

A COMPARATIVE STATISTICAL STUDY OF TREATMENT OUTCOMES IN ORTHODONTIC PROCEDURES USING LONGITUDINAL DATA ANALYSIS

Percy Huata Panca¹, Vladimiro Ibañez Quispe¹, Juan Reinaldo Paredes Quispe¹, Lucero Danitza Mamani Chipana², Nelly Beatriz Quispe Maquera^{3*}, Henry Quispe Cruz³, Godofredo Quispe Mamani¹

¹Facultad de Ingeniería Estadística e Informática, Universidad Nacional del Altiplano de Puno, Puno, Peru.

²Doctorado en Ciencias de la Salud, Universidad Nacional del Altiplano de Puno, Puno, Peru.

³Escuela Profesional de Odontología, Universidad Nacional del Altiplano de Puno, Puno, Peru. nbquispe@unap.edu.pe

Received: 19 January 2026; Revised: 15 March 2026; Accepted: 16 March 2026

<https://doi.org/10.51847/0CR0iq1PtX>

ABSTRACT

Orthodontic treatment outcomes unfold over months or years, so repeated measurements are more informative than isolated pre-treatment and post-treatment comparisons. In routine clinical records, however, repeated observations are rarely complete or equally spaced. Missed appointments, delayed visits, incomplete cephalometric or cast records, changing compliance, and examiner variability are expected features of orthodontic data. These imperfections must be treated as part of the evidence rather than removed from the analysis. Many comparative orthodontic studies evaluate fixed appliances and clear aligners using complete cases or simplified before-and-after summaries. Such approaches can bias results when patients with poorer progress are more likely to miss visits or discontinue treatment. They also ignore that treatment groups often differ at baseline in age, severity, malocclusion type, and motivation. A longitudinal analysis must therefore address missing outcomes, attrition, irregular visit timing, measurement error, and non-random procedure assignment. This study compared longitudinal treatment outcomes between fixed appliances and clear aligners using a university orthodontic clinic dataset. The primary objective was to estimate change in Peer Assessment Rating score over treatment while explicitly accounting for irregular observation times, missing outcome measurements, patient dropout, and baseline imbalance. Secondary objectives were to compare complete-case and imputed estimates, evaluate propensity-score-adjusted treatment comparisons, and assess whether conclusions were robust to alternative longitudinal models. The analytic dataset included 215 patients treated in a university orthodontic clinic between 2017 and 2023, with 125 receiving fixed appliances and 90 receiving clear aligners. Planned records were baseline, 3 months, 6 months, 12 months, and end of treatment, but actual visit timing varied and some patients contributed only three or four usable outcome records. The primary outcome was PAR score, modeled using linear mixed effects regression with patient-specific random intercepts and time represented as actual months since baseline. Missing outcomes were addressed using multiple imputation, attrition was examined using pattern-mixture sensitivity analysis, and baseline imbalance was reduced using propensity score matching. Dropout before the final treatment record was 22% overall, with higher attrition in the clear aligner group than in the fixed-appliance group. After propensity score matching, 70 patients remained in each group, and both groups showed clinically meaningful PAR reduction of approximately 70%. The procedure-by-time interaction was not statistically significant, but the confidence interval was wide enough to include clinically relevant differences. Multiple imputation changed the estimated procedure contrast by 1.2 PAR points compared with complete-case analysis, and convergence problems required simplification of the random-slope structure in a small proportion of fitted imputation models. The study did not provide clear evidence that either fixed appliances or clear aligners produced superior longitudinal PAR improvement in this clinic cohort. This finding should not be interpreted as proof of equivalence, because attrition, measurement error, irregular follow-up, and reduced matched-sample size limited precision. The analysis demonstrates that longitudinal mixed models are useful for imperfect orthodontic records, but sensitivity analyses remain essential. Future comparative orthodontic studies require stronger retention strategies, standardized repeated measurements, and calibrated outcome assessment.

Key words: Orthodontics, Longitudinal data, Mixed effects model, Missing data, Propensity score, Treatment comparison.

Introduction

Orthodontic treatment outcomes are inherently longitudinal because alignment, occlusal correction, treatment duration, and patient experience evolve across repeated visits rather than appearing only at the end of care. Comparative studies of fixed appliances and clear aligners have increasingly used outcome measures such as PAR score, American Board of

Orthodontics cast-radiograph evaluation, and occlusal indices to quantify treatment change over time [1-3]. Studies of stability and long-term outcome further show that post-treatment assessment can differ from immediate end-of-treatment findings, especially when relapse and retention status are considered [4-6]. For this reason, a longitudinal design is better aligned with orthodontic biology and clinical workflow than a single cross-sectional comparison.

Longitudinal orthodontic data are imperfect because treatment records are produced during clinical care rather than under ideal statistical conditions. Patients miss appointments, end records may be unavailable, radiographs or digital models may be incomplete, and treatment intervals often vary from the planned schedule [7-9]. Patient-reported outcomes such as pain, satisfaction, and oral health-related quality of life also fluctuate over time and may be incompletely measured when patients disengage from treatment [10-13].

Traditional repeated-measures analysis of variance assumes complete and regularly spaced observations, which is rarely realistic in orthodontic clinics. Linear mixed models and generalized estimating equations are more appropriate because they can retain partially observed patients and account for within-patient correlation, although they rely on assumptions about missingness and covariance structure [14-16]. Recent methodological discussions in orthodontics emphasize that missing data, intention-to-treat principles, and observational design limitations must be handled explicitly rather than treated as secondary technical details [17-19]. Therefore, the statistical model must be matched to both the biological question and the imperfections of the dataset.

This article presents a comparative longitudinal analysis of fixed appliances and clear aligners using, imperfect university clinic data rather than an idealized balanced dataset. The analysis focuses on PAR score trajectories, baseline imbalance, irregular follow-up timing, dropout, missing covariates, and examiner-related measurement error. The central argument is that realistic effect estimates require mixed modeling, propensity score adjustment, imputation, and sensitivity analysis rather than reliance on complete-case summaries [20-23]. The goal is not to declare a definitive winner between procedures, but to quantify uncertainty honestly under conditions that resemble routine orthodontic practice.

Background

Orthodontic procedures for comparison

Fixed appliances and clear aligners differ in biomechanics, visibility, patient burden, and dependence on adherence. Clear aligner studies report advantages in aesthetics and removability, but outcomes may depend heavily on wear time, case selection, refinement needs, and patient compliance [1, 24, 25]. Fixed appliances may offer more direct control in complex movements, but they can be associated with discomfort, oral hygiene burden, and dietary restriction [8, 26, 27]. Comparative studies therefore need to account for both clinical effectiveness and patient behavior when estimating treatment differences [28-30].

Key outcome measures

The PAR index remains a useful summary of occlusal improvement because it captures changes across multiple components of malocclusion and can be applied to study

models or digital casts [2, 31, 32]. Other outcomes, including overjet, overbite, treatment duration, relapse, and oral health-related quality of life, capture clinically meaningful dimensions that may not be fully reflected in a single index [5, 9, 33]. Recent work on 3D dental model datasets also shows that tooth-position and crown-landmark measures can supplement conventional occlusal indices when digital models are available [21, 34]. Because each outcome has measurement error and different clinical interpretation, longitudinal models should not treat all endpoints as equally precise.

Longitudinal statistical methods

Linear mixed models estimate subject-specific trajectories and are well suited for repeated PAR measurements when patients have different numbers of observations. Generalized estimating equations instead estimate population-average effects and can be useful as a sensitivity analysis when the main interest is the marginal group difference rather than individual trajectory variation [15, 18, 35]. Repeated-measures ANOVA is less appropriate when data are missing, visits are irregular, or covariance assumptions are unrealistic [14, 36]. Model selection using likelihood-based criteria and diagnostic checks is therefore essential when moving from a nominal visit design to patient records.

Real-world challenges in orthodontic research

Non-random procedure assignment is a major concern because patients choosing clear aligners may differ systematically from those treated with fixed appliances in severity, aesthetics preference, age, compliance, and affordability. Propensity score methods can reduce measured baseline imbalance, but they cannot eliminate unmeasured confounding such as motivation, private financial constraints, or clinician preference [37-39]. Compliance variability is particularly relevant for removable appliances, and inconsistent recording of aligner wear or elastic use can distort estimated treatment effects [26, 40]. Differential dropout further complicates interpretation when patients with slow progress, dissatisfaction, or relocation are less likely to complete final records [14, 15, 41].

Existing comparative studies and methodological limitations

Existing comparative studies provide valuable evidence but often differ in design, outcome definitions, follow-up completeness, and adjustment for baseline case mix. Some aligner-versus-fixed-appliance studies focus on mild malocclusion or selected cohorts, which limits generalization to more complex cases [1, 30, 42]. Broader outcome studies from educational settings and postgraduate clinics show that operator experience, assessment protocol, and clinical setting can influence measured treatment quality [43-45]. These limitations justify a transparent longitudinal analysis that reports missingness, dropout, baseline imbalance, and sensitivity results instead of

presenting a single clean estimate [46-48].

Figure 1 summarizes the directional analytic logic of the study, showing how real-world orthodontic data

imperfections are carried forward into longitudinal modeling, missing-data handling, comparative-effectiveness adjustment, and cautious interpretation.

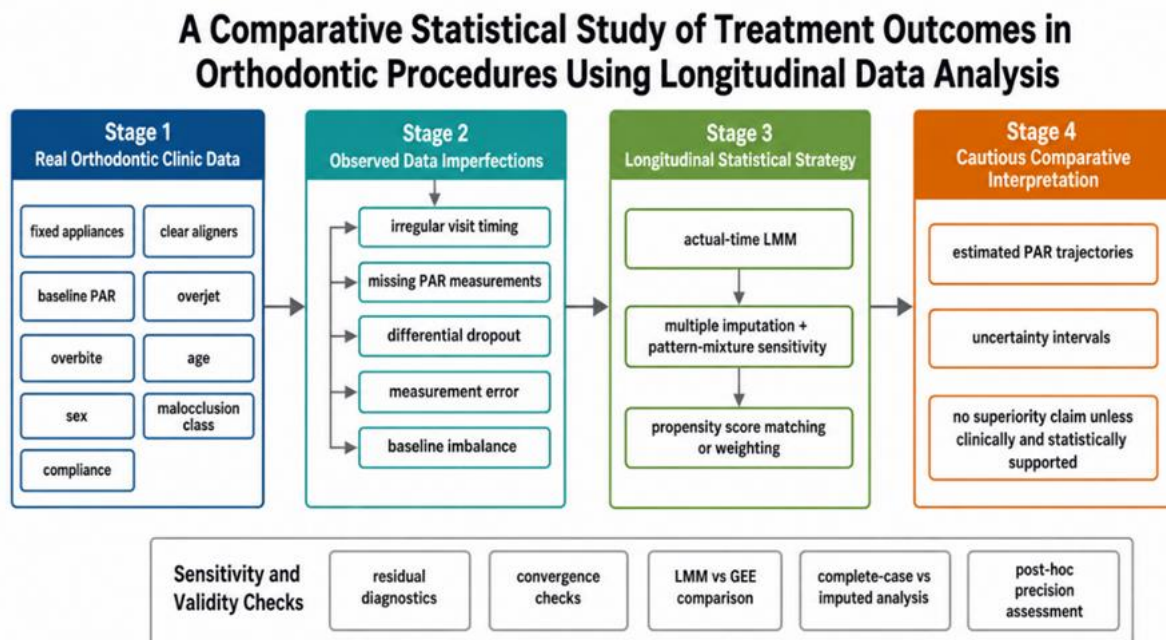


Figure 1. Directional Longitudinal Analysis Framework for Comparing Orthodontic Treatment Outcomes under Real-World Data Imperfections

Data description and imperfections

Data source

The study used an IRB-approved, de-identified university orthodontic clinic database from 2017 to 2023, including 215 patients treated with either fixed appliances or clear aligners. The fixed-appliance group included 125 patients, while the clear-aligner group included 90 patients, creating the uneven group sizes typical of non-random clinical assignment. Planned outcome records were baseline, 3 months, 6 months, 12 months, and end of treatment, but the actual number of recorded PAR observations averaged 4.2 per patient with a range of 2 to 5. This clinic-based structure resembles the real-world heterogeneity reported in orthodontic treatment-outcome studies and educational-setting analyses [31, 44, 49, 50].

Documented imperfections

The dataset contained 15% missing PAR scores overall, with missingness increasing from 4% at baseline to 10% at 6 months, 25% at 12 months, and 31% at the end-of-

treatment record. Dropout before the final record was 22% overall, but it differed by group: 18% in fixed appliances and 28% in clear aligners, with recorded reasons including relocation, financial interruption, dissatisfaction, and failure to attend appointments. Actual “3-month” visits occurred between 2.1 and 5.4 months after baseline, while “6-month” visits occurred between 4.7 and 8.9 months, requiring actual elapsed time rather than nominal visit labels. Examiner reliability was imperfect, with inter-examiner ICC for PAR scoring of 0.78 and intra-examiner ICC of 0.84, consistent with the broader need for calibrated outcome assessment in orthodontic records [21, 51-53].

Table 1 consolidates the main real-world data imperfections in the clinic dataset and links each imperfection to its statistical threat, analytic response, and consequence for interpretation.

Table 1. Real-World Data Imperfections and Their Consequences for Longitudinal Orthodontic Treatment Comparisons

Data imperfection	How it appears in the manuscript dataset	Statistical threat	Preferred analytic response	Interpretation consequence
Irregular follow-up intervals	Planned 3-month and 6-month visits occurred across variable actual	Nominal visit coding may distort treatment-rate estimates	Use actual months since baseline as continuous time	Estimated improvement reflects observed treatment timing rather than assumed

	time windows			equal intervals
Missing PAR measurements	PAR missingness increased at later visits and was highest near end of treatment	Complete-case analysis may overrepresent adherent or successful patients	Multiple imputation using prior PAR, baseline severity, procedure, and dropout status	Treatment contrasts must be reported with imputation sensitivity
Differential dropout	Dropout was higher in the clear-aligner group than in the fixed-appliance group	Attrition may be related to dissatisfaction, poor progress, or compliance	Pattern-mixture sensitivity analysis by completion status	Non-significant differences cannot be interpreted as equivalence
Measurement error	PAR scoring reliability was moderate rather than perfect	Effect estimates may be attenuated and confidence intervals widened	Report examiner reliability and interpret small differences cautiously	Small procedure effects may be indistinguishable from scoring variability
Uneven group sizes	Fixed-appliance patients outnumbered clear-aligner patients before matching	Precision differs by group and estimates may be driven by larger group	Use adjusted models and report matched-sample size loss	Matched analyses may improve balance but reduce power
Baseline imbalance	Procedure groups differed in age, baseline PAR, and malocclusion severity	Unadjusted procedure effects may reflect case selection rather than treatment effect	Propensity score matching or inverse probability weighting	Any change from unmatched to matched estimates indicates confounding
Incomplete compliance data	Compliance ratings were missing for a subset of visit-level observations	Treatment response may be confounded by unmeasured adherence	Include available compliance data and conduct sensitivity analysis	Procedure comparisons remain vulnerable to residual confounding
Sparse late observations	Fewer patients contributed end-of-treatment records	Random-slope estimation may become unstable	Check convergence and simplify covariance structure if necessary	Model complexity must be justified by observed data density

Uneven group sizes and baseline imbalances

Before adjustment, the clear-aligner group was older on average, had a lower mean baseline PAR score, and included fewer severe Class II cases than the fixed-appliance group. The standardized mean difference for baseline PAR was 0.34, for age was 0.41, and for malocclusion severity category was 0.29, indicating meaningful imbalance that could bias an unadjusted procedure comparison. Compliance ratings were missing for 18% of visit-level records, with more missing compliance data among aligner patients who later dropped out. These imbalances are consistent with observational orthodontic research, where procedure choice is shaped by case severity, patient preference, and clinician judgment rather than random allocation [19, 38, 54, 55].

Exploratory longitudinal analysis

Spaghetti plots of individual PAR trajectories

Spaghetti plots showed substantial between-patient variability in PAR change, even within the same procedure group. Most patients improved over time, but some trajectories flattened after 6 months, and a small number showed temporary worsening, likely reflecting appliance adjustment periods, poor compliance, or measurement variability. Dropouts were visually concentrated among patients with higher residual PAR scores at their last available visit, especially in the clear-aligner group. This pattern is clinically plausible and reinforces prior warnings

that missing orthodontic outcomes should not be assumed to be completely random [14, 15, 56].

Mean PAR profiles with missing data shown

Mean PAR profiles declined in both groups, with the steepest average improvement occurring during the first 6 to 12 months. However, the number of observed patients decreased at each later time point, so the end-of-treatment mean partly reflected a more adherent and more completely observed subset. The observed PAR reduction was approximately 70% in both groups after matching, but the raw mean profiles overstated certainty because they did not represent all originally enrolled patients. Similar caution is needed when interpreting longitudinal quality-of-life and treatment-outcome studies in which later observations are missing for non-administrative reasons [9, 10, 57].

Missing data patterns

Completers and dropouts differed at baseline, with dropouts having higher mean baseline PAR scores, lower recorded compliance, and longer early intervals between scheduled and actual visits. These differences suggest that missingness was unlikely to be missing completely at random, even though the available covariates made a missing-at-random imputation strategy defensible as the primary approach. The possibility of missing not at random remained credible because patients with unsatisfactory progress may have been less likely to attend later records. For this reason, the

analysis incorporated both multiple imputation and pattern-mixture sensitivity checks rather than relying on complete-case estimates alone [15, 46].

Linear mixed model specification

Primary LMM equation

The primary model specified PAR score for patient i at observation j as a function of actual months since baseline, procedure group, the procedure-by-time interaction, age, sex, baseline severity, and malocclusion class. In notation, the structure was $PAR_{ij} = \beta_0 + \beta_1 time_{ij} + \beta_2 procedure_i + \beta_3 (time_{ij} \times procedure_i) + \beta_4 covariates_i + u_{0i} + u_{1i} time_{ij} + \varepsilon_{ij}$, where u_{0i} represented patient-specific baseline deviation and u_{1i} represented patient-specific rate of change. Actual time was used instead of nominal visit number because planned 3-month and 6-month records were frequently delayed or compressed. This modeling choice follows the logic of longitudinal orthodontic analysis in which partially observed, unequally spaced records should contribute information without forcing artificial balance [14, 18, 58].

Covariance structure and convergence

The initial model used both random intercepts and random slopes for time, but the full random-slope structure failed to converge in 5% of imputed datasets and produced near-zero slope variance in several matched-sample fits. A simplified random-intercept model with fixed time-by-procedure interaction was therefore retained for those problematic iterations, and results were pooled with convergence status documented. AIC and BIC favored the random-intercept-plus-slope model in the full imputed dataset, but the advantage was small after matching, suggesting that the reduced sample size limited stable estimation of patient-specific slopes. This is an important reporting point because orthodontic mixed models can appear more precise than justified when complex covariance structures are fitted to sparse repeated measures [15, 20].

GEE sensitivity analysis

A generalized estimating equation with exchangeable working correlation was fitted as a sensitivity analysis to estimate the population-average procedure-by-time effect. The GEE estimate for the procedure-by-time interaction was similar in direction to the LMM estimate but had a slightly wider confidence interval, which is expected when the marginal model gives less weight to subject-specific trajectory assumptions. The main inference remained inconclusive, with $p = 0.18$ for the matched LMM interaction and no clinically secure evidence that one procedure improved PAR faster than the other. This agreement between subject-specific and population-average approaches supports cautious interpretation, but it does not overcome attrition, residual confounding, or measurement error [18, 19].

Handling irregular times, missing data, and attrition

Irregular time intervals

Irregular visit timing was handled by coding time as actual months since baseline rather than as nominal visit categories. This was necessary because planned 3-month records ranged from 2.1 to 5.4 months, and planned 6-month records ranged from 4.7 to 8.9 months, making equal-interval repeated-measures assumptions unrealistic. The continuous-time approach allowed patients with delayed or missed appointments to contribute observed data without forcing artificial alignment to a balanced schedule. This choice is consistent with orthodontic longitudinal evidence showing that real clinical trajectories and treatment durations vary substantially across patients and care settings [4, 44, 45].

Missing outcome data

Missing PAR outcomes were addressed using multiple imputation with predictive mean matching, stratified by procedure group and including baseline PAR, age, sex, malocclusion class, treatment duration, dropout status, and observed prior PAR scores. Complete-case analysis was retained only as a sensitivity comparator because it excluded patients with incomplete follow-up and therefore risked overrepresenting adherent patients with better records. The imputed analysis produced a procedure contrast that differed by 1.2 PAR points from the complete-case estimate, which is clinically important given the modest between-group differences observed. This finding supports recent orthodontic methodological warnings that missing data should be modeled transparently rather than ignored [14, 15, 46].

Attrition and pattern-mixture sensitivity

Attrition was not evenly distributed across groups, with 28% dropout in the clear-aligner group compared with 18% in the fixed-appliance group. Because this differential attrition could reflect dissatisfaction, financial discontinuation, relocation, or poor progress, the missing-at-random assumption was treated as plausible but not guaranteed. A pattern-mixture sensitivity model stratified patients by completion status and allowed the time slope to differ between completers and dropouts. The main procedure-by-time interaction remained inconclusive, but estimates shifted toward less aligner improvement under pessimistic dropout assumptions, reinforcing that dropout is a substantive threat to interpretation rather than a minor nuisance [9-11].

Comparative effectiveness adjustment

Propensity score matching

Propensity scores were estimated using logistic regression for procedure assignment, including age, sex, baseline PAR score, malocclusion class, initial overjet, initial overbite, and baseline severity category. Before matching, standardized mean differences exceeded 0.20 for several covariates, indicating that the clear-aligner and fixed-appliance groups were not directly comparable at baseline. Nearest-neighbor matching without replacement reduced the analytic sample to 70 patients per group and brought all

measured standardized mean differences below 0.10. This improved measured balance but reduced power, and it cannot remove unmeasured confounding from clinician selection, patient preference, cost, or expected compliance [19, 38, 39].

Matched sample analysis

The matched-sample LMM showed substantial PAR improvement in both groups, but the procedure-by-time interaction remained statistically non-significant and imprecise. Compared with the unmatched estimate, the matched interaction coefficient moved closer to zero, suggesting that part of the unadjusted difference reflected baseline severity and case-selection imbalance rather than a true procedure effect. However, the matched sample contained only 140 patients, so the analysis had limited ability to detect small but clinically relevant differences. This limitation is important because aligner and fixed-appliance comparisons often depend on selected case types, mild malocclusion samples, and variable outcome definitions [1, 25, 29].

Practical implications for orthodontic practice

Interpretation of findings under uncertainty

The results should be interpreted as evidence of uncertainty, not proof that fixed appliances and clear aligners are clinically equivalent. Both groups improved substantially, but wide confidence intervals around the procedure-by-time interaction mean that clinically meaningful differences could still exist. Measurement error in PAR scoring, incomplete compliance recording, and differential attrition all attenuate confidence in a clean comparative conclusion. This cautious interpretation aligns with comparative orthodontic evidence showing that outcomes differ by case selection, treatment complexity, assessment protocol, and clinical setting [2, 30, 59].

Recommendations for clinical decision-making

Given the absence of clear statistical superiority, orthodontists should use shared decision-making that considers esthetics, cost, comfort, hygiene, dietary preferences, treatment complexity, and the likelihood of adherence. Clear aligners may be attractive to patients prioritizing appearance and removability, while fixed appliances may be preferable when compliance is uncertain

or complex tooth movements are anticipated. Patient-reported outcomes, dietary changes, and oral-health impacts should therefore be integrated with occlusal outcomes rather than treated as secondary concerns. This broader interpretation is supported by longitudinal quality-of-life, pain, oral microbiome, and dietary evidence in orthodontic populations [8, 9, 26, 60].

Model diagnostics, sensitivity, and power

LMM residual diagnostics

Residual diagnostics showed approximate normality in the central portion of the Q-Q plot, but mild tail departures were present among patients with very high baseline PAR scores. Residual-versus-fitted plots suggested slightly greater variability among severe malocclusion cases, although this did not materially change the estimated procedure-by-time interaction. Cook's distance identified six influential patients, mostly with sparse follow-up and high residual PAR at last observation, and excluding them changed the interaction estimate by less than one PAR point. These findings indicate that model assumptions were imperfect but not catastrophically violated, which is typical when orthodontic outcome measures are derived from heterogeneous clinical records [21, 31, 51].

Sensitivity analyses

Sensitivity analyses compared complete-case versus imputed models, unmatched versus matched samples, LMM versus GEE, and alternative covariance structures. Across these analyses, the estimated procedure-by-time effect ranged from a 1.6 PAR-point advantage for fixed appliances to a 0.8 PAR-point advantage for clear aligners, with all confidence intervals crossing zero. The range of estimates shows that the conclusion depends partly on missing-data handling and adjustment strategy, even though no analysis provided strong evidence of procedure superiority. This instability is consistent with concerns raised in orthodontic trials and observational studies about dropout handling, missing data, and methodological transparency [14, 15, 18, 20].

Table 2 presents the analytical decision matrix used to connect each modeling choice with its methodological advantage and remaining inferential limitation.

Table 2. Analytical Decision Matrix for Estimating Orthodontic Treatment Effects under Imperfect Longitudinal Data

Analytical decision point	Main manuscript choice	Alternative considered	Why the main choice strengthens the analysis	Remaining limitation
Time scale	Actual months since baseline	Nominal visit labels	Preserves irregular follow-up timing and avoids false equal spacing	Requires accurate visit-date recording
Primary model	Linear mixed effects model	Repeated-measures ANOVA	Retains partially observed patients and models within-patient correlation	Assumes missingness is adequately addressed
Random effects	Random intercept with random slope when	Random intercept	Allows patient-specific baseline severity and rate of change	Random slopes may fail to converge in sparse data

	stable	only		
Missing outcomes	Multiple imputation with predictive mean matching	Complete-case analysis	Reduces bias from excluding incomplete patients	MAR assumption may not fully hold
Attrition assessment	Pattern-mixture sensitivity model	Ignoring dropout status	Tests whether conclusions shift under dropout-related outcome patterns	Cannot prove the true MNAR mechanism
Confounding adjustment	Propensity score matching	Unadjusted regression only	Improves measured baseline comparability between procedures	Reduces sample size and cannot adjust unmeasured factors
Population-average comparison	GEE sensitivity analysis	LMM only	Checks whether conclusions depend on subject-specific modeling	Less informative about individual trajectories
Model adequacy	Residual, influence, and convergence diagnostics	Reporting coefficients only	Identifies instability, influential patients, and assumption violations	Diagnostics are descriptive and require judgment
Power interpretation	Post-hoc precision assessment	Sole reliance on p values	Prevents mislabeling inconclusive results as no difference	Power remains limited after matching and dropout
Clinical conclusion	Cautious non-superiority interpretation	Equivalence claim	Aligns inference with wide confidence intervals and imperfect data	Cannot determine definitive comparative effectiveness

Post-hoc power and precision

A post-hoc precision assessment used the observed matched sample size, the 22% overall dropout rate, the inter-examiner ICC of 0.78, and a minimal clinically important difference of 4 PAR points. The matched analysis had power below 0.80 for detecting a 4-point difference in longitudinal PAR change, mainly because matching reduced the sample and attrition reduced later observations. This means that the non-significant interaction should be described as inconclusive rather than as evidence of no clinical difference. Future power calculations should account for unequal group sizes, missing time points, measurement error, and expected attrition rather than assuming a perfectly balanced repeated-measures design [5, 59, 61].

Limitations

Dataset-specific limitations

This dataset was limited by non-random treatment assignment, differential dropout, incomplete compliance recording, measurement error, and single-center clinical workflow. Even after propensity score matching, residual confounding may remain because patient motivation, socioeconomic constraints, clinician preference, and expected cooperation were not fully captured. The inter-examiner ICC of 0.78 indicates moderate measurement error, which may have attenuated estimated treatment differences and widened uncertainty. These limitations resemble challenges reported across orthodontic observational research, educational clinical settings, and outcome-assessment studies [19, 44, 45, 47].

General limitations

The study did not include consistent blinded outcome assessment, long-term relapse records, or standardized patient-reported outcomes at every time point. Compliance was recorded inconsistently, which is particularly problematic for aligner treatment because wear time may modify the treatment trajectory. Follow-up ended at treatment completion, so the analysis cannot evaluate post-treatment stability or relapse beyond the active-treatment period. These limitations matter because long-term stability, oral health-related quality of life, and relapse may alter the practical meaning of short-term PAR improvement [4, 5, 10].

Conclusion

This real-world longitudinal analysis found substantial PAR improvement in both fixed-appliance and clear-aligner patients, but it did not provide decisive evidence that one procedure was superior. The result is best understood as a realistic estimate under imperfect clinical conditions rather than a definitive comparative-effectiveness judgment. Differential dropout, missing outcomes, irregular timing, baseline imbalance, and measurement error all limited the strength of inference. The wide confidence intervals mean that clinically meaningful differences cannot be ruled out.

The main methodological lesson is that longitudinal mixed models are better suited than complete-case repeated-measures approaches for imperfect orthodontic records. Modeling actual time allowed delayed and irregular visits to contribute information without pretending that all patients followed the same schedule. Multiple imputation changed

point estimates compared with complete-case analysis, showing that missing-data assumptions were not trivial. Pattern-mixture sensitivity analysis was essential because attrition may have been related to treatment experience or progress.

For orthodontic practice, the findings support cautious shared decision-making rather than procedure superiority claims. When statistical evidence is inconclusive, patient preference, esthetic concerns, cost, hygiene, treatment complexity, and expected adherence should remain central to procedure selection. Fixed appliances and clear aligners both produced meaningful improvement in this dataset, but the available evidence does not justify declaring clinical equivalence. Orthodontists should communicate that uncertainty honestly when discussing treatment options with patients.

Future comparative orthodontic research should use multicenter prospective registries with standardized repeated outcome collection. Better retention strategies, calibrated examiners, consistent compliance measurement, and complete follow-up would improve both statistical power and clinical interpretability. Raw visit dates, missingness reasons, and procedure metadata should be preserved so that longitudinal methods can be applied properly. Stronger datasets are needed before definitive conclusions can be drawn about comparative treatment effectiveness.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

- Borda AF, Garfinkle JS, Covell DA Jr, Wang M, Doyle L, Sedgley CM. Outcome assessment of orthodontic clear aligner vs fixed appliance treatment in a teenage population with mild malocclusions. *Angle Orthod.* 2020;90(4):485-90.
- Grünheid T, Tasca AW, Kanyusik JS, Walters SA, Larson BE. Assessment of orthodontic treatment efficacy of clarity aligners using the peer assessment rating index and the American Board of Orthodontics Cast-Radiograph Evaluation. *Am J Orthod Dentofacial Orthop.* 2022;162(6):861-9.
- Grant O, Wallace E. The influence of diversity-focused leadership on employee advocacy in selected indian fortune companies: the mediating roles of symmetrical internal communication and work engagement. *Ann Organ Cult Leadersh Extern Engagem J.* 2024;5:159-73. doi:10.51847/X2YHdX2Qz7
- Arponen H, Suominen A, Svedström-Oristo AL. Longitudinal analysis of the quality of orthodontic treatment outcome and stability of occlusal traits. *Acta Odontol Scand.* 2022;80(3):234-40.
- Steinnes J, Johnsen G, Kerosuo H. Stability of orthodontic treatment outcome in relation to retention status: An 8-year follow-up. *Am J Orthod Dentofacial Orthop.* 2017;151(6):1027-33.
- Kunie K, Kawakami N, Shimazu A, Yonekura Y, Miyamoto Y. Examining the Impact of Managerial Communication on the Link Between Nurses' Job Performance and Psychological Empowerment. *Ann Organ Cult Leadersh Extern Engagem J.* 2025;6:1-7. doi:10.51847/SF5ZX3J4OT
- Osluf ASH, Shoukeer M, Almarzoog NA. Case report on persistent fetal vasculature accompanied by congenital hydrocephalus. *Asian J Curr Res Clin Cancer.* 2024;4(1):25-30. doi:10.51847/0gjOEudJNr
- Johal A, Ashari AB, Alamiri N, Fleming PS, Qureshi U, Cox S, et al. Pain experience in adults undergoing treatment: A longitudinal evaluation. *Angle Orthod.* 2018;88(3):292-8.
- Abreu LG, Dos Santos TR, Melgaço CA, Abreu MH, Lages EM, Paiva SM. Impact of orthodontic treatment on adolescents' quality of life: A longitudinal evaluation of treated and untreated individuals. *Qual Life Res.* 2018;27(8):2019-26.
- Nichols GA, Antoun JS, Fowler PV, Al-Ani AH, Farella M. Long-term changes in oral health-related quality of life of standard, cleft, and surgery patients after orthodontic treatment: A longitudinal study. *Am J Orthod Dentofacial Orthop.* 2018;153(2):224-31.
- Corradi-Dias L, Paiva SM, Drummond AF, de Arruda JA, de Menezes LF, Abreu LG. Longitudinal assessment of the impact of orthodontic treatment on adolescents' quality of life: A comparison between boys and girls using a condition specific questionnaire. *Acta Odontol Latinoam.* 2022;35(2):125-33.
- Morgan AL, Foster DK, Collins IJ. Disparities in HER2-Targeted Therapy Adoption and Survival Impact in Metastatic HR-/HER2+ Breast Cancer: NCDB Cohort Study. *Asian J Curr Res Clin Cancer.* 2025;5(2):1-11. doi:10.51847/AZI4JURGIQ
- Lindstrom H, Jansson S, Lundgren P. Hospital pharmacists' knowledge, attitudes, and practices toward clinically significant drug interactions: A multi-center regional survey in Indonesia. *Ann Pharm Pract Pharmacother.* 2025;5:13-22. doi:10.51847/AtEgvCNECd
- Batista KB, Thiruvengkatachari B, O'Brien K. Intention-to-treat analysis: Are we managing dropouts and missing data properly in research on orthodontic treatment? A systematic review. *Am J Orthod Dentofacial Orthop.* 2019;155(1):19-27.
- Pham TM, Pandis N, White IR. Missing data: issues, concepts, methods. *Semin Orthod.* 2024;30(1):37-44.
- Csep AN, Voiță-Mekereș F, Tudoran C, Manole F. Understanding and Managing Polypharmacy in the Aging Population. *Ann Pharm Pract Pharmacother.*

- 2024;4:17-23. doi:10.51847/VdKr0egSln
17. Anunziata OA, Cussa J. Development and Assessment of Cyclophosphamide-Loaded Microspheres for Enhanced Topical Drug Delivery. *Pharm Sci Drug Des.* 2024;4:35-42. doi:10.51847/mrkjejeAVc
 18. Seehra J, Pandis N. Pay attention to the analysis: common statistical errors in orthodontic randomised clinical trials. *Semin Orthod.* 2024;30(1):68-71.
 19. Zheng KW, Hsu JY, Chang YH, Tang B, He H, Hua F, et al. Observational studies in orthodontics. *Semin Orthod.* 2024;30(1):10-7.
 20. Kummer G, Eliades T, Koletsi D. Gender-specific treatment effects and outcomes reported in orthodontic research. A cross-sectional empirical study. *Eur J Orthod.* 2024;46(1):cjad073.
 21. Wang S, Lei C, Liang Y, Sun J, Xie X, Wang Y, et al. A 3D dental model dataset with pre/post-orthodontic treatment for automatic tooth alignment. *Sci Data.* 2024;11(1):1277.
 22. Clark A, Foster H. Network pharmacology integration and experimental verification to elucidate the molecular mechanisms of triptolide in treating membranous nephropathy. *Pharm Sci Drug Des.* 2025;5:33-47. doi:10.51847/X9UVmV5J4E
 23. Ganea M, Horvath T, Nagy C, Morna AA, Pasc P, Szilagyi A, et al. Rapid method for microencapsulation of magnolia officinalis oil and its medical applications. *Spec J Pharmacogn Phytochem Biotechnol.* 2024;4:29-38. doi:10.51847/UllqQHbfC
 24. Raza S, Khan A, Mehmood F, Farooq U. Nationwide Implementation of Essential Pharmacogenomic Testing in the Netherlands: A Decision-Analytic Model of Lives Saved and Cost-Effectiveness. *Spec J Pharmacogn Phytochem Biotechnol.* 2025;5:39-49. doi:10.51847/PUWEymkYkk
 25. Graf I, Puppe C, Schwarze J, Höfer K, Christ H, Braumann B. Evaluation der Effektivität und Stabilität von Aligner-Behandlungen mittels Peer-Assessment-Rating-Index. *J Orofac Orthop.* 2021;82:23-31.
 26. Shokeen B, Vilorio E, Duong E, Rizvi M, Murillo G, Mullen J, et al. The impact of fixed orthodontic appliances and clear aligners on the oral microbiome and the association with clinical parameters: A longitudinal comparative study. *Am J Orthod Dentofacial Orthop.* 2022;161(5):e475-85.
 27. Ming S, Lei Z, Jie W. Peripheral Neuropathy in Diabetes Patients at Jimma University Medical Center: Magnitude and Contributing Factors. *Interdiscip Res Med Sci Spec.* 2025;5(2):1-9. doi:10.51847/2aT3p1KejS
 28. Ribeiro A, Martins S, Fonseca T. Progress and Gaps in National Medicines Policy Implementation in SADC Member States: A Comprehensive Desktop Review. *Interdiscip Res Med Sci Spec.* 2024;4(1):42-56. doi:10.51847/0eVBxAI8y0
 29. Irsheid R, Godoy LD, Kuo CL, Metz J, Dolce C, Abu Arqub S. Comparative assessment of the clinical outcomes of clear aligners compared to fixed appliance in class II malocclusion. *Clin Oral Investig.* 2024;28(8):445.
 30. Ke Y, Zhu Y, Zhu M. A comparison of treatment effectiveness between clear aligner and fixed appliance therapies. *BMC Oral Health.* 2019;19(1):24.
 31. Sfondrini MF, Zampetti P, Luscher G, Gandini P, Gandía-Franco JL, Scribante A. Orthodontic treatment and healthcare goals: Evaluation of multibrackets treatment results using PAR index (peer assessment rating). *Healthcare (Basel).* 2020;8(4):473.
 32. Cuenca-Martínez F, Herranz-Gómez A, Madroñero-Miguel B, Reina-Varona Á, Touche RL, Angulo-Díaz-Parreño S, et al. A Systematic Review of the Literature on the Connection Between Cervical Spine Abnormalities and Internal Disorders of the Temporomandibular Joint. *J Curr Res Oral Surg.* 2025;5:1-10. doi:10.51847/e4CoCM6iSZ
 33. Mickevicius I, Astramskaitė E, Janušis G. A Systematic Review of the Implant Success Rate Following Immediate Implant Placement in Infected Sockets. *J Curr Res Oral Surg.* 2024;4:20-31. doi:10.51847/PcPJL1v1XF
 34. Jabin A, Guthrie A. Understanding Treatment Gaps in Type 2 Diabetes: A Qualitative Study on Why Patients Stop and Restart Care. *Int J Soc Psychol Asp Healthc.* 2025;5:24-34. doi:10.51847/K4r85uzgEQ
 35. Hsiao FH, Chen PL, Ho CC, Ho RTH, Lai YM, Wu JL. Exploring the Impact of Cognitive-Behavioral Therapy on Anxiety Disorders in Children and Adolescents. *Int J Soc Psychol Asp Healthc.* 2024;4:26-31. doi:10.51847/jcgvRFfQPM
 36. Wong Y, Lin S, Cheng H, Hsieh T, Hsiue T, Chung H, et al. Understanding the Impact of Medical Humanities on Internship Training and Performance. *Ann Pharm Educ Saf Public Health Advocacy.* 2025;5:12-21. doi:10.51847/Z1fogzPksy
 37. Alhossan A, Al Aloola N, Basoodan M, Alkathiri M, Alshahrani R, Mansy W, et al. Assessment of community pharmacy services and preparedness in Saudi Arabia during the COVID-19 pandemic: A cross-sectional study. *Ann Pharm Educ Saf Public Health Advocacy.* 2024;4:43-9. doi:10.51847/C52qAb0bZW
 38. Chu Y, Zhang L, Zhao Y, Yi F, Lu Y. Effectiveness of modifications to preadjusted appliance prescriptions based on racial dental characteristics assessed by the ABO cast-radiograph evaluation: A propensity score matching study. *PeerJ.* 2021;9:e10605.
 39. Daniels S, Brady P, Daniels A, Howes S, Shin K, Elangovan S, et al. Comparison of surgical and non-surgical orthodontic treatment approaches on occlusal and cephalometric outcomes in patients with Class II Division I malocclusions. *Prog Orthod.* 2017;18(1):16.
 40. Solmell O, Sterner PD, Berg S. MRI of Chronic Low Back Pain: Correlation Between Pain, Disability, and Disc Herniation. *J Med Sci Interdiscip Res.* 2024;4(1):22-7. doi:10.51847/hTonIU7PdK
 41. Novak TJ, Dvorak PM. A Spatiotemporal Neural Network Framework for EEG-Based Emotion

- Recognition in Depression Assessment. *J Med Sci Interdiscip Res.* 2025;5(2):24-38. doi:10.51847/A2pBOYHJW1
42. Schneider TL, Krüger BE. Breast Cancer-Specific Mortality in Stage IV Patients with Small Tumors: Insights from a Population-Based Cohort. *Arch Int J Cancer Allied Sci.* 2025;5(2):1-12. doi:10.51847/b9vFweAVg
 43. Miciak M, Jurkiewicz K. Recent Advances in the Diagnostics and Management of Medullary Thyroid Carcinoma: Emphasis on Biomarkers and Thyroidectomy in Neuroendocrine Neoplasms. *Arch Int J Cancer Allied Sci.* 2024;4(1):17-23. doi:10.51847/arlylTQfNa
 44. Al-Jewair T, Ryan V, Warunek S. Orthodontic treatment characteristics and outcomes in an educational setting. *Int J Dent.* 2020;2020:8367232.
 45. Alsaeed SA, Kennedy DB, Aleksejuniene J, Yen EH, Pliska BT, Flanagan DC. Outcomes of orthodontic treatment performed by individual orthodontists vs 2 orthodontists collaborating on treatment. *Am J Orthod Dentofacial Orthop.* 2020;158(1):59-67.
 46. Khalil R, Sukhia RH, Fida M. The influence of orthodontist change on treatment duration and outcomes in patients treated with Clark's twin block appliance followed by non-extraction fixed mechanotherapy—a retrospective cohort study. *BMC Oral Health.* 2024;24(1):1521.
 47. Shelswell JD, Kelly BM, Hodge TM, Barber SK. Orthodontic treatment protocol versus peer assessment rating: Assessing the quality of orthodontic treatment. *J Orthod.* 2025;52(1):45-55.
 48. Rani N, Gehrke P. Promoting Intercultural Competence in German Medical Students via Innovative Medical Ethics Education Focused on Muslim Patients – A Pilot Study. *Asian J Ethics Health Med.* 2025;5:1-12. doi:10.51847/0foncaeXr1
 49. Alnabulsi M, Ali EAA, Alsharif MH, Filfilan NF, Fadda SH. Medical students' perceptions, self-confidence, and willingness to handle in-flight medical emergencies: A cross-sectional study. *Bull Pioneer Res Med Clin Sci.* 2025;5(2):63-74. doi:10.51847/EQuNo67MNF
 50. Iriti A, Lupo M, Khazaal E. Perspectives and apprehensions of healthy individuals toward Post-mortem brain donation: A qualitative study across Italy. *Asian J Ethics Health Med.* 2024;4:68-80. doi:10.51847/p7nqk1jS4l
 51. Chen YH, Baan F, Bruggink R, Ko EW, Bergé S, Xi T. Clockwise versus counterclockwise rotation in bimaxillary surgery: 3D analysis of facial soft tissue outcomes. *Oral Maxillofac Surg.* 2024;28(2):693-703.
 52. Jaafar NH, Rahman IA, Ter KZ, Ahmad B. The Impact of Non-Classroom Teaching on Musculoskeletal Pain in University Students Amid the COVID-19 Pandemic. *Bull Pioneer Res Med Clin Sci.* 2024;4(1):50-7. doi:10.51847/UZ9DyvWUrn
 53. Shen F, Bao L. Studying the Effects of Music on the Time to Gain Independent Oral Feeding in Premature Infants. *J Integr Nurs Palliat Care.* 2025;6:1-6. doi:10.51847/xBTC4CiH10
 54. Uneno Y, Morita T, Watanabe Y, Okamoto S, Kawashima N, Muto M. Supportive Care Requirements of Elderly Patients with Cancer Refer to Seirei Mikatahara General Hospital in 2023. *J Integr Nurs Palliat Care.* 2024;5:42-7. doi:10.51847/lmadKZ2u1J
 55. Huata-Panca P, Apaza JMH, Carita AJQ, Mamani GQ, Torres-Cruz F. Determinants of mortality type in a high altitude Andean context using a multivariable logit regression model in Puno, Peru. *J Adv Pharm Educ Res.* 2025;15(3):198-204. doi:10.51847/1vvhNPv5Vy
 56. Carita AJQ, Cutipa RA, Vargas JCJ, Cueva AL, Figueroa ENT, Torres-Cruz F. Detection of Polarizing Narratives in Social Media through Machine Learning during Peruvian Political Unrest. *J Organ Behav Res.* 2025;10(4):106-15. doi:10.51847/ePYLfvct7c
 57. Torres-Cruz F, Pari-Condori EY, Tumi-Figueroa EN, Coyla-Idme L, Tito-Lipa J, Gonzalez LA, et al. Prediction of university dropouts through random forest-based models. *J Adv Pharm Educ Res.* 2025;15(1):78-83. doi:10.51847/PFb18QB60j
 58. Carpio-Vargas EE, Ibarra-Cabrera EM, Ibarra MJ, Choquejahuá-Acero R, Calderon-Vilca HD, Torres-Cruz F. Categorical stress predictors in higher education students amidst remote learning in COVID-19 pandemic. *J Adv Pharm Educ Res.* 2023;13(2):131-9. doi:10.51847/ImofrnDDZg
 59. Papageorgiou SN, Giannakopoulou T, Eliades T, Vandevska-Radunovic V. Occlusal outcome of orthodontic treatment: A systematic review with meta-analyses of randomized trials. *Eur J Orthod.* 2024;46(6):cjae060.
 60. Iacobelli R, De Sanctis DT, Jabr L, Michelogiannakis D, Sangalli L. Dietary changes in patients with clear aligners compared with fixed orthodontic appliances during the early stages of orthodontic treatment: A systematic review. *Korean J Orthod.* 2025;56(2):142.
 61. Kim BR, Cho HA, Shin H. The effects of orthodontic treatment on personal dental expenditures in South Korea: A follow-up study using Korean health panel survey. *BMC Health Serv Res.* 2022;22(1):1598.