Statistical Analysis Plan (SAP)

Validation Study of a Digital Measurement Device for Central Hand Representation in Children with Neonatal Brachial Plexus Palsy (NBPP)

NCT Number

NCT06950879

Researchers:
L. Vanhoorebeeck, N. Vermassen, E. Staelens, Prof. Dr. R. Van Der Looven
Date: November 19, 2024
Version: 1.2

1. Objectives of the Analysis

The primary objective of the statistical analysis is to evaluate the reliability of the HandUZ device in measuring implicit finger length (FL) in children diagnosed with Neonatal Brachial Plexus Palsy (NBPP). Specifically, the study aims to assess:

Intrarater reliability – the consistency of repeated measurements taken by the same examiner across different time points.

Interrater reliability – the agreement between two independent raters conducting the same measurement protocol.

A secondary objective is to examine trial-related improvement, defined as progressive changes in estimated FL over successive trials, which could impact the reliability of measurements.

2. Data Preprocessing

Prior to statistical analysis, data preprocessing was performed to ensure data integrity and prepare the dataset for analysis. The following steps were taken:

Visual inspection of each hand map to identify trials that could not be reconstructed due to technical issues or incorrect landmark placement.

Exclusion of trials (7 out of 204, i.e., 3.4%) with unusable data, while retaining the second valid trial if only one of a pair was excluded.

Identification and treatment of missing values in accordance with predefined handling procedures.

Verification of the correct assignment of trials to the respective raters and measurement rounds to preserve the study protocol structure.

3. Descriptive Statistics

Descriptive statistics were computed for:

- Participant demographics (age, sex, lesion characteristics, handedness).
- Actual finger length (in cm) per hand.
- Estimated finger lengths derived from the HandUZ measurements.

- Percentage overestimation (OE) of FL using the formula: OE (%) = 100 × [(Estimated FL – Actual FL) / Actual FL].
- Standard deviations, minimum and maximum values, and ranges were calculated to assess interindividual variability.

4. Assumption Checks

Before performing parametric tests, statistical assumptions were evaluated:

- Normality of the estimated FL data at each measurement point was tested using the Shapiro-Wilk test.
- Visual inspection of Q-Q plots was used to confirm approximate normal distribution.
- Sphericity was assessed using Mauchly's test in the repeated-measures ANOVA. Where this assumption was violated, Greenhouse-Geisser corrections were applied to adjust the degrees of freedom and maintain valid significance testing.

5. Trial-related improvement Analysis

To assess whether children's performance improved or changed due to increased task familiarity, a learning effect analysis was performed:

A repeated-measures ANOVA was conducted separately for each hand using the four trials administered by Rater 1 (M1, M2, M5, M6). The independent variable was 'time', with four within-subject levels representing the measurement rounds.
When Mauchly's test indicated a violation of sphericity, the Greenhouse-Geisser correction was applied. Polynomial contrast analysis was used to test for linear and higher-order trends.
Visual inspection of estimated marginal means was used to identify systematic trends and interpret borderline findings.

6. Reliability Analyses

6.1. Intrarater Reliability

Intrarater reliability was assessed using Intraclass Correlation Coefficients (ICCs), based on a two-way mixed-effects model, single measures, with absolute agreement (ICC(3,1)).

Two independent sets of measurements by Rater 1 were averaged:
- Average 1 = Mean of M1 and M5
- Average 2 = Mean of M2 and M6

ICCs were calculated separately for the left and right hands.
Confidence intervals (95%) and F-tests were reported to assess the significance of the

ICCs. Interpretation followed Koo and Li's guidelines:

- ICC < 0.5 = poor
- 0.5 ≤ ICC < 0.75 = moderate
- 0.75 ≤ ICC < 0.90 = good
- ICC ≥ 0.90 = excellent

6.2. Interrater Reliability

Interrater reliability was evaluated using the same ICC model (ICC(3,1)). Average 3 (M2 and M5 by Rater 1) was compared with Average 4 (M3 and M4 by Rater 2). Separate ICCs were computed for each hand. Confidence intervals and significance levels were reported to quantify rater agreement.

7. Handling of Multiple Testing

This validation study focused on a limited number of predefined analyses, thus adjustments for multiple comparisons were not applied. Nevertheless, the potential for Type I error was minimized by restricting analyses to those aligned with the stated study objectives. Effect sizes (partial η^2) were reported alongside p-values for ANOVA results to aid interpretation and assess the magnitude of learning effects.

8. Software

All statistical analyses were performed using IBM SPSS Statistics (Version 29.0.2.0, IBM Corp., Armonk, NY, USA). Data preprocessing and visualization of hand maps were conducted using custom-built software in Python (Spider 3.0).

9. Significance Thresholds

All statistical tests were two-tailed. The threshold for statistical significance was set at:

- $\alpha = 0.05$
- Confidence intervals were set at 95%.
- For repeated-measures ANOVA, partial eta squared (η^2) was reported as the measure of effect size.