Original Research Paper

The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability

Nicholas G LaRocca, Lynn D Hudson, Richard Rudick, Dagmar Amtmann, Laura Balcer, Ralph Benedict, Robert Bermel, Ih Chang, Nancy D Chiaravalloti, Peter Chin, Jeffrey A Cohen, Gary R Cutter, Mat D Davis, John DeLuca, Peter Feys, Gordon Francis, Myla D Goldman, Emily Hartley, Raj Kapoor, Fred Lublin, Gary Lundstrom, Paul M Matthews, Nancy Mayo, Richard Meibach, Deborah M Miller, Robert W Motl, Ellen M Mowry, Rob Naismith, Jon Neville, Jennifer Panagoulias, Michael Panzara, Glenn Phillips, Ann Robbins, Matthew F Sidovar, Kathryn E Smith, Bjorn Sperling, Bernard MJ Uitdehaag and Jerry Weaver; for the Multiple Sclerosis Outcome Assessments Consortium (MSOAC)

Abstract

Background: The Multiple Sclerosis Outcome Assessments Consortium (MSOAC) was formed by the National MS Society to develop improved measures of multiple sclerosis (MS)-related disability.

Objectives: (1) To assess the current literature and available data on functional performance outcome measures (PerfOs) and (2) to determine suitability of using PerfOs to quantify MS disability in MS clinical trials.

Methods: (1) Identify disability dimensions common in MS; (2) conduct a comprehensive literature review of measures for those dimensions; (3) develop an MS Clinical Data Interchange Standards Consortium (CDISC) data standard; (4) create a database of standardized, pooled clinical trial data; (5) analyze the pooled data to assess psychometric properties of candidate measures; and (6) work with regulatory agencies to use the measures as primary or secondary outcomes in MS clinical trials.

Conclusion: Considerable data exist supporting measures of the functional domains ambulation, manual dexterity, vision, and cognition. A CDISC standard for MS (http://www.cdisc.org/therapeutic#MS) was published, allowing pooling of clinical trial data. MSOAC member organizations contributed clinical data from 16 trials, including 14,370 subjects. Data from placebo-arm subjects are available to qualified researchers. This integrated, standardized dataset is being analyzed to support qualification of disability endpoints by regulatory agencies.

Correspondence to:
N LaRocca
290 Sabino Rd., West Bath, ME 04530, USA.
Nicholas.LaRocca@nmss.org

Nicholas G LaRocca
Kathryn E Smith
National Multiple Sclerosis Society, New York, NY, USA

Lynn D Hudson
Emily Hartley
Gary Lundstrom
Jon Neville
Ann Robbins
Critical Path Institute, Tucson, AZ, USA

Richard Rudick
Ih Chang
Glenn Phillips
Bjorn Sperling
Biogen, Cambridge, MA, USA

Dagmar Amtmann
University of Washington, Seattle, WA, USA

Laura Balcer
NYU Langone Medical Center, New York, NY, USA

Ralph Benedict
University at Buffalo, Buffalo, NY, USA

Robert Bermel
Jeffrey A Cohen
Deborah M Miller
Cleveland Clinic, Cleveland, OH, USA

Nancy D Chiaravalloti
John DeLuca
Kessler Foundation, West Orange, NJ, USA

Peter Chin
Genentech, San Francisco, CA, USA

Introduction

The need for better measures of MS disability has been recognized for decades. In 1993, the National Multiple Sclerosis Society (NMSS) convened an international workshop on the topic.[1] One result was a task force, charged with recommending outcome assessment methods that might improve on the Kurtzke[2] Expanded Disability Status Scale (EDSS). The task force

recommended quantitative neurological performance testing as opposed to clinical rating scales such as EDSS, largely because performance outcome measures (PerfOs) have superior psychometric properties. The task force recommended the Timed 25-Foot Walk (T25FW) as a measure of walking speed, the 9-Hole Peg Test (9HPT) as an upper extremity dexterity measure, and the Paced Auditory Serial Addition Test

Gary R Cutter
UAB School of Public
Health, Birmingham, AL,
USA

Mat D Davis
Teva Pharmaceuticals, Green
Valley, PA, USA

Peter Feys
Faculty of Medicine and Life
Sciences, Hasselt University,
Hasselt, Belgium

**Gordon Francis
Richard Meibach Jerry
Weaver**
Novartis Pharmaceuticals,
East Hanover, NJ, USA

Myla D Goldman
University of Virginia,
Charlottesville, VA, USA

Raj Kapoor
University College London,
London, UK

Fred Lublin
Icahn School of Medicine
at Mount Sinai, New York,
NY, USA

Paul M Matthews
Imperial College London,
London, UK

Nancy Mayo
McGill University, Montreal,
QC, Canada

Robert W Motl
University of Alabama at
Birmingham, Birmingham,
AL, USA

Ellen M Mowry
Johns Hopkins University,
Baltimore, MD, USA

Rob Naismith
Washington University in St.
Louis, St. Louis, MO, USA

**Jennifer Panagoulias
Michael Panzara**
Sanofi Genzyme, Cambridge,
MA, USA

Matthew F Sidovar
Acorda Therapeutics, Inc.,
Ardsley, NY, USA; KES
Business Consulting LLC,
Lyme, CT, USA

Bernard MJ Uitdehaag
VU University Medical
Center, Amsterdam, The
Netherlands

(PASAT; 3-second version) as a measure of cognitive processing speed.[3] The task force also urged the academic community to develop a test for visual function, because high contrast letter acuity was not sensitive to change. The task force recommended inclusion of three PerfOs, together called the Multiple Sclerosis Functional Composite (MSFC) for inclusion in future trials. What followed was inclusion of MSFC in most prospectively designed clinical trials conducted by industry and academia and development of Low Contrast Letter Acuity (LCLA) as a more sensitive measure of MS-related visual impairment.[4] Many placebo-controlled clinical trials demonstrated treatment effects on the MSFC score. However, complexities related to the reference population used to create standardized scores and difficulty assigning clinical meaningfulness to z-score changes limited use of the MSFC as a primary outcome measure for registration trials.[5,6]

In view of perceived limitations of the MSFC approach, and with recognition of the continuing need for better clinical measures of MS-related disability, the Multiple Sclerosis Outcome Assessments Consortium (MSOAC) was established in 2012 to accelerate the development of therapies for MS.[7] MSOAC established the concept of interest (COI) for meaningful treatment benefit as "MS disability," or simply "disability," characterized as neurological or neuropsychological impairments that result in limitations in activities and restrictions in participation or life roles, caused by MS, that are understood to be important by the person with MS. Frequent interactions with the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) served to shape the consortium's research plan and guide efforts to select PerfOs (https://www.fda.gov/downloads/drugs/guidances/ucm230597.pdf) and to determine suitability of using specific PerfOs to quantify MS disability in MS clinical trials. The context of use (COU) for the selected PerfO was use as primary or secondary endpoints in clinical trials of treatments intended to slow or stop the worsening of disability in MS.

MSOAC first defined a conceptual framework for disability measures in MS, drawing on the International Classification of Functioning, Disability and Health (ICF) core sets for MS.[8] Early on, MSOAC members highlighted the need for a visual measure to include as part of a multi-dimensional outcome measure and expressed a preference for the Symbol Digit Modalities Test (SDMT) over PASAT as a measure of processing speed because of accumulating experience with both tests. Also, MSOAC members agreed to focus on dimensions of MS that lent themselves to simple, objective, and reliable measurement and not focus on crucial dimensions of MS (e.g. pain, fatigue) that were inherently patient self-reported. A systematic literature review was conducted to assess published evidence on measures for walking speed, manual dexterity, vision, and information processing speed. This literature provided support for the ability of performance measures for these domains to capture how people with MS feel and function.[9–12]

Key to the MSOAC goal is analysis of the prospectively acquired data from multiple clinical trials. This paper details the methods used to establish the MSOAC database and the Statistical Analysis Plan (SAP) that is presently being applied to assess the clinical meaningfulness of different performance measures. Future papers will report on the results of these analyses.

Methods and initial results

Establishing a consortium

MSOAC is organized and managed by the Critical Path Institute (C-Path; https://c-path.org/programs/msoac/). With input from NMSS, C-Path established the membership agreements and engaged a wide spectrum of stakeholders, including persons with MS, advocacy organizations, clinical researchers, industry sponsors, regulators, and other governmental agencies, all working together with standard development organizations, contract research organizations, and data managers (Supplementary Table 1). C-Path supplied expertise for development of therapeutic area data standards and the remapping of legacy data to the Clinical Data Interchange Standards Consortium (CDISC) data standard accepted by the FDA and used by C-Path for analytic purposes. C-Path staff also provided regulatory expertise to guide each step through the FDA's[13] Drug Development Tool and EMA's Novel Methodologies qualification processes for PerfO qualification (Table 1).

In addition to contributing data, many MSOAC members participated in a Coordinating Committee, which served as the governing body. Working groups were established to focus on (1) Defining Disability, (2) Data Standards and Integration, (3) Clinical Outcome Assessments, (4) Regulatory, (5) Literature Review, (6) Statistics, and (7) Voice of the Patient (VOP).

Selecting domains of function from the ICF core sets for MS

In a series of in-person meetings and teleconferences, the Defining Disability Workgroup examined ICF domains for MS.[8] Inclusion and exclusion criteria (Table 2) were developed and applied to the ICF domains and to the associated measures of those domains. The Workgroup used the specified COU to provide a contextual anchor for the selection process.

Table 1. Glossary.

Acronym	Term
9HPT	9-Hole Peg Test—a brief, standardized, quantitative test of upper extremity function
ADaM	Analysis Data Tabulation Model—the CDISC model that defines standards for analysis datasets
BDI	Beck Depression Inventory—a 21-question multiple-choice self-report inventory to measure the severity of depression
CDASH	Clinical Data Acquisition Standards Harmonization—describes the recommended data collection fields for 16 domains, including demographics, adverse events, and other domains common to most therapeutic areas and clinical research phases (https://www.cdisc.org/standards/foundational/cdash)
CDE	Common Data Element—data element that is common to multiple datasets across different studies: https://www.commondataelements.ninds.nih.gov/MS.aspx#tab=Data_StandardsNLM/NIH
CDISC	Clinical Data Interchange Standards Consortium is a non-profit Standards Development Organization (SDO): https://www.cdisc.org
CFAST	Coalition For Accelerating Standards and Therapies—formed by CDISC and C-Path to focus on therapeutic area data standards and analysis standards: https://www.cdisc.org/partnerships/cfast
ClinRO	Clinician-reported outcome measure—a measurement based on a report that comes from a trained healthcare professional after observation of a patient's health condition. Most ClinRO measures involve a clinical judgment or interpretation of the observable signs, behaviors, or other manifestations related to a disease or condition.[14]
COA	Clinical Outcome Assessment—assessment of a clinical outcome can be made through report by a clinician, a patient, a non-clinician observer or through a performance-based assessment. There are four types of COAs: clinician-reported outcome, observer-reported outcome, patient-reported outcome, and performance outcome.[14]
COI	Concept(s) of Interest (COI) for meaningful treatment benefit—a description of the meaningful aspect of patient experience that will represent the intended benefit of treatment (e.g. presence/severity of symptoms, limitations in performance of daily activities).[14]
COU	Context of Use—A statement that fully and clearly describes the way the medical product development tool is to be used and the medical product development-related purpose of the use.[14]
EDSS	Expanded Disability Status Scale—a clinician-reported outcome measure of disability in MS
FSS	Functional Systems Scores—a clinician-reported measure of pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, and cerebral (or mental) activity
LCLA	Low Contrast Letter Acuity
MIC	Minimal Important Change—smallest change in score in the domain of interest which patients perceive as important.[15]
MID	Minimal Important Difference—the difference observed between groups that are known to differ on the construct of interest in an important way.[15]
MS	Multiple Sclerosis
MSFC	Multiple Sclerosis Functional Composite—a three-part, standardized, quantitative assessment instrument for assessing mobility, dexterity and cognition in clinical studies of MS
PASAT	Paced Auditory Serial Addition Test—a test used to assess capacity and rate of information processing and sustained and divided attention
PPMS	Primary Progressive Multiple Sclerosis
PerfO	Performance Outcome Measure—a measurement based on a task(s) performed by a patient according to instructions that is administered by a healthcare professional.[14]
PRO	Patient-Reported Outcome Measure—a measurement based on a report that comes directly from the patient (i.e. study subject) about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else.[14]
Qualification	Qualification—a conclusion, based on a formal regulatory process that within the stated context of use, a medical product development tool can be relied upon to have a specific interpretation and application in medical product development and regulatory review.[14]

(Continued)

Table 1. (Continued)

Acronym	Term
RRMS	Relapsing Remitting Multiple Sclerosis
SAP	Statistical Analysis Plan
SDMT	Symbol Digit Modalities Test
SDTM	Study Data Tabulation Model—provides a standardized, predefined collection of domains for clinical data submissions
SF-36	Short Form (36) Health Survey—a 36-item patient-reported survey of patient health
SPMS	Secondary Progressive Multiple Sclerosis
T25FW	Timed 25-Foot Walk—a quantitative mobility and leg function performance test based on a timed 25-walk
TAPSC	Therapeutic Area standards Program Steering Committee—an operations group in CFAST focused on therapeutic area data standards
WHO ICF	World Health Organization International Classification of Functioning, Disability and Health

Table 2. Inclusion and exclusion criteria for selecting ICF domains.

Inclusion criteria	Exclusion criteria
A domain must represent something related to MS that affects a significant proportion of people with MS.	The domain is not thought to relate to activities, functions, or roles that are important to people with MS in their everyday lives.
A domain must be something that can be measured objectively and that does not rely entirely on patient-reported symptoms.	The domain is not commonly affected in people with MS (e.g. hearing).
A domain must be something that can be measured easily, with minimal equipment, and in a reasonable amount of time.	The domain does not change over time or vary depending on MS severity.
A domain must be something that affects a real-life function that is meaningful to a person with MS.	The domain cannot be objectively assessed (e.g. fatigue or pain).
A domain should preferably be one for which accessible data exist from MS clinical trials.	The function related to the domain cannot be quantified or cannot be measured using practical test procedures (e.g. sexual function).

An important component of this process was the Workgroup's mapping of the ICF domains to activities of daily living that are limited by MS. Several rounds of reviews were needed to reduce the candidate domains to a smaller set of finalist domains. The Workgroup then utilized a numerical rating system to arrive at a consensus concerning the most appropriate domains to be considered. This final set was then discussed with the Coordinating Committee, which endorsed the recommendations. The Workgroup then proceeded to identify the most appropriate performance measures to assess each of the domains.

Domains selected from the core and comprehensive ICF domains are shown in Table 3. Domains that did not represent common MS symptoms were eliminated with the understanding that the domain may be affected in a limited number of MS patients. Because the workgroup was only focused on objectively measurable domains, domains that could only be assessed by patient reports were eliminated with the understanding

that certain of these domains (e.g. fatigue, depression) represent significant issues in MS. Given the COU, that is, large clinical trials, certain domains (e.g. gait pattern functions) were considered of value but too complex to incorporate in such studies. Both memory and speed of information processing were considered for inclusion as measures of cognition. Evidence from the literature indicated that speed of information processing is involved in memory and has a stronger relationship to real-life activities such as employment. Therefore, speed of information processing was selected as the most useful cognitive domain. The final domains selected reflect a core set of real-life functions meaningful to MS patients for which data exist in clinical trial datasets and in the scientific literature.

Literature review and extraction methods

A related activity to further define the COI of disability in MS focused on the four domains selected by the Defining Disability Workgroup: ambulation, arm

Table 3. Activities of daily living limited by disability in MS mapped to ICF domains.

Activities of daily living limited by disability in MS	Bodily function involved in the activity of daily living	ICF domain[a]	Comments	Possible neuro performance measures
1. Remembering to take medications[8,16]	Cognition: learning, retention, and recall of information	b144	One of the most frequently impaired cognitive functions in MS patients but complex and time-consuming to measure.	1. California Verbal Learning Test 2. Brief Visuospatial Reminder Test 3. 7/24 Spatial Recall Test
2. Keeping up with conversations[8,17–19]	Cognition: speed and accuracy of processing information	b1600 b164	Very practical to measure with a good deal of literature to support it. In addition, it is a function that MS patients complain about and that is related to activities and participation.	1. Symbol Digit Modalities Test 2. Paced Serial Additions Test
3. Seeing someone crossing the street[8,20,21]	Vision: recognizing people and objects	b120	Basic to many daily activities and practical, sensitive tests are available.	Low Contrast Letter Acuity
4. Reading a newspaper[8]	Vision: reading	b120	Basic to many daily activities and practical, sensitive tests are available.	Low Contrast Letter Acuity
5. Walking quickly to be on time for an appointment[8,22–24]	Ambulation: walking at different speeds	d450 b730	Frequently affected in MS and easily measured in varied clinical settings.	Timed 25-Foot Walk
6. Using a knife and fork, writing, and using a computer keyboard[8,25–27]	Coordination: fine hand use	d440 d445 b760	Often affected in MS and can interfere with a wide variety of important daily functions	9-Hole Peg Test

References in column 1 document the importance of the bodily function to people living with MS.
[a]ICF Brief Core and Comprehensive Domains: b120: seeing functions; b760: voluntary movement functions; b144: memory; d440: fine hand use; b164: higher level cognitive functions; d445: hand and arm use; b730: muscle power functions; d450: walking; b1600: pace of thought.

Table 4. Literature review research questions.

1. What are the most common symptoms or impairments caused by MS?

2. Which MS symptoms or impairments are the most challenging for people with MS?

3. Which symptoms or impairments are most likely to be altered by treatment and/or are predictive of future worsening?

4. What daily activities are compromised by each of the symptoms or impairments of MS?

5. What validated measures exist to evaluate MS symptoms or impairments?

6. What are the psychometric properties of these measures, including dimensionality, reliability (reproducibility, internal consistency, inter-rater agreement, etc.), validity, objectivity, sensitivity to differences and change, predictive validity, and clinical meaningfulness, among others?

7. How feasible (cost, complexity, timeliness, etc.) are these measures for use in large clinical trials?

8. How have these measures performed to date in the context of clinical trials?

9. How adequate is the published evidence supporting the utilization of these measures, based on standard criteria for level of evidence?

10. What is the evidence concerning what constitutes the size of a change or difference in each measure that is both perceptible to a person living with MS and that constitutes an important difference in day-to-day function?
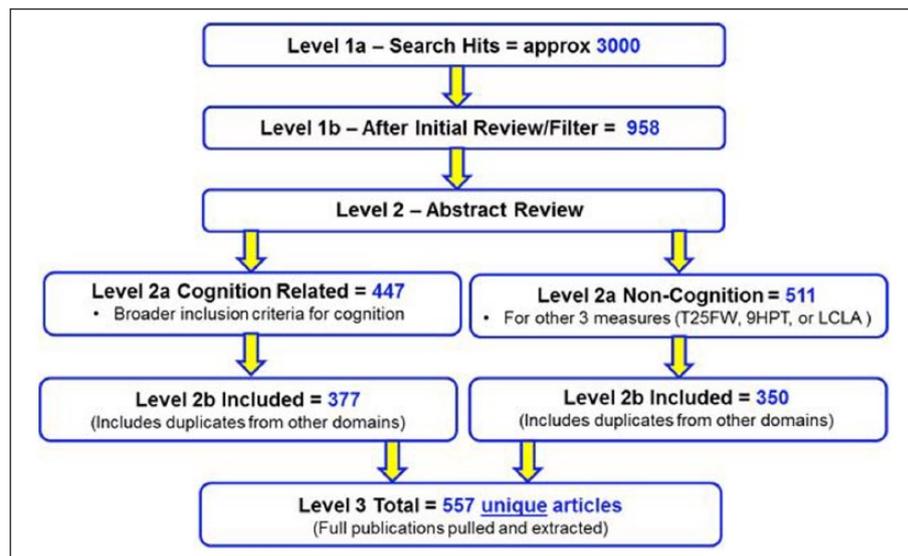
Figure 1. Overview of literature review results.

The literature review was performed in three consecutive levels. For Level 2, the workgroup applied a structured algorithm to review abstracts and categorize each article into "Included" or "Excluded" (see Abstract Filtering Criteria in Supplementary Table 3), with reasons for exclusion. Abstracts that included a measure of cognition received a code of "keep" or "exclude" only. Abstracts that included measures of vision, manual dexterity, or ambulation received a code of "high," "medium," and "low" importance or "exclude." All articles meeting the inclusion criteria were entered into the Data Extraction Table (Supplementary Table 4) and ranked to identify articles of most relevance to study objectives. Additional exclusions were also identified at this stage. For those articles that met the inclusion criteria, subject matter experts (SMEs) carried out the initial analysis and a separate group reviewed the analysis. A meeting was held to adjudicate any differences in the determinations of the SMEs. The Data Extraction Table (Supplementary Table 4) contains information on a total of 564 reviewed publications.

dexterity, vision, and cognition. Research questions were developed (Table 4) that could be addressed through an extensive literature review. Search parameters were designed to identify articles on performance measures relevant to domains of interest. In addition to the T25FW, 9HPT, LCLA, and SDMT, alternate measures used in the four domains were included in the literature search as well as articles that would combine domains in a disability assessment.

The literature review was performed in three levels (Figure 1). Parameters and search terms were defined (Supplementary Table 2) in Level 1 and abstract filtering criteria (Supplementary Table 3) were applied during the Level 2 Review. In reviewing the initial search, the Literature Review Workgroup identified a number of key papers that had been missed because key words and abstracts did not always include the performance measure search terms. Alternative search criteria increased

the number of abstracts identified to approximately 9000. Broadening the search criteria captured the missing articles but also identified many other articles unrelated to the scope of the project. The Literature Review Workgroup decided to use an enrichment technique that allowed the addition of subject matter experts (SMEs)-recommended papers that should definitely be in the review. This combined "enriched search" approach identified approximately 3000 papers.

Based on the results from the literature review, the SDMT was selected as the measure of choice[9,28,29] for processing speed. The workgroup considered potential measures of vision and decided on LCLA, utilizing 1.25% and 2.5% contrast Sloan charts, based on its strong performance in recent clinical trials.[11] Walking was considered as essential for inclusion, and the workgroup decided on walking speed as the most appropriate measure based in large part on the extensive use of the T25FW in clinical trials as part of the MSFC.[12,30] Finally, the workgroup endorsed the inclusion of a measure of manual dexterity to assess upper extremity function including coordination. The 9HPT, also part of the MSFC, was endorsed as the most appropriate measure in this domain based on its successful use in numerous clinical trials.[10,30]

Articles analyzed by the Literature Review Workgroup (see Data Extraction Table, Supplementary Table 4) were drawn on for recently published review articles, which summarized the utility and validity of each recommended measure.[9–12] Authors of the reviews determined which of the identified articles should be included in the reviews. Using the vision domain as an example, the actual search parameters were not specifically designed to assess vision in all its aspects in MS, nor even LCLA when used in other non-MS settings. In addition, background and technical references (e.g. information on optical coherence tomography (OCT), visual evoked potentials (VEP) etc.) were included in the vision publication that support the use of LCLA in MS but that were not part of the formal literature search. A similar approach was used for published reviews of the other domains.

Developing a CDISC therapeutic area data standard for MS

To allow aggregation of data from clinical trials, a common data standard for MS had to be developed and data from each trial remapped to that standard. The process for creating the first MS data standard was instituted through the Coalition for Accelerating Standards and Therapies (CFAST), an initiative formed by CDISC and C-Path to create and maintain data

standards in therapeutic areas important to public health. In general, the process mirrors that of other Standards Development Organizations (SDOs), including the International Organization for Standardization (ISO), Health Level 7 (HL7), and Integrating the Healthcare Enterprise (IHE). In brief, the sequential steps include scoping/charter, modeling and producing a draft standard, initial review and comment disposition, final public review and comment, disposition, and publication. The comment disposition ensures that those who contribute to the development process know how the comments were resolved to produce the resulting consensus-based standard.

C-Path submitted the scoping proposal for approval to the Therapeutic Area Program Steering Committee (TAPSC) that is organized by CFAST. The scoping proposal included a brief description of the project, including background information and proposed deliverables. Following approval of the scoping proposal, a detailed project proposal was submitted to TAPSC. The charter contained detailed information on the proposed standard, including focus populations, proposed team members/roles and other resources, stakeholder engagement considerations, concepts in scope, and a gap analysis of these concepts versus existing CDISC standards. The TAPSC reviewed and approved the charter. The Data Standards and Integration Workgroup developed the concept model and drafted the data standard, which was subsequently subjected to two rounds of review, including a public comment process. Revisions were incorporated, and a separate group of data standard experts carried out the final review and approval.

The Workgroup drew on the information content of the "common data elements (CDEs)" for MS that were developed through National Institute of Neurological Disorders and Stroke (NINDS)-supported efforts to identify those biomedical concepts that would form the MS CDISC data standard. Though CDEs guide researchers with recommendations on what should be captured and ensure consistent definitions of the captured content, they do not stand alone as data standards. A complete data standard also specifies how the collected data are represented in a database. Data standard specifications must also account for the often complex relationships between individual data elements to ensure that reviewers can construct accurate analyses involving multiple data elements which may exist in more than one table in the database.

Some of the retained CDEs were also further refined into comprehensive concepts—represented visually as concept maps—that described their origins in
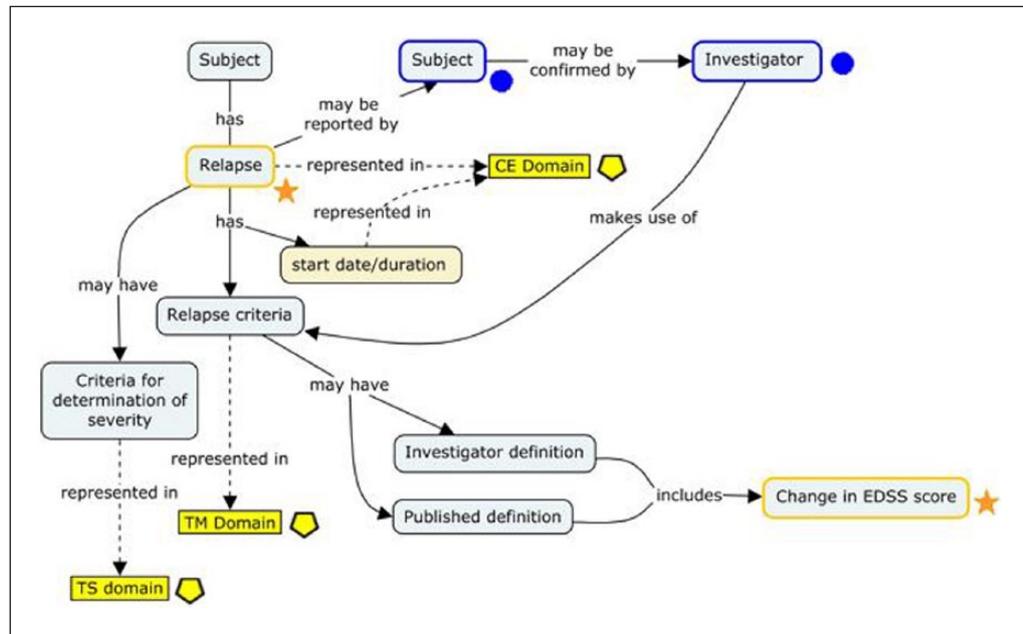
Figure 2. A concept map representing relapse in MS.
Multiple pathways can lead to the conclusion that relapse has occurred in patients with MS, including variations in relapse criteria and criteria for determination of severity. These criteria are typically anchored on changes in EDSS score. The resulting data standard accommodates this variation and specifies where in the CDISC data model (SDTM) this information can be found (represented by yellow boxes). CE is the Clinical Events domain; TM is the Trial Milestones domain; TS is the Trial Summary domain. Symbols correspond to the Bioinformatics Research Information Domain Grid (BRIDG) model concept classes (not discussed). Stars in orange represent the "Observation Result" class; circles in blue represent the "Assessor" class; and pentagons in yellow represent SDTM domains.

study-related processes and their interrelationships to other data elements. One such concept map ("relapse") is presented in Figure 2. Standard development workgroup discussions revealed that multiple pathways can lead to the conclusion that relapse has occurred in patients with MS, including variations in relapse criteria and criteria for determining severity. These criteria are typically, but not always, anchored on changes in EDSS score. The resulting data standard accommodates this variation and specifies where in the CDISC data model (SDTM) this information can be found (represented by yellow boxes). In the CDISC Therapeutic Area User Guide (TAUG) for MS v1.0, this concept map is followed by more explicit mock data examples showing how these data are represented and how they are linked to each other in a relational database (http://www.cdisc.org/therapeutic#MS; Figure 3).

During the development of the MS data standard, it was recognized that the organization of data within SDTM would benefit from the creation of two additional SDTM domains: (1) Functional Tests (FT), which includes performance measures such as the T25FW and 9HPT, and (2) Ophthalmic Examinations (OE), which includes the LCLA findings. The

SDTM domains used for the MSOAC database are shown in Table 5.

Acquiring, standardizing, and pooling data from MS clinical trials

To facilitate sharing of clinical trial data, C-Path developed two legal agreements that govern MSOAC membership and data contributions. Following execution of the legal agreements, MSOAC acquired 16 datasets from consortium industry and academic members (Table 6) and remapped the data to the new CDISC data standard for MS (Figure 4). The standardized data consisting of control and treatment arms of clinical trials formed the MSOAC database. The database includes information on a range of performance measures from 14,370 study subjects. Baseline descriptive statistics for age, sex, race, treatment arms, and disease severity as assessed by EDSS are shown in Figure 5.

As a resource for the research community, a database containing the placebo arms of MS clinical trials was also established (https://c-path.org/programs/msoac/). C-Path staff secured permission for the inclusion of ~2500 individual patient records that are part of the overall MSOAC database and developed the infrastructure to support the
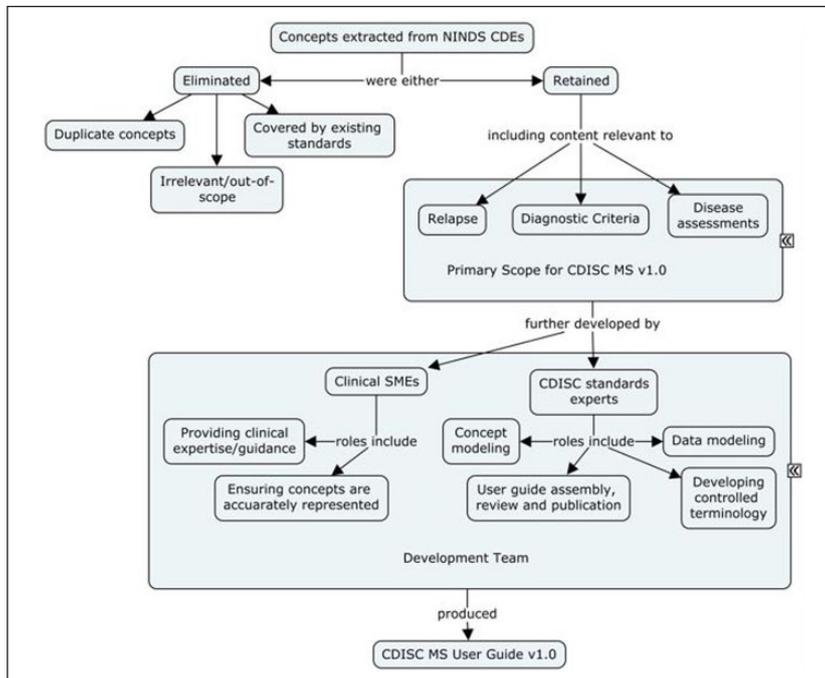
Figure 3. Process used for development of CDISC data standards for MS, v1.0.

A working group of MS subject matter experts and CDISC data standards experts reviewed the NINDS Common Data Elements for MS (https://www.commondataelements.ninds.nih.gov/MS.aspx#tab=Data_StandardsNLM/NIH). Concepts extracted from these CDEs were either retained or eliminated based on their relevance and whether or not they were already represented in CDISC SDTM standards. The concepts retained formed the scope for v1.0 of the MS CDISC User Guide, and those were further developed by CDISC data modelers in consultation with clinical SMEs. Development work included concept maps, data modeling examples, and controlled terminology. Controlled terminology was developed in collaboration with the NCI Enterprise Vocabulary Service (EVS). Examples of controlled terminology developed for MS include more than 500 terms registered in support of coding the various items and scores for the clinical outcome assessments and functional tests developed as a part of this project. The draft user guide was assembled as the output of this working group.

Table 5. Study data tabulation model (SDTM) domains used for the MSOAC database.

SDTM domain	Abbreviation	Observation class	Contents
Clinical Events	CE	Events	MS symptoms and relapse events and other events
Concomitant/Prior Medications	CM	Interventions	Betaseron, dexamethasone, glatiramer acetate, interferon, methylprednisolone, prednisolone, prednisone, etc.
Demographics	DM	Special purpose	Age, gender, race, trial arm, country
Disposition	DS	Events	Informed consent, randomization, reason for early withdrawal
Findings About Clinical Events	FACE	Findings sub-class	Number of relapses, relapses requiring hospitalization or steroids, result of relapse diagnosis tests, etc.
Findings About Medical History	FAMH	Findings sub-class	Number of relapses 1, 2, or 3 years before study start or since MS diagnosis; experienced acute relapse
Functional Tests	FT	Findings	T25FW, 9HPT, PASAT, SDMT
Medical History	MH	Events	MS diagnosis and pre-study symptoms, general medical history
Ophthalmic Examinations	OE	Findings	Visual acuity (low and high contrast)
Physical Examination	PE	Findings	General physical exam
Questionnaires	QS	Findings	BDI-FS, BDI-II, EDSS, FS scores, MSNQ, Neurological Change Questionnaire, RAND-36, SF-36, SF-12
Reproductive System Findings	RP	Findings	Pregnancy test
Subject Characteristics	SC	Findings	Dominant hand
Subject Disease Milestones	SM	Special purpose	MS relapse events
Trial Disease Milestones	TM	Trial design	Definitions of MS relapse

The Functional Tests (FT) and Ophthalmic Examinations (OE) domains were developed as a result of the therapeutic area data standard for MS. PASAT and SDMT are cognitive function tests that are included in the FT domain.

Table 6. Source datasets in the MSOAC database.

CT.gov no.[a]	Description	N	MS Type	EDSS	FSS	T25FW	9HPT	PASAT	SDMT	LCLA	SF-36	BDI-II
NCT00027300	AFFIRM	939	RRMS	√	√	√	√	√	NO	√	√	NO
NCT00030966	SENTINEL	1196	RRMS	√	√	√	√	√	NO	√	√	√
NCT00127530	MS-F203	301	ALL	√	NO	√	NO	NO	NO	NO	NO	NO
NCT00134563	TEMSO	1086	RRMS	√	√	√	√	√	NO	NO	√	NO
NCT00211887	COMBIRX	1008	RRMS	√	√	√	√	√	NO	√	√	NO
NCT00289978	FREEDOMS	1272	RRMS	√	√	√	√	√	NO	NO	NO	NO
NCT00297232	STRATA	1094	RRMS	√	√	NO	NO	NO	√	NO	NO	BDI-FS
NCT00340834	TRANSFORMS	1292	RRMS	√	√	√	√	√	NO	√	NO	NO
NCT00355134	FREEDOMS II	1083	RRMS	√	√	√	√	√	NO	√	NO	NO
NCT00483652	MS-F204	239	ALL	√	NO	√	NO	NO	NO	NO	NO	NO
NCT00530348	CARE-MS 1	563	RRMS	√	√	√	√	√	NO	√	√	NO
NCT00548405	CARE-MS 2	798	RRMS	√	√	√	√	√	NO	√	√	NO
NCT00869726	MAESTRO	610	SPMS	√	√	√	√	√	NO	NO	√	NO
NCT00906399	ADVANCE	1512	RRMS	√	√	√	√	√	√	√	SF-12	√
N/A[b]	PROMISE	943	PPMS	√	√	√	√	√	NO	NO	√	NO
N/A[b]	IMPACT	434	SPMS	√	√	√	√	√	NO	NO	√	√

The outcome measures that are included in the MSOAC database from each study are indicated by a check; No indicates that data from a particular measure are not included in the database. N is the number of subjects in a dataset; for the MS Type, "All" includes RRMS, SPMS, and PPMS.
[a]CT.gov refers to the Clinical Trials.gov website where clinical trials are registered.
[b]Study does not have a Clinical Trials.gov identifier.
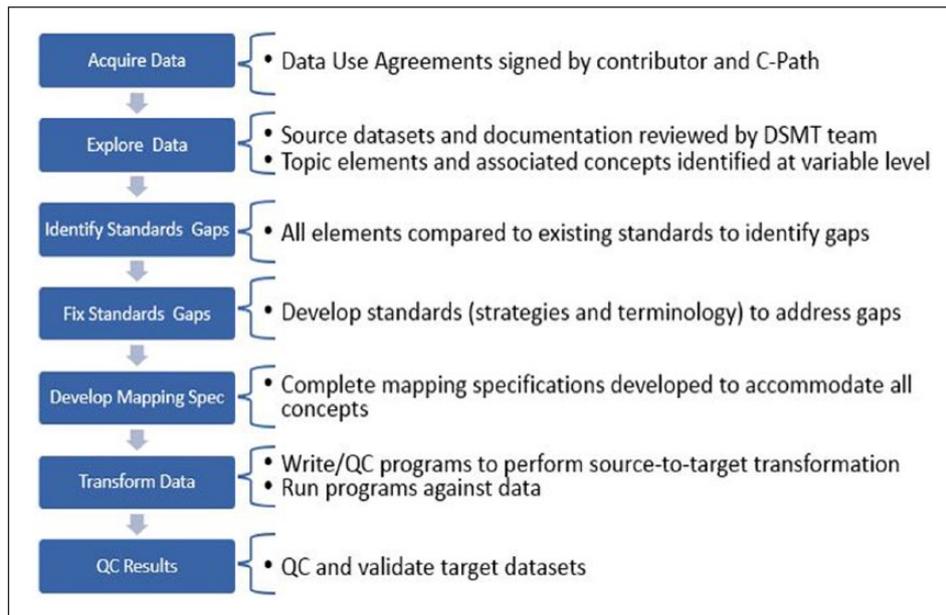
Figure 4. Steps in data mapping.
Data contributors agreed to provide data that meet the requirement of either a "Limited Dataset" or a "De-identified Dataset" in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor requirements (45 CFR 164 Subpart E, https://www.gpo.gov/fdsys/pkg/CFR-2011-title45-vol1/pdf/CFR-2011-title45-vol1-part164.pdf). C-Path policy is to exceed that minimum requirement whenever possible by further de-identifying the data after receiving it so that, in addition to meeting HIPAA Safe Harbor requirements, the data do not contain the year portion (which is allowed by Safe Harbor) of any dates. Instead, the timing of events is represented as time in relation to the study medication start date (defined as study day 1).
Data were standardized using the CDISC Study Data Tabulation Model (SDTM). This process was carried out by C-Path data managers in three stages: (1) logically planning how the data would fit into SDTM, (2) programmatically remapping the data, and (3) validating the mapped data. The purpose of the validation step was two-fold: to ensure that the meaning of the data was accurately preserved and to check for SDTM compliance. The latter was done with the aid of the Pinnacle 21 Validator tool. Once all the data were in SDTM format, the data were pooled together into an aggregated dataset that was provided to the CRO, Premier Research.
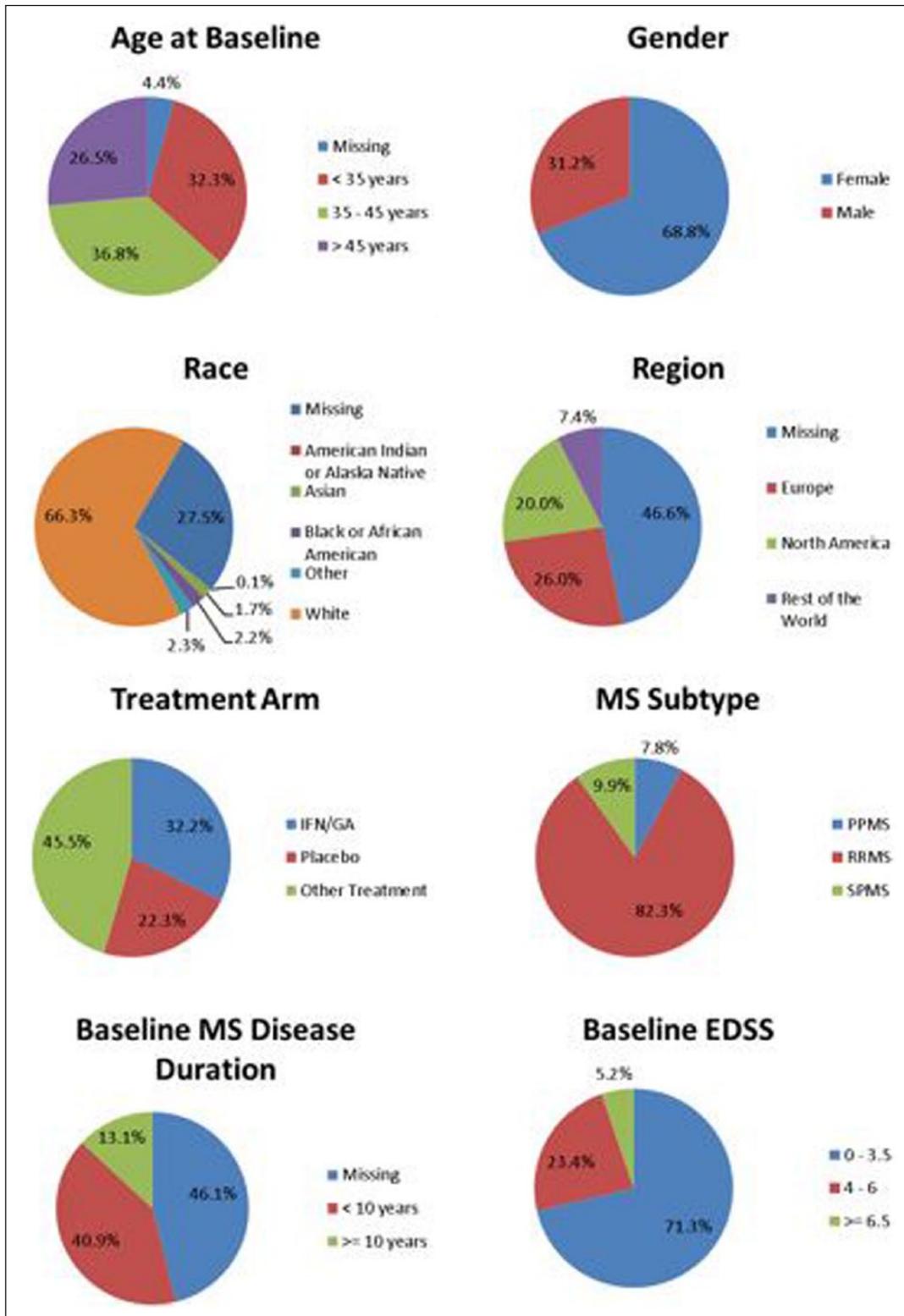
Figure 5. Baseline descriptive statistics for the pooled subjects in the MSOAC Database.
Through the data contribution agreement, contributors specified the level of access to the data (e.g. C-Path staff and contractors, MSOAC members) and which data they will contribute. Some contributors provided all of the data collected for a given study while others provided only the data relevant to the MSOAC analyses. Consequently, some information, including demographics on race and geographic region, is missing.
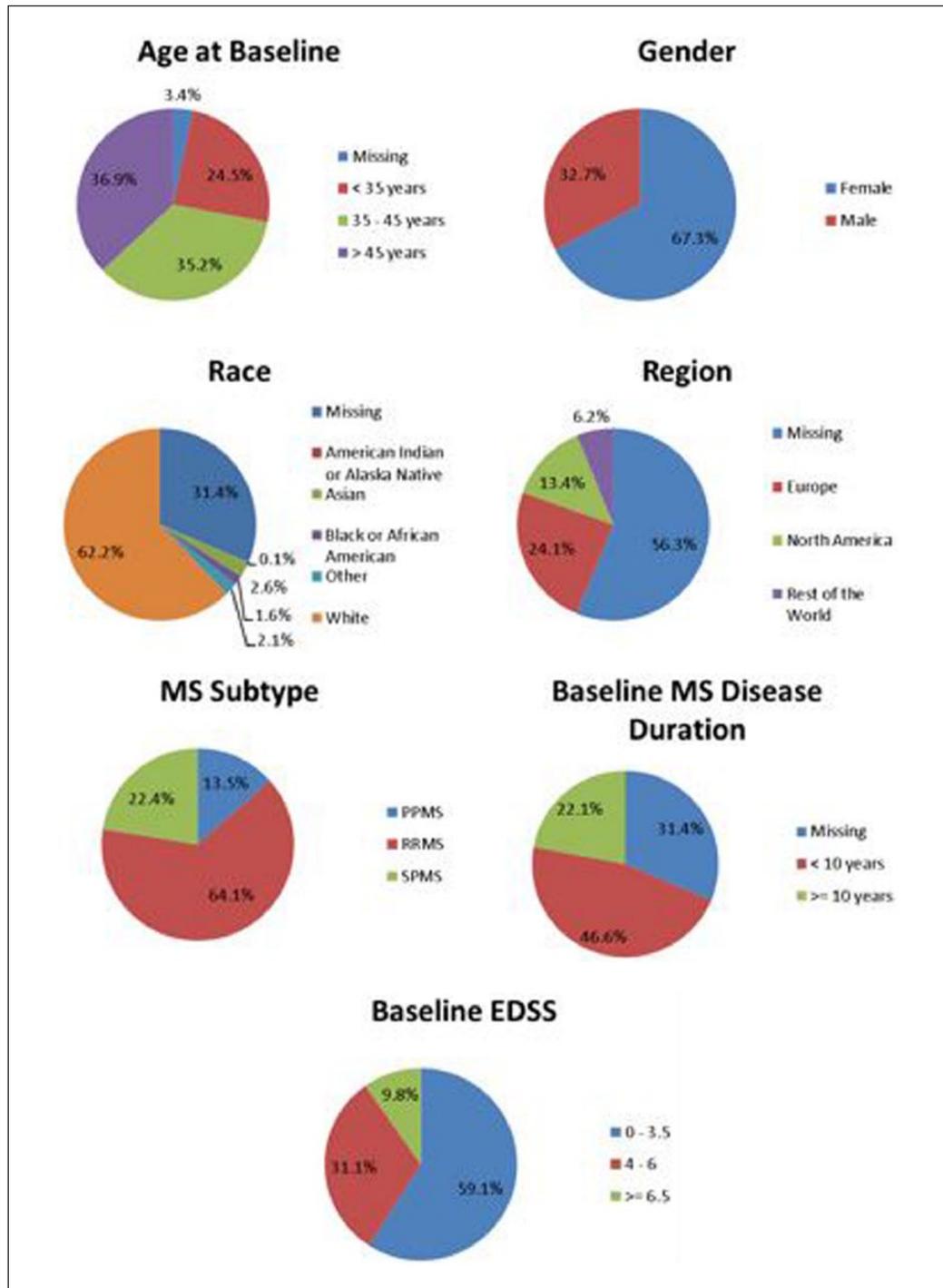
Figure 6. Baseline descriptive statistics for the pooled subjects in the placebo-arm database.
The data contribution agreement for each study stated whether the placebo-arm data could be made available to qualified researchers. For the studies where permission was granted, the standardized records for the placebo-arm subjects were copied and pooled together into a separate dataset. These records were further anonymized by removing the study IDs and using a random number generator to assign all of the subjects a new ID number. To request access to the database, use the following link: https://c-path.org/programs/msoac/.

storage, security, access requests, data use agreements, and access approvals, including a standing Review Board. Baseline descriptive statistics of this placebo-arm database are shown in Figure 6.

Analyzing data in the MSOAC database

MSOAC's Statistics Workgroup developed the SAP, incorporating both regulatory feedback and recommendations from MSOAC members on

which functional domains to examine, the analyses to be performed, and the optimal approach for incorporating the VOP. MSOAC members living with MS reviewed the initial plans for measuring clinically meaningful aspects of disability and identified gaps in the approach. A literature review provided insights on what aspects of disability are of most importance to people with MS and what performance measures adequately capture those concepts. Four performance measures were selected for detailed analysis, based on literature review and availability of PerfOs in the MSOAC database: the T25FW for ambulation, the 9HPT for manual dexterity, LCLA for vision (1.25% and 2.5% contrast), and both the SDMT and the PASAT for cognition. As detailed in the supplementary material, the following attributes were assessed for each measure: floor or ceiling effects, test–retest reliability, change over time, construct validity, convergent validity, extent of practice effects, known-group validity, sensitivity to change, and the minimum clinically important change in performance scores. Both the placebo arm and the treatment arm of the aggregated data were used for the statistical analyses. Results from the statistical analyses will be reported separately.

Conclusion

MSOAC was formed to develop more sensitive methods to measure whether a drug effectively reduces disability worsening in MS, based on the belief that acceptance of a more sensitive and precise, yet meaningful measure would accelerate progress in developing effective MS therapies. Therefore, the primary purpose for MSOAC was to qualify disability performance measures as primary or secondary endpoints for MS clinical trials submitted to the FDA and the EMA. Qualification of the SDMT as a measure of information processing speed is underway at the FDA, and qualification of all four performance measures (SDMT, T25FW, 9HPT, and LCLA) is in process at the EMA. In addition, given the preference for simple, reproducible performance tests, the consortium recognized that the same outcome measures could be useful within medical practice to grade MS severity and monitor patients over time, potentially harmonizing the metrics used in clinical trials and clinical practice. By including the same operationally defined, quantitative measures in clinical trials and healthcare settings, it should be possible to use "real world data" to augment clinical trials, test interventions in less controlled settings, and realize the potential of the "learning health system."[31]

MSOAC is a global effort, with members from 5 advocacy organizations, 2 regulatory agencies and 1 other governmental agency, 12 pharmaceutical companies, 23 academic institutions, 4 consultant groups, and 4 non-profit organizations. By sharing data and expertise in teams of volunteers that reported to the MSOAC Coordinating Committee (i.e. the Data Standards and Integration Workgroup, Defining Disability Workgroup, Clinical Outcome Assessments Workgroup, Regulatory Advisory Workgroup, Literature Review Workgroup, Statistical Workgroup, and VOP Workgroup), consortium members have delivered the following: (1) the first TAUG for MS, which is freely available at http://www.cdisc.org/therapeutic#MS; (2) a standardized database of 14,370 trial subjects for use in qualification of new PerfOs; (3) a placebo-arm database for use by the research community; (4) an extensive review of the literature on performance measures relevant to MS;[9–12] and (5) analyses of performance measure data for submission to the FDA and the European Medicines Agency for qualification. An approach to assess clinical meaningfulness of differences in the four measures by directly engaging persons with MS is also underway. Termed the VOP, this effort will contribute evidence toward the clinical meaningfulness of walking speed, manual dexterity, visual acuity, and speed of information processing in the lives of people with MS.

The consortium approach is not without challenges. Sharing data proved difficult or impossible for several members. All but one participant were willing to provide the needed copyright permissions for incorporation of scales into the MS CDISC data standard. Stakeholders were initially divided on the research plan, including optimal approaches to establish clinical meaningfulness.

Generating a CDISC standard for MS was a milestone that allowed pooling of clinical trial data for MSOAC's analysis and regulatory submission. The CDISC standard also provided a new tool to the entire MS community, which is of value now that all drug trial data submitted to the FDA[32] and Pharmaceuticals and Medical Devices Agency (PMDA) must be in CDISC format. Another consortium objective that benefits the research community was the creation of a separate database containing the placebo-arm data from registration trials (https://c-path.org/programs/msoac/). Most importantly, MSOAC's proposed outcome measure, once qualified by the EMA and the FDA, will be adopted by drug developers to demonstrate treatment benefit of

therapies designed to slow progression of disability and promote improvement in MS. MSOAC illustrates the potential for pre-competitive, cooperative, consortium-driven progress in drug development tools that benefit both sponsors and the broader MS community.

References

1. Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: A critical analysis. *Mult Scler* 1995; 1(1): 37–47.

2. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An Expanded Disability Status Scale (EDSS). *Neurology* 1983; 33(11): 1444–1452.

3. Rudick R, Antel J, Confavreux C, et al. Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996; 40(3): 469–479.

4. Balcer LJ, Baier ML, Cohen JA, et al. Contrast letter acuity as a visual component for the multiple sclerosis functional composite. *Neurology* 2003; 61(10): 1367–1373.

5. Ontaneda D, LaRocca N, Coetzee T, et al.; NMSS MSFC Task Force. Revisiting the multiple sclerosis

functional composite: Proceedings from the National Multiple Sclerosis Society (NMSS) Task Force on Clinical Disability Measures. *Mult Scler* 2012; 18(8): 1074–1080.

6. Fox RJ, Thompson A, Baker D, et al. Setting a research agenda for progressive multiple sclerosis: The International Collaborative on Progressive MS. *Mult Scler* 2012; 18(11): 1534–1540.

7. Rudick RA, Larocca N, Hudson LD, et al. Multiple Sclerosis Outcome Assessments Consortium: Genesis and initial project plan. *Mult Scler* 2014; 20(1): 12–17.

8. Coenen M, Cieza A, Freeman J, et al.; Members of the Consensus Conference. The development of ICF Core Sets for multiple sclerosis: Results of the International Consensus Conference. *J Neurol* 2011; 258(8): 1477–1488.

9. Benedict RH, DeLuca J, Phillips G, et al.; Multiple Sclerosis Outcome Assessments Consortium. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 721–733.

10. Feys P, Lamers I, Francis G, et al.; Multiple Sclerosis Outcome Assessments Consortium. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler* 2017; 23: 711–720.

11. Balcer LJ, Raynowska J, Nolan R, et al.; Multiple Sclerosis Outcome Assessments Consortium. Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 734–747.

12. Motl RW, Cohen JA, Benedict R, et al.; Multiple Sclerosis Outcome Assessments Consortium. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 704–710.

13. FDA. Clinical Outcome Assessment Qualification Program, https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm284077.htm; European Medicines Agency. Scientific advice and protocol assistance—Qualification of novel methodologies for medicine development, http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000319.jsp

14. The BEST (Biomarkers, EnpointS and other Tools) resource, https://www.ncbi.nlm.nih.gov/books/NBK338448/

15. Mayo NE. *Dictionary of quality of life and health outcomes measurement*. 1st ed. Milwaukee, WI: International Society for Quality of Life Research, 2015.

16. Rao SM, Leo GJ, Ellington L, et al. Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. *Neurology* 1991; 41(5): 692–696.

17. Strober LB, Christodoulou C, Benedict RHB, et al. Unemployment in multiple sclerosis: The contribution of personality and disease. *Mult Scler* 2012; 18(5): 647–653.

18. Sosnoff JJ, Balantrapu S, Pilutti LA, et al. Cognitive processing speed is related to fall frequency in older adults with multiple sclerosis. *Arch Phys Med Rehabil* 2013; 94(8): 1567–1572.

19. Strober LB, Rao SM, Lee JC, et al. Cognitive impairment in multiple sclerosis: An 18 year follow-up study. *Mult Scler Relat Disord* 2014; 3(4): 473–481.

20. Balcer LJ, Galetta SL, Polman CH, et al. Low-contrast acuity measures visual improvement in phase 3 trial of natalizumab in relapsing MS. *J Neurol Sci* 2012; 318(1–2): 119–124.

21. Feaster HT and Bruce JM. Visual acuity is associated with performance on visual and non-visual neuropsychological tests in multiple sclerosis. *Clin Neuropsychol* 2011; 25(4): 640–651.

22. Motl RW, McAuley E, Wynn D, et al. Physical activity, self-efficacy, and health-related quality of life in persons with multiple sclerosis: Analysis of associations between individual-level changes over one year. *Qual Life Res* 2013; 22(2): 253–261.

23. Pilutti LA, Dlugonski D, Sandroff BM, et al. Gait and six-minute walk performance in persons with multiple sclerosis. *J Neurol Sci* 2013; 334(1–2): 72–76.

24. Sidovar MF, Limone BL, Lee S, et al. Mapping the 12-item multiple sclerosis walking scale to the EuroQol 5-dimension index measure in North American multiple sclerosis patients. *BMJ Open* 2013; 3(5): e002798.

25. Bosma L, Kragt JJ, Polman CH, et al. Walking speed, rather than Expanded Disability Status Scale, relates to long-term patient-reported impact in progressive MS. *Mult Scler* 2013; 19(3): 326–333.

26. Costelloe L, O'Rourke K, McGuigan C, et al. The longitudinal relationship between the patient-reported Multiple Sclerosis Impact Scale and the clinician-assessed Multiple Sclerosis Functional Composite. *Mult Scler* 2008; 14(2): 255–258.

27. Polman CH and Rudick RA. The multiple sclerosis functional composite: A clinically meaningful measure of disability. *Neurology* 2010; 74(suppl. 3): S8–S15.

28. Goverover Y, Strober L, Chiaravalloti N, et al. Factors that moderate activity limitation and participation restriction in people with multiple sclerosis. *Am J Occup Ther* 2015; 69(2): 6902260020p1–6902260020p9.

29. Morrow SA, Drake A, Zivadinov R, et al. Predicting loss of employment over three years in multiple sclerosis: Clinically meaningful cognitive decline. *Clin Neuropsychol* 2010; 24(7): 1131–1145.

30. Rudick RA, Polman CH, Cohen JA, et al. Assessing disability progression with the Multiple Sclerosis Functional Composite. *Mult Scler* 2009; 15(8): 984–997.

31. Califf RM, Robb MA, Bindman AB, et al. Transforming evidence generation to support health and health care decisions. *N Engl J Med* 2016; 375(24): 2395–2400.

32. FDA references for the CDISC data standards—Study Data for Submission to CDER and CBER, https://www.fda.gov/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/ucm248635.htm; PMDA references for the CDISC data standards requirements, http://www.pmda.go.jp/files/000153708.pdf#page=6