-1.2	-3.4, 1.1	0.288
Occlusal contacts	14.7 (4.9)	12.3 (3.4)	2.5	-0.4, 5.3	0.091
Occlusal relationship	7.4 (3.2)	7.8 (2.1)	-0.4	-2.0, 1.2	0.642
Interproximal contacts	0.2 (0.9)	0.0 (0.0)	0.2	-0.2, 0.6	0.234
Total	52.6 (10.8)	50.6 (7.2)	2.0	-4.1, 8.1	0.503
Adj Total (Scott et. al)	34.6	32.6	2.0		

[1] Mean (SD)
[2]Paired t-test

Appendix VI. Final PAR Scores, Perfect Pairs (N=22)

Variable	Aligner Mean (SD)[1]	Fixed Mean (SD)[1]	Difference[2]	95% CI[2]	p-value[2]
Crowding Mx	0.3 (0.7)	0.3 (0.7)	-0.1	-0.5, 0.3	0.665
Crowding Mn	0.2 (0.4)	0.3 (0.5)	-0.2	-0.4, 0.1	0.162
AP	3.2 (0.8)	2.9 (0.9)	0.4	-0.1, 0.9	0.141
Transverse	0.3 (0.5)	0.1 (0.2)	0.2	-0.1, 0.5	0.176
Vertical	0.0 (0.1)	0.0 (0.0)	0.0	0.0, 0.1	0.162
Overjet	0.8 (2.1)	0.5 (1.8)	0.3	-1.0, 1.6	0.665
Overbite	1.3 (1.2)	0.3 (0.6)	1.0	0.4, 1.7	0.003

Centerline	0.0 (0.0)	0.2 (0.9)	-0.2	-0.6, 0.2	0.329
Total PAR Score	6.0 (3.2)	4.7 (2.6)	1.4	-0.6, 3.3	0.152

[1] Mean (SD)
[2]Paired t-test

Variable	**Aligner Median (IQR)[1]**	**Fixed Median (IQR)[1]**	**p-value[2]**
Crowding Mx	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.527
Crowding Mn	0.0 (0.0, 0.0)	0.0 (0.0, 1.0)	0.198
AP	3.0 (3.0, 4.0)	3.0 (2.0, 3.5)	0.180
Transverse	0.0 (0.0, 0.5)	0.0 (0.0, 0.0)	0.243
Vertical	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.346
Overjet	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.766
Overbite	2.0 (0.0, 2.0)	0.0 (0.0, 0.0)	0.007
Centerline	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	>0.999
Total PAR Score	5.5 (4.0, 6.5)	4.0 (3.0, 5.0)	0.235

[1] Mean (SD)
[2]Wilcoxon signed rank test with continuity correction

Appendix VII. Treatment Efficiency Characteristics, Perfect Pairs (N=22)

Variable	**Aligner N=22[1]**	**Fixed N=22[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Treatment time (months)	26.3 (8.3)	28.1 (6.2)	-1.8	-5.5, 2.0	0.346[2]
Number of visits	12.3 (4.5)	24.0 (8.6)	-11.7	-14.4, -9.0	<0.001[2]
Number of emergency visits					
0	25 (83.3%)	13 (43.3%)			0.029[3]
1	2 (6.7%)	7 (23.3%)			
2+	3 (10.0%)	10 (33.3%)			

[1] Mean (SD)
[2]Paired t-test, [3]McNemar-Bowker test

Appendix VIII. Cephalometric Changes, Perfect Pairs (N=22)

Variable	**Aligner Mean (SD)**	**Fixed Mean (SD)**	**Difference**	**95% CI**	**p-value**	**Regression p-value**
ANB, °	0.1 (2.0)	0.3 (1.2)	-0.2	-1.1, 0.8	0.678	0.513
SN-MP. °	0.0 (2.8)	0.1 (2.0)	-0.2	-2.0, 1.7	0.864	0.927
Overbite, mm	1.3 (3.5)	-0.4 (3.2)	1.7	-0.5, 3.9	0.126	0.001
U1-NA, °	-16.0 (8.1)	-14.9 (6.3)	-1.1	-5.6, 3.5	0.626	0.396
U1-PP, °	-16.6 (7.3)	-15.1 (5.6)	-1.5	-5.8, 2.7	0.457	0.137
L1-NB, °	-10.0 (6.4)	-2.9 (7.0)	-7.1	-10.3, -3.8	<0.001	<0.001

IMPA, °	-9.9 (6.6)	-2.7 (7.1)	-7.1	-10.5, -3.8	<0.001	<0.001
U1-NA, mm	-7.8 (5.5)	-7.4 (4.7)	-0.4	-3.0, 2.1	0.721	0.936
L1-NB, mm	-5.0 (3.4)	-4.3 (2.8)	-0.7	-2.2, 0.7	0.305	0.246

Appendix IX. Molar Cephalometric Changes, All Pairs (N=30)

Variable	**Aligner Mean (SD)**	**Fixed Mean (SD)**	**Difference**	**95% CI**	**p-value**	**Regression p-value**
PP-U6, mm	-0.7 (8.6)	-0.2 (5.0)	-0.5	-4.5, 3.5	0.805	0.802
PP-L6, mm	-0.4 (10.2)	-0.7 (6.3)	0.2	-4.7, 5.2	0.920	0.860
U6 M root, °	-3.3 (6.3)	0.5 (7.1)	-3.7	-7.3, -0.2	0.041	0.044
L6 M root, °	1.2 (6.5)	3.7 (5.0)	-2.5	-5.3, 0.3	0.079	0.012
PTV-U6, mm	3.6 (8.1)	2.4 (7.4)	1.2	-2.5, 4.9	0.505	0.442
PTV-L6, mm	2.0 (9.2)	1.1 (7.8)	0.9	-2.9, 4.7	0.634	0.795

Appendix X. Available Space, Perfect Pairs (w/ premolar extractions N=20-22)

Variable	**Aligner Mean (SD)[1]**	**Fixed Mean (SD)[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Upper right quadrant	6.6 (0.7)	6.4 (0.8)	0.2	-0.3, 0.6	0.427
Upper left quadrant	6.6 (0.7)	6.7 (0.6)	-0.1	-0.5, 0.2	0.418
Maxilla (average)	13.2 (1.0)	13.1 (1.3)	0.0	-0.6, 0.7	0.886
Lower left quadrant	5.9 (1.0)	6.4 (0.9)	-0.5	-1.0, 0.1	0.095
Lower right quadrant	5.8 (1.6)	6.3 (0.8)	-0.5	-1.3, 0.3	0.226
Mandible (average)	5.8 (1.6)	6.3 (0.8)	-0.5	-1.3, 0.3	0.226

[1]Mean (SD)
[2]Paired t-test

Appendix XI. Root Parallelism, Perfect Pairs with First Premolars Only (N=20-22)

Variable	**Aligner Mean (SD)[1]**	**Fixed Mean (SD)[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Upper right quadrant	-6.5 (6.6)	7.0 (5.9)	-13.5	-16.9, -10.1	<0.001
Upper left quadrant	-5.6 (5.8)	4.3 (5.6)	-9.9	-13.0, -6.7	<0.001
Maxilla (average)	-6.0 (5.0)	5.6 (4.8)	-11.7	-14.2, -9.1	<0.001

Lower left quadrant	-14.5 (6.8)	-0.7 (7.6)	-13.8	-18.1, -7.4	<0.001
Lower right quadrant	-12.1 (7.9)	0.7 (7.6)	-12.8	-17.1, -5.7	<0.001
Mandible (average)	-13.0 (7.0)	-0.4 (6.8)	-12.7	-17.2, -7.0	<0.001

[1]Mean (SD)
[2]Paired t-test

Appendix XII. Root Parallelism, Perfect Pairs (w/ premolar extractions N=20-22)

Variable	**Aligner Mean (SD)[1]**	**Fixed Mean (SD)[1]**	**Difference[2]**	**95% CI[2]**	**p-value[2]**
Upper right quadrant	-6.5 (6.6)	7.0 (5.9)	-13.5	-17.3, -9.6	<0.0001
Upper left quadrant	-5.6 (5.8)	4.3 (5.6)	-9.9	-13.3, -6.4	<0.001
Maxilla (average)	-6.0 (5.0)	5.6 (4.8)	-11.7	-14.6, -8.7	<0.001
Lower left quadrant	-15.3 (6.8)	-2.5 (8.4)	-12.8	-17.6, -7.9	<0.001
Lower right quadrant	-12.7 (7.7)	-1.3 (8.6)	-11.4	-16.6, -6.2	<0.001
Mandible (average)	-14.0 (6.8)	-1.9 (7.4)	-12.1	-16.6, -7.5	<0.001

[1]Mean (SD)
[2]Paired t-test