NEuRoMS
Neuropsychological Evaluation and Rehabilitation in Multiple Sclerosis

Nottinghamshire Healthcare NHS
NHS Foundation Trust

NWORTH
North Wales Organisation for Randomised Trials in Health & Social Care
Sefydliad Hap-Dreialon Iechyd a Gofal Cymdeithasol Gogledd Cymru

PRIFYSGOL BANGOR UNIVERSITY

University of Nottingham
UK | CHINA | MALAYSIA

SCHE Swansea Centre for Health Economics
Research led, Analysis Driven, Real World Perspective

Prifysgol Abertawe
Swansea University

STATISTICAL AND HEALTH ECONOMIC ANALYSIS PLAN

FOR NEuRoMS Work Package 4

Date: 14/09/2026

Version: 2

North Wales Organisation for Randomised Trials in Health (NWORTH)
Bangor University, College of Health & Behavioural Sciences,
Institute of Medical & Social Care Research, Ardudwy, Normal Site, Holyhead Road, Bangor, Gwynedd LL57 2PZ
Telephone: 01248 388095 Email: nworth@bangor.ac.uk Website: http://nworth-ctu.bangor.ac.uk/

Statistical and Health Economic Analysis plan for the Neuropsychological Evaluation and Rehabilitation in Multiple Sclerosis (NEuRoMS): A mixed methods pragmatic multicentre Randomised Controlled Trial (RCT) with nested health economic and process evaluations and an implementation study (Phase 3: Work Packages 4 and 5)

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.

FUNDED BY
NIHR | National Institute for Health and Care Research

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 7, dated 16/04/2026.

Chief Investigator	Professor Roshan das Nair
Statistical co-applicant	Dr Zoë Hoare
Health economic co-applicant	Prof Deb Fitzsimmons
Programme Manager	Dr Gogem Topcu
Data Manager	Dr Nia Goulden
Trial Statistician	Dr Nia Goulden
Health economist	Dr Shaun Harris

Nia Goulden

Statistician Author's Name **Date and Signature**

Dr Zoë Hoare

Second Statistical Advisor **Date and Signature**

Dr Shaun Harris

Health Economist Author **Date and Signature**

Professor Deborah Fitzimmons

Second Health Economic 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	14/08/2026	N. Goulden	New	
2		N. Goulden S. Harris	6.5.6 6.6.14 Appendix 2	Added information regarding EQ-5D-5L scoring Sensitivity Analysis reworded to be consistent with SAP section Additional information regarding assumptions of models

Acronyms and definition of terms

Acronym	Meaning
A-Level	Advanced Level
AP	Assistant Psychologist
BNF	British National Formulary
CEAC	Cost-Effectiveness Acceptability Curve
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
COReC	Care and Outcomes Research Centre
CONSORT	Consolidation Standards of Reporting Trials
CRF	Case Report Form
CTU	Clinical Trials Unit
DMEC	Data Monitoring and Ethics Committee
DMT	Disease Modifying Therapy
EQ-5D-3L	EuroQol EQ-5D Questionnaire -3 Level Version
EQ-5D-5L	EuroQol EQ-5D Questionnaire -5 Level Version
GAD	Generalized Anxiety Disorder
GAS	Goal Attainment Scale
GCSE	General Certificate of Secondary Education
ICECAP-A	ICEpop CAPability measure for Adults
ICER	Incremental Cost Effectiveness Ratio
IT	Information Technology
MCID	Minimum Clinically Important Difference
MFIS	Modified Fatigue Impact Scale
MS	Multiple Sclerosis
MSIS	Multiple Sclerosis Impact Scale

MSIS-Psych	Multiple Sclerosis Impact Scale – Psychological Subscale
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
NHSCII	NHS cost inflation index
NICE	National Institute for Health and Care Excellence
NVQ	National Vocational Qualification
NWORTH	North Wales Organisation for Randomised Trials in Health
ONS	Office for National Statistics
OT	Occupational Therapist
PDQ	Perceived Deficits Questionnaire
PHQ	Patient Health Questionnaire
PMG	Programme Management Group
PMM	Predictive Mean Matching
PPI	Patient and Public Involvement
PSC	Programme Steering Committee
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
SD	Standard Deviation
SHEAP	Statistical and Health Economic Analysis Plan
SST	Symbol Substitution Task
SOP	Standard Operating Procedure
SPSS	Statistical Package for the Social Sciences
SUQ	Service Use Questionnaire
WCT	Word Colour Task
WP	Work Package
WTP	Willingness-to-pay

Table of Contents

Acronyms and definition of terms 4
1. Statistical and Health Economic Analysis Plan Authorship 8
2. Introduction 8
2.1 Background and Rationale 8
2.2 Trial Objectives 9
2.3 Trial Population 9
2.4 Trial Design 10
2.5 NEuRoMS Intervention 10
3. Statistical Principles 13
3.1 Sample Size Justification 13
3.2 Randomisation 13
3.3 Levels of Confidence 13
3.4 Protocol Violations and Deviations 14
3.5 Missing Data 14
3.6 Outliers 14
4. Data 15
4.1 Part 1 Screening Data 15
4.2 Part 2 Clinical Outcomes 15
4.3 Definitions and Calculations of Outcome Measures 17
4.4 Unblinding 17
4.5 Safety Data 17
5. Statistical Analyses 18
5.2 Recruitment and Retention 18
5.3 Descriptive statistics 20
5.3 Part 1 Analysis 20
5.4 Part 2 Outcome Measures 21
5.5 Check of Assumptions 24
6. Health Economic Analyses 25
6.1. Aim and objective of economic evaluation 25
6.2. Overview of economic analysis 25
6.3. Jurisdiction and Perspective 25
6.4. Time Horizon 25
6.5. Economic data collection and management 25
6.5.1. Identification of resources 25
6.5.2. Measurement of resource-use data 26
6.5.3. Valuation of resource-use data 26
6.5.4. Identification of outcomes 28
6.5.5. Measurement of outcomes 28
6.5.6. Valuation of outcomes 28
6.6. Economic Data Analysis 28

6.6.1. Analysis population 28
6.6.2. Timing of analysis 28
6.6.3. Measurement of outcomes 29
6.6.4. Discount rates for costs and benefits 29
6.6.5. Cost-effectiveness threshold 29
6.6.6. Statistical Decision Rules 29
6.6.7. Analysis of resource use and costs 29
6.6.8. Analysis of outcomes 29
6.6.9. Data cleaning for analysis 30
6.6.10. Missing data 30
6.6.11. Analysis of cost-effectiveness 30
6.6.12. Sampling uncertainty 31
6.6.13. Subgroup analyses 31
6.6.14. Sensitivity analyses 31
6.6.15. Modelling of longer-term cost-effectiveness 32
6.7. Reporting/publishing standards 33
7. Software 33
7. References 33
8. Appendices 38

1. Statistical and Health Economic Analysis Plan Authorship

The statistical analysis plan has been authored by Dr Nia Goulden, Trial Statistician. The health economic analysis plan has been authored by Dr Shaun Harris and Professor Deborah Fitzsimmons. There has also been input from Dr Zoë Hoare (Principal Statistician), Dr Gogem Topcu (Programme Manager), Dr Grzegorz Suldecki (Trial Manager) and Professor Roshan das Nair (Chief Investigator). The draft plan will be circulated to the Programme Management Group, Data Monitoring and Ethics Committee and Programme Steering Committee for comments before being signed off. All statistical analysis will be completed by Nia Goulden and overseen by Dr Zoë Hoare. All Health economic analysis will be completed by Dr Shaun Harris with researcher support and overseen by Prof. Deb Fitzsimmons.

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 (1–4). By not intervening early, disability accrues, and people with MS are less likely to benefit from rehabilitation (5), and costs for the people with MS, their families, the NHS and society are likely to escalate.

As part of the NEuRoMS screening and management pathway, all patients will complete online cognitive screening as part of newly introduced routine clinical care. Based on pre-determined cut-offs and data from Work Package 2, Work Package 3, and Work Package 6 (validation and norming studies), 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 referred to the trial, and if they consent to participate, will be randomly allocated to receive either a manualised NEuRoMS rehabilitation intervention led by an Assistant Psychologist (AP)/Research Nurse/Assistant Occupational Therapist (OT) (i.e., intervention providers) plus usual care or usual care only.

This document details the statistical and health economic analysis plan for Work Package 4. In this work package, we will conduct a Randomised Controlled Trial (RCT) to evaluate the new screening and management pathway.

Part 1 is an observational study of those who receive cognitive screening, using routinely collected clinical data to assess the frequency and extent of cognitive problems. Part 2 uses a multicentre, individually randomised, pragmatic, parallel

group RCT design with embedded health economic evaluation to test the clinical and cost-effectiveness of the NEuRoMS rehabilitation intervention.

2.2 Trial Objectives

The primary objective is to determine whether receiving the NEuRoMS rehabilitation intervention in addition to usual care is associated with reduced impact of cognitive problems, as measured on the MS Impact Scale (MSIS-Psych) when compared to usual care alone, 3 months post-randomisation.

Secondary objectives are:

Part 1 – Cognitive Screening:

1.a. Assess the frequency and severity of cognitive problems, based on Symbol Substitution Task (SST) and/or Word Colour Task (WCT).

1.b. Explore the relationship between the demographics and screening usage data collected, and cognitive screening performance.

1.c. Explore whether there are different neuropsychological phenotypes (i.e., clusters of participants based on measurable traits such as self-reported cognitive problems, fatigue and mental health) based on the screening data.

Part 2 – Randomised Controlled Trial of Intervention for Cognitive Problems:

2.a. Evaluate whether the NEuRoMS rehabilitation intervention in addition to usual care is associated with improvements in participant's mood, objective and subjective cognitive impairment, quality of life, function, self-efficacy, fatigue, work related issues, Disease Modifying Therapy (DMT) adherence, and individual goal attainment.

2.b. Evaluate cost-effectiveness of the intervention.

2.3 Trial Population

Eligibility Criteria

All individuals: Able and willing to give consent and able to communicate in English. Participant information sheets can be provided in Welsh upon request but the standardised materials and tests to be used require communication in English since these have not yet been developed for other languages.

Part 1: People with MS:

- Diagnosis of MS
- Aged 18 years or above

Part 2: People with MS:

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

- Mild or moderate problems will be identified by the clinical team based on the screening scores, the clinical and contextual information collected as part of the cognitive screening, and the patient-clinician conversations on cognitive problems experienced.
- Thresholds for mild and moderate cognitive problems in people with MS will be defined based on norms compared to control participants (people without MS) defined by (6) on the level of performance on SST and/or WCT*:
    - Mild cognitive problems: Scores of 1 SD below the mean or lower, and higher than 2 SD below the mean;
    - Moderate cognitive problems: Scores of 2 SD below the mean or lower, and higher than 2.5 SD below the mean.

      *People with scores in mild or moderate ranges in either cognitive test (SST and/or WCT) would be considered eligible to participate.

Exclusion criteria

All participants: Lack mental capacity to consent to participate in the study.
RCT participants only:

- Currently receiving neuropsychological intervention for cognitive problems.
- Received NEuRoMS intervention during WP2ii or WP3 (earlier work packages in the NEuRoMS programme).

2.4 Trial Design

Part 1 involves an observational study of those who receive screening 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 seven MS clinics as part of routine clinical care.

Part 2 involves a multi centre, individually randomised, pragmatic parallel group RCT, with embedded health economic evaluation, evaluating the clinical and cost-effectiveness of the NEuRoMS rehabilitation intervention by comparing NEuRoMS intervention programme plus usual care to usual care alone for people with MS with mild and moderate cognitive problems (Figure 1).

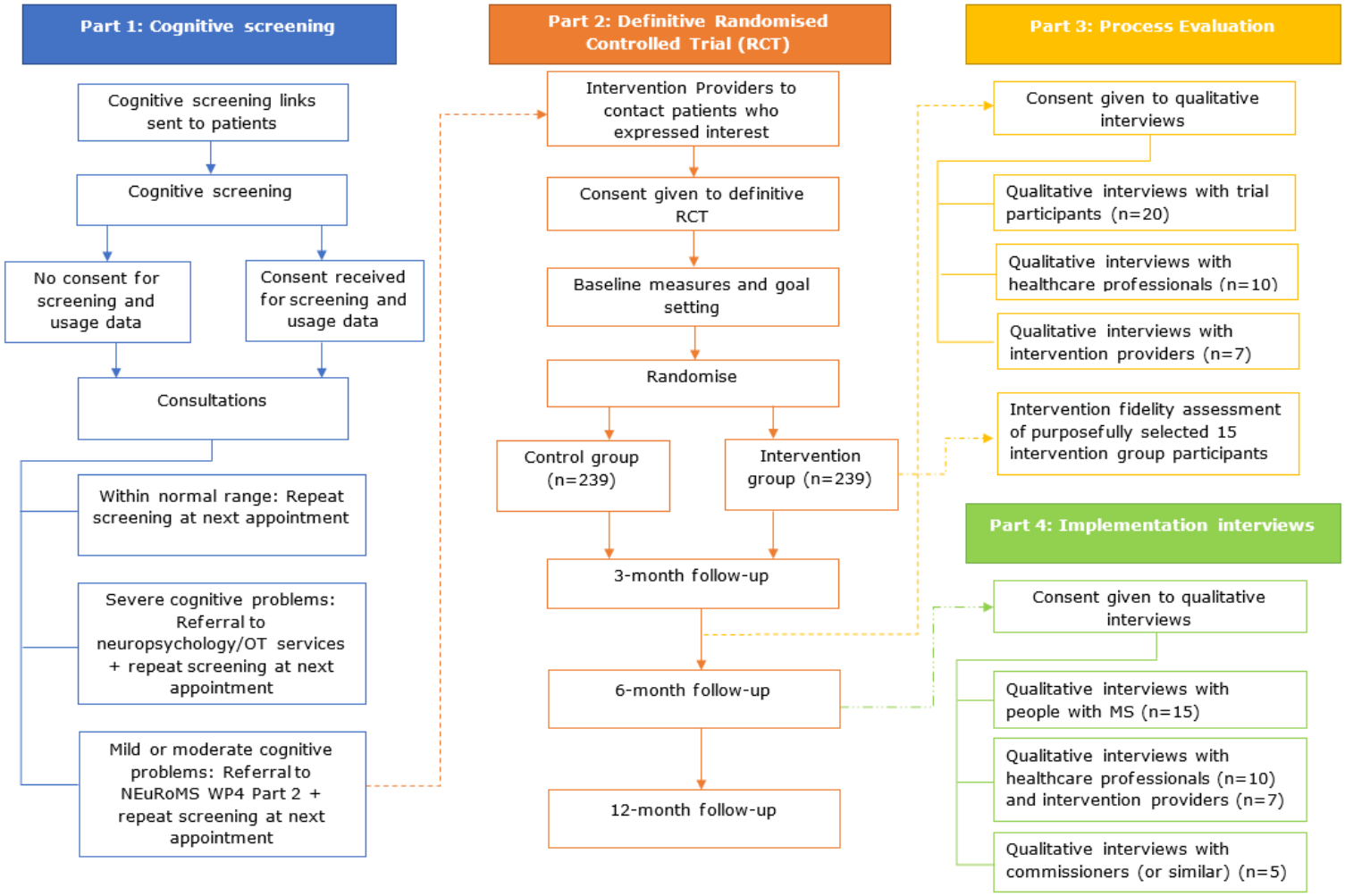
2.5 NEuRoMS Intervention

The NEuRoMS rehabilitation intervention is multi-faceted, involving various components (i.e., information provision, setting an action plan to achieve goals) and a range of strategies and techniques (e.g., psychoeducation, compensatory strategies, boosting cognitive reserve) and resources. 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 intervention sessions are finished.

The intervention is manualised and will be delivered by a trained therapist, also known as intervention providers, with psychology, nursing or occupational therapy backgrounds (e.g., AP, Research Nurse, or Assistant OT), under the supervision of a clinical psychologist or OT. In-person, videoconferencing and telephone, and hybrid delivery options will be available. To support remote delivery, digital content, including brief online videos, will be available to complement the Intervention Workbook. The duration of the intervention will be up to 4 hours, spread across up to 6 sessions. We anticipate these sessions to occur over a 2-3 month period, based on patient availability.

Figure 1: Design of the trial



3. Statistical Principles

3.1 Sample Size Justification

Part 2: The primary outcome is the self-reported psychological impact of MS on everyday life, and will be measured using the MSIS-Psych (7). This subscale taps items that relate to psychological wellbeing associated with MS, which our PPI partners have identified as being important to them and which are likely to be affected by the intervention. We are aware of MCID for the physical subscale of MSIS (MCID=8, based on 20 items), but not the psychological subscale. We have determined that a 3-point difference on this subscale (based on 9 items) would be clinically meaningful, based on the CRAMMS trial (8) and interviews we have had with PPI members. A 3-point difference would mean a change of one point (e.g., from 'moderately' to 'a little bit') on 3 items (e.g., 'worries related to your MS', 'problems concentrating', or 'lack of confidence') or a change of up to 3 points (e.g., from 'quite a bit' to 'not at all') on one item.

Based on a two sample *t*-test, a total sample of 478 people with MS will have 90% power to detect a 3-point difference with a standard deviation of 9, which provides an effect size of 1/3 (9) on the MSIS-Psych (7) at a 5% significance level, two-tailed, with 20% attrition (based on findings from our CRAMMS trial; (8)). Based on a conservative estimate of 30% having mild or moderate cognitive deficits (based on previous WPs) and 40% consenting (based on WP3), ≈3979 people with MS will need to be screened for eligibility. We believe this will be achievable, with a recruitment pool of ≈7040 screened patients in all sites in Part 1 over the 22-month recruitment period.

3.2 Randomisation

Once the participant has been recruited and consented, and has completed the baseline questionnaires, the intervention provider or a member of the research team will individually randomise the participant to intervention or control groups (ratio 1:1). Randomisation will be stratified by site and cognitive impairment level. Randomisation will be through a secure online dynamic adaptive algorithm (10), developed, hosted and maintained by NWORTH CTU.

3.3 Levels of Confidence

All applicable statistical tests will be two-sided and will be performed considering a 5% significance level, and all confidence intervals presented will be 95%.

3.4 Protocol Violations and Deviations

The definition of a protocol violation is an intended failure to adhere to the protocol such as wrong treatment being prescribed or administered or 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. Other examples of protocol deviations would be errors in entering data in the randomisation system, participants being assigned to the incorrect treatment and researchers being unblinded by participants during visits.

3.5 Missing Data

When the questionnaires are collected, the researcher will check for any missing data items. The researcher will contact the participant to complete any missing data .

For missing items within an outcome measure, any published guidelines for that outcome measure will be utilised. If the level of missing data is less than 5%, no action will be taken to account for missing data, in accordance with guidance by Jakobsen and colleagues (11). Otherwise, in accordance with guidelines by Bono and colleagues (12), if the number of missing items on an outcome measure is between 5 and 20%, the missing value for the item will be substituted by the mean score for that item for all participants at the same timepoint and in the same allocated group. If missing data on the measure is higher than 20%, the score will not be calculated via this method and analysis will be completed using the data that have been collected.

For missing total scores multiple imputations will be used. The number of imputations will be dependent on the percentage of missing data. Multiple imputation will be completed in Stata (13). Covariates which are to be included in the analysis models will be incorporated in the imputation model along with any factors deemed to be predictors of missingness.

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, they 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 WP4 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 Clinical Outcomes

Primary outcome

The primary outcome for the RCT is the self-reported psychological impact of MS on everyday life, and will be measured using the Psychological Subscale of the Multiple Sclerosis Impact Scale (MSIS-Psych; (7)). MSIS-Psych will be assessed at 3, 6 and 12 months post-randomisation to assess immediate and long-term effects of the intervention. The primary endpoint is at 3 months post-randomisation.

Secondary outcomes

Secondary outcomes to be assessed in relation to each of the secondary objectives are:

Outcomes for secondary objectives 1a and 1b:

Participant scores from cognitive screening measures (completed at a single time-point at screening) including an online version of the: 1) Symbol Digit Modalities Test (Symbol Substitution Task; SST), 2) Stroop task (Word Colour Task; WCT).

Demographic, clinical features and usage information will be collected as part of cognitive screening (at a single-time point at screening) to describe the sample, understand usage and how the participants performed on the screening tasks, and to explore the relationship between the screening scores and the demographics and usage information. Demographics, clinical features and usage information completed at screening include: age, gender, education, ethnicity, Index of Multiple Deprivation, MS type and duration, hand preference, COVID status, fatigue level, vision and hand problems while completing the tasks, type of device used to complete the screening (e.g., tablet, mobile phone, laptop), and setting the screening completed at (e.g. online at home, online in-clinic).

Other usage data that will be collected through electronic Case Report Forms (CRFs) include time taken to complete the screening (in minutes), number of patients who require reminders and extra support (telephone/in-clinic) to complete the screening, and number and content of contact between service providers and patient participants.

Outcomes for secondary objective 1c:

Clinical data collected as part of usual care: Patient scores from the three measures from the Multiple Sclerosis Quality of Life Inventory for self-reported cognitive problems, fatigue and mental health (14), measured during screening as part of usual care: 1) Modified Fatigue Impact scale-5 item (15); 2) Mental Health Inventory-5 item (16); and 3) Perceived Deficits Questionnaire-5 item (17).

Outcomes for secondary objectives 2a and 2b:

The following secondary outcome measures will be used to capture information about the patient at baseline and to assess outcomes at 3, 6 and 12 months after randomisation. Secondary outcomes are:

- Self-reported (subjective) cognitive impairment as measured using the Perceived Deficits Questionnaire – 20-items [PDQ-20; (18))].
- Objective cognitive impairment as measured using the Symbol Substitution Task (SST). Participants' Part 1 SST scores will be used as baseline, and SST will be repeated at 3-, 6- and 12-month follow-ups.
- Quality of life as measured using the Multiple Sclerosis Impact Scale [MSIS-29; (7)] and EQ-5D-5L (19).
- Mood as measured using the Patient Health Questionnaire 9-items (20)]; Generalized Anxiety Disorder-7 (21)].
- Function as measured using the Nottingham Extended Activities of Daily Living Scale [NEADL;(22)].
- Self-efficacy as measured using the Multiple Sclerosis Self-efficacy Scale [MSSE (23)].
- Fatigue as measured using the Modified Fatigue Impact Scale-5 item version [MFIS-5;(15); (24))].
- Work-related issues as measured using the Multiple Sclerosis Work Difficulties Questionnaire short form (25) and a bespoke single-item question asking to what extent work has been impacted by cognitive problems. For participants who are in education, a bespoke single-item question will be asked, to measure to what extent education has been impacted by cognitive problems. All work and education related questions will only be asked to those participants who indicated they are currently at work and/or in education in the demographics form. The employment status question on the demographics form will be repeated at all follow-up time points to check if their status has changed since baseline.
- Medication adherence as measured using a bespoke single-item question asking to what extent medication adherence has been impacted by cognitive problems.
- Cost-effectiveness as measured by the:

    - Service-use questionnaire, based on a measure used in other MS trials (8) adapted for use in the NEuRoMS project and refined in WP3 feasibility study.
    - EQ-5D-5L (19).
    - MSIS-8D (using MSIS-29) to derive utilities.
    - ICECAP-A [ICEpop CAPability measure for Adults; (26)] measure of capability and wellbeing for health economic evaluation.
- (Please note: MSIS-29 and EQ-5D-5L will be collected only once per time point, but will be used for two purposes: 1) for clinical evaluation, and 2) for cost-effectiveness evaluation.)
- Goal Attainment Scale (GAS): Each Part 2 participant will be asked to set between two and three personal short- and long-term goals that they would like to achieve by the end of the study (at least one short- and one long-term goal, with an optional third goal). Participants will set their goals after completing the baseline questionnaires, before randomisation, with the assistance of the intervention provider (face-to-face/phone/ video-conferencing/email). At 3-, 6-, and 12-month follow-ups, each participant will be asked to evaluate their individual goals in terms of the degree to which each goal has been met on a four-point Likert scale (i.e., 'not met at all', 'met a little', 'mostly met', and 'fully met').

4.3 Definitions and Calculations of Outcome Measures

The data collected as part of the trial will include the measures listed in section 4.2. See the table in appendix 1 for full details.

4.4 Unblinding

The final unblinding for results will take place after all blinded analysis statistical and health economic 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, the DMEC and PSC.

4.5 Safety Data

Any safety events from this trial will be reported with details of severity, cause and outcome included. This will be completed overall and per group. The number and

percentage of patients that have been affected by an adverse event will also be presented. No formal statistical testing will be undertaken on these data.

5. Statistical Analyses

5.1 Analysis Time Frame

TASK	EXPECTED DATE
First participant recruited	November 2023
Final participant recruited	August 2025
Final follow-up completed	October 2026
Data cleaning completed	November 2026
* Analysis completed	November 2026 – January 2027

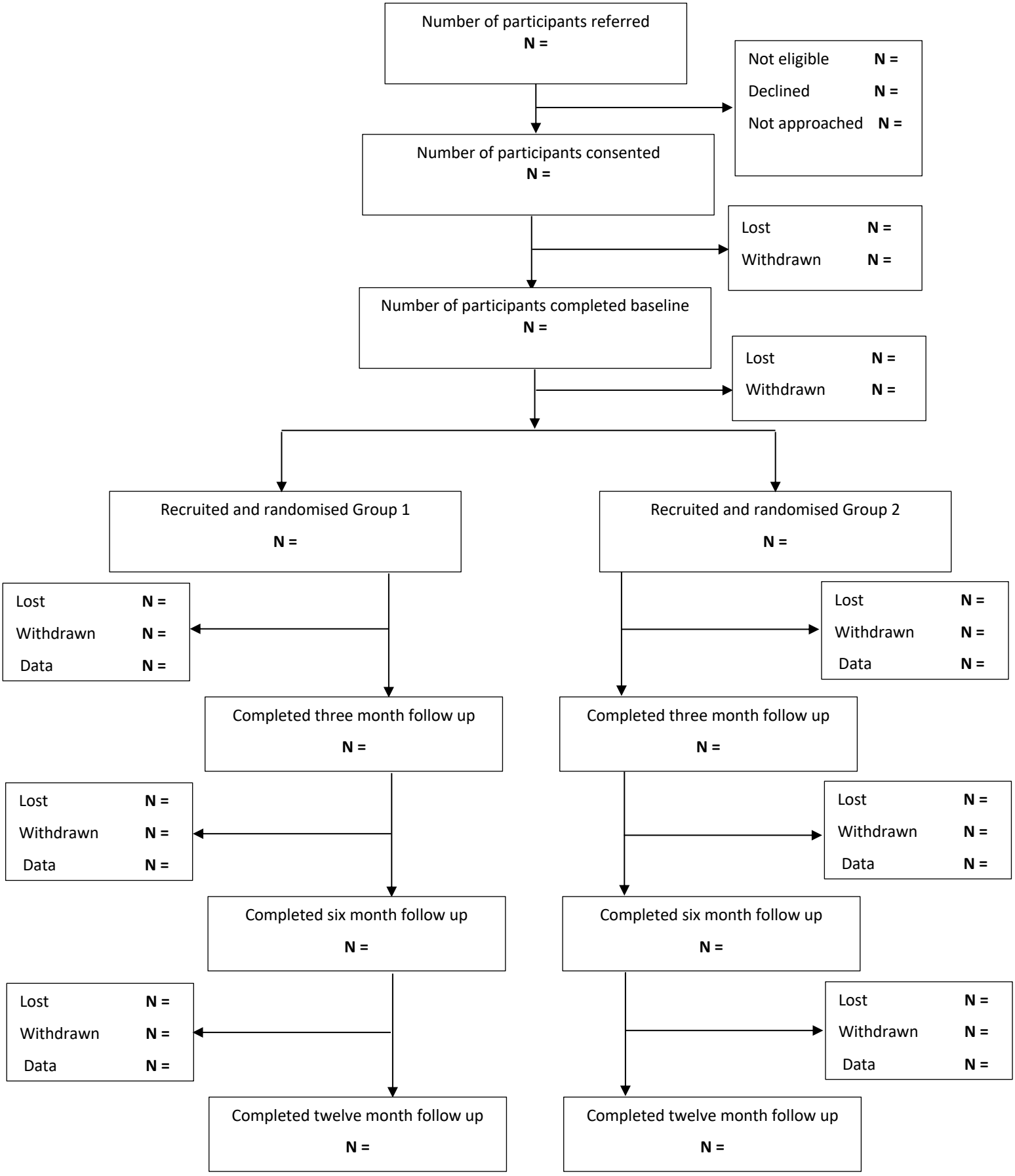
*Analysis will require approx. 2 months from the date of data locking. Delays in responding and dealing with data queries and data entry may result in delays to analysis delivery.

5.2 Recruitment and Retention

The CONSORT diagram for NEuRoMS WP4 is presented in Figure 2. The analysis will consider the items from the CONSORT checklist (27) 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 attrition rates and withdrawal rates will be reported per group.

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

Figure 2: CONSORT diagram for NEuRoMS WP4



5.3 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 will be reported with counts and related percentages.

5.3 Part 1 Analysis

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

The proportion of participants falling within normal range, or presenting with mild, moderate and severe cognitive impairment based on the screening battery will be calculated.

The SST task is 90 seconds long. The total number of items, correct items and incorrect items at 15, 30, 45, 60, 75 and 90 seconds will be collected. The WCT has three components or '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, 75 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 (28,29) will be used for the WCT data.

Regression based norms

There will be additional analysis to evaluate the thresholds adjusted for age, gender and level of education. For the SST and WCT, multiple linear regression will be employed with the task result as dependent variable, age as a covariate and gender and level of education as factors. This will allow regression-based norms to be applied to data. For both tests, means and standard deviations will be calculated from the analysis output to calculate thresholds adjusted for age, gender and level of education as defined above.

Neuropsychological phenotypes predictors

The SST, WCT, Modified Fatigue Impact scale-5 item score, the Mental Health Inventory-5 item score and the Perceived Deficits Questionnaire-5 will be analysed using linear regression in order to determine predictors of the score, the score from the task will be the dependent variable. The model will include age as a covariate and gender and level of education as factors.

Effect sizes will be estimated using adjusted scores from the model, standard errors and 95% confidence intervals.

5.4 Part 2 Outcome Measures

All analysis will be guided by the principle of intention-to-treat. This means that data will be analysed with participants in the group they were allocated to by the randomisation algorithm, regardless of whether they received the allocated intervention or not.

The following sensitivity analyses will also be run in order to assess consistency of results from multiple imputation:

1) Complete Case, including participants who have complete data only.

Primary Analysis for 3 month follow up primary endpoint

The primary outcome, the MSIS-Psych, as well as the PDQ-20, SST, MSIS-29, MSSE, PHQ-9, GAD-7, NEADL, MFIS-5, GAS, MSWDQ and single item questions about work, education and medication adherence will be analysed using a multi-level mixed effects linear regression model, to assess scores at the three month follow up as this is the primary endpoint. The score at this time point will be the dependent variable. The model will include the baseline score and age as covariates, as well as cognitive impairment group (Mild or Moderate), gender (Woman, Man, Other, Prefer not to say), level of education (Below GCSE, GCSE, A-Level, NVQ, Degree, Higher Degree, Not known), type of MS (Relapsing-remitting MS, Primary Progressive MS, Secondary Progressive MS, Not known) and allocated treatment group (Control or NEuRoMS intervention) as fixed factors. The model will also include site (Cardiff, Leeds, Lincoln, London, Newcastle, Nottingham, Sheffield) as a random effect. The assumed covariance matrix will be the independent covariance matrix.

Effect sizes will be estimated using adjusted mean differences from the model, standard errors and 95% confidence intervals.

The primary effect estimate will be the adjusted mean difference for the MSIS-Psych at 3 month follow up. The primary hypothesis, of a significant difference between groups for the MSIS-Psych at 3 month follow up, will be tested by assessing the p-value for the allocated group in the analysis model. If allocated group has a p-value less than

0.05 then it will be possible to state that there is a significant difference between the groups, otherwise it will not be possible to state that there is a significant difference. The hypothesis of a significant difference between groups will be evaluated in the same way for all outcomes.

Secondary analysis for secondary endpoints

The MSIS-Psych, PDQ-20, SST, MSIS-29, MSSE, PHQ-9, GAD-7, NEADL, MFIS-5, GAS, MSWDQ and single item questions about work, education and medication adherence will be analysed using a multilevel mixed effects linear regression model, to assess changes over time and group differences. The model will include the baseline score and age as covariates, as well as cognitive impairment group (Mild or Moderate), gender (Woman, Man, Other, Prefer not to say), 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 fixed factors. The model will also include allocated treatment group (Control or NEuRoMS intervention), time (6 month follow up, 12 month follow up) and a time*group interaction, as well as participant nested within site (Cardiff, Leeds, Lincoln, London, Newcastle, Nottingham, Sheffield) as a random effect. The assumed covariance matrix will be the independent covariance matrix.

If the time*group interaction is found to be non-significant then the interaction will not be included in the analysis model for ease of interpretation.

Effect sizes will be estimated using adjusted mean differences from the model, standard errors and 95% confidence intervals.

Using Interval-level metrics for the MSIS-Psych

The primary and secondary analysis will be repeated for the MSIS-Psych, the primary outcome, using interval-level metrics found from Rasch analysis and reported by Young and colleagues (30).

Subgroup Analysis

The primary analysis will be repeated with addition of an interaction term for allocated group*cognitive impairment group. This will assess whether the pattern of results is different for those with mild cognitive impairment and moderate cognitive impairment.

Sensitivity Analysis – (Low number of participants in categories)

Within the gender variable, it may be the case that the 'Other' and 'Prefer not to say' categories are poorly represented. In the event that there are fewer than 25 participants, and there is a significant ($p<0.05$) effect for either category, a sensitivity analysis would be conducted on the primary and secondary analysis excluding the data for 'Other' and 'Prefer not to say' to check that there are no spurious significant effects as a result of small category sizes.

If there is a similar effect seen for any other category then this approach will also need to be undertaken. The same may be true for 'Not known' level of education, for example.

Sensitivity Analysis – Follow up visit Data Collection

Some participants may complete data earlier or later than the target date, all data will be included for the primary and secondary analyses. Sensitivity analysis, following the same format as the primary and secondary analyses for the primary outcome, will be conducted to include data from visits which were completed within two weeks of the target date. The two week window has been chosen to reflect a reasonable timeframe for completing the visits as required. For each visit this means:

1. Three month follow up; Target date is 91 days after randomisation, therefore visits 77-105 days after randomisation will be included;
2. Six month follow up; Target date is 182 days after randomisation, therefore visits 168-196 days after randomisation will be included;
3. Twelve month follow up; Target date is 365 days after randomisation, therefore visits 351-379 days after randomisation will be included.

Note that participant data will not be excluded for this purpose if one visit is within the acceptable time and another not. For example, if the six month follow up is completed within the acceptable time but the twelve month follow up not, the six month follow up data will be included in the analysis and twelve month follow up data will not be included in the analysis.

If there is a significant effect of group for a secondary outcome, this sensitivity analysis will also be conducted for that secondary outcome.

Sensitivity Analysis – Intervention Engagement

Sensitivity analysis, following the same format as the primary and secondary analyses for the primary outcome, will be conducted to exclude data from participants in the intervention group who did not complete their first intervention session before the three month follow up. Therefore, any participant in the intervention group who did

not complete the first intervention session by their three month follow up will not be included in this analysis. Note that all participants in the control group will be retained.

If there is a significant effect of group for a secondary outcome, this sensitivity analysis will also be conducted for that secondary outcome.

Since this analysis requires knowledge of intervention delivery, this analysis will need to be completed after the trial statistician has been unblinded.

Sensitivity Analysis – Follow up visit Data Collection and Intervention Engagement

The sensitivity analysis for follow up timing will be repeated, excluding participants from the intervention group who did not complete their first intervention session before the three month follow up as described in the sensitivity analysis for intervention engagement.

5.5 Check of Assumptions

A check of the assumptions required of the statistical tests will be performed, see tables in appendix.

6. Health Economic Analyses

6.1. Aim and objective of economic evaluation

The aim of the economic evaluation is to evaluate whether the NEuRoMS intervention is cost-effective compared to usual care for people with MS.

The primary objective is to evaluate the cost-effectiveness of the NEuRoMS intervention compared to usual care from an NHS and Personal Social Services (PSS) perspective for people with MS at 12-months after randomisation.
The secondary objectives will be to:

- Assess the cost-effectiveness of the NEuRoMS intervention compared to usual care in patients with MS utilising an MS specific quality of life measure
- Assess the primary cost-effectiveness at 3-months post-randomisation, to match the primary statistical analysis endpoint.
- Examine the cost-effectiveness from a wider perspective (including directly incurred patient and carer costs, and productivity losses).

6.2. Overview of economic analysis

The within-trial cost-effectiveness analysis will use individual patient data from the NEuRoMS trial. The analytical approaches will take the form of cost-effectiveness and cost-utility analyses. The primary analysis will consist of a cost-utility analysis (CUA) considering the comparative differences in costs and quality-adjusted life years (QALYs). Incremental cost-effectiveness ratios (ICERs) will be calculated as a ratio between the difference in mean costs and mean effects. The CEA will use the primary statistical outcome measure (MSIS-Psy) with ICERs expressed as the additional cost per unit improvement in MSIS-Psy score.

6.3. Jurisdiction and Perspective

The primary health economic analysis will be conducted from UK NHS and Personal Social Services perspective which is predominantly publicly funded and free at the point of access. A broader partial societal perspective including patient and carer costs will be conducted as an additional sensitivity analysis. This will include direct health and non-health costs, and an estimation of lost productivity, using a human capital approach to estimate costs from the Office for National Statistics (ONS) sources.

6.4. Time Horizon

The primary economic analysis will compare the costs and effects for both the NEuRoMS intervention and usual care arms for the 12-month follow-up period.

6.5. Economic data collection and management

6.5.1. Identification of resources

Relevant resources identified as important include:

1. Resources used for the training and delivery of the NEuRoMS intervention
2. Primary health care use including consultations with key healthcare professionals (e.g. GP, community nurse and healthcare assistant)
3. Use of social services including contacts with local authority and social care services
4. Secondary care use including MS-related outpatient, day case, inpatient, and accident and emergency visits
5. Use of MS related medications
6. Hours of employment, education, and normal activities lost, and hours of informal care by family and friends

6.5.2. Measurement of resource-use data

Intervention and delivery costs

Costs relating to the intervention will be determined by calculating the costs of staff time and any resources and materials used in delivering the NEuRoMS intervention. This includes any resources associated with training, delivery, and administration, but excludes any costs which would not be a component of the NEuRoMS in a standard clinical setting, including those associated with research.

Primary and secondary health care use

Patient self-completed questionnaires will be used to collect: (1) MS-related primary care and community based health care contacts; (2) MS-related secondary care activity (e.g. outpatient, inpatient, A&E visits); (3) MS-related social service use; and (4) MS-related prescribed medications.

Patient incurred costs

For the broader societal perspective directly incurred MS-related costs (e.g. private MS-related health and care contacts, travel to services), lost hours of employment and education, normal activities, and informal care will be obtained from patient self-completed questionnaires.

6.5.3. Valuation of resource-use data

All resources will be valued in monetary terms using appropriate UK unit costs or participant valuation estimated at the time of analysis. When unit costs are not available for the year of analysis, it will be inflated to current prices using the NHS cost inflation index (NHSCII) as published in the unit cost series.

Table 1: Resource use data measurement and valuation

Resource Use	**Item**	**Measurement**	**Unit cost source**	**Unit cost calculations/assumptions**
Intervention costs	Staffing required for providing intervention, resources utilised	Discussions with trial team	COReC, National Cost Collection (31) (32)	Duration of sessions, unit cost of additional resources
Primary healthcare	Number of: GP contacts; nurse contacts; practitioner contacts; community care contacts	Patient self-reported questionnaire	COReC	e.g. consultation time, type (face-to-face/virtual)
Secondary healthcare	Number of hospital inpatient admissions; number of nights in hospital	Patient self-reported questionnaire	National Cost Collection	Estimated cost per admission
	Number of hospital outpatient visits	Patient self-reported questionnaire	National Cost Collection	Weighted average of outpatient visits
	Number of emergency department/urgent care unit visits	Patient self-reported questionnaire	National Cost Collection	Accident and emergency visit cost
Prescribed medications	Number of medications prescribed	Patient self-reported questionnaire	British National Formulary (33), Prescription cost analysis (33)	Based on dosage, frequency, and duration
Work, education and productivity	Hours of work, employment, and normal activities lost	Patient self-reported questionnaire	Office of National Statistics (34)	Based on average hours per week
Informal care costs	Hours of additional support/care by family/friends	Patient self-reported questionnaire	Office of National Statistics	Based on average hours per week

6.5.4. Identification of outcomes

The primary economic outcome measure will be quality adjusted life years (QALYs) derived from the EQ-5D-5L quality of life instrument (35) completed at 12-months. A secondary cost-utility analysis using QALYs derived from the MSIS-8D will also be evaluated. Secondary outcomes for the economic evaluation will include a cost-effectiveness analysis on the primary outcome of MSIS-Psych, and a capability adjusted economic evaluation using ICEpop CAPability measure for Adults (ICECAP-A) (36).

6.5.5. Measurement of outcomes

Outcomes will be collected at baseline, 3-, 6-, and 12-months post-randomisation by questionnaires completed online (using REDCap), using postal paper copies, or over the phone. Baseline data will be collected once the participant has been recruited and consented.

6.5.6. Valuation of outcomes

Participant EQ-5D-5L profiles will be valued using the validated EQ-5D-5L UK value set as recommended by the National Institute for Health Care (37,38). As a sensitivity analysis, EQ-5D-5L responses will be mapped to the UK EQ-5D-3L value set using the validated crosswalk(36).The MSIS-8D, a disease specific quality of life measure for people with MS, will be derived from the MSIS-29 using the system developed by Goodwin et al. (39). Utility values from both the EQ-5D-5L and MSIS-8D will be transformed to QALYs using the area-under-the-curve approach.
Scores from secondary outcomes will be reported in their natural units.

6.6. Economic Data Analysis

6.6.1. Analysis population

The full analysis set will include all participants, analysed in the groups which they were randomised, which is in accordance with the “intention to treat” (ITT) principle.

6.6.2. Timing of analysis

In accordance with the statistical analysis (section 3.6), the final analyses will be conducted after all participants have reached 12-month follow-up and the database is locked. The health economic analysis will be conducted blinded to allocation. Resource use, cost, and outcome data will be cleaned, validated, and analysed using anonymised group labels prior to unblinding. Treatment allocation will be revealed at the results meeting as described in Section 4. Intervention-specific costs will then be attributed to the relevant study groups, enabling calculation of final cost-effectiveness

outcomes, including incremental cost-effectiveness ratios (ICERs) and related uncertainty analyses.

6.6.3. Measurement of outcomes

The primary analysis will be conducted once all participants have completed 12-months follow-up.

6.6.4. Discount rates for costs and benefits

There will be no discounting of costs or effects as the trial final endpoint does not exceed 12-months.

6.6.5. Cost-effectiveness threshold

The estimated mean QALYs and costs associated with both the NEuRoMS intervention and usual care arms will be combined with a feasible range of values for decision makers' willingness-to-pay (λ), to obtain the distribution of net benefits at different levels of λ. Cost-utility analyses derived from EQ-5D-5L and MSIS-8D responses will use a cost-effectiveness threshold of £20,000 and £30,000 per QALY.

6.6.6. Statistical Decision Rules

Mean differences in costs, QALYs and net benefits between the treatment groups will be estimated with associated 95% confidence intervals.

6.6.7. Analysis of resource use and costs

Mean resource use will be estimated and presented by trial arm for each resource use category (e.g. outpatient visits, GP appointments, medication use etc). Standard deviations and the number of patients included in each category by arm will also be presented. Appropriate regression techniques will be used to estimate adjusted mean costs and the difference in adjusted mean costs (and their associated 95% confidence intervals) between the trial arms. The economic evaluation will be aligned with the statistical analysis, including using the same adjustment variables and modelling framework (e.g. mixed models to account for site effects).

If costs are not normally distributed, they may be modelled using Poisson or gamma distributions.

6.6.8. Analysis of outcomes

The primary economic outcome in the economic evaluation is Quality-Adjusted Life Years (QALYs). QALYs for each participant will be derived from utility values using the area-under-the-curve approach.

Appropriate regression techniques (e.g. seemingly unrelated regression) will be used to estimate QALYs and the difference in adjusted mean QALYs and their associated 95% confidence intervals between trial arms. Regression will be adjusted for the baseline values, and the stratification variable (e.g. site) Additional covariates may be incorporated into the analysis model (see Section 5.4) if the statistical analysis identifies important prognostic factors or reveals evidence of baseline imbalance.

6.6.9. Data cleaning for analysis

Face validity tests will be conducted on data (e.g. to identify misspelt text, implausible values) and checked against the source documents. Corrections identified will be documented in the Stata/R code. Data variables not required for the economic analysis and duplicate data entries will be dropped from the dataset.

6.6.10. Missing data

In accordance with the statistical analysis missing items within outcome measures will be dealt with in accordance with published guidelines. The appropriate method for dealing with missing data will depend on the prevalence and likely mechanism of missingness. As missingness in health economic data may be related to values of resource use, costs and health outcomes, failure to appropriately account for missingness could introduce bias within the formal analyses. Where data is considered missing at random (MAR), multiple imputation using chained equations will be used to impute missing items. Imputations will be combined using Rubins rules (40).

Item non-response on the EQ-5D-5L questionnaire will not be individually imputed. Partially completed questionnaires will be treated in the same manner as those which are fully incomplete with utility values imputed. Due to the predefined range of outcome scores from the EQ-5D-5L and MSIS-8D, and the typically skewed distribution of cost data, predictive mean matching (PMM) will be used to ensure imputed values are consistent with observed data.

6.6.11. Analysis of cost-effectiveness

If distributions of the QALYs and costs permit, we will adopt a seemingly unrelated regression model to account for potential correlation between costs and outcomes, while adjusting for baseline covariates.

For each perspective, if neither arm is dominant (i.e. less expensive and more effective), ICERs will be created using the outputs from the appropriate regression. These outputs will also be used to estimate the incremental net monetary benefit (INMB) statistic at the standard NICE willingness-to-pay thresholds of £20,000 and £30,000 per QALY.

For secondary outcomes, differences in mean outcomes alongside differences in mean costs will be examined. If neither the intervention or usual care group is dominant then an ICER will be calculated, reporting the incremental cost per unit change in outcome score.

6.6.12. Sampling uncertainty

Uncertainty will be addressed by combining the multiple imputation analysis with non-parametric bootstrapping. Bootstrap resamples will be drawn within each imputed dataset (e.g. 50 bootstrap samples from each of 20 imputed datasets to generate 1,000 bootstrap replications), and estimates combined across imputations using Rubin rules. The joint distribution of incremental costs and QALYs will be used to generate cost-effectiveness acceptability curves (CEACs) for a range of willingness-to-pay thresholds. This assesses the probability that the NEuRoMS intervention is cost-effective compared to usual care at a range of willingness-to-pay thresholds.

6.6.13. Subgroup analyses

In accordance with the statistical analysis, the primary and secondary cost-utility analyses (EQ-5D-5L and MSIS-8D) will be repeated separately for those with mild cognitive impairment and those with moderate cognitive impairment. These subgroup analyses are exploratory in nature, as the study is not powered to detect statistically significant differences within these groups. Instead, they will provided indicative insights of whether the cost-effectiveness of the intervention varies according to the severity of cognitive impairment.

Further exploratory sub-group analyses for the primary cost-utility outcomes will be conducted if there is sufficient evidence of interaction between cognitive impairment groups (mild versus moderate impairment), consistent with the planned statistical analysis. The primary economic analysis will be repeated with the addition of an interaction term between treatment group and cognitive impairment group (allocated group*cognitive impairment group).

6.6.14. Sensitivity analyses

Uncertainty in the methodological choices made for the present economic evaluation will be assessed through sensitivity analyses:

1. As explained in 6.3, a broader societal perspective will be undertaken on both the primary cost-utility (EQ-5D-5L), secondary cost-utility (MSIS-8D), and the secondary cost-effectiveness analyses to evaluate the impact of including patient and carer costs, and lost productivity.
2. Cost-effectiveness results derived from the EQ-5D will be evaluated using utility values derived from the crosswalk of EQ-5D-5L profiles to the EQ-5D-3L

value set to assess the robustness of the results to the utility valuation methodology.

3. Exploring the impact on the ICER of plausible changes to key methodological assumptions to understand how changes in the methodological assumptions impact on the cost-effectiveness result. This may include but is not limited to alternative sources or ranges of key unit costs, different approaches to the derivation of utility scores (e.g. EQ-5D-5L crosswalk vs. value set mapping), and exclusion of training costs in the calculation of intervention costs. These analyses will help assess the robustness of the cost-utility findings under different plausible scenarios.
4. In accordance with the statistical analysis (5.4), an additional sensitivity analysis will be conducted on the primary cost-utility (EQ-5D-5L) outcome to assess the impact of follow-up timing. Specifically, this analysis will exclude participants who completed their follow-up assessments outside the valid visit window. The valid visit window, defined as ±14 days from the scheduled follow-up date, will be applied in a consistent manner to the statistical analysis across the 3-month, 6-month, and 12-month follow-ups.
5. Exclusion of participants with insufficient intervention engagement. In accordance with the statistical analysis (5.4), a sensitivity analysis will be conducted on the primary cost-utility (EQ-5D-5L) analysis including only participants who completed the full intervention content prior to the three-month follow-up. Note that all participants in the control group will be retained.

6.6.15. Modelling of longer-term cost-effectiveness

If the trial demonstrates appropriate evidence of effectiveness, a decision analytic model will be constructed to explore longer-term cost-effectiveness. Structured literature searches will be used to obtain appropriate inputs beyond the trial with all assumptions verified by the trial team. Using the Markov model developed within the CRAMMS trial (41) as a potential starting point, we will construct a model based on the trial comparators with the choice of plausible time horizons (e.g. lifetime) agreed, using rules on whether sufficient data/assumptions can be made to project the trial effects into the future and if there is sufficient information available within the literature. A full modelling plan will be developed based on recommended good practice (42). A base-case ICER (cost per QALY) will be produced (and if appropriate lifetime horizon), with sensitivity analyses undertaken. We would consider the model important exploratory analysis, seeking to derive robust estimates of the longer-term cost-effectiveness whilst assessing the associated uncertainty. The development of the model depends on the trial finding and unblinding to determine

the feasibility of long-term extrapolation, and may be reported separately from the primary trial-based economic evaluation.

Whilst resources are not available within the budget requested to quantify the impact of the intervention (and pathway) reconfiguration on service capacity, introduction of cognitive screening at sites where screening is not currently part of treatment as usual, and work-load planning (for example), we will document these issues for consideration by commissioners as part of informing the implementation strategy in WP5 as selected scenarios. We will produce a trial-based budget impact analysis, and extrapolate to the UK population to estimate the likely impact of the intervention on NHS budgets, conducted according to recommended good practice (43).

6.7. Reporting/publishing standards

The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines (44) will be followed when reporting the health economic evaluation in a format appropriate for stakeholders and policymakers. Any deviations in the final health economic analysis from the published HEAP will be documented and justified in the final published report.

7. Software

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

7. References

1. Croft A, Mynors G, Roberts M, Doncaster D, Bowen A. MS Forward View: A consensus for the future of MS services. 2016.
2. MS Trust. Rehabilitation medicine consultants and MS: Report of a survey of rehabilitation consultants working in MS in the UK for the MS Forward View Project. 2015.
3. Mynors G, Bowen A, Doncaster D. MS Specialist Nursing in the UK 2016: Report on progress towards equitable provision. 2016.
4. Roberts M, Mynors G, Bowen A. Improving services for people with advanced MS. 2016.

5. Giovannoni G, Butzkueven H, Dhib-Jalbut S, Hobart J, Kobelt G, Pepper G, et al. Brain health: time matters in multiple sclerosis. Multiple Sclerosis and Related Disorders. 2016. doi:10.1016/j.msard.2016.07.003
6. Lezak MD. Neuropsychcological Assessment. 3rd ed. Oxford University Press; 1995.
7. Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The multiple sclerosis impact scale (MSIS-29) a new patient-based outcome measure. Brain. 2001. doi:10.1093/brain/124.5.962 PubMed PMID: 11335698.
8. Lincoln NB, Bradshaw LE, Constantinescu CS, Day F, Drummond AE, Fitzsimmons D, et al. Group cognitive rehabilitation to reduce the psychological impact of multiple sclerosis on quality of life: The CRAMMS RCT. Health Technol Assess (Rockv). 2020. doi:10.3310/hta24040 PubMed PMID: 31934845.
9. Jones KH, Ford D V., Jones PA, John A, Middleton RM, Lockhart-Jones H, et al. The physical and psychological impact of multiple sclerosis using the MSIS-29 via the web portal of the UK MS Register. PLoS One. 2013 Jan 31;8(1). doi:10.1371/JOURNAL.PONE.0055422 PubMed PMID: 23383186.
10. Russell D, Hoare ZSJ, Whitaker R, Whitaker CJ, Russell IT. Generalized method for adaptive randomization in clinical trials. Stat Med. 2011;30(9):922–34. doi:10.1002/sim.4175 PubMed PMID: 21284014.
11. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - A practical guide with flowcharts. BMC Med Res Methodol. 2017 Dec 6;17(1):1–10. doi:10.1186/s12874-017-0442-1 PubMed PMID: 29207961.
12. Bono C, Ried LD, Kimberlin C, Vogel B. Missing data on the Center for Epidemiologic Studies Depression Scale: A comparison of 4 imputation techniques. Research in Social and Administrative Pharmacy. 2007 Mar;3(1):1–27. doi:10.1016/j.sapharm.2006.04.001 PubMed PMID: 17350555.
13. Carlin JB, Galati JC, Royston P. A new framework for managing and analyzing multiply imputed data in Stata. Stata Journal. 2008;8(1):49–67. doi:10.1177/1536867x0800800104
14. Ritvo PG, Fischer JS, Miller DM, Andrews H, Paty DW, LaRocca NG. MSQLI Multiple Sclerosis Quality of Life Inventory: A User's Manual Developed by: The

Consortium of Multiple Sclerosis Centers Health Services Research Subcommittee.

15. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: Initial validation of the fatigue impact scale. Clinical Infectious Diseases. 1994. doi:10.1093/clinids/18.Supplement_1.S79 PubMed PMID: 8148458.
16. Veit CT, Ware JE. The structure of psychological distress and well-being in general populations. J Consult Clin Psychol. 1983. doi:10.1037/0022-006X.51.5.730
17. Sullivan MJL, Edgley K, Dehoux E. A survey of multiple sclerosis. Part 1: Perceived cognitive problems and compensatory strategy use. Canadian Journal of Rehabilitation. 1990.
18. Sullivan M, Edgley K, Dehoux E. A survey of multiple sclerosis: I. Perceived cognitive problems and compensatory strategy use. undefined. 1990.
19. 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). Quality of Life Research. 2011. doi:10.1007/s11136-011-9903-x PubMed PMID: 21479777.
20. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a Brief Depression Severity Measure. J Gen Intern Med. 2001;16(9):606. doi:10.1046/J.1525-1497.2001.016009606.X PubMed PMID: 11556941.
21. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. Arch Intern Med. 2006 May 22;166(10):1092–7. doi:10.1001/ARCHINTE.166.10.1092 PubMed PMID: 16717171.
22. Nicholl CR, Lincoln NB, Playford ED. The reliability and validity of the Nottingham Extended Activities of Daily Living Scale in patients with multiple sclerosis. Multiple Sclerosis. 2002. doi:10.1191/1352458502ms827oa PubMed PMID: 12356202.
23. Rigby SA, Domenech C, Thornton EW, Tedman S, Young CA. Development and validation of a self-efficacy measure for people with multiple sclerosis: The

Multiple Sclerosis self-efficacy scale. Multiple Sclerosis. 2003. doi:10.1191/1352458503ms870oa PubMed PMID: 12617272.

24. D'Souza E. Modified Fatigue Impact Scale - 5-item version (MFIS-5). Occup Med (Lond). 2016 Apr 1;66(3):256–7. doi:10.1093/OCCMED/KQV106 PubMed PMID: 27016749.
25. Honan CA, Brown RF, Hine DW. The Multiple Sclerosis Work Difficulties Questionnaire (MSWDQ): Development of a shortened scale. Disabil Rehabil. 2014. doi:10.3109/09638288.2013.805258 PubMed PMID: 23786346.
26. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. Quality of Life Research. 2012. doi:10.1007/s11136-011-9927-2 PubMed PMID: 21598064.
27. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. International Journal of Surgery. 2012;10(1):28–55. doi:10.1016/j.ijsu.2011.10.001 PubMed PMID: 22036893.
28. Golden CJ. Stroop Color and Word Test: A Manual for Clinical and Experimental Uses. Chicago, IL: Stoelting; 1978.
29. Golden CJ, Freshwater SM. The Stroop Color and Word Test : A Manual for Clinical and Experimental Uses [Internet]. Chicago, IL: Stoelting; 2002 [cited 2023 Apr 11]. Available from: https://www.worldcat.org/title/stroop-color-and-word-test-a-manual-for-clinical-and-experimental-uses/oclc/811788223
30. Young CA, Rog DJ, Sharrack B, Tanasescu R, Kalra S, Chhetri SK, et al. Physical and psychological aspects of multiple sclerosis: Revisiting the Multiple Sclerosis Impact Scale (MSIS-29). Mult Scler. 2024 Nov 1;30(13):1630. doi:10.1177/13524585241288393 PubMed PMID: 39474866.
31. Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, et al. Unit Costs of Health and Social Care 2024 Manual [Internet]. 2025 Apr 8 [cited 2025 Apr 16]. Available from: https://www.pssru.ac.uk/unitcostsreport/
32. NHS England » National Cost Collection for the NHS [Internet]. [cited 2025 Apr 16]. Available from: https://www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/

33. Prescription Cost Analysis – England 2023/24 | NHSBSA [Internet]. [cited 2025 Apr 16]. Available from: https://www.nhsbsa.nhs.uk/statistical-collections/prescription-cost-analysis-england/prescription-cost-analysis-england-202324
34. Additional guidance for the 2025 Annual Survey of Hours and Earnings Secure Electronic File Transfer - Office for National Statistics [Internet]. [cited 2025 Apr 16]. Available from: https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/additionalguidanceforthe2020annualsurveyofhoursandearnings
35. 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 Dec;20(10):1727–36. doi:10.1007/S11136-011-9903-X PubMed PMID: 21479777.
36. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res. 2012 Feb;21(1):167–76. doi:10.1007/S11136-011-9927-2 PubMed PMID: 21598064.
37. Overview | Interim methods statement: implementing the EQ-5D-5L value set | Guidance | NICE.
38. Overview | NICE technology appraisal and highly specialised technologies guidance: the manual | Guidance | NICE.
39. Goodwin E, Green C, Spencer A. Estimating a Preference-Based Index for an Eight-Dimensional Health State Classification System for Multiple Sclerosis. Value in Health. 2015 Dec 1;18(8):1025–36. doi:10.1016/J.JVAL.2015.10.004 PubMed PMID: 26686787.
40. Donald B. Rubin. Multiple Imputation for Nonresponse in Surveys | Wiley [Internet]. 2004 [cited 2025 Apr 16]. Available from: https://www.wiley.com/en-gb/Multiple+Imputation+for+Nonresponse+in+Surveys-p-9780471655749
41. Lincoln NB, Bradshaw LE, Constantinescu CS, Day F, Drummond AE, Fitzsimmons D, et al. Group cognitive rehabilitation to reduce the psychological impact of multiple sclerosis on quality of life: the CRAMMS RCT. Health Technol Assess. 2020;24(4):1. doi:10.3310/HTA24040 PubMed PMID: 31934845.

42. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Value in Health. 2012 Sep 1;15(6):796–803. doi:10.1016/J.JVAL.2012.06.012 PubMed PMID: 22999128.
43. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget Impact Analysis—Principles of Good Practice: Report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health. 2014 Jan 1;17(1):5–14. doi:10.1016/J.JVAL.2013.08.2291 PubMed PMID: 24438712.
44. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMC Med. 2022 Dec 1;20(1). doi:10.1186/S12916-021-02204-0/TABLES/2 PubMed PMID: 35022047.

Referenced documents:

1. NEuRoMS WP4 Data Management Plan
2. NWORTH SOP 5.03 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 4 = Almost Always, total score and subscale scores are sum of scores for relevant items The PDQ-5 consists of items 4, 9, 11, 14 and 16, the total score is the sum of the item scores	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 = nearly 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 = Nearly 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
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

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
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'. Items 3, 4, 9, 10, 12, and 14 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 0 'Never' to 4 '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
Three single-item questions for medication adherence, work and education	Extent that medication adherence, work and 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

Outcome measure	Domains covered	Scoring	Subscales	Missing value rules	Thresholds
MFIS-5	Measure of fatigue	Scale from 0 = Never to 4 = Almost Always, total score is sum of item scores	None	None	None
MHI-5	Measure of mental health	Scale from 6 = All of the time to 1 = None of the time, total score is sum of item scores Items 1, 3, 5 need to be reverse scored	None	None	None
GAS	Goal attainment scale – measures the degree to which each goal has been met	Four-point Likert scale (i.e., 'not met at all', 'met a little', 'mostly met', and 'fully met')	None	None	None

Appendix 2 – Assumptions of Linear Refression and Multilevel Mixed Effects Linear Regression Models

Assumption	How to check	Action if violated
1. Residuals of model are approximately normally distributed 2. Homogeneity of variance	1. Shapiro-Wilk test for normality of residuals 2. Levene's robust test statistic for equality of variance	If either condition violated, consider transforming the data in order to satisfy the assumptions of the test. 1. 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 2. 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
1. The covariate is linearly related to the dependent variable at each level of the independent variable 2. There is homogeneity of regression slopes (there is no interaction between the independent variable and the covariate)	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 linear mixed effects model
		In the event of the assumptions not being met we will also consider use of a generalised linear model with gamma distribution and inverse link function in the first instance. If the generalised linear model with gamma distribution and inverse link function are not appropriate to satisfy the violations of assumptions of the linear mixed model then alternative distributions and link functions will be considered. For Likert scales, if the distributional characteristics indicate that an ordinal model would be more appropriate then an ordinal model will be used.

Appendix 3 – Health economic figures and tables

This table provides an indication of the tables that the health economic analysis will produce. This list is not definitive; additions and amendments to this list may be made in the final report.

- Table 6.1.: Resources collected and valuation
- Table 6.2.: Missing data percentage by variable and trial arm
- Table 6.3.: Resources and costs associated with the delivery of the NEuRoMS intervention and usual care
- Table 6.4.: Resource use by trial arm
- Table 6.5.: Summary of unadjusted outcomes by time point and trial arm
- Table 6.6.: Primary cost-utility analysis from EQ-5D-5L
- Table 6.7.: Secondary cost-utility analysis from MSIS-8D
- Table 6.8.: Secondary analyses (e.g. cost-effectiveness/cost-consequence analysis from ICECAP-A)
- Table 6.9.: Sensitivity analysis on primary and secondary cost-utility analyses
- Table 6.10.: Sensitivity analysis on secondary outcomes
- Table 6.11.: Longer term cost-effectiveness
- Figure 6.1.: CEAC for primary cost-utility analysis from EQ-5D-5L
- Figure 6.2.: CEAC for secondary cost-utility analysis from MSIS-8D
- Figure 6.3.: CEAC for longer-term cost utility analysis