Statistical Analysis Plan 2.01.2022

Assessment of the Impact of the Probiotic Limosilactobacillus Reuteri Prodentis (DSM 17938 and ATCC PTA 5289) on Clinical and Microbiological Parameters in Orthodontically Treated Patients with Concomitant Gingivitis

1. Materials and methods

1.1 Study objections The primary objective of this observational study was to evaluate the longitudinal effects of probiotic therapy as an adjunct to nonsurgical periodontal treatment on key clinical parameters in patients with periodontitis, including probing depth, full mouth plaque index, and bleeding on probing. The secondary objective was to describe the dynamics of nine microbiological parameters (including three key periodontal pathogens: Porphyromonas gingivalis, Treponema denticola, Tannerella forsythia) over time in both the probiotic treatment and control groups.

1.2 Statistical analysis Significance level, descriptive statistics and statistical tests A significance level of $\alpha = 0.05$ was adopted for all inferential tests, with two-tailed p- values reported to assess statistical significance. Descriptive statistics for numerical variables were reported as median (first quartile [Q1], third quartile [Q3]) with minimum and maximum values [min, max], supplemented by 95% confidence intervals for medians. For categorical variables, frequencies were summarized as counts (n) and percentages (%), with 95% confidence intervals for proportions. Comparisons between two independent groups were performed using Pearson's chi-squared test for categorical variables and the Wilcoxon rank sum for numerical variables (Pearson, 1900; Wilcoxon, 1945). Confidence intervals for proportions were calculated using the Wilson score method, which provides reliable coverage for small samples and proportions near 0 or 1 (Wilson, 1927). Confidence intervals for medians were derived using a distribution-free approach based on the Wilcoxon signed-rank statistic (Hodges & Lehmann, 1963).

Propensity score weighting

2 Entropy balancing was employed to derive propensity weights, achieving baseline covariate balance between the probiotic treatment and control groups. This enabled unbiased estimation of the average treatment effect (ATE) over time by aligning means of key covariates: sex (male), age (years), probing depth (mm), full mouth plaque index (%), and bleeding on probing (%). The method mitigated confounding by taking into account the clinical reliability of measurements, the availability of comprehensive baseline data, and the balanced sample sizes across groups, thereby facilitating reliable longitudinal group comparisons. Propensity weights were derived through entropy balancing (Hainmueller, 2012). Weight ranges were examined to confirm they fell within appropriate bounds that avoid excessive variance inflation – typically, ratios of maximum to minimum weights below 10 are considered acceptable to maintain estimation stability (Cole & Hernán, 2008).

Statistical Analysis Plan 2.01.2022

Assessment of the Impact of the Probiotic Limosilactobacillus Reuteri Prodentis (DSM 17938 and ATCC PTA 5289) on Clinical and Microbiological Parameters in Orthodontically Treated Patients with Concomitant Gingivitis

Weight diagnostics included assessment of the coefficient of variation, mean absolute deviation, entropy, and verification of no zero weights to ensure well-behaved distributions. Effective sample sizes (ESS) were calculated,

representing the equivalent unweighted sample size yielding comparable precision (Kish, 1965). Post-weighting ESS values were compared to unweighted sizes to evaluate efficiency loss, with reductions below 10% generally deemed acceptable for preserving statistical power. Covariate balance was evaluated using standardized mean differences (SMD), computed with pooled standard deviations for robustness. A threshold of Statistical Analysis Plan 0.10 was applied to denote acceptable balance, as values exceeding this may indicate residual confounding (Austin, 2009). Unweighted and weighted SMDs were compared to assess improvements in balance. To visualize these changes, a Love plot was constructed, displaying absolute SMDs before and after balancing, with a threshold line at 0.10 to highlight acceptable levels (Zhang et al., 2019). Regression analysis Clinical outcomes (probing depth, full mouth plaque index, and bleeding on probing) were analyzed using mixed-effects models to estimate the average treatment effect over time. Propensity score weights derived from entropy balancing (Hainmueller, 2012) were incorporated to adjust for baseline covariate imbalances between the probiotic treatment and control groups. The models included fixed effects for treatment group (probiotic vs. control), time point (baseline vs. 3-month follow-up), and their interaction.

3 For probing depth, a mixed model for repeated measures (MMRM) with Gaussian distribution and an unstructured covariance matrix was used (Refer to Equation [1]) to account for within-subject correlations (Mallinckrodt et al., 2008; Cnaan et al., 1997). For full mouth plaque index and bleeding on probing (bounded proportions), beta regression mixed models (Refer to equations [2.1, 2.2]) were employed with a logit link function and random intercepts for subjects (Ferrari & Cribari-Neto, 2004). Proportions were adjusted using the Smithson and Verkuilen transformation to avoid boundary issues (Smithson & Verkuilen, 2006). All models incorporated the propensity score weights to enhance causal inference. The equation form of the MMRM for probing depth (mm) was defined according to [1]: where Y_{ij} is the outcome for subject i at time j , G_i is the treatment group indicator, T_j is the time indicator, u_i represents the subject-specific random effects with an unstructured covariance matrix, and ϵ_{ij} is the residual error assumed normally distributed.

For the beta regression mixed models applied to full mouth plaque index and bleeding on probing (proportions), the models were described by equations [2.1, 2.2]: where μ ij is the expected proportion, ϕ is the precision parameter, and u i is the random intercept for subject i. Post-hoc contrasts were performed uniformly across all models to evaluate treatment effects. Estimated marginal means and 95% confidence intervals were computed at each time point for both groups. Contrasts assessed group differences at baseline and 3-month follow- up, as well as changes over time within each group (3-month follow-up minus baseline). Inference relied on a Wald t-distribution with Satterthwaite degrees of freedom approximation (Satterthwaite, 1946). The changes within groups over time were quantified using Cohen's d effect size, with 95% confidence intervals, where values of 0.2, 0.5, and 0.8 are conventionally interpreted as small, medium, and large effects, respectively (Cohen, 1988).

1.3 Characteristic of the statistical tool

4 Analyses were conducted using the R Statistical language (version 4.5.2; R Core Team, 2025) on Windows 11 Pro 64 bit (build 26100), using the packages gridExtra (version 2.3; Auguie B, 2017), glmmTMB (version 1.1.13; Brooks ME et al., 2017), ggalluvial (version Statistical Analysis Plan 0.12.5; Brunson JC, Read QD, 2023), rio (version 1.2.4; Chan C et al., 2023), cobalt (version 4.6.1; Greifer N, 2025), WeightIt (version 1.5.1; Greifer N, 2025), rstatix (version 0.7.3; Kassambara A, 2025), emmeans (version 2.0.0; Lenth R, Piaskowski J, 2025), parameters (version 0.28.3; Lüdecke D et al., 2020), report (version 0.6.2; Makowski D et al., 2023), mmrm (version 0.3.17; Sabanes Bove D et al., 2026), gtsummary (version 2.4.0; Sjoberg D et al., 2021), ggplot2 (version 4.0.1; Wickham H, 2016), and dplyr (version 1.1.4; Wickham H et al., 2023).