STATISTICAL ANALYSIS PLAN

rTMS Combined with Motor Training for the Treatment of Upper Limb Motor Dysfunction in Stroke Patients

National Clinical Trial (NCT) Identified Number: NCT06752499

May 7, 2025

TABLE OF CONTENTS

1. Synopsis of the Study

This Phase I clinical trial aims to evaluate whether repetitive transcranial magnetic stimulation (rTMS) applied concurrently with motor training elicits greater motor improvement, as measured by the Fugl-Meyer Assessment of the Upper Extremity (FMA-UE), compared to non-concurrent rTMS or sham concurrent rTMS. Assessments will occur immediately after the 2-week intervention and at 1- and 3-month follow-ups. The central hypothesis is that concurrent rTMS with motor training enhances plasticity in the engaged motor circuits. However, it remains unknown whether precisely timed rTMS with individualized targeting is sufficient to produce a superior therapeutic response while maintaining a comparable safety and tolerability profile.

OBJECTIVES	ENDPOINTS	JUSTIFICATION FOR ENDPOINTS
Primary		
To determine whether there is an initial overall treatment effect among 3 groups: ➢ Concurrent rTMS (online) ➢ Non-concurrent rTMS (offline) ➢ Sham rTMS (sham) Efficacy is measured at day 105 after the initial (day 0) of the intervention.	FM-UE scale	FM-UE scale is the primary outcome and a measure of functional motor impairment
	WMFT time score	WMFT is a secondary outcome and a measure of functional motor activity
	ARAT score	ARAT is a secondary outcome and a measure of functional motor activity
	MBI score	MBI is a secondary outcome and a measure of the quality of daily life
	PSQI score	PSQI is a secondary outcome and a measure of the quality of sleep
Secondary		
To confirm that the proposed intervention is safe, tolerable, and feasible to administer in a multi-site trial setting	Rate of adverse events	Safety needs to be monitored in a phase I+ II study
	Feasibility	How many subjects withdrew from the study as a result of adverse effects?
	Treatment tolerability	80% of treatment completion rate across sites without unexplained/unresolved variability is a measure of feasibility.
Tertiary/Exploratory		
To examine whether CST-FA (structural assessment of integrity of descending motor tract) is correlated with changes in FM-UE scale, and evaluate the	sMRI [T1 and T2, Diffusion tensor imaging (DTI))	Structural aspects of motor tracts change after treatment. This change can be captured by comparing CST-FA measures before and after treatment

utility of these measures as biomarkers for subject selection criteria in the future confirmatory Phase II and III study.		
To examine changes in functional connectivity between different regions related to motor circuits.	fMRI	fMRI provides a functional measure of changes in functional connectivity within motor-related circuits and serves as an indicator that correlates with improvements in behavioral scales.
The objective is to identify EEG-based biomarkers and determine whether they exhibit a correlation with the behavioral scales.	EEG	Resting-state EEG can be used to test for changes in band power. For TMS evoked potentials, changes in the N100 component can be examined. In addition, other TEP components should also be analyzed to determine if they exhibit similar alterations.
To investigate whether there is an improvement in the upper limb motor trajectory of patients.	Videos	The patient's kinematic parameters can be measured by analyzing the trajectory graph generated during motor training by computer vision. Both the task duration and the variance of the movement trajectory can be calculated to evaluate concurrent rTMS efficacy.
MEPs (functional assessment of integrity of descending motor tract)	MEPs	MEPs is a functional measure integrity of descending motor tract using TMS that has shown promise in predicting motor outcomes.
To investigate whether short-term or long-term improvements occur in the descending motor pathways.	EMG	This metric is designed to quantify neuromuscular adaptations by integrating two key components: the change in electromyographic (EMG) amplitude and the similarity of muscle synergy patterns. By concurrently tracking these features, it provides a comprehensive assessment of muscular control strategies, capturing both short-term neuromuscular activation changes and long-term neuroplastic reorganization throughout the intervention and follow-up periods.

2. Study Design

The trial will test an overall hypothesis that a combination of rTMS applied concurrent with motor performance, will lead to a greater sustained motor improvement on day 105 (± 5 days) after the start of the intervention [day 0 (± 2 days)] as compared to non-concurrent or sham stimulation. Sustained benefits will be assessed at day 15 (± 2 days) and day 45 (± 5 days) after the start of the intervention. This multicenter, phase I, sham-controlled 3-arm study will allocate 60 subjects in 3 arms (concurrent, non-concurrent, or sham), and treat subjects with the assigned paradigm of rTMS for about 30-50 minutes for 10 sessions over a 2-week period.

3. Sample Size and Power

We initially conducted a single-center, randomized, double-blind, controlled trial at Ruijin Hospital. Sample size estimation was performed using PASS 2021 software, with a two-sided α level of 0.025, designed to test the significance in the outcome differences between concurrent rTMS group and non-concurrent rTMS group or sham rTMS group. Based on the data from the preliminary study, we anticipated an effect size of 1.5 in Fugl-Meyer Assessment of the Upper Extremity (FMA-UE) scores between the concurrent rTMS and sham groups. Under these assumptions, a minimum of 18 participants (9 per group) was required to achieve 80% power. An equal sample size of 9 was also planned for the non-concurrent group. Considering the low dropout rate among in-patients (~10%), we aimed to recruit 10 participants per group, totaling 30 patients. However, following stratification by motor evoked potential (MEP) status, the sample size in each subgroup became insufficient. To further evaluate the generalizability of individualized concurrent rTMS targeting in MEP-negative patients, the trial will be expanded to a multicenter study. Two additional hospitals (Yangzhi Hospital and Changhai Hospital) will be included, with a focus on prioritizing enrollment of MEP-negative patients for individualized stimulation. To compare the effects of individualized versus non-individualized stimulation across all rTMS conditions, each group included both subtypes. Ultimately, a total of 60 subjects will be enrolled across the three participating centers. The planned recruitment includes 30 subjects from Ruijin Hospital, 15 from Yangzhi Rehabilitation Hospital, and 15 from Changhai Hospital.

4. Definition of Analysis Samples

4.1 Modified-full-analysis set (mFAS) Sample

All analyses will be conducted under the mFAS principle. The mFAS included all enrolled participants who initiated the intervention and had baseline and at least one post-baseline assessment.

4.2 Per Protocol (PP) Sample

To estimate the efficacy of concurrent rTMS under optimal conditions, a per-protocol analysis will be performed. This analysis will be restricted to a cohort of high-adherence participants, defined as those who completed the full course of ten rTMS sessions, attended all post-treatment FM-UE assessments, and did not experience confounding events such as recurrent stroke or eligibility violations. The intent is to isolate the maximum potential treatment effect in the absence of major protocol deviations.

5. General Statistical Considerations

5.1 Allocation Procedures

This pilot study adopts a non-randomized allocated design. Participants will be allocated using a convenience allocation method, where assignment to either the intervention or control group is based on the order of enrollment. Recruitment priority will be given to fulfilling the sample size requirements of the intervention group.

5.2 Missing Data

The primary ITT analysis included all subjects who were assigned to the study. Therefore, for subjects who did not complete the FM-UE scale assessment, the change in their FM-UE scores from the time of allocation—defined as the primary outcome—needed to be considered. Each subject will be assessed using the FM-UE scale at the time of allocation, as well as on days15, 45, and 105. Each FM-UE assessment will be scored by their respective research site. The measurements used in the model are defined as the scores provided by the research sites. The primary analysis will employ a linear mixed-effects repeated measures model to appropriately handle subjects with missing data. This method has been demonstrated to be more effective than the "Last Observation Carried Forward" (LOCF) method or "Multiple Imputation" (MI)[1]. For analyses requiring the inclusion of covariates, similar techniques will be considered, in line with current best practices for handling missing data[2-4]. As a sensitivity analysis, if more than 10% of subjects miss all follow-up visits, the multiple imputation method will be used instead. The multiple imputation model will include necessary covariates but will exclude the treatment factor.

6. Analysis: Primary Aim

6.1 Overview

The goal of the primary analysis is to determine whether there is a persistent treatment effect among 3 groups (concurrent rTMS, non-concurrent rTMS, and sham rTMS) as measured at day 105 after the initiation of the intervention.

6.2 Primary Outcome Analysis: Long-Term Outcome at Day 105

The primary outcome assesses whether there is a treatment difference in the Fugl-Meyer upper extremity scale, a measure of motor impairment.

6.2.1 Endpoint

The primary endpoint is defined as the change in the site assessed Fugl-Meyer upper extremity score from the baseline assessment (before allocation) and the day 105 post-intervention visit.

6.2.2 Statistical Hypotheses

The primary null hypothesis is that there is no difference in change in FM-UE on day 105 after initiation (day 0) of the 2-week intervention among the three treatment groups.

H_0: $\mu_{concurrent,105d} = \mu_{non\text{-}concurrent,105d} = \mu_{sham,105d}$

H_A: The means are not all equal

where $\mu_{concurrent,105d}$ represents the mean change in FM-UE Scale in the concurrent rTMS group at Day 105, and $\mu_{non\text{-}concurrent,105d}$, $\mu_{sham,105d}$ represents the same value in the non-concurrent rTMS group and sham rTMS group respectively. The null hypothesis is tested at type I error rate of 5% against the alternative that at least one pair of means is not equal.

6.2.3 Analysis Method

The change in the Fugl-Meyer Assessment for the Upper Extremity (FMA-UE) score will be analyzed using a generalized linear mixed-effects repeated measures model. The dependent variable is the change from baseline at each post-intervention visit. The model will include fixed effects for treatment arm, baseline FMA-UE score, time since stroke onset, study visit, and the affected hemisphere, with the primary treatment effect estimated as the fitted value at Day 105. Model assumptions, including residual normality, will be rigorously evaluated. Should violations be detected, variable transformations will be applied in accordance with established statistical guidelines. If the primary null hypothesis is rejected, all pairwise comparisons among treatment arms will be conducted using a Tukey-type or Bonferroni correction for multiple testing.

6.3 Secondary Outcomes Analysis

Aim 1 will also evaluate secondary null hypotheses concerning the least-squares mean change from baseline to Day 105 in the Wolf Motor Function Test (WMFT) time score, Action Research Arm Test (ARAT), Modified Barthel Index (MBI), and Pittsburgh Sleep Quality Index (PSQI). A hierarchical testing procedure will be employed to control the family-wise Type I error rate. Specifically, if the primary null hypothesis (regarding the FM-UE) is rejected, the secondary hypothesis for the MBI will be tested. Should that also be rejected, subsequent tests will be conducted for the WMFT time score, ARAT,

and PSQI, in a pre-specified sequence, using a Tukey or Bonferroni-type gatekeeping procedure[5, 6].

6.3.1 Endpoints

Wolf Motor Functional Test (WMFT): The functional outcome will be defined as the change from baseline in the final time score of the more affected limb, as assessed locally. The baseline measurement is taken at the initial visit prior to allocation, with follow-up assessments occurring on Days 15, 45, and 105. The final time score is calculated as the median time to complete all performed timed tasks. Any task not fully executed within the 120-second maximum allowed time is assigned a censored value of 121 seconds. The primary endpoint for this outcome is the change at Day 105. Given the expected right-skewed distribution of time scores based on previous studies, a logarithmic transformation will be applied prior to analysis. The results will be back-transformed into seconds for final interpretation and reporting.

Action Research Arm Test (ARAT): The functional outcome measure will be defined as the change in the total ARAT score from the baseline assessment (measured at the first baseline visit prior to randomization) to each of the follow-up visits (day 15, 45, and 105) in the more affected limb as scored by the local assessor. The ARAT consists of 19 items across four domains (grasp, grip, pinch, and gross movement), each scored on a 4-point ordinal scale (0 = no movement, 3 = normal performance). The total score ranges from 0 to 57, with higher scores indicating better upper limb function. For this outcome, the primary outcome of interest is the change at day 15. The distribution of ARAT scores is typically skewed toward lower values in more severely affected patients; if non-normality is present, appropriate non-parametric or transformed analyses will be considered.

Modified Barthel Index (MBI): The functional outcome measure will be defined as the change in the MBI total score from the baseline assessment to each of the follow-up visits (day 15, 45, and 105), as scored by the local assessor. The MBI assesses performance in 10 domains of activities of daily living (feeding, bathing, grooming, dressing, bowel/bladder control, toilet use, transfers, mobility, and stair climbing). Each item is scored in increments of independence, with the total score ranging from 0 to 100, where higher scores indicate greater independence. For this outcome, the primary outcome of interest is the change at day 15. As the MBI often exhibits ceiling effects in patients with mild disability, sensitivity analyses may be conducted to assess robustness of results in subgroups stratified by baseline severity.

Pittsburgh Sleep Quality Index (PSQI): The outcome measure will be defined as the change in the PSQI global score from the baseline assessment to each of the follow-up visits (day 15, 45, and 105), as assessed by the local evaluator.

The PSQI is a self-reported questionnaire consisting of 19 items grouped into 7 components (subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction). Each component is scored from 0 (no difficulty) to 3 (severe difficulty), yielding a global score ranging from 0 to 21, with higher scores indicating poorer sleep quality. For this outcome, the primary outcome of interest is the change at day 15. Because PSQI scores are ordinal and may demonstrate skewness, analyses will use methods appropriate for non-normally distributed outcomes; results will be reported on the original scale for interpretability.

6.3.2 Analysis Method

The change from baseline in the outcome will be analyzed using a linear mixed-effects repeated measures model. The model will include fixed effects for visit (categorical: Days 15, 45, 105), treatment arm, baseline score of the outcome, time from stroke onset, and study site. To account for within-subject correlation across visits, an autoregressive order 1 (AR1) covariance structure will be applied. Additionally, an unstructured covariance structure will be used to model correlation among subjects within the same site. Model assumptions will be evaluated, and variables will be transformed, if necessary, consistent with established statistical practices. If the primary null hypothesis is rejected, all pairwise treatment comparisons will be conducted using a Dunnett-type correction for multiple testing[7, 8].

6.4 Secondary Analyses: Subgroup Analyses by Sex

No prespecified subgroup analyses stratified by sex were planned for this trial. Differential effects of rTMS by sex are not anticipated, and therefore the trial is not powered specifically for these subgroup analyses, but these subgroup analyses will allow any unexpected large variations to be identified and regardless of outcome will assist with planning a Phase II and III trial which will account for differences if identified here.

6.5 Secondary Analyses: Short- and Mid-Term Outcomes (Day 15 and Day 45)

Key secondary endpoints include changes from baseline to Day 15 and Day 45. A hierarchical testing strategy will be applied: the primary hypothesis at Day 105 will be tested first at a two-sided $\alpha=0.05$; only if significant, subsequent tests at Day 15 and Day 45 will be performed in a prespecified order to control the family-wise error rate.

7. Analysis: Secondary Aims

7.1 Overview

This analysis aims to evaluate the intervention's safety, tolerability, and multi-

site feasibility. Safety will be assessed by comparing the incidence of clinically important adverse events between groups. Tolerability will be measured by the rate of discontinuations attributed to discomfort or adverse events. Feasibility will be demonstrated if $\geqslant$80% of enrolled subjects complete the treatment protocol without evidence of unexplained or unresolved inter-site variability.

7.2 Safety

7.2.1 Analysis Method

Adverse events (AEs), including serious adverse events (SAEs), will be collected throughout the intervention period. After the intervention period, only SAEs and adverse events assessed as possibly or definitely related to the study will be collected. Periodic safety reports summarizing all adverse events, including any SAEs, will be provided to the Medical Safety Monitor (MSM). All reportable clinical safety endpoints and SAEs will be summarized by AE code (as specified on the AE case report form), including the frequency of events, number of subjects affected, severity, and relationship to study treatment. Clinically important adverse safety events are defined as any of the following occurring during the intervention period (up through the reporting window for the Day 15 visit) and assessed as possibly or definitely related to the intervention at any time:

- Severe headache
- Epileptiform discharges on EEG
- Clinical seizure

The proportion of subjects experiencing each of these events will be presented in the closed report, stratified by treatment arm, along with two-sided 95% confidence intervals and unadjusted risk differences for each pairwise comparison between stimulation arms. In the Phase I dose-escalation study, none of the subjects experienced any of these clinically important adverse events.

7.3 Tolerability

7.3.1 Analysis Method

Treatment Tolerability: Treatment tolerability[9-11] will be primarily assessed by the proportion of patients who discontinued the study due to adverse events or discomfort. Group differences in withdrawal rates will be compared across the three study arms using a chi-square test (or Fisher's exact test if expected counts are small). Tolerability will thus be defined as the ability of participants to complete the assigned treatment without premature discontinuation attributable to side effects. Descriptive summaries (counts and percentages) of reasons for withdrawal will be presented for transparency. The study will be deemed tolerable if this proportion does not exceed a pre-specified threshold of 15%.

7.4 Feasibility

7.4.1 Assessment of Feasibility

Treatment Adherence: the primary feasibility measure[9-11] for treatment adherence will be defined as at least 80% of enrolled participants completing at least 8 out of the 10 scheduled rTMS treatment sessions. Completion of the treatment protocol is defined as attendance at the session, regardless of the exact number of motor training tasks completed within that session. Operational feasibility, including unexplained or unresolved variability across sites, will be continuously monitored as part of the trial management plan. Sites with recurrent deviations or excessive variability may be retrained or removed. No formal statistical testing will be conducted for this measure.

8. Exploratory Analysis

8.1 Overview

The list below outlines the planned analysis methods for the pre-specified exploratory outcomes. The usual verification of variable and model assumptions (e.g., normality and homoscedasticity) and goodness of fit assessments will accompany each analysis. These analysis results will be treated as supportive evidence (or lack thereof) of the treatment effect rather than conclusive evidence. The significance of each test is determined at the two-sided alpha level of 0.05 of type I error.

8.2 Exploratory Aim: Responder Analysis

In addition to the primary analysis of the mean change from baseline in the FM-UE, it is of clinical importance to report the proportion of patients in each arm who achieve a minimal level of improvement. An improvement from baseline to the 105-day visit will be defined as a change on the FM-UE of ≥ 5.25 [12-14], which is considered to be a meaningful clinically important difference (MCID). Furthermore, current exploratory research suggests that key subscales or elements of the FM-UE may be more impactful to recovery. Responder analysis using the relevant portions or weighting may also be considered using the same approach if applicable.

8.3 Exploratory Aim 1

This study aims to investigate the correlation of a comprehensive set of biomarkers with changes in the Fugl-Meyer Assessment of the Upper Extremity (FMA-UE) score in stroke patients. The biomarkers of interest include kinematic parameters, electroencephalographic (EEG) band power, TMS-EEG evoked potentials, a key objective is to evaluate the utility of these measures as potential biomarker-based selection criteria for participant stratification in future phase II or III confirmatory clinical trials.

8.3.1 Outcome Measures

Kinematic parameters were employed to quantitatively assess the movement trajectory of the paretic upper limb in stroke patients, with a focus on metrics such as movement variance and task time. A computer vision-based analysis was conducted using the MM Pose open-source pose estimation toolkit. Two-dimensional joint trajectories were extracted from video recordings of task performance. From these trajectories, key kinematic parameters—including the range of motion and measures of movement smoothness (e.g., jerk metric)—were derived to provide an objective comparison of motor performance across study groups.

Electroencephalographic (EEG) band power will be analyzed to investigate cortical oscillatory activity in stroke patients. Resting-state EEG signals will be acquired and processed to extract power spectral densities within standard frequency bands: delta (δ, 1-4 Hz), theta (θ, 4-8 Hz), alpha (α, 8-13 Hz), beta (β, 13-30 Hz), and gamma (γ, >30 Hz). The relative power of each band will be calculated to assess the balance of neural oscillations. This analysis will aim to identify oscillatory biomarkers, such as abnormal alpha/beta power reductions in the sensorimotor cortex, which will be statistically correlated with motor impairment and recovery potential.

TMS-EEG evoked potentials will be acquired to directly probe cortical reactivity and effective connectivity in stroke patients. Single-pulse TMS will be applied to the primary motor cortex (M1), while concurrent EEG will record the immediate cortical responses. Key components of the TMS-evoked potential (TEP), such as the early N15/P30 and late N100/P180 complex, will be analyzed for their amplitude and latency. These metrics will be examined for their correlation with motor function and their potential to serve as biomarkers of cortical integrity and plasticity following intervention.

8.3.2 Analysis

Kinematic Parameters: Movement kinematics will be quantified using parameters such as range of motion, movement smoothness (quantified by jerk metric), and velocity profile. The analysis will focus on the following aspects for each kinematic metric: 1. The change in the parameter from baseline to Day 15 during standardized motor tasks. 2.The between-limb difference (affected side vs. unaffected side) in movement quality at Day 15. Each kinematic outcome will be adjusted for its corresponding baseline value. Summary statistics will first be used to characterize movement patterns across the three treatment groups. Formal statistical testing will then be performed to assess between-group differences in the magnitude of kinematic changes. Furthermore, correlation analysis will examine the relationship between intervention-induced changes in kinematic parameters and changes in FM-UE scores.

Electroencephalographic (EEG) Band Power: Spectral power analysis will be conducted for standard frequency bands (delta, theta, alpha, beta, and gamma) across three groups. The analysis will focus on: The change in relative band power from baseline to Day 15 in the ipsilesional sensorimotor cortex. The interhemispheric asymmetry in specific band power (e.g., sensorimotor rhythm) at Day 15. All spectral metrics will be adjusted for baseline values. Initial exploratory analysis will characterize spectral patterns across groups, followed by formal testing for between-group differences in spectral power changes. Correlation analysis will then investigate whether changes in specific band power are associated with improvements in FM-UE scores.

TMS-EEG Evoked Potentials (TEPs): TEP components (including N15, P30, N45, and N100) will be analyzed for amplitude and latency characteristics across three groups. The analysis will examine: 1. The change in TEP component features from baseline to Day 15 following stimulation of the ipsilesional M1. 2. The change in alpha oscillations at Day 15 following stimulation of the ipsilesional M1. All TEP measures will be adjusted for baseline values. Summary statistics will first describe TEP characteristics across groups, followed by statistical testing for between-group differences in TEP modulation. Finally, correlation analysis will explore the relationship between changes in TEP features and changes in FM-UE scores.

8.4 Exploratory Aim 2

This study aims to investigate whether intervention-induced changes in motor function are associated with functional or structural alterations in the corticospinal tract and functional reorganization of the motor network. These alterations will be assessed using a multimodal approach, including motor evoked potentials (MEPs), electromyography (EMG), and resting-state functional connectivity measured by fMRI. A secondary objective is to evaluate the utility of these neurophysiological and imaging metrics as potential biomarkers for enriching patient selection in future Phase II/III clinical trials.

8.4.1 Outcome Measures

Presence or Absence of Motor Evoked Potentials (MEPs): This measure serves as an indicator of the integrity of the corticospinal tract following stroke. MEP status will be defined as a binary variable (present/absent) and assessed exclusively on the more affected side. MEPs will be considered present (positive) if the resting motor threshold (RMT) on the affected side is ≤100% of the maximum stimulator output (MSO); otherwise, they will be classified as absent (negative). Should statistical analysis reveal that no more than 10% of subjects are categorized as MEP absent, an alternative metric—the RMT asymmetry index—will be adopted. This index is calculated as (|affected side RMT| - |unaffected side RMT|) / (|affected side RMT|＋|unaffected side RMT|)

and will be analyzed as a continuous variable.

Electromyography (EMG) Amplitude: The amplitude of the EMG signal on the more affected side reflects the extent of corticospinal tract and motor neuron activation during both motor tasks alone and motor tasks synchronized with TMS. During each formal treatment session, patients will wear an 8-channel EMG system to record activity from eight specified muscles. The primary EMG outcome is defined as the change in MEP amplitude on the more affected side from baseline to the end of the treatment period.

Corticospinal Tract (CST) – Fractional Anisotropy (CST-FA): The CST-FA value, which ranges from 0 to 1, is used to assess the microstructural integrity of the tract, with values closer to 1 indicating healthy, well-aligned fibers. This metric will be recorded for both the ipsilesional and contralesional hemispheres at baseline and on Day 15.

Functional Connectivity Measured by fMRI: This metric evaluates changes in functional connectivity between distinct brain regions within the motor circuit. It aims to determine whether TMS-synchronized motor training enhances functional integration among motor network nodes, thereby helping to elucidate the neural mechanisms underpinning motor recovery.

8.4.2 Analysis

Presence or Absence of Motor Evoked Potentials (MEP): The proportion of subjects with measurable Motor Evoked Potentials (MEPs) in each of the three treatment groups will first be compared using appropriate parametric or nonparametric tests.

Laterality quotient (LQ): Cortical excitability imbalance was quantified using the laterality quotient (LQ), calculated as:

$$LQ = (RMT_{contralesional} - RMT_{ipsilesional}) / (RMT_{contralesional} + RMT_{ipsilesional}) \times 100$$

A positive LQ indicates greater excitability in the ipsilesional hemisphere (lower RMT), a negative LQ indicates greater excitability in the contralesional hemisphere, and an LQ of zero reflects symmetric cortical excitability.

Electromyographic (EMG) Activity: For EMG metrics (including root mean square and integrated EMG), the analysis across three groups will focus on the following aspects: 1. The change in muscle activity from baseline to Day 15 during standardized motor tasks. 2. The task-related modulation of muscle activity (Task vs. Task+rTMS) at each treatment day. 3. The similarity of muscle synergy patterns among the three treatment groups. Each of these outcome measures will be adjusted for its corresponding baseline value. Summary statistics will first be employed to characterize the patterns of neuromuscular

changes across the three treatment groups. Formal statistical testing will then be conducted to assess between-group differences in the magnitude of these changes. Furthermore, correlation analysis will be performed to examine the relationship between the intervention-induced changes in EMG metrics and changes in the FM-UE score.

LesM1 to M1 and CST-FA: For each of these metrics (lesM1 connectivity and CST-FA), the analysis will focus on the following differences: The change in the metric from baseline to Day 15 on the lesioned side. The side-to-side difference (lesioned side vs. contralateral side) at Day 15. Each of these outcome measures will be adjusted for its corresponding baseline value (i.e., the covariate used during randomization). Similarly, summary statistics will first be employed to assess whether there are differences in the magnitude of these changes among the three treatment groups. This will be followed by further analysis to determine if the observed changes are correlated with changes in the FM-UE score.

Functional connectivity measured by fMRI: Resting-state fMRI will be used to assess changes in functional connectivity within the motor network. The analysis will focus on two primary questions: The within-group change in FC strength from baseline to post-intervention (e.g., Day 15) specifically between key motor regions, such as the ipsilesional primary motor cortex (M1) and ipsilesional primary somatosensory cortex (S1). The between-group difference in the magnitude of this FC change among the three treatment groups. Seed-based connectivity analysis will be performed, and the resulting FC values will be adjusted for baseline connectivity. The association between the intervention-induced change in FC and the change in FM-UE score will then be investigated to determine its relevance to motor recovery.

9. Go, No-Go Approach and Conclusion at the End of the Trial

Evidence from this proposed Phase I trial will support progression to Phase II if the following criteria are met: (a) the Phase I trial is deemed feasible; (b) a statistically significant difference in treatment effects is observed at the 5% significance level [primary null hypothesis], with changes in the FM-UE scale indicating treatment benefit (i.e., concurrent rTMS demonstrates a greater treatment effect than non-concurrent rTMS and sham rTMS); and (c) at least one active treatment group showing a treatment benefit is found to be safe and tolerable. Once these criteria are satisfied and rTMS is considered suitable for a Phase II trial, the selection of which group to advance for further testing will be based on a comprehensive review of all collected information. This will include secondary pairwise hypothesis testing of the primary endpoint using a Dunnet-type procedure to control the type I error rate [11,12], hypothesis testing of key secondary endpoints (such as WMFT score or time, ARAT score, MBI

score, and PSQI score) using a sequential gatekeeping approach to control the type I error rate [13,14], as well as observed safety profiles, patient tolerability, and feasibility of administration. It should be noted that these design considerations are not solely determined by statistical rules. For illustrative purposes, Section 9.1 presents potential outcomes from several hypothetical trials, along with corresponding recommended decisions.

9.1 Example Scenarios

	Feas.	Mean Treatment Effect			Primary P-Value	Safe	Tolerable	Conclusion
		concurrent	Non-concurrent	sham				
A	N							No-Go: The trial is considered infeasible because ＜80% subjects completed at least 8 out of 10 sessions.
B	Y	3.3	4.4	4.4	0.52	Y	Y	No-Go: The study will not proceed to Phase II, because the p-value is not significant. Therefore, the study results do not support the additional investigation.
C	Y	0.1	2.8	4.1	0.04	Y	Y	No-Go: Although we reject the null hypothesis of no difference, the mean treatment effect in the concurrent group do not demonstrate a benefit.
D	Y	12.1	9.7	4.3	＜0.001	Y	N (concurrent)	Go: We reject the primary null hypothesis and conclude that at least one treatment arm is different, however since there were significantly decreased tolerability in the concurrent group, a Phase II comparing non-concurrent vs. Sham would be proposed.
E	Y	12.1	9.7	4.3	＜0.001	Y	Y	Go: We will reject the primary null hypothesis and conclude that at least one treatment arm is different. Both arms are safe, tolerable, and demonstrate a signal of improvement at day 105.
E.1	Y	12.1	9.7	4.3	＜0.001	Y	Y	Requires Discussion: Although the trial meets the Go criteria (we reject the primary null hypothesis and conclude that at least one treatment arm is different), neither WMFT, ARAT, MBI or PSQI shows any indications of efficacy. Ad Hoc exploratory analysis would be required to explain this discrepancy before the PIs feel a Phase II trial would be supported by peer review.

E.2	Y	12.1	9.7	4.3	＜0.001	Y	Y	Requires Discussion: There is sufficient evidence that rTMS active arm is better than a sham. However, there is not a strong difference between concurrent and non-concurrent groups in the secondary outcome (ARAT, WMFT, MBI and PSQI). In this case, we will proceed with concurrent rTMS.
E.3	Y	12.1	9.7	4.3	＜0.001	Y	Y	Requires Discussion: The evidence for efficacy is the same as above. However, the ARAT, WMFT, MBI and PSQI has additional benefits in functional and QOL improvement. In this case, we will proceed with concurrent rTMS.

10. References

1. Siddiqui, O., *MMRM versus MI in Dealing with Missing Data—A Comparison Based on 25 NDA Data Sets.* Journal of Biopharmaceutical Statistics, 2011. **21**(3): p. 423-436.
2. Sullivan, T.R., et al., *Should multiple imputation be the method of choice for handling missing data in randomized trials?* Statistical Methods in Medical Research, 2016. **27**(9): p. 2610-2626.
3. Marshall, A., et al., *Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study.* BMC Medical Research Methodology 2010.
4. Lee, K.J., et al., *Multiple imputation for missing data in a longitudinal cohort study: a tutorial based on a detailed case study involving imputation of missing outcome data.* International Journal of Social Research Methodology, 2016. **19**(5): p. 575-591.
5. Dmitrienko, A., W.W. Offen, and P.H. Westfall, *Gatekeeping strategies for clinical trials that do not require all primary effects to be significant.* Statistics in Medicine, 2003. **22**(15): p. 2387-2400.
6. Dmitrienko, A., A.C. Tamhane, and B.L. Wiens, *General Multistage Gatekeeping Procedures.* Biometrical Journal, 2008. **50**(5): p. 667-677.
7. Hsu, J.C., *The Factor Analytic Approach to Simultaneous Inference in the General Linear Model.* Journal of Computational and Graphical Statistics, 1992.
8. Hsu, J.C. and B. NELSON, *Multiple Comparisons in the General Linear Model.* Journal of Computational and Graphical Statistics, 1998.
9. Arain, M., et al., *What is a pilot or feasibility study? A review of current practice and editorial policy.* BMC Medical Research Methodology, 2010. **10**(1).
10. Lancaster, G.A. and L. Thabane, *Guidelines for reporting non-randomised pilot and feasibility studies.* Pilot and Feasibility Studies, 2019. **5**(1).
11. Eldridge, S.M., et al., *CONSORT 2010 statement: extension to randomised pilot and feasibility trials.* Pilot and Feasibility Studies, 2016. **2**(1).
12. Stephen J. Page, George D. Fulk, and P. Boyne, *Clinically Important Differences for the Upper-Extremity Fugl-Meyer Scale in People With Minimal to Moderate Impairment Due to Chronic Stroke.* Physical Therapy, 2012.
13. Narayan Arya, K., R. Verma, and R.K. Garg, *Estimating the Minimal Clinically Important Difference of an Upper Extremity Recovery Measure in Subacute Stroke Patients.* Topics in Stroke Rehabilitation, 2015. **18**(sup1): p. 599-610.
14. Shogo hiragami, O., R. Yu inoue, and R. Kazuhiro harada, *Minimal clinically important difference for the Fugl-Meyer assessment of the upper extremity in convalescent stroke patients with moderate to severe hemiparesis.* The Journal of Physical Therapy Science, 2019.