Statistical Analysis Plan

1. Study Overview

This Statistical Analysis Plan (SAP) describes the planned statistical analyses for the observational cross-sectional study investigating the relationship between Timed Walking Test (TWT) performance, executive function, and freezing of gait (FOG) severity in people with Parkinson's disease (PD).

The study will include two groups:

- **PD group:** individuals with idiopathic PD and FOG.
- **Healthy control group:** age-matched adults without neurological disorders or clinically significant gait impairment.

Each participant will undergo a single assessment visit. The primary analyses will focus on the PD group. Comparisons between the PD and healthy control groups will be considered secondary analyses.

The primary analyses will address whether TWT performance provides information about executive function beyond standard gait performance, whether the TWT-B minus TWT-A difference is more strongly associated with executive function than TWT-A performance alone, and whether FOG severity is associated with TWT performance independently of standard gait performance.

2. Study Aims and Hypotheses

Aim 1. To determine whether the Timed Walking Test provides additional information about cognitive performance beyond standard single-task and dual-task gait assessments.

H1a: TWT-B performance will be independently associated with executive function, assessed using the Trail Making Test B−A (TMT B−A), after accounting for standard gait performance.

H1b: The difference in completion time between TWT-B and TWT-A will show a stronger association with executive function (TMT B−A) than TWT-A completion time alone.

Aim 2. To investigate the relationship between TWT performance and freezing of gait severity.

H2: Greater FOG severity, assessed using the Giladi–ClinRO, will be associated with longer TWT completion times, even after accounting for baseline single-task gait performance. TWT-A and TWT-B will be analysed separately.

Secondary and exploratory objectives

- To investigate the associations between TWT performance and global cognitive function, assessed using the Montreal Cognitive Assessment (MoCA).
- To explore the associations between TWT performance and working memory and attention, assessed using Digit Span.

- To compare TWT performance and gait measures between participants with PD and healthy controls.

3. Analysis Population

The primary analysis population will consist of participants with PD who meet the study eligibility criteria and have completed the relevant clinical, neuropsychological, gait, and TWT assessments.

Healthy controls will be included in secondary between-group comparisons.

For each analysis, participants will be included if valid data are available for all variables required by the respective statistical model. The number of participants included in each analysis will be reported.

4. Outcome Measures and Variable Definitions

4.1. Primary outcome measures

Executive function

Executive function will primarily be operationalized using the difference between Trail Making Test Part B and Part A completion times (TMT B−A). Higher values indicate poorer performance.

TMT B−A will be calculated as:

TMT B−A = TMT-B completion time − TMT-A completion time.

TWT performance

TWT performance will be assessed using completion time in two conditions:

- **TWT-A:** sequential numerical condition.
- **TWT-B:** alternating numerical-alphabetical condition.

Two recorded trials will be performed for each condition. The arithmetic mean of the two valid trials will be used for all primary analyses:

- Mean TWT-A time = (TWT-A trial 1 + TWT-A trial 2) / 2.
- Mean TWT-B time = (TWT-B trial 1 + TWT-B trial 2) / 2.

Lower completion times indicate better performance.

The TWT difference score will be calculated as:

TWT difference = mean TWT-B time − mean TWT-A time.

A higher difference score indicates a greater increase in completion time during the alternating condition compared with the numerical condition.

Standard gait performance

Standard single-task gait performance will be represented by the completion time of the 10-m gait assessment. This variable will be used as the primary gait covariate in H1a and H2.

The exact operational definition of the 10-m gait assessment, including whether it is a 10-Metre Walk Test or a 10-metre Timed Up and Go protocol, must be confirmed in the final version of the SAP.

4.2. FOG severity

FOG severity will be assessed using the Giladi–ClinRO measure, as specified in the study protocol.

The total or summary score defined by the selected scoring procedure will be used as the primary FOG severity variable. Higher scores will represent greater FOG severity, according to the scoring direction of the instrument.

4.3. Secondary cognitive measures

Global cognitive function will be assessed using the Czech version of the Montreal Cognitive Assessment (MoCA).

Working memory and attention will be assessed using Digit Span.

Additional neuropsychological measures, including Judgment of Line Orientation, CLOX, Stroop Test, and Verbal Fluency Test, may be included in exploratory analyses where appropriate.

4.4. Additional clinical and gait variables

The following variables will be described and may be used in exploratory or sensitivity analyses:

- Age and sex.
- PD duration.
- Years of formal education.
- MDS-UPDRS Part II item 2.13.
- MDS-UPDRS Part III.
- PDQ-8.
- Falls Efficacy Scale International (FES-I).
- Additional gait parameters derived from the APDM Opal inertial measurement units.

The primary models will remain limited to the covariates specified in the hypotheses. Additional covariates will not be introduced into the primary models solely on the basis of statistical significance in univariable analyses.

5. Data Preparation and Descriptive Statistics

Data will be checked for completeness, plausibility, and consistency before statistical analysis.

Continuous variables will be summarized using mean and standard deviation (SD) for approximately normally distributed data, or median and interquartile range (IQR) for skewed data.

Categorical variables will be summarized using frequencies and percentages.

Distributional characteristics will be assessed using graphical methods, including histograms and Q–Q plots, and appropriate statistical tests where relevant.

The distribution of model residuals, rather than the distribution of each variable alone, will be considered when evaluating the assumptions of linear regression.

The number of valid observations and missing values will be reported for all variables used in the analyses.

No imputation of missing data is planned for the primary analyses. Analyses will be conducted using available complete observations for the variables included in each respective model.

6. Primary Statistical Analyses

6.1. H1a: Independent association between TWT-B performance and executive function

The objective of H1a is to determine whether TWT-B performance is associated with executive function beyond the information provided by standard gait performance.

A hierarchical multiple linear regression analysis will be performed in the PD group.

The dependent variable will be TMT B−A completion time.

The following nested models will be evaluated:

Model 1:

$\text{TMT B}-\text{A} = \beta_0 + \beta_1 (\text{10-m gait time}) + \varepsilon$

Model 2:

$\text{TMT B}-\text{A} = \beta_0 + \beta_1 (\text{10-m gait time}) + \beta_2 (\text{mean TWT-B time}) + \varepsilon$

The primary parameter of interest will be the regression coefficient for mean TWT-B time in Model 2.

The incremental explanatory contribution of TWT-B will be evaluated by comparing Model 2 with Model 1 using the change in R^2 (ΔR^2) and the corresponding nested-model comparison.

The regression coefficient, 95% confidence interval (CI), p-value, R^2, adjusted R^2, and ΔR^2 will be reported.

A statistically significant contribution of TWT-B after adjustment for standard gait performance will support H1a.

Regression assumptions will be assessed by examining linearity, residual distribution, homoscedasticity, influential observations, and multicollinearity. If substantial deviations from model assumptions are identified, appropriate sensitivity analyses, such as robust standard errors or alternative model specifications, will be considered and clearly reported.

6.2. H1b: Comparison of associations between TWT-A and the TWT difference score with executive function

The objective of H1b is to determine whether the TWT difference score is more strongly associated with executive function than TWT-A performance alone.

Spearman's rank correlation coefficients will be calculated for the following associations:

1. Mean TWT-A completion time and TMT B−A.
2. TWT difference score (mean TWT-B minus mean TWT-A) and TMT B−A.

The two correlation coefficients are statistically dependent because they are calculated in the same participants and share the same outcome variable.

Therefore, their difference will be evaluated directly using a paired bootstrap procedure. Participants will be resampled with replacement, preserving the pairing of all variables within each participant. Both Spearman correlation coefficients and their difference will be recalculated in each bootstrap sample.

The bootstrap distribution of the difference between the two correlation coefficients will be used to estimate a 95% confidence interval and assess whether the association involving the TWT difference score is stronger than the association involving TWT-A.

The number of bootstrap resamples will be specified in the final analysis implementation. The direction and magnitude of the difference between the correlation coefficients, together with its confidence interval and p-value where calculated, will be reported.

H1b will be supported if the association between the TWT difference score and TMT B−A is significantly stronger than the association between TWT-A and TMT B−A in the hypothesized direction.

6.3. H2: Association between FOG severity and TWT performance

The objective of H2 is to investigate whether greater FOG severity is associated with poorer TWT performance independently of standard single-task gait performance.

TWT-A and TWT-B will be analysed separately to distinguish the association between FOG severity and performance in the numerical and alternating conditions.

Two separate regression models will be evaluated:

Model A: TWT-A

Mean TWT-A time = $\beta_0 + \beta_1$ (10-m gait time) + β_2 (Giladi–ClinRO score) + ε

Model B: TWT-B

Mean TWT-B time = $\beta_0 + \beta_1$ (10-m gait time) + β_2 (Giladi–ClinRO score) + ε

The primary parameter of interest in each model will be the regression coefficient for FOG severity.

Regression coefficients, 95% CIs, p-values, R^2, and adjusted R^2 will be reported.

A positive association between FOG severity and TWT completion time, after adjustment for standard gait performance, will support H2.

The two TWT conditions will be interpreted separately. The analyses will therefore establish whether the association with FOG severity is present in TWT-A, TWT-B, or both.

Regression assumptions and potential influential observations will be assessed as described for H1a. If the assumptions are not adequately met, appropriate sensitivity analyses will be considered.

7. Secondary and Exploratory Analyses

7.1. Global cognitive function

Associations between MoCA scores and the following TWT variables will be investigated:

- Mean TWT-A completion time.
- Mean TWT-B completion time.
- TWT difference score.

Spearman's rank correlation coefficients will be used as the primary measure of association. Correlation coefficients, 95% CIs where available, and p-values will be reported.

These analyses will be considered secondary and will not replace the primary analyses using TMT B−A.

7.2. Working memory and attention

Associations between Digit Span performance and TWT-A, TWT-B, and the TWT difference score will be explored using appropriate correlation analyses.

The direction of scoring will be considered when interpreting the results.

These analyses will be considered exploratory.

7.3. Comparisons between PD participants and healthy controls

TWT performance and standard gait measures will be compared between the PD and healthy control groups.

Between-group comparisons will include:

- Mean TWT-A completion time.
- Mean TWT-B completion time.

- TWT difference score.
- Standard gait performance.

The choice of statistical test will depend on the distribution of the data and the actual matching procedure.

For independently sampled groups, independent-samples t-tests or Mann–Whitney U tests will be used as appropriate. If participants have been individually matched, the paired structure will be taken into account using paired statistical tests or corresponding regression methods.

Effect estimates and 95% CIs will be reported where appropriate.

These analyses will be considered secondary.

8. Multiple Comparisons

The primary hypothesis-testing family will comprise the following four tests:

1. H1a: incremental contribution of TWT-B to the regression model.
2. H1b: direct comparison of the two dependent correlation coefficients.
3. H2: association between FOG severity and TWT-A.
4. H2: association between FOG severity and TWT-B.

The Holm procedure will be used to control the family-wise error rate across these four primary tests.

Secondary and exploratory analyses will be interpreted with appropriate caution. Where multiple related correlations are evaluated, adjusted p-values will be reported using the Holm procedure or another explicitly specified correction appropriate to the analysis family.

Unadjusted and adjusted p-values will be distinguished clearly in the reporting of results.

9. Statistical Significance and Reporting

All statistical tests will be two-sided.

A nominal significance level of $\alpha = 0.05$ will be used. For the primary hypothesis-testing family, statistical significance will be determined using Holm-adjusted p-values.

Results will be reported with effect estimates and 95% confidence intervals wherever possible. Statistical significance will not be interpreted in isolation from the magnitude and precision of the estimated effects.

The number of participants included in each analysis will be reported.

10. Statistical Software

Statistical analyses will be performed using Python – Visual Studio Code software.

All data processing, statistical models, and derived variables will be documented to support reproducibility.

11. Sample Size and Power Calculation

The planned sample size is 30-35 patients with PD. An a priori power calculation was performed using effect estimates informed by a preliminary pilot study involving 25 participants. The pilot study demonstrated associations between cognitive performance and TWT-A and TWT-B completion times. However, as TMT B−A was not assessed in the pilot study, the expected incremental explained variance (ΔR^2) for the primary regression analysis (H1a) could not be directly estimated.

Therefore, a sensitivity analysis was performed using different assumptions about the effect size, expressed as Cohen's f^2. Based on the pilot findings and the expected moderate-to-large association between cognitive performance and TWT performance, the following sample size estimates were obtained:

Cohen's f^2	Effect size	Required sample size
0.15	Medium	55
0.35	Large	25

The calculations indicate that a sample of approximately 55 participants would be required to detect a medium effect, whereas approximately 25 participants would be sufficient to detect a large effect, under the assumptions used in the power analysis.

The target sample size will therefore be determined with consideration of the expected effect size and feasibility of recruitment. The results of the primary analyses will be interpreted in the context of these sample size assumptions.

12. Deviations from the Statistical Analysis Plan

Any deviations from this SAP will be documented, justified, and reported transparently.

Analyses not specified as primary in this document will be identified as secondary, exploratory, or post hoc, as appropriate.