STATISTICAL ANALYSIS PLAN

FOR NEuRoMS Work Package 3

Date: 18/07/2023

Version: 1

North Wales Organisation for Randomised Trials in Health (NWORTH)

Bangor University, College of Health & Behavioural Sciences,

Institute of Medical & Social Care Research, Y Wern, Normal Site, Holyhead Road, Bangor, Gwynedd LL57 2PZ

Telephone: 01248 388095 Email: nworth@bangor.ac.uk Website: http://nworth-ctu.bangor.ac.uk/

Statistical analysis plan (SAP) for the Neuropsychological Evaluation and Rehabilitation in Multiple Sclerosis (NEuRoMS): Multicentre Feasibility Randomised Controlled Trial and Fidelity Evaluation (Phase 2: Work Package 3 Feasibility Study)

Funding Source:

This study/project is funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research Programme (project reference RP-PG--0218-20002). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

IRAS number: 287115

Funder ref: NIHR Programme Grants for Applied research (ref no. RP-PG-0218-20002)

This document has been written based on information contained in the study protocol version 9, dated 05/05/2023.

Chief Investigator:	Professor Roshan das Nair
Statistical co-applicant	Dr Zoë Hoare
Programme Manager	Dr Gogem Topcu
Data Manager	Dr Nia Goulden
Trial Statistician	Dr Nia Goulden

Nia Goulden

Author's Name **Date and Signature**

Dr Zoë Hoare

Second Statistical Advisor **Date and Signature**

Professor Roshan das Nair

Chief Investigator **Date and Signature**

Document History

Updated version no.	Effective date	Authorship	Section changed	Summary of changes
1		N. Goulden	New	

Acronyms and definition of terms

Acronym	Meaning
ANOVA	Analysis of Variance
AP	Assistant Psychologist
CONSORT	Consolidation Standards of Reporting Trials
CRN	Clinical Research Network
CTU	Clinical Trials Unit
DMEC	Data Monitoring and Ethics Committee
GAD	Generalized Anxiety Disorder
ICECAP-A	ICEpop CAPability measure for Adults
MCID	Minimum Clinically Important Difference
MS	Multiple Sclerosis
MSIF	Multiple Sclerosis International Federation
MSIS	Multiple Sclerosis Impact Scale
MSSE	Multiple Sclerosis Self-Efficacy Scale
MSWDQ	Multiple Sclerosis Work Difficulties Questionnaire
NEADL	Nottingham Extended Activities of Daily Living
NHS	National Health Service
NEuRoMS	Neuropsychological Evaluation and Rehabilitation in Multiple Sclerosis
NWORTH	North Wales Organisation for Randomised Trials in Health
OT	Occupational Therapist
PDQ	Perceived Deficits Questionnaire
PHQ	Patient Health Questionnaire
RCT	Randomised Controlled Trial
SAP	Statistical Analysis Plan
SST	Symbol Substitution Task
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
WCT	Word Colour Task

Table of Contents

1. Statistical Analysis Plan Authorship

The analysis plan has been authored by Dr Nia Goulden, Trial Statistician. There has also been input from Dr Zoë Hoare (Principal Statistician), Dr Gogem Topcu (Programme Manager), Dr Jacqueline Mhizha-Murira (Research Fellow) and Professor Roshan das Nair (Chief Investigator). The draft plan will be circulated to the Programme Management Group, in particular Denise Kendrick, Shirley Thomas and Deborah Fitzsimmons, and Programme Steering Committee for comments before being signed off. All of the statistical analysis will be completed by Dr Nia Goulden and overseen by Dr Zoë Hoare.

2. Introduction

2.1 Background and Rationale

Currently, staff at NHS Multiple Sclerosis (MS) clinics do not routinely screen and provide sufficient support for all patients with MS who present with cognitive problems (Croft et al., 2016; MS Trust, 2015; Mynors et al., 2016; Roberts et al., 2016). By not intervening early, disability accrues, and people with MS are less likely to benefit from rehabilitation (Giovannoni et al., 2016), and costs for the people with MS, their families, the NHS and society are likely to escalate.

Previous NEuRoMS work packages have developed a screening tool to identify the level of cognitive impairment in patients with MS. As part of a new screening and management pathway, all patients will complete cognitive screening as part of newly introduced routine clinical care. Cognitive problems from the screening will be categorised as 'within normal range', 'mild cognitive problems', 'moderate cognitive problems' or 'severe cognitive problems'. Those with mild or moderate cognitive problems will be offered a manualised NEuRoMS intervention led by an Assistant Psychologist (AP)/Research Nurse/Assistant Occupational Therapist (OT). The intervention will be tailored to screening profile and individual needs, but will focus on: psychoeducation, internal and external compensatory strategies; environmental modifications; and the importance of dealing with low mood and fatigue.

NEuRoMS (Neuropsychological Evaluation and Rehabilitation in Multiple Sclerosis) is a 6-year programme of research that develops over five work packages in which people with MS have routine cognitive screening assessments conducted online and those with mild or moderate cognitive problems are offered cognitive rehabilitation.

This document details the statistical analysis plan for Work Package 3. In this work package, we will conduct a feasibility trial to determine the feasibility of conducting a definitive RCT for evaluating the new screening and management pathway.

Part 1 of the feasibility trial will test the cognitive screening pathway, part 2 will assess the acceptability and feasibility of a definitive Randomised Controlled Trial (RCT) and conduct a fidelity evaluation, and part 3 will conduct qualitative interviews to determine feasibility and acceptability of a definitive RCT. Since part 3 will not be analysed with quantitative analysis this part of the feasibility trial is not described further in this statistical analysis plan.

2.2 Aim

2.3 Trial Objectives

The primary objective is to assess the feasibility of conducting a definitive RCT to investigate the clinical and cost-effectiveness of the NEuRoMS intervention in reducing the impact of cognitive problems in daily life amongst people with MS, and the acceptability of the intervention.

Secondary objectives which will be evaluated using quantitative statistical analysis are:

<u>Part 1 – Testing cognitive screening pathway</u>

1. Assess the frequency of 'within normal range', 'mild cognitive deficits', 'moderate cognitive deficits' and 'severe cognitive deficits' and thus the size of the target population (potentially eligible participants for a future definitive RCT) based on Symbol Substitution Task (SST) and/or Word Colour Task (WCT). Evaluated using data on usage and participants scores on the cognitive screening measures – see section 5.3 for methods.

<u>Part 2 – Acceptability, Feasibility RCT and fidelity evaluation:</u>

1. Identify the necessary parameters and methods to undertake a clinical and cost-effectiveness analysis in a future definitive trial. This will be achieved by collecting data for the feasibility trial and assessing these measures and outcomes. Methods and data described in sections 5.4, 5.5 and 5.6 using data on feasibility of trial procedures.
2. Assess acceptability of data collection tools, processes, data completeness and follow-up rates, and determine suitability of outcome measures. This will be achieved by collecting data for the feasibility trial and assessing these outcomes. Methods described in sections 5.4, 5.5 and 5.6 using data on feasibility of trial procedures, completion rates of outcome measures and patient preferences for completion.

The health economic analysis will be presented within a separate, complementary health economic analysis plan (HEAP).

2.4 Trial Design

Part 1 involves an observational study of those who receive screening and support for cognitive problems, using routinely collected clinical data. In Part 1, cognitive screening, a new clinical procedure introduced as part of the NEuRoMS cognitive screening and management pathway, will be incorporated within three MS clinics as part of routine clinical care.

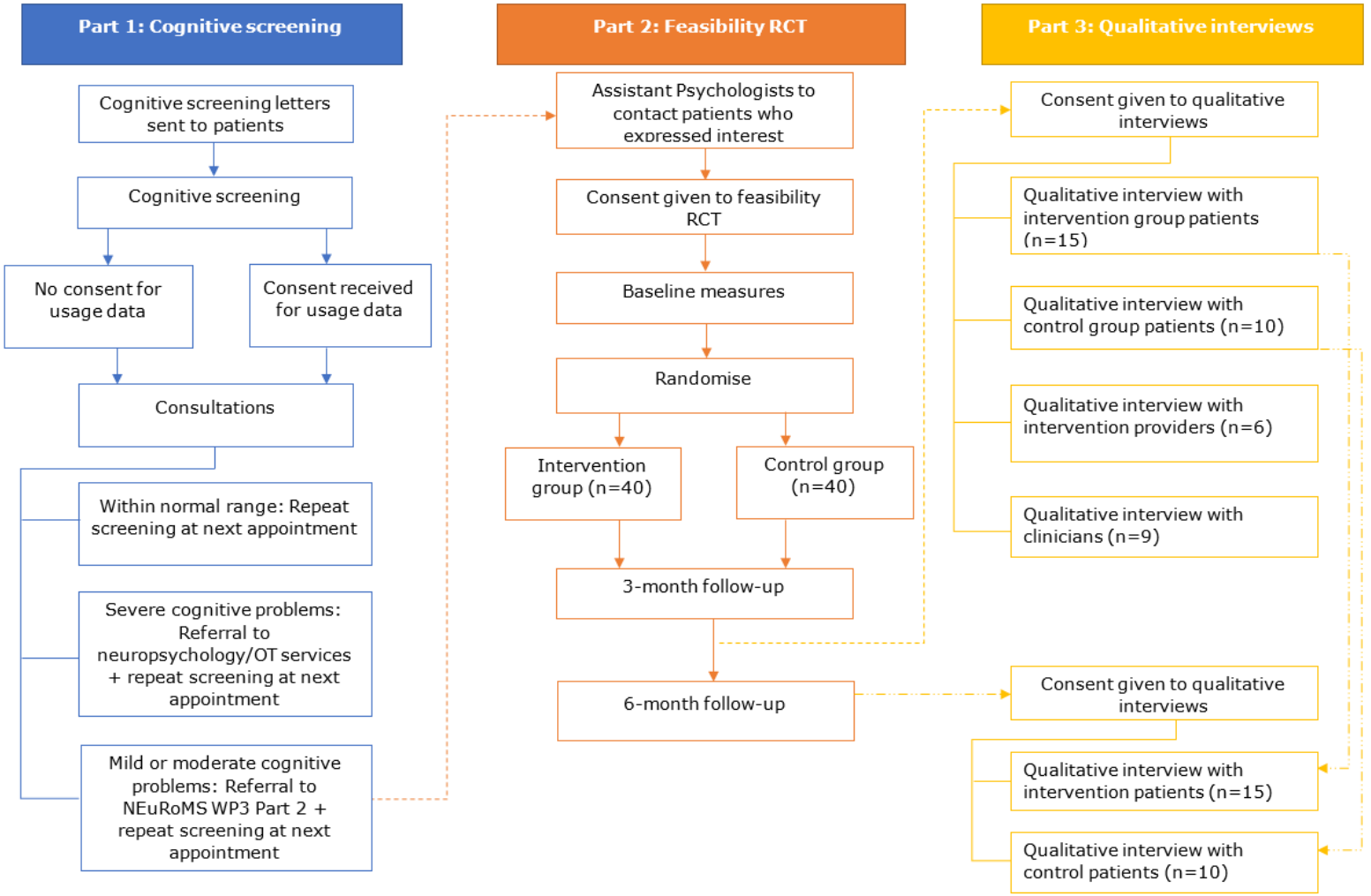
In Part 2, a parallel group, feasibility, multi-centre RCT with nested fidelity evaluation will evaluate the feasibility of undertaking a definitive trial comparing NEuRoMS intervention programme plus usual care to usual care only, amongst people with MS with mild and moderate cognitive problems (Figure 1).

2.5 NEuRoMS Intervention

The NEuRoMS intervention is multi-faceted, involving various components (i.e., information provision, goal setting) and a range of strategies and techniques (e.g., psychoeducation, compensatory strategies, boosting cognitive reserve). The intervention is person-centred, tailored to the needs and lifestyle of each participant, and aims to help people with MS cope with and manage cognitive problems by establishing strategies that can be maintained once the intervention sessions are finished.

The intervention will be delivered by a trained therapist (AP, Research Nurse, or Assistant OT), under the supervision of a clinical psychologist or OT. Face-to-face (dependent on Government and NHS COVID-19 advice), videoconferencing and telephone delivery options will be available. The intervention provider will attend training and receive monthly supervision via telephone/videoconferencing with a clinical psychologist.

Figure 1: Design of the feasibility trial



2.6 Study Population

Inclusion Criteria

All individuals: able and willing to give consent and able to communicate in English.

Part 1 – Testing cognitive screening pathway:

People with MS:

- Diagnosis of MS
- Aged 18 years or above

Part 2 – Acceptability, Feasibility RCT and fidelity evaluation:

People with MS:

- Diagnosis of MS
- Received cognitive screening and mild or moderate cognitive problems identified (Part 1)
- Aged 18 years or above

Exclusion Criteria

All participants:

Do not have mental capacity to consent to take part in the study

Part 2 participants only:

- Currently receiving neuropsychological intervention for cognitive problems
- Received NEuRoMS intervention during WP2ii

3. Statistical Principles

3.1 Sample Size Justification

Based on our experiences in an earlier work package (WP2) and current clinic throughput, we estimate that ~1405 patients will receive cognitive screening as part of their newly introduced clinical care. We will continue to screen until at least 40-50% of patients fully complete both SST and WCT.

Sample size for feasibility RCTs range between 24-50 (Hooper, 2019). We believe that 60-80 participants will enable us to optimally address the aims of this study and provide us with parameter estimates to confirm our sample size calculations for the definitive trial. Recruitment for this stage will stop once we have randomised up to 80 people, which does not account for attrition. We believe we will be able to recruit this

number of participants in the timeframe April 2022 – January 2023 based on our recruitment figures from previous MS trials and patient throughput in clinics (average completion of screening is ~50 patients per month per clinic; based on what we have learnt during the previous work package).

The sample size calculation for the full RCT, Work Package 4, is based on the Multiple Sclerosis Impact Scale, Psychological Subscale (MSIS-Psych). The sample size calculation is based on a mean difference of 3 points between control and intervention groups, and a standard deviation of 9. This mean difference and standard deviation, and a sample size of 80 (40 in control group and 40 in intervention group), have been used to calculate the precision of the sample using the prec_meandiff command in R. The width of the 95% confidence interval is 8.0, with a lower bound of -1.0 and an upper bound of 7.0.

3.2 Randomisation

Once the participant has been recruited, consented and completed the baseline questionnaires, they will be individually randomised to control or intervention groups (ratio 1:1 stratified by NHS site), using an online dynamic adaptive algorithm (Russell et al., 2011), developed and maintained by NWORTH CTU.

3.3 Levels of Confidence

All confidence intervals presented will be 95% and two-sided.

3.4 Protocol Violations and Deviations

The definition of a protocol violation is an intended failure to adhere to the protocol such as incorrect data being collected and documented. A protocol deviation is an unintended failure to adhere to the protocol; examples include errors in applying inclusion/exclusion criteria or missed follow-up visits due to error. A table containing any protocol violations or deviations will be summarised within the final analysis.

3.5 Missing Data

Data entered directly into REDCap (the electronic data storage tool) will have no missing data as settings will be put in place so that participants must provide an answer. However, some questionnaires will be completed on paper and returned to the team, which may have missing data. Participants will be contacted by a blinded member of the research team to complete any missing data on questionnaires.

For this feasibility study there will be no imputation of missing data. Descriptive statistics will be produced to describe the level of missing data, and this will be used as an indicator of the appropriateness of any measures for a definitive future RCT.

3.6 Outliers

Data will be checked for outliers by plotting the distribution of the data and with box and whisker plots. Outliers identified from the statistical analyses will be examined by rechecking the data. No outliers will be discarded if they can be verified or are within range. If any outliers are dropped from the dataset, it will be reported and full reasoning given.

4. Data

For full details on the data collection, flow and storage please refer to the current version of the NEuRoMS WP3 Data Management Plan.

4.1 Part 1 Screening Data

The scores from the SST and/or WCT at screening will assess the participant's level of cognitive impairment.

4.2 Part 2 Feasibility Outcomes

Feasibility metrics to determine whether a definitive RCT will be feasible are:

1. Recruitment and retention;
2. Acceptance, adherence to and fidelity of the intervention.

Progression criteria are specified in section 5.4.

4.3 Part 2 Proposed Clinical Outcomes

The following measures will also be used to capture information about the patient at baseline and to assess outcomes at 3- and 6-months after randomisation, and analysed according to the principles outlined in this statistical analysis plan:

- Cognitive impairment (Perceived Deficits Questionnaire [PDQ-20 ; (Sullivan et al., 1990)]) ;
- Quality of life (Multiple Sclerosis Impact Scale [MSIS-29; (Hobart et al., 2001)]);
- Mood (Patient Health Questionnaire-9 [PHQ-9; (Kroenke et al., 2001)]; Generalized Anxiety Disorder-7 [GAD-7; (Spitzer et al., 2006)]; Whooley Questions for depression screening (Whooley et al., 1997));
- Function (Nottingham Extended Activities of Daily Living Scale [NEADL; (Nicholl et al., 2002)]);
- Self-efficacy (Multiple Sclerosis Self-efficacy Scale [MSSE; (Rigby et al., 2003)]);

- Work-related issues (Multiple Sclerosis Work Difficulties Questionnaire short form [MSWDQ; (Honan et al., 2014)]);
- Two single-item questions asking to what extent work and medication adherence has been impacted by cognitive problems. The single-item Work question has been added here to determine whether the single-item question can be used in the definitive trial instead of the MS Work Difficulties Questionnaire. This may make the questionnaire set smaller for the definitive trial.

4.4 Definitions and Calculations of Outcome Measures

The data collected as part of the feasibility will include the measures listed in section 4.3. See the table in appendix 1 for full details.

4.5 Unblinding

The final unblinding for results will take place after all blinded analysis, as stipulated in this plan, have been completed. The unblinding form (found in the Appendix of SOP 5.03 Randomisation systems) will be completed by the trial statistician and handed to the NWORTH IT team who will then provide the group details. The group allocations will be revealed at a results meeting which may include members of the PMG and PSC.

5. Statistical Analyses

NEuRoMS WP3 is a feasibility trial. The analyses described will therefore be exploratory, and are not intended to be definitive. The aim of conducting these analyses is to guide and refine the analysis for the definitive trial, Work Package 4.

5.1 Analysis Time Frame

TASK	EXPECTED DATE
First participant recruited	March 2022
Final participant recruited	January 2023
Final follow up completed	July 2023
Data cleaning completed	July 2023
Analysis completed	August 2023

5.2 Recruitment and Retention

The analysis will consider the items from the CONSORT checklist for randomised pilot and feasibility trials (Eldridge et al., 2016) to ensure that all topics are being covered. Values for eligibility rates, recruitment rates, attrition rates and withdrawal rates will be reported using the flow data collected within the study. This will be evaluated overall and per group.

Furthermore, details on reasons for ineligibility and non-recruitment will be reported within a table along with their related patient frequencies and percentages. Information on withdrawals and non-respondents will be presented including reasons where applicable and time points during the trial.

5.3 Part 1 Analysis

The levels of cognitive impairment based on the scores from the SST and/or WCT will be defined as follows:

1. Within normal range: Scores higher than 1.5 standard deviations below the mean;
2. Mild cognitive deficits: Scores of 1.5 standard deviations below the mean or lower, and higher than 2.5 standard deviations below the mean;
3. Moderate cognitive deficits: Scores of 2.5 standard deviations below the mean or lower, and higher than 3 standard deviations below the mean;
4. Severe cognitive deficits: Scores of 3 standard deviations below the mean or lower.

The proportion of participants with no, mild, moderate and severe cognitive impairment from the screening battery will be calculated.

The SST task is 90 seconds long. For people with MS who complete the cognitive screening using the WP3 screening links the total number of items, correct items and incorrect items at 30 seconds and 60 seconds will also be collected. For people with MS who complete the cognitive screening using the WP3 screening links the WCT task will be completed. The WCT has three conditions (i.e., word identification, colour identification and incongruent conditions). Each condition is 90 seconds long, and the total number of items, correct items and incorrect items at 15, 30, 45, 60 and 90 seconds will be collected. We will also collect the time taken to answer the first 50 items, as well as the number of correct items and number of incorrect items. A scoring method adapted from the Golden scoring method (Golden, 1978; Golden & Freshwater, 2002) will be used for the WCT data.

Demographic influences on screening data

The screening includes the SST, the WCT, and a set of brief questionnaires assessing mood, fatigue and self-reported cognitive function from the Multiple Sclerosis Quality of Life Inventory (Ritvo et al., n.d.): the Mental Health Inventory 5 items, Modified Fatigue Impact Scale 5 items and Perceived Deficits Questionnaire 5 items. Separate regression analyses will be performed with the score for each task as dependent variables, together with age as a covariate and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) as factors. Note that for this analysis the SST score at 30, 60 and 90 seconds will be used, and the WCT scoring method adapted from the Golden scoring method (Golden, 1978; Golden & Freshwater, 2002) for the WCT at 15, 30, 45, 60 and 90 seconds.

Investigating whether the SST and WCT can be shorter

For the SST, regression analyses will be applied to determine whether the scores at 30 seconds and/or 60 seconds are predictive of the score at 90 seconds. The purpose of this is to determine whether the task could be shorter in a definitive trial. Separate regression analyses will be conducted for data at 30 seconds and 60 seconds. The dependent variable will be the score at 90 seconds, with covariates including the scores at 30 seconds/60 seconds. Age will be included as a covariate, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) as factors in the analysis model. Separate correlation analyses will also be applied to the scores at 90 seconds with the scores at 30 seconds and 60 seconds. In order to consider using a shorter task the regression coefficient for the shorter time should have a p-value less than 0.05 and the correlation coefficient should have a p-value less than 0.05.

For the WCT data scored at different timepoints, separate regression analyses will be conducted for data at 15, 30, 45, 60 seconds for each of the different conditions of the task. The dependent variable will be the score at 90 seconds, with covariates including the scores at 15/30/45/60 seconds. Age will be included as a covariate, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) as factors in the analysis model. Separate correlation analyses will also be applied to the scores at 90 seconds with the scores at 15, 30, 45 and 60 seconds. In order to consider using a shorter task the regression coefficient for the shorter time should have a p-value less than 0.05 and the correlation coefficient should have a p-value less than 0.05.

Association between SST and WCT

In order to determine the association between the SST and WCT scores, a correlation will be computed between the SST and WCT scores at 90 seconds.

Influence of mood and fatigue on the SST and WCT

A partial correlation analysis will be computed to determine variability of the SST and WCT at 90 seconds explained by mood (MHI) and fatigue (MFIS). Separate partial correlations will be computed for the SST and WCT, controlling for MHI and MFIS.

5.4 Feasibility Outcomes

The outcome measures relating to recruitment, retention and adherence/fidelity will be assessed using Red/Amber/Green (RAG) criteria, as defined below:

1. Confirmation of adequate recruitment for a definitive trial: Go/Green: Average of 12 participants a month or more recruited; Review/Amber: Average of 6-11 participants a month recruited; Stop/Red: Average of five participants or less a month recruited;
2. Confirmation of adequate retention for the definitive trial, the number of participants who complete the six month follow up: Go/Green: 80% or more participants retained; Review/Amber: 50-79% participants retained; Stop/Red: 49% or less of participants retained;
3. Suitability of outcome measures will be determined by the level of completeness. This will include assessing how many participants have fully completed the outcome measure. In addition it will be necessary to consider the number of participants who have missing data for an outcome measure, but it is possible to compute a score using missing value rules. Potential key outcome measures (such as MSIS-Psychological subscale) for the definitive RCT, will be deemed appropriate if minimum success criteria are achieved, or if we can identify solutions to overcome any identified issue: These criteria are Go/Green: ≥ 80%; Review/Amber: 50-79%; Stop/Red: ≤ 49% for completion rates of these potential key outcome measures.

Progression will be possible with some amber and red results if it is possible to demonstrate how the difficulty could be addressed and rectified e.g. if recruitment was too low each month, how the recruitment rate will be increased.
For point 3., red and amber results would not prevent progression to a definitive trial but would indicate that the outcome measure is not suitable for use in a definitive trial. Use of the outcome measure in a definitive trial would depend on the ability to demonstrate a plan for improving completion of the outcome measure.

5.5 Descriptive Statistics

Descriptive statistics of the data will be presented. All continuous measures will be reported with mean values and standard deviations provided that data are normally distributed, otherwise the median and interquartile range will be used. Categorical variables reported with counts and related percentages.

5.6 Outcome Measures

In order to obtain effect sizes for a definitive RCT, and determine the suitability of analysis models, the analysis models that would be used in a definitive RCT will need to be tested. All analysis will be guided by the principle of intention to treat.

Baseline and Outcome measures

The same measures will be collected at baseline, 3-month follow up and 6-month follow up.

A repeated measures analysis of variance (ANOVA) will be used to assess whether there is a change between time points and whether there is a difference between control and intervention groups. The model will include the allocated group (Control or NEuRoMS intervention), cognitive impairment group (Mild or Moderate) and the site (Nottingham, Cardiff, London) as factors, as well as age as a covariate, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) and type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) as factors. Effect sizes will be estimated using adjusted mean differences from the model and standard errors.

A general linear model will be fitted for the data at 3-month follow up adjusted for baseline score, allocation group and stratification variable (site). This will be repeated with the 6-month follow up data. Effect sizes will be estimated using adjusted mean differences from the model and standard errors.

The repeated measures ANOVA will be used to assess the change over all time points, and the general linear model will assess the change from baseline to the each follow up time point, to be able to focus on the change to the primary time point.

Comparison of a single work/education question and the MSWDQ

We will evaluate the use of a single-item question instead of the full MSWDQ. Correlation analysis will be computed between the total score of the MSWDQ and the

single-item question. A regression analysis will be applied with the single work/education question as a dependent variable and the MSWDQ score as a covariate, together with age and time since diagnosis as covariates, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) and type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) as factors. In order to consider replacing the MSWDQ with the single question, the correlation should have a p-value less than 0.05 and the regression coefficient for the MSWDQ should have a p-value less than 0.05. This decision will be made together with clinical judgement regarding the appropriateness of replacing the single item question with the MSWDQ.

Comparison of PHQ-2 and PHQ-9

We will evaluate the use of PHQ-2 instead of the PHQ-9. Correlation analysis will be computed between the total score of the PHQ-9 and the total score of the PHQ-2. A regression analysis will be applied with the PHQ-9 score as a dependent variable and the PHQ-9 score as a covariate, together with age and time since diagnosis as covariates, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) and type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) as factors. In order to consider replacing the PHQ-9 with the PHQ-2, the correlation should have a p-value less than 0.05 and the regression coefficient for the PHQ-9 should have a p-value less than 0.05. This decision will be made together with clinical judgement regarding the appropriateness of replacing the PHQ-9 with the PHQ-2.

Comparison of GAD-2 and GAD-7

We will evaluate the use of GAD-2 instead of the GAD-7. Correlation analysis will be computed between the total score of the GAD-7 and the total score of the GAD-2. A regression analysis will be applied with the GAD-2 score as a dependent variable and the GAD-7 score as a covariate, together with age and time since diagnosis as covariates, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) and type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) as factors. In order to consider replacing the GAD-7 with the GAD-2, the correlation should have a p-value less than 0.05 and the regression coefficient for the GAD-7 should have a p-value less than 0.05. This decision will be made together with

clinical judgement regarding the appropriateness of replacing the GAD-7 with the GAD-2.

Comparison of Whooley questions and the PHQ-9

We will evaluate the use of Whooley questions instead of the full PHQ-9. Correlation analysis will be computed between the total score of the GAD-7 and the total score of the GAD-2. A regression analysis will be applied with the total Whooley score as a dependent variable and the PHQ-9 score as a covariate, together with age and time since diagnosis as covariates, and gender (Woman, Man, Other, Prefer not to say) and level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known) and type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) as factors. In order to consider replacing the PHQ-9 with the Whooley questions, the correlation should have a p-value less than 0.05 and the regression coefficient for the PHQ-9 should have a p-value less than 0.05. This decision will be made together with clinical judgement regarding the appropriateness of replacing the PHQ-9 with the Whooley questions.

Impact of objective and subjective measures of cognition, mood and fatigue on work

Work difficulties are being measured by the MSWDQ and the single item work/education question. For the total score of the MSWDQ and the single item question, separately, regression analysis will be conducted with PDQ20 score (subjective cognition), SST score (objective cognition), PHQ-9 and GAD-7 scores (mood) and MSIS-29 question 23 (fatigue) as covariates.

5.7 Check of Assumptions

A check of the assumptions required of the statistical tests will be performed.

Regression Analysis

For the regression we will check that the residuals from the model are normally distributed. A scatterplot will be produced of the standardised residuals against the predicted values to test for homoscedasticity. There will also be a check that there is a linear relationship between the dependent variable and each of the independent variables, and that there is no multicollinearity by checking the variance inflation factor of the model.

Repeated Measures ANOVA

For the repeated measures ANOVA, the dependent variable should be approximately normally distributed at each level it is measured. We will also check for sphericity, that the variance of the dependent variable is equal across all levels. In the event that these assumptions are violated we will consider transformation of the variable or use of an appropriate non-parametric test such as the Friedman test.

General Linear Model

For general linear model tests we will check that the residuals of the models are approximately normally distributed and that there is homogeneity of variance. A scatterplot will be produced of the standardised residuals against the predicted values to test for homoscedasticity. In the event of these assumptions being violated we will consider transforming the data in order to satisfy the assumptions of the test. In the event of the residuals being skewed to the right, or the variance increasing with an increasing independent variable, a transformation using a lower power, such as square root, cube root or a log transformation will be used. In the event of the residuals being skewed to the left, or variance decreasing with increasing independent variable, a transformation using a higher power, such as square root or cube root transformation will be used. This should also address any violations of non-normality of data or homogeneity of variance.

We will need to produce scatterplots of the covariates against the dependent variables for each level of the independent variables. This should show a linear relationship between the covariate and dependent variable. The lines of best fit should be parallel, so that there is no interaction between the covariate and the independent variable. In the event of this assumption being violated it will be necessary to add the interaction term of the covariate and dependent variable to the general linear model.

In the event of the assumptions of general linear model not being met we will also consider use of a generalised linear model with appropriate distribution and link function.

6. Software

All quantitative analysis will be completed using SPSS, Stata and R.

7. References

Croft, A., Mynors, G., Roberts, M., Doncaster, D., & Bowen, A. (2016). *MS Forward View: A consensus for the future of MS services*.

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., Lancaster, G. A., O'Cathain, A., Altman, D., Bretz, F., Campbell, M., Cobo, E., Craig, P., Davidson, P., Groves, T., Gumedze, F., Hewison, J., Hirst, A., Hoddinott, P., … Tugwell, P. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies*. https://doi.org/10.1186/s40814-016-0105-8

Giovannoni, G., Butzkueven, H., Dhib-Jalbut, S., Hobart, J., Kobelt, G., Pepper, G., Sormani, M. P., Thalheim, C., Traboulsee, A., & Vollmer, T. (2016). Brain health: time matters in multiple sclerosis. In *Multiple Sclerosis and Related Disorders*. https://doi.org/10.1016/j.msard.2016.07.003

Golden, C. J. (1978). *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses.* Stoelting.

Golden, C. J., & Freshwater, S. M. (2002). *The Stroop Color and Word Test : A Manual for Clinical and Experimental Uses*. Stoelting. https://www.worldcat.org/title/stroop-color-and-word-test-a-manual-for-clinical-and-experimental-uses/oclc/811788223

Hobart, J., Lamping, D., Fitzpatrick, R., Riazi, A., & Thompson, A. (2001). The multiple sclerosis impact scale (MSIS-29) a new patient-based outcome measure. *Brain*. https://doi.org/10.1093/brain/124.5.962

Honan, C. A., Brown, R. F., & Hine, D. W. (2014). The Multiple Sclerosis Work Difficulties Questionnaire (MSWDQ): Development of a shortened scale. *Disability and Rehabilitation*. https://doi.org/10.3109/09638288.2013.805258

Hooper, R. (2019). *Justifying sample size for a feasibility study*. https://www.rds-london.nihr.ac.uk/wpcms/wp-content/uploads/2019/02/Justifying-sample-size-for-feasibility-study-updated-22-Feb-2019.pdf

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

MA, W., AL, A., J, M., & WS, B. (1997). Case-finding instruments for depression. Two questions are as good as many. *Journal of General Internal Medicine*, *12*(7), 439–445. https://doi.org/10.1046/J.1525-1497.1997.00076.X

MS Trust. (2015). *Rehabilitation medicine consultants and MS: Report of a survey of rehabilitation consultants working in MS in the UK for the MS Forward View Project.*

Mynors, G., Bowen, A., & Doncaster, D. (2016). *MS Specialist Nursing in the UK 2016: Report on progress towards equitable provision*.

Nicholl, C. R., Lincoln, N. B., & Playford, E. D. (2002). The reliability and validity of the Nottingham Extended Activities of Daily Living Scale in patients with multiple sclerosis. *Multiple Sclerosis*. https://doi.org/10.1191/1352458502ms827oa

Rigby, S. A., Domenech, C., Thornton, E. W., Tedman, S., & Young, C. A. (2003). Development and validation of a self-efficacy measure for people with multiple sclerosis: The Multiple Sclerosis self-efficacy scale. *Multiple Sclerosis*. https://doi.org/10.1191/1352458503ms870oa

Ritvo, P. G., Fischer, J. S., Miller, D. M., Andrews, H., Paty, D. W., & LaRocca, N. G. (n.d.). *MSQLI Multiple Sclerosis Quality of Life Inventory: A User's Manual Developed by: The Consortium of Multiple Sclerosis Centers Health Services Research Subcommittee*.

Roberts, M., Mynors, G., & Bowen, A. (2016). *Improving services for people with advanced MS.*

Russell, D., Hoare, Z. S. J., Whitaker, R., Whitaker, C. J., & Russell, I. T. (2011). Generalized method for adaptive randomization in clinical trials. *Statistics in Medicine*, *30*(9), 922–934. https://doi.org/10.1002/sim.4175

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*. https://doi.org/10.1001/archinte.166.10.1092

Sullivan, M., Edgley, K., & Dehoux, E. (1990). A survey of multiple sclerosis: I. Perceived cognitive problems and compensatory strategy use. *Undefined*.

Referenced documents:

1. NEuRoMS WP3 Data Management Plan

2. NWORTH SOP 5.04 Randomisation systems

8. Appendices

Appendix 1 – Outcome measures summary table

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
PDQ	Questions to determine an assessment of cognitive dysfunction in patients with MS	Scale from 0 = Never to 5 = Almost Always, total score and subscale scores are sum of scores for relevant items	1. Attention – Q1, Q5, Q9, Q13, Q17 2. Retrospective memory – Q2, Q6, Q10, Q14, Q18 3. Prospective memory – Q3, Q7, Q11, Q15, Q19 4. Planning and organisation - Q4, Q8, Q12, Q16, Q20	Mean-substitution of missing items within a subscale provided at least 80% of subscale has been completed	None
MSIS-29	Questions to determine the impact of MS on day to day life in the past two weeks	Scale from 1 = Not at all to 5 = Extremely, total score and subscale scores are sum of scores for relevant items.	1. Physical – Q1 to Q20 2. Psychological – Q21 to Q29	Mean-substitution of missing items within a subscale provided at least 50% of subscale has been completed	None

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
PHQ-9	Questions to determine whether they have been bothered by symptoms of depression in the past two weeks	Scale from 0 = Not at all to 3 = Almost every day, total score is sum of scores for relevant items	None	None	0 - 4 = None 5 - 9 = Mild depression 10 - 14 = Moderate depression 15 - 19 = Moderately severe depression 20 - 27 = Severe depression From PHQ and GAD-7 Instructions Manual
GAD-7	Questions to determine whether they have been bothered by symptoms of anxiety in the past two weeks	Scale from 0 = Not at all to 3 = Almost every day, total score is sum of scores for relevant items	None	None	0 – 5 = None 6 - 10 = Mild Anxiety 11 - 15 = Moderate Anxiety 16 – 21 = Severe Anxiety From PHQ and GAD-7 Instructions Manual
Whooley Questions for Depression	Two questions to evaluate level of depression	“Yes” or “No”	None	None	“Yes” to one or both questions requires further evaluation, “No” to both questions indicates not depressed

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
NEADL	Questions regarding level of disability	Scored 0 “Not at all”, 0 “With help”, 1 “On my own with difficulty” and 1 “On my own” , total score and subscale scores are sum of scores for relevant items	1. Mobility – Q1 to Q6 2. Kitchen – Q7 to Q11 3. Domestic – Q12 to Q16 4. Leisure – Q17 to Q22	None	None
MSSE	Self-efficacy measure for patients with multiple sclerosis	Each of 14 item is scored on a 6-point Likert scale, ranging from ‘Strongly Disagree’ to ‘Strongly Agree’. Some items need to be reverse scored. The total score is the sum of the item scores.	None	None	None
MSWDQ	Work difficulties in patients with MS	Each of 23 items scored on a 5-point Likert scale, ranging from ‘Never’ to ‘Almost always’, total score and subscale scores are sum of scores for relevant items	1. Psychological/Cognitive barriers – Q2, Q3, Q4, Q6, Q7, Q10, Q13, Q15, Q16, Q19, Q22 2. Physical barriers – Q1, Q5, Q8, Q11, Q14, Q18, Q29 3. External barriers – Q12, Q17, Q21, Q23	None	None

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
Two single-item questions for medication adherence and work/education	Extent that medication adherence and work/education has been impacted by MS	5-point Likert scale ranging from 'Never' to 'Almost always'. There is also an additional option to select if the question is not applicable i.e. not taking medication or not in work/education.	None	None	None