BMJ Open

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029565
Article Type:	Protocol
Date Submitted by the Author:	31-Jan-2019
Complete List of Authors:	Seebacher, Barbara; Medical University of Innsbruck, Clinical Department of Neurology Reindl, Markus; Medical University of Innsbruck, Clinical Department of Neurology Zamarian, Laura; Medical University of Innsbruck, Clinical Department of Neurology Kuisma, Raija; University of Brighton, School of Health Sciences Mills, Roger J; Lancashire Teaching Hospitals NHS Foundation Trust, Department of Neurology Kircher, Simone; University Hospital of Innsbruck, Clinical Department of Neurology Brenneis, Christian; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Ehling, Rainer; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Deisenhammer, Florian; Medical University of Innsbruck, Clinical Department of Neurology
Keywords:	Multiple sclerosis < NEUROLOGY, self-efficacy, Unidimensional Self-efficacy Scale for Multiple Sclerosis, German version, cross-cultural adaption for Austria, Rasch analysis

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Barbara Seebacher[1], Markus Reindl[1], Laura Zamarian[1], Raija Kuisma[2], Roger J Mills[3], Simone Kircher[4], Christian Brenneis[5, 6], Rainer Ehling[5, 6], Florian Deisenhammer[1]

[1]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria

[2]School of Health Sciences, University of Brighton, Eastbourne, UK

[3]Department of Neurology, Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

[4]University Hospital of Innsbruck, Clinical Department of Neurology, Innsbruck, Austria

[5]Department of Neurology, Clinic for Rehabilitation Münster, Münster, Austria

[6]Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster, Münster, Austria

Correspondence to Dr Barbara Seebacher; Clinical Department of Neurology, Medical University of Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria; phone +43.512.504.82499; barbara.seebacher@i-med.ac.at

Word count: 3981 words

ABSTRACT

Introduction: Self-efficacy refers to individuals' confidence in their ability to perform relevant tasks to accomplish desired goals. This is independent of their actual abilities. In people with multiple sclerosis (PwMS), self-efficacy has been shown to powerfully influence motivation and health-related behaviour, such as adherence to prescribed treatment or physical activity. So far, a rigorously tested German language self-efficacy questionnaire for PwMS has not been found.

Methods: Based on Bandura's concept of self-efficacy and international guidelines for questionnaire development, this study will translate the original Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) into German. The patient-led development of the pre-final German version will involve a forward-backward translation process, synthesis of translations, expert committee review and consensus with the original test developers. At two centres in the region of Tyrol, Austria, content and face validity and cultural adaption for Austria will be established using face-to-face semi-structured cognitive interviews of 30 PwMS. A further 292 PwMS with minimal to severe disability will be tested at two time-points to validate the German USE-MS (USE-MS-G).

Analysis: Mixed methods analyses will be applied. Interviews will be transcribed and analysed employing qualitative content analysis. External validity will be explored using Spearman's Rank correlation coefficients of the USE-MS-G with the 13-item Resilience Scale, General Self-Efficacy Scale, Multiple Sclerosis International Quality of Life questionnaire, Hospital Anxiety and Depression Scale and MS-specific Neurological Fatigue Index. Test-retest reliability, internal consistency and floor and ceiling effects will be evaluated. Internal validity will be examined using Rasch analysis.

Ethics and dissemination: Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Results from this study will be disseminated to the participants via mail and MS Societies (Austria, Germany, Switzerland and UK), and to clinicians and researchers through peer-reviewed publications and conferences.

Study registration: ISRCTN Registry; trial ID ISRCTN14843579; prospectively registered on 02. 01. 2019; http://www.isrctn.com/ISRCTN14843579

Study protocol, version 1, 31.1.2019

Keywords: multiple sclerosis; self-efficacy; Unidimensional Self-efficacy Scale for Multiple Sclerosis; German version; cross-cultural adaption for Austria; Rasch analysis

ARTICLE SUMMARY

Strengths and limitations of this study

- This study protocol describes the German translation of the original English language Unidimensional Self-efficacy Scale for Multiple Sclerosis (USE-MS), upon permission of the scale developers and applying international recommendations.
- Consistent with the conceptual framework of the English USE-MS, Bandura's concept of self-efficacy will be adhered to.
- Employing a patient-led process in phase 1, 30 people with MS (PwMS) will be interviewed about the pre-final German USE-MS, to establish face and content validity and cultural adaption for PwMS in Austria.
- In phase 2, the German USE-MS will be validated in a larger sample of 292 PwMS.
- Applying classical test theory and Rasch analysis approaches, internal and external validity, internal consistency and test retest reliability will be explored.

INTRODUCTION

Multiple sclerosis (MS) is one of the most common neurological diseases in young adults worldwide, with increasing prevalence[1]. MS is characterised by a wide variety of symptoms and different disease courses[2]. Despite the development of novel disease modifying drugs and neurorehabilitation strategies, the unpredictability of the disease with psychological distress, losses in social contact and quality of life (QoL) are concerning for PwMS. However, individuals' self-knowledge can modulate their approach to day-to-day activities. According to Bandura's social cognitive theory, psychosocial functioning is regulated by reciprocal interactions between behaviour, personal factors and environmental conditions[3]. Self-regulation and intrinsic motivation enable individuals to set and pursue their own goals, observe and evaluate themselves in relation to attained goals[4]. Bandura defined self-efficacy as individuals' beliefs regarding their capability to perform significant tasks, to achieve goals that are meaningful for their daily lives[3]. Self-efficacy beliefs considerably influence people's feelings, thoughts and motivation[5] while, notably, being independent of their physical performance[5]. Such a concept appears important for people with disabilities because it may shape their motivation to initiate and adhere to treatment, particularly when facing side effects.

Perceived self-efficacy influences health-related behaviour such as adhering to medication[6] or engaging in physical activity in PwMS[7]. Health status evaluations of responses to rehabilitation and steroid treatment after an MS relapse can be predicted by self-efficacy levels[8]. Also, higher self-efficacy levels are associated with better long-term perceived cognitive functioning[9] and QoL[10, 11]. PwMS who report higher perceived self-efficacy also state lower levels of fatigue, depression and anxiety[12]. Recent evidence has provided insight into the importance of self-management and intrinsic motivation for motor learning[13]. Recognising the relevance of self-efficacy especially for people with disabilities, valid and reliable measurement tools are still needed for its assessment. Three generic self-efficacy scales were found in the literature[7, 14-16]. However, generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS. The initial impact of a diagnosis of MS, in addition to the manifold symptoms and necessity of managing a progressive disease may affect individuals' self-efficacy perceptions. Studies

demonstrated that the capability to effectively solve problems, consistent with higher self-efficacy levels, is strongly associated with PwMS' psychological adaptation to their disability[17], supporting the choice of a disease-specific over a generic self-efficacy questionnaire. MS-specific self-efficacy scales include the Liverpool Self-efficacy Scale (LSES)[18], Multiple Sclerosis Self-Efficacy Scale (MSSS)[19], MS Self-Efficacy Scale (MSSE)[20], Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS)[21] and University of Washington Self-Efficacy Scale for people with disabilities[22].

Following current guidelines, patients should be involved in the translation and development process of disease specific questionnaires, to ensure the scale reflects their experiences[23]. LSES and MSSS development used in-depth patient interviews while the USE-MS consists of items from both the LSES and MSSS. Bandura's concept of self-efficacy is reflected in the wording of all three questionnaires. The USE-MS study sample was the largest thereof (N=303), and only the USE-MS was exposed to Rasch analysis assessing internal construct validity, in addition to conventional external construct validity and reliability testing. Fit to the Rasch model was demonstrated, and good external validity and reliability[21]. Consequently, the USE-MS appears to be appropriate for use in clinical practice and research. However, so far no validated German language version of the USE-MS is available. The purpose of this study will therefore be to translate the USE-MS into German and validate the German language version in a larger sample of PwMS.

METHODS

Study aims

The first aim of this patient led study is to translate the original English USE-MS, developed by Young et al. (2012) into German, based on international guidelines.

The second aim is to establish face and content validity and cultural adaption of the German version for PwMS in Austria, using individual semi-structured cognitive interviews.

The third aim is to evaluate internal and external validity, internal consistency and test-retest reliability of the German USE-MS (USE-MS-G), using classical test theory and Rasch analysis.

Study design

This will be a bi-centre prospective cross-sectional translation and validation study with repeated measures, consisting of Phase 1 and Phase 2. The SPIRIT 2013 and SPIRIT-PRO Extension checklist for study protocols[24] is presented as Supplementary File 1.

Study setting and timeline

Locations will be the outpatient MS-Clinic of the Clinical Department of Neurology, Medical University of Innsbruck, Austria and Department of Neurology, Clinic for Rehabilitation Münster, Austria.

The expected overall study duration is 33 months, from 01.02.2019 to 31.10. 2021.

Participants and recruitment

A random cross-sectional cohort of patients with clinically definite MS will be recruited from the two centres. The study will be advertised in the MS-Clinic, the Rehabilitation Centre and on the Austrian MS Society website. Further interested PwMS will be examined for eligibility by neurologists at the two study locations. Severely disabled PwMS (EDSS ≥8) will be offered home visits to enable their participation. Written informed consent will be obtained either by neurologists (FD, CB and RE) or the first author (BS). Participants may withdraw from the study at any time and for any reasons without prejudice. Outpatient participants will be reimbursed for travel expenses only. Inclusion and exclusion criteria are listed in Table 1.

Table 1 Inclusion and exclusion criteria

Inclusion criteria	**Exclusion criteria**
People with any MS phenotype according to the version of the McDonald's criteria valid at the time of diagnosis	Concomitant diseases which may affect subjective self-efficacy ratings (e.g. malignant diseases, other neurological or psychiatric disorders)
Aged 18 years or over	A relapse of MS within the last two months
Any ethnicity	Any medication change within four weeks prior to the study
Disability status score on the Expanded Disability Status Scale (EDSS) of 0 to 9.0	A relapse between testing 2 and 3 would necessitate the exclusion of the participant.
Able to speak and understand German language, or German language as the first spoken language	

References: McDonald's criteria[25-27]; Expanded Disability Status Scale[28].

Ethics approval, permissions and dissemination plan

Ethical approval for both centres was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Due to the absence of an intervention, no insurance policy is required for this study and no harm to participants is expected.

Permission to translate into German and validate the original USE-MS[21] was provided by the test developers who hold the copyright for the USE-MS-G.

Results from this study will be disseminated to the participants via mail and MS Societies (Austria, UK). Findings will be disseminated to clinicians and researchers through peer-reviewed publications and conferences.

Patient and public involvement

In Phase 1, patients will be lay members of the expert committee to consolidate all the translations and back translations of the USE-MS. Their role regarding the item and response option wording and sentence structure will be crucial, as the final questionnaire should be understood by PwMS. Patients will also be involved using face-to-face cognitive interviewing, to gain insight into their views about the clarity of the wording, meaning and completeness of the questions of the pre-final USE-MS-G. The Austrian MS (recruitment) and MS Research Societies (funding) will be involved in this study, with whom the findings will be shared as soon as available (patient magazine, meetings). The findings will also be disseminated to the UK MS Society and MS Trust.

Sample size

Phase 1

Using a power of 80% (β=0.2) and a 5% type I error rate, a sample size of 27 participants is necessary to detect a problem within a questionnaire, which occurs at a prevalence of 0.06% (1 in 1666.67)[29]. Thus, including a 10-12% attrition rate, 30 participants will be required. As recommended for qualitative interviewing, it is expected that 30 participants will be sufficient to reach saturation[30].

Phase 2

Rasch analysis sample size requirements are predicated upon the degree of precision required for estimating item and person difficulties. Regardless of targeting, one can be 99% confident that a sample size of 243 participants is adequately large to obtain a (high) precision of ±0.5 log odd units (logits). Good targeting provided, a sample size of 108 people would be sufficient[21, 31]. Hence, including a 20% attrition rate, 292 participants will be aimed at in this study.

Outcomes and data collection

Assessments used in this study were developed using patient involvement and/or recommended by governmental or patient organisations (Supplementary File 2). Study outcomes and methods for their assessment are presented in Figure 1. Participant characteristics and assessments used at all time-points are shown in Table 2.

Figure 1 around here

Figure 1 Study outcomes and their assessment

Figure legend: GSE: General Self-Efficacy Scale; RS-13: Resilience Scale, short version; MusiQol: Multiple Sclerosis International Quality of Life questionnaire; HADS: Hospital Anxiety and Depression Scale; NFI-MS: Neurological Fatigue Index.

Table 2 Participant characteristics and assessments used in this study

Participant characteristics and assessments (assessments will be collected in a random order to avoid order effect)	**Phase 1**	**Phase 2**	
	T1	**T2**	**T3**
Participant identifier (ID)	**X**	**X**	**X**
Age	**X**	**X**	
Gender	**X**	**X**	
MS phenotype[1]	**X**	**X**	
Disease duration	**X**	**X**	
Expanded Disability Status Scale (EDSS)[2]	**X**	**X**	
Disease modifying treatment (DMT)[3]	**X**	**X**	**X**
(Pre-final) German version of Unidimensional Self-Efficacy Scale for Multiple Sclerosis	**X**	**X**	**X**
Qualitative cognitive interview	**X**		
Resilience Scale, short version		**X**	**X**
General Self-Efficacy Scale		**X**	**X**
Multiple Sclerosis International Quality of Life questionnaire		**X**	**X**
Hospital Anxiety and Depression Scale		**X**	**X**
Neurological Fatigue Index		**X**	**X**

[1]Relapsing-remitting; primary progressive; secondary progressive multiple sclerosis[32]

[2]EDSS groups: 0-4.0; 4.5-6.5; 7.0-7.5; 8.0-9.0[28]

[3](a) No DMTs; (b) low effective DMTs: interferon-b 1a, interferon-b 1a, interferon-b 1b, pegylated interferon-b 1a, glatiramer acetate, dimethyl fumarate, teriflunomide, azathioprin, intravenous immunoglobulins; (c) high effective DMTs: alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, cyclophosphamide, mitoxantrone,

rituximab[33, 34]

At recruitment, disability will be assessed by neurologists (FD, CB or RE) using the EDSS, ranging from 0 to 10, with higher scores representing higher levels of disability[28]. Although psychometric validation studies criticised its low responsiveness to changes, the EDSS has no floor or ceiling effects[35], has been shown to be valid and reliable[36] and is therefore recommended for use in clinical studies[37].

Excellent internal and external validity and reliability of the original USE-MS has been shown[21]. Scoring of the USE-MS draws results from all 12 items while items 7, 8, 10, 11 and 12 are inversely scored. Higher numbers represent stronger self-efficacy beliefs in participants[21]. The USE-MS includes a 4-point Likert scale (1= strongly disagree to 4=strongly agree).

To assess external construct validity, the following questionnaires will be administered:

The validated German version[38] of the 10-item General Self-Efficacy Scale (GSE)[15] is a self-administered 4-point Likert scale with a summary score ranging from “not at all true” to “exactly true”. The total GSE score ranges between 10 and 40, higher scores signifying greater self-efficacy. Psychometric testing demonstrated high internal consistency, moderate concurrent validity and unidimensionality[15].

The validated German version[39] of the 13-item Resilience Scale (RS-13)[40], based on the 25-item Resilience Scale[41] will be used. RS-13 item scores from a 7-point Likert scale are added up, indicating low (13-66 points), moderate (67-72 points) or high (73-91 points) resilience[40]. The German RS-13 showed high internal consistency and moderate test-retest reliability. Confirmatory factor analysis indicated an acceptable model fit[40].

The validated German version[42] of the 31-item Multiple Sclerosis International Quality of Life (MusiQol) questionnaire[43] will be employed. Response options use a 6-point Likert scale, from 1= “never/not at all” to 5= “always/very much” and 6= “not applicable”. Negatively worded item scores are reversed, and for each participant mean scores for each dimension of the item scores are calculated. All nine dimension scores are linearly transformed to a 0-100 scale, their mean representing the global index score, 0 indicating the worst level of health-related QoL and 100 the best. Psychometric testing

showed satisfactory internal and external validity and acceptable reliability for all MusiQol dimensions[43].

The validated German version[44] of the 14-item Hospital Anxiety Depression Scale (HADS)[45] will be used. The HADS is a self-report questionnaire with a 4-point Likert scale and a 42 point maximum, higher scores representing higher levels of anxiety or depression. Odd items are added to score the anxiety subscale (0-21 points), and even items are added to generate the depression subscale (0-21 points). Testing of the German version demonstrated good internal consistency and acceptable test-retest reliability [44]. The two-factor structure of the scale was confirmed[44].

The validated German version[46] of the 23-item Neurological Fatigue Index (NFI-MS) will be used[47]. Four factors of the NFI-MS were confirmed by principal component analysis and explained 62% of the variance. The four subscales and total scale showed acceptable responsiveness[48], good test-retest reliability, moderate convergent validity and fit to Rasch model expectations[47]. Items are scored on a 4-point Likert scale from 0= “strongly disagree” to 3= “strongly agree”. For scoring, the following item values are added: 1-8= “physical subscale”; 9-12= “cognitive subscale”; 13-18= “relief by diurnal sleep or rest subscale”; 19-23= “abnormal nocturnal sleep and sleepiness subscale”; and 1-7, 9, 11-12 = “physical and cognitive summary score”[47].

Assessments will be performed by trained physiotherapists holding a Master’s (SK) and PhD degree (BS) and a clinical neuropsychologist (LZ). The number of participants who decline to participate or drop out will be recorded, together with reasons (CONSORT flow chart). Any health problems will be recorded.

Phase 1: data will be collected at one time-point (Testing 1, T1), with an expected duration of 45-60 minutes.

Phase 2: for the test-retest reliability assessment, data will be collected at two time-points and will last 60-90 minutes: Testing 2 (T2) and Testing 3 (T3), 14-21 days after T2[47, 49].

Data management

With regard to confidentiality, the Austrian and Tyrolean Data Protection Acts will be adhered to. Double data entry and range checks for data values will be used. For qualitative content analysis, double coding of the data set will be performed. Only the

research team will have access to the data. All personal data will be anonymised by a participant ID. Data and files will be saved on a password protected computer, will not be transferred via emails and will be only used for the purposes for which they were collected. Participants will be informed about their right to disclosure for their own data even if these data lack clinical utility. Anonymised data will be kept for 15 years following completion of the study. Blank data collection forms can be requested from the corresponding author.

Study procedures

This study will follow the Beaton et al. guidelines for the cross-cultural adaptation of patient reported outcomes[50] and its enhanced version from the University of Leeds, UK.

Phase 1

Stage 1: Forward translation of the items, response options, instructions and scoring information into German will be performed by three independent translators; translator 1 is a medical professional and informed about self-efficacy, while translators 2 and 3 have no medical knowledge and are “naïve” to self-efficacy. Translators are bilingual German native speakers and will create a written report for all translations (T1, T2 and T3), which will then be compared, to distinguish any wording differences or ambiguities[51].

Stage 2 will be a synthesis of T1-3 into T-123. Involving a fourth, unbiased person, the three versions will be discussed with the translators and any discrepancies solved by consensus. A revised questionnaire and comprehensive report will be produced[50].

Stage 3: Backward translation of T-123 into English will be done by three bilingual English native speakers who are blind to the original version. Translators are “naïve” to self-efficacy and medicine, to minimise bias[51]. Vague wording, obvious inconsistencies or theoretical errors in the translations shall be detected. A report for each version, TB1, TB2 and TB3, will be written by the translators. To maximise comprehension, language will be used which can be understood by a 12 year old[52, 53], indicated by a Flesch reading-ease score of 80-90[54]. The German Flesch value=180-ASL-(58,5*ASW), where ASL=average sentence length and ASW=average number of syllables per word[54].

Stage 4: Considering written documentations, an Expert Committee will review and integrate all versions of the questionnaire, involving instructions and scoring documentation, and develop the pre-final version of the USE-MS-G. The Expert Committee will consist of three neurologists, two physiotherapists, a neuropsychologist, a methodologist, two language professionals, the translators, three lay PwMS and the translation synthesis recorder. The Expert Committee will be in close contact with the original USE-MS developers. A written report of the consensus process will be created. Decision-making will be based on guidelines to accomplish cross-cultural equivalence between the original and German versions in four areas[51], shown in Figure 2.

Figure 2 around here

Figure 2 Cross-cultural equivalence areas to be achieved between original and German USE-MS (USE-MS-G); adapted from[51]

Stage 5: Pretesting of the pre-final USE-MS-G will be performed in 30 PwMS, involving completion of the scale and face-to-face cognitive interviews. Cognitive interviewing will be used to evaluate whether survey questions are easily comprehended, response categories match natural responses, and if people are motivated to respond truthfully and accurately[55-57]. Leading questions will be avoided to minimise bias. Enquiries for comprehension and meaning will be used, and repetition of content by patients[55, 57]. Probing will be applied to explore cognitive processes such as memory, underlying reasons for certain responses and overall level of difficulty or confidence[56]. Verbal probes, following Willis' model, will be used immediately after the questions[58]: (a) standardised, anticipated probes: scripted; (b) standardised, conditional probes: scripted, but will be used only if activated by certain participant behaviors such as hesitation[59]; (c) non-standardised, spontaneous probes: flexible, at researcher's digression; and (d) non-standardised, emergent probes: applied in reaction to participant behaviour[60]. The interview guide is presented in Table 3. Recording and field notes will be used, reviewed for inconsistencies or gaps shortly before the end of the interview.

Table 3 Questions used for semi-structured interview

Participants will be given sufficient time to complete the pre-final German USE-MS.	
1.	Having read the questions in the questionnaire, what are your thoughts about them?
2.	Would you please repeat this question in your own words?
3.	What do you think this question is asking?
4.	What do you think about that particular question?
5.	What do you think about the wording of this question, in terms of its clarity?
6.	How easy or hard was this to answer?
7.	How sure are you of your answer?
8.	Could you talk me through your answers in more detail?
9.	What were you thinking of when you answered this question?
10.	Do you have any other comments?
11.	If responses from participants are somewhat unclear, the interviewer asks: “Why so?”
12.	Should a participant hesitate, the interviewer conveys: “You spent some time answering that question - what were you thinking about?”

Adapted from [56] and [61]

An overview of study procedures is presented in Figure 3.

Figure 3 around here

Figure 3 Flowchart of the study procedures

Figure legend: MS: multiple sclerosis; T1 (2; 3): testing 1 (2; 3).

Phase 2

The USE-MS-G will be validated in a larger sample of 292 PwMS who will complete the above described questionnaires at T1 and T2.

Data analyses

Mixed methods data analyses will be used.

Phase 1- Qualitative analyses

Interviews will be transcribed and analysed using qualitative content analysis (QCA)[62, 63] using QDA MINER LITE software (Provalis Research, Montreal, Canada) and adhering to the Consolidated Criteria for Reporting Qualitative Research (COREQ)[64]. Analysis steps will be performed as follows[65-68]:

- Data organisation based on the research question
- Identification of recurring ideas, concepts, themes and words
- Development of a coding frame (requirements: unidimensionality, mutual exclusivity of subcategories within dimensions, exhaustiveness of subcategories and saturation, where each subcategory is used at least once)
- Selection of relevant material, structuring, marking and segmentation of text sections, based on Bandura's concept of self-efficacy and the original USE-MS, to identify main and subcategories
- Definition, naming and characterisation of categories and decision rules, to enable consistent assignment of data segments
- Illustration of categories and subcategories using citations
- Creation of a data matrix, followed by quantitative data analysis (descriptive statistics, e.g. frequencies)
- Report
- Rigor and credibility will be maximised by[69-71]

- systematic and consistent approach throughout the analysis
- revision and expansion of the coding frame
- double coding of the whole dataset by two independent researchers (10-14 days after initial coding)
- checking for researcher effects (reflexivity)[72]

Phase 2 - Quantitative analyses

Descriptive statistics and reliability estimates will be performed using IBM SPSS software, release 25.0 (IBM Corporation, Armonk, NY, USA). Rasch Analysis will be conducted with RUMM2030 software[73]. Statistical significance is defined as two-tailed p value <0.05.

Missing data will be treated as follows:

1) Missing data should be avoided by checking questionnaires for missing item responses and asking participants for completion.
2) Rasch analysis calculates an estimate from all available data and does not require a complete data set[74].

Test-retest reliability

Test-retest reliability will be evaluated using Lin's concordance correlation coefficient (r_c) between T2 and T3 (0-1)[75, 76]. R_cs will be calculated with their 95% confidence intervals (CI). Values of <0 will be considered to indicate poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1 almost perfect agreement[77].

External validity

It is hypothesised that scores on the German USE-MS will demonstrate moderate to high positive correlations with scales assessing conceptually similar constructs (convergent validity; with the GSE, RS-13 and MusiQol) and moderate to high negative correlations with scales measuring divergent constructs (divergent validity; with the HADS and NFI-MS); Spearman's Rank correlation coefficients of 0.3-0.49 being considered low, 0.5-0.69 moderate and ≥0.7 strong[78].

Internal validity - Rasch analysis

Rasch analysis[79] assumes the probability of a person endorsing an item is a logistic function of the difference between the "person ability" (perceived self-efficacy) and the "item difficulty" (level of self-efficacy) expressed[80]. Item characteristic curves, arranged on the log-odds units (logit) scale, will be used to visualise the probability of a person's correct response in relation to the item difficulty[81].

The polytomous Rasch model will be chosen for this study, suitable for scales with multiple response categories for their items[82]. A significant likelihood ratio test signifying inconsistent distance between response category thresholds would require the use of Masters' unrestricted (partial credit) model[83], otherwise Andrich's rating scale model[84]. Category thresholds are located centrally between two adjacent categories where either response is equally likely[82, 85]. The 4-point USE-MS includes three thresholds.

Ordered item category thresholds

Category probability curves will be inspected, checking regular distribution and monotonic advance of measures across categories[85].

Targeting

Targeting refers to the degree to which the scale captures the full range of self-efficacy. Inspecting person-item threshold distribution maps, the mean location score for the respondents will be compared with the default items zero value. A well-targeted scale is centred around zero logits (±0.5 logits), corresponding to the scale's item of mean difficulty[86].

The proportion of floor and ceiling effects will be monitored, considered noteworthy if >5%[87].

Local independence

Local independence means there should be no associations between the items. Inspection of the correlation matrix of item standardised residuals should show Pearson's correlations of <0.2 above the mean value of the matrix as a whole.

Unidimensionality

Unidimensionality as a Rasch model requirement allows a summary score measurement of a single construct. Using a PCA of the residuals, positively and

negatively loadings of the first component will be identified, generating two subsets and separate person estimates. Independent t-tests will explore significant differences[88]. If less than 5% of t-tests are significant or the lower bound of the binominal CI overlaps 5%, unidimensionality is supported[89, 90].

Fit to the Rasch model

Different fit statistics will seek to determine if the assumption of a probabilistic ordering of items is satisfied:

(a) Summary Chi-square interaction statistics and individual item Chi-square statistics are expected to be non-significant (Bonferroni-adjusted p-values for the number of items)[47].

(b) Individual person and item fit residuals are expected to be between ±2.5 (99% CI)[91].

(c) Person and summary item fit residuals reflect perfect model fit if their mean and standard deviation are close to 0 and 1, respectively[92, 93].

Reliability

Reliability is indicated by the person separation index (PSI; range 0-1)[94] and Cronbach's alpha (missing data excluded), which should be ≥0.85 for individual use or 0.70 for group use[47, 95].

Invariance and differential item functioning (DIF)

Invariance means that all persons completing a questionnaire, regardless of their ability (or self-efficacy), recognise the difficulty in identical items[94]. Any likelihood of differently scored items between the groups violates the assumption of invariance, called DIF[96, 97]. Absence of DIF will be tested in gender (female; male), age (quartile groups) and disease duration (quartile groups) and indicated by a non-significant ANOVA of the residuals (5% alpha with Bonferroni correction) where the group is the main factor[97, 98]. Any observed DIF will be examined to know whether it cancels out at the test level[96].

If model fit is not achieved, an iterative stepwise procedure will be initiated, involving

strategies for combining response categories, stepwise deletion of the worst fitting item, testlet (superitem) construction and adjusting for DIF as appropriate[99].

Acknowledgements

The authors acknowledge the permission and support from the original USE-MS developers from The Walton Centre for Neurology and Neurosurgery, Liverpool, UK from the University of Leeds, UK, particularly Professor CA Young and Mike Horton.

Author contributions

All authors critically and substantially revised the manuscript and approved the current version to be submitted for publication. BS devised and designed the study and drafted the manuscript. RK developed the qualitative data analysis plan. LZ substantially contributed to the conception and design of the study. MR provided input on the study methodology and quantitative analysis. RM provided relevant advice on the Rasch analysis. SK substantially contributed to the development of the study protocol. FD is a study manager at his centre and substantially contributed to the development of the study protocol. CB is a study manager at his centre and contributed to the design of the study protocol. RE is a study manager at his centre and substantially contributed to the design and development of the study protocol.

Funding

This work was supported by the Austrian MS Research Society (no grant number). Funders had no role in the study design, decision to publish or preparation of the manuscript.

Competing Interests

None declared.

Data sharing statement

Data generated by this research that support any publications will be made available upon reasonable request as soon as possible. In addition, meaningful data from this research will be available online as a “scientific use file”, wherever legally and ethically possible.

References

1. Pugliatti M, Rosati G, Carton H, et al. The epidemiology of multiple sclerosis in Europe. European Journal 0f Neurology 2006;**13**(7):700-22 doi: 10.1111/j.1468-1331.2006.01342.x[published Online First: Epub Date]|.
2. Compston A, Confavreux C, Lassmann H, et al. *McAlpine´s multiple sclerosis*. 4th ed ed. London: Elsevier, 2006.
3. Bandura A. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs: Prentice-Hall, 1986.
4. Bandura A, Schunk DH. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology 1981;**41**(3):586-98
5. Bandura A. Self-efficacy. In: Ramachaudran VS, ed. Encyclopedia of human behavior. New York: Academic Press, 1994:71-81.
6. Jongen PJ, Lemmens WA, Hoogervorst EL, et al. Glatiramer acetate treatment persistence - but not adherence - in multiple sclerosis patients is predicted by health-related quality of life and self-efficacy: a prospective web-based patient-centred study (CAIR study). Health and quality of life outcomes 2017;**15**(1):50 doi: 10.1186/s12955-017-0622-z[published Online First: Epub Date]|.
7. Casey B, Uszynski M, Hayes S, et al. Do multiple sclerosis symptoms moderate the relationship between self-efficacy and physical activity in people with multiple sclerosis? Rehabilitation psychology 2018;**63**(1):104-10 doi: 10.1037/rep0000190[published Online First: Epub Date]|.
8. Riazi A, Thompson AJ, Hobart JC. Self-efficacy predicts self-reported health status in multiple sclerosis. Multiple Sclerosis Journal 2004;**10**(1):61-6
9. Hughes AJ, Beier M, Hartoonian N, et al. Self-efficacy as a longitudinal predictor of perceived cognitive impairment in individuals with multiple sclerosis. Arch Phys Med Rehabil 2015;**96**(5):913-9 doi: 10.1016/j.apmr.2015.01.008[published Online First: Epub Date]|.
10. Motl RW, Snook EM. Physical Activity, Self-Efficacy, and Quality of Life in Multiple Sclerosis. Annals of Behavioral Medicine 2008;**35**(1):111-15 doi: 10.1007/s12160-007-9006-7[published Online First: Epub Date]|.
11. Stuifbergen AK, Seraphine A, Roberts G. An explanatory model of health promotion and quality of life in chronic disabling conditions. Nursing research 2000;**49**(3):122-9
12. Motl RW, McAuley E, Snook EM, et al. Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. Psychology Health and Medicine 2009;**14**(1):111-24 doi: 10.1080/13548500802241902[published Online First: Epub Date]|.
13. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. Psychonomic bulletin & review 2016;**23**(5):1382-414 doi: 10.3758/s13423-015-0999-9[published Online First: Epub Date]|.
14. French MA, Moore MF, Pohlig R, et al. Self-efficacy Mediates the Relationship between Balance/Walking Performance, Activity, and Participation after Stroke. Topics in stroke rehabilitation 2016;**23**(2):77-83 doi: 10.1080/10749357.2015.1110306[published Online First: Epub Date]|.
15. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, eds. Measures in health psychology: A user's portfolio. Causal and control beliefs. Windsor, UK: NFER-NELSON, 1995:35-37.
16. Motl RW, Balto JM, Ensari I, et al. Self-efficacy and Walking Performance in Persons With Multiple Sclerosis. Journal of neurologic physical therapy : JNPT 2017;**41**(2):114-18 doi: 10.1097/NPT.0000000000000172[published Online First: Epub Date]|.
17. Schmitt MM, Goverover Y, Deluca J, et al. Self-efficacy as a predictor of self-reported physical, cognitive, and social functioning in multiple sclerosis. Rehabilitation psychology 2014;**59**(1):27-34 doi: 10.1037/a0035288[published Online First: Epub Date]|.
18. Airlie J, Baker GA, Smith SJ, et al. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. Clinical rehabilitation 2001;**15**(3):259-

65 doi: 10.1191/026921501668362643[published Online First: Epub Date]|.
19. Rigby SA, Domenech C, Thornton EW, et al. Development and validation of a self-efficacy measure for people with multiple sclerosis: the Multiple Sclerosis Self-efficacy Scale. Mult Scler 2003;**9**(1):73-81 doi: 10.1191/1352458503ms870oa[published Online First: Epub Date]|.
20. Schwartz CE, Coulthard-Morris L, Zeng Q, et al. Measuring self-efficacy in people with multiple sclerosis: a validation study. Arch Phys Med Rehabil 1996;**77**(4):394-8
21. Young CA, Mills RJ, Woolmore J, et al. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. Mult Scler 2012;**18**(9):1326-33 doi: 10.1177/1352458512436592[published Online First: Epub Date]|.
22. Amtmann D, Bamer AM, Cook KF, et al. University of Washington self-efficacy scale: a new self-efficacy scale for people with disabilities. Arch Phys Med Rehabil 2012;**93**(10):1757-65 doi: 10.1016/j.apmr.2012.05.001[published Online First: Epub Date]|.
23. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi journal of anaesthesia 2017;**11**(Suppl 1):S80-S89 doi: 10.4103/sja.SJA_203_17[published Online First: Epub Date]|.
24. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. JAMA : the journal of the American Medical Association 2018;**319**(5):483-94 doi: 10.1001/jama.2017.21903[published Online First: Epub Date]|.
25. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. The Lancet Neurology 2018;**17**(2):162-73 doi: 10.1016/s1474-4422(17)30470-2[published Online First: Epub Date]|.
26. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol 2011;**69**(2):292-302 doi: 10.1002/ana.22366[published Online First: Epub Date]|.
27. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 2001;**50**(1):121-7
28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). Neurology 1983;**33**(11):1444-52
29. Perneger TV, Courvoisier DS, Hudelson PM, et al. Sample size for pre-tests of questionnaires. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2015;**24**(1):147-51 doi: 10.1007/s11136-014-0752-2[published Online First: Epub Date]|.
30. Streiner DL, Norman GR. Devising the items. In: Streiner D, Norman GR, eds. Health Measurement Scales: A Practical Guide to their Development and Use. 4th ed. Oxford, UK: Oxford University Press, 2008.
31. Linacre JM. Sample size and item calibration stability. Rasch Measurement Transactions 1994;**7**(4):328
32. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. Neurology 1996;**46**(4):907-11
33. Montalban X, Gold R, Thompson AJ, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. Mult Scler 2018;**24**(2):96-120 doi: 10.1177/1352458517751049[published Online First: Epub Date]|.
34. Diener H-C, Weimar CH. Diagnose und Therapie der Multiplen Sklerose,Gültigkeit der Leitlinie nach Überprüfung durch das Leitliniensekretariat verängert bis 29.2.2017. In: Diener H-C, Weimar C, eds. Leitlinien für Diagnostik und Therapie in der Neurologie. Stuttgart: Kommission "Leitlinien" der Deutschen Gesellschaft für Neurologie, Thieme Verlag, 2012/2017.
35. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods

to clinical intuition. Brain : a journal of neurology 2000;**123 (Pt 5)**:1027-40
36. Sharrack B, Hughes RA, Soudain S, et al. The psychometric properties of clinical rating scales used in multiple sclerosis. Brain : a journal of neurology 1999;**122 (Pt 1)**:141-59
37. Meyer-Moock S, Feng YS, Maeurer M, et al. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. BMC Neurol 2014;**14**(1):58 doi: 10.1186/1471-2377-14-58[published Online First: Epub Date]|.
38. Schwarzer R, Jerusalem ME. *Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin, 1999.
39. Schumacher J, Leppert K, Gunzelmann T, et al. Die Resilienzskala – Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie 2004;**53**:16-39
40. Leppert K, Koch B, Brähler E, et al. Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13. Klinische Diagnostik und Evaluation 2008;**2**:226–43
41. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. Journal of nursing measurement 1993;**1**(2):165-78
42. Flachenecker P, Vogel U, Simeoni MC, et al. [MusiQol: international questionnaire investigating quality of life in multiple sclerosis: validation results for the German subpopulation in an international comparison]. Der Nervenarzt 2011;**82**(10):1281-9 doi: 10.1007/s00115-011-3276-9[published Online First: Epub Date]|.
43. Simeoni M, Auquier P, Fernandez O, et al. Validation of the Multiple Sclerosis International Quality of Life questionnaire. Mult Scler 2008;**14**(2):219-30 doi: 10.1177/1352458507080733[published Online First: Epub Date]|.
44. Petermann F. Hospital Anxiety and Depression Scale, Deutsche Version (HADS-D). Zeitschrift für Psychiatrie, Psychologie und Psychotherapie 2011;**59**(3):251-53 doi: 10.1024/1661-4747/a000077[published Online First: Epub Date]|.
45. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta psychiatrica Scandinavica 1983;**67**(6):361-70
46. NFI-MS Neurologischer Fragebogen zur Müdigkeit. NFI-MS Austria/German - Version of 30 Sep 13 - Mapi. ID7555 / NFI-MS_AU1.0_deu-AT.doc 2010.
47. Mills RJ, Young CA, Pallant JF, et al. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). Health and quality of life outcomes 2010;**8**:22 doi: 10.1186/1477-7525-8-22[published Online First: Epub Date]|.
48. Mills RJ, Calabresi M, Tennant A, et al. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). Mult Scler 2013;**19**(4):502-5 doi: 10.1177/1352458512457840[published Online First: Epub Date]|.
49. Marx RG, Menezes A, Horovitz L, et al. A comparison of two time intervals for test-retest reliability of health status instruments. Journal of clinical epidemiology 2003;**56**(8):730-35 doi: 10.1016/s0895-4356(03)00084-2[published Online First: Epub Date]|.
50. Beaton D, Bombardier C, Guillemin F, et al. *Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures*. Toronto, Canada: Institute for Work and Health, 2007.
51. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. Journal of clinical epidemiology 1993;**46**(12):1417-32
52. Brislin RW, Lonner WJ, Thorndike RM. Questionnaire wording and translation. Cross-cultural Research Methods. New York: John Wiley, 1973:32-58.
53. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health and quality of life

outcomes 2003;**1**:79 doi: 10.1186/1477-7525-1-79[published Online First: Epub Date]|.
54. Flesch R. How to Write Plain English. Secondary How to Write Plain English 2016.
55. Willis G. *Cognitive Interviewing and Questionnaire Design: A Training Manual. Cognitive Methods Staff Working Paper Series No 7*. Hyattsville, MD, USA: Office of Research and Methodology, National Center for Health Statistics, 1994.
56. Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys. Meeting of the American Statistical Association; 1999 8. - 12. August; Baltimore, Maryland.
57. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. Public Opinion Quarterly 2007;**71**(2):287-311 doi: 10.1093/poq/nfm006[published Online First: Epub Date]|.
58. Oksenberg L, Cannell CF, Kalton G. New Strategies for Pretesting Survey Questions. Journal of Official Statistics 1991;**7**:349–65
59. Conrad F, Blair J. Data Quality in Cognitive Interviews: The Case for Verbal Reports. In: Presser S, Rothgeb JM, Couper MP, et al., eds. Methods for Testing and Evaluating Survey Questionnaires. Hoboken, NJ: John Wiley and Sons, 2004.
60. Willis G. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage, 2005.
61. Olsson JE, Ekblad S, Bertilson BC, et al. Swedish adaptation of the General Medical Council's multisource feedback questionnaires: a qualitative study. International journal of medical education 2018;**9**:161-69 doi: 10.5116/ijme.5af6.c209[published Online First: Epub Date]|.
62. Berelson B. *Content Analysis in Communication Research*. Glencoe: Free Press, 1952.
63. Lasswell HD. *Power and Personality*. New York, NY: W. W. Norton & Company, 1948.
64. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua 2007;**19**(6):349-57 doi: 10.1093/intqhc/mzm042[published Online First: Epub Date]|.
65. Schreier M. *Qualitative Content Analysis in Practice*. London, UK: SAGE Publications, 2012.
66. Bryman A, Burgess RGE. *Analyzing Qualitative Data*. Abingdon, Oxon: Routledge, 2007.
67. Denzin NK, Lincoln YSE. *The SAGE Handbook of Qualitative Research*. 5th ed. Thousand Oaks, CA: SAGE Publications, 2018.
68. Bryman A. *Social Research Methods*. 4th ed. New York, NY: Oxford University Press, 2012.
69. Cypress BS. Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. Dimensions of critical care nursing : DCCN 2017;**36**(4):253-63 doi: 10.1097/DCC.0000000000000253[published Online First: Epub Date]|.
70. Noble H, Smith J. Issues of validity and reliability in qualitative research. Evidence-based nursing 2015;**18**(2):34-5 doi: 10.1136/eb-2015-102054[published Online First: Epub Date]|.
71. Leung L. Validity, reliability, and generalizability in qualitative research. Journal of family medicine and primary care 2015;**4**(3):324-7 doi: 10.4103/2249-4863.161306[published Online First: Epub Date]|.
72. Miles HB, Huberman AM. *Qualitative Data Analysis*. Thousand Oaks, CA: Sage, 1994.
73. Andrich D, Lyne A, Sheridon B, et al. *RUMM 2030*. Perth: RUMM Laboratory, 2009.
74. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health technology assessment 2009;**13**(12):iii, ix-x, 1-177 doi: 10.3310/hta13120[published Online First: Epub Date]|.
75. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;**45**(1):255-68
76. Lin LI-K. Corrections. Biometrics 2000;**56**:324 25
77. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;**33**(1):159-74
78. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. 5th ed ed. Boston: Houghton Mifflin, 2003.

79. Rasch G. *Probabilistic models for some intelligence and attainment tests (revised and expanded ed.)*. Chicago The University of Chicago Press, 1960.
80. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press, 1980.
81. Andrich D. *Rasch Models for Measurement* Beverly Hills: SAGE, 1988.
82. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis and rheumatism 2007;**57**(8):1358-62 doi: 10.1002/art.23108[published Online First: Epub Date]|.
83. Masters G. Rasch model for partial credit scoring. Psychometrika 1982;**47**:149–74
84. Andrich D. Rating formulation for ordered response categories. Psychometrika 1978;**43**:561
85. Linacre JM. Optimizing rating scale category effectiveness. Journal of applied measurement 2002;**3**(1):85-106
86. Khadka J, Pesudovs K, McAlinden C, et al. Reengineering the glaucoma quality of life-15 questionnaire with rasch analysis. Investigative ophthalmology & visual science 2011;**52**(9):6971-7 doi: 10.1167/iovs.11-7423[published Online First: Epub Date]|.
87. Fisher WPJ. Rating Scale Instrument Quality Criteria. Rasch Measurement Transactions 2007;**21**(1):1095
88. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of applied measurement 2002;**3**(2):205-31
89. Tennant A, Pallant JF. Unidimensionality matters! (a tale of two Smiths?). Rasch Measurement Transactions 2006;**20**:1048-51
90. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. PloS one 2014;**9**(6):e99485 doi: 10.1371/journal.pone.0099485[published Online First: Epub Date]|.
91. Andrich D. *Rasch models for measurement series: quantitative applications in the social sciences no. 68*. London: Sage, 1988.
92. Twiss J, McKenna SP, Graham J, et al. Applying Rasch analysis to evaluate measurement equivalence of different administration formats of the Activity Limitation scale of the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR). Health and quality of life outcomes 2016;**14**:57 doi: 10.1186/s12955-016-0462-2[published Online First: Epub Date]|.
93. Milinis K, Tennant A, A. Young C. Rasch analysis of SF-Qualiveen in multiple sclerosis. Neurourology and Urodynamics 2017;**36**(4):1161-66 doi: 10.1002/nau.23081[published Online First: Epub Date]|.
94. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 2nd ed. London: Lawrence Erlbaum Associates, 2007.
95. Linacre JM. Data Variance Explained by Rasch Measures. Rasch Measurement Transactions 2006;**20**(1):1045-54
96. Tennant A, Pallant JF. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. Rasch Measurement Transactions 2007;**20**(4):1082-84
97. Holland PW, Wainer H. *Differential Item Functioning* Hillsdale, N. J.: Lawrence Erlbaum, 1993.
98. Grimby G. Useful reporting of DIF Rasch Meas Transactions 1998;**12**:651
99. Siegert RJ, Jackson DM, Tennant A, et al. Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research. Journal of rehabilitation medicine : official journal of the UEMS European Board of Physical and Rehabilitation Medicine 2010;**42**(4):302-9 doi: 10.2340/16501977-0511[published Online First: Epub Date]|.

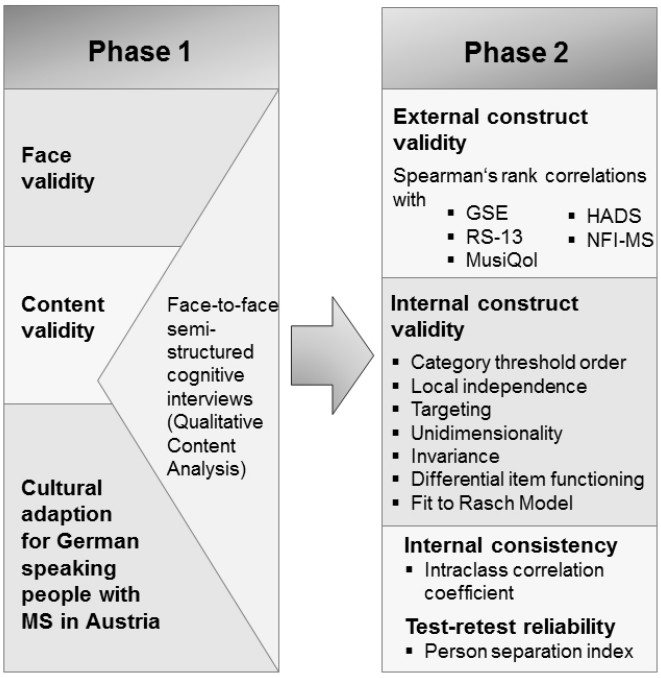

Figure 1

254x190mm (96 x 96 DPI)

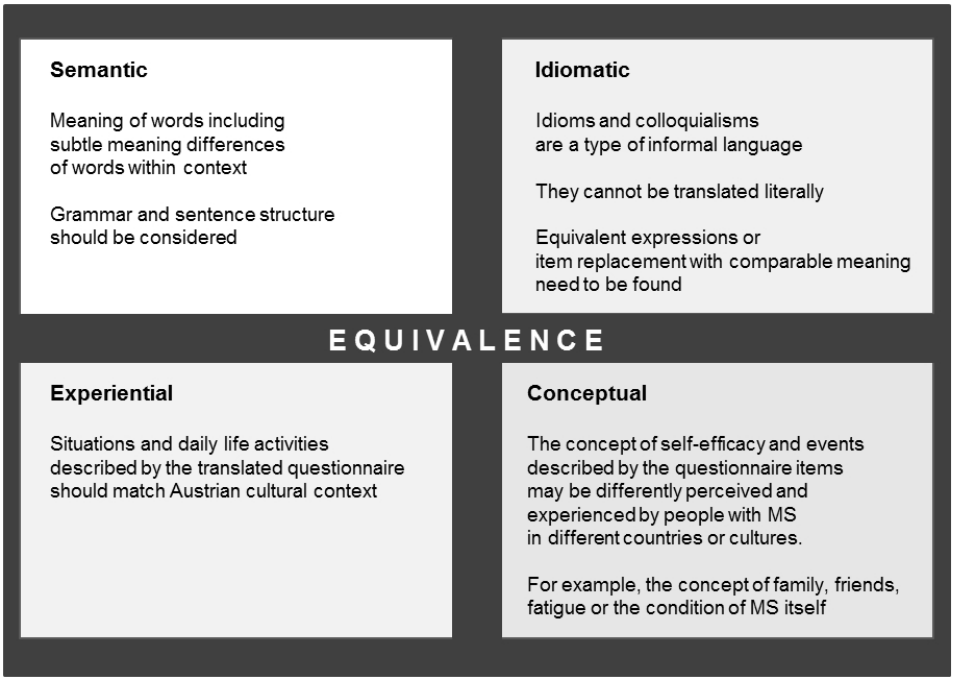

Figure 2

254x190mm (96 x 96 DPI)

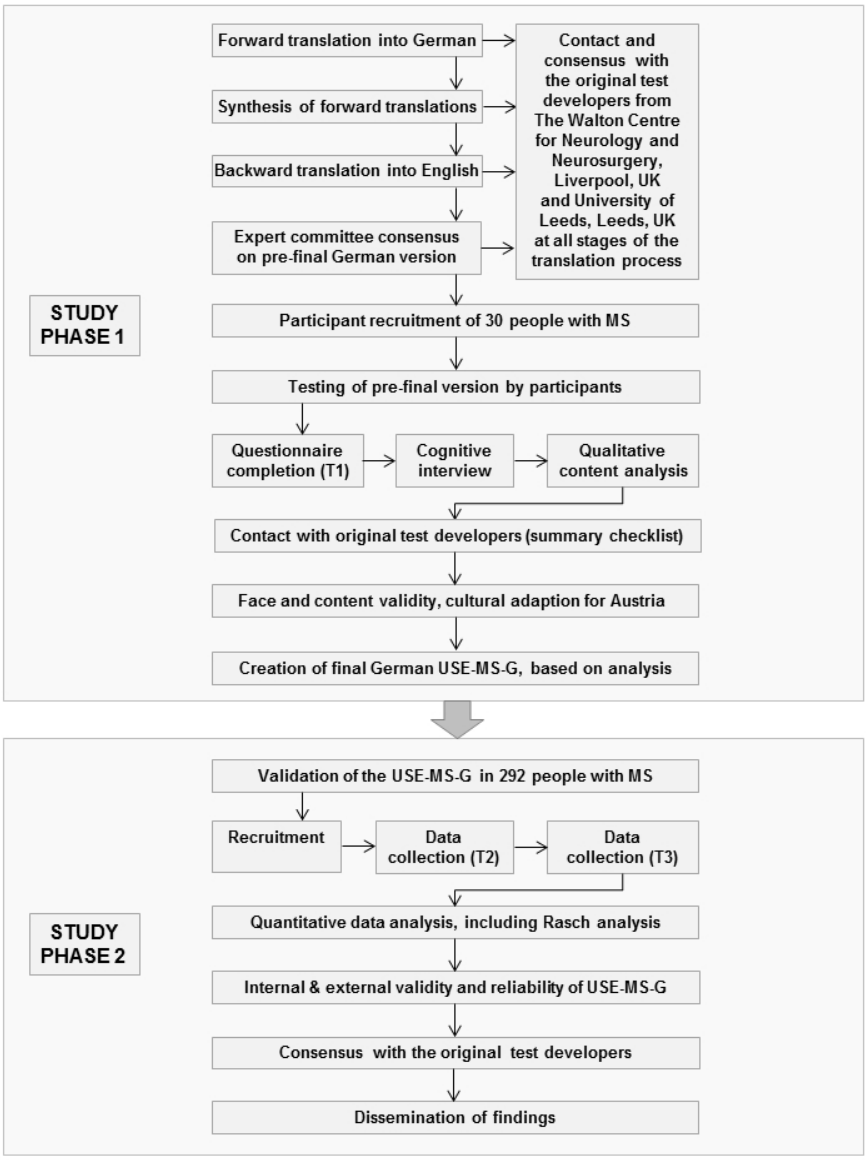

Figure 3

190x254mm (96 x 96 DPI)

SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol
Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. JAMA : the journal of the American Medical Association 2018;319(5):483-94 doi: 10.1001/jama.2017.21903[published Online First: Epub Date])

Section/item	ItemNo	Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.
Administrative information					
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym			1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry			2
	2b	All items from the World Health Organization Trial Registration Data Set			2; 22
Protocol version	3	Date and version identifier			2
Funding	4	Sources and types of financial, material, and other support			22

Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors			1; 22
	5b	Name and contact information for the trial sponsor	SPIRIT-5a-PRO Elaboration	Specify the individual(s) responsible for the PRO content of the trial protocol.	BS
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities			8
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)			7; 13; 14; 22
Introduction					
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	SPIRIT-6a-PRO Extension	Describe the PRO-specific research question and rationale for PRO assessment and summarize PRO findings in relevant studies.	5-6

	6b	Explanation for choice of comparators			11-13; Supplementary File 2
Objectives	7	Specific objectives or hypotheses	SPIRIT-7-PRO Extension	State specific PRO objectives or hypotheses (including relevant PRO concepts/domains).	6
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)			6-7
Methods: Participants, interventions, and outcomes					
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained			7
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	SPIRIT-10-PRO Extension	Specify any PRO-specific eligibility criteria (eg, language/reading requirements or prerandomization completion of PRO). If PROs will not be collected from the entire study sample, provide a rationale and describe the method for obtaining the PRO subsample.	7-8

Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered			NA
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)			7-8 (concerning assessments) Interventions: NA
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)			7; 13
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial			7-8
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	SPIRIT-12-PRO Extension	Specify the PRO concept/domains used to evaluate the intervention (eg, overall health-related quality of life, specific domain, specific symptom) and, for each one, the analysis metric (eg, change from baseline, final value, time to event) and the principal time point or period of interest.	9-13

Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	SPIRIT-13-PRO Extension	Include a schedule of PRO assessments, providing a rationale for the time points, and justifying if the initial assessment is not prerandomization. Specify time windows, whether PRO collection is prior to clinical assessments, and if using multiple questionnaires, whether order of administration will be standardized.	9-13
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	SPIRIT-14-PRO Extension	When a PRO is the primary end point, state the required sample size (and how it was determined) and recruitment target (accounting for expected loss to follow-up). If sample size is not established based on the PRO end point, then discuss the power of the principal PRO analyses.	8-9
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size			7
Methods: Assignment of interventions (for controlled trials)					
Allocation:					NA

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions			NA
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned			NA
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions			NA
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how			NA

	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial			NA
Methods: Data collection, management, and analysis					
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	SPIRIT-18a (i)-PRO Extension	Justify the PRO instrument to be used and describe domains, number of items, recall period, and instrument scaling and scoring (eg, range and direction of scores indicating a good or poor outcome). Evidence of PRO instrument measurement properties, interpretation guidelines, and patient acceptability and burden should be provided or cited if available, ideally in the population of interest. State whether the measure will be used in accordance with any user manual and specify and justify deviations if planned.	9-13
			SPIRIT-18a (ii)-PRO Extension	Include a data collection plan outlining the permitted mode(s) of administration (eg, paper, telephone, electronic, other) and setting (eg, clinic, home, other).	7-8; 15
			SPIRIT-18a (iii)-PRO Extension	Specify whether more than 1 language version will be used and state whether translated versions have been developed using currently recommended methods.	Total manuscript

			SPIRIT-18a (iv)-PRO Extension	When the trial context requires someone other than a trial participant to answer on his or her behalf (a proxy-reported outcome), state and justify the use of a proxy respondent. Provide or cite evidence of the validity of proxy assessment if available.	NA
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	SPIRIT-18b (i)-PRO Extension	Specify PRO data collection and management strategies for minimizing avoidable missing data.	19
			SPIRIT-18b (ii)-PRO Elaboration	Describe the process of PRO assessment for participants who discontinue or deviate from the assigned intervention protocol.	19
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol			13-14; 18-19; 23

Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	SPIRIT- 20a-PRO Elaboration	State PRO analysis methods, including any plans for addressing multiplicity/type I (α) error.	18-22
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)			20-22
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	SPIRIT- 20c-PRO Elaboration	State how missing data will be described and outline the methods for handling missing items or entire assessments (eg, approach to imputation and sensitivity analyses).	19
Methods: Monitoring					
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed			13; 17-18; 22

	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial			13
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	SPIRIT- 22-PRO Extension	State whether or not PROs data will be monitored during the study to inform the clinical care of individual trial participants and, if so, how this will be managed in a standardized way. Describe how this process will be explained to participants; eg, in the participant information sheet and consent form.	13-17
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor			17
Ethics and dissemination					
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval			8

Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)			NA
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)			7
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable			NA
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial			8; 12-13
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site			22
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators			12-13

Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation			NA
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions			8; 23
	31b	Authorship eligibility guidelines and any intended use of professional writers			22
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code			23
Appendices					
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates			23
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable			NA

Abbreviations: SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; PRO, patient-reported outcome.
*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license and is reproduced with permission.

Supplementary File 2

Except good to excellent psychometric properties, the selection of outcome measures depended also on patient involvement during their development, patient acceptability of the tools or recommendations from relevant governmental organisations or MS Societies. The MusiQol was developed using individual semi-structured interviews of 1992 PwMS in five countries, followed by qualitative content analysis. The MS-EDGE Outcome Measure Task Force reviewed and recommended the MusiQol for use in PwMS[1]. Similarly, both the German long and short versions of the Resilience Scale are based on face-to-face interviews of 2031 people. The GSE was exposed to repeated validation studies in thousands of participants where excellent validity and reliability were confirmed. A study explored the validity of the HADS in PwMS using interviews by an interviewer who was blind to the HADS scores [2]. The acceptability of the German HADS to various patient populations was tested and confirmed [3]. The UK National Institute for Health and Care Excellence (NICE) 2018 Guidelines for MS[4] recommended the HADS anxiety subscale as a sensitive and specific anxiety screening tool, based on a review[5]. The NFI-MS was developed through a two-stage process in 635 PwMS; quantitative validation and qualitative interviews with patients. The NFI-MS was reviewed and recommended by the MS International Federation (MSIF)[6].

References

1. Potter K, Cohen ET, Allen DD, et al. Outcome measures for individuals with multiple sclerosis: recommendations from the American Physical Therapy Association neurology section task force. Phys Ther 2014;**94**(5):593-608 doi: 10.2522/ptj.20130149[published Online First: Epub Date]|.
2. Watson TM, Ford E, Worthington E, et al. Validation of Mood Measures for People with Multiple Sclerosis. Int J MS Care 2014;**16**:105–09
3. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. Journal of psychosomatic research 1997;**42**(1):17-41
4. Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186). Secondary Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186) 2018.

https://www.nice.org.uk/guidance/CG186/documents/surveillance-review-proposal.

5. Litster B, Fiest KM, Patten SB, et al. Screening Tools for Anxiety in People with Multiple Sclerosis: A Systematic Review. Int J MS Care 2016;**18**(6):273-81 doi: 10.7224/1537-2073.2016-004[published Online First: Epub Date]|.
6. MSIF. Fatigue and MS. MS in focus 2012;**1**:1-28

BMJ Open

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029565.R1
Article Type:	Protocol
Date Submitted by the Author:	26-Jun-2019
Complete List of Authors:	Seebacher, Barbara; Medical University of Innsbruck, Clinical Department of Neurology Mills, Roger J; Walton Centre NHS Foundation Trust, Department of Neurology Reindl, Markus; Medical University of Innsbruck, Clinical Department of Neurology Zamarian, Laura; Medical University of Innsbruck, Clinical Department of Neurology Kuisma, Raija; University of Brighton, School of Health Sciences Kircher, Simone; University Hospital of Innsbruck, Clinical Department of Neurology Brenneis, Christian; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Ehling, Rainer; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Deisenhammer, Florian; Medical University of Innsbruck, Clinical Department of Neurology
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Rehabilitation medicine, Patient-centred medicine
Keywords:	Multiple sclerosis < NEUROLOGY, Self Efficacy, Patient Reported Outcome Measures, Austria, Cross-Cultural Comparison, Validation Studies as Topic

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Barbara Seebacher[1], Roger J Mills[2], Markus Reindl[1], Laura Zamarian[1], Raija Kuisma[3], Simone Kircher[4], Christian Brenneis[5, 6], Rainer Ehling[5, 6], Florian Deisenhammer[1]

[1]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria

[2]Walton Centre Foundation Trust, Department of Neurology, Liverpool, UK

[3]School of Health Sciences, University of Brighton, Eastbourne, UK

[4]University Hospital of Innsbruck, Clinical Department of Neurology, Innsbruck, Austria

[5]Department of Neurology, Clinic for Rehabilitation Münster, Münster, Austria

[6]Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster, Münster, Austria

Correspondence to Dr Barbara Seebacher; Clinical Department of Neurology, Medical University of Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria; phone +43.512.504.82499; barbara.seebacher@i-med.ac.at

Word count: 3960 words excluding declarations; 4206 words including declarations

ABSTRACT

Introduction: Self-efficacy refers to individuals' confidence in their ability to perform relevant tasks to accomplish desired goals. This is independent of their actual abilities. In people with multiple sclerosis (MS), self-efficacy has been shown to powerfully influence motivation and health-related behaviour, such as adherence to prescribed treatment or physical activity. So far, a rigorously tested German language self-efficacy questionnaire for people with MS is missing.

Methods: The purpose of this study is to translate the original Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) into German and to validate the German USE-MS (USE-MS-G). Based on Bandura's concept of self-efficacy and international guidelines for questionnaire development, the patient-led development of the pre-final German version will involve a forward-backward translation process, synthesis of translations, expert committee review and consensus with the original test developers. At two centres in Tyrol, Austria, content and face validity and cultural adaption for Austria will be established using face-to-face semi-structured cognitive interviews of 30 people with MS. A further 292 people with MS with minimal to severe disability will be tested at two time-points to validate the USE-MS-G.

Results: Mixed methods analyses will be applied. Interviews will be transcribed and analysed employing qualitative content analysis. External validity will be explored using Spearman's Rank correlation coefficients of the USE-MS-G with the 13-item Resilience Scale, General Self-Efficacy Scale, Multiple Sclerosis International Quality of Life questionnaire, Hospital Anxiety and Depression Scale and MS-specific Neurological Fatigue Index. Test-retest reliability, internal consistency and floor and ceiling effects will be evaluated. Internal validity will be examined using Rasch analysis.

Ethics and dissemination: Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Results from this study will be disseminated to the participants and MS Societies, and to clinicians and researchers through peer-reviewed publications and conferences.

Study registration: ISRCTN Registry; trial ID ISRCTN14843579; prospectively registered on 02. 01. 2019; http://www.isrctn.com/ISRCTN14843579

Study protocol, revision 1, 26.6.2019

Keywords: Multiple Sclerosis; Self Efficacy; Patient Reported Outcome Measures; Austria; Cross-Cultural Comparison; Validation Studies as Topic.

ARTICLE SUMMARY

Strengths and limitations of this study

- This study protocol describes the German translation of the original English language Unidimensional Self-efficacy Scale for Multiple Sclerosis (USE-MS), upon permission of the scale developers and applying international recommendations.
- Consistent with the conceptual framework of the English USE-MS, Bandura's concept of self-efficacy will be adhered to.
- Employing a patient-led process in phase 1, 30 people with MS (PwMS) will be interviewed about the pre-final German USE-MS, to establish face and content validity and cultural adaption for PwMS in Austria.
- In phase 2, the German USE-MS will be validated in a larger sample of 292 PwMS.
- Applying classical test theory and Rasch analysis approaches, internal and external validity, internal consistency and test retest reliability will be explored.

INTRODUCTION

Multiple sclerosis (MS) is one of the most common neurological diseases in young adults worldwide, with increasing prevalence[1]. MS is characterised by a wide variety of symptoms and different disease courses[2]. Despite the development of novel disease modifying drugs and neurorehabilitation strategies, the unpredictability of the disease with psychological distress, losses in social contact and quality of life (QoL) are concerning for PwMS. However, individuals' self-knowledge can modulate their approach to day-to-day activities. According to Bandura's social cognitive theory, psychosocial functioning is regulated by reciprocal interactions between behaviour, personal factors and environmental conditions[3]. Self-regulation and intrinsic motivation enable individuals to set and pursue their own goals, observe and evaluate themselves in relation to attained goals[4]. Bandura defined self-efficacy as individuals' beliefs regarding their capability to perform significant tasks, to achieve goals that are meaningful for their daily lives[3]. Self-efficacy beliefs considerably influence people's feelings, thoughts and motivation[5] while, notably, being independent of their physical performance[5]. Such a concept appears important for people with disabilities because it may shape their motivation to initiate and adhere to treatment, particularly when facing side effects.

Perceived self-efficacy influences health-related behaviour such as adhering to medication[6] or engaging in physical activity in PwMS[7]. Health status evaluations of responses to rehabilitation and steroid treatment after an MS relapse can be predicted by self-efficacy levels[8]. Also, higher self-efficacy levels are associated with better long-term perceived cognitive functioning[9] and QoL[10 11]. PwMS who report higher perceived self-efficacy also state lower levels of fatigue, depression and anxiety[12]. Recent evidence has provided insight into the importance of self-management and intrinsic motivation for motor learning[13]. Recognising the relevance of self-efficacy especially for people with disabilities, valid and reliable measurement tools are still needed for its assessment. Three generic self-efficacy scales were found in the literature[7 14-16]. However, generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS. The initial impact of a diagnosis of MS, in addition to the manifold symptoms and necessity of managing a progressive disease may affect individuals' self-efficacy perceptions. Studies demonstrated that the

capability to effectively solve problems, consistent with higher self-efficacy levels, is strongly associated with PwMS' psychological adaptation to their disability[17], supporting the choice of a disease-specific over a generic self-efficacy questionnaire. MS-specific self-efficacy scales include the Liverpool Self-efficacy Scale (LSES)[18], Multiple Sclerosis Self-Efficacy Scale (MSSS)[19], MS Self-Efficacy Scale (MSSE)[20], Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS)[21] and University of Washington Self-Efficacy Scale for people with disabilities[22].

Following current guidelines, patients should be involved in the translation and development process of disease specific questionnaires, to ensure the scale reflects their experiences[23]. LSES and MSSS development used in-depth patient interviews while the USE-MS consists of items from both the LSES and MSSS. Bandura's concept of self-efficacy is reflected in the wording of all three questionnaires. The USE-MS study sample was the largest thereof (N=303), and only the USE-MS was exposed to Rasch[24] [25] analysis assessing internal construct validity, in addition to conventional external construct validity and reliability testing. Fit to the Rasch model was demonstrated, and good external validity and reliability[21]. Consequently, the USE-MS appears to be appropriate for use in clinical practice and research. However, so far no validated German language version of the USE-MS is available. The purpose of this study will therefore be to translate the USE-MS into German and validate the German language version in a larger sample of PwMS.

METHODS

Study aims

The first aim of this patient led study is to translate the original English USE-MS, developed by Young et al. (2012) into German, based on international guidelines.

The second aim is to establish face and content validity and cultural adaption of the German version for PwMS in Austria, using individual semi-structured cognitive interviews.

The third aim is to evaluate internal and external validity, internal consistency and test-retest reliability of the German USE-MS (USE-MS-G), using classical test theory and Rasch analysis .

Study design

This will be a bi-centre prospective cross-sectional translation and validation study with repeated measures, consisting of Phase 1 and Phase 2. The SPIRIT 2013 and SPIRIT-PRO Extension checklist for study protocols[26] is presented as Supplementary File 1.

Study setting and timeline

Locations will be the outpatient MS-Clinic of the Clinical Department of Neurology, Medical University of Innsbruck, Austria and Department of Neurology, Clinic for Rehabilitation Münster, Austria.

The expected overall study duration is 33 months, from 01.02.2019 to 31.10. 2021.

Participants and recruitment

A random cross-sectional cohort of patients with clinically definite MS will be recruited from the two centres. Adult (≥18 years) people of any ethnicity and with any MS phenotype according to the McDonald's criteria[27-29] version valid at the time of diagnosis will be included in the study. Their disability status score on the Expanded Disability Status Scale (EDSS)[30] may range from 0 (no disability) to 9.0 (severe disability). Patients will be included if they are able to speak and understand German language. Exclusion criteria are concomitant diseases which may affect subjective self-efficacy ratings (e.g. malignant diseases, other neurological or psychiatric disorders), a relapse of MS within the last two months or any medication change within four weeks prior to the study. A relapse between testing 2 and 3 would necessitate the exclusion of the participant.

The study will be advertised in the MS-Clinic, the Rehabilitation Centre and on the Austrian MS Society website. Further interested PwMS will be examined for eligibility by neurologists at the two study locations. Severely disabled PwMS (EDSS ≥8) will be offered home visits to enable their participation. Written informed consent will be obtained by the first author (BS) who is not involved in the treatment of the patients. Participants may withdraw from the study at any time and for any reasons without prejudice. Outpatient participants will be reimbursed for travel expenses only.

Patient and public involvement

In Phase 1, patients will be lay members of the expert committee to consolidate all the translations and back translations of the USE-MS. Their role regarding the item and response option wording and sentence structure will be crucial, as the final

questionnaire should be understood by PwMS. Patients will also be involved using face-to-face cognitive interviewing, to gain insight into their views about the clarity of the wording, meaning and completeness of the questions of the pre-final USE-MS-G. The Austrian MS (recruitment) and MS Research Societies (funding) will be involved in this study, with whom the findings will be shared as soon as available (patient magazine, meetings). The findings will also be disseminated to the UK MS Society and MS Trust.

Sample size

Phase 1

Patients will be recruited until saturation is achieved. Saturation is a standard term in qualitative methodology to signify the point when the analysis of data from new participants reveals no further emergent qualitative themes. Saturation is typically achieved after 10-30 people have been interviewed but is determined by the nature of the analysis and the participants themselves[31].

Phase 2

Rasch analysis sample size requirements are predicated upon the degree of precision required for estimating item and person difficulties. Regardless of targeting, one can be 99% confident that a sample size of 243 participants is adequately large to obtain a (high) precision of ±0.5 log odd units (logits). Good targeting provided, a sample size of 108 people would be sufficient[21 32]. Using the formula N=n/(1-(z/100)) where n is the calculated number of participants and z the expected attrition rate of 15-20%, a total sample size of 286-304 participants will be aimed at in this study.

Outcomes and data collection

Assessments used in this study were developed using patient involvement and/or recommended by governmental or patient organisations (Supplementary File 2). Study outcomes and methods for their assessment are presented in Figure 1. Participant characteristics and assessments used at all time-points are shown in Table 1.

Figure 1 around here

Figure 1 Study outcomes and their assessment

Figure legend: GSE: General Self-Efficacy Scale; RS-13: Resilience Scale, short version; MusiQol: Multiple Sclerosis International Quality of Life questionnaire; HADS:

Hospital Anxiety and Depression Scale; NFI-MS: Neurological Fatigue Index.

Table 1 Participant characteristics and assessments used in this study

Participant characteristics and assessments (assessments will be collected in a random order to avoid order effect)	**Phase 1**	**Phase 2**	
	T1	**T2**	**T3**
Participant identifier (ID)	**X**	**X**	**X**
Age	**X**	**X**	
Gender	**X**	**X**	
MS phenotype[1]	**X**	**X**	
Disease duration	**X**	**X**	
Expanded Disability Status Scale (EDSS)[2]	**X**	**X**	
Disease modifying treatment (DMT)[3]	**X**	**X**	**X**
(Pre-final) German version of Unidimensional Self-Efficacy Scale for Multiple Sclerosis	**X**	**X**	**X**
Qualitative cognitive interview	**X**		
Resilience Scale, short version		**X**	**X**
General Self-Efficacy Scale		**X**	**X**
Multiple Sclerosis International Quality of Life questionnaire		**X**	**X**
Hospital Anxiety and Depression Scale		**X**	**X**
Neurological Fatigue Index		**X**	**X**

[1]Relapsing-remitting; primary progressive; secondary progressive multiple sclerosis[33]

[2]EDSS groups: 0-4.0; 4.5-6.5; 7.0-7.5; 8.0-9.0[30]

[3](a) No DMTs; (b) low effective DMTs: interferon-b 1a, interferon-b 1a, interferon-b 1b, pegylated interferon-b 1a, glatiramer acetate, dimethyl fumarate, teriflunomide, azathioprin, intravenous immunoglobulins; (c) high effective DMTs: alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, cyclophosphamide, mitoxantrone, rituximab[34 35]

At recruitment, disability will be assessed by neurologists (FD, CB or RE) using the EDSS, ranging from 0 to 10, with higher scores representing higher levels of disability[30].

Although psychometric validation studies criticised its low responsiveness to changes, the EDSS has no floor or ceiling effects[36], has been shown to be valid and reliable[37] and is therefore recommended for use in clinical studies[38].

Excellent internal and external validity and reliability of the original USE-MS has been shown[21]. Scoring of the USE-MS draws results from all 12 items while items 5, 7, 8, 9 and 11 are reversed scored. Higher numbers represent stronger self-efficacy beliefs in participants[21]. The USE-MS includes a 4-point Likert scale (0= strongly disagree to 3=strongly agree).

To assess external construct validity, the following questionnaires will be administered:

The validated German version[39] of the 10-item General Self-Efficacy Scale (GSE)[15] is a self-administered 4-point Likert scale with a summary score ranging from “not at all true” to “exactly true”. The total GSE score ranges between 10 and 40, higher scores signifying greater self-efficacy. Psychometric testing demonstrated high internal consistency, moderate concurrent validity and unidimensionality[15].

The validated German version[40] of the 13-item Resilience Scale (RS-13)[41], based on the 25-item Resilience Scale[42] will be used. RS-13 item scores from a 7-point Likert scale are added up, indicating low (13-66 points), moderate (67-72 points) or high (73-91 points) resilience[41]. The German RS-13 showed high internal consistency and moderate test-retest reliability. Confirmatory factor analysis indicated an acceptable model fit[41].

The validated German version[43] of the 31-item Multiple Sclerosis International Quality of Life (MusiQol) questionnaire[44] will be employed. Response options use a 6-point Likert scale, from 1= “never/not at all” to 5= “always/very much” and 6= “not applicable”. Negatively worded item scores are reversed, and for each participant mean scores for each dimension of the item scores are calculated. All nine dimension scores are linearly transformed to a 0-100 scale, their mean representing the global index score, 0 indicating the worst level of health-related QoL and 100 the best. Psychometric testing showed satisfactory internal and external validity and acceptable reliability for all MusiQol dimensions[44].

The validated German version[45] of the 14-item Hospital Anxiety Depression Scale

(HADS)[46] will be used. The HADS is a self-report questionnaire with a 4-point Likert scale and a 42 point maximum, higher scores representing higher levels of anxiety or depression. Items 2, 4, 7, 9, 12 and 14 are reversed scored, odd items are added to score the anxiety subscale (0-21 points) and even items are added to generate the depression subscale (0-21 points). Testing of the German version demonstrated good internal consistency and acceptable test-retest reliability [45]. The two-factor structure of the scale was confirmed[45].

The validated German version[47] of the 23-item Neurological Fatigue Index (NFI-MS) will be used[48]. Four factors of the NFI-MS were confirmed by principal component analysis and explained 62% of the variance. The four subscales and total scale showed acceptable responsiveness[49], good test-retest reliability, moderate convergent validity and fit to Rasch model expectations[48]. Items are scored on a 4-point Likert scale from 0= “strongly disagree” to 3= “strongly agree”. For scoring, the following item values are added: 1-8= “physical subscale”; 9-12= “cognitive subscale”; 13-18= “relief by diurnal sleep or rest subscale”; 19-23= “abnormal nocturnal sleep and sleepiness subscale”; and 1-7, 9, 11-12 = “physical and cognitive summary score”[48].

Assessments will be performed by trained physiotherapists holding a Master’s (SK) and PhD degree (BS) and a clinical neuropsychologist (LZ). The number of participants who decline to participate or drop out will be recorded, together with reasons (CONSORT flow chart). Any health problems will be recorded.

Phase 1: data will be collected at one time-point (Testing 1, T1), with an expected duration of 45-60 minutes.

Phase 2: for the test-retest reliability assessment, data will be collected at two time-points and will last 60-90 minutes: Testing 2 (T2) and Testing 3 (T3), 14-21 days after T2[48 50].

Data management

With regard to confidentiality, the Austrian and Tyrolean Data Protection Acts will be adhered to. Double data entry and range checks for data values will be used. For qualitative content analysis, double coding of the data set will be performed. Only the research team will have access to the data. All personal data will be codified by a participant ID. Data and files will be saved on a password protected computer, will not be

transferred via emails and will be only used for the purposes for which they were collected. Participants will be informed about their right to disclosure for their own data even if these data lack clinical utility. Codified data will be kept for 15 years following completion of the study. Blank data collection forms can be requested from the corresponding author.

Study procedures

This study will follow the Beaton et al. guidelines for the cross-cultural adaptation of patient reported outcomes[51] and its enhanced version from the University of Leeds, UK.

Phase 1

Stage 1: Forward translation of the items, response options, instructions and scoring information into German will be performed by three independent translators; translator 1 is a medical professional and informed about self-efficacy, while translators 2 and 3 have no medical knowledge and are "naïve" to self-efficacy. Translators are bilingual German native speakers and will create a written report for all translations (T1, T2 and T3), which will then be compared, to distinguish any wording differences or ambiguities[52].

Stage 2 will be a synthesis of T1-3 into T-123. Involving a fourth, unbiased person, the three versions will be discussed with the translators and any discrepancies solved by consensus. A revised questionnaire and comprehensive report will be produced[51].

Stage 3: Backward translation of T-123 into English will be done by three bilingual English native speakers who are blind to the original version. Translators are "naïve" to self-efficacy and medicine, to minimise bias[52]. Vague wording, obvious inconsistencies or theoretical errors in the translations shall be detected. A report for each version, TB1, TB2 and TB3, will be written by the translators. To maximise comprehension, language will be used which can be understood by a 12 year old[53 54], indicated by a Flesch reading-ease score of 80-90[55]. The German Flesch value=180-ASL-(58,5*ASW), where ASL=average sentence length and ASW=average number of syllables per word[55].

Stage 4: Considering written documentations, an Expert Committee will review and integrate all versions of the questionnaire, involving instructions and scoring documentation, and develop the pre-final version of the USE-MS-G. The Expert Committee will consist of three neurologists, two physiotherapists, a neuropsychologist,

a methodologist, two language professionals, the translators, three lay PwMS and the translation synthesis recorder. The Expert Committee will be in close contact with the original USE-MS developers. A written report of the consensus process will be created. Decision-making will be based on guidelines to accomplish cross-cultural equivalence between the original and German versions in four areas[52], shown in Figure 2.

Figure 2 around here

Figure 2 Cross-cultural equivalence areas to be achieved between original and German USE-MS (USE-MS-G); adapted from[52]

Stage 5: Pretesting of the pre-final USE-MS-G will be performed in 30 PwMS, involving completion of the scale and face-to-face cognitive interviews. Cognitive interviewing will be used to evaluate whether survey questions are easily comprehended, response categories match natural responses, and if people are motivated to respond truthfully and accurately[56-58]. Leading questions will be avoided to minimise bias. Enquiries for comprehension and meaning will be used, and repetition of content by patients[56 58]. Probing will be applied to explore cognitive processes such as memory, underlying reasons for certain responses and overall level of difficulty or confidence[57]. Verbal probes, following Willis' model, will be used immediately after the questions[59]: (a) standardised, anticipated probes: scripted; (b) standardised, conditional probes: scripted, but will be used only if activated by certain participant behaviors such as hesitation[60]; (c) non-standardised, spontaneous probes: flexible, at researcher's digression; and (d) non-standardised, emergent probes: applied in reaction to participant behaviour[61]. The interview guide is presented in Table 2. Recording and field notes will be used, reviewed for inconsistencies or gaps shortly before the end of the interview.

Table 2 Questions used for semi-structured interview

Participants will be given sufficient time to complete the pre-final German USE-MS.	
1.	Having read the questions in the questionnaire, what are your thoughts about them?
2.	Would you please repeat this question in your own words?
3.	What do you think this question is asking?
4.	What do you think about that particular question?
5.	What do you think about the wording of this question, in terms of its clarity?
6.	How easy or hard was this to answer?
7.	How sure are you of your answer?
8.	Could you talk me through your answers in more detail?
9.	What were you thinking of when you answered this question?
10.	Do you have any other comments?
11.	If responses from participants are somewhat unclear, the interviewer asks: “Why so?”
12.	Should a participant hesitate, the interviewer conveys: “You spent some time answering that question - what were you thinking about?”

Adapted from [57] and [62]

An overview of study procedures is presented in Figure 3.

Figure 3 around here

Figure 3 Flowchart of the study procedures

Figure legend: MS: multiple sclerosis; T1 (2; 3): testing 1 (2; 3).

Phase 2

The USE-MS-G will be validated in a larger sample of 292 PwMS who will complete the above described questionnaires at T1 and T2.

Data analyses

Mixed methods data analyses will be used.

Phase 1- Qualitative analyses

Interviews will be transcribed and analysed using qualitative content analysis (QCA)[63 64] using QDA MINER LITE software (Provalis Research, Montreal, Canada) and adhering to the Consolidated Criteria for Reporting Qualitative Research (COREQ)[65]. Analysis steps will be performed as follows[66-69]:

- Data organisation based on the research question
- Identification of recurring ideas, concepts, themes and words
- Development of a coding frame (requirements: unidimensionality, mutual exclusivity of subcategories within dimensions, exhaustiveness of subcategories and saturation, where each subcategory is used at least once)
- Selection of relevant material, structuring, marking and segmentation of text sections, based on Bandura's concept of self-efficacy and the original USE-MS, to identify main and subcategories
- Definition, naming and characterisation of categories and decision rules, to enable consistent assignment of data segments
- Illustration of categories and subcategories using citations
- Creation of a data matrix, followed by quantitative data analysis (descriptive statistics, e.g. frequencies)
- Report
- Rigor and credibility will be maximised by[70-72]

- systematic and consistent approach throughout the analysis
- revision and expansion of the coding frame
- double coding of the whole dataset by two independent researchers (10-14 days after initial coding)
- checking for researcher effects (reflexivity)[73]

Phase 2 - Quantitative analyses

Descriptive statistics and reliability estimates will be performed using IBM SPSS software, release 25.0 (IBM Corporation, Armonk, NY, USA). Rasch Analysis will be conducted with RUMM2030 software[74]. Statistical significance is defined as two-tailed p value <0.05.

Missing data will be treated as follows:

1) Missing data should be avoided by checking questionnaires for missing item responses and asking participants for completion.
2) Rasch analysis calculates an estimate from all available data and does not require a complete data set[75].

Test-retest reliability

Test-retest reliability will be evaluated using Lin's concordance correlation coefficient (r_c) between T2 and T3 (0-1)[76 77]. R_cs will be calculated with their 95% confidence intervals (CI). Values of <0 will be considered to indicate poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1 almost perfect agreement[78]. The data will be racked for the analysis of the concordance correlation coefficient, and stacked for differential item functioning (DIF) by time point.

External validity

It is hypothesised that scores on the German USE-MS will demonstrate moderate to high positive correlations with scales assessing conceptually similar constructs (convergent validity; with the GSE, RS-13 and MusiQol) and moderate to high negative correlations with scales measuring divergent constructs (divergent validity; with the HADS and NFI-MS); Spearman's Rank correlation coefficients of 0.3-0.49 being considered low, 0.5-0.69 moderate and ≥0.7 strong[79].

Internal validity - Rasch analysis

Rasch analysis[25] assumes the probability of a person endorsing an item is a logistic function of the difference between the "person ability" (perceived self-efficacy) and the "item difficulty" (level of self-efficacy) expressed[24]. Item characteristic curves, arranged on the log-odds units (logit) scale, will be used to visualise the probability of a person's correct response in relation to the item difficulty[80].

The polytomous Rasch model will be chosen for this study, suitable for scales with multiple response categories for their items[81]. A significant likelihood ratio test signifying inconsistent distance between response category thresholds would require the use of Masters' unrestricted (partial credit) model[82], otherwise Andrich's rating scale model[83]. Category thresholds are located centrally between two adjacent categories where either response is equally likely[81 84]. The 4-point USE-MS includes three thresholds.

Ordered item category thresholds

Category probability curves will be inspected, checking regular distribution and monotonic advance of measures across categories[84].

Targeting

Targeting refers to the degree to which the scale captures the full range of self-efficacy. Inspecting person-item threshold distribution maps, the mean location score for the respondents will be compared with the default items zero value. A well-targeted scale is centred around zero logits (±0.5 logits), corresponding to the scale's item of mean difficulty[85].

The proportion of floor and ceiling effects will be monitored, considered noteworthy if >5%[86].

Local independence

Local independence means there should be no associations between the items. Inspection of the correlation matrix of item standardised residuals should show Pearson's correlations of <0.2 above the mean value of the matrix as a whole.

Unidimensionality

Unidimensionality as a Rasch model requirement allows a summary score

measurement of a single construct. Using a PCA of the residuals, positively and negatively loadings of the first component will be identified, generating two subsets and separate person estimates. Independent t-tests will explore significant differences[87]. If less than 5% of t-tests are significant or the lower bound of the binominal CI overlaps 5%, unidimensionality is supported[88 89].

Fit to the Rasch model

Different fit statistics will seek to determine if the assumption of a probabilistic ordering of items is satisfied:

(a) Summary Chi-square interaction statistics and individual item Chi-square statistics are expected to be non-significant (Bonferroni-adjusted p-values for the number of items)[48].

(b) Individual person and item fit residuals are expected to be between ±2.5 (99% CI)[90].

(c) Person and summary item fit residuals reflect perfect model fit if their mean and standard deviation are close to 0 and 1, respectively[91 92].

Reliability

Reliability is indicated by the person separation index (PSI; range 0-1)[93] and Cronbach's alpha (missing data excluded), which should be ≥0.85 for individual use or 0.70 for group use[48 94].

Invariance, differential item functioning and differential test functioning

Invariance means that all persons completing a questionnaire, regardless of their ability (or self-efficacy), recognise the difficulty in identical items[93]. Any likelihood of differently scored items between the groups violates the assumption of invariance, called DIF[95 96]. The data will be pooled with a data set from the UK development sample and tested for invariance by language to equate the language versions. Absence of DIF will be tested in gender (female; male), age (quartile groups), disease duration (quartile groups), language (English, German), time point (retest) and centre and indicated by a non-significant ANOVA of the residuals (5% alpha with Bonferroni correction) where the group is the main factor[96 97]. Any observed DIF will be examined to know whether it cancels out at the test level[95]. If there are many items displaying DIF by language, Differential Test Functioning (DTF) will be performed.

If model fit is not achieved, an iterative stepwise procedure will be initiated, involving strategies for combining response categories, stepwise deletion of the worst fitting item, testlet (superitem) construction and adjusting for DIF as appropriate[98].

Ethics approval, permissions and dissemination plan

Ethical approval for both centres was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Due to the absence of an intervention, no insurance policy is required for this study and no harm to participants is expected.

Permission to translate into German and validate the original USE-MS[21] was provided by the test developers who hold the copyright for the USE-MS-G.

Results from this study will be disseminated to the participants via mail and MS Societies (Austria, UK). Findings will be disseminated to clinicians and researchers through peer-reviewed publications and conferences.

Acknowledgements

The authors acknowledge the permission and support from the original USE-MS developers from The Walton Centre for Neurology and Neurosurgery, Liverpool, UK from the University of Leeds, UK, particularly Professor CA Young and Mike Horton.

Author contributions

All authors critically and substantially revised the manuscript and approved the current version to be submitted for publication. BS devised and designed the study and drafted the manuscript. RM provided relevant advice on the Rasch analysis. RK developed the qualitative data analysis plan. LZ substantially contributed to the conception and design of the study. MR provided input on the study methodology and quantitative analysis. SK substantially contributed to the development of the study protocol. FD is a study manager at his centre and substantially contributed to the development of the study protocol. CB is a study manager at his centre and contributed to the design of the study protocol. RE is a study manager at his centre and substantially contributed to the design and development of the study protocol.

Funding

This work was supported by the Austrian MS Research Society (no grant number). Funders had no role in the study design, decision to publish or preparation of the manuscript.

Competing Interests

None declared.

Data sharing statement

Data generated by this research that support any publications will be made available upon reasonable request as soon as possible. It will be considered submitting these data to the Open Science initiative once future analyses related to this data set are completed.

References

1. Pugliatti M, Rosati G, Carton H, et al. The epidemiology of multiple sclerosis in Europe. *European Journal Of Neurology* 2006;13(7):700-22. doi: 10.1111/j.1468-1331.2006.01342.x
2. Compston A, Confavreux C, Lassmann H, et al. McAlpine´s multiple sclerosis. 4th ed ed. London: Elsevier 2006.
3. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs: Prentice-Hall 1986.
4. Bandura A, Schunk DH. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology* 1981;41(3):586-98.
5. Bandura A. Self-efficacy. In: Ramachaudran VS, ed. Encyclopedia of human behavior. New York: Academic Press, 1994:71-81.
6. Jongen PJ, Lemmens WA, Hoogervorst EL, et al. Glatiramer acetate treatment persistence - but not adherence - in multiple sclerosis patients is predicted by health-related quality of life and self-efficacy: a prospective web-based patient-centred study (CAIR study). *Health Qual Life Outcomes* 2017;15(1):50. doi: 10.1186/s12955-017-0622-z
7. Casey B, Uszynski M, Hayes S, et al. Do multiple sclerosis symptoms moderate the relationship between self-efficacy and physical activity in people with multiple sclerosis? *Rehabil Psychol* 2018;63(1):104-10. doi: 10.1037/rep0000190
8. Riazi A, Thompson AJ, Hobart JC. Self-efficacy predicts self-reported health status in multiple sclerosis. *Multiple Sclerosis Journal* 2004;10(1):61-6.
9. Hughes AJ, Beier M, Hartoonian N, et al. Self-efficacy as a longitudinal predictor of perceived cognitive impairment in individuals with multiple sclerosis. *Arch Phys Med Rehabil* 2015;96(5):913-9. doi: 10.1016/j.apmr.2015.01.008
10. Motl RW, Snook EM. Physical Activity, Self-Efficacy, and Quality of Life in Multiple Sclerosis. *Annals of Behavioral Medicine* 2008;35(1):111-15. doi: 10.1007/s12160-007-9006-7
11. Stuifbergen AK, Seraphine A, Roberts G. An explanatory model of health promotion and quality of life in chronic disabling conditions. *Nursing Research* 2000;49(3):122-9.
12. Motl RW, McAuley E, Snook EM, et al. Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. *Psychology Health and Medicine* 2009;14(1):111-24. doi: 10.1080/13548500802241902
13. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for

learning: The OPTIMAL theory of motor learning. *Psychon Bull Rev* 2016;23(5):1382-414. doi: 10.3758/s13423-015-0999-9
14. French MA, Moore MF, Pohlig R, et al. Self-efficacy Mediates the Relationship between Balance/Walking Performance, Activity, and Participation after Stroke. *Top Stroke Rehabil* 2016;23(2):77-83. doi: 10.1080/10749357.2015.1110306
15. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, eds. Measures in health psychology: A user's portfolio Causal and control beliefs. Windsor, UK: NFER-NELSON 1995:35-37.
16. Motl RW, Balto JM, Ensari I, et al. Self-efficacy and Walking Performance in Persons With Multiple Sclerosis. *J Neurol Phys Ther* 2017;41(2):114-18. doi: 10.1097/NPT.0000000000000172
17. Schmitt MM, Goverover Y, Deluca J, et al. Self-efficacy as a predictor of self-reported physical, cognitive, and social functioning in multiple sclerosis. *Rehabil Psychol* 2014;59(1):27-34. doi: 10.1037/a0035288
18. Airlie J, Baker GA, Smith SJ, et al. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. *Clin Rehabil* 2001;15(3):259-65. doi: 10.1191/026921501668362643
19. Rigby SA, Domenech C, Thornton EW, et al. Development and validation of a self-efficacy measure for people with multiple sclerosis: the Multiple Sclerosis Self-efficacy Scale. *Mult Scler* 2003;9(1):73-81. doi: 10.1191/1352458503ms870oa
20. Schwartz CE, Coulthard-Morris L, Zeng Q, et al. Measuring self-efficacy in people with multiple sclerosis: a validation study. *Arch Phys Med Rehabil* 1996;77(4):394-8.
21. Young CA, Mills RJ, Woolmore J, et al. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Mult Scler* 2012;18(9):1326-33. doi: 10.1177/1352458512436592
22. Amtmann D, Bamer AM, Cook KF, et al. University of Washington self-efficacy scale: a new self-efficacy scale for people with disabilities. *Arch Phys Med Rehabil* 2012;93(10):1757-65. doi: 10.1016/j.apmr.2012.05.001
23. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth* 2017;11(Suppl 1):S80-S89. doi: 10.4103/sja.SJA_203_17
24. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press 1980.
25. Rasch G. Probabilistic models for some intelligence and attainment tests (revised and expanded ed.). Chicago The University of Chicago Press 1960.
26. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA* 2018;319(5):483-94. doi: 10.1001/jama.2017.21903
27. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology* 2018;17(2):162-73. doi: 10.1016/s1474-4422(17)30470-2
28. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 2011;69(2):292-302. doi: 10.1002/ana.22366 [published Online First: 2011/03/10]
29. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50(1):121-7.
30. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983;33(11):1444-52. [published Online First: 1983/11/01]
31. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity* 2018;52(4):1893-907. doi:

10.1007/s11135-017-0574-8
32. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions* 1994;7(4):328.
33. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 1996;46(4):907-11.
34. Montalban X, Gold R, Thompson AJ, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler* 2018;24(2):96-120. doi: 10.1177/1352458517751049
35. Diener H-C, Weimar CH. Diagnose und Therapie der Multiplen Sklerose,Gültigkeit der Leitlinie nach Überprüfung durch das Leitliniensekretariat verängert bis 29.2.2017. In: Diener H-C, Weimar C, eds. Leitlinien für Diagnostik und Therapie in der Neurologie. Stuttgart: Kommission "Leitlinien" der Deutschen Gesellschaft für Neurologie, Thieme Verlag 2012/2017.
36. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain* 2000;123 (Pt 5):1027-40.
37. Sharrack B, Hughes RA, Soudain S, et al. The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* 1999;122 (Pt 1):141-59.
38. Meyer-Moock S, Feng YS, Maeurer M, et al. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurology* 2014;14(1):58. doi: 10.1186/1471-2377-14-58
39. Schwarzer R, Jerusalem ME. Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin 1999.
40. Schumacher J, Leppert K, Gunzelmann T, et al. Die Resilienzskala – Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 2004;53:16-39.
41. Leppert K, Koch B, Brähler E, et al. Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13. *Klinische Diagnostik und Evaluation* 2008;2:226–43.
42. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. *J Nurs Meas* 1993;1(2):165-78.
43. Flachenecker P, Vogel U, Simeoni MC, et al. [MusiQol: international questionnaire investigating quality of life in multiple sclerosis: validation results for the German subpopulation in an international comparison]. *Nervenarzt* 2011;82(10):1281-9. doi: 10.1007/s00115-011-3276-9
44. Simeoni M, Auquier P, Fernandez O, et al. Validation of the Multiple Sclerosis International Quality of Life questionnaire. *Mult Scler* 2008;14(2):219-30. doi: 10.1177/1352458507080733
45. Petermann F. Hospital Anxiety and Depression Scale, Deutsche Version (HADS-D). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 2011;59(3):251-53. doi: 10.1024/1661-4747/a000077
46. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67(6):361-70.
47. NFI-MS Neurologischer Fragebogen zur Müdigkeit. *NFI-MS Austria/German - Version of 30 Sep 13 - Mapi ID7555 / NFI-MS_AU10_deu-ATdoc* 2010.
48. Mills RJ, Young CA, Pallant JF, et al. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outcomes* 2010;8:22. doi: 10.1186/1477-7525-8-22
49. Mills RJ, Calabresi M, Tennant A, et al. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). *Mult Scler* 2013;19(4):502-5. doi: 10.1177/1352458512457840
50. Marx RG, Menezes A, Horovitz L, et al. A comparison of two time intervals for test-retest reliability of

health status instruments. *Journal of Clinical Epidemiology* 2003;56(8):730-35. doi: 10.1016/s0895-4356(03)00084-2
51. Beaton D, Bombardier C, Guillemin F, et al. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures. Toronto, Canada: Institute for Work and Health 2007.
52. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417-32.
53. Brislin RW, Lonner WJ, Thorndike RM. Questionnaire wording and translation. Cross-cultural Research Methods. New York: John Wiley 1973:32-58.
54. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79. doi: 10.1186/1477-7525-1-79
55. Flesch R. How to Write Plain English Canterbury, UK: University of Canterbury. Archived from the original on July 12, 2016; 2016 [accessed 4 August 2018.
56. Willis G. Cognitive Interviewing and Questionnaire Design: A Training Manual. Cognitive Methods Staff Working Paper Series No 7. Hyattsville, MD, USA: Office of Research and Methodology, National Center for Health Statistics 1994.
57. Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys. Meeting of the American Statistical Association; 1999 8. - 12. August; Baltimore, Maryland.
58. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly* 2007;71(2):287-311. doi: 10.1093/poq/nfm006
59. Oksenberg L, Cannell CF, Kalton G. New Strategies for Pretesting Survey Questions. *Journal of Official Statistics* 1991;7:349–65.
60. Conrad F, Blair J. Data Quality in Cognitive Interviews: The Case for Verbal Reports. In: Presser S, Rothgeb JM, Couper MP, et al., eds. Methods for Testing and Evaluating Survey Questionnaires. Hoboken, NJ: John Wiley and Sons 2004.
61. Willis G. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage 2005.
62. Olsson JE, Ekblad S, Bertilson BC, et al. Swedish adaptation of the General Medical Council's multisource feedback questionnaires: a qualitative study. *Int J Med Educ* 2018;9:161-69. doi: 10.5116/ijme.5af6.c209
63. Berelson B. Content Analysis in Communication Research. Glencoe: Free Press 1952.
64. Lasswell HD. Power and Personality. New York, NY: W. W. Norton & Company 1948.
65. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042
66. Schreier M. Qualitative Content Analysis in Practice. London, UK: SAGE Publications 2012.
67. Bryman A, Burgess RGE. Analyzing Qualitative Data. Abingdon, Oxon: Routledge 2007.
68. Denzin NK, Lincoln YSE. The SAGE Handbook of Qualitative Research. 5th ed. Thousand Oaks, CA: SAGE Publications 2018.
69. Bryman A. Social Research Methods. 4th ed. New York, NY: Oxford University Press 2012.
70. Cypress BS. Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. *Dimens Crit Care Nurs* 2017;36(4):253-63. doi: 10.1097/DCC.0000000000000253
71. Noble H, Smith J. Issues of validity and reliability in qualitative research. *Evid Based Nurs* 2015;18(2):34-5. doi: 10.1136/eb-2015-102054
72. Leung L. Validity, reliability, and generalizability in qualitative research. *J Family Med Prim Care* 2015;4(3):324-7. doi: 10.4103/2249-4863.161306
73. Miles HB, Huberman AM. Qualitative Data Analysis. Thousand Oaks, CA: Sage 1994.
74. Andrich D, Lyne A, Sheridon B, et al. RUMM 2030. Perth: RUMM Laboratory 2009.

75. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13(12):iii, ix-x, 1-177. doi: 10.3310/hta13120
76. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45(1):255-68.
77. Lin LI-K. Corrections. *Biometrics* 2000;56:324 25.
78. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.
79. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed ed. Boston: Houghton Mifflin 2003.
80. Andrich D. Rasch Models for Measurement Beverly Hills: SAGE 1988.
81. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57(8):1358-62. doi: 10.1002/art.23108
82. Masters G. Rasch model for partial credit scoring. *Psychometrika* 1982;47:149–74.
83. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978;43:561.
84. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3(1):85-106.
85. Khadka J, Pesudovs K, McAlinden C, et al. Reengineering the glaucoma quality of life-15 questionnaire with rasch analysis. *Invest Ophthalmol Vis Sci* 2011;52(9):6971-7. doi: 10.1167/iovs.11-7423
86. Fisher WPJ. Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions* 2007;21(1):1095.
87. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3(2):205-31.
88. Tennant A, Pallant JF. Unidimensionality matters! (a tale of two Smiths?). *Rasch Measurement Transactions* 2006;20:1048-51.
89. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLOS ONE* 2014;9(6):e99485. doi: 10.1371/journal.pone.0099485
90. Andrich D. Rasch models for measurement series: quantitative applications in the social sciences no. 68. London: Sage 1988:68-86.
91. Twiss J, McKenna SP, Graham J, et al. Applying Rasch analysis to evaluate measurement equivalence of different administration formats of the Activity Limitation scale of the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR). *Health Qual Life Outcomes* 2016;14:57. doi: 10.1186/s12955-016-0462-2
92. Milinis K, Tennant A, A. Young C. Rasch analysis of SF-Qualiveen in multiple sclerosis. *Neurourology and Urodynamics* 2017;36(4):1161-66. doi: 10.1002/nau.23081
93. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. London: Lawrence Erlbaum Associates 2007.
94. Linacre JM. Data Variance Explained by Rasch Measures. *Rasch Measurement Transactions* 2006;20(1):1045-54.
95. Tennant A, Pallant JF. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions* 2007;20(4):1082-84.
96. Holland PW, Wainer H. Differential Item Functioning Hillsdale, N. J.: Lawrence Erlbaum 1993.
97. Grimby G. Useful reporting of DIF *Rasch Meas Transactions* 1998;12:651.
98. Siegert RJ, Jackson DM, Tennant A, et al. Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research. *J Rehabil Med* 2010;42(4):302-9. doi: 10.2340/16501977-0511

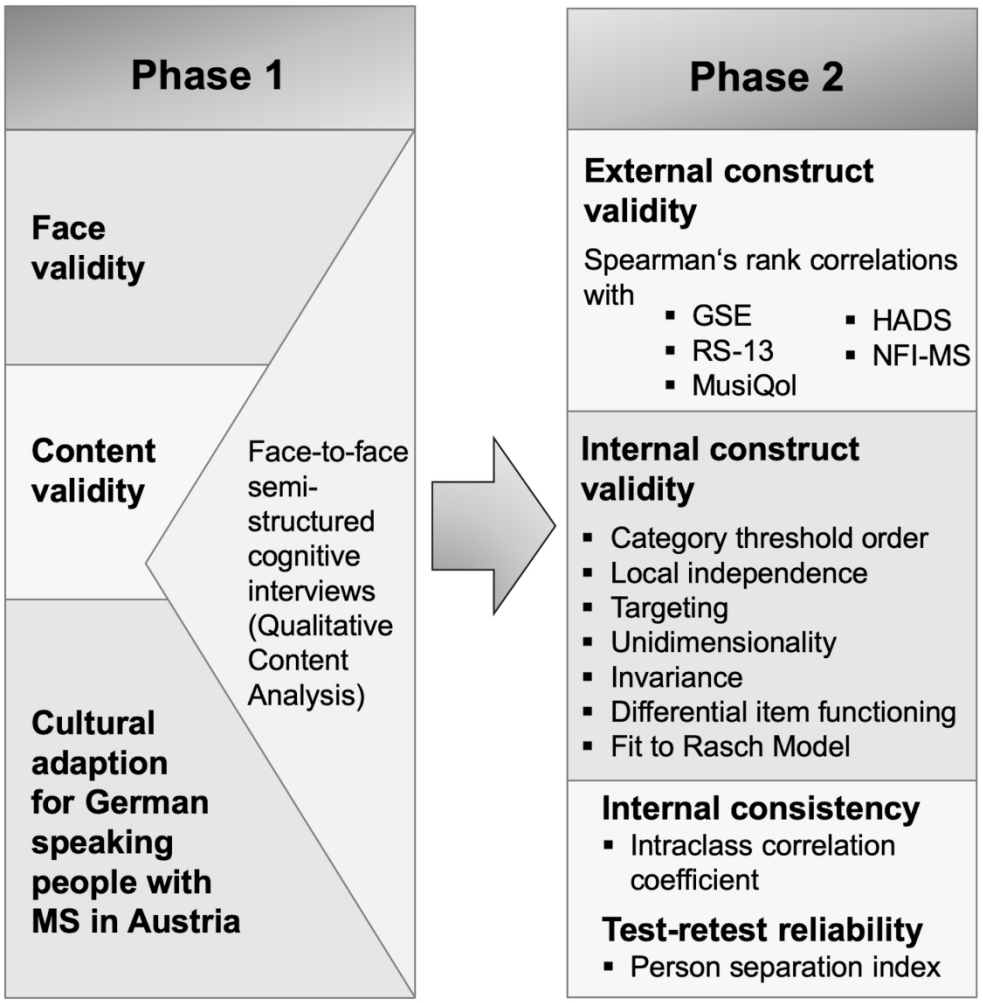

Figure 1

170x173mm (300 x 300 DPI)

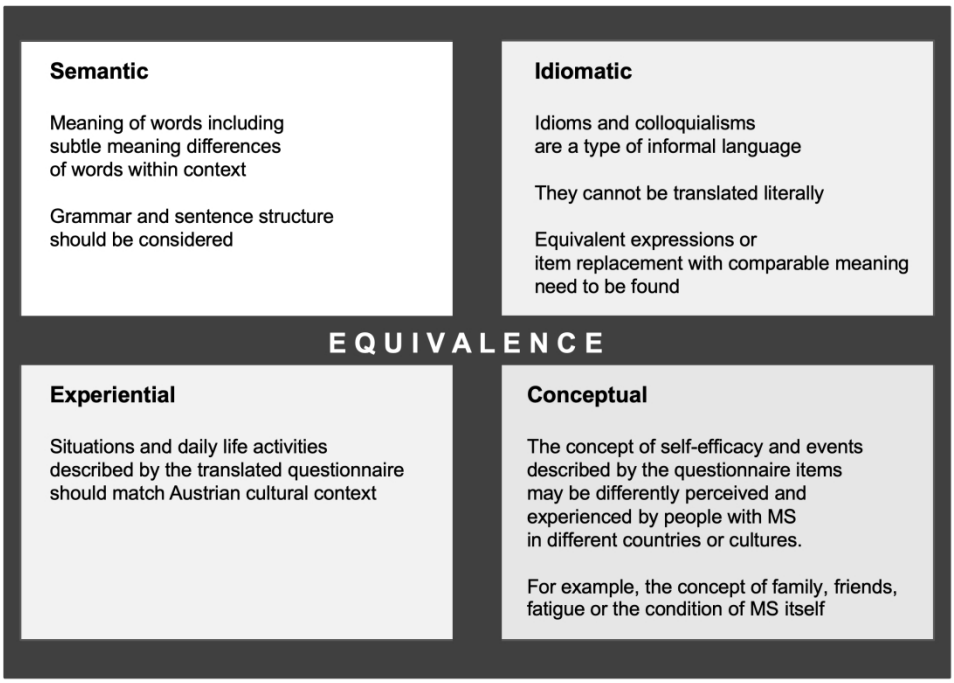

Figure 2

254x190mm (250 x 250 DPI)

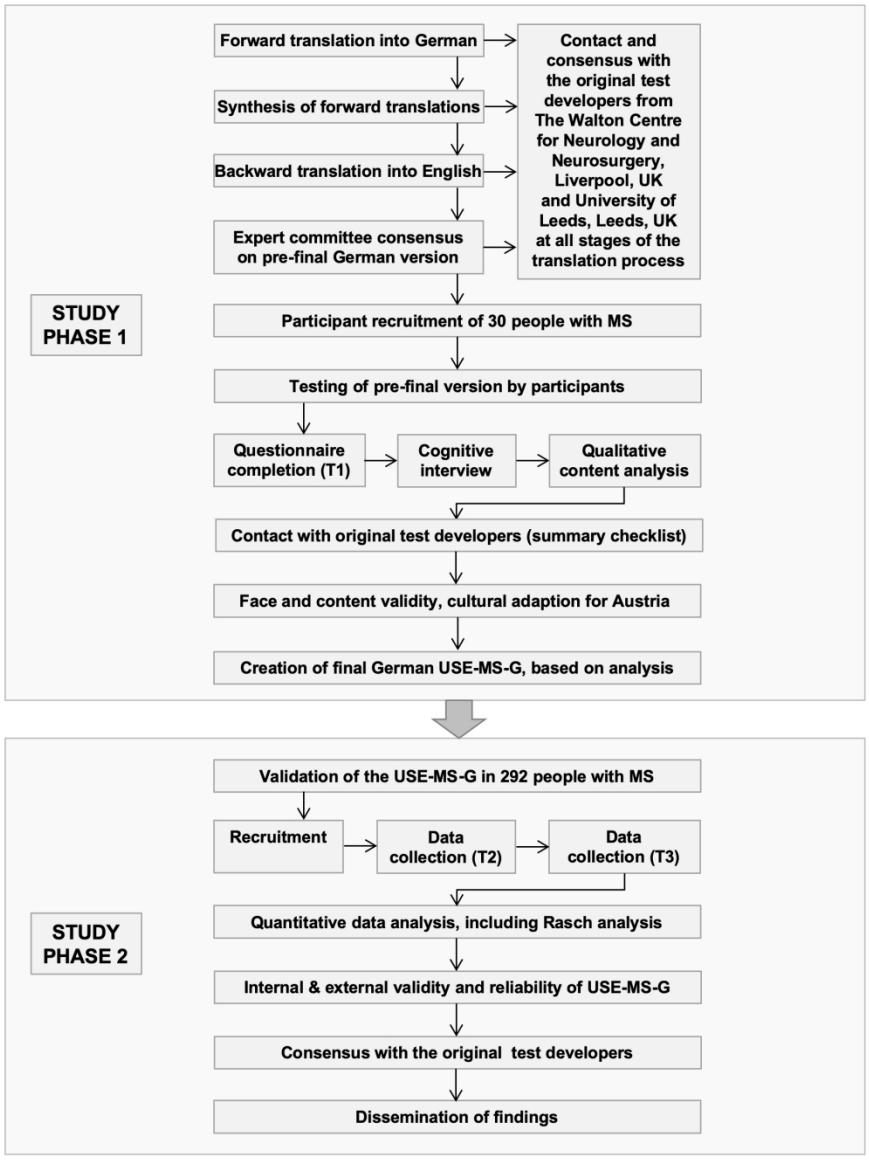

Figure 3

190x254mm (250 x 250 DPI)

SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol
Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. JAMA : the journal of the American Medical Association 2018;319(5):483-94 doi: 10.1001/jama.2017.21903[published Online First: Epub Date])

Section/item	ItemNo	Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.
Administrative information					
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym			1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry			2
	2b	All items from the World Health Organization Trial Registration Data Set			2: described in registry
Protocol version	3	Date and version identifier			3
Funding	4	Sources and types of financial, material, and other support			20

Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors			1; 20
	5b	Name and contact information for the trial sponsor	SPIRIT-5a-PRO Elaboration	Specify the individual(s) responsible for the PRO content of the trial protocol.	BS
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities			20
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)			7; 11-13
Introduction					
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	SPIRIT-6a-PRO Extension	Describe the PRO-specific research question and rationale for PRO assessment and summarize PRO findings in relevant studies.	5-6

	6b	Explanation for choice of comparators			10-11; Supplementary File 2
Objectives	7	Specific objectives or hypotheses	SPIRIT-7-PRO Extension	State specific PRO objectives or hypotheses (including relevant PRO concepts/domains).	6
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)			6-7
Methods: Participants, interventions, and outcomes					
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained			7
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	SPIRIT-10-PRO Extension	Specify any PRO-specific eligibility criteria (eg, language/reading requirements or prerandomization completion of PRO). If PROs will not be collected from the entire study sample, provide a rationale and describe the method for obtaining the PRO subsample.	7

Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered			NA
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)			7-8 (concerning assessments) Interventions: NA
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)			7; 13
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial			7
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	SPIRIT-12-PRO Extension	Specify the PRO concept/domains used to evaluate the intervention (eg, overall health-related quality of life, specific domain, specific symptom and, for each one, the analysis metric (eg, change from baseline, final value, time to event) and the principal time point or period of interest.	8-11; Figure 1

Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	SPIRIT-13-PRO Extension	Include a schedule of PRO assessments, providing a rationale for the time points, and justifying if the initial assessment is not prerandomization. Specify time windows, whether PRO collection is prior to clinical assessments, and if using multiple questionnaires, whether order of administration will be standardized.	7; 11; Table 1; Figure 3
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	SPIRIT-14-PRO Extension	When a PRO is the primary end point, state the required sample size (and how it was determined) and recruitment target (accounting for expected loss to follow-up). If sample size is not established based on the PRO end point, then discuss the power of the principal PRO analyses.	8
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size			7
Methods: Assignment of interventions (for controlled trials)					
Allocation:					NA

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions			NA
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned			NA
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions			NA
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how			NA

	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial			NA
Methods: Data collection, management, and analysis					
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	SPIRIT-18a (i)-PRO Extension	Justify the PRO instrument to be used and describe domains, number of items, recall period, and instrument scaling and scoring (eg, range and direction of scores indicating a good or poor outcome). Evidence of PRO instrument measurement properties, interpretation guidelines, and patient acceptability and burden should be provided or cited if available, ideally in the population of interest. State whether the measure will be used in accordance with any user manual and specify and justify deviations if planned.	8; 11-12
			SPIRIT-18a (ii)-PRO Extension	Include a data collection plan outlining the permitted mode(s) of administration (eg, paper, telephone, electronic, other) and setting (eg, clinic, home, other).	7
			SPIRIT-18a (iii)-PRO Extension	Specify whether more than 1 language version will be used and state whether translated versions have been developed using currently recommended methods.	Total manuscript

			SPIRIT-18a (iv)-PRO Extension	When the trial context requires someone other than a trial participant to answer on his or her behalf (a proxy-reported outcome), state and justify the use of a proxy respondent. Provide or cite evidence of the validity of proxy assessment if available.	NA
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	SPIRIT-18b (i)-PRO Extension	Specify PRO data collection and management strategies for minimizing avoidable missing data.	16
			SPIRIT-18b (ii)-PRO Elaboration	Describe the process of PRO assessment for participants who discontinue or deviate from the assigned intervention protocol.	7; 16
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol			11-12; 15-16

Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	SPIRIT- 20a-PRO Elaboration	State PRO analysis methods, including any plans for addressing multiplicity/type I (α) error.	16-19
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)			16; 18-19
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	SPIRIT- 20c-PRO Elaboration	State how missing data will be described and outline the methods for handling missing items or entire assessments (eg, approach to imputation and sensitivity analyses).	16
Methods: Monitoring					
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed			NA

	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial			NA
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	SPIRIT- 22-PRO Extension	State whether or not PRO data will be monitored during the study to inform the clinical care of individual trial participants and, if so, how this will be managed in a standardized way. Describe how this process will be explained to participants; eg, in the participant information sheet and consent form.	NA 11
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor			NA
Ethics and dissemination					
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval			19

Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)			NA
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)			7
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable			NA
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial			11-12
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site			20
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators			11-12

Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation			NA
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions			8; 19-20
	31b	Authorship eligibility guidelines and any intended use of professional writers			19
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code			20
Appendices					
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates			12
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable			NA

Abbreviations: SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; PRO, patient-reported outcome.
*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license and is reproduced with permission.

Supplementary File 2

Except good to excellent psychometric properties, the selection of outcome measures depended also on patient involvement during their development, patient acceptability of the tools or recommendations from relevant governmental organisations or MS Societies. The MusiQol was developed using individual semi-structured interviews of 1992 PwMS in five countries, followed by qualitative content analysis. The MS-EDGE Outcome Measure Task Force reviewed and recommended the MusiQol for use in PwMS[1]. Similarly, both the German long and short versions of the Resilience Scale are based on face-to-face interviews of 2031 people. The GSE was exposed to repeated validation studies in thousands of participants where excellent validity and reliability were confirmed. A study explored the validity of the HADS in PwMS using interviews by an interviewer who was blind to the HADS scores [2]. The acceptability of the German HADS to various patient populations was tested and confirmed [3]. The UK National Institute for Health and Care Excellence (NICE) 2018 Guidelines for MS[4] recommended the HADS anxiety subscale as a sensitive and specific anxiety screening tool, based on a review[5]. The NFI-MS was developed through a two-stage process in 635 PwMS; quantitative validation and qualitative interviews with patients. The NFI-MS was reviewed and recommended by the MS International Federation (MSIF)[6].

References

1. Potter K, Cohen ET, Allen DD, et al. Outcome measures for individuals with multiple sclerosis: recommendations from the American Physical Therapy Association neurology section task force. Phys Ther 2014;**94**(5):593-608 doi: 10.2522/ptj.20130149[published Online First: Epub Date]|.
2. Watson TM, Ford E, Worthington E, et al. Validation of Mood Measures for People with Multiple Sclerosis. Int J MS Care 2014;**16**:105–09
3. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. Journal of psychosomatic research 1997;**42**(1):17-41
4. Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186). Secondary Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186) 2018.

https://www.nice.org.uk/guidance/CG186/documents/surveillance-review-proposal.

5. Litster B, Fiest KM, Patten SB, et al. Screening Tools for Anxiety in People with Multiple Sclerosis: A Systematic Review. Int J MS Care 2016;**18**(6):273-81 doi: 10.7224/1537-2073.2016-004[published Online First: Epub Date]|.
6. MSIF. Fatigue and MS. MS in focus 2012;**1**:1-28

BMJ Open

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029565.R2
Article Type:	Protocol
Date Submitted by the Author:	27-Jul-2019
Complete List of Authors:	Seebacher, Barbara; Medical University of Innsbruck, Clinical Department of Neurology Mills, Roger J; Walton Centre NHS Foundation Trust, Department of Neurology Reindl, Markus; Medical University of Innsbruck, Clinical Department of Neurology Zamarian, Laura; Medical University of Innsbruck, Clinical Department of Neurology Kuisma, Raija; University of Brighton, School of Health Sciences Kircher, Simone; University Hospital of Innsbruck, Clinical Department of Neurology Brenneis, Christian; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Ehling, Rainer; Clinic for Rehabilitation Münster, Department of Neurology; Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster Deisenhammer, Florian; Medical University of Innsbruck, Clinical Department of Neurology
Primary Subject Heading:	Neurology
Secondary Subject Heading:	Rehabilitation medicine, Patient-centred medicine
Keywords:	Multiple sclerosis < NEUROLOGY, Self Efficacy, Patient Reported Outcome Measures, Austria, Cross-Cultural Comparison, Validation Studies as Topic

GERMAN TRANSLATION, CULTURAL ADAPTION AND VALIDATION OF THE UNIDIMENSIONAL SELF-EFFICACY SCALE FOR MULTIPLE SCLEROSIS: STUDY PROTOCOL

Barbara Seebacher[1], Roger J Mills[2], Markus Reindl[1], Laura Zamarian[1], Raija Kuisma[3], Simone Kircher[4], Christian Brenneis[5, 6], Rainer Ehling[5, 6], Florian Deisenhammer[1]

[1]Clinical Department of Neurology, Medical University of Innsbruck, Innsbruck, Austria

[2]Walton Centre Foundation Trust, Department of Neurology, Liverpool, UK

[3]School of Health Sciences, University of Brighton, Eastbourne, UK

[4]University Hospital of Innsbruck, Clinical Department of Neurology, Innsbruck, Austria

[5]Department of Neurology, Clinic for Rehabilitation Münster, Münster, Austria

[6]Karl Landsteiner Institut für Interdisziplinäre Forschung am Reha Zentrum Münster, Münster, Austria

Correspondence to Dr Barbara Seebacher; Clinical Department of Neurology, Medical University of Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria; phone +43.512.504.82499; barbara.seebacher@i-med.ac.at

Word count: 3972 words excluding declarations; 4218 words including declarations

ABSTRACT

Introduction: Self-efficacy refers to individuals' confidence in their ability to perform relevant tasks to accomplish desired goals. This is independent of their actual abilities. In people with multiple sclerosis (MS), self-efficacy has been shown to powerfully influence motivation and health-related behaviour, such as adherence to prescribed treatment or physical activity. So far, a rigorously tested German language self-efficacy questionnaire for people with MS is missing.

Methods: The purpose of this study is to translate the original Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS) into German and to validate the German USE-MS (USE-MS-G). Based on Bandura's concept of self-efficacy and international guidelines for questionnaire development, the patient-led development of the pre-final German version will involve a forward-backward translation process, synthesis of translations, expert committee review and consensus with the original test developers. At two centres in Tyrol, Austria, content and face validity and cultural adaption for Austria will be established using face-to-face semi-structured cognitive interviews of 30 people with MS. A further 292 people with MS with minimal to severe disability will be tested at two time-points to validate the USE-MS-G.

Results: Mixed methods analyses will be applied. Interviews will be transcribed and analysed employing qualitative content analysis. External validity will be explored using Spearman's Rank correlation coefficients of the USE-MS-G with the 13-item Resilience Scale, General Self-Efficacy Scale, Multiple Sclerosis International Quality of Life questionnaire, Hospital Anxiety and Depression Scale and MS-specific Neurological Fatigue Index. Test-retest reliability, internal consistency and floor and ceiling effects will be evaluated. Internal validity will be examined using Rasch analysis.

Ethics and dissemination: Ethical approval was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Results from this study will be disseminated to the participants and MS Societies, and to clinicians and researchers through peer-reviewed publications and conferences.

Study registration: ISRCTN Registry; trial ID ISRCTN14843579; prospectively registered on 02. 01. 2019; http://www.isrctn.com/ISRCTN14843579

Study protocol, revision 1, 26.6.2019

Keywords: Multiple Sclerosis; Self Efficacy; Patient Reported Outcome Measures; Austria; Cross-Cultural Comparison; Validation Studies as Topic.

ARTICLE SUMMARY

Strengths and limitations of this study

- This study protocol describes the German translation of the original English language Unidimensional Self-efficacy Scale for Multiple Sclerosis (USE-MS), upon permission of the scale developers and applying international recommendations.
- Consistent with the conceptual framework of the English USE-MS, Bandura's concept of self-efficacy will be adhered to.
- Employing a patient-led process in phase 1, 30 people with MS (PwMS) will be interviewed about the pre-final German USE-MS, to establish face and content validity and cultural adaption for PwMS in Austria.
- In phase 2, the German USE-MS will be validated in a larger sample of 292 PwMS.
- Applying classical test theory and Rasch analysis approaches, internal and external validity, internal consistency and test retest reliability will be explored.

INTRODUCTION

Multiple sclerosis (MS) is one of the most common neurological diseases in young adults worldwide, with increasing prevalence[1]. MS is characterised by a wide variety of symptoms and different disease courses[2]. Despite the development of novel disease modifying drugs and neurorehabilitation strategies, the unpredictability of the disease with psychological distress, losses in social contact and quality of life (QoL) are concerning for PwMS. However, individuals' self-knowledge can modulate their approach to day-to-day activities. According to Bandura's social cognitive theory, psychosocial functioning is regulated by reciprocal interactions between behaviour, personal factors and environmental conditions[3]. Self-regulation and intrinsic motivation enable individuals to set and pursue their own goals, observe and evaluate themselves in relation to attained goals[4]. Bandura defined self-efficacy as individuals' beliefs regarding their capability to perform significant tasks, to achieve goals that are meaningful for their daily lives[3]. Self-efficacy beliefs considerably influence people's feelings, thoughts and motivation[5] while, notably, being independent of their physical performance[5]. Such a concept appears important for people with disabilities because it may shape their motivation to initiate and adhere to treatment, particularly when facing side effects.

Perceived self-efficacy influences health-related behaviour such as adhering to medication[6] or engaging in physical activity in PwMS[7]. Health status evaluations of responses to rehabilitation and steroid treatment after an MS relapse can be predicted by self-efficacy levels[8]. Also, higher self-efficacy levels are associated with better long-term perceived cognitive functioning[9] and QoL[10 11]. PwMS who report higher perceived self-efficacy also state lower levels of fatigue, depression and anxiety[12]. Recent evidence has provided insight into the importance of self-management and intrinsic motivation for motor learning[13]. Recognising the relevance of self-efficacy especially for people with disabilities, valid and reliable measurement tools are still needed for its assessment. Three generic self-efficacy scales were found in the literature[7 14-16]. However, generic questionnaires may not adequately cover the construct of self-efficacy in a chronic neurological disease like MS. The initial impact of a diagnosis of MS, in addition to the manifold symptoms and necessity of managing a progressive disease may affect individuals' self-efficacy perceptions. Studies demonstrated that the

capability to effectively solve problems, consistent with higher self-efficacy levels, is strongly associated with PwMS' psychological adaptation to their disability[17], supporting the choice of a disease-specific over a generic self-efficacy questionnaire. MS-specific self-efficacy scales include the Liverpool Self-efficacy Scale (LSES)[18], Multiple Sclerosis Self-Efficacy Scale (MSSS)[19], MS Self-Efficacy Scale (MSSE)[20], Unidimensional Self-Efficacy Scale for Multiple Sclerosis (USE-MS)[21] and University of Washington Self-Efficacy Scale for people with disabilities[22].

Following current guidelines, patients should be involved in the translation and development process of disease specific questionnaires, to ensure the scale reflects their experiences[23]. LSES and MSSS development used in-depth patient interviews while the USE-MS consists of items from both the LSES and MSSS. Bandura's concept of self-efficacy is reflected in the wording of all three questionnaires. The USE-MS study sample was the largest thereof (N=303), and only the USE-MS was exposed to Rasch[24] [25] analysis assessing internal construct validity, in addition to conventional external construct validity and reliability testing. Fit to the Rasch model was demonstrated, and good external validity and reliability[21]. Consequently, the USE-MS appears to be appropriate for use in clinical practice and research. However, so far no validated German language version of the USE-MS is available. The purpose of this study will therefore be to translate the USE-MS into German and validate the German language version in a larger sample of PwMS.

METHODS

Study aims

The first aim of this patient led study is to translate the original English USE-MS, developed by Young et al. (2012) into German, based on international guidelines.

The second aim is to establish face and content validity and cultural adaption of the German version for PwMS in Austria, using individual semi-structured cognitive interviews.

The third aim is to evaluate internal and external validity, internal consistency and test-retest reliability of the German USE-MS (USE-MS-G), using classical test theory and Rasch analysis .

Study design

This will be a bi-centre prospective cross-sectional translation and validation study with repeated measures, consisting of Phase 1 and Phase 2. The SPIRIT 2013 and SPIRIT-PRO Extension checklist for study protocols[26] is presented as Supplementary File 1.

Study setting and timeline

Locations will be the outpatient MS-Clinic of the Clinical Department of Neurology, Medical University of Innsbruck, Austria and Department of Neurology, Clinic for Rehabilitation Münster, Austria.

The expected overall study duration is 33 months, from 01.02.2019 to 31.10. 2021.

Participants and recruitment

A random cross-sectional cohort of patients with clinically definite MS will be recruited from the two centres. Adult (≥18 years) people of any ethnicity and with any MS phenotype according to the McDonald's criteria[27-29] version valid at the time of diagnosis will be included in the study. Their disability status score on the Expanded Disability Status Scale (EDSS)[30] may range from 0 (no disability) to 9.0 (severe disability). Patients will be included if they are able to speak and understand German language. Exclusion criteria are concomitant diseases which may affect subjective self-efficacy ratings (e.g. malignant diseases, other neurological or psychiatric disorders), a relapse of MS within the last two months or any medication change within four weeks prior to the study. A relapse between testing 2 and 3 would necessitate the exclusion of the participant.

The study will be advertised in the MS-Clinic, the Rehabilitation Centre and on the Austrian MS Society website. Further interested PwMS will be examined for eligibility by neurologists at the two study locations. Severely disabled PwMS (EDSS ≥8) will be offered home visits to enable their participation. Written informed consent will be obtained by the first author (BS) who is not involved in the treatment of the patients. Participants may withdraw from the study at any time and for any reasons without prejudice. Outpatient participants will be reimbursed for travel expenses only.

Patient and public involvement

In Phase 1, patients will be lay members of the expert committee to consolidate all the translations and back translations of the USE-MS. Their role regarding the item and response option wording and sentence structure will be crucial, as the final

questionnaire should be understood by PwMS. Patients will also be involved using face-to-face cognitive interviewing, to gain insight into their views about the clarity of the wording, meaning and completeness of the questions of the pre-final USE-MS-G. The Austrian MS (recruitment) and MS Research Societies (funding) will be involved in this study, with whom the findings will be shared as soon as available (patient magazine, meetings). The findings will also be disseminated to the UK MS Society and MS Trust.

Sample size

Phase 1

Patients will be recruited until saturation is achieved. Saturation is a standard term in qualitative methodology to signify the point when the analysis of data from new participants reveals no further emergent qualitative themes. Saturation is typically achieved after 10-30 people have been interviewed but is determined by the nature of the analysis and the participants themselves[31].

Phase 2

Rasch analysis sample size requirements are predicated upon the degree of precision required for estimating item and person difficulties. Regardless of targeting, one can be 99% confident that a sample size of 243 participants is adequately large to obtain a (high) precision of ±0.5 log odd units (logits). Good targeting provided, a sample size of 108 people would be sufficient[21 32]. Using the formula N=n/(1-(z/100)) where n is the calculated number of participants and z the expected attrition rate of 15-20%, a total sample size of 286-304 participants will be aimed at in this study.

Outcomes and data collection

Assessments used in this study were developed using patient involvement and/or recommended by governmental or patient organisations (Supplementary File 2). Study outcomes and methods for their assessment are presented in Figure 1. Participant characteristics and assessments used at all time-points are shown in Table 1.

Figure 1 around here

Figure 1 Study outcomes and their assessment

Figure legend: GSE: General Self-Efficacy Scale; RS-13: Resilience Scale, short version; MusiQol: Multiple Sclerosis International Quality of Life questionnaire; HADS:

Hospital Anxiety and Depression Scale; NFI-MS: Neurological Fatigue Index.

Table 1 Participant characteristics and assessments used in this study

Participant characteristics and assessments (assessments will be collected in a random order to avoid order effect)	**Phase 1**	**Phase 2**	
	T1	**T2**	**T3**
Participant identifier (ID)	**X**	**X**	**X**
Age	**X**	**X**	
Gender	**X**	**X**	
MS phenotype[1]	**X**	**X**	
Disease duration	**X**	**X**	
Expanded Disability Status Scale (EDSS)[2]	**X**	**X**	
Disease modifying treatment (DMT)[3]	**X**	**X**	**X**
(Pre-final) German version of Unidimensional Self-Efficacy Scale for Multiple Sclerosis	**X**	**X**	**X**
Qualitative cognitive interview	**X**		
Resilience Scale, short version		**X**	**X**
General Self-Efficacy Scale		**X**	**X**
Multiple Sclerosis International Quality of Life questionnaire		**X**	**X**
Hospital Anxiety and Depression Scale		**X**	**X**
Neurological Fatigue Index		**X**	**X**

[1]Relapsing-remitting; primary progressive; secondary progressive multiple sclerosis[33]

[2]EDSS groups: 0-4.0; 4.5-6.5; 7.0-7.5; 8.0-9.0[30]

[3](a) No DMTs; (b) low effective DMTs: interferon-b 1a, interferon-b 1a, interferon-b 1b, pegylated interferon-b 1a, glatiramer acetate, dimethyl fumarate, teriflunomide, azathioprin, intravenous immunoglobulins; (c) high effective DMTs: alemtuzumab, cladribine, fingolimod, natalizumab, ocrelizumab, cyclophosphamide, mitoxantrone, rituximab[34 35]

At recruitment, disability will be assessed by neurologists (FD, CB or RE) using the EDSS, ranging from 0 to 10, with higher scores representing higher levels of disability[30].

Although psychometric validation studies criticised its low responsiveness to changes, the EDSS has no floor or ceiling effects[36], has been shown to be valid and reliable[37] and is therefore recommended for use in clinical studies[38].

Excellent internal and external validity and reliability of the original USE-MS has been shown[21]. Scoring of the USE-MS draws results from all 12 items while items 5, 7, 8, 9 and 11 are reversed scored. Higher numbers represent stronger self-efficacy beliefs in participants[21]. The USE-MS includes a 4-point Likert scale (0= strongly disagree to 3=strongly agree).

To assess external construct validity, the following questionnaires will be administered:

The validated German version[39] of the 10-item General Self-Efficacy Scale (GSE)[15] is a self-administered 4-point Likert scale with a summary score ranging from “not at all true” to “exactly true”. The total GSE score ranges between 10 and 40, higher scores signifying greater self-efficacy. Psychometric testing demonstrated high internal consistency, moderate concurrent validity and unidimensionality[15].

The validated German version[40] of the 13-item Resilience Scale (RS-13)[41], based on the 25-item Resilience Scale[42] will be used. RS-13 item scores from a 7-point Likert scale are added up, indicating low (13-66 points), moderate (67-72 points) or high (73-91 points) resilience[41]. The German RS-13 showed high internal consistency and moderate test-retest reliability. Confirmatory factor analysis indicated an acceptable model fit[41].

The validated German version[43] of the 31-item Multiple Sclerosis International Quality of Life (MusiQol) questionnaire[44] will be employed. Response options use a 6-point Likert scale, from 1= “never/not at all” to 5= “always/very much” and 6= “not applicable”. Negatively worded item scores are reversed, and for each participant mean scores for each dimension of the item scores are calculated. All nine dimension scores are linearly transformed to a 0-100 scale, their mean representing the global index score, 0 indicating the worst level of health-related QoL and 100 the best. Psychometric testing showed satisfactory internal and external validity and acceptable reliability for all MusiQol dimensions[44].

The validated German version[45] of the 14-item Hospital Anxiety Depression Scale

(HADS)[46] will be used. The HADS is a self-report questionnaire with a 4-point Likert scale and a 42 point maximum, higher scores representing higher levels of anxiety or depression. Items 2, 4, 7, 9, 12 and 14 are reversed scored, odd items are added to score the anxiety subscale (0-21 points) and even items are added to generate the depression subscale (0-21 points). Testing of the German version demonstrated good internal consistency and acceptable test-retest reliability [45]. The two-factor structure of the scale was confirmed[45].

The validated German version[47] of the 23-item Neurological Fatigue Index (NFI-MS) will be used[48]. Four factors of the NFI-MS were confirmed by principal component analysis and explained 62% of the variance. The four subscales and total scale showed acceptable responsiveness[49], good test-retest reliability, moderate convergent validity and fit to Rasch model expectations[48]. Items are scored on a 4-point Likert scale from 0= “strongly disagree” to 3= “strongly agree”. For scoring, the following item values are added: 1-8= “physical subscale”; 9-12= “cognitive subscale”; 13-18= “relief by diurnal sleep or rest subscale”; 19-23= “abnormal nocturnal sleep and sleepiness subscale”; and 1-7, 9, 11-12 = “physical and cognitive summary score”[48].

Assessments will be performed by trained physiotherapists holding a Master’s (SK) and PhD degree (BS) and a clinical neuropsychologist (LZ). The number of participants who decline to participate or drop out will be recorded, together with reasons (CONSORT flow chart). Any health problems will be recorded.

Phase 1: data will be collected at one time-point (Testing 1, T1), with an expected duration of 45-60 minutes.

Phase 2: for the test-retest reliability assessment, data will be collected at two time-points and will last 60-90 minutes: Testing 2 (T2) and Testing 3 (T3), 14-21 days after T2[48 50].

Data management

With regard to confidentiality, the Austrian and Tyrolean Data Protection Acts will be adhered to. Double data entry and range checks for data values will be used. For qualitative content analysis, double coding of the data set will be performed. Only the research team will have access to the data. All personal data will be codified by a participant ID. Data and files will be saved on a password protected computer, will not be

transferred via emails and will be only used for the purposes for which they were collected. Participants will be informed about their right to disclosure for their own data even if these data lack clinical utility. Codified data will be kept for 15 years following completion of the study. Blank data collection forms can be requested from the corresponding author.

Study procedures

This study will follow the Beaton et al. guidelines for the cross-cultural adaptation of patient reported outcomes[51] and its enhanced version from the University of Leeds, UK.

Phase 1

Stage 1: Forward translation of the items, response options, instructions and scoring information into German will be performed by three independent translators; translator 1 is a medical professional and informed about self-efficacy, while translators 2 and 3 have no medical knowledge and are "naïve" to self-efficacy. Translators are bilingual German native speakers and will create a written report for all translations (T1, T2 and T3), which will then be compared, to distinguish any wording differences or ambiguities[52].

Stage 2 will be a synthesis of T1-3 into T-123. Involving a fourth, unbiased person, the three versions will be discussed with the translators and any discrepancies solved by consensus. A revised questionnaire and comprehensive report will be produced[51].

Stage 3: Backward translation of T-123 into English will be done by three bilingual English native speakers who are blind to the original version. Translators are "naïve" to self-efficacy and medicine, to minimise bias[52]. Vague wording, obvious inconsistencies or theoretical errors in the translations shall be detected. A report for each version, TB1, TB2 and TB3, will be written by the translators. To maximise comprehension, language will be used which can be understood by a 12 year old[53 54], indicated by a Flesch reading-ease score of 80-90[55]. The German Flesch value=180-ASL-(58,5*ASW), where ASL=average sentence length and ASW=average number of syllables per word[55].

Stage 4: Considering written documentations, an Expert Committee will review and integrate all versions of the questionnaire, involving instructions and scoring documentation, and develop the pre-final version of the USE-MS-G. The Expert Committee will consist of three neurologists, two physiotherapists, a neuropsychologist,

a methodologist, two language professionals, the translators, three lay PwMS and the translation synthesis recorder. The Expert Committee will be in close contact with the original USE-MS developers. A written report of the consensus process will be created. Decision-making will be based on guidelines to accomplish cross-cultural equivalence between the original and German versions in four areas[52], shown in Figure 2.

Figure 2 around here

Figure 2 Cross-cultural equivalence areas to be achieved between original and German USE-MS (USE-MS-G); adapted from[52]

Stage 5: Pretesting of the pre-final USE-MS-G will be performed in 30 PwMS, involving completion of the scale and face-to-face cognitive interviews. Cognitive interviewing will be used to evaluate whether survey questions are easily comprehended, response categories match natural responses, and if people are motivated to respond truthfully and accurately[56-58]. Leading questions will be avoided to minimise bias. Enquiries for comprehension and meaning will be used, and repetition of content by patients[56 58]. Probing will be applied to explore cognitive processes such as memory, underlying reasons for certain responses and overall level of difficulty or confidence[57]. Verbal probes, following Willis' model, will be used immediately after the questions[59]: (a) standardised, anticipated probes: scripted; (b) standardised, conditional probes: scripted, but will be used only if activated by certain participant behaviors such as hesitation[60]; (c) non-standardised, spontaneous probes: flexible, at researcher's digression; and (d) non-standardised, emergent probes: applied in reaction to participant behaviour[61]. The interview guide is presented in Table 2. Recording and field notes will be used, reviewed for inconsistencies or gaps shortly before the end of the interview.

Table 2 Questions used for semi-structured interview

Participants will be given sufficient time to complete the pre-final German USE-MS.	
1.	Having read the questions in the questionnaire, what are your thoughts about them?
2.	Would you please repeat this question in your own words?
3.	What do you think this question is asking?
4.	What do you think about that particular question?
5.	What do you think about the wording of this question, in terms of its clarity?
6.	How easy or hard was this to answer?
7.	How sure are you of your answer?
8.	Could you talk me through your answers in more detail?
9.	What were you thinking of when you answered this question?
10.	Do you have any other comments?
11.	If responses from participants are somewhat unclear, the interviewer asks: “Why so?”
12.	Should a participant hesitate, the interviewer conveys: “You spent some time answering that question - what were you thinking about?”

Adapted from [57] and [62]

An overview of study procedures is presented in Figure 3.

Figure 3 around here

Figure 3 Flowchart of the study procedures

Figure legend: MS: multiple sclerosis; T1 (2; 3): testing 1 (2; 3).

Phase 2

The USE-MS-G will be validated in a larger sample of 292 PwMS who will complete the above described questionnaires at T1 and T2.

Data analyses

Mixed methods data analyses will be used.

Phase 1- Qualitative analyses

Interviews will be transcribed and analysed using qualitative content analysis (QCA)[63 64] using QDA MINER LITE software (Provalis Research, Montreal, Canada) and adhering to the Consolidated Criteria for Reporting Qualitative Research (COREQ)[65]. Analysis steps will be performed as follows[66-69]:

- Data organisation based on the research question
- Identification of recurring ideas, concepts, themes and words
- Development of a coding frame (requirements: unidimensionality, mutual exclusivity of subcategories within dimensions, exhaustiveness of subcategories and saturation, where each subcategory is used at least once)
- Selection of relevant material, structuring, marking and segmentation of text sections, based on Bandura's concept of self-efficacy and the original USE-MS, to identify main and subcategories
- Definition, naming and characterisation of categories and decision rules, to enable consistent assignment of data segments
- Illustration of categories and subcategories using citations
- Creation of a data matrix, followed by quantitative data analysis (descriptive statistics, e.g. frequencies)
- Report
- Rigor and credibility will be maximised by[70-72]

- systematic and consistent approach throughout the analysis
- revision and expansion of the coding frame
- double coding of the whole dataset by two independent researchers (10-14 days after initial coding)
- checking for researcher effects (reflexivity)[73]

Phase 2 - Quantitative analyses

Descriptive statistics and reliability estimates will be performed using IBM SPSS software, release 25.0 (IBM Corporation, Armonk, NY, USA). Rasch Analysis will be conducted with RUMM2030 software[74]. Statistical significance is defined as two-tailed p value <0.05.

Missing data will be treated as follows:

1) Missing data should be avoided by checking questionnaires for missing item responses and asking participants for completion.
2) Rasch analysis calculates an estimate from all available data and does not require a complete data set[75].

Test-retest reliability

Test-retest reliability will be evaluated using Lin's concordance correlation coefficient (r_c) between T2 and T3 (0-1)[76 77]. R_cs will be calculated with their 95% confidence intervals (CI). Values of <0 will be considered to indicate poor, 0–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial and 0.81–1 almost perfect agreement[78]. The data will be racked for the analysis of the concordance correlation coefficient, and stacked for differential item functioning (DIF) by time point.

External validity

It is hypothesised that scores on the German USE-MS will demonstrate moderate to high positive correlations with scales assessing conceptually similar constructs (convergent validity; with the GSE, RS-13 and MusiQol) and moderate to high negative correlations with scales measuring divergent constructs (divergent validity; with the HADS and NFI-MS); Spearman's Rank correlation coefficients of 0.3-0.49 being considered low, 0.5-0.69 moderate and ≥0.7 strong[79].

Internal validity - Rasch analysis

Rasch analysis[25] assumes the probability of a person endorsing an item is a logistic function of the difference between the "person ability" (perceived self-efficacy) and the "item difficulty" (level of self-efficacy) expressed[24]. Item characteristic curves, arranged on the log-odds units (logit) scale, will be used to visualise the probability of a person's correct response in relation to the item difficulty[80].

The polytomous Rasch model will be chosen for this study, suitable for scales with multiple response categories for their items[81]. A significant likelihood ratio test signifying inconsistent distance between response category thresholds would require the use of Masters' unrestricted (partial credit) model[82], otherwise Andrich's rating scale model[83]. Category thresholds are located centrally between two adjacent categories where either response is equally likely[81 84]. The 4-point USE-MS includes three thresholds.

Ordered item category thresholds

Category probability curves will be inspected, checking regular distribution and monotonic advance of measures across categories[84].

Targeting

Targeting refers to the degree to which the scale captures the full range of self-efficacy. Inspecting person-item threshold distribution maps, the mean location score for the respondents will be compared with the default items zero value. A well-targeted scale is centred around zero logits (±0.5 logits), corresponding to the scale's item of mean difficulty[85].

The proportion of floor and ceiling effects will be monitored, considered noteworthy if >5%[86].

Local independence

Local independence means there should be no associations between the items. Inspection of the correlation matrix of item standardised residuals should show Pearson's correlations of <0.2 above the mean value of the matrix as a whole.

Unidimensionality

Unidimensionality as a Rasch model requirement allows a summary score

measurement of a single construct. Using a PCA of the residuals, positively and negatively loadings of the first component will be identified, generating two subsets and separate person estimates. Independent t-tests will explore significant differences[87]. If less than 5% of t-tests are significant or the lower bound of the binominal CI overlaps 5%, unidimensionality is supported[88 89].

Fit to the Rasch model

Different fit statistics will seek to determine if the assumption of a probabilistic ordering of items is satisfied:

(a) Summary Chi-square interaction statistics and individual item Chi-square statistics are expected to be non-significant (Bonferroni-adjusted p-values for the number of items)[48].

(b) Individual person and item fit residuals are expected to be between ±2.5 (99% CI)[90].

(c) Person and summary item fit residuals reflect perfect model fit if their mean and standard deviation are close to 0 and 1, respectively[91 92].

Reliability

Reliability is indicated by the person separation index (PSI; range 0-1)[93] and Cronbach's alpha (missing data excluded), which should be ≥0.85 for individual use or 0.70 for group use[48 94].

Invariance, differential item functioning and differential test functioning

Invariance means that all persons completing a questionnaire, regardless of their ability (or self-efficacy), recognise the difficulty in identical items[93]. Any likelihood of differently scored items between the groups violates the assumption of invariance, called DIF[95 96]. The data will be pooled with a data set from the UK development sample and tested for invariance by language to equate the language versions. Absence of DIF will be tested in gender (female; male), age (quartile groups), disease duration (quartile groups), language (English, German), time point (retest) and centre and indicated by a non-significant ANOVA of the residuals (5% alpha with Bonferroni correction) where the group is the main factor[96 97]. Any observed DIF will be examined to know whether it cancels out at the test level[95]. If there are many items displaying DIF by language, Differential Test Functioning (DTF) will be performed.

If model fit is not achieved, an iterative stepwise procedure will be initiated, involving strategies for combining response categories, stepwise deletion of the worst fitting item, testlet (superitem) construction and adjusting for DIF as appropriate[98].

Ethics approval, permissions and dissemination plan

Ethical approval for both centres was received from the Ethics Committee of the Medical University of Innsbruck, Austria (reference number EK1260/2018; 13.12.2018). Due to the absence of an intervention, no insurance policy is required for this study and no harm to participants is expected.

Permission to translate into German and validate the original USE-MS[21] was provided by the test developers who hold the copyright for the USE-MS-G.

Results from this study will be disseminated to the participants via mail and MS Societies (Austria, UK). Findings will be disseminated to clinicians and researchers through peer-reviewed publications and conferences.

Acknowledgements

The authors acknowledge the permission and support from the original USE-MS developers from The Walton Centre for Neurology and Neurosurgery, Liverpool, UK from the University of Leeds, UK, particularly Professor CA Young and Mike Horton.

Author contributions

All authors critically and substantially revised the manuscript and approved the current version to be submitted for publication. BS devised and designed the study and drafted the manuscript. RM provided relevant advice on the Rasch analysis. RK developed the qualitative data analysis plan. LZ substantially contributed to the conception and design of the study. MR provided input on the study methodology and quantitative analysis. SK substantially contributed to the development of the study protocol. FD is a study manager at his centre and substantially contributed to the development of the study protocol. CB is a study manager at his centre and contributed to the design of the study protocol. RE is a study manager at his centre and substantially contributed to the design and development of the study protocol.

Funding

This work was supported by the Austrian MS Research Society (no grant number). Funders had no role in the study design, decision to publish or preparation of the manuscript.

Competing Interests

None declared.

Data sharing statement

Data generated by this research that support any publications will be made available upon reasonable request as soon as possible. It will be considered submitting these data to the Open Science initiative once future analyses related to this data set are completed. The informed consent form will include the consent to controlled data sharing.

References

1. Pugliatti M, Rosati G, Carton H, et al. The epidemiology of multiple sclerosis in Europe. *European Journal Of Neurology* 2006;13(7):700-22. doi: 10.1111/j.1468-1331.2006.01342.x
2. Compston A, Confavreux C, Lassmann H, et al. McAlpine´s multiple sclerosis. 4th ed ed. London: Elsevier 2006.
3. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs: Prentice-Hall 1986.
4. Bandura A, Schunk DH. Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology* 1981;41(3):586-98.
5. Bandura A. Self-efficacy. In: Ramachaudran VS, ed. Encyclopedia of human behavior. New York: Academic Press, 1994:71-81.
6. Jongen PJ, Lemmens WA, Hoogervorst EL, et al. Glatiramer acetate treatment persistence - but not adherence - in multiple sclerosis patients is predicted by health-related quality of life and self-efficacy: a prospective web-based patient-centred study (CAIR study). *Health Qual Life Outcomes* 2017;15(1):50. doi: 10.1186/s12955-017-0622-z
7. Casey B, Uszynski M, Hayes S, et al. Do multiple sclerosis symptoms moderate the relationship between self-efficacy and physical activity in people with multiple sclerosis? *Rehabil Psychol* 2018;63(1):104-10. doi: 10.1037/rep0000190
8. Riazi A, Thompson AJ, Hobart JC. Self-efficacy predicts self-reported health status in multiple sclerosis. *Multiple Sclerosis Journal* 2004;10(1):61-6.
9. Hughes AJ, Beier M, Hartoonian N, et al. Self-efficacy as a longitudinal predictor of perceived cognitive impairment in individuals with multiple sclerosis. *Arch Phys Med Rehabil* 2015;96(5):913-9. doi: 10.1016/j.apmr.2015.01.008
10. Motl RW, Snook EM. Physical Activity, Self-Efficacy, and Quality of Life in Multiple Sclerosis. *Annals of Behavioral Medicine* 2008;35(1):111-15. doi: 10.1007/s12160-007-9006-7
11. Stuifbergen AK, Seraphine A, Roberts G. An explanatory model of health promotion and quality of life in chronic disabling conditions. *Nursing Research* 2000;49(3):122-9.
12. Motl RW, McAuley E, Snook EM, et al. Physical activity and quality of life in multiple sclerosis: intermediary roles of disability, fatigue, mood, pain, self-efficacy and social support. *Psychology Health and Medicine* 2009;14(1):111-24. doi: 10.1080/13548500802241902

13. Wulf G, Lewthwaite R. Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychon Bull Rev* 2016;23(5):1382-414. doi: 10.3758/s13423-015-0999-9
14. French MA, Moore MF, Pohlig R, et al. Self-efficacy Mediates the Relationship between Balance/Walking Performance, Activity, and Participation after Stroke. *Top Stroke Rehabil* 2016;23(2):77-83. doi: 10.1080/10749357.2015.1110306
15. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, eds. Measures in health psychology: A user's portfolio Causal and control beliefs. Windsor, UK: NFER-NELSON 1995:35-37.
16. Motl RW, Balto JM, Ensari I, et al. Self-efficacy and Walking Performance in Persons With Multiple Sclerosis. *J Neurol Phys Ther* 2017;41(2):114-18. doi: 10.1097/NPT.0000000000000172
17. Schmitt MM, Goverover Y, Deluca J, et al. Self-efficacy as a predictor of self-reported physical, cognitive, and social functioning in multiple sclerosis. *Rehabil Psychol* 2014;59(1):27-34. doi: 10.1037/a0035288
18. Airlie J, Baker GA, Smith SJ, et al. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. *Clin Rehabil* 2001;15(3):259-65. doi: 10.1191/026921501668362643
19. Rigby SA, Domenech C, Thornton EW, et al. Development and validation of a self-efficacy measure for people with multiple sclerosis: the Multiple Sclerosis Self-efficacy Scale. *Mult Scler* 2003;9(1):73-81. doi: 10.1191/1352458503ms870oa
20. Schwartz CE, Coulthard-Morris L, Zeng Q, et al. Measuring self-efficacy in people with multiple sclerosis: a validation study. *Arch Phys Med Rehabil* 1996;77(4):394-8.
21. Young CA, Mills RJ, Woolmore J, et al. The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Mult Scler* 2012;18(9):1326-33. doi: 10.1177/1352458512436592
22. Amtmann D, Bamer AM, Cook KF, et al. University of Washington self-efficacy scale: a new self-efficacy scale for people with disabilities. *Arch Phys Med Rehabil* 2012;93(10):1757-65. doi: 10.1016/j.apmr.2012.05.001
23. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth* 2017;11(Suppl 1):S80-S89. doi: 10.4103/sja.SJA_203_17
24. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press 1980.
25. Rasch G. Probabilistic models for some intelligence and attainment tests (revised and expanded ed.). Chicago The University of Chicago Press 1960.
26. Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA* 2018;319(5):483-94. doi: 10.1001/jama.2017.21903
27. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology* 2018;17(2):162-73. doi: 10.1016/s1474-4422(17)30470-2
28. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 2011;69(2):292-302. doi: 10.1002/ana.22366 [published Online First: 2011/03/10]
29. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50(1):121-7.
30. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983;33(11):1444-52. [published Online First: 1983/11/01]
31. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its

conceptualization and operationalization. *Quality & quantity* 2018;52(4):1893-907. doi: 10.1007/s11135-017-0574-8

32. Linacre JM. Sample size and item calibration stability. *Rasch Measurement Transactions* 1994;7(4):328.
33. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 1996;46(4):907-11.
34. Montalban X, Gold R, Thompson AJ, et al. ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis. *Mult Scler* 2018;24(2):96-120. doi: 10.1177/1352458517751049
35. Diener H-C, Weimar CH. Diagnose und Therapie der Multiplen Sklerose,Gültigkeit der Leitlinie nach Überprüfung durch das Leitliniensekretariat verängert bis 29.2.2017. In: Diener H-C, Weimar C, eds. Leitlinien für Diagnostik und Therapie in der Neurologie. Stuttgart: Kommission "Leitlinien" der Deutschen Gesellschaft für Neurologie, Thieme Verlag 2012/2017.
36. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain* 2000;123 (Pt 5):1027-40.
37. Sharrack B, Hughes RA, Soudain S, et al. The psychometric properties of clinical rating scales used in multiple sclerosis. *Brain* 1999;122 (Pt 1):141-59.
38. Meyer-Moock S, Feng YS, Maeurer M, et al. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC Neurology* 2014;14(1):58. doi: 10.1186/1471-2377-14-58
39. Schwarzer R, Jerusalem ME. Skalen zur Erfassung von Lehrer- und Schülermerkmalen: Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin: Freie Universität Berlin 1999.
40. Schumacher J, Leppert K, Gunzelmann T, et al. Die Resilienzskala – Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 2004;53:16-39.
41. Leppert K, Koch B, Brähler E, et al. Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13. *Klinische Diagnostik und Evaluation* 2008;2:226–43.
42. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. *J Nurs Meas* 1993;1(2):165-78.
43. Flachenecker P, Vogel U, Simeoni MC, et al. [MusiQol: international questionnaire investigating quality of life in multiple sclerosis: validation results for the German subpopulation in an international comparison]. *Nervenarzt* 2011;82(10):1281-9. doi: 10.1007/s00115-011-3276-9
44. Simeoni M, Auquier P, Fernandez O, et al. Validation of the Multiple Sclerosis International Quality of Life questionnaire. *Mult Scler* 2008;14(2):219-30. doi: 10.1177/1352458507080733
45. Petermann F. Hospital Anxiety and Depression Scale, Deutsche Version (HADS-D). *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie* 2011;59(3):251-53. doi: 10.1024/1661-4747/a000077
46. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67(6):361-70.
47. NFI-MS Neurologischer Fragebogen zur Müdigkeit. *NFI-MS Austria/German - Version of 30 Sep 13 - Mapi ID7555 / NFI-MS_AU10_deu-ATdoc* 2010.
48. Mills RJ, Young CA, Pallant JF, et al. Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS). *Health Qual Life Outcomes* 2010;8:22. doi: 10.1186/1477-7525-8-22
49. Mills RJ, Calabresi M, Tennant A, et al. Perceived changes and minimum clinically important difference of the Neurological Fatigue Index for multiple sclerosis (NFI-MS). *Mult Scler* 2013;19(4):502-5. doi: 10.1177/1352458512457840

50. Marx RG, Menezes A, Horovitz L, et al. A comparison of two time intervals for test-retest reliability of health status instruments. *Journal of Clinical Epidemiology* 2003;56(8):730-35. doi: 10.1016/s0895-4356(03)00084-2
51. Beaton D, Bombardier C, Guillemin F, et al. Recommendations for the Cross-Cultural Adaptation of the DASH & QuickDASH Outcome Measures. Toronto, Canada: Institute for Work and Health 2007.
52. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46(12):1417-32.
53. Brislin RW, Lonner WJ, Thorndike RM. Questionnaire wording and translation. Cross-cultural Research Methods. New York: John Wiley 1973:32-58.
54. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes* 2003;1:79. doi: 10.1186/1477-7525-1-79
55. Flesch R. How to Write Plain English Canterbury, UK: University of Canterbury. Archived from the original on July 12, 2016; 2016 [accessed 4 August 2018.
56. Willis G. Cognitive Interviewing and Questionnaire Design: A Training Manual. Cognitive Methods Staff Working Paper Series No 7. Hyattsville, MD, USA: Office of Research and Methodology, National Center for Health Statistics 1994.
57. Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys. Meeting of the American Statistical Association; 1999 8. - 12. August; Baltimore, Maryland.
58. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. *Public Opinion Quarterly* 2007;71(2):287-311. doi: 10.1093/poq/nfm006
59. Oksenberg L, Cannell CF, Kalton G. New Strategies for Pretesting Survey Questions. *Journal of Official Statistics* 1991;7:349–65.
60. Conrad F, Blair J. Data Quality in Cognitive Interviews: The Case for Verbal Reports. In: Presser S, Rothgeb JM, Couper MP, et al., eds. Methods for Testing and Evaluating Survey Questionnaires. Hoboken, NJ: John Wiley and Sons 2004.
61. Willis G. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage 2005.
62. Olsson JE, Ekblad S, Bertilson BC, et al. Swedish adaptation of the General Medical Council's multisource feedback questionnaires: a qualitative study. *Int J Med Educ* 2018;9:161-69. doi: 10.5116/ijme.5af6.c209
63. Berelson B. Content Analysis in Communication Research. Glencoe: Free Press 1952.
64. Lasswell HD. Power and Personality. New York, NY: W. W. Norton & Company 1948.
65. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19(6):349-57. doi: 10.1093/intqhc/mzm042
66. Schreier M. Qualitative Content Analysis in Practice. London, UK: SAGE Publications 2012.
67. Bryman A, Burgess RGE. Analyzing Qualitative Data. Abingdon, Oxon: Routledge 2007.
68. Denzin NK, Lincoln YSE. The SAGE Handbook of Qualitative Research. 5th ed. Thousand Oaks, CA: SAGE Publications 2018.
69. Bryman A. Social Research Methods. 4th ed. New York, NY: Oxford University Press 2012.
70. Cypress BS. Rigor or Reliability and Validity in Qualitative Research: Perspectives, Strategies, Reconceptualization, and Recommendations. *Dimens Crit Care Nurs* 2017;36(4):253-63. doi: 10.1097/DCC.0000000000000253
71. Noble H, Smith J. Issues of validity and reliability in qualitative research. *Evid Based Nurs* 2015;18(2):34-5. doi: 10.1136/eb-2015-102054
72. Leung L. Validity, reliability, and generalizability in qualitative research. *J Family Med Prim Care* 2015;4(3):324-7. doi: 10.4103/2249-4863.161306
73. Miles HB, Huberman AM. Qualitative Data Analysis. Thousand Oaks, CA: Sage 1994.

74. Andrich D, Lyne A, Sheridon B, et al. RUMM 2030. Perth: RUMM Laboratory 2009.
75. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009;13(12):iii, ix-x, 1-177. doi: 10.3310/hta13120
76. Lin LI-K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45(1):255-68.
77. Lin LI-K. Corrections. *Biometrics* 2000;56:324 25.
78. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.
79. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed ed. Boston: Houghton Mifflin 2003.
80. Andrich D. Rasch Models for Measurement Beverly Hills: SAGE 1988.
81. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57(8):1358-62. doi: 10.1002/art.23108
82. Masters G. Rasch model for partial credit scoring. *Psychometrika* 1982;47:149–74.
83. Andrich D. Rating formulation for ordered response categories. *Psychometrika* 1978;43:561.
84. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3(1):85-106.
85. Khadka J, Pesudovs K, McAlinden C, et al. Reengineering the glaucoma quality of life-15 questionnaire with rasch analysis. *Invest Ophthalmol Vis Sci* 2011;52(9):6971-7. doi: 10.1167/iovs.11-7423
86. Fisher WPJ. Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions* 2007;21(1):1095.
87. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3(2):205-31.
88. Tennant A, Pallant JF. Unidimensionality matters! (a tale of two Smiths?). *Rasch Measurement Transactions* 2006;20:1048-51.
89. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLOS ONE* 2014;9(6):e99485. doi: 10.1371/journal.pone.0099485
90. Andrich D. Rasch models for measurement series: quantitative applications in the social sciences no. 68. London: Sage 1988:68-86.
91. Twiss J, McKenna SP, Graham J, et al. Applying Rasch analysis to evaluate measurement equivalence of different administration formats of the Activity Limitation scale of the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR). *Health Qual Life Outcomes* 2016;14:57. doi: 10.1186/s12955-016-0462-2
92. Milinis K, Tennant A, A. Young C. Rasch analysis of SF-Qualiveen in multiple sclerosis. *Neurourology and Urodynamics* 2017;36(4):1161-66. doi: 10.1002/nau.23081
93. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. 2nd ed. London: Lawrence Erlbaum Associates 2007.
94. Linacre JM. Data Variance Explained by Rasch Measures. *Rasch Measurement Transactions* 2006;20(1):1045-54.
95. Tennant A, Pallant JF. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions* 2007;20(4):1082-84.
96. Holland PW, Wainer H. Differential Item Functioning Hillsdale, N. J.: Lawrence Erlbaum 1993.
97. Grimby G. Useful reporting of DIF *Rasch Meas Transactions* 1998;12:651.
98. Siegert RJ, Jackson DM, Tennant A, et al. Factor analysis and Rasch analysis of the Zarit Burden Interview for acquired brain injury carer research. *J Rehabil Med* 2010;42(4):302-9. doi: 10.2340/16501977-0511

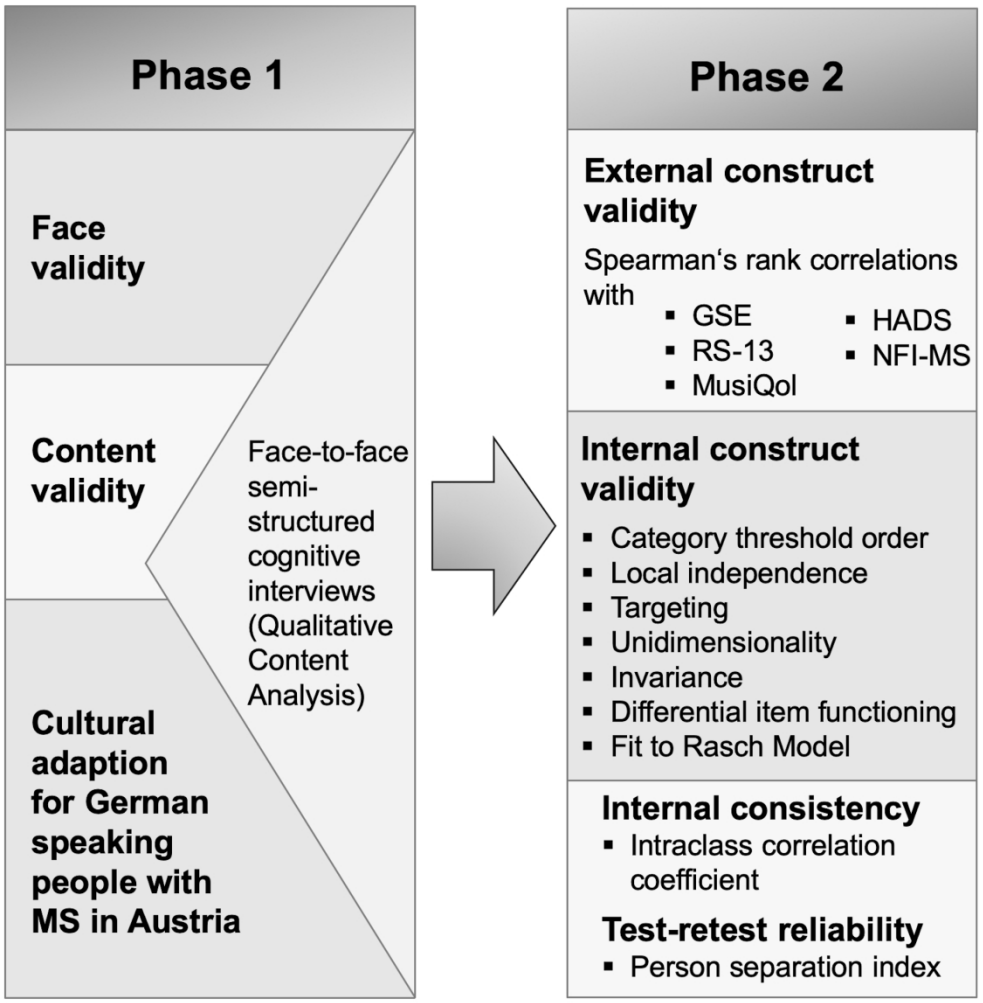

Figure 1

170x173mm (300 x 300 DPI)

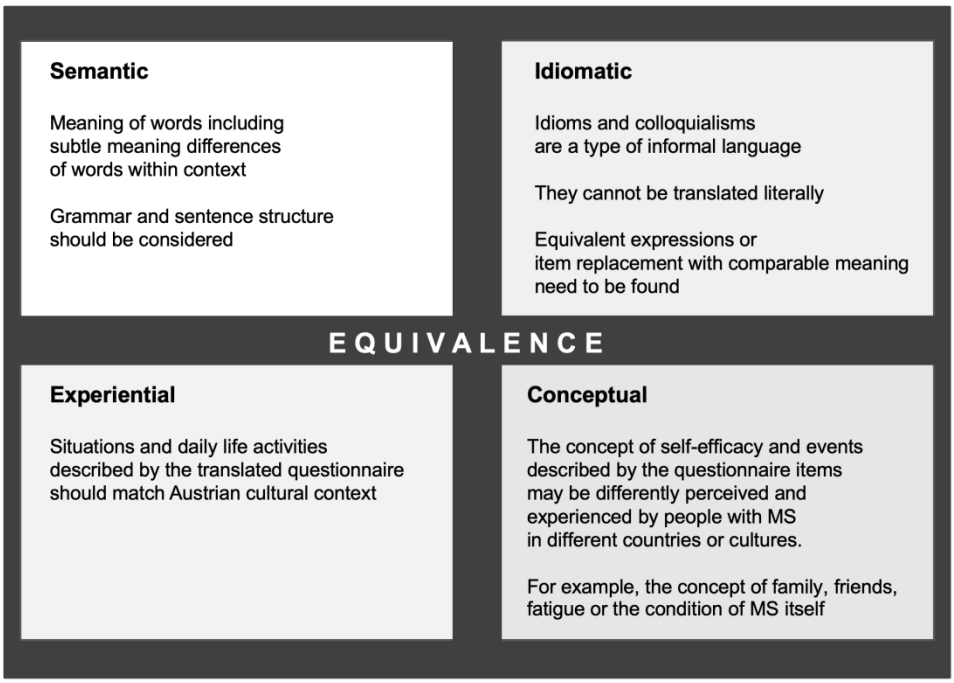

Figure 2

254x190mm (250 x 250 DPI)

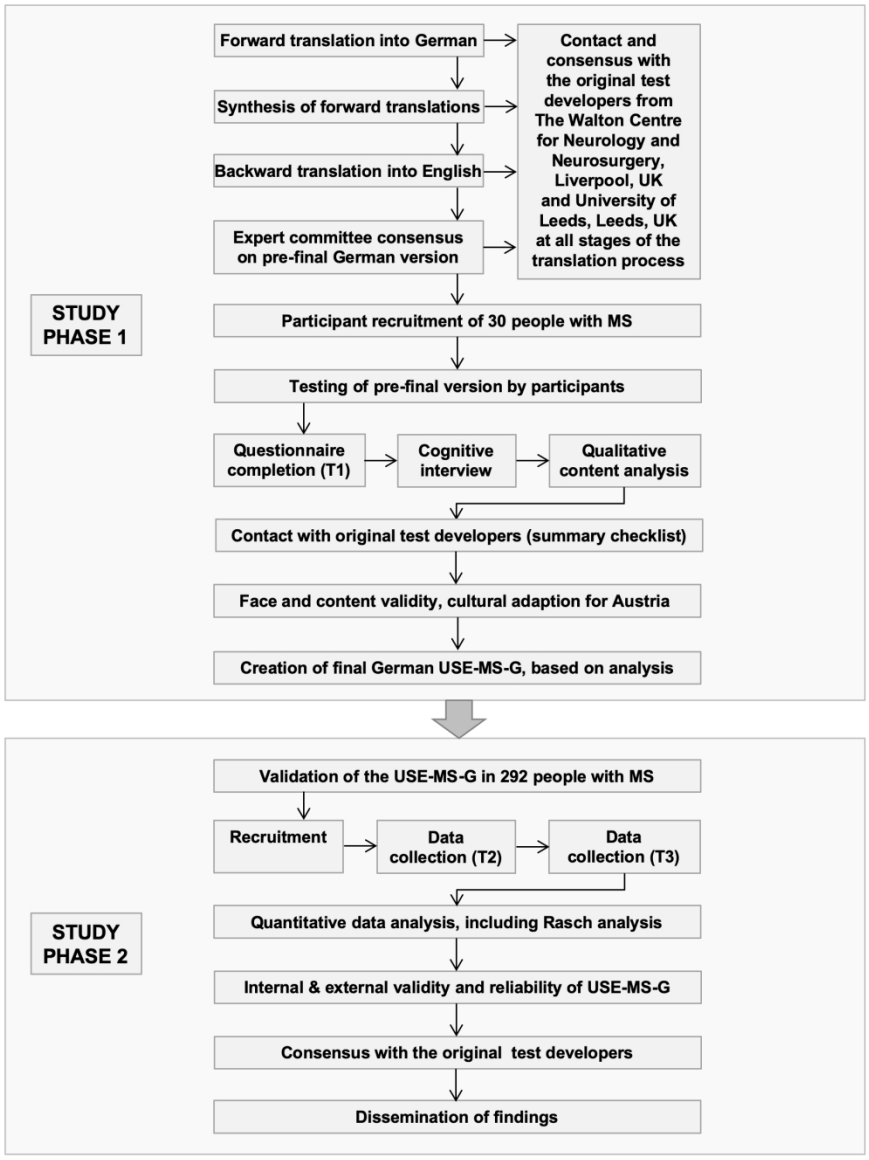

Figure 3

190x254mm (250 x 250 DPI)

SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol
Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. JAMA : the journal of the American Medical Association 2018;319(5):483-94 doi: 10.1001/jama.2017.21903[published Online First: Epub Date])

Section/item	ItemNo	Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.
Administrative information					
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym			1
Trial registration	2a	Trial identifier and registry name. If not yet registered, name of intended registry			2
	2b	All items from the World Health Organization Trial Registration Data Set			2: described in registry
Protocol version	3	Date and version identifier			3
Funding	4	Sources and types of financial, material, and other support			20

Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors			1; 20
	5b	Name and contact information for the trial sponsor	SPIRIT-5a-PRO Elaboration	Specify the individual(s) responsible for the PRO content of the trial protocol.	BS
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities			20
	5d	Composition, roles, and responsibilities of the coordinating centre, steering committee, endpoint adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see Item 21a for data monitoring committee)			7; 11-13
Introduction					
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	SPIRIT-6a-PRO Extension	Describe the PRO-specific research question and rationale for PRO assessment and summarize PRO findings in relevant studies.	5-6

	6b	Explanation for choice of comparators			10-11; Supplementary File 2
Objectives	7	Specific objectives or hypotheses	SPIRIT-7-PRO Extension	State specific PRO objectives or hypotheses (including relevant PRO concepts/domains).	6
Trial design	8	Description of trial design including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)			6-7
Methods: Participants, interventions, and outcomes					
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected. Reference to where list of study sites can be obtained			7
Eligibility criteria	10	Inclusion and exclusion criteria for participants. If applicable, eligibility criteria for study centres and individuals who will perform the interventions (eg, surgeons, psychotherapists)	SPIRIT-10-PRO Extension	Specify any PRO-specific eligibility criteria (eg, language/reading requirements or prerandomization completion of PRO). If PROs will not be collected from the entire study sample, provide a rationale and describe the method for obtaining the PRO subsample.	7

Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered			NA
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)			7-8 (concerning assessments) Interventions: NA
	11c	Strategies to improve adherence to intervention protocols, and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)			7; 13
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial			7
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	SPIRIT-12-PRO Extension	Specify the PRO concept/domains used to evaluate the intervention (eg, overall health-related quality of life, specific domain, specific symptom and, for each one, the analysis metric (eg, change from baseline, final value, time to event) and the principal time point or period of interest.	8-11; Figure 1

Participant timeline	13	Time schedule of enrolment, interventions (including any run-ins and washouts), assessments, and visits for participants. A schematic diagram is highly recommended (see Figure)	SPIRIT-13-PRO Extension	Include a schedule of PRO assessments, providing a rationale for the time points, and justifying if the initial assessment is not prerandomization. Specify time windows, whether PRO collection is prior to clinical assessments, and if using multiple questionnaires, whether order of administration will be standardized.	7; 11; Table 1; Figure 3
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	SPIRIT-14-PRO Extension	When a PRO is the primary end point, state the required sample size (and how it was determined) and recruitment target (accounting for expected loss to follow-up). If sample size is not established based on the PRO end point, then discuss the power of the principal PRO analyses.	8
Recruitment	15	Strategies for achieving adequate participant enrolment to reach target sample size			7
Methods: Assignment of interventions (for controlled trials)					
Allocation:					NA

Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enrol participants or assign interventions			NA
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned			NA
Implementation	16c	Who will generate the allocation sequence, who will enrol participants, and who will assign participants to interventions			NA
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts), and how			NA

	17b	If blinded, circumstances under which unblinding is permissible, and procedure for revealing a participant's allocated intervention during the trial			NA
Methods: Data collection, management, and analysis					
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and a description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known. Reference to where data collection forms can be found, if not in the protocol	SPIRIT-18a (i)-PRO Extension	Justify the PRO instrument to be used and describe domains, number of items, recall period, and instrument scaling and scoring (eg, range and direction of scores indicating a good or poor outcome). Evidence of PRO instrument measurement properties, interpretation guidelines, and patient acceptability and burden should be provided or cited if available, ideally in the population of interest. State whether the measure will be used in accordance with any user manual and specify and justify deviations if planned.	8; 11-12
			SPIRIT-18a (ii)-PRO Extension	Include a data collection plan outlining the permitted mode(s) of administration (eg, paper, telephone, electronic, other) and setting (eg, clinic, home, other).	7
			SPIRIT-18a (iii)-PRO Extension	Specify whether more than 1 language version will be used and state whether translated versions have been developed using currently recommended methods.	Total manuscript

			SPIRIT-18a (iv)-PRO Extension	When the trial context requires someone other than a trial participant to answer on his or her behalf (a proxy-reported outcome), state and justify the use of a proxy respondent. Provide or cite evidence of the validity of proxy assessment if available.	NA
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	SPIRIT-18b (i)-PRO Extension	Specify PRO data collection and management strategies for minimizing avoidable missing data.	16
			SPIRIT-18b (ii)-PRO Elaboration	Describe the process of PRO assessment for participants who discontinue or deviate from the assigned intervention protocol.	7; 16
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values). Reference to where details of data management procedures can be found, if not in the protocol			11-12; 15-16

Statistical methods	20a	Statistical methods for analysing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	SPIRIT- 20a-PRO Elaboration	State PRO analysis methods, including any plans for addressing multiplicity/type I (α) error.	16-19
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)			16; 18-19
	20c	Definition of analysis population relating to protocol non-adherence (eg, as randomised analysis), and any statistical methods to handle missing data (eg, multiple imputation)	SPIRIT- 20c-PRO Elaboration	State how missing data will be described and outline the methods for handling missing items or entire assessments (eg, approach to imputation and sensitivity analyses).	16
Methods: Monitoring					
Data monitoring	21a	Composition of data monitoring committee (DMC); summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol. Alternatively, an explanation of why a DMC is not needed			NA

	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial			NA
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	SPIRIT- 22-PRO Extension	State whether or not PRO data will be monitored during the study to inform the clinical care of individual trial participants and, if so, how this will be managed in a standardized way. Describe how this process will be explained to participants; eg, in the participant information sheet and consent form.	NA 11
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and the sponsor			NA
Ethics and dissemination					
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board (REC/IRB) approval			19

Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, REC/IRBs, trial participants, trial registries, journals, regulators)			NA
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorised surrogates, and how (see Item 32)			7
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable			NA
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained in order to protect confidentiality before, during, and after the trial			11-12
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site			20
Access to data	29	Statement of who will have access to the final trial dataset, and disclosure of contractual agreements that limit such access for investigators			11-12

Ancillary and post-trial care	30	Provisions, if any, for ancillary and post-trial care, and for compensation to those who suffer harm from trial participation			NA
Dissemination policy	31a	Plans for investigators and sponsor to communicate trial results to participants, healthcare professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data sharing arrangements), including any publication restrictions			8; 19-20
	31b	Authorship eligibility guidelines and any intended use of professional writers			19
	31c	Plans, if any, for granting public access to the full protocol, participant-level dataset, and statistical code			20
Appendices					
Informed consent materials	32	Model consent form and other related documentation given to participants and authorised surrogates			12
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable			NA

Abbreviations: SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; PRO, patient-reported outcome.
*It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items. Amendments to the protocol should be tracked and dated. The SPIRIT checklist is copyrighted by the SPIRIT Group under the Creative Commons "Attribution-NonCommercial-NoDerivs 3.0 Unported" license and is reproduced with permission.

Supplementary File 2

Except good to excellent psychometric properties, the selection of outcome measures depended also on patient involvement during their development, patient acceptability of the tools or recommendations from relevant governmental organisations or MS Societies. The MusiQol was developed using individual semi-structured interviews of 1992 PwMS in five countries, followed by qualitative content analysis. The MS-EDGE Outcome Measure Task Force reviewed and recommended the MusiQol for use in PwMS[1]. Similarly, both the German long and short versions of the Resilience Scale are based on face-to-face interviews of 2031 people. The GSE was exposed to repeated validation studies in thousands of participants where excellent validity and reliability were confirmed. A study explored the validity of the HADS in PwMS using interviews by an interviewer who was blind to the HADS scores [2]. The acceptability of the German HADS to various patient populations was tested and confirmed [3]. The UK National Institute for Health and Care Excellence (NICE) 2018 Guidelines for MS[4] recommended the HADS anxiety subscale as a sensitive and specific anxiety screening tool, based on a review[5]. The NFI-MS was developed through a two-stage process in 635 PwMS; quantitative validation and qualitative interviews with patients. The NFI-MS was reviewed and recommended by the MS International Federation (MSIF)[6].

References

1. Potter K, Cohen ET, Allen DD, et al. Outcome measures for individuals with multiple sclerosis: recommendations from the American Physical Therapy Association neurology section task force. Phys Ther 2014;**94**(5):593-608 doi: 10.2522/ptj.20130149[published Online First: Epub Date]|.
2. Watson TM, Ford E, Worthington E, et al. Validation of Mood Measures for People with Multiple Sclerosis. Int J MS Care 2014;**16**:105–09
3. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. Journal of psychosomatic research 1997;**42**(1):17-41
4. Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186). Secondary Surveillance of Multiple sclerosis in adults: management (NICE guideline CG186) 2018.

https://www.nice.org.uk/guidance/CG186/documents/surveillance-review-proposal.

5. Litster B, Fiest KM, Patten SB, et al. Screening Tools for Anxiety in People with Multiple Sclerosis: A Systematic Review. Int J MS Care 2016;**18**(6):273-81 doi: 10.7224/1537-2073.2016-004[published Online First: Epub Date]|.
6. MSIF. Fatigue and MS. MS in focus 2012;**1**:1-28