

Study Name:

Effect of Kangaroo Mother Care during the first 72 hours of life on early growth and breastfeeding in normal birth weight newborns: Protocol for a Randomised Controlled Trial

Statistical Analysis Plan

General Principles: All statistical analyses will be performed using R software (R Foundation for Statistical Computing, <https://www.r-project.org/>), Python (Python Software Foundation, <https://www.python.org/>) or STATA (StataCorp, College Station, Texas). The primary analysis will be conducted on an Intention-to-Treat (ITT) basis, including all randomised mother-infant dyads in the groups to which they were originally assigned, regardless of adherence to the intervention or protocol deviations. Statistical significance will be defined as a two-sided p-value ≤ 0.05 .

Analysis Populations:

- **Intention-to-Treat (ITT) Population:** All randomised participants. This will serve as the primary analysis population to estimate the efficacy of the intervention under real-world conditions.
- **Per-Protocol (PP) Population:** A secondary analysis will be performed restricting the intervention group to "compliers", i.e., participants with high fidelity to the protocol. Compliance is defined as achieving ≥ 8 hours of daily skin-to-skin contact with exclusive breastfeeding performed in the KMC position in more than 80% of episodes during the hospital stay (first 48 hours).

Descriptive Statistics: Baseline demographic and clinical characteristics will be summarised by treatment group. Continuous variables will be presented as means with standard deviations (SD) or medians with interquartile ranges (IQR) based on the normality of distribution (assessed via the Shapiro-Wilk test). Categorical variables will be presented as frequencies and percentages. Baseline balance will be assessed clinically, and any significant imbalances will be considered for inclusion as covariates in adjusted models.

Analysis of Primary Outcomes

- **Percentage Weight Loss & Weight Gain Velocity (Continuous):** The mean percentage weight loss at 48 hours and weight gain velocity during the newborn period will be compared between the intervention and control groups using Linear Mixed-Effects Models (LMM). The model will include the treatment group as a fixed effect and the study site as a random effect to account for stratification. To increase the precision of the estimated treatment effect, the LMMs will be adjusted for prespecified covariates that are strong biological predictors of the outcome: maternal parity (primiparous vs. multiparous), infant sex, and baseline birthweight. If residuals are non-normally distributed, 95% Confidence Intervals (CI) will be derived using Bias-Corrected and Accelerated Bootstrapping (1,000 replications).
- **Quality of Breastfeeding (Binary & Ordinal):**
 - Primary evaluation (Binary) – As per the study hypothesis, BBAT scores will be dichotomised into "Good" (≥ 7) vs. "Moderate-to-Poor" (< 7). The proportion of dyads with moderate-to-poor scores will be compared using Mixed-Effects Logistic Regression, with study site included as a random intercept and adjustment for parity, infant sex and baseline birthweight. Adjusted odds Ratios (aOR) and 95% CIs will be reported.
 - Sensitivity analysis (Ordinal scale) – To leverage the full information of the 0–8 scale, the raw BBAT score will be analysed using Mixed-Effects Ordinal Logistic Regression (Proportional Odds Model). This analysis will assess the intervention's effect on achieving higher quality scores across the entire scale.

Analysis of Secondary Outcomes

- **Binary Outcomes (e.g., Exclusive Breastfeeding, Maternal Depression, PSBI):** Analysed using mixed-effects logistic regression or Chi-square tests where modelling is not feasible due to low event rates.
- **Continuous Outcomes (e.g., Bonding Scores, Self-Efficacy):** Analysed using linear mixed-effects models or independent t-tests (with robust standard errors) as appropriate.

Handling of Missing Data: Missingness patterns will be examined. If missing data for primary outcomes exceeds 5%, Multiple Imputation by Chained Equations (MICE) will be employed, assuming data are Missing At Random (MAR). Sensitivity analyses will be conducted using Complete Case Analysis to check the robustness of the results.