2023-08

Do clinical trials prepare to fail by failing to prepare? An examination of MS trials and recommendations for patient-reported outcome measure selection

Hobart, J

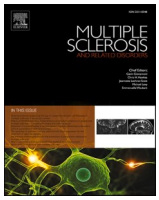
10.1016/j.msard.2023.104788
Multiple Sclerosis and Related Disorders
Elsevier

Multiple Sclerosis and Related Disorders 76 (2023) 104788

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

journal homepage: www.elsevier.com/locate/msard



Review article

Do clinical trials prepare to fail by failing to prepare? An examination of MS trials and recommendations for patient-reported outcome measure selection

Jeremy Hobart [a,*], Tanuja Chitnis [b], Jiwon Oh [c], Laurie Burke [d], Miriam King [e], Pamela Vo [e], Jo Vandercappellen [e], Andrew Lloyd [f]

[a] Peninsula Schools of Medicine and Dentistry, University of Plymouth, Plymouth, UK
[b] Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA
[c] Division of Neurology, St Michael's Hospital, University of Toronto, Toronto, ON, Canada
[d] LORA Group LLC, Normal, IL, USA
[e] Novartis Pharma AG, Basel, Switzerland
[f] Acaster Lloyd Consulting Ltd, London, UK

ARTICLE INFO

Keywords:
Clinical assessment
Clinical measurement
Clinical outcome assessment
Clinical trial
Cost effectiveness research
Medical decision-making
Multiple sclerosis
Patient-reported outcome (PRO)
PRO measure

ABSTRACT

Background: Many clinical trials use patient-reported outcome (PRO) measures, which can influence treatment decision-making, drug approval and label claims. Given that many PRO measure options exist, and there are conceptual and contextual complexities with PRO measurement, we aimed to evaluate how and why specific PRO measures have been selected for pivotal multiple sclerosis (MS) clinical trials. Specifically, we aimed to identify the reasons documented for PRO measure selection in contemporary phase III MS disease-modifying treatment (DMT) clinical trials.

Methods: We searched for phase III clinical trials of MS DMTs published between 2015 and 2021 and evaluated trial protocols, or primary publications where available, for PRO measure selection information. Specifically, we examined study documents for their clarification of clinical concepts measured, definitions of concepts measured, explanations of which PRO measures were considered, why specific PRO measures were chosen, and trade-offs in PRO measure selection.

Results: We identified 1705 abstracts containing 61 unique phase III MS DMT clinical trials. We obtained and examined 27/61 trial protocols. Six protocols were excluded: four contained no mention of PRO measures and two contained redacted sections preventing adequate assessment, leaving 21 protocols for assessment. For the remaining 34 trials (61–27), we retrieved 31 primary publications; 15 primary publications mentioned the use of a PRO measure. None of the 36 clinical trials that mentioned the use of PRO measures (21 protocols and 15 primary publications) documented clear PRO or clinical outcome assessment (COA) measurement strategies, provided clear justifications for PRO selection, or reasons why specific PRO measures were selected when alternatives existed.

Conclusion: PRO measure selection for clinical trials is not evidence-based or underpinned by structured systematic approaches. This represents a critical area for study design improvement as PRO measure results directly affect patient care, PRO measurement has conceptual and contextual complexities, and there is a wide range of options when selecting a PRO measure. We recommend trial designers use formal approaches for PRO measure selection to ensure PRO measurement-based decisions are optimised. We provide a simple, logical, five-stage approach for PRO measure selection in clinical trials.

* Corresponding author.
E-mail address: jeremy.hobart@plymouth.ac.uk (J. Hobart).

https://doi.org/10.1016/j.msard.2023.104788
Received 14 April 2023; Received in revised form 24 May 2023; Accepted 1 June 2023
Available online 7 June 2023

1. Introduction

Pivotal multiple sclerosis (MS) treatment trials frequently include clinical variables best measured using patient reports. Common examples are fatigue, walking ability, and life quality. These variables are measured using self-reported rating scales or questionnaires. The important distinction between the clinical variables for measurement (patient-reported outcomes, PROs), and the methods for their measurement (PRO measures) is often unclear. This article concerns how and why clinical trialists choose PRO measures.

Initially, PROs were exploratory endpoints (Freeman et al., 2001). Now, they are typically secondary (Naismith et al., 2020) or primary (Hobart et al., 2019; Hupperts et al., 2022) endpoints, and subjective methods for interpreting objective measurements (Fisk et al., 2005; Goodman et al., 2010, 2009). Consequently, data generated by PRO measures increasingly influence drug approvals, public expenditure, and personalised clinical decision-making (Butcher et al., 2020). These critical implications mean there can be no justification, scientific or ethical, for weak measurement; or, more precisely, for using weaker-than-is-available PRO measurement as trial designers typically have a choice of PRO measures from which to select. This emphasises the importance of optimizing PRO measure selection so that measured changes accurately approximate real changes in the most meaningful outcome within the context of the targeted population, study design, and analysis plan.

The chosen PRO measure affects the study results because PRO measures differ in structural and performance aspects. For example, PRO measures purporting to be reliable and valid measures of the same clinical concept/variable (e.g., fatigue) can differ in development quality (Close et al., 2023), item number and content, and item response category number and nature. These factors affect PRO measurement performance (validity, reliability, range, precision, ability to detect change). Consequently, different PRO measures, seemingly measuring the same concepts, could reach substantially different conclusions.

PRO measures also have context-dependent performance characteristics. For example, PRO measures have limited measurement ranges. Therefore, the distribution of scores on a PRO measure may have implications for the potential to measure change, depending on the context of use. For example, in the EXPAND study (Kappos et al., 2018), baseline 12-item MS walking scale (MSWS-12v2) scores were skewed (Hobart et al., 2022). Half of participants (44%) were located in the most disabled quartile of the scale range (MSWS-12v2 total score range 32–42). This observed score distribution is not surprising, as EXPAND participants had secondary progressive MS (SPMS). However, the importance lies in the interaction between the PRO (walking ability), PRO measure (MSWS-12v2), patient sample (walking disabled people progressively worsening), conceptualisation of siponimod's treatment effect (aiming to slow progression), and study design (entry criteria, time to endpoint, analysis plan).

There are other context-dependent PRO measurement issues. These include: the nature of the concept of interest (e.g., walking ability), which cannot be assumed to be static across contexts of use; item response dependence, the influence of Time 1 item scores on subsequent timepoint scores (Andrich et al., 2012; Hobart et al., 2022); and differential item functioning (Dib et al., 2017) to assess different PRO measure performance across groups (e.g., treatments, genders and cultures). All these empiric issues underscore the value and importance of carefully selecting 'the best' of existing relevant PRO measures for clinical trials.

2. Objective

To identify phase III MS clinical trial disease-modifying treatment (DMT) studies that used PRO measures and assess whether rationales and justifications for PRO measure selection were provided.

3. Method

3.1. Literature search

Table 1 shows our literature search terms and criteria. We searched for phase III MS DMT clinical trials published between January 2015 and November 2021. This time window was chosen to reflect recent studies, the increasing focus on PRO measures, and to enable a large enough sample size for meaningful inferences. For these trials we attempted to access the full clinical trial protocols, from either ClinicalTrials.gov or directly from trial sponsors. When full trial protocols could not be acquired, we retrieved the study's primary publications.

3.2. Analysis of pro measure selection

From each protocol or primary publication meeting our inclusion criteria, we extracted data concerning PRO measure selection. Table 2 shows our six assessment criteria and simple bespoke scoring system (no

Table 1
Search terms and inclusion criteria.

Databases searched (January 2015 – November 2021)	BIOSIS Previews, Cochrane Database of Systematic reviews, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment, NHS Economic Evaluation Database, Embase, and Medline
Search terms (Search terms were used alone and in combination)	multiple sclerosis (relapsing or relapsing remitting or secondary progressive or primary progressive), RMS, RRMS, SPMS, PPMS, clinical trial, control groups, cross-over studies, double-blind method, random allocation, single-blind method, randomization, randomized controlled trial, randomized controlled trial (topic), controlled clinical trial, controlled clinical trial (topic), controlled clinical trial, controlled clinical trial (topic), controlled clinical trial, controlled clinical trial (topic), clinical trials as topic, single blind procedure, double blind procedure, placebo, placebo effect, randomized, randomization, random, blind, mask, case, report, study, abstract, comment, editorial, letter, news, report, alemtuzumab, alks 8700, amiloride, antegren, anti alpha4 integrin, antivla4, ATX-MS-1467, avonex, aubagio, bexarotene, BG00012, BG 12, BG12, campath, cladribine, clemastine, copaxone, copolymer, co polymer, cop1, daclizumab, dimethylfumarate, diroximel fumarate, DMF, evobrutinib, extavia, fenebrutinib, fingolimod, fluoxetine, fty 720, fty720, fumaric acid dimethyl ester, glatiramer, ibudilast, idebenone, IFN beta, IFN 2 beta, interferon 2 beta, interferon-beta, laquinimod, lemtrada, mabcampath, masitinib, MD1003, mitoxantrone, natalizumab, Nerventra, ocrelizumab, ocrevus, ofatumumab, opicinumab, orelabrutinib, Ozanimod, plegridy, ponesimod, rebif, remibrutinib, riluzole, SAR442168, simvastatin, siponimod, tecfidera, temelimab, teriflunomide, tolebrutinib, tysabri, ublituximab, zinbryta, English language, human, Phase III.
Identification of study protocols and inclusion criteria	• Screened for phase III trials of MS disease-modifying therapies (DMTs) • Excluded abstracts not reporting a phase III MS DMT trial • NCT numbers and/or study names for the retained records obtained • Reviewed trial publications and clinical trial databases to identify study protocols, and also contacted study sponsors via email to request a copy of trial protocols • Excluded protocols that did not mention the use of a PRO measure

Table 2
Analysis of PRO measure selection process and scoring system used.

PRO measure selection criteria	Scoring system
1 Was a PRO measure selection strategy documented?	'Yes' or 'No'. Whether or not there were any basis, logic, reasoning, rationale, justification, motivation, or any other explanation or account, provided in the associated publication as to why the particular PRO measure was selected for use.
2 Was it clear which variables were being measured and why?	'Good' (2): A clear description is provided for what variable is intended for measurement, and exactly why in the specific context. 'Partial' (1): A basic description of the variable being measured is provided, and a brief explanation why. 'None' (0): The protocol simply states the variable being measured, with no further explanation of why this variable is important in the context of the trial.
3 Were the clinical variables intended for measurement defined clearly?	Good' (2): A clear and comprehensive definition is provided for the variable being measured. 'Partial' (1): A basic description of the variable is provided. 'None' (0): The variable is mentioned but no description is provided.
4 Was it clear which PRO measures were considered for inclusion?	'Good' (2): A comprehensive level of information is provided. 'Partial' (1): A minimal level of information is provided. 'None' (0): No information is available.
5 Was there a clear explanation why specific PRO measures were chosen from available candidates?	'Good' (2): A strong justification is provided for why the specific instrument was selected, including why it is more appropriate than other available instruments. 'Partial' (1): A good rationale for why the instrument was chosen, and a basic justification as to why it was chosen over other instruments available. 'None' (0): No justification for why it is the most appropriate for this setting.
6 Were the measurement trade-offs associated with the choice of PRO measure documented?	'Good' (2): A thorough description is given of the trade-offs associated with the specific selected instrument. 'Partial'(1): A brief description is given of the trade-offs associated with the specific selected instrument. 'None' (0): No description of the trade-offs with this instrument is provided.

established scoring exists): 'Good' (2), 'Partial' (1), 'None' (0).

Our first assessment criterion asks whether an overall outcome measurement strategy (clear documentation of which outcomes were chosen for measurement and why) was documented *a priori*. This would include all outcomes measured as endpoints (primary, secondary, exploratory), not simply PROs and their PRO measures, that provide the framework for measurement, and lay the foundation for the more detailed and specific information on each clinical variable.

Our five other assessment criteria focus on PROs. They underpin a logical path for instrument choice when measuring complex clinical variables, where multiple measurement options exist, and different PRO measures have different structures, properties, and context-of-use-specific issues. The five questions are: was it clearly documented which PROs were being measured and why; were the PROs intended for measurement defined (PRO measurement scores need to represent a stated concept); was it clear which PRO measures were considered; was there an explanation why a specific PRO measure was chosen instead of others; were the trade-offs associated with different PRO measures explained?

4. Results

4.1. Literature search

Fig. 1 shows our literature search returned 1705 abstracts. Supplementary Table S1 lists the 61 unique phase III MS DMT clinical trials we identified published between January 2015 and November 2021.

4.1.1. Clinical trial protocols

We retrieved 27/61 protocols (44%): 10 directly from study sponsors, 17 from ClinicalTrials.gov. Six protocols were excluded: four contained no details of PRO measures, two contained redacted sections. ClinicalTrials.gov records for these six trials did not mention PRO measure use (Supplementary Table S1).

Table 3 shows the remaining 21 clinical trials, listed in alphabetic order, mentioning PRO measures in their trial protocols. For the remaining 34 trials, we examined their primary publications and ClinicalTrials.gov records.

4.1.2. Primary publications

We retrieved the primary publications for 31/34 (91%) clinical trials where trial protocols were not available. No primary publications were available for three currently ongoing clinical trials (ENSEMBLE [NCT03085810], HERCULES [NCT04411641] and PERSEUS [NCT04458051]). Only abstracts for these trials were retrieved from the literature. Although none of the abstracts mentioned PRO measure use, ClinicalTrials.gov records indicated all three trials used PRO measures. Supplementary Table S1 details these studies and the PRO measures used.

Of the 31 primary publications available, 15 clinical trials mentioned PRO measures (Table 4). The remaining 16 primary publications, and their ClinicalTrials.gov records, implied no PRO measures were used. Of note, ClinicalTrials.gov records contain considerable variability in the level of clinical trial detail provided. PRO measures may have been used but not documented in primary publications, abstracts, or ClinicalTrials.gov records.

Overall, based on available information, 39/61 (64%) MS clinical trials reported between 2015 and 2021 used PRO measures.

4.2. Analysis of PRO measure selection

Below we provide separate results for the clinical trial protocols and primary publications.

4.2.1. Clinical trial protocols

Table 3 summarizes the information derived from the 21 clinical trial protocols. Relevant information on each clinical trial is provided, and our assessments based on the six criteria are listed in Table 2.

Overall, we rated nearly all criteria as 'None' (score=0) (82/87; 94% PRO measures) as the protocols did not provide robust justifications for the selection of individual PRO measures. For the remaining 5/87 (6%) PRO measures, we rated the quality of the documented rationale as 'Partial' (score=1).

The description of the PRO measures in the protocols assessed was often not provided, or not specific, providing limited or no information to justify their use in the trials. For example, the CASTING protocol (NCT02861014) used the Treatment Satisfaction Questionnaire for Medication (TSQM II). When assessing whether the protocols clarified the variables for measurement, and why the variables were being measured, the protocol states that "TSQM II was used to characterize patient satisfaction with treatment." When considering whether the protocols explained specifically why the chosen PRO measure was selected from those considered, the protocol states "TSQM has been validated using a national panel study of chronic disease." Neither statement adequately answers the questions we have posed.

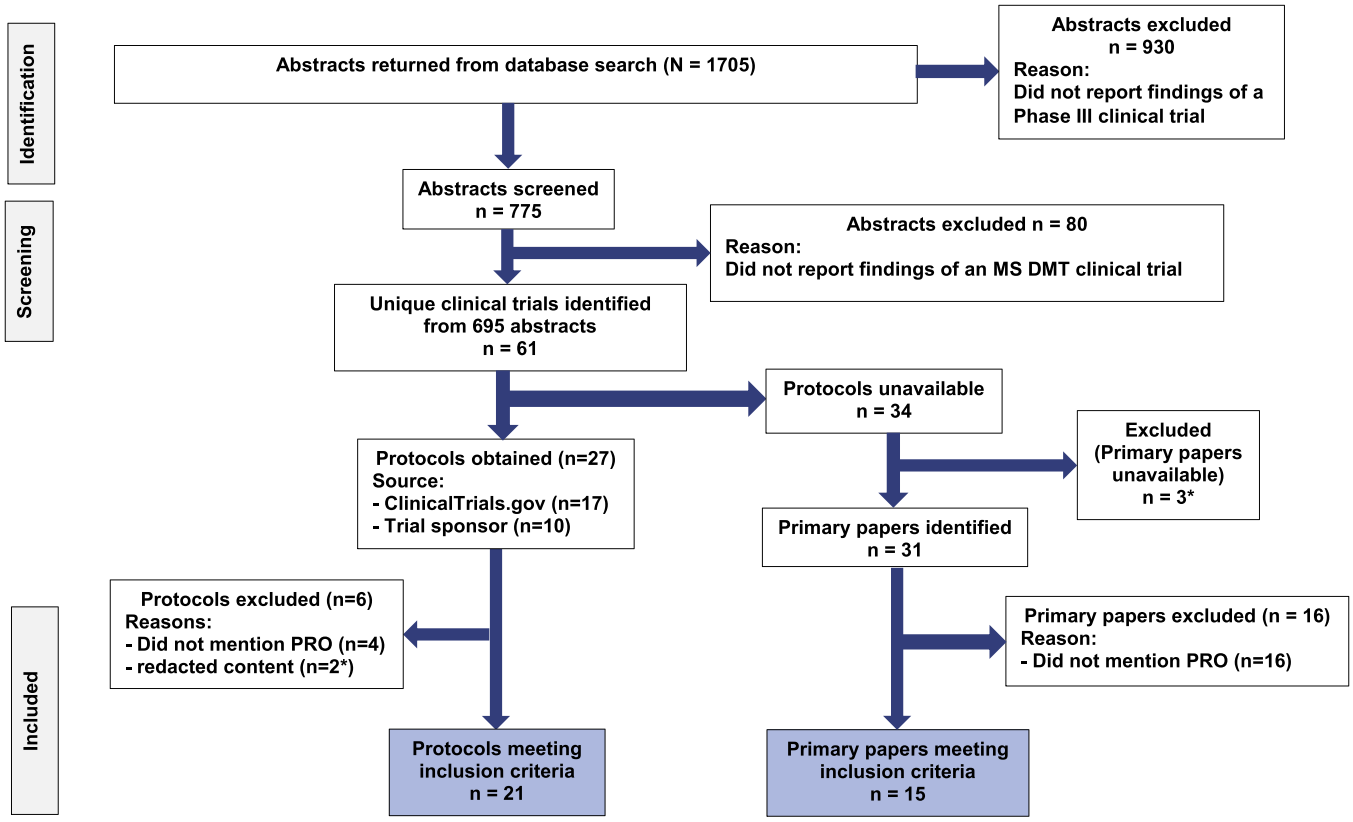


Fig. 1. Flow diagram of literature search process and articles meeting inclusion criteria.
*Conference abstracts were retrieved for the unavailable primary papers and the clinical trial protocols with redacted content. None mentioned the use of a PRO measure. **DMT** disease-modifying treatment, **MS** multiple sclerosis, **PRO** patient-reported outcome.

Was an outcome measurement strategy documented?. Table 3 shows that while 0/21 clinical trial protocols provided an overall outcome measurement, or PRO measure selection strategy, some information/reasoning was provided in one study. ARTIOS (NCT04353492) provided the statement: "PRO measures are included in the study to provide an empirical assessment from the subject's perspective of the benefits of treatment that cannot be gained from magnetic resonance imaging (MRI), Expanded Disability Status Scale (EDSS), or relapse measurement." Whilst this is a general justification for using PRO measures, it does not provide the foundations of a PRO measure selection strategy as there is no information as to why specific concepts were chosen for measurement, nor why specific instruments were chosen to measure those concepts. There are many different clinical concepts which cannot be gained from MRI, EDSS, or relapse measurement, and those concepts can be measured using many different methods, some of which are PRO measures of which there are likely many options. In many cases it could be assumed that the PRO measure was added to support future cost-effectiveness, or other economic or healthcare decision-making research. These points underscore the need for, and importance of, clarification of PRO measure selection.

Was it clear which variables were being measured and why?. Table 3 shows that none of the protocols clearly identified which clinical variables were being measured and why they had been chosen. Though it seems logical that the starting point for choosing a measurement instrument is a clarification of what to measure, this was not stated in any of the protocols we reviewed.

Table 5 lists the PRO measures used in the trials we assessed, the documented concepts they purport to measure (from the original instrument development publications), and the number of trials in which they were used. In almost all cases, there is a description of the concept/s the instrument is purported to measure, but no clear definition of the concepts measured, or the aspects of the concepts measured. For example, Patient Reported Indices in Multiple Sclerosis (PRIMUS) assesses "MS symptoms, activities and quality of life". It is unclear which of the many MS-related symptoms, activities and aspects of life quality are assessed and why. Without specific information it is difficult, if not impossible, to justify the suitability of the PRO measure and its comparison with others.

Excluding FLOODLIGHT and telephone interviews, which are not individual PRO measures *per se*, a total of 32 different PRO measures were used in the 21 trial protocols assessed. Columbia Suicide Severity Rating Scale (C-SSRS; 10/21 protocols), Multiple Sclerosis Impact Scale (MSIS-29; 9/21 protocols) and EQ-5D (7/21 protocols) were the most used PRO measures in the assessed trials. It is not surprising that the C-SSRS is the most widely used PRO measure as FDA mandates suicide risk is measured and describes the C-SSRS as the "gold standard". Table 5 shows the different versions of the same PRO measures grouped together; these are C-SSRS/eC-SSRS, EQ-5D-3L/EQ-5D-5L, 36-item short form health survey (SF-36/SF-36v2), TSQM II/TSQM v1.4/TSQM-9, and Work Productivity and Activity Impairment questionnaire (WPAI/WPAI:MS).

Were the clinical variables intended for measurement defined clearly?. Table 3 shows that 0/21 protocols clearly defined the clinical variables intended for measurement. Table 6 provides exemplars of some of the statements provided in protocols. None of these are clear or partial definitions (i.e., 'Good' [2] or 'Partial' [1]) of the variables intended for measurement. Rather, they are simply statements of fact or assumption. For example, the RAM-MS trial protocol (NCT03477500) reports "The Fatigue Severity Scale (FSS) is designed to differentiate fatigue from clinical depression, since both share same symptoms."

Table 3
Summary of the PRO measures and their selection process documented in the clinical trial protocols.

Study name	Patient population	DMT	Study design	PRO measure	Was a PRO measure selection strategy documented? (Y/N)*	Rationale for PRO measure selection	Was it clear which variables were being measured and why?	Were the clinical variables intended for measurement defined clearly?	Was it clear which PRO measures were considered for inclusion?	Was there a clear explanation why specific PRO measures were chosen from available candidates?	Were the measurement trade-offs associated with the choice of PRO measure documented?
ARTIOS NCT04353492	RMS transitioning from fumarate-based RMS-approved therapies or fingolimod	Ofatumumab	Single-arm, prospective, multicentre, open-label	FLOODLIGHT	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				FSMC		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				HADS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSIS-29		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TSQM v1.4		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ASCLEPIOS I NCT02792218	RMS	Ofatumumab vs teriflunomide	Randomised, double-blind, double-dummy, parallel-group	C-SSRS	N	None (0)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				EQ-5D		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSIS-29		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				WPAI:MS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ASCLEPIOS II NCT02792231	RMS	Ofatumumab vs teriflunomide	Randomised, double-blind, double-dummy, parallel-group	C-SSRS	N	Partial (1)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				EQ-5D		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSIS-29		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Telephone interview		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				WPAI:MS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ASSESS NCT01633112	RRMS	Fingolimod vs glatiramer acetate	Randomised, rater- and dose-blinded	C-SSRS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSIS-29		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				PRIMUS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TSQM v1.4		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
CASTING NCT02861014	RRMS	Ocrelizumab	Open-label	MSIS-29	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Patient Diary (optional)		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SymptoMScreen		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Telephone interviews		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TSQM II		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				WPAI		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
CHORDS NCT02637856	RRMS	Ocrelizumab	Open-label	MSIS-29	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SATMED-Q		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TSQM II		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ENDORSE NCT00835770	RRMS	BG00012	Double-blind, multicentre extension	EQ-5D	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-36		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
EXPAND NCT01665144	SPMS	Siponimod (BAF312)	Randomised, double-blind, parallel-group, placebo-controlled, multicentre, variable treatment duration	C-SSRS	N	Partial (1)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				EQ-5D		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSIS-29 (CP only)		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MSWS-12		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
OLIKOS NCT04486716	RMS	Ofatumumab	Single-arm, prospective, multicentre	C-SSRS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				PGIC		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-12		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TFQ		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				TSQM-9		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
OPERA I NCT01247324	RMS	Ocrelizumab vs interferon β 1A (REBIF)	Randomised, double-blind, double dummy, parallel-group	CES-D	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				C-SSRS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				EQ-5D		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MFIS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)

(*continued on next page*)

Table 3 (*continued*)

Study name	Patient population	DMT	Study design	PRO measure	Was a PRO measure selection strategy documented? (Y/N)*	Rationale for PRO measure selection	Was it clear which variables were being measured and why?	Were the clinical variables intended for measurement defined clearly?	Was it clear which PRO measures were considered for inclusion?	Was there a clear explanation why specific PRO measures were chosen from available candidates?	Were the measurement trade-offs associated with the choice of PRO measure documented?
				SF-36v2		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Telephone interview		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Patient's Assessment of Treatment Benefit		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
OPERA II NCT01412333	RMS	Ocrelizumab vs interferon β 1A (REBIF)	Randomised, double-blind, double dummy, parallel-group	CES-D	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				C-SSRS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				EQ-5D		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MFIS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-36v2		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Telephone interview		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Patient's Assessment of Treatment Benefits		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
OPTIMUM NCT02425644	RMS	Ponesimod vs teriflunomide	Randomised, double-blind, parallel-group, multicentre, active-controlled, superiority	FSIQ-RMS	N	Partial (1)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				PGI-S		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-36v2		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Patient Preference Questionnaire		Partial (1)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				WPAI:MS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ORATORIO NCT01194570	PPMS	Ocrelizumab	Randomised, multicentre, parallel-group, double-blind, placebo-controlled	Patient's Assessment of Treatment Benefit	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				MFIS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-36		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Telephone interviews		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
PARADIGMS NCT01892722	MS (Paediatric)	Fingolimod vs interferon β−1A	A two-year, open-label, rater-blind, randomised, multicentre, active-controlled	C-SSRS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				Peds-QL		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
POINT NCT02907177	RMS	Ponesimod	Randomised, multicentre, double-blind, parallel-group, add-on, superiority	eC-SSRS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				FSIQ-RMS		Partial (1)	None (0)	None (0)	None (0)	Partial (1)	None (0)
				Relapse Assessment Questionnaire		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				SF-36v2		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
				WPAI:MS		None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
RADIANCE NCT02047734	RMS	RPC1063	Randomised, multicentre, double-blind, placebo-controlled (Part A) and double-blind, double-	MSQOL-54	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)

(*continued on next page*)

Table 3 (*continued*)

Study name	Patient population	DMT	Study design	PRO measure	Was a PRO measure selection strategy documented? (Y/N)*	Rationale for PRO measure selection	Was it clear which variables were being measured and why?	Were the clinical variables intended for measurement defined clearly?	Was it clear which PRO measures were considered for inclusion?	Was there a clear explanation why specific PRO measures were chosen from available candidates?	Were the measurement trade-offs associated with the choice of PRO measure documented?
			dummy, active-controlled (Part B), parallel-group								
RAM-MS NCT03477500	RRMS	Alemtuzumab vs cyclophosphamide and anti-thymocyte globuline	Prospective, multicentre, interventional, unblinded, randomised, parallel-group	EQ-5D-5L FSS MSIS-29	N	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)
SPI2 NCT02936037	PMS	MD1003	Randomised, double-blind, placebo-controlled	C-SSRS MSQOL-54 CAREQOL-MS CGI-I	N	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)
STHENOS NCT04788615	RMS	Ofatumumab	Open-label, rater-blind, randomised, multicentre, parallel-arm, active-comparator	MHI-5 MSIS-29 MSTCQ Social life and activities impact TSQM v1.4 Work productivity questionnaire FSIQ-RMS	N	None (0) None (0) None (0) None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0) None (0) None (0) None (0)
ULTIMATE I NCT03277261	RMS	Ublituximab vs teriflunomide	Randomised, multicentre, double-blind, double-dummy, active-controlled	MSQoL54 (inclusive of SF36) FIS	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
ULTIMATE II NCT03277248	RMS	Ublituximab vs teriflunomide	Randomised, multicentre, double-blind, double-dummy, active-controlled	MSQoL54 (inclusive of SF36) FIS	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)

*Whether or not there were any basis, logic, reasoning, rationale, justification, motivation, or any other explanation or account, provided in the associated publication as to why the particular PRO measure was selected for use.

The information provided in the trial protocols has been rated as 'None' (0), 'Partial' (1) or 'Good' (2). Where no information was available, this was categorised as 'None' (0). 'Partial' (1) indicates that a minimal level of information was provided, and 'Good' (2) indicates that a comprehensive level of information was provided.

Abbreviations: CAREQOL-MS Caregiver Health-Related Quality-of-Life in Multiple Sclerosis, CES-D Center for Epidemiologic Studies Depression Scale, CGI-I Clinical Global Impression of Improvement Scale, C-SSRS Columbia Suicide Severity Rating Scale, DMT disease-modifying treatment, eC-SSRS Electronic self-rated version of the Columbia-Suicide Severity Rating Scale, EQ-5D EuroQol Group health status measure (3-level version), FIS Fatigue Impact Scale, FLOODLIGHT Smartphone-based remote tracking device, FSIQ-RMS Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis, FSMC Fatigue Scale for Motor and Cognitive Functions, FSS Fatigue Severity Scale, HADS Hospital Anxiety and Depression Scale, MFIS Modified Fatigue Impact Scale, MHI-5 Mental health inventory – 5 Item, MSIS-29 Multiple Sclerosis Impact Scale, MSQOL-54 Multiple Sclerosis Quality-of-Life (54-item instrument), MSTCQ Multiple sclerosis treatment concerns questionnaire, MSWS-12 Multiple Sclerosis Walking Scale 12, Peds-QL Pediatric Quality-of-Life Inventory, PGIC Patient Global Impression of Change, PGI-S Patient's Global Impression of Severity of Fatigue, PMS Progressive multiple sclerosis, PPMS Primary progressive multiple sclerosis, PRIMUS Patient Reported Indices in Multiple Sclerosis, PRO patient-reported outcome, RMS Relapsing multiple sclerosis, RRMS Relapsing-remitting multiple sclerosis, SATMED-Q The Treatment Satisfaction with Medicines Questionnaire, SF-12 12-Question health questionnaire, SF-36 36-item generic health status measure, SF-36v2 36-item generic health status measure version 2, SPMS Secondary progressive multiple sclerosis, TFQ Trial Feedback Questionnaire, TSQM II Treatment Satisfaction Questionnaire for Medication II, TSQM-1.4 Treatment Satisfaction Questionnaire for Medication Version 1.4, TSQM-9 Treatment Satisfaction Questionnaire for Medication (9-items), WPAI:MS Work Productivity and Activity Impairment questionnaire for multiple sclerosis.

Was it clear which PRO measures were considered for inclusion?. Table 3 shows that 0/21 protocols provided a clear overview of which instruments were considered for measuring the concept (the specific goal of measurement) of interest. Several PRO measure descriptions merely provided a statement that measures were used in accordance with regulatory guidelines, but no rationales were provided based on theoretical/conceptual reasons as to why the chosen instrument was most suitable. Several protocols justified the use of efficacy and safety endpoints, but this was lacking for the PRO endpoints, demonstrating a disparity in the focus placed on the sets of outcomes.

Was there a clear explanation why specific PRO measures were chosen from available candidates?. Table 3 shows that for 81/87 (93%) of assessed PROs, no clear explanation was provided as to why these specific PRO measures were chosen from those available. In several protocols using MSIS-29, no explicit justifications were given as to why this PRO measure was selected above others. The protocols state that MSIS-29 "is considered a reliable, valid and responsive PRO measure that complements other indicators of disease severity used to improve our understanding of the impact of MS." These examples demonstrate that the rationale for PRO measure selection was not included and could be described in more detail to clarify which concepts related to "disease severity" are measured by the MSIS-29, and justify whether or not those concepts are useful and are well measured in the specific context of use.

Tables 3, 7 and 8 show six PRO measures (6/87; 7% of the total PROs assessed) from five protocols where we rated the information documented as 'Partial' (1) justifications for PRO measure selection; however, these 'Partial' justifications do not provide clarification as to why these individual PRO measures were chosen over other available PRO measures. Of these, perhaps the best information was given for the selection of the Patient Preference Questionnaire from OPTIMUM (NCT02425644) which aimed to "capture patient preferences for selected treatment outcomes for use as an additional input to healthcare decisions. An increased understanding of individual values and preferences is the basis for shared decision-making, which in turn encourages patient compliance and health outcomes". However, this statement leaves many relevant questions unanswered.

The only publication to clarify why a specific PRO measure is chosen over those available is shown in the development of the Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis (FSIQ-RMS) indicated in Table 5. In this PRO measure's development publication, the authors state: "Although available PRO instruments have been used to measure fatigue in MS patients, review of their measurement properties suggests shortcomings in terms of current standards for PRO instrument development. For instance, the 9-item Fatigue Severity Scale (FSS) and the 21-item Modified Fatigue Impact Scale (MFIS) do not fit the assumption of unidimensionality, and so studies using their global scores may need to be re-evaluated" (Hudgens et al., 2019). Whilst this justification seems reasonable, only two of many fatigue PRO measures are mentioned, and no steps were initiated or head-to-head comparisons reported to show the new instrument's superiority in their context of use.

Table 7 shows the justifications provided by OPTIMUM (NCT02425644) and POINT (NCT02907177) for using the FSIQ-RMS, which is that the development of the FSIQ-RMS was in accordance with FDA requirements (FDA et al., 2009). However, this is a statement about the development of FSIQ-RMS, from the development paper, rather than an objective critique of the suitability of the FSIQ-RMS. There is no definition of fatigue, conceptualization of how the active treatment might influence fatigue, explanation or empiric evidence why the FSIQ-RMS was considered preferable or superior to competing fatigue PRO measures which also purport to provide conceptually strong, reliable and valid fatigue measurement.

Were the measurement trade-offs associated with the choice of PRO measures explained?. Only the OPTIMUM (NCT02425644) and POINT (NCT02907177) studies provided information about comparison instruments, via the FSIQ-RMS development publication (Hudgens et al., 2019). This was to justify FSIQ-RMS's development. However, there was no consideration of the trade-offs of using the FSIQ-RMS above other fatigue PRO measures, which were deemed (assumed) to be inappropriate.

4.2.2. Primary publications

When clinical trial protocols were unavailable, we retrieved the primary publications associated with these trials to assess whether a PRO measure selection strategy was described. Fifteen of the 31 primary publications (48%) mention PRO measure use. Table 4 shows these 15 trials, the PRO measures used, and selection information documented. None of the primary publications provided a rationale for why the PRO measures had been selected.

We recognize journal articles have tight word limits and that these details might be sacrificed. Also, primary publications may not be the most appropriate platform for discussing PRO measure selection strategies. However, this should be considered when interpreting data derived from these publications. For these reasons, we also assessed the secondary publications of five randomly chosen trials (ACAPELLA, ASCEND, TEMSO, TOPIC, TOWER). No additional information was found to that reported below.

Was an outcome measurement strategy documented?. Table 4 shows that none of the primary publications documented a PRO measure selection strategy.

The EVOLVE-MS-II primary publication provided the most information. The head-to-head study evaluated the gastrointestinal (GI) tolerability of diroximel fumarate (DRF) versus dimethyl fumarate (DMF) in adult patients with RMS (NCT03093324; Naismith et al., 2020). The authors provide good information for the potential suitability of the Individual GI Symptom and Impact Scale (IGISIS) and Global GI Symptom and Impact Scale (GGISIS). Specifically, the IGISIS assesses "the incidence, intensity, onset, duration, and functional impact of five key individual GI symptoms: nausea, vomiting, upper abdominal pain, lower abdominal pain, and diarrhea… In the DMF pivotal DEFINE/CONFIRM trials, these specific GI symptoms were among the most commonly reported adverse effects (AEs) and were the most common GI AEs leading to treatment discontinuation … The GGISIS is designed to assess the overall intensity of five GI symptoms (nausea, vomiting, upper abdominal pain, lower abdominal pain, and diarrhea) experienced during the previous 24 h, the level of interference and functional impact on work and daily activities, and how bothersome GI symptoms were for patients." Nevertheless, there was no discussion of alternative PRO measures, and it would seem logical to pilot the IGISIS and GGISIS, which were adaptations of existing PRO measures, in a relevant sample of people with MS before using in a phase III clinical trial, particularly one evaluating gastrointestinal tolerability.

Was it clear which variables were being measured and why?. Table 4 shows that none of the primary publications provided clear explanations of which variables were being measured. Several of these publications stated the PRO measure used but did not provide any details of the variable being measured. The EVOLVE-MS-II primary publication provided partial information, as described above.

Were the clinical variables intended for measurement defined clearly?. Table 4 shows that none of the primary publications provided explicit definitions of the variables for measurement. The EVOLVE-MS-II primary publication provided partial information, as described above.

Table 4
Summary of the PRO measures and their selection process documented in the clinical trial from the primary publications of clinical trials when no trial protocol was acquired.

Study name	Patient population	DMT	Study design	PRO measure	Was a PRO measure selection strategy documented? (Y/N)*	Rationale for PRO measure selection	Was it clear which variables were being measured and why?	Were the clinical variables intended for measurement defined clearly?	Was it clear which PRO measures were considered for inclusion?	Was there a clear explanation why specific PRO measures were chosen from available candidates?	Were the measurement trade-offs associated with the choice of PRO measures explained?
AB07002 Vermersch et al., (2022)	PPMS or relapse-free SPMS	Masitinib vs placebo	Randomised, double-blind, multicentre, 2 parallel-group, placebo-controlled	EQ-VAS MSQOL-54	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
ACAPELLA Montalban et al. (2017)	PPMS (Adults)	Ocrelizumab vs placebo	Randomised, parallel-group, double-blind, multicentre, placebo-controlled	SF- 36	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
ALLOW Naismith et al. (2019)	RMS	Peginterferon β−1A	Open-label	FLS-S PDDS	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
ASCEND Kapoor et al. (2018)	SPMS Natalizumab-naive patients	Natalizumab vs placebo	Randomised, double-blind, placebo-controlled trial (part 1) with an optional 2-year open-label extension (part 2)	ABILHAND MSIS-29 WPAI	N	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)
DAYBREAK Cree et al. (2022)	RMS	RPC1063	Randomised, double-blind, double-dummy, active-controlled, multicentre, parallel-group	C-SSRS** HADS**	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
DECIDE Benedict et al. (2018)	RRMS	Daclizumab vs interferon β 1a	Randomised, double-blind, multicentre, parallel-group, monotherapy, active-control	MSIS-29**	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
EVOLVE-MS-I Wray et al. (2022)	RRMS	Diroximel fumarate	Open-label	EQ-5D-5L SF-12	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
EVOLVE-MS-II Naismith et al. (2020)	RRMS	ALKS 8700 vs dimethyl fumarate	Randomised, double-blind, head-to-head	GGISIS IGISIS	N	Partial (1) Partial (1)	Partial (1) Partial (1)	Partial (1) Partial (1)	Partial (1) Partial (1)	None (0) None (0)	None (0) None (0)
FREEDOMS II Calabresi et al. (2014)	RRMS	Fingolimod vs placebo	Randomised, double-blind, multicentre, placebo-controlled, parallel-group	EQ-5D mFIS PRIMUS	N	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)	None (0) None (0) None (0)
INFORMS Lublin et al. (2016)	PPMS	Fingolimod vs placebo	Randomised, double-blind, multicentre, placebo-controlled, parallel-group	EQ-5D MSWS-12 PRIMUS U-FIS	N	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)	None (0) None (0) None (0) None (0)
SUNBEAM Comi et al. (2019)	RMS	Ozanimod vs interferon β−1A	Randomised, double-blind, double-dummy, multicentre, active-	MSQOL-54	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)

(*continued on next page*)

Table 4 (*continued*)

Study name	Patient population	DMT	Study design	PRO measure	Was a PRO measure selection strategy documented? (Y/N)*	Rationale for PRO measure selection	Was it clear which variables were being measured and why?	Were the clinical variables intended for measurement defined clearly?	Was it clear which PRO measures were considered for inclusion?	Was there a clear explanation why specific PRO measures were chosen from available candidates?	Were the measurement trade-offs associated with the choice of PRO measures explained?
			controlled, parallel-group								
TEMSO O'Connor et al. (2011)	RMS	Teriflunomide vs placebo	Randomised, placebo-controlled study	FIS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
TENERE Vermersch et al. (2014)	RMS	Teriflunomide vs interferon β−1A	Rater-blinded	FIS TSQM	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)
TOPIC Miller et al. (2014)	First clinical episode suggestive of MS	Teriflunomide vs placebo	Randomised, double-blind, multicentre, placebo-controlled, parallel-group	FIS	N	None (0)	None (0)	None (0)	None (0)	None (0)	None (0)
TOWER Confavreux et al. (2014)	RMS	Teriflunomide vs placebo	Randomised, double-blind, placebo-controlled	FIS SF-36	N	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)	None (0) None (0)

Primary publications were retrieved where clinical trial protocols were unavailable. PRO measurement strategy as defined by our PRO measure selection analysis set out in Table 2. *Whether or not there were any basis, logic, reasoning, rationale, justification, motivation, or any other explanation or account, provided in the associated publication as to why the particular PRO measure was selected for use. **PRO measure mentioned in ClinicalTrials.gov study summary but not in primary publication.

The PRO information provided in the primary publications has been rated as 'None' (0), 'Partial' (1) or 'Good' (2). Where no information was available, this was categorised as 'None' (0). 'Partial' (1) indicates that a minimal level of information was provided, and 'Good' (2) indicates that a comprehensive level of information was provided.

Abbreviations: ABILHAND semi-structured item-response questionnaire that measures manual ability according to an individual's perceived difficulty performing daily bimanual tasks, C-SSRS Columbia Suicide Severity Rating Scale, DMT disease-modifying treatment, EQ-5D EuroQol Group health status measure (3-level version), EQ-5D EuroQol Group health status measure (3-level version), EQ-5D-5L EuroQol Group health status measure 5-level version, EQ-VAS EuroQol Group health status measure visual analogue scale, FIS Fatigue Impact Scale, FLS-S Flu-Like Symptoms Score, GGISIS Gastrointestinal Symptom and Impact Scale, HADS Hospital Anxiety and Depression Scale, IGISIS Individual Gastrointestinal Symptom and Impact Scale, mFIS modified Fatigue Impact Scale, MSIS-29 Multiple Sclerosis Impact Scale, MSQOL-54 Multiple Sclerosis Quality-of-Life (54-item instrument), MSWS-12 Multiple Sclerosis Walking Scale, PDDS Patient-Determined Disease Steps, PPMS Primary progressive multiple sclerosis, PRIMUS Patient Reported Indices in Multiple Sclerosis, PRO patient-reported outcome, RMS Relapsing multiple sclerosis, RRMS relapsing-remitting multiple sclerosis, SF-12 12-item short-form health survey, SF-36 36-item generic health status measure, SPMS Secondary progressive multiple sclerosis, TSQM Treatment Satisfaction Questionnaire for Medication, U-FIS Unidimensional Fatigue Impact Scale, WPAI Work Productivity and Activity Impairment questionnaire.

Table 5
Patient-reported assessment methods and the concepts they purport to measure.

Patient-reported assessment method	Concept patient-reported assessment method purports to measure, according to instrument development publications	Number of protocols describing the patient-reported assessment method
CAREQOL-MS	"…to measure caregiver health-related quality of life (HRQOL) in MS…" The CAREQOL-MS was a useful instrument to measure caregiver quality of life in multiple sclerosis - PubMed (nih.gov)	1
CES-D	"…measure depressive symptomatology in the general population." v01n3p385.pdf (umn.edu)	2
CGI-I	"Severity of illness….takes into account all available information, including a knowledge of the patient's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the patient's ability to function." The Clinical Global Impressions Scale - PMC (nih.gov) ECDEU assessment manual for psychopharmacology (1976 edition) | Open Library	1
C-SSRS and eC-SSRS	"…quantify the severity of suicidal ideation and behavior. The authors examined the psychometric properties of the scale." The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults - PubMed (nih.gov) Feasibility and validation of a computer-automated Columbia-Suicide Severity Rating Scale using interactive voice response technology - PubMed (nih.gov)	10
EQ-5D and EQ-5D-5L	"…instrument for describing and valuing health. It is based on a descriptive system that defines health in terms of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression…designed to measure decrements in health." Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D) - PubMed (nih.gov) CHE Discussion Paper 136.pdf (york.ac.uk) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5 L) - PubMed (nih.gov)	7
FIS	"…improve our understanding of the effects of fatigue on quality of life. The FIS was constructed to include three subscales to assess perceived fatigue impact on cognitive functioning (10 items), physical functioning (10 items) and psychosocial functioning (20 items)." Measuring the functional impact of fatigue: initial validation of the fatigue impact scale - PubMed (nih.gov)	2
FLOODLIGHT*	"…captures reliable and clinically relevant measures of functional impairment in MS…assessed the functional ability across three key domains affected by MS: cognition, upper extremity function, and gait and balance." Adherence and Satisfaction of Smartphone- and Smartwatch-Based Remote Active Testing and Passive Monitoring in People With Multiple Sclerosis: Nonrandomized Interventional Feasibility Study - PMC (nih.gov) A smartphone sensor-based digital outcome assessment of multiple sclerosis - PubMed (nih.gov)	1
FSIQ-RMS	"…assess fatigue symptoms relevant to patients within the spectrum of RMS and the relevant impact of these symptoms on patients' lives, in accordance with the FDA PRO guidance." Development and Validation of the FSIQ-RMS: A New Patient-Reported Questionnaire to Assess Symptoms and Impacts of Fatigue in Relapsing Multiple Sclerosis - PubMed (nih.gov)	3
FSMC	"…for the assessment of MS-related cognitive and motor fatigue…provides differential quantification and graduation of cognitive and motor fatigue… focuses on the two main reported domains of fatigue (i.e. cognitive and motor)…represents a new patient reported outcome measure for measuring mental and physical fatigue." The Fatigue Scale for Motor and Cognitive Functions (FSMC): validation of a new instrument to assess multiple sclerosis-related fatigue - PubMed (nih.gov)	1
FSS	"…detect clinically predicted changes in fatigue over time…assesses disabling fatigue across two different clinical disorders…the scale was successful in identifying features of fatigue that are specific to the medically ill." The fatigue severity scale. Application to patients with multiple sclerosis and systemic lupus erythematosus - PubMed (nih.gov)	1
HADS	"…a reliable instrument for detecting states of depression and anxiety in the setting of an hospital medical outpatient clinic. The anxiety and depressive subscales are also valid measures of severity of the emotional disorder…self-assessment mood scale specifically designed for use in non-psychiatric hospital departments." The hospital anxiety and depression scale - PubMed (nih.gov)	1
MFIS	"… a multidimensional scale developed to assess the perceived impact of fatigue on a variety of daily activities. The items of the MFIS can be aggregated into three subscales (physical, cognitive, and psychosocial)…" Measuring the functional impact of fatigue: initial validation of the fatigue impact scale - PubMed (nih.gov) Microsoft Word - NEW_MSQLI_Cover.doc (nationalmssociety.org)	3
MHI-5	"…measure of psychological distress and well-being, developed for use in general populations." The structure of psychological distress and well-being in general populations - PubMed (nih.gov)	1
MSIS-29	"…patient-based outcome measure of the impact of multiple sclerosis suitable for clinical trials and epidemiological studies. The MSIS-29 is a measure of the physical and psychological impact of multiple sclerosis from the patients' perspective." The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-basedoutcome measure - PubMed (nih.gov)	9
MSQOL-54	"…self-report measure of HRQOL for MS that combines the strengths of generic and disease-targeted approaches to HRQOL measurement…assesses HRQOL for individuals with a chronic neurological condition, multiple sclerosis, using 54 items that define 12 multiple-item scales." A health-related quality of life measure for multiple sclerosis - PubMed (nih.gov)	4

(*continued on next page*)

Table 5 (*continued*)

Patient-reported assessment method	Concept patient-reported assessment method purports to measure, according to instrument development publications	Number of protocols describing the patient-reported assessment method
MSTCQ	"…Multiple Sclerosis Treatment Concerns Questionnaire (MSTCQ) and pain measures after introduction of the new Rebiject II™ injection system would allow determination of changes perceived by patients…assessed patient perceptions of the multiple domains associated with use of an injection device for IFN-β−1a." Patient satisfaction with an injection device for multiple sclerosis treatment - PubMed (nih.gov)	1
MSWS-12	"…a patient-based measure of walking ability in MS. The MSWS-12 satisfies standard criteria as a reliable and valid patient-based measure of the impact of MS on walking…multi-item rating scale of walking that combines patients' perspectives with psychometric methods and is suitable for epidemiologic studies, clinical trials, and routine data collection for audit purposes." Measuring the impact of MS on walking ability: the 12-Item MS Walking Scale (MSWS-12) - PubMed (nih.gov)	1
Patient's Assessment of Treatment Benefit	No specific information for broad description of PRO instrument.	3
Patient Diary (optional)	No specific information for broad description of PRO instrument.	1
Patient Preference Questionnaire	No specific information for broad description of PRO instrument.	1
Peds-QL	"… assesses patient's and parent's perceptions of HRQOL in pediatric patients with chronic health conditions using the pediatric cancer as an exemplary model…integrated core and modular measure of HRQOL for pediatric chronic health conditions." The PedsQL: measurement model for the pediatric quality of life inventory - PubMed (nih.gov)	1
PGIC	"The CGI was developed for use in NIMH-sponsored clinical trials to provide a brief, stand-alone assessment of the clinician's view of the patient's global functioning prior to and after initiating a study medication…takes into account all available information, including a knowledge of the patient's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the patient's ability to function." The Clinical Global Impressions Scale - PMC (nih.gov) ECDEU assessment manual for psychopharmacology (1976 edition) | Open Library	1
PGI-S	"…designed to identify and assess symptoms of fatigue with both reliability and validity for use in clinical practice and research." The Clinical Global Impressions Scale - PMC (nih.gov) ECDEU assessment manual for psychopharmacology (1976 edition) | Open Library Development of a clinical global impression scale for fatigue - PubMed (nih.gov)	1
PRIMUS	"…assess MS symptoms, activities, and quality of life…aid the assessment of the impact of MS from the patient's perspective. The opportunity was also taken to generate scales of symptoms (impairment) and activity limitations that could be used as summary measures in clinical studies." The development of patient-reported outcome indices for multiple sclerosis (PRIMUS) - PubMed (nih.gov)	1
Relapse Assessment Questionnaire	"…evaluate relapse symptoms in patients with multiple sclerosis (MS) and their impact on daily functioning, as well as response to treatment…Part 1 consists of 7 questions that evaluate relapse symptoms, impact on activities of daily living (ADL), overall functioning, and response to treatment for previous relapses. Part 2 consists of 7 questions that evaluate treatment response in terms of symptom relief, functioning, and tolerability." Assessing Relapse in Multiple Sclerosis Questionnaire: Results of a Pilot Study - PMC (nih.gov)	1
SATMED-Q	"…measuring satisfaction with treatment with medicines. The questionnaire was designed to be used in chronic patients undergoing pharmacological treatment for any disease." Development and validation of the "Treatment Satisfaction with Medicines Questionnaire" (SATMED-Q) - PubMed (nih.gov)	1
SF-12	"…12-item short form (SF-12) health survey summary measure and 8 scale profile in comparison with SF-36 summary measures and scales." A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity - PubMed (nih.gov)	1
SF-36 and SF-36v2	"constructed to survey health status in the Medical Outcomes Study…designed for use in clinical practice and research, health policy evaluations, and general populations surveys…assesses eight health concepts 1) limitations in physical activities because of health problems; 2) limitations in social activities because of physical or emotional problems; 3) limitations in usual role activities because of physical health problems; 4) bodily pain; 5) general mental health (psychological distress and well-being); 6) limitations in usual role activities because of emotional problems; 7) vitality (energy and fatigue); and 8) general health perceptions." The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection - PubMed (nih.gov)	6
Social life and activities impact	No specific information for broad description of PRO instrument.	1
SymptoMScreen	"…enable patients to …efficiently communicate symptom severity in multiple domains… SymptoMScreen, an in-house developed tool for rapid assessment of MS symptom severity in routine practice…developed "SymptoMScreen," a battery of 7-point Likert scales for 12 distinct domains commonly affected by MS: mobility, dexterity, body pain, sensation, bladder function, fatigue, vision, dizziness, cognition, depression, and anxiety." SymptoMScreen: A Tool for Rapid Assessment of Symptom Severity in MS Across Multiple Domains - PubMed (nih.gov)	1
Telephone interviews*	No specific information for broad description of PRO instrument.	5
TFQ	"…a questionnaire to provide a structured approach to evaluate patients' experience of clinical trial participation…Assessing the clinical trial experience from the patient perspective using a robust questionnaire may offer potential to improve trial design and ensure subjects stay engaged throughout the trial process."	1

(*continued on next page*)

Table 5 (*continued*)

Patient-reported assessment method	Concept patient-reported assessment method purports to measure, according to instrument development publications	Number of protocols describing the patient-reported assessment method
	Development of a Patient-Led End of Study Questionnaire to Evaluate the Experience of Clinical Trial Participation - PubMed (nih.gov)	
TSQM II, TSQM v1.4 and TSQM-9	"…a general measure of patients' satisfaction with medication…a psychometrically sound and valid measure of the major dimensions of patients' satisfaction with medication…may also be a good predictor of patients' medication adherence across different types of medication and patient populations…provides a way of evaluating and comparing patients' satisfaction with various types and forms of medications." Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease - PubMed (nih.gov)	6
WPAI and WPAI:MS	"…WPAI measures of time missed from work, impairment of work and regular activities due to overall health and symptoms…The Work Productivity and Activity Impairment (WPAI) questionnaire elicited the number of days and hours missed from work, days and hours worked, days during which performing work was difficult and the extent to which the individual was limited at work (work impairment) during the past 7 days. The extent of work loss and impairment, attributable to both poor health and the symptom or problem specific by the respondent, was elicited." The validity and reproducibility of a work productivity and activity impairment instrument - PubMed (nih.gov) "To characterize work productivity in relapsing multiple sclerosis (MS) by using a work productivity scale and to identify associations between work productivity and disability, depression, fatigue, anxiety, cognition, and health-related quality of life." Work Productivity in Relapsing Multiple Sclerosis Associations with Disability, Depression, Fatigue, Anxiety, Cognition, and Health-Related Quality of Life (core.ac.uk)	6

Note: *FLOODLIGHT consists of multiple instruments and is not a single PRO measure. Telephone interviews are not considered as PRO measures. Both FLOODLIGHT and telephone interviews have been included in the table as patient-reported assessment methods. Quotes for the concepts measured are taken from the original instruments' development publications. The link to each instrument's development publication is beneath each quote.

Table 6
Exemplars of some of the statements provided in protocols.

Study	PRO measure	Statements provided in the protocol
ARTIOS	FLOODLIGHT FSMC HADS MSIS-29 TSQM v1.4	HADS: "Depression and anxiety are reported to have a severe negative impact on patients with MS and are associated with a reduction in health-related quality-of-life…The HADS has been found to have high sensitivity and specificity in relation to clinical interview and to other mood rating scales in people with MS." FSMC: "Fatigue in the context of multiple sclerosis (MS) is a complex symptom with still unknown pathophysiology. The Fatigue Scale for Motor and Cognitive Functions (FSMC) is a 20 item scale developed as a measure of cognitive and motor fatigue for people with MS…Sensitivity and specificity scores allow reliable assessment and the statistically identified cutoff values provide detailed quantification of fatigue in clinical routine… Improving fatigue in patients with MS is difficult and drug trials have shown mixed results. Given the high impact on employment and quality of life, the scale will be utilized."
OLIKOS	C-SSRS PGIC SF-12 TFQ TSQM-9	TFQ: "TFQ responses may be used by the sponsor to understand where improvements can be made in the clinical trial process.."
PARADIGMS	C-SSRS Peds-QL	Peds-QL: "Generic multidimensional health-related quality of life will be assessed with the Pediatric Quality of Life Inventory."
RAM-MS	EQ-5D-5L FSS MSIS-29	FSS: "The Fatigue Severity Scale (FSS) is designed to differentiate fatigue from clinical depression, since both share same symptoms."
STHENOS	FSIQ-RMS MHI-5 MSIS-29 MSTCQ Social life and activities impact TSQM v1.4 Work productivity questionnaire	MSTCQ: "The MSTCQ will be used to assess patient satisfaction with their treatment injections…"

Abbreviations: C-SSRS Columbia Suicide Severity Rating Scale, EQ-5D-5L European Quality-of-Life – 5 Dimensions 5 Level, FIS Fatigue Impact Scale, FLOODLIGHT Smartphone-based remote tracking device, FSIQ-RMS Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis, FSMC Fatigue Scale for Motor and Cognitive Functions, FSS Fatigue Severity Scale, HADS Hospital Anxiety and Depression Scale, MHI-5 Mental health inventory – 5 Item, MSIS-29 Multiple Sclerosis Impact Scale, MSTCQ Multiple sclerosis treatment concerns questionnaire, Peds-QL Paediatric Quality-of-Life Inventory, PGIC Patient Global Impression of Change, PGI-S Patient's Global Impression of Severity of Fatigue, SF-12 12-Question health questionnaire, TFQ Trial Feedback Questionnaire, TSQM v1.4 Treatment Satisfaction Questionnaire for Medication Version 1.4, TSQM-9 Treatment Satisfaction Questionnaire for Medication (9-items).

Was it clear which PRO measures were considered for inclusion?. Table 4 shows that while the primary publications reported the PRO measure used, none mentioned whether other PRO measures were considered. The EVOLVE-MS-II primary publication provided partial information, as described above.

Was there a clear explanation why specific PRO measures were chosen from available candidates?. Table 4 shows that none of the primary publications documented why individual PRO measures were chosen above others.

Table 7
PRO measures given a score of 'Partial' for the quality of rationale for their selection in the clinical trial protocols.

Study name	PRO measure	Details provided in the protocol
ASCLEPIOS I/II	C-SSRS*	"C-SSRS data mapped to Columbia Classification Algorithm for Suicide assessment (C-CASA) as per FDA guidance on suicidality."
	C-SSRS*	
EXPAND	C-SSRS*	"A validated version of the C-SSRS is used to capture self-reported C-SSRS data via an interactive voice response telephone system (eC-SSRS). The eC-SSRS uses a detailed branched logic algorithm to perform the C-SSRS patient interview, evaluating each patient's suicidality ideation and behavior in a consistent manner. The use of the eC-SSRS (or equivalent) to detect suicidal ideation or behavior is currently mandated in studies of CNS active drugs."
OPTIMUM	FSIQ-RMS	"The development of FSIQ-RMS is in accordance with the requirements set forth in the Final Guidance to the Industry on Subject Reported Outcomes: Use in Medical Product Development to Support Label Claims [FDA 2009a]."
	Patient Preference Questionnaire	To "capture patient preferences for selected treatment outcomes for use as an additional input to healthcare decisions. An increased understanding of individual values and preferences is the basis for shared decision-making, which in turn encourages patient compliance and health outcomes."
POINT	FSIQ-RMS	"The development of FSIQ-RMS is in accordance with the requirements set forth in the Final Guidance to the Industry on Subject Reported Outcomes: Use in Medical Product Development to Support Label Claims [FDA 2009a]."

Abbreviations: CNS central nervous system, C-SSRS Columbia Suicide Severity Rating Scale, FSIQ-RMS Fatigue Symptoms and Impacts Questionnaire – Relapsing Multiple Sclerosis. *The FDA mandates the measurement of suicide risk using a method that meets FDA reporting requirements. The FDA describes C-SSRS as the 'gold standard'. We recommend a PRO measure selection process.

Table 8
Assessment of PRO measure selection reported in clinical trial protocols.

Criterion		Number of PRO measures		
	Yes/No	None	Partial	Good
Was a PRO measure selection strategy documented? Yes or No	No: 21/21 (100%)	–	–	–
Was it clear which variables were being measured and why?	–	87 (100%)	0	0
Were the clinical variables intended for measurement defined clearly?	–	87 (100%)	0	0
Was it clear which PRO measures were considered for inclusion?	-	87 (100%)	0	0
Was there a clear explanation why specific PRO measures were chosen from available candidates?	-	81 (93%)	6* (7%)	0
Were the measurement trade-offs associated with the choice of PRO measure explained?	-	87 (100%)	0	0

A total of 21 clinical trial protocols were assessed, with 87 PRO measures described. Refer to the scoring system in Table 2 for definitions of 'None' (0), 'Partial' (1) and 'Good' (2). *Protocols documenting a 'Partial' (1) explanation why specific PRO measures were chosen from the available candidates (PRO measures in brackets): ASCLEPIOS I/II (C-SSRS), EXPAND (C-SSRS), OPTIMUM (FSIQ-RMS and Patient Preference Questionnaire), POINT (FSIQ-RMS).

Were the measurement trade-offs associated with the choice of PRO measures explained?. Table 4 shows that none of the primary publications documented the trade-offs associated with their choice of PRO measures. This is not a surprise as none documented clear definitions of the concepts for measurement, nor justifications why the chosen PRO measures were selected above potential alternatives.

5. Discussion

None of the pivotal phase III MS clinical trials reviewed provided explicit rationales or justifications underpinning PRO measure selection. Trials providing limited reasonings tended to base PRO measure selection on previous trials. Measured concepts were not clearly defined. Explanations why specific concepts were chosen were limited. Explicit rationales and justifications may have underpinned PRO measure selection in these studies, but the information is not documented. PRO measure selection processes may have been more rigorous than our results imply, but more implicit and tacit than explicit. We suspect this is unlikely; our recent study showed fatigue PRO measure use was not related to PRO measure development quality (Close et al., 2023).

Several reasons may explain this situation. First, there is no obligation to provide these data in clinical trial protocols or publications. Second, there is no specific guidance on PRO measure selection in the many PRO recommendations that exist (Butcher et al., 2020; Calvert et al., 2013a, 2013b, 2021; Close et al., 2023; Patrick et al., 2007; Rothman et al., 2009; US FDA, 2009, 2018, 2022a; https://www.healthmeasures.net/index.php), nor at the common data elements site sponsored by NINDS (www.commondataelements.ninds.nih.gov). Whilst these guidelines and resources have evolved over time and provide useful information and a starting point, they focus on aspects of PRO measure performance, development requirements and reporting. None provide clinical trialists with explicit guidance on PRO measure selection. Supplementary Fig. S1 shows FDA's roadmap to patient-focused outcomes measurement in clinical trials. This rarely cited diagram includes highly relevant information for PRO measure selection, but in our opinion provides neither explicit guidance nor enough detail (FDA, 2022b).

A third reason why limited PRO measure selection guidance exists may be a general under-recognition of measurement issues associated with PRO measures. Specifically, there may be an under-appreciation of the impact of different PRO measure development qualities, structures (items and number of item response categories and content), performance characteristics (range, precision, error) and context-dependent factors (score distributions, response dependence, differential item functioning). Also, for these reasons, there is a misplaced over-interpretation of the statement "reliable and valid measure of…". These under-recognitions and over-interpretations may reflect the limited availability of comprehensive head-to-head comparisons, *post hoc* examinations in clinical trial data, and appropriately critical PRO appraisals. As such, some clinicians are unfamiliar with PRO measurement issues and see measurement science research as perplexing. Consequently, valuable fundamental research is published in journals that clinicians are less likely to access (Stenner et al., 1983), and written in less clinician-accessible language (Andrich., 2011).

A fourth reason is that some researchers prefer to use the same PRO measures across studies for comparability between treatment options. This is inadvisable given that, as discussed above, measures developed for one context cannot be assumed valid and reliable for another. It seems that this is of little concern to some economic researchers [Brazier et al., 2023; Perfetto et al., 2023]. Related to this fourth reason is the mandated use of PRO measures. For example, the FDA mandates suicide risk is measured. However, it is important to note the FDA does not mandate the specific PRO measure used. It mandates using a method meeting their reporting requirements. The FDA describes the C-SSRS as the 'gold standard'. However, the suitability of the C-SSRS in any context needs to be considered along with any other PRO measure. We find it hard to think the FDA would not accept a reasoned evidence-based argument as to why another suicide risk PRO measure was chosen. Hence, we recommend a PRO measure selection process.

With this recognition, it is evident that this current situation should be rectified, as the negative ramifications of weak PRO measurement are

too important. We recommend clinical trialists employ strategic and formal approaches to PRO measure selection, to maximize the possibility that trial results approximate real clinical effects, and document their approach in trial protocols and pivotal publications, albeit as appendices and supplementary information. Indeed, we strongly recommend these be made regulatory and scientific requirements. A more critical academic appraisal of PRO studies is needed so that clinicians are more familiar with the pitfalls, and the weaknesses of the field are exposed. Without these academic and regulatory efforts, developments in the quality of the health measurement field will continue to be slow and fragmented.

5.1. PRO measure selection strategy recommendation

Logically, a PRO measure selection strategy has five stages. First, clarify and justify the specific concepts for measurement, the PROs, within the specific context of use. Second, identify the pool of PRO measures from which to choose. Third, shortlist a set of candidate PRO measures for more detailed examination, based on published information of their development. Fourth, compare the performance characteristics of the short-listed PRO measures, head-to-head, in a suitable sample. Fifth, synthesize the information and make a reasoned decision. This logical, five-stage process provides the evidence required to select the best PRO measure for the objective at hand, and, if necessary, the platform for modification or new measure development.

Stage 1, the clarification of the specific concepts of interest and context of use, heavily underpins the subsequent stages. A meaningful search for PRO measures cannot be conducted until trialists have a clear understanding of the concepts they intend to measure and the context in which measurement will be conducted. Context of use should also include explicitly stated study hypotheses. All too frequently trialists cite ambiguous umbrella terms, such as quality-of-life, health status, well-being, disability, and functioning. These terms do not enable an evaluation of the suitability of a PRO measure or its item content. We recommend clinical trialists define their concepts of interest, very explicitly, and conduct qualitative research to understand the concepts that are important to patients.

Clinical trialists are very familiar with their contexts of use, especially the expected sample characteristics, disease natural history, and hypothesized treatment effects. However, we suggest trialists are less familiar with the PRO measurement implications associated with their contexts of use. A simple exemplar is the EXPAND study (siponimod versus placebo in people with SPMS) skewed MSWS-12v2 baseline score distribution, in the context of a sample where there are more walking disabled people with progressive disease and a treatment hypothesized to have an anti-progressive effect (Hobart et al., 2022). Conceptually, participant walking ability was expected (and shown) to worsen over time. Therefore, the MSWS-12v2 score distribution skewness was expected (and shown) to worsen over time, resulting in increasing proportions of EXPAND participants located in the upper quartile (i.e., worse walking ability) of the MSWS-12v2 score range where the scale's ability to detect change is constrained by its fixed measurement range (Hobart et al., 2022). Consequently, EXPAND's MSWS-12v2-measured walking ability changes, and treatment group differences, were almost certainly underestimates of 'true' walking ability changes. Had EXPAND used a PRO measure better targeted for decline in walking ability, all other measurement issues being equal, the magnitude of the treatment effect (i.e., a slowing of walking disability progression) would have been better detected. This does not mean the MSWS-12v2 is a bad measure, just that in the EXPAND study context of use the MSWS-12v2 likely underestimated the treatment effects leading to type II measurement error (i.e., a false negative). This could have been anticipated and mitigated. This exemplar shows that scientifically solid PRO measures have context-dependent limitations with notable implications.

In Stage 2 we recommend clinical trialists search for all potential PRO measures purporting to measure their concept of interest. For each PRO measure, trialists should evaluate its development and item content – two related but different evaluations. PRO development differs in method and quality. It can be evaluated against guidance (Close et al., 2023; FDA et al., 2009). In essence, key issues are the definition of the concepts measured, the strength of the conceptual underpinnings, the method of item generation, the nature and extent of patient involvement, and how the final item set was achieved. Patient involvement in item development through qualitative methods is crucial for content validity. It is difficult to see how the development of a PRO measure without patient involvement could be considered content valid, unless strongly supported by *post hoc* qualitative research in patients. Unfortunately, for many PRO measures this information is not well documented.

Guidelines for PRO measure development provide useful frameworks for evaluation. However, in our opinion, they miss a key step – the articulation of a measurement concept as a set of items. We have coined the term "item set analysis", discussed in detail elsewhere (Close et al., 2023). In brief, there should be an explicit link between the overall concept, domains, subdomains, items and scores generated. This requires clear definitions of concepts, domains and subdomains, and clarity of how and why the items that are combined to form scores adequately represent those subdomains. When this is the case, the link from score through item to subdomains and concept is explicit. This link is rarely clear. Even when there is a conceptual framework, there seems to be a disconnect between the components of the framework and the item sets that generate scores. We recommend careful and greater consideration is given to the item content of subdomains.

Stage 3 is short-listing. When PRO measures have been extracted, their development papers reviewed and critiqued, their development process evaluated and their item content considered, a short-list of suitable candidates can be identified. One way of formalizing short-listing is to use Consensus Standards for Measurement Instruments (COSMIN) scoring, and select the highest scoring candidates. Although useful, we identified limitations in COSMIN's rating process (Close et al., 2023). Specifically, what constitutes adequacy for PRO definitions, conceptualisations and qualitative work. Perhaps the most important limitation is the absence of the item set analysis; the degree to which a set of items generating a score maps a variable, discussed above.

Stage 4 is a head-to-head comparison of short-listed PRO measures in a sample representative of the context of use. This enables a comparison of measurement properties. We recommend using modern psychometric methods, Rasch measurement theory (RMT) or item response theory (IRT), rather than traditional methods based on classical test theory (CTT). RMT is our preference. This provides an hypothesis test against which to examine observed PRO measure data, enables the diagnosis of measurement weaknesses, and provides a strong platform for measurement improvement.

Stage 5 synthesises all the information from Stages 1 to 4 to reach a rational, evidence-based, decision. We anticipate, in many circumstances, that this will not be straightforward and there will be trade-offs. However, the process will lead to consideration of the problem and opportunities for better measurement.

Our recommendations may seem labor intensive. However, we think this would be a misinterpretation. Several areas of coordinated research could provide the MS community with a body of work to underpin PRO measure selection in clinical trials. For example, publicly available repositories of PRO measures purporting to measure variables, with cataloging of concept definitions, conceptualisations, and items. Targeted, empiric, head-to-head comparisons in samples pertinent to MS trials (e.g., RMS, primary progressive MS [PPMS], SPMS, advanced MS) would provide an evidence base of relative performance and trade-offs.

Currently, PRO measurement issues appear to be secondary considerations. This parallels the history of medical statistics which used to be an afterthought but is now integral to the initial stages of clinical trial design. The same is required of a measurement strategy. It is also important not to conflate measurement methods, which concern the

generation of measurements, with statistical analysis which involves the analysis of measurements. As such, measurement methodological issues are prior to statistics. Without high quality PRO measurement, type II errors will pervade our trials. The implications for MS care, MS science development and individual patients are far too great for these identified issues not to be considered seriously.

6. Conclusions

PRO measure selection in multi-million-dollar pivotal MS clinical trials that dictate patient care, drug licensing and label claims currently lacks evidence. We believe this is also a common problem in clinical trials in other therapy areas. Widespread type II error from clinical trials can be avoided in future by adopting a robust PRO measure selection strategy. Widespread recognition of this issue and subsequent evaluation, critique, and documentation are required to optimize PRO measurement and their utility in pivotal clinical trials.

Role of funding source

Medical writing support for the review was funded by Novartis Pharma AG.

Declaration of Competing Interest

Jeremy Hobart has received consulting fees, honoraria, support to attend meetings or research support from Acorda, Asubio, Bayer Schering, Biogen Idec, F. Hoffmann-La Roche, Genzyme, Merck Serono, Novartis, Oxford PharmaGenesis and Teva. Disclosures do not show a conflict with the work being presented. Tanuja Chitnis has received compensation for consulting from Biogen, Novartis Pharmaceuticals, Roche Genentech, and Sanofi Genzyme. She has received research support from Brainstorm Cell Therapeutics, EMD Serono, I-Mab Biopharma, Mallinckrodt ARD, the National Institutes of Health, National MS Society, Novartis Pharmaceuticals, Octave Bioscience, Roche Genentech, Sumaira Foundation, Tiziana Life Sciences, and US Department of Defense. Disclosures do not conflict with the work being presented. Jiwon Oh has received research support from Biogen-Idec, Eli-Lilly, EMD-Serono, and Roche; and fees for consulting or speaking from Biogen-Idec, Bristol Myers Squibb, EMD-Serono, Novartis, Roche, and Sanofi-Genzyme. Laurie Burke has past and ongoing research support and contracts from various non-profit organisations and for-profit companies that do not conflict with this work. Andrew Lloyd works for and holds stock in Acaster Lloyd Consulting Ltd which has received fees from Novartis. Disclosures do not show a conflict with the work being presented. Pamela Vo is an employee of Novartis Pharma AG. Miriam King and Jo Vandercappellen were employees of Novartis Pharma AG during the analysis of this study and manuscript development.

Acknowledgments

The authors would like to thank David McMinn, PhD (Novartis CONEXTS, Ireland) and Barbara Chan, PhD (Novartis CONEXTS, Ireland) for providing medical writing support, which was funded by Novartis Pharma AG.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2023.104788.

References

Andrich, D., 2011. Rating scales and Rasch measurement. Expert. Rev. Pharmacoecon. Outcomes Res. 11 (5), 571–585.

Andrich, D., Humphrey, S., Marais, I., 2012. Quantifying local, response dependence between two polytomous items using the Rasch model. Appl. Psychol. Meas. 36 (4), 309–324.

Benedict, R.H., Cohan, S., Lynch, S.G., Riester, K., Wang, P., Castro-Borrero, W., Elkins, J., Sabatella, G., 2018. Improved cognitive outcomes in patients with relapsing-remitting multiple sclerosis treated with daclizumab beta: results from the DECIDE study. Mult. Scler. 24 (6), 795–804.

Brazier, J., Peasgood, T., Mukuria, C., Luo, N., Mulhern, B., Pickard, A.S., Augustovski, F., Greiner, W., Engel, L., 2023. Author reply. Value Health 26 (3), 437–440.

Butcher, N.J., Mew, E.J., Monsour, A., Chan, A.W., Moher, D., Offringa, M., 2020. Outcome reporting recommendations for clinical trial protocols and reports: a scoping review. Trials 21 (1), 620.

Calabresi, P.A., Radue, E.W., Goodin, D., Jeffery, D., Rammohan, K.W., Reder, A.T., Vollmer, T., Agius, M.A., Kappos, L., Stites, T., Li, B., Cappiello, L., von Rosenstiel, P., Lublin, F.D., 2014. Safety and efficacy of fingolimod in patients with relapsing-remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Neurol. 13 (6), 545–556.

Calvert, M., Blazeby, J., Altman, D.G., Revicki, D.A., Moher, D., Brundage, M.D., Group, C.P., 2013a. Reporting of patient-reported outcomes in randomized trials: the consort PRO extension. JAMA 309 (8), 814–822.

Calvert, M., Brundage, M., Jacobsen, P.B., Schunemann, H.J., Efficace, F., 2013b. The consort patient-reported outcome (PRO) extension: implications for clinical trials and practice. Health Qual. Life Outcomes 11, 184.

Calvert, M., King, M., Mercieca-Bebber, R., Aiyegbusi, O., Kyte, D., Slade, A., Chan, A.W., Basch, E., Bell, J., Bennett, A., Bhatnagar, V., Blazeby, J., Bottomley, A., Brown, J., Brundage, M., Campbell, L., Cappelleri, J.C., Draper, H., Dueck, A.C., Ells, C., Frank, L., Golub, R.M., Griebsch, I., Haywood, K., Hunn, A., King-Kallimanis, B., Martin, L., Mitchell, S., Morel, T., Nelson, L., Norquist, J., O'Connor, D., Palmer, M., Patrick, D., Price, G., Regnault, A., Retzer, A., Revicki, D., Scott, J., Stephens, R., Turner, G., Valakas, A., Velikova, G., von Hildebrand, M., Walker, A., Wenzel, L., 2021. SPIRIT-PRO extension explanation and elaboration: guidelines for inclusion of patient-reported outcomes in protocols of clinical trials. BMJ Open 11 (6), e045105.

Close, J., Vandercappellen, J., King, M., Hobart, J., 2023. Measuring fatigue in multiple sclerosis: there may be trouble ahead. Neurol. Ther. Manuscr. Accept.

Comi, G., Kappos, L., Selmaj, K.W., Bar-Or, A., Arnold, D.L., Steinman, L., Hartung, H.P., Montalban, X., Kubala Havrdova, E., Cree, B.A.C., Sheffield, J.K., Minton, N., Raghupathi, K., Ding, N., Cohen, J.A., SUNBEAM Study Investigators, 2019. Safety and efficacy of ozanimod versus interferon beta-1a in relapsing multiple sclerosis (SUNBEAM): a multicentre, randomised, minimum 12-month, phase 3 trial. Lancet Neurol. 18 (11), 1009–1020.

Confavreux, C., O'Connor, P., Comi, G., Freedman, M.S., Miller, A.E., Olsson, T.P., Wolinsky, J.S., Bagulho, T., Delhay, J.L., Dukovic, D., Truffinet, P., Kappos, L., TOWER Trial Group., 2014. Oral teriflunomide for patients with relapsing multiple sclerosis (TOWER): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Neurol. 13 (3), 247–256.

Cree, B.A., Selmaj, K.W., Steinman, L., Comi, G., Bar-Or, A., Arnold, D.L., Hartung, H.P., Montalban, X., Havrdova, E.K., Sheffield, J.K., Minton, N., Cheng, C.Y., Silva, D., Kappos, L., Cohen, J.A., 2022. Long-term safety and efficacy of ozanimod in relapsing multiple sclerosis: up to 5 years of follow-up in the DAYBREAK open-label extension trial. Mult. Scler. 28 (12), 1944–1962.

Dib, H., Tamam, Y., Terzi, M., Hobart, J., 2017. Testing patient-reported outcome measurement equivalence in multinational clinical trials: an exemplar using the 12-item Multiple Sclerosis Walking Scale. Mult. Scler. J. Exp. Transl. Clin. 3 (3) https://doi.org/10.1177/2055217317728740.

Fisk, J.D., Brown, M.G., Sketris, I.S., Metz, L.M., Murray, T.J., Stadnyk, K.J., 2005. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. J. Neurol. Neurosurg. Psychiatry 76 (1), 58–63.

Freeman, J.A., Thompson, A.J., Fitzpatrick, R., Hutchinson, M., Miltenburger, C., Beckmann, K., Dahlke, F., Kappos, L., Polman, C., Pozzilli, C., European Study Group on Interferon-beta1b in Secondary Progressive, M.S., 2001. Interferon-beta1b in the treatment of secondary progressive MS: impact on quality of life. Neurology 57 (10), 1870–1875.

Goodman, A.D., Brown, T.R., Edwards, K.R., Krupp, L.B., Schapiro, R.T., Cohen, R., Marinucci, L.N., Blight, A.R., MSF204 Investigators, 2010. A phase 3 trial of extended release oral dalfampridine in multiple sclerosis. Ann. Neurol. 68 (4), 494–502.

Goodman, A.D., Brown, T.R., Krupp, L.B., Schapiro, R.T., Schwid, S.R., Cohen, R., Marinucci, L.N., Blight, A.R., Fampridine MS-F203 Investigators, 2009. Sustained-release oral fampridine in multiple sclerosis: a randomised, double-blind, controlled trial. Lancet 373 (9665), 732–738.

Hobart, J., Vo, P., Ryan, S., Arnould, S., Burke, L., 2022. Patient-reported outcomes in multiple sclerosis clinical trials: measurement lessons from the EXPAND study. Mult. Scler. 28 (3), 130–691.

Hobart, J., Ziemssen, T., Feys, P., Linnebank, M., Goodman, A.D., Farrell, R., Hupperts, R., Blight, A.R., Englishby, V., McNeill, M., Chang, I., Lima, G., Elkins, J., ENHANCE Study Investigators, 2019. Assessment of clinically meaningful improvements in self-reported walking ability in participants with multiple sclerosis: results from the randomized, double-blind, phase III enhance trial of prolonged-release fampridine. CNS Drugs 33 (1), 61–79.

Hudgens, S., Schuler, R., Stokes, J., Eremenco, S., Hunsche, E., Leist, T.P., 2019. Development and validation of the FSIQ-RMS: a new patient-reported questionnaire to assess symptoms and impacts of fatigue in relapsing multiple sclerosis. Value Health 22 (4), 453–466.

Hupperts, R., Gasperini, C., Lycke, J., Ziemssen, T., Feys, P., Xiao, S., Acosta, C., Koster, T., Hobart, J., 2022. Efficacy of prolonged-release fampridine versus placebo

on walking ability, dynamic and static balance, physical impact of multiple sclerosis, and quality of life: an integrated analysis of mobile and enhance. Ther. Adv. Neurol. Disord. 15 https://doi.org/10.1177/17562864221090398.

Kapoor, R., Ho, P.R., Campbell, N., Chang, I., Deykin, A., Forrestal, F., Lucas, N., Yu, B., Arnold, D.L., Freedman, M.S., Goldman, M.D., Hartung, H.P., Havrdova, E.K., Jeffery, D., Miller, A., Sellebjerg, F., Cadavid, D., Mikol, D., Steiner, D., ASCEND Investigators, 2018. Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): a phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension. Lancet Neurol. 17 (5), 405–415.

Kappos, L., Bar-Or, A., Cree, B.A.C., Fox, R.J., Giovannoni, G., Gold, R., Vermersch, P., Arnold, D.L., Arnould, S., Scherz, T., Wolf, C., Wallstrom, E., Dahlke, F., Investigators, E.C., 2018. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. Lancet 391 (10127), 1263–1273.

Lublin, F., Miller, D.H., Freedman, M.S., Cree, B.A.C., Wolinsky, J.S., Weiner, H., Lubetzki, C., Hartung, H.P., Montalban, X., Uitdehaag, B.M.J., Merschhemke, M., Li, B., Putzki, N., Liu, F.C., Haring, D.A., Kappos, L., INFORMS Study Investigators, 2016. Oral fingolimod in primary progressive multiple sclerosis (INFORMS): a phase 3, randomised, double-blind, placebo-controlled trial. Lancet 387 (10023), 1075–1084.

Miller, A.E., Wolinsky, J.S., Kappos, L., Comi, G., Freedman, M.S., Olsson, T.P., Bauer, D., Benamor, M., Truffinet, P., O'Connor, P.W., TOPIC Study Group., 2014. Oral teriflunomide for patients with a first clinical episode suggestive of multiple sclerosis (TOPIC): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Neurol. 13 (10), 977–986.

Montalban, X., Hauser, S.L., Kappos, L., Arnold, D.L., Bar-Or, A., Comi, G., de Seze, J., Giovannoni, G., Hartung, H.P., Hemmer, B., Lublin, F., Rammohan, K.W., Selmaj, K., Traboulsee, A., Sauter, A., Masterman, D., Fontoura, P., Belachew, S., Garren, H., Mairon, N., Chin, P., Wolinsky, J.S., for the ORATORIO Clinical Investigators, 2017. Ocrelizumab versus placebo in primary progressive multiple sclerosis. N. Engl. J. Med. 376 (3), 209–220.

Naismith, R.T., Hendin, B., Wray, S., Huang, D., Gaudenzi, F., Dong, Q., Sperling, B., Mann, M., Werneburg, B., 2019. Patients transitioning from non-pegylated to pegylated interferon beta-1a have a low risk of new flu-like symptoms: ALLOW phase 3b trial results. Mult. Scler. J. Exp. Transl. Clin. 5 (1) https://doi.org/10.1177/2055217318822148.

Naismith, R.T., Wundes, A., Ziemssen, T., Jasinska, E., Freedman, M.S., Lembo, A.J., Selmaj, K., Bidollari, I., Chen, H., Hanna, J., Leigh-Pemberton, R., Lopez-Bresnahan, M., Lyons, J., Miller, C., Rezendes, D., Wolinsky, J.S., on behalf of The EVOLVE-MS-2 Study Group, 2020. Diroximel fumarate demonstrates an improved gastrointestinal tolerability profile compared with dimethyl fumarate in patients with relapsing-remitting multiple sclerosis: results from the randomized, double-blind, phase III evolve-ms-2 study. CNS Drugs 34 (2), 185–196.

O'Connor, P., Wolinsky, J.S., Confavreux, C., Comi, G., Kappos, L., Olsson, T.P., Benzerdjeb, H., Truffinet, P., Wang, L., Miller, A., Freedman, M.S., TEMSO Trial Group., 2011. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. N. Engl. J. Med. 365 (14), 1293–1303.

Patrick, D.L., Burke, L.B., Powers, J.H., Scott, J.A., Rock, E.P., Dawisha, S., O'Neill, R., Kennedy, D.L., 2007. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 10 (Suppl 2), S125–S137.

Perfetto, E.M., Burke, L., Love, T.R., Schrandt, M.S., Hobart, J., 2023. Measuring health and well-being: we need to get it right for patients, with patients. Value Health 26 (3), 435–437.

Rothman, M., Burke, L., Erickson, P., Leidy, N.K., Patrick, D.L., Petrie, C.D., 2009. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification pro task force report. Value Health 12 (8), 1075–1083.

Stenner, A.J., Smith, M., Burdick, D.S., 1983. Toward a theory of construct definition. J. Educ. Meas. 20 (4), 305–316.

US FDA, CDER., CBER., CDRH., 2009. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. December 2009.

US FDA, 2018. Methods to identify what is important to patients and select, develop or modify fit-for-purpose clinical outcomes assessments. Patient-focused drug development guidance public workshop: october 15-16, 2018.

US FDA, 2022. Principles for selecting, developing, modifying, and adapting patient-reported outcome instruments for use in medical device evaluation. Guidance for industry and Food and Drug Administration staff, and other stakeholders. Document issued on January 26, 2022.

US FDA, 2022. Roadmap to patient-focused outcome measurement in clinical trials. https://www.fda.gov/media/87004/download. (Accessed 18 November 2022).

Vermersch, P., Brieva-Ruiz, L., Fox, R.J., Paul, F., Ramio-Torrenta, L., Schwab, M., Moussy, A., Mansfield, C., Hermine, O., Maciejowski, M., AB07002 Study Group, 2022. Efficacy and safety of masitinib in progressive forms of multiple sclerosis: a randomized, phase 3, clinical trial. Neurol. Neuroimmunol. Neuroinflamm. 9 (3), e1148.

Vermersch, P., Czlonkowska, A., Grimaldi, L.M., Confavreux, C., Comi, G., Kappos, L., Olsson, T.P., Benamor, M., Bauer, D., Truffinet, P., Church, M., Miller, A.E., Wolinsky, J.S., Freedman, M.S., O'Connor, P., TENERE Trial Group., 2014. Teriflunomide versus subcutaneous interferon beta-1a in patients with relapsing multiple sclerosis: a randomised, controlled phase 3 trial. Mult. Scler. 20 (6), 705–716.

Wray, S., Then Bergh, F., Wundes, A., Arnold, D.L., Drulovic, J., Jasinska, E., Bowen, J. D., Negroski, D., Naismith, R.T., Hunter, S.F., Gudesblatt, M., Chen, H., Lyons, J., Shankar, S.L., Kapadia, S., Mendoza, J.P., Singer, B.A., 2022. Efficacy and safety outcomes with diroximel fumarate after switching from prior therapies or continuing on DRF: results from the phase 3 EVOLVE-MS-1 Study. Adv. Ther. 39 (4), 1810–1831.

NCT00835770: BG00012 monotherapy safety and efficacy extension study in multiple sclerosis (MS) (ENDORSE). BG00012 Monotherapy Safety and Efficacy Extension Study in Multiple Sclerosis (MS) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01194570: A study of ocrelizumab in participants with primary progressive multiple sclerosis. A Study of Ocrelizumab in Participants With Primary Progressive Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01247324: A study of ocrelizumab in comparison with interferon beta-1a (Rebif) in participants with relapsing multiple sclerosis. A Study of Ocrelizumab in Comparison With Interferon Beta-1a (Rebif) in Participants With Relapsing Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01412333: A study of ocrelizumab in comparison with interferon beta-1a (Rebif) in participants with relapsing multiple sclerosis. A Study of Ocrelizumab in Comparison With Interferon Beta-1a (Rebif) in Participants With Relapsing Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01633112: MS study evaluating safety and efficacy of two doses of fingolimod versus copaxone (ASSESS). MS Study Evaluating Safety and Efficacy of Two Doses of Fingolimod Versus Copaxone - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01665144: Exploring the efficacy and safety of siponimod in patients with secondary progressive multiple sclerosis (EXPAND). Exploring the Efficacy and Safety of Siponimod in Patients With Secondary Progressive Multiple Sclerosis (EXPAND) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT01892722: Safety and efficacy of fingolimod in pediatric patients with multiple sclerosis. Safety and efficacy of fingolimod in pediatric patients with multiple sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02047734: Efficacy and safety study of ozanimod in relapsing multiple sclerosis (radiance). Efficacy and Safety Study of Ozanimod in Relapsing Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02425644: Oral ponesimod versus teriflunomide in relapsing multiple sclerosis (optimum). Oral Ponesimod Versus Teriflunomide In Relapsing MUltiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02637856: A study of ocrelizumab in participants with relapsing remitting multiple sclerosis (RRMS) who have had a suboptimal response to an adequate course of disease-modifying treatment (DMT). A Study of Ocrelizumab in Participants With Relapsing Remitting Multiple Sclerosis (RRMS) Who Have Had a Suboptimal Response to an Adequate Course of Disease-Modifying Treatment (DMT) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02792218: Efficacy and safety of ofatumumab compared to teriflunomide in patients with relapsing multiple sclerosis (ASCLEPIOS I). Efficacy and Safety of Ofatumumab Compared to Teriflunomide in Patients With Relapsing Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02792231: Efficacy and safety of ofatumumab compared to teriflunomide in patients with relapsing multiple sclerosis. (ASCLEPIOS II). Efficacy and Safety of Ofatumumab Compared to Teriflunomide in Patients With Relapsing Multiple Sclerosis. - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02861014: A study of ocrelizumab in participants with relapsing remitting multiple sclerosis (RRMS) who have had a suboptimal response to an adequate course of disease-modifying treatment (DMT). A Study of Ocrelizumab in Participants With Relapsing Remitting Multiple Sclerosis (RRMS) Who Have Had a Suboptimal Response to an Adequate Course of Disease-Modifying Treatment (DMT) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02907177: Clinical study to compare the efficacy and safety of ponesimod to placebo in subjects with active relapsing multiple sclerosis who are treated with dimethyl fumarate (Tecfidera®) (POINT). Clinical Study to Compare the Efficacy and Safety of Ponesimod to Placebo in Subjects With Active Relapsing Multiple Sclerosis Who Are Treated With Dimethyl Fumarate (Tecfidera®) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT02936037: Effect of MD1003 in progressive multiple sclerosis (SPI2) (SPI2). Effect of MD1003 in Progressive Multiple Sclerosis (SPI2) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT03277248: Study to assess the efficacy and safety of ublituximab in participants with relapsing forms of multiple sclerosis (RMS) (ULTIMATE II). Study to Assess the Efficacy and Safety of Ublituximab in Participants With Relapsing Forms of Multiple Sclerosis (RMS) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT03277261: Study to assess the efficacy and safety of ublituximab in participants with relapsing forms of multiple sclerosis (RMS) (ULTIMATE 1). Study to Assess the Efficacy and Safety of Ublituximab in Participants With Relapsing Forms of Multiple Sclerosis (RMS) (ULTIMATE 1) - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT03477500: RCT comparing autologous hematopoietic stem cell transplantation versus alemtuzumab in MS (RAM-MS). RCT Comparing Autologous Hematopoietic Stem Cell Transplantation Versus Alemtuzumab in MS - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT04353492: An open-label study evaluating ofatumumab treatment effectiveness and PROs in subjects with RMS transitioning from fumarate-based RMS approved therapies or fingolimod to ofatumumab (ARTIOS). An Open-label Study Evaluating Ofatumumab Treatment Effectiveness and PROs in Subjects With RMS Transitioning

From Fumarate-based RMS Approved Therapies or Fingolimod to Ofatumumab - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT04486716: A single arm study evaluating the efficacy, safety and tolerability of ofatumumab in patients with relapsing multiple sclerosis (OLIKOS). A Single Arm Study Evaluating the Efficacy, Safety and Tolerability of Ofatumumab in Patients With Relapsing Multiple Sclerosis - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).

NCT04788615: Open label randomized multicenter to assess efficacy & tolerability of ofatumumab 20mg vs. first line DMT in RMS (STHENOS). Open Label Randomized Multicenter to Assess Efficacy & Tolerability of Ofatumumab 20mg vs. First Line DMT in RMS - Full Text View - ClinicalTrials.gov (accessed 07 June 2022).