

STATISTICAL ANALYSIS PLAN

Effectiveness of a WeChat Mini-Program–Delivered Parent-Focused Behavioral Sleep Promotion Intervention for Preschool Children: A Randomized Waitlist-Controlled Trial

Version	1.1
Date Finalised	August 25, 2026
Trial Registration	ClinicalTrials.gov: NCT number pending
Principal Investigator	Dr. Ying DAI, School of Nursing, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China
Statistician(s)	Dr. Ying DAI
Ethics Approval	2025108A01
Funding	National Natural Science Foundation of China (NSFC)

This Statistical Analysis Plan (SAP) prespecifies the statistical analyses for the Sleep Island randomized controlled trial. It is intended to be finalized before database lock and before examination of treatment-effect results. Any subsequent deviations from this SAP will be documented, justified, and identified as post hoc in study reports.

1. Background and Objectives

1.1 Background

Sleep problems are common during the preschool years and are influenced by modifiable parental practices, bedtime routines, parent-child interactions, and the home sleep environment. Sleep Island is a WeChat mini-program adapted from an Intervention Mapping-based parent-child behavioral sleep promotion program. The digital intervention includes electronic sleep monitoring, graphical sleep-trend visualization, sleep-behavior feedback, sleep-health education, and individualized recommendations. Personalized recommendations are generated using a large language model supported by retrieval-augmented generation and are reviewed by the research team before release. A prior feasibility study supported progression to a longer controlled evaluation.

1.2 Objectives

The primary objective is to determine whether the 4-week Sleep Island intervention, compared with routine sleep-health education alone, improves objectively measured nocturnal sleep duration in preschool children.

The secondary objectives are to evaluate intervention effects on actigraphy-measured sleep efficiency, actigraphy-measured night wakings, parent-reported child sleep problems measured using the Children's Sleep Habits Questionnaire (CSHQ), and maintenance of effects on nocturnal sleep duration at 3 and 6 months. The trial will also describe implementation outcomes, including sleep-diary engagement, intervention adherence, technical problems, and researcher revision of large language model-generated individualized recommendations.

The newly developed Parent Knowledge, Attitudes, Self-Efficacy, and Beliefs about Preschool Sleep Scale (PKASB-PS) will be analyzed as a trial outcome only if its final item set, scoring rules, and version are frozen before outcome data collection. Otherwise, analyses involving this scale will be treated as exploratory psychometric analyses rather than confirmatory intervention-effect analyses.

2. Trial Design

The Sleep Island trial is a single-center, two-arm, parallel-group randomized waitlist-controlled trial. A total of 122 parent-child dyads will be randomized in a 1:1 ratio to either routine sleep-health education plus access to the Sleep Island WeChat mini-program for 4 weeks or a waitlist control condition receiving routine sleep-health education without Sleep Island access during the controlled intervention period.

The original sample-size calculation estimated that 102 evaluable families were required, with 51 participants per group, to achieve 80% statistical power at a two-sided significance level of 0.05 for the primary sleep-duration outcome. Allowing for approximately 20% attrition, the target enrollment is 122 parent-child dyads.

Assessments will occur at baseline, immediately after the 4-week intervention, 3 months after intervention completion, and 6 months after intervention completion. Because participants will know whether they have access to Sleep Island, participant blinding is not feasible.

3. Outcomes

3.1 Primary Outcome

The primary outcome is mean actigraphy-measured nocturnal sleep duration, expressed in minutes per night. Children will undergo wrist actigraphy monitoring for 7 consecutive days at each

assessment. Mean nocturnal sleep duration will be calculated across valid recording nights. The primary endpoint is the between-group difference in change in mean nocturnal sleep duration from baseline to 4 weeks.

3.2 Secondary Outcomes

Actigraphy-measured nocturnal sleep duration at 3 and 6 months will use the same derivation as the primary outcome and will evaluate maintenance of the intervention effect.

Actigraphy-measured sleep efficiency will be calculated as the mean nocturnal sleep efficiency across valid recording nights and expressed as a percentage. Higher values indicate better sleep efficiency.

Actigraphy-measured night wakings will be calculated as the mean number of nocturnal awakenings across valid recording nights and expressed as awakenings per night. Higher values indicate more frequent night waking.

Parent-reported child sleep problems will be assessed using the validated Chinese version of the Children's Sleep Habits Questionnaire. The CSHQ total sleep disturbance score will be used as the prespecified questionnaire outcome, with higher scores indicating greater sleep problems.

3.3 Other Prespecified Outcomes

Implementation outcomes will be evaluated within the intervention group. These outcomes are the number of days with completed electronic sleep diary entries during the 4-week intervention, the proportion of participants meeting the prespecified intervention-adherence criterion, the proportion of participants experiencing at least one technical problem while using Sleep Island, and the proportion of large language model-generated individualized recommendations requiring researcher revision before release.

3.4 Exploratory PKASB-PS Outcomes

If the final PKASB-PS is established before trial outcome data collection, the Knowledge, Attitudes, Self-Efficacy, and Beliefs domains will be analyzed separately. A combined total score will not be used unless psychometric evaluation supports interpretation of a higher-order total construct.

4. Statistical Principles

4.1 Analysis Population

The primary and secondary efficacy analyses will follow the intention-to-treat principle. All randomized participants will be analyzed according to their randomized treatment assignment regardless of adherence to the allocated intervention or completeness of follow-up. Implementation outcomes that are applicable only to Sleep Island use will be summarized among participants randomized to the intervention group who received access to the mini-program.

4.2 Significance Level and Multiplicity

There is one primary outcome and one primary endpoint. The primary hypothesis will be tested using a two-sided significance level of 0.05. Secondary outcomes will be considered supportive. No formal adjustment for multiple comparisons will be applied to secondary outcomes; effect estimates, 95% confidence intervals, and exact P values will be reported and interpreted in the context of multiplicity. Exploratory PKASB-PS and implementation analyses will not be used to establish the primary efficacy conclusion.

4.3 Missing Data

The amount and pattern of missing data will be summarized by randomized group and assessment time. For longitudinal continuous outcomes, the primary mixed-effects models will use all available outcome observations under a missing-at-random assumption without restricting the analysis to participants with complete follow-up. No single-value imputation, including last observation carried forward, will be used for the primary analysis. If missing primary-outcome data are substantial or differ meaningfully between randomized groups, sensitivity analyses using multiple imputation will be conducted.

4.4 Baseline Characteristics

Baseline demographic and clinical characteristics will be summarized separately by randomized group. Continuous variables will be reported using means and standard deviations or medians and interquartile ranges, as appropriate. Categorical variables will be reported using frequencies and percentages. Statistical significance tests will not be performed to compare randomized groups at baseline because the purpose of the baseline table is descriptive rather than inferential.

5. Primary Outcome Analysis

5.1 Descriptive Analysis

Mean nocturnal sleep duration will be summarized by randomized group and assessment time using the number of participants with available data, the mean, standard deviation, and, where informative, the standard error and 95% confidence interval.

5.2 Primary Inferential Model

The primary inferential analysis will use a linear mixed model for repeated measurements. The model will include fixed effects for treatment group, assessment time treated as a categorical variable, and the treatment group-by-time interaction, together with a participant-specific random intercept to account for within-participant correlation. The assessment times will be baseline, 4 weeks, 3 months, and 6 months.

The primary treatment contrast will estimate the difference between the intervention and control groups in change in mean nocturnal sleep duration from baseline to 4 weeks. The treatment effect will be reported as the estimated mean difference in change, 95% confidence interval, and two-sided P value. Group-by-time contrasts at 3 and 6 months will be treated as secondary maintenance analyses.

5.3 Model Assumptions

Residual distributions and model fit will be examined using appropriate diagnostic procedures. If substantial departures from model assumptions are identified, suitable transformations or robust alternative models may be considered. Any analysis that departs from the prespecified primary model will be identified as a sensitivity analysis rather than replacing the primary analysis without documentation.

6. Secondary Outcome Analyses

6.1 Nocturnal Sleep Duration at 3 and 6 Months

The longitudinal linear mixed model specified for the primary outcome will also provide estimates of intervention effects at the 3- and 6-month follow-up assessments. Prespecified contrasts will estimate

the between-group difference in change from baseline at each of these follow-up assessments and will be interpreted as maintenance analyses.

6.2 Actigraphy-Measured Sleep Efficiency

Sleep efficiency will be analyzed using a linear mixed model containing treatment group, categorical assessment time, treatment group-by-time interaction, and a participant-specific random intercept. Between-group differences in change from baseline will be estimated at 4 weeks, 3 months, and 6 months. Results will be presented as mean differences in percentage points with 95% confidence intervals.

6.3 Actigraphy-Measured Night Wakings

The mean number of night wakings per night will first be summarized descriptively. Because this variable represents count data, its distribution will be evaluated before inferential analysis. If the distribution is adequately approximated by a normal continuous distribution, a linear mixed model will be used. If clear count-type skewness or overdispersion is present, a generalized linear mixed model with an appropriate count distribution, such as the negative binomial distribution, will be fitted. The choice of distribution will be based on the observed form and dispersion of the data and not on treatment-effect significance.

6.4 Children's Sleep Habits Questionnaire

The CSHQ total score will be analyzed using a linear mixed model containing treatment group, assessment time, treatment group-by-time interaction, and a participant-specific random intercept. Between-group differences in change from baseline will be estimated at 4 weeks, 3 months, and 6 months. Higher CSHQ scores indicate greater child sleep problems; therefore, reductions from baseline indicate improvement.

6.5 PKASB-PS

If the final PKASB-PS is frozen before trial outcome data collection, each finalized PKASB-PS subscale will be treated as a separate outcome. Descriptive statistics will be reported by randomized group and assessment time. Longitudinal change will be evaluated using linear mixed models containing treatment group, time, treatment group-by-time interaction, and a participant-specific random intercept. Because the PKASB-PS is newly developed, these analyses will be considered secondary or exploratory and will not alter the primary efficacy conclusion.

7. Other Prespecified Outcome Analyses

7.1 Sleep Diary Engagement

The number of calendar days with completed electronic sleep diary entries during the 4-week intervention will be summarized among intervention participants using the mean and standard deviation or the median and interquartile range, as appropriate.

7.2 Intervention Adherence

The proportion of intervention participants meeting the prespecified Sleep Island adherence criterion will be reported as a percentage with a 95% confidence interval. The exact adherence criterion will be finalized before database lock and before examination of treatment-effect results.

7.3 Technical Problems

The proportion of intervention participants experiencing at least one technical problem during the intervention will be calculated as the number of participants experiencing one or more technical problems divided by the number of intervention participants who access Sleep Island, multiplied by 100. The percentage and corresponding 95% confidence interval will be reported.

7.4 Researcher Revision of Large Language Model-Generated Recommendations

The proportion of individualized recommendations requiring researcher revision before release will be calculated as the number of recommendations requiring revision divided by the total number of large language model-generated recommendations reviewed, multiplied by 100. The percentage and corresponding 95% confidence interval will be reported. Reasons for revision will additionally be summarized descriptively using prespecified categories that may include factual inaccuracy, inconsistency with the intervention protocol, unclear or inappropriate wording, insufficient personalization, and other reasons.

8. Sensitivity and Subgroup Analyses

8.1 Missing-Data Sensitivity Analysis

If clinically meaningful missingness is observed for the primary endpoint, multiple imputation will be performed under a missing-at-random assumption. The primary treatment contrast from the imputed datasets will be compared with the primary mixed-model estimate. Where warranted by the observed pattern of missingness, additional sensitivity analyses may evaluate departures from the missing-at-random assumption.

8.2 Complete-Case Sensitivity Analysis

A complete-case analysis restricted to participants with available primary-outcome data at baseline and 4 weeks will be performed as a sensitivity analysis. This analysis will be supplementary and will not replace the intention-to-treat analysis.

8.3 Subgroup Analyses

No confirmatory subgroup analyses are prespecified. Any subgroup analyses conducted subsequently will be clearly identified as exploratory.

9. Handling of the Baseline Characteristics Table

No significance tests will be applied to baseline characteristics. The baseline table will describe the randomized groups and the study population. Any notable numerical imbalance will be acknowledged descriptively and, if clinically important and prespecified before treatment-effect analysis, may be considered in a sensitivity analysis. Baseline imbalance alone will not be used as a data-driven basis for changing the primary model.

10. Reporting of Results

For continuous outcomes, results will generally include observed means and standard deviations, estimated marginal means where applicable, estimated mean differences, 95% confidence intervals, and two-sided P values. For binary implementation outcomes, results will include the numerator, denominator, percentage, and 95% confidence interval. For count outcomes analyzed using generalized linear mixed models, an appropriate effect measure such as a rate ratio will be reported with a 95% confidence interval.

Exact P values will be reported to three decimal places, except values below 0.001, which will be reported as $P < 0.001$. Primary, secondary, and exploratory analyses will be clearly distinguished in all reports and publications.

11. Software

Statistical analyses will be conducted using R (version 4.5.2 or higher). Linear and generalized linear mixed-effects models will be fitted using appropriate validated R packages, such as lme4 and lmerTest, and estimated marginal means and prespecified contrasts may be obtained using emmeans. The R version and final package versions used for the analysis will be documented in the final statistical report.