Perturbation training for improving reactive stepping responses to prevent falls in Parkinson's Disease (ReSteP) – a randomized controlled trial

- Statistical Analysis Plan -

Principal investigator:

Prof. Dr. Simon Steib
Institute for Sports and Sports Sciences
Heidelberg University
E-Mail: simon.steib@issw.uni-heidelberg.de

Sub-investigators:

Dr. Eunyoung Kwag
Institute for Sports and Sports Sciences
Heidelberg University
E-Mail: eunyoung.kwag@issw.uni-heidelberg.de

Pia Rueß, M.Sc.
Institute for Sports and Sports Sciences
Heidelberg University
E-Mail: pia.ruess@issw.uni-heidelberg.de

Collaborators (alphabetical order):

Dr. med. Andreas Becker (srh Kurpfalzkrankenhaus Heidelberg)

PD Dr. med. Julian Conrad (Section for Neurodegeneration, Neurological Clinic, University Medical Centre Mannheim, Heidelberg University)

Dr. Marietta Kirchner (Institute for Medical Biometry, Dept of Clinical Studies, Heidelberg University Hospital)

Prof. Dr. med. Rebecca Schüle (Section for Neurodegenerative Diseases and Movement Disorders, Dept of Neurology, Heidelberg University Hospital and Medical Faculty Heidelberg)

PD Dr. med. Joachim Wolf (Dept of Neurology, Brüderklinikum Julia Lanz Mannheim)

Date: 13.11.2026
Version 1.1

Table of contents

1 Summary 3
2 Study objectives, hypotheses, and outcomes 3
2.1 Primary objective and outcome 3
2.2 Secondary objectives and outcomes 3
2.3 Exploratory outcomes 4
2.4 Descriptive outcomes 4
3 Study design 4
3.1 Sample size justification 4
3.2 Randomization, stratification, and blinding 4
4 Study population 5
5 Data handling 6
5.1 Missing data 6
6 Analysis populations 8
7 Statistical analysis 8
7.1 Primary outcome analysis 16
7.2 Key secondary outcome analysis 16
7.3 Other continuous secondary and exploratory outcomes 16
7.4 Fall-related outcomes 16
7.5 Safety analysis 17
7.6 Interim analyses 17
8 Revision history 18
9 Appendix 19
10 References 22

1 Summary

Falls are a major clinical concern in people with Parkinson's disease (PwPD), who have approximately twice the fall risk of age-matched adults. Impairments in gait, postural stability, and particularly reactive balance contribute substantially to fall risk in PwPD, as the ability to rapidly generate appropriate motor responses to unexpected perturbations is impaired (Hulzinga et al., 2021; Murueta-Goyena et al., 2024). Perturbation-based balance training (PBT) specifically targets reactive balance by repeatedly exposing individuals to controlled, unexpected external perturbations, thereby providing task-specific practice of rapid balance recovery responses. Initial studies in PwPD have reported improvements in balance performance, gait ability, and fall rates following several weeks of PBT (Klamroth et al., 2016; Smania et al., 2010; Steib et al., 2017). However, the available evidence is largely based on small sample sizes and exhibits methodological limitations. To address these gaps, this randomized controlled trial will evaluate the efficacy of treadmill-based PBT for improving reactive balance in PwPD. Secondary objectives are to investigate its effects on falls and other clinically relevant balance, mobility, motor, non-motor and psychological measures, and to evaluate the intervention's feasibility and acceptability.

2 Study objectives, hypotheses, and outcomes

2.1 Primary objective and outcome

The primary objective of this study is to evaluate the efficacy of treadmill-based PBT for improving reactive balance in response to gait perturbations in PwPD.

Main hypothesis: PBT will lead to greater improvements in reactive balance responses to laboratory-induced gait perturbations (higher score in the modified Dynamic Stepping Threshold Test; mDSTT) immediately post-intervention compared to conventional treadmill training (CTT).

2.2 Secondary objectives and outcomes

The key secondary objective is to evaluate the efficacy of treadmill-based PBT for improving reactive balance in response to stance perturbations in PwPD.

Key secondary hypothesis: PBT will result in greater improvements in reactive balance responses to laboratory-induced stance perturbations (higher score in the Stepping Threshold Test; STT) immediately post-intervention compared to CTT.

Further secondary outcomes will be used to examine the feasibility and acceptability of PBT, its effects on real-life fall incidence, and the retention of training effects at 6-month follow-ups. Additional outcomes will assess transfer to reactive and dynamic balance, gait performance, psychological outcomes including balance confidence and fear of falling, and Parkinson's disease-related motor symptoms.

2.3 Exploratory outcomes

Exploratory analyses will aim to improve the understanding of the underlying mechanisms of reactive balance by examining changes in neuromuscular activity and trunk movements in response to perturbations. Associations between changes in reactive balance performance and various potential predictors (e.g., cognitive and executive functions, physical performance and mobility outcomes) will be examined. Furthermore, health-related quality of life, real-life physical activity, sleep quality, and mobility behavior will be explored.

2.4 Descriptive outcomes

Descriptive outcomes include participants' demographic characteristics such as age, sex, education level, and history of falls, as well as disease-related characteristics, including i.a. disease duration, Hoehn and Yahr stage and the presence of freezing of gait.

3 Study design

This study is designed as an assessor-blinded, randomized controlled trial.

3.1 Sample size justification

The target sample size was determined pragmatically, based on the resources available and the number of participants considered feasible to recruit. This estimate was informed by the research group's established clinical and healthcare network in the study region, the planned intervention procedures, and the regional prevalence of Parkinson's disease. On this basis, recruitment of 90 participants was considered feasible. Drawing on prior randomized controlled trials conducted by the research group in comparable cohorts (Wanner et al., 2021), as well as comparable studies from other groups (Nørgaard et al., 2023; Pelicioni et al., 2023), an overall dropout rate of approximately 20% was expected over the course of the study, yielding an anticipated evaluable sample of N = 72 participants. Using G*Power (version 3.1.9.7), a sensitivity power analysis on the primary endpoint revealed that a total sample of 72 participants would provide 80% power to detect a between-group effect of f=0.33, at a two-sided significance level of α=0.05, using an ANCOVA adjusted for baseline (T0) values and disease stage (Hoehn & Yahr). This effect size was considered plausible in light of previous work in individuals with Parkinson's disease (Steib et al., 2017) and intervention studies in neurologically healthy older adults (Bhatt et al., 2012; Brüll et al., 2023; Hezel et al., 2025; Kurz et al., 2016; Okubo et al., 2019; Pai et al., 2014).

3.2 Randomization, stratification, and blinding

Participants will be randomly allocated to the intervention or active control group in a 1:1 ratio using stratified permuted-block randomization with variable block sizes. Stratification is applied according to Hoehn & Yahr stage (1-2 vs. 3) to balance disease severity across study groups, as this may influence balance performance (Harro et al., 2018; Santos et al., 2023). A randomization sequence is computer-generated using the Sealed Envelope online blocked randomization list generator (version 1.23.1; Sealed Envelope Ltd., London, UK)

prior to the start of the study. Randomization takes place only after completion of all baseline assessments. To prevent selection bias, randomization is performed by research staff who are not involved in data collection or outcome assessments, thereby ensuring allocation concealment. The trial employs an assessor-blinded design, whereby the research staff responsible for conducting outcome assessments are blinded to group allocation to minimize detection bias, while participants and intervention providers are inevitably aware of treatment assignment due to the nature of the interventions.

4 Study population

Inclusion criteria:

- Parkinson's Disease diagnosed (MDS diagnostic criteria)
- Stage of the Parkinson's Disease: Hoehn & Yahr ≤ 3
- Increased fall risk: ≥1 fall in the previous 12 months or ≥ 12 seconds in the Timed Up and Go Test
- Stable medication for at least 4 weeks
- Age: 40 to 85 years

Exclusion criteria:

- Unable to provide informed consent
- Montreal Cognitive Assessment (MoCA) score ≤ 17 and significant functional impairment resulting from cognitive impairment (e.g., inability to manage medications or finances independently)
- Atypical Parkinsonism
- Comorbidities: Neurological, cardiovascular, pulmonary, orthopedic or psychiatric contraindications for the planned treadmill interventions
- Deep brain stimulation
- Concurrent enrollment in other interventional clinical trials involving pharmacological treatment and/or structured physical training programs
- Prior experience with structured perturbation training programs systematically involving gait perturbations in the past 2 years
- Inability to walk on the treadmill without assistance for at least 2 minutes at 70% of habitual overground gait speed (assessed during screening)
- Body weight > 135 kg (exceeds the treadmill manufacturer's specified weight limit)

5 Data handling

Prior to lock of the database for subsequent statistical analysis, the analysis dataset will undergo predefined data-quality checks, including checks for completeness, valid ranges, internal consistency, duplicate records, and plausibility of derived variables. Identified discrepancies will be reviewed against the source data and corrected where appropriate. All data corrections and decisions regarding data queries will be documented. Implausible values will not be excluded solely on the basis of their extremeness if they are confirmed to represent valid observations. Prior to inferential analyses, data will be descriptively and graphically inspected for distributions, outliers, missing-data patterns, and plausibility. Model diagnostics will be performed for the prespecified statistical models, including graphical assessment of residual distributions. All statistical analyses are planned to be performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria) and/or SPSS (IBM Corp, Armonk, NY, United States). The alpha level for statistical significance will be set at $p \leq .05$, unless otherwise stated. Additionally, effect sizes will be reported.

5.1 Missing data

The number and proportion of missing observations will be reported for each outcome, assessment time point (T0, T1, and T2), and treatment group. Where available, the reasons for missingness will be documented and summarized by treatment group. The extent and patterns of missing data may be examined, where appropriate, to support the interpretation of the results and inform decisions regarding additional sensitivity analyses.

Primary and key secondary outcome

Missing post-baseline outcome measurements at T1 and/or T2 will not be imputed explicitly in the primary and key secondary linear mixed model for repeated measures (MMRM) analysis, which will use all available observed outcome measurements under the missing-at-random assumption (MAR).

If baseline modified DSTT or STT scores at T0 are missing, 50 imputed datasets will be generated using predictive mean matching (PMM). The imputation model will include treatment group, H&Y stage, available T1/T2 outcomes, and relevant baseline variables. Estimates will be pooled using Rubin's rules. As a sensitivity analysis, missing T1 and T2 outcomes, including cases where both are missing, will be imputed using 50 datasets under the MAR assumption. The same MMRM will be fitted to each dataset, and estimates will be pooled using Rubin's rules. No imputation is planned for H&Y stage, as complete data are expected. A supplementary complete-case analysis will include participants with observed modified DSTT or STT scores at both T0 and T1.

Secondary outcomes

The extent and patterns of missing data will be summarised by study group and assessment time point, with reasons documented where available. No single imputation method will be applied routinely.

For continuous outcomes analysed using MMRM, including Mini-BESTest, TUG, gait speed, 2MWT, FES-I, ABC Scale, PACES, and MDS-UPDRS Part III scores, all available observed measurements will be included without prior imputation, assuming data are missing at random (MAR) conditional on the variables included in the model.

For questionnaire-based outcomes, including PACES, ABC Scale, and acceptability measures, incomplete responses will be handled according to the relevant instrument-specific scoring guidance (see Table 2 in Appendix). Missing items will not routinely be imputed. For clinical assessments, like the Mini-BESTest, no total score will be calculated if one or more items are missing. Domain-specific scores will be calculated, if all items are available for the respective domain. For part III of the MDS-UPDRS, the motor score will be calculated if up to seven randomly missing items are present (Goetz et al., 2015). Assessment exceeding this threshold will be considered missing.

For binary, categorical, count, rate, and time-to-event and fall-related outcomes, missing data will be handled according to the planned analysis method. Missing observations will not automatically be interpreted as zero events or absence of an outcome. For fall-related outcomes, periods with missing records will be distinguished from periods with confirmed absence of falls, and available valid observation time will be used where appropriate.

Qualitative interview data will be analysed using the available material; unavailable interviews and incomplete data will be documented where possible.

Exploratory outcomes

Missing data for exploratory outcomes will be summarised by study group and assessment time point, with reasons documented where available. No routine imputation will be performed.

For continuous outcomes, including PDQ-39, NMSS, cognitive test scores, trunk kinematics, and EMG-derived measures, analyses will use available observed data according to the planned statistical model. For repeated continuous outcomes analysed using MMRM or linear mixed models, this approach assumes data are missing at random (MAR), conditional on the variables included in the model.

For sensor-derived outcomes from Axivity AX6 and ActiGraph LEAP, data completeness and quality will be considered when deriving activity, mobility, and sleep measures. Measurements with insufficient or unreliable data will be treated as missing where appropriate; missing sensor data will not routinely be imputed.

For subgroup assessments, including Go/No-go and Stop-signal tasks, missingness and the number of participants with available data will be reported. Missing data will be handled according to the outcome type and planned analysis.

Descriptive outcomes

Descriptive outcomes will be summarised using available observed data. The number of available observations will be reported for each variable, and missing values will not be imputed solely for descriptive purposes. Denominators will be reported where relevant.

6 Analysis populations

Intention-to-treat analysis set (ITT):

In the ITT analysis set, all participants will be analyzed in the group to which they were randomized. The analysis of all endpoints, unless noted otherwise, will be conducted on the ITT analysis set.

Per-protocol analysis set (PP):

Participants who complete the study period without major protocol violations will be included in the PP analysis. Expected major protocol violations include:

1. Completion of <75% of the scheduled training sessions
2. Minimum exposure to the defining components of the allocated intervention <75% (i.e., in the intervention group, <75% of the planned gait and stance perturbations, respectively, and in the control group, <75% of the planned treadmill walking time)

Complete-case analysis set:

Only participants who qualify for the ITT population with available data at T0, T1 and T2 will be included in the complete-case analysis.

7 Statistical analysis

All analyses will follow the intention-to-treat principle. A supplementary per-protocol and compete case analysis will be conducted. The outcome variables will be assessed at baseline (T0, within 2 weeks prior to intervention start), post-intervention (T1, within 1 week after end of intervention) and 6-month follow-up (T2, 24 weeks after end of intervention). Real-life falls will be monitored up to 12 months after end of intervention (T3, 52 weeks after end of intervention).

Table 1. Overview of all study variables and their respective methods of analysis.

Primary outcome					
Outcome	**Variable**	**Assessment tool**	**Time of assessment**	**Type of variable**	**Methods of analysis**
Reactive balance (gait perturbations)	mDSTT total score	Modified Dynamic Stepping Threshold Test (mDSTT)	T0, T1	Continuous	MMRM
Key secondary outcome					
Outcome	**Variable**	**Assessment tool**	**Time of assessment**	**Type of variable**	**Methods of analysis**
Reactive balance (stance perturbations)	STT total score	Stepping Threshold Test (STT)	T0, T1	Continuous	MMRM
Secondary outcomes					
Outcome	**Variable**	**Assessment tool**	**Time of assessment**	**Type of variable**	**Methods of analysis**
Feasibility	Training session completion rate	Feasibility documentation	T0 - T1	Proportion	Descriptive or Welch’s t-test
	Dropout rate	Feasibility documentation	T0 - T1	Binary / proportion	Descriptive or Risk ratio with 95% CI; Fisher’s exact or chi-square test
	Perceived physical and mental load	NASA-Task Load Index item	T0 - T1	Ordinal (Likert-type scale)	Mann–Whitney U test

	Total number of perturbations completed	Training records	T0 - T1 (only intervention group)	Count / Proportion	Descriptive
	Intensity of perturbations completed	Training records	T0 - T1 (only intervention group)	Ordinal	Descriptive
Safety	Incidence and severity of adverse events	Exercise Harms Reporting Method, the Common Terminology Criteria for Adverse Events, and informed questionnaire and event documentation	T0 - T3	Count (number of AEs); categorical (severity and type)	Descriptive
Acceptability	General acceptability score	Theoretical framework of acceptability-based questionnaire	T1	Ordinal (single item)	Mann–Whitney U test
	Acceptability sum score	Theoretical framework of acceptability-based questionnaire	T1	Continuous	Welch’s t-test or Mann–Whitney U test
	Enjoyment score	Physical Activity Enjoyment Scale (German version)	T1	Continuous	Welch’s t-test or Mann–Whitney U test
	Interview findings	Semi-structured interviews	T1	Qualitative	Thematic analysis

Real-life falls	Fall rate per person-year	Fall diaries supplemented by telephone interviews	T0 - T3	Count rate	Negative binomial or Poisson regression; IRR with 95% CI
	Rate of injurious falls	Fall protocol supplemented by telephone interviews	T0 - T3	Count rate	Negative binomial or Poisson regression; IRR with 95% CI
	Proportion of participants who experienced at least one fall (≥1 fall)	Fall diaries supplemented by telephone interviews	T0 - T3	Count and proportion	Descriptive and robust Poisson regression; RR with 85% CI
	Total number of falls	Fall diaries supplemented by telephone interviews	T0 - T3	Count	Descriptive
	Non-fallers, single-fallers, recurrent fallers	Fall diaries supplemented by telephone interviews	T0 - T3	Three-category categorical	Descriptive
	Time to first fall	Fall diaries supplemented by telephone interviews	T0 - T3	Time-to-event	Kaplan–Meier curves or Cox regression; HR with 95% CI
	Activity-adjusted falls	Fall diaries and activity monitoring	T0 - T3	Count / exposure-adjusted rate	Negative binomial regression or Poisson regression with an appropriate

					activity-exposure offset; IRR with 95% CI
Reactive balance (gait perturbations) – retention effects	mDSTT total score	Modified Dynamic Stepping Threshold Test (mDSTT)	T1, T2	Continuous	MMRM
Reactive balance (stance perturbations) – retention effects	STT total score	Stepping Threshold Test (STT)	T1, T2	Continuous	MMRM
Reactive balance (gait perturbations) – direction-specific performance	Direction-specific mDSTT scores	mDSTT	T0, T1, T2	Continuous	MMRM
Reactive balance (gait perturbations) – direction-specific performance	STT subscores	STT	T0, T1, T2	Continuous	MMRM
Balance performance	Mini-BESTest total and domain scores	Mini-BESTest	T0, T1, T2	Continuous	MMRM

Mobility	TUG completion time (s)	Timed Up and Go Test (TUG)	T0, T1, T2	Continuous	MMRM
Gait performance	Habitual and fast gait speed (m/s)	10-Meter Walk Test	T0, T1, T2	Continuous	MMRM
	Two-minute walking capacity (m)	2-Minute Walk Test	T0, T1, T2	Continuous	MMRM
Concerns about falling	FES-I total score	Falls Efficacy Scale-International (FES-I)	T0, T1, T2	Continuous	MMRM
Balance confidence	ABC total score	Activities-specific Balance Confidence (ABC) Scale	T0, T1, T2	Continuous	MMRM
Disease-related motor symptoms	UPDRS motor score	Part III of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS-III)	T0, T1, T2	Continuous	MMRM
Exploratory outcomes					
Outcome	**Variable**	**Assessment tool**	**Time of assessment**	**Type of variable**	**Methods of analysis**
Health-related quality of life	PDQ-39 total score	Parkinson's Disease Questionnaire-39 (PDQ-39)	T0, T2	Continuous	MMRM

Disease-related non-motor symptoms	NMSQuest total score	Non-Motor Symptoms Questionnaire for Parkinson's Disease (NMSQuest)	T0, T2	Continuous	MMRM
Real-world mobility	Real-world digital mobility outcomes, e.g., daily walking amount, walking speed	Inertial measurement unit at L5 (Axivity AX6)	T0, T1, T2	Continuous	MMRM
Physical activity	Actigraphy-based time spent in sedentary, light, and moderate-to-vigorous activity	Wrist-worn accelerometer (ActiGraph LEAP)	T0, T1, T2	Continuous	MMRM
Sleep	Sleep quality outcomes, e.g., total sleep time, sleep efficiency	Wrist-worn accelerometer (ActiGraph LEAP)	T0, T1, T2	Continuous	MMRM
Executive functions	Composite z-score derived from four executive functions tests	Go/No-go test (commission error rate); Stop-signal test (stop-signal reaction time); Trail Making test (completion time difference between Parts B and A); Backward Corsi Block test (total score)	T0, T1 (subgroup)	Continuous	MMRM (subgroup); Linear regression or Pearson correlation

Neuromuscular activity	Lower-limb muscle activation and co-contraction parameters, e.g., agonist-antagonist co-contraction patterns	Surface EMG during mDSTT and STT	T0 (subgroup), T1 (subgroup), T2 (subgroup)	Continuous	MMRM; Linear regression or Pearson correlation
Compensatory balance reactions	Composite score of the perturbations at the individual highest anterior and mediolateral intensity level during the modified Dynamic Stepping Threshold Test	Adapted Compensatory Arm and Leg Movements (A-CALM) scale	T0, T1, T2	Continuous	MMRM
	Composite score of the anterior, posterior, and mediolateral perturbations at the individual multiple stepping threshold intensity level during the Stepping Threshold Test	Compensatory Arm and Leg Movements (CALM) scale	T0, T1, T2	Continuous	MMRM
	Trunk angle and angular velocity	Axivity AX6 IMUs at L5 and sternum during mDSTT and STT	T0, T1, T2	Continuous	MMRM; Linear regression or Pearson correlation

CI: Confidence Interval; EMG: Electromyography; HR: Hazard Ratio; IRR: Incidence Rate Ratio; MMRM: Mixed Model for Repeated Measures

7.1 Primary outcome analysis

The primary outcome will be analyzed using a linear mixed model for repeated measures (MMRM), using the observed modified DSTT score at each of the two post-intervention timepoints (T1 and T2) as the repeated dependent variable. Fixed effects will include treatment group (PBT and CTT), timepoint (T1 and T2), group x timepoint interaction, modified DSTT score at baseline (T0) as a covariate, and H&Y stage as a covariate. Within-participant correlation between repeated measurements will be modelled using an unstructured residual covariance matrix, with participant ID specified as the subject variable. If this model fails to converge, a first-order autoregressive (AR(1)) covariance structure will be used as a fallback. The primary intervention effect will be tested using the prespecified contrast estimating the baseline-adjusted between group difference in modified DSTT score at T1.

The analysis will be conducted on the full analysis set according to the ITT principle. As a supplementary analysis, the primary analysis model will be extended to additionally adjust for age and sex and the primary endpoint will also be analyzed in the per-protocol dataset. A complete-case analysis will additionally be performed.

7.2 Key secondary outcome analysis

Group difference in reactive balance in response to stance perturbations at T1, assessed with the STT, is defined as the key secondary outcome. This outcome will be tested in a confirmatory manner using a fixed-sequence (hierarchical) testing procedure at the same overall significance level of $\alpha = 0.05$, without any further alpha-splitting between the two outcomes. Under this fixed-sequence procedure, the key secondary outcome (STT) will be tested in a confirmatory manner only if the primary outcome is statistically significant. If the primary outcome is not significant, testing stops; the STT comparison will still be reported, but descriptively/exploratory only, regardless of its *p*-value. For statistical analysis of the key secondary outcome, an MMRM will be used analogously to the primary outcome analysis.

7.3 Other continuous secondary and exploratory outcomes

Other continuous secondary and exploratory outcomes, including retention measures for the primary and key secondary outcomes, will be analyzed using the same MMRM framework, with the corresponding baseline outcome value included as a covariate.

7.4 Fall-related outcomes

Fall outcomes will be assessed at 6 and 12 months of follow-up. For each treatment group, the total number of falls and number of falls with and without injuries during the post-intervention follow-up period (T1 – T3) will be summarized as an incidence rate (IR), expressed as the total number of falls and falls with and without injuries per person-time and calculated by dividing the total number of falls and falls with and without injuries by the total person-time at risk. The proportion of participants who experienced at least one fall and who experienced at least one injurious fall during this period will also be reported, calculated as

the number of participants who experienced at least one fall and who experienced at least one injurious fall divided by the total number of participants with available follow-up data. Both measures will be reported with their respective 95% confidence intervals (CIs). Participants with at least 6 months of available fall diary data will be included in the 6-month and/or 12-month IR analyses, as applicable, with the observed follow-up time taken into account. Fall proportions will be calculated separately at each time point and will include only participants with complete fall diary data for the corresponding follow-up period (6 or 12 months). To account for risk of exposure time for falls, the number of falls will additionally be calculated with adjustment for physical activity, as assessed during the seven-day physical activity monitoring.

Between-group comparisons for fall rate, rate of injurious falls and activity-adjusted falls will be analyzed using negative binomial regression and expressed as an incidence rate ratio (IRR). Overdispersion will be assessed to determine the appropriate count regression model; if there is no evidence of overdispersion, Poisson regression will be used instead.

Between-group differences in the proportion of participants who experienced at least one fall will be analyzed using a generalized linear model with a binomial distribution and a log link function to estimate the risk ratio (RR).

Time to first fall will be analyzed using Kaplan–Meier survival curves, with between-group differences assessed using the log-rank test. A Cox proportional hazards regression model will be used to estimate the hazard ratio (HR) and corresponding 95% CI for the intervention effect. Time will be calculated from the end of the intervention to the first fall, with participants who do not experience a fall censored at the end of the 12-month follow-up period or at the date of their last available fall assessment. The proportional hazards assumption will be assessed.

7.5 Safety analysis

Safety will be assessed using structured questionnaires and event documentation during the training phase and throughout the intervention period. Safety outcomes will be analyzed descriptively.

7.6 Interim analyses

No formal interim analysis is planned. In particular, no interim analysis will be conducted for the purpose of early efficacy or futility stopping, sample-size re-estimation, treatment-arm selection, or other modification of the trial design. Preliminary analyses of accumulated data may be conducted during the trial for purposes such as scientific communication. Such analyses will not influence trial conduct or the prespecified statistical analysis and will be clearly identified as preliminary in any dissemination.

8 Revision history

Version	Date	Section(s) affected	Description of changes
1.0	30.10.2026	-	Initial version of the SAP
1.1	13.11.2026	Statistical Analysis – Table 1 – Exploratory Outcomes	– Refinement of the exploratory physical activity outcome measure to focus the analysis on the most relevant physical activity intensity parameters, in response to editorial feedback from ClinicalTrials.gov and to improve transparency – Revision of the exploratory executive function outcome measure from separate test outcomes to one composite score, following editorial feedback from ClinicalTrials.gov to enhance transparency and clarity – Editorial clarification of the two exploratory compensatory balance reaction measures assessed with the A-/ CALM scale to improve the transparency and reproducibility of the outcome measure definition, in response to ClinicalTrials.gov review comments

9 Appendix

Table 2. Overview of the questionnaires used in this study and their respective scoring methods and handling of missing items.

Questionnaire	Scoring method	Handling of missing items
TFA-based acceptability questionnaire	– The items „burden“ and „opportunity costs” are reversed scored, so that a higher score always indicates higher acceptability – Acceptability mean score: mean score of the 7 TFA items → minimum 7 (lowest acceptability) to maximum 35 (highest acceptability) – Acceptability score: score of the general acceptability item → minimum 1 (lowest acceptability) to maximum 5 (highest acceptability)	– If ≥1 of the 7 TFA items is missing, no acceptability mean score will be calculated – If the general acceptability item is missing, no acceptability score will be reported
PACES	– Negatively worded items are reverse coded, so that a higher score always indicates higher enjoyment – Enjoyment score: mean score of the items → minimum 1 (lowest enjoyment) to maximum 5 (highest enjoyment)	– If ≥1 of the items is missing, no enjoyment score will be calculated
FES-I	– FES-I total score: sum score of all 16 items → minimum 16 (no concern about falling) to maximum 64 (severe concern about falling)	– If ≥5 items are missing, no FES-I total score will be calculated – If ≤4 items are missing, the sum of the items which have been completed is calculated, and this score will be divided by the number of items

		completed and multiplied by 16 and rounded up to the nearest whole number
ABC Scale	– ABC total score: mean score of the 16 items → minimum 0% (no confidence) to maximum 100% (full confidence)	– If ≥5 items are missing, no ABC total score will be calculated – If ≤4 items are missing, the mean score of the items which have been completed is calculated
PDQ-39	– Dimension total scores: sum score of the corresponding dimension items, transformed to a scale from 0 (never have difficulty) and 100 (always have difficulty) by multiplying the dimension sum score by 100 and dividing this value by the maximum possible value of the dimension sum score o Mobility (10 items) o Activities of daily living (6 items) o Emotional well-being (6 items) o Stigma (4 items) o Social support (3 items) o Cognitions (4 items) o Communication (3 items) o Bodily discomfort (3 items) – PDQ-39 SI: mean of the dimension total scores → minimum 0 (never have difficulty) to maximum 100 (always have difficulty)	– If ≥50% of the items in a dimension have been answered, missing responses are replaced with the mean of the available responses within that dimension; after replacing missing responses with the mean of the available responses within each dimension, the sum scores for the eight dimensions are calculated by summation and transformation, as described before

NMSQuest	– NMSQuest total score: number of non-motor symptoms present → minimum 0 (no non-motor symptoms present) to maximum 30 (all of the 30 non-motor symptoms present)	– If ≥1 of the items is missing, no NMSQuest total score will be calculated

10 References

Bhatt, T., Yang, F., & Pai, Y.-C. (2012). Learning to resist gait-slip falls: Long-term retention in community-dwelling older adults. *Archives of Physical Medicine and Rehabilitation*, *93*(4), 557–564. https://doi.org/10.1016/j.apmr.2011.10.027

Brüll, L., Hezel, N., Arampatzis, A., & Schwenk, M. (2023). Comparing the Effects of Two Perturbation-Based Balance Training Paradigms in Fall-Prone Older Adults: A Randomized Controlled Trial. *Gerontology*, *69*(7), 910–922. https://doi.org/10.1159/000530167

Fiorenzato, E., Cauzzo, S., Weis, L., Garon, M., Pistonesi, F., Cianci, V., Nasi, M. L., Vianello, F., Zecchinelli, A. L., Pezzoli, G., Reali, E., Pozzi, B., Isaias, I. U., Siri, C., Santangelo, G., Cuoco, S., Barone, P., Antonini, A., & Biundo, R. (2024). Optimal MMSE and MoCA cutoffs for cognitive diagnoses in Parkinson's disease: A data-driven decision tree model. *Journal of the Neurological Sciences*, *466*, 123283. https://doi.org/10.1016/j.jns.2024.123283

Goetz, C. G., Luo, S., Wang, L., Tilley, B. C., LaPelle, N. R., & Stebbins, G. T. (2015). Handling missing values in the MDS-UPDRS. *Movement Disorders*, *30*(12), 1632–1638. https://doi.org/10.1002/mds.26153

Harro, C. C., Kelch, A., Hargis, C., & DeWitt, A. (2018). Comparing Balance Performance on Force Platform Measures in Individuals with Parkinson's Disease and Healthy Adults. *Parkinson's Disease*, *2018*, 6142579. https://doi.org/10.1155/2018/6142579

Hezel, N., Buchner, T., Becker, C., Bauer, J. M., Sloot, L. H., Steib, S., & Werner, C. (2025). Dose-response relationship of treadmill perturbation-based balance training for improving reactive balance in older adults at risk of falling: Results of the FEATURE randomized controlled pilot trial. *European Review of Aging and Physical Activity*, *22*(1), 8. https://doi.org/10.1186/s11556-025-00375-w

Hulzinga, F., Rond, V. de, Vandendoorent, B., Gilat, M., Ginis, P., D'Cruz, N., Schlenstedt, C., & Nieuwboer, A. (2021). Repeated Gait Perturbation Training in Parkinson's Disease and Healthy Older Adults: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, *15*, 732648. https://doi.org/10.3389/fnhum.2021.732648

Klamroth, S., Steib, S., Gaßner, H., Goßler, J., Winkler, J., Eskofier, B [Bjoern], Klucken, J., & Pfeifer, K. (2016). Immediate effects of perturbation treadmill training on gait and postural control in patients with Parkinson's disease. *Gait & Posture*, *50*, 102–108. https://doi.org/10.1016/j.gaitpost.2016.08.020

Kurz, I., Gimmon, Y., Shapiro, A., Debi, R., Snir, Y., & Melzer, I. (2016). Unexpected perturbations training improves balance control and voluntary stepping times in older adults - a double blind randomized control trial. *BMC Geriatrics*, *16*, 58. https://doi.org/10.1186/s12877-016-0223-4

Murueta-Goyena, A., Muiño, O., & Gómez-Esteban, J. C. (2024). Prognostic factors for falls in Parkinson's disease: A systematic review. *Acta Neurologica Belgica*, *124*(2), 395–406. https://doi.org/10.1007/s13760-023-02428-2

Nørgaard, J. E., Andersen, S., Ryg, J., Stevenson, A. J. T., Andreasen, J., Oliveira, A. S., Danielsen, M. B., & Jorgensen, M. G. (2023). Effect of Treadmill Perturbation-Based Balance Training on Fall Rates in Community-Dwelling Older Adults: A Randomized Clinical Trial. *JAMA Network Open*, *6*(4), e238422. https://doi.org/10.1001/jamanetworkopen.2023.8422

Okubo, Y., Sturnieks, D. L., Brodie, M. A., Duran, L., & Lord, S. R. (2019). Effect of Reactive Balance Training Involving Repeated Slips and Trips on Balance Recovery Among Older Adults: A Blinded Randomized Controlled Trial. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *74*(9), 1489–1496. https://doi.org/10.1093/gerona/glz021

Pai, Y.-C., Bhatt, T., Yang, F., & Wang, E. (2014). Perturbation training can reduce community-dwelling older adults' annual fall risk: A randomized controlled trial. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *69*(12), 1586–1594. https://doi.org/10.1093/gerona/glu087

Pelicioni, P. H. S., Lord, S. R., Menant, J. C., Chaplin, C., Canning, C., Brodie, M. A., Sturnieks, D. L., & Okubo, Y. (2023). Combined Reactive and Volitional Step Training Improves Balance Recovery and Stepping Reaction Time in People With Parkinson's Disease: A Randomised Controlled Trial. *Neurorehabilitation and Neural Repair*, 15459683231206743. https://doi.org/10.1177/15459683231206743

Santos, G. V., d'Alencar, M. S., Helene, A. F., Roque, A. C., Miranda, J. G. V., & Piemonte, M. E. P. (2023). A non-expensive bidimensional kinematic balance assessment can detect early postural instability in people with Parkinson's disease. *Frontiers in Neurology*, *14*, 1243445. https://doi.org/10.3389/fneur.2023.1243445

Smania, N., Corato, E., Tinazzi, M., Stanzani, C., Fiaschi, A., Girardi, P., & Gandolfi, M. (2010). Effect of balance training on postural instability in patients with idiopathic Parkinson's disease. *Neurorehabilitation and Neural Repair*, *24*(9), 826–834. https://doi.org/10.1177/1545968310376057

Steib, S., Klamroth, S., Gaßner, H., Pasluosta, C., Eskofier, B [Björn], Winkler, J., Klucken, J., & Pfeifer, K. (2017). Perturbation During Treadmill Training Improves Dynamic Balance and Gait in Parkinson's Disease: A Single-Blind Randomized Controlled Pilot Trial. *Neurorehabilitation and Neural Repair*, *31*(8), 758–768. https://doi.org/10.1177/1545968317721976

Wanner, P., Winterholler, M., Gaßner, H., Winkler, J., Klucken, J., Pfeifer, K., & Steib, S. (2021). Acute exercise following skill practice promotes motor memory consolidation in Parkinson's disease. *Neurobiology of Learning and Memory*, *178*, 107366. https://doi.org/10.1016/j.nlm.2020.107366