UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Statistical Analysis Plan

Version: 1.0 Date: 03.08.2023

Development and Evaluation of a Web-based Programme on Relapse Management for People With Multiple Sclerosis

POWER@MS2

Study Protocol	Version 1.8 from 29.11.2021
EudraCT No.	-
Clinicaltrials.gov-No. / DRKS - No.	-
Internal Protocol ID No.	-
NCT-No.	NCT04233970
SAP Version	1.0/03.08.2023
SAP Revisions	
Development Phase	
Sponsor	University Medical Centre Hamburg Eppendorf Martinistraße 52 20246 Hamburg, Germany
Coordinating Investigator	Prof Dr. Christoph Heesen Institute of Neuroimmunology and Multiple Sclerosis University Medical Centre Hamburg Eppendorf Martinistraße 52 20246 Hamburg

Approval of the Statistical Analysis Plan

Development and Evaluation of a Web-based Programme on Relapse Management for People With Multiple Sclerosis (POWER@MS2)

Protocol Version No: 1.0/03.08.2023

Coordinating Investigator
Prof. Dr. Christoph Heesen

4/8/23
Date


Signature

Biostatistician
Prof. Dr. Tim Friede

03.08.23
Date


Signature

Table of Contents

List of Figures

No table of figures entries found.

List of Tables

List of Abbreviations

BSA-F	*Physical Activity, Exercise and Sport Questionnaire (Bewegungs- und Sportaktivität Fragebogen)*
CG	Control group
CI	Confidence interval
CPS	*Control Preference Scale*
EDSS	*Expanded Disability Status Scale*
EQ-5D	*EuroQol 5-Dimension*
HAQUAMS	*Hamburg Quality of Life in Multiple Sclerosis Scale*
HADS	*Hospital Anxiety and Depression Scale*
HARAS	*Hamburg Relapse Assessment Scale*
IG	Intervention group
MS	Multiple sclerosis
MMRM	Mixed model for repeated measurements
OR	Odds ratio
PAM	*Patient Activation Measure*
PBMS	*Planned Behaviour in MS Scale*
PwMS	People with MS
RCT	Randomised controlled trial
UNDS	*United Kingdom Neurological Disability Scale*

1 Introduction

This document has been written based on the study protocol version 1.8 dated 29th of November 2021. A design paper was published (Rahn *et al,* 2021). This statistical analysis plan focusses on the evaluation by a randomised controlled trial (RCT). Additionally, a mixed-methods process evaluation and a health economic evaluation will be carried out.

1.1 Background and Rationale

Multiple sclerosis (MS) is a chronic inflammatory, degenerative disease of the central nervous system affecting about 200,000 mostly young people in Germany. MS manifests initially with relapses in about 85% of cases. In Germany, the dominating management approach is an intravenous therapy with high-dose corticosteroids. This is one of the main reasons for hospital admissions of people with MS (PwMS) in Germany.

POWER@MS2 aims to evaluate whether a web-based relapse management programme will positively change relapse management and strengthen autonomy in people with multiple sclerosis.

1.2 Objectives and Endpoints

POWER@MS2 investigates the hypothesis that the web- and evidence-based relapse management programme results in more autonomous relapse treatment decision-making by PwMS.

Clinical endpoints are evaluated by the trial statistician, while endpoints related to health economic aspects of the trial are evaluated by the Health Economics team and therefore not specified in this SAP. Primary, secondary and safety endpoints are listed in Table 1.

Table 1 Objectives and related endpoints

	Objective	**Endpoint**
Primary	To determine whether the programme leads to a decrease in corticosteroid relapse treatment and less intravenous administered corticosteroids in case of steroid treatment.	Proportion of relapses not treated or treated with oral corticosteroids during the follow-up.
Secondary	To determine whether the programme results in more autonomous relapse treatment decisions.	Differences in decision autonomy measured by CPS.
	To determine whether the programme leads to fewer relapses.	Annualised relapse rate.
	To determine whether the programme leads to an increase in risk-knowledge.	Changes in Risk Knowledge.
	To determine whether the programme leads to more empowerment.	Changes in empowerment measured by PAM from baseline to end of study.

	To determine whether the programme leads to a more autonomous steroid treatment decision-making	Changes in PBMS relapses.
Safety	Assessment of safety of the intervention	Changes in HAQUAMS. Changes in HADS. Changes in EDSS. Changes in UNDS.

1.3 Primary objective and endpoint

To determine the impact of the programme on corticosteroid treatment decision in case of a relapse, the proportion of relapses and their treatment with oral corticosteroids is assessed. Two experienced, blinded neurologists will independently rate all relapses at end of study as definite, possible or no relapses based on the information assessed in the 3-monthly telephone interviews.

1.4 Secondary objective and endpoint

Secondary endpoints are recorded as listed in Table 1. The annualised relapse rate is calculated based on the standardised assessment of relapses during the 3-monthly phone interviews. For further details, see Section 5.1.

2 Study methods

2.1 Trial design

Power@MS2 is a national, parallel group, superiority, investigator and participant blinded RCT. Participants will be randomized to the web-based programme (intervention group, IG) or a standard information on relapse management programme (control group, CG).

Participating physicians and MS centres will not be provided with any information about group assignments. Furthermore, outcome assessors are blinded. Blinding of the participants is pursued, but only to a limited extent. It cannot be prevented that patients discuss the intervention contents with their physician and might realise the group assignment.

2.2 Randomization

PwMS are randomized in a 1:1 allocation ratio to IG or CG stratified by the centre using block randomization through a computer-generated system in secuTrial®.

2.3 Sample Size

The unit of analysis for the primary endpoint is the occurrence of relapses. Eighty-one relapses per group yield in a power of 85% at two-sided significance level of 5% given proportions of 78% and 56% of orally treated or non-treated relapses in IG and CG. It is expected that this relapse rate can be observed in 170 PwMS with 1 to 2 years follow-up, corresponding to an annual rate of 0.64. The dropout rate is expected to be at 10%. Therefore, 188 PwMS will be randomized.

The blinded sample size recalculation report based on a blinded data export from 10th of September 2021 comprising events of 107 patients at that time with 51 possible and definite relapses (Asendorf, Kloidt and Friede, 2021).

In the first definition of the primary endpoint events, only definite relapses were counted. In the second definition, possible relapses were counted additionally. The first definition resulted in 20 primary endpoint events, while the second resulted in 51 primary endpoint events.

According to the estimated relapse rate per year of 0.8775 (95%-CI: [0.6632, 1.1612]), 125 PwMS need to be followed up for two years. Until 20th of October 2021, 92 patient-years were observed and 58 patient-years documented. With an extension of recruitment for further 3 months and flexible follow-up, the goal of 250 patient-years can be reached. Based on the inclusion of further 39 patients and assuming no change in drop-out rate, 294.3 patient-years can be reached, but only 277 years will be documented as documentation is performed every three months. Therefore, 146 PwMS should be sufficient to reach a power of 85%.

2.4 Framework

All endpoints are tested for superiority of the intervention programme over control.

2.5 Statistical Interim Analyses and Stopping Guidance

As no relevant adverse events are unlikely, no stopping guidelines are planned.

2.6 Timing of the Final Analysis

The final analysis will take place when all outcomes have been collected and the database is locked.

2.7 Timing of Outcome Assessments

Data will be obtained at different time points over a period of at least 12 months and at most 36 months (on average 24 months). The trial will end when 109 relapses have been documented or after 125 PwMS have reached 24 months of follow-up. The measurement time points are calculated based on the randomization date. t_{-1} takes place before enrolment and t_0 before randomization, t_1 to t_{12} are scheduled every three months. Deviations from the ideal time point are allowed to the degree of ± 21 days. For all participants, who have not reached t_{12}, the final telephone interview t_x takes place after the final participant reaches t_8. Table 2 displays the assessment of variables within these time points.

Table 2 Assessments and measurement time points

Instrument

	Screenin	Baseline	Allocatio	Post allocation												
	t_{-1}	t_0		t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_x
Month	-1	0		3	6	9	12	15	18	21	24	27	30	33	36	X
Eligibility screen	X															
Informed consent	X															
Sociodemographic data	X	X														
Randomization			X													
Relapse history				X	X	X	X	X	X	X	X	X	X	X	X	X
Relapse questions				(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
Decision autonomy and satisfaction				(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)	(X)
CPS relapse				X												
PAM and empowerment scale		X		X			X				X		X		X	X
HAQUAMS		X					X									
HADS		X					X									
PBMS relapses				X			X									
UNDS		X					X	X	X	X	X	X	X	X	X	X
Relapse risk knowledge				X			X									
Physician visit including EDSS		X					X									
EQ-5D-5L		X		X			X				X		X		X	X
Health economic parameters		X		X	X	X	X	X	X	X	X	X	X	X	X	X

(x): in case of relapse

3 Statistical Principles

3.1 Confidence intervals and p-values

All tests will be performed two-sided on a significance level of 5%. Confidence intervals (CI) will be reported with 95% confidence level.

3.2 Adherence and protocol deviations

In order to ensure patient adherence, the usage of the IG and CG programme is monitored. Participants will be contacted by phone every three months.

3.3 Analysis populations

Primary analysis is planned following the intention-to-treat principle. The intention-to-treat population contains all randomised patients. For the primary endpoint, only definite and possible relapses are considered.

4 Trial population

4.1 Screening data

Screening data will be reported and described within a CONSORT flowchart.

4.2 Eligibility

Persons aged between 18 and 65 years with suspected or diagnosed relapsing-remitting MS with ≥ 1 relapse in the last year and/or ≥ 2 relapses in the last 2 years will be included. Only PwMS having access to the internet, being fluent in German and providing signed informed consent will be included.

PwMS being allergic to steroids as well as PwMS with primary or secondary progressive MS, an acute relapse, severe cognitive deficit, severe visual impairment or severe psychiatric disorder will be excluded. Participation in the parallel study POWER@MS1 or in the former training programme on relapses will lead to exclusion.

4.3 Recruitment

Recruitment numbers will be reported and described within a CONSORT flow diagram.

4.4 Withdrawal/follow-up

Participants may leave the study at any time and may withdraw consent. Reasons for study withdrawal will be documented using a CONSORT flowchart.

4.5 Baseline patient characteristics

All baseline patient characteristics will be summarized using descriptive statistics (e.g. mean, standard deviation, median, IQR for continuous variables and frequencies (percentages) for categorical variables) and appropriate graphical methods (e.g. boxplots, histograms, barplots) depending on the data type.

5 Analysis

5.1 Outcome definitions

A full list of outcomes and their timing is described in Section 2.7.

Ten adapted items of the **Patient Activation Measure (PAM13)** are used to assess empowerment (Zill, 2013). The items of the PAM13 range from 1 (Disagree at all) to 4 (Fully agree). The raw score is the sum of the items divided by the number of answered questions (excepting non-applicable items) and multiplied with 10. Afterwards the raw score is transformed to a range from 0 to 100 (Moljord, 2015).

10-item PAM calculation expressed as formula: $100 \times \frac{\left(\left(\frac{sum\ score}{Number\ of\ answered\ questions} \times 10\right) - 10\right)}{(40-10)}$.

An adapted empowerment scale with three of the five items from Bann *et al* (2010) was used to assess empowerment. The items are answered on the 3-point scale: ‘‘no,’’ ‘‘yes, a little,’’ or ‘‘yes, a lot.’’.

Control Preference Scale (CPS) is used to assess preferred and realised role preference in steroid treatment decision-making. It consists of five roles ranging from A (the individual making the treatment decisions), over C (the individual making the decisions jointly with the physician) to E (the physician making the decisions). Satisfaction with the decision will be measured in a telephone interview within 3 months in case a decision for or against steroid treatment has been made.

Hamburg Quality of Life in MS Scale Version 10.0 (HAQUAMS) is a questionnaire for measuring quality of life in PwMS (Gold *et al.*, 2001). It consists of 44 items with mostly 5-point-likert scale. Twenty-eight of the items are subdivided into the six subscales:

- Fatigue: Items 6, 7, 8, 9
- Cognition: Items 10, 11, 12,13
- Lower extremity: Items 17,18,19,20
- Upper extremity: Items 21, 22, 23, 24, 25
- Communication: Items 29, 30, 31, 32, 33, 34
- Mood: Items 36, 37, 38, 39, 40.

Mean subscale scores are calculated, the total score is built by calculating the mean of the subscale score means: *HAQUAMS total score = mean(mean(upper extremity), mean(lower extremity), mean(fatigue), mean(cognition), mean(communication), mean(mood))*. Subscale and total score range from one to five; high scores indicate low quality of life. For missing items, mean substitution is allowed.

Planned Behaviour in MS Scale (PBMS relapses) is adapted to steroid decision-making and is used to measure behaviour strategies in case of relapses (Kasper, 2012). The 18-item questionnaire consists of six sub-domains. Items are measured on a four-step Likert-scale format ranging from “strongly disagree” (0) to “strongly agree” (3). Eight items are negatively framed and their polarity is reversed (Items: 5, 8, 9, 10, 13, 16, 17, 18). Three items each are allocated into the six sub-domains as follows:

1a) Expectations regarding the effects of immunotherapy: Items 1, 7, 13

1b) Importance of the expected effects: Items 4, 10, 16

2a) Assumptions regarding the expectations of relevant social partners: Items 2, 8, 14

2b) Importance of the social norm: Items 5, 11, 17

3a) Assumptions regarding the control belief: Items 3, 9, 15

3b) Importance of the control belief: Items 6, 12, 18

Scale means are calculated for the sub-domains, whereby missing values are interpolated if necessary. The three main domain scores are obtained by the multiplication of the corresponding two sub-domains, e.g. 1a) x 1b). The total score is the mean of the three products. A higher score indicates a more critical attitude on short-term high-dose corticosteroids during relapses.

Expanded Disability Status Scale (EDSS) is determined by the treating neurologist in order to assess impairment (Kurtzke, 1983). It ranges from 0 to 10, with higher scores indicating greater disability.

United Kingdom Neurological Disability Scale (UNDS) will be used to assess impairment (Heesen, 2007). It consists of the 12 domains *cognitive disorders*, *mood*, *vision*, *communication*, *swallowing, upper extremity, lower extremity, bladder, intestine, fatigue, sexuality* and *pain* with sub-scores ranging from 0 to 5. The score, the sum of the 12 sub-scores, ranges from 0 to 60, with higher scores indicating greater disability. If the *sexuality score* is unknown, it can be imputed by the mean of *lower limb*, *bladder* and *intestine score* rounded to the nearest integer.

Hospital Anxiety and Depression Scale (HADS) is a self-reported questionnaire and is used as a measure for depression and anxiety (Zigmond, 1983). It consists of two 7-item subscales, one for anxiety and one for depression, with a maximum score of 21 for each subscale (each item is scoring from 0 to 3), with higher scores representing higher levels of anxiety or depression. A maximum of one missing item per subscale can be estimated by the rounded mean of the six existing items of the same subscale.

Relapse risk knowledge is used to assess risk knowledge and has been adapted to reflect current evidence (Köpke, 2009). The questionnaire consists of nine multiple-choice questions and one free-response question. The score is obtained by the number of correct answers and ranges from 0 to 10, with higher scores indicating better knowledge.

Hamburg Relapse Assessment Scale (HARAS) is used to assess relapses in more detail (Köpke, 2011). 24 items of the questionnaire are used and each item is scored on a 5-point Likert scale, with higher scores indicating greater risk of relapse. The total score is calculated by summing the item scores.

5.2 Analysis methods

In the analysis of primary and secondary endpoints, centres with less than 10 patients are pooled into one centre.

The primary endpoint is analysed using a generalised linear model with mixed effects and logit link function. Subject-specific effects are modelled as random effects, while intervention group and study centre are included as fixed effects. In case of nonconvergence due to a flat likelihood in the variance of the subject-specific effects near 0, the model will be simplified by removing the random subject-specific effects resulting in a generalized linear model (logistic regression). In this case correction of standard errors by using generalized estimating equations (GEE) will be explored. The intervention effect is reported as odds ratio (OR) with 95% CI and p-value testing the null hypothesis of no intervention effect ($H_0: OR = 1$).

```
library(MASS)

prim.mod <- glmer(oral_treated ~ intervention + site + (1|Patient.Id),
                  data = primEnd, family = binomial(link = "logit"))
summary(prim.mod)
emmeans(prim.mod ~ intervention, type = "response")
```

Longitudinal assessments of the quality of life and impairment are analysed employing Gaussian linear models for repeated measures with intervention group, time, intervention-by-time interaction and study centre as factors and, if the instrument was assessed at baseline, its baseline score as a covariate (mixed model for repeated measurements, MMRM).

The error terms are assumed to follow a multivariate normal distribution with unstructured covariance matrix. Least squares mean changes from baseline will be reported for both groups with 95% CI as well as the difference between the least-squares intervention group means with 95% CI and p-value testing the null hypothesis of no treatment effect ($H_0: \delta = 0$).

```
library(lme4)
library(car)
library(emmeans)

var_sec <- lmer(var ~ baseline_var + intervention + time + intervention * time +
                  site + (1|Patient.Id), data = secEndpoints)
summary(var_sec)
res <- Anova(var_sec, type = "lll", test.statistic = "F")
res
# report least square group differences
tmt.means <- emmeans(var_sec, ~ intervention * time)
contrast(tmt.means)
pairs(tmt.means)
```

Safety endpoints that are raised at baseline and at another visit will be analysed using group mean comparisons between the intervention groups (δ) adjusted for baseline assessments and centre in analysis of covariance (ANCOVA) models. Least square group differences will be reported with 95% CIs and p-values testing the null hypothesis of no intervention effect ($H_0: \delta = 0$).

```
library(emmeans)


mod.sec <- lm(var ~ intervention + var_baseline + site,
                        data = secEndpoints)
# report least square group differences
tmt.means <- emmeans(mod.sec, ~ intervention)
tmt.means
pairs(tmt.means)
```

For the Control Preference Scale, ordinal regression with logit link will be used with intervention and study centre as covariates. The model assumptions will be checked.

```
library(ordinal)
library(emmeans)

# Ordinal regression for the Control Preference Scale
mod.sec <- clm(cps ~ intervention + site, link = 'logit',
                  data = secEndpoints)
summary(mod.sec)

emmeans(mod.sec, ~ cps | intervention, mode = 'prob')

# Check model assumptions
nominal_test(mod.sec) # test partial proportional odds assumption
scale_test(mod.sec) # test for signs of scale effects
```

5.3 Missing data

In case of missing data, all PwMS will be analysed following the intention-to-treat principle. To handle missing data in baseline variables or follow-up assessments, multiple imputation models will be applied. For the knowledge questionnaires, we will follow questionnaire specific guidance.

5.4 Additional analyses

Effects of age, gender, level of education, centre and level of impairment will be explored in subgroup and regression analyses.

5.5 Harms

Safety measures are applied to control for anxiety, depression and disease-specific quality of life.

5.6 Statistical software

All analyses will be performed in the current version of R or alternatively in SAS (version 9.4 or higher). All R packages are used in their current version and will be reported in the statistical report. For data visualization, *ggplot2* is used. Multiple imputation is conducted using the package *mice*.

5.7 References

1. Rahn, A.C., Wenzel, L., Icks, A. et al. Development and evaluation of an interactive web-based decision-making programme on relapse management for people with multiple sclerosis (POWER@MS2)—study protocol for a randomised controlled trial. Trials 22, 139 (2021). https://doi.org/10.1186/s13063-021-05059-1
2. Asendorf T, Kloidt AM and Friede T, Power@MS2 – Verblindeter Review der Fallzahlplanungsannahmen, 2021
3. Bann CM, Sirois FM, Walsh EG. Provider support in complementary and alternative medicine: exploring the role of patient empowerment. J Altern Complement Med. 2010;16(7):745–52
4. EuroQol Research Foundation. EQ-5D-5L User Guide, 2019. Available from: https://euroqol.org/publications/user-guides.
5. Gold SM, Heesen C, Schulz H, et al. Disease specific quality of life instruments in multiple sclerosis: validation of the Hamburg quality of life questionnaire in multiple sclerosis (HAQUAMS). Mult Scler 2001;7:119–30
6. Heesen C, Reich C, Borchers C, Gold S, Schulz K. Patient based scoring instruments in multiple sclerosis –German validation of the United Kingdom Disability Scale and the Patient Generated Index. Neurol Rehab. 2007;13:24 –36.
7. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727 –36.
8. Kasper J, Köpke S, Fischer K, Schäffler N, Backhus I, Solari A, et al. Applying the theory of planned behaviour to multiple sclerosis patients ' decisions on disease modifying therapy--questionnaire concept and validation. BMC Med Inform Decis Mak. 2012;12:60
9. Köpke S, Kasper J, Mühlhauser I, Nubling M, Heesen C. Patient education program to enhance decision autonomy in multiple sclerosis relapse management: a randomized-controlled trial. Mult Scler. 2009;15(1):96–104.
10. Köpke S, Richter T, Kasper J, Mühlhauser I, Flachenecker P, Heesen C. Implementation of a patient education program on multiple sclerosis relapse management. Patient Education and Counseling. 2011.
11. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33(11):1444 –52.
12. Moljord, I. E. O., Lara-Cabrera, M. L., Perestelo-Pérez, L., Rivero-Santana, A., Eriksen, L., & Linaker, O. M. (2015). Psychometric properties of the Patient Activation Measure-13 among out-patients waiting for mental health treatment: A validation study in Norway. In Patient Education and Counseling (Vol. 98, Issue 11, pp. 1410–1417). Elsevier BV.
13. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67:361–70
14. Zill JM, Dwinger S, Kriston L, Rohenkohl A, Harter M, Dirmaier J. Psychometric evaluation of the German version of the Patient Activation Measure (PAM13). BMC Public Health. 2013;13:1027.