Paper DS14

Multiple Sclerosis Trial Unpacked: Standards, Endpoints and Programming Experience

Daryna Teteriuk, Intego Group, LLC, Kyiv, Ukraine

ABSTRACT

Multiple Sclerosis (MS) remains one of the most complex therapeutic areas in clinical research, where data standards and statistical accuracy play a crucial role. This article offers a practical roadmap for statistical programmers who are either new to MS studies or looking to deepen their expertise. Readers will discover how the nature of the disease shapes trial design, how CDISC standards and agency guidelines bring consistency and traceability, and which study endpoints are most common in the field. Rather than focusing on theory alone, the author highlights real-world challenges and lessons learned, drawing on more than five years of hands-on programming experience in MS trials. By connecting standards, study design, and reporting, the article provides insights to help programmers bridge raw datasets to meaningful outputs.

INTRODUCTION

Multiple sclerosis (MS) is a neurological condition that affects the brain and spinal cord - the central nervous system, which regulates many essential bodily functions. MS is characterized by clinical relapses (periods of new or worsening neurological symptoms) and progressive neurological disability over time. It is an autoimmune disease, and while there is currently no cure, available therapies can meaningfully reduce disease activity and improve outcomes. Treatment focuses on reducing inflammation, managing symptoms, and preventing relapses through disease-modifying therapies (DMTs), with the goal of slowing disease progression and improving quality of life. MS can affect people of all ages, but it is more common in young adults and in females. Recent global estimates suggest that approximately 2.9 million people worldwide are living with MS in 2023, with a female-to-male ratio of 3:1. The incidence of multiple sclerosis appears to be increasing, and the average age at diagnosis is around 32 years. Beyond its clinical impact, MS is associated with reduced quality of life, treatment, and economic burden.

Clinical trials in MS have been highly productive, contributing to the development of numerous treatments. The trials focus on reducing inflammation - commonly measured through relapse rate and MRI lesion outcomes - and on slowing disability progression. Advances in imaging (MRI) and biomarkers such as serum neurofilament light chain (sNfL) have further strengthened the ability to monitor disease activity and treatment response. Based on available information from ClinicalTrials.gov, more than 3,500 studies involving MS have been registered. Of these, nearly 2,000 are listed as completed, around 2,500 are interventional, and roughly 1,000 are observational. While these numbers may vary depending on search criteria and updates to the registry, they highlight the scale of ongoing MS research and the continued need to develop and refine treatments for the many people affected.

The main clinical subtypes of MS include clinically isolated syndrome (CIS), relapsing-remitting MS (RRMS), secondary progressive MS (SPMS), and primary progressive MS (PPMS). RRMS is the most common subtype and is characterized by acute attacks followed by periods of partial or complete recovery. Approximately 85% of individuals with MS are present with RRMS; therefore, the examples in this article will focus primarily on RRMS studies, while the underlying principles are applicable across other MS subtypes as well.

These features make MS a good setting to discuss how standards, endpoints, and programming decisions connect - from raw data to analysis-ready datasets and clear, traceable outputs.

DATA STANDARDIZATION AND SUBMISSION READINESS

A practical first step in statistical programming is data standardization - making the study data CDISC-compliant to support a successful regulatory submission (e.g., to the FDA, PMDA, or EMA). At the outset, the standards to be used and their versions should be agreed. This typically includes SDTM and the SDTM Implementation Guide (SDTMIG) with define.xml, as well as ADaM and the ADaM Implementation Guide with define.xml. It also includes controlled terminology (e.g., CDISC SDTM and ADaM terminology), SDTMIG and ADaMIG questionnaire (QRS) supplements that provide information on how to structure the data in a standard format and the dictionary versions used for coding adverse events and medical history (MedDRA) and concomitant medications (WHO Drug). This agreement should be documented early as it drives mapping rules, dataset structures, validation strategy, and the final define.xml package.

In addition, the following considerations are important:

- Availability of the sponsor's internal standards. Many organizations maintain internal guidance that extends CDISC recommendations, especially for edge cases that are not fully covered in general documents. These rules help ensure consistent implementation across studies.

- Use of the most recent CDISC documentation available (when feasible). Updated standards support more efficient data flow from collection through submission and enable smoother data exchange between sponsors, CROs, and technology partners.
- Prior studies and potential pooling. When studies are intended to be integrated, aligning standards early improves cross-study comparability and reduces downstream rework.
- Applicability of the Study Data Technical Conformance Guide. This provides specifications, recommendations, and general considerations for submitting standardized study data using FDA-supported data standards.
- Availability and relevance of Therapeutic Area User Guides (TAUGs). TAUGs are intended to supplement SDTM and the SDTMIG by describing therapy-specific implementation approaches. For MS trials, the TAUG aligns with SDTM v1.4 and SDTMIG v3.2 and focuses on a subset of domains with MS-specific guidance. If a later SDTM/SDTMIG version is used, priority should be given to the foundational CDISC standards and their implementation guides. TAUG familiarity is helpful, but it does not replace working knowledge of SDTM and is not sufficient on its own to produce complete, SDTM-compliant regulatory submissions for MS clinical trial data.

MS TAUG

A large portion of MS trial data can be collected and standardized using existing CDASH and SDTM conventions without any MS-specific customization. This includes common clinical trial domains such as Demographics, Adverse Events, Concomitant Medications, Disposition, Exposure, Laboratory Tests, Vital Signs, Medical History, and related trial design domains. In practice, these areas are usually driven by the sponsor's general standards and SDTMIG, not by a therapeutic area guide.

The Multiple Sclerosis TAUG (v1.0) was created to illustrate how MS-specific concepts can be represented consistently with SDTM and controlled terminology. In MS, TAUG examples are most helpful when the study includes disease-specific assessments that are not straightforward to model using "generic" SDTM patterns. Typical examples include functional tests and questionnaires (e.g., PRO instruments), visual outcomes (visual acuity / contrast sensitivity), diagnosis and disease characteristics, relapses, optical coherence tomography (OCT), and visual evoked potential (VEP). The key point for programmers is that TAUG is an example-based supplement. It helps align how MS concepts are represented, but it does not replace SDTM/SDTMIG knowledge and it does not cover every data type collected in MS trials.

Historically, MS TAUG examples relied on domains that were not fully mature at the time of publication. Over time, SDTMIG versions have expanded and formalized many body-system findings domains (including Nervous System Findings [NV] and Ophthalmic Examinations [OE]). Practically, this means that when you implement TAUG concepts today, you should always reconcile the TAUG example with the SDTMIG version you are using and follow the SDTMIG as the primary reference. Some TAUG examples propose a modeling approach for data types that are still evolving. That is useful, but it can change as SDTMIG guidance changes. A good example is the Morphology (MO) domain: SDTMIG v3.4 explicitly decommissioned MO. If older MS TAUG examples map OCT-derived retinal thickness measures (e.g., RNFL thickness) to MO, those mappings should be updated. In current practice, ophthalmic device-derived measures typically fit better in an eye-focused findings structure (often OE) or another appropriate body-system findings domain, depending on the SDTMIG version and sponsor conventions.
MS trials rely heavily on MRI for disease activity and treatment effect interpretation, but not every MS data type is fully illustrated in TAUG examples. In later endpoint sections of this paper, TAUG-aligned examples will be used where they add value (particularly for relapse-related outcomes, disability/functional measures), while other endpoints will be described using standard SDTM/ADaM principles.

ENDPOINTS IN MULTIPLE SCLEROSIS TRIALS

Multiple sclerosis (MS) is commonly evaluated using endpoints such as relapse frequency, disability progression, and MRI activity. In addition, studies assess "no evidence of disease activity" (NEDA), a composite concept that combines these three components. From a statistical perspective, analyzing these endpoints can present several challenges, particularly due to differing data structures, assessment schedules, and interrelationships between measures.

The most frequently used endpoints can be grouped as follows:

- Clinical outcomes
  - Relapses: relapse rate; annualized relapse rate (ARR)
  - Disability and function: Expanded Disability Status Scale (EDSS); Confirmed Disability Progression (CDP); Timed 25-Foot Walk (T25FW); 9-Hole Peg Test (9HPT); Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS); Hospital Anxiety and Depression Scale (HADS); Fatigue Severity Scale (FSS)
- Biological markers
  - MRI-based measures: gadolinium-enhancing T1 lesions; T2 lesions; active lesions; combined unique active (CUA) lesions; brain volume; brain atrophy
  - Other biomarkers: cerebrospinal fluid (CSF) oligoclonal bands; neurofilament light chain (NfL) in CSF

- Composite measures
  - Disease activity and progression composites: NEDA-3; NEDA-4; No Evidence of Progression or Active Disease (NEPAD)
  - Relapse-independent progression constructs: Progression Independent of Relapses (PIRA); composite PIRA (cPIRA); modified composite PIRA (mcPIRA)
  - Relapse-related worsening constructs: Relapse Worsening Activity (RAW); composite RAW (cRAW); modified composite RAW (mcRAW)
  - Disability accumulation constructs: Confirmed Disability Accumulation (CDA); composite CDA (cCDA); modified composite CDA (mcCDA)
- Patient-reported outcomes (PROs)
  - Multiple Sclerosis Quality of Life-54 (MSQOL-54)
  - Treatment Satisfaction Questionnaire for Medication, version 1.4 (TSQM v1.4)
- Associations across endpoint types
  - Correlations between clinical outcomes and PROs, as well as between clinical, imaging, and biomarker endpoints

The following subsections examine these endpoints in more detail.

RELAPSES

In MS trials, a relapse (also called an attack or exacerbation) is typically defined as new or clearly worsening neurologic symptoms that last at least 24 hours, occur in the absence of fever or active infection, and are separated from a prior relapse by a period of clinical stability (often at least 30 days).

In most MS trials, relapses are not collected as adverse events (AEs). Instead, they are captured on dedicated eCRF pages and used primarily for efficacy analyses. The forms below are examples of basic collection for MS onset history and relapses during the study.

History of MS onset	
MS Onset Date (first symptom)	DD-MMM-YYYY
Major systems effected	□ Pyramidal Function □ Cerebellar Function □ Brain Stem Function □ Sensory Function □ Bowel and Bladder Function □ Visual (optic) Function □ Cerebral (mental) Function □ Other, specify _____
Second Attack date (first relapse)	DD-MMM-YYYY
Diagnosis date	DD-MMM-YYYY

MS Relapse	
Onset of new relapse	DD-MMM-YYYY
Stabilization/resolution date	DD-MMM-YYYY
Was steroid treatment required?	○ Yes ○ No
Was the subject hospitalized?	○ Yes Reason for hospitalization ____ Hospital admission date DD-MMM-YYYY Hospital discharge date DD-MMM-YYYY ○ No
Major systems effected	□ Pyramidal Function □ Cerebellar Function □ Brain Stem Function □ Sensory Function □ Bowel and Bladder Function □ Visual (optic) Function □ Cerebral (mental) Function □ Other, specify _____
Severity	○ Mild ○ Moderate ○ Severe
Was the relapse qualifying or non-qualifying according to the protocol definition?	○ Qualifying ○ Non-qualifying

From an SDTM perspective, RRMS onset is typically represented in Medical History (MH), while additional clinical details (e.g., functional system affected, second attack, diagnosis date) are stored in Findings About (FA). Relapse events are represented in Clinical Events (CE), with supporting details captured in FA. Hospitalization information is

stored in Hospitalization (HO). If certain data points cannot be represented in a standard domain and are not appropriate for a supplemental qualifier structure, SUPP-- may be used (e.g., reason for hospitalization). Relationships across domains are documented in the RELREC dataset.

Example below is for subject MS01-001 who has MS onset in 1999-03, with pyramidal function affected; the second attack occurred on 1993-12-17, and the diagnosis date was 1994-01-20. After study enrollment, the subject experienced a mild relapse starting on 2012-05-28 and resolving on 2012-06-18. Sensory and brainstem functions were affected, and the relapse was classified as qualifying.

mh.xpt

STUDYID	DOMAIN	USUBJID	MHLNKID	MHTERM	MHDECOD
ABC123	MH	MS01-001	1	RRMS	Relapsing Remitting Multiple Sclerosis
MHCAT	MHSCAT	MHSTDTC			
PRIMARY DIAGNOSIS	ONSET COURSE	1999-03			

ce.xpt

STUDYID	DOMAIN	USUBJID	CELNKID	CETERM	CESEV
ABC123	CE	MS01-001	2	Multiple Sclerosis Relapse	MILD
CESTDTC	CEENDTC				
2012-05-28	2012-06-18				

fa.xpt

STUDYID	DOMAIN	USUBJID	FALNKID	FATESTCD	FATEST
ABC123	FA	MS01-001	1	LOCONSET	Localization of Onset
ABC123	FA	MS01-001	1	DATE	Date
ABC123	FA	MS01-001	1	DATE	Date
ABC123	FA	MS01-001	2	LOCONSET	Localization of Onset
ABC123	FA	MS01-001	2	LOCONSET	Localization of Onset
ABC123	FA	MS01-001	2	QUAL	Qualifying Relapse
FAOBJ	FACAT	FASCAT	FAORRES	FALOC	FADTC
RRMS	PRIMARY DIAGNOSIS	ONSET COURSE	Y	PYRAMIDAL FUNCTION	1999-03
Second Attack (First Relapse)	PRIMARY DIAGNOSIS	ONSET COURSE	1993-12-17		1999-03
Diagnosis	PRIMARY DIAGNOSIS	ONSET COURSE	1994-01-20		1999-03
Multiple Sclerosis Relapse	MS RELAPSE REPORT		Y	SENSORY FUNCTION	2012-05-28
Multiple Sclerosis Relapse	MS RELAPSE REPORT		Y	BRAIN STEM FUNCTION	2012-05-28
Was the Relapse Qualifying or Non-Qualifying According to the Protocol Definition	MS RELAPSE REPORT		QUALIFYING		2012-05-28

ho.xpt

STUDYID	DOMAIN	USUBJID	HOLNKID	HOTERM	HOPRESP	HOOCUR	HOSTDTC	HOSTDTC
ABC123	HO	MS01-001	2	HOSPITAL	Y	Y	2012-06-01	2012-06-15

relrec.xpt

STUDYID	RDOMAIN	USUBJID	IDVAR	IDVARVAL	RELTYPE	RELID
ABC123	MH		MHLNKID		ONE	1
ABC123	FA		FALNKID		MANY	1
ABC123	CE		CELNKID		ONE	2
ABC123	FA		FALNKID		MANY	2
ABC123	HO		HOLNKID		ONE	2

For ADaM, relapse information is often consolidated across SDTM domains into a single analysis dataset using an occurrence structure, so that each relapse is represented by one row containing the relevant attributes. A separate BDS dataset may be created for summary endpoints such as number of relapses, relapse rate, ARR, and other study-specific parameters. Depending on sponsor preference and analysis requirements, individual relapse data and summary parameters may also be combined within a single BDS dataset. In such cases, separate parameters are created for each data element (e.g., onset date, severity), and a grouping variable is added to link records associated with the same relapse.

Below is an example of ADaM BDS dataset that includes both individual relapse-level data and summary parameters. A relapse identifier (for example, ANLKREID) is used to group records belonging to the same relapse. A timing variable such as ASDT is often used alongside the identifier to support sorting and interpretation, particularly when

subjects have multiple events. As a matter of good practice, SCR variables are recommended to support traceability and enable clear linkage back to the originating SDTM records. For partial relapse onset or resolution dates, imputation rules should be prospectively defined and documented in the SAP. This is critical because relapse timing can directly affect: relapse counts within analysis periods, person-time calculations for ARR, classification of events relative to treatment exposure and time-to-event derivations. For time-to-event analyses (e.g., time to first relapse), an ADTTE dataset should be produced.

adrel.xpt

STUDYID	USUBJID	ANLKID	ASTDT	PARAMCD	RARAM	AVAL	AVALC
ABC123	MS01-001	REL1	28MAY2012	RQUAL	Qualifying Relapse		QUALIFYING
ABC123	MS01-001	REL1	28MAY2012	RHOCCUR	Hospitalization Occurrence		Y
ABC123	MS01-001	REL1	28MAY2012	RCEENDTC	Resolution date		2012-06-18
…							
ABC123	MS01-001	REL2	20DEC2012	RQUAL	Qualifying Relapse		NON-QUALIFYING
ABC123	MS01-001	REL2	20DEC2012	RHOCCUR	Hospitalization Occurrence		N
…							
ABC123	MS01-001			RELNUM	Number of All Relapses	2	
ABC123	MS01-001			RELNUMH	Number of All Relapses leading to hospitalization	1	
ABC123	MS01-001			ARR	Annualized Relapse Rate	1.08	

Annualized relapse rate (ARR) is one of the most frequently reported summary endpoints in relapsing MS trials. Unadjusted ARR is defined as the number of relapses divided by follow-up time expressed in person-years at risk. The unadjusted population ARR is often presented as the total relapse count divided by total person-years, but it may also be derived as the mean of individual participant-level ARRs.

Adjusted ARR is typically estimated using a Poisson regression model or, when over-dispersion is present, a negative binomial regression model. A standard specification models relapse count as the dependent variable, includes log(follow-up time) as an offset, and adjusts for clinically relevant baseline covariates such as sex, age, and baseline EDSS. In SAS®, this can be implemented using PROC GENMOD, as illustrated below.

```
ods output ParameterEstimates=est;
ods output Estimates=res;
proc genmod data=inds;
class bledsscat(ref='ref_value');/*all categorical variables used as fixed effects to
be mentioned with corresponding reference values*/
model relapse=age bledsscat /dist=poisson link=log offset=time_ln /*time_ln is log of
the follow-up time*/
estimate "mean" intercept 1 age &mean_age bledsscat &mean_edss1 &mean_edss0;
run;
```

Other commonly used relapse endpoints include total relapse count, relapse rate over a prespecified period, and time to first relapse. When the analysis aims to use all relapse events rather than only the first, recurrent event methods can be applied. One option is the mean cumulative function (MCF), which summarizes the average cumulative number of relapses per subject over time and provides a clear view of relapse accumulation across the follow-up period.

DISABILITY PROGRESSION: EDSS AND CDP

The Expanded Disability Status Scale (EDSS) is the most widely used disability measure in MS clinical trials. It was introduced by Kurtzke in 1983 and remains a standard way to quantify neurologic disability over time. EDSS is scored from 0 (normal neurologic exam) to 10 (death due to MS) in 0.5-step increments. The scale has three broad ranges that reflect different clinical constructs: the lower range (up to 4.0) measures impairment in eight functional systems, the mid-range (4.0–7.0) focuses on ambulatory function, and the top range (7.0–9.5) largely assesses the ability to carry out activities of daily living. The graph below shows the distribution of patients over the EDSS.

EDSS is typically collected as a clinician-rated assessment at scheduled visits, using a dedicated form. The example below is based on an aCRF structure illustrated in SDTMIG QRS supplement.

Expanded Disability Status Scale	
Is EDSS completed?	○ Yes ○ No
If yes, please provide date of assessment	DD-MMM-YYYY
Expanded Disability Status Scale	□ 0.0 - Normal neurological exam (all grade 0 in all Functional System (FS) scores) □ 1.0 - No disability, minimal signs in one FS* (i.e., grade 1). □ 1.5 - No disability, minimal signs in more than one FS* (more than 1 FS grade 1). □ 2.0 - Minimal disability in one FS (one FS grade 2, others 0 or 1). □ 2.5 - Minimal disability in two FS (two FS grade 2, others 0 or 1).

	□ 3.0 - Moderate disability in one FS (one FS grade 3, others 0 or 1) or mild disability in three or four FS (three or four FS grade 2, others 0 or 1) though fully ambulatory. □ 3.5 - Fully ambulatory but with moderate disability in one FS (one grade 3) and one or two FS grade 2; or two FS grade 3 (others 0 or 1) or five grade 2 (others 0 or 1). □ 4.0 - Fully ambulatory without aid, self-sufficient, up and about some 12 hours a day despite relatively severe disability consisting of one FS grade 4 (others 0 or 1), or combination of lesser grades exceeding limits of previous steps. □ 4.5 - Fully ambulatory without aid, up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance; characterized by relatively severe disability consisting of one FS grade 4 (others 0 or 1) or combinations of lesser grades exceeding limits of previous steps; able to walk without aid or rest some 300 meters. □ 5.0 - Ambulatory without aid or rest for about 200 meters; disability severe enough to impair full daily activities (e.g., to work a full day without special provisions); (Usual FS equivalents are one grade 5 alone, others 0 or 1; or combinations of lesser grades usually exceeding specifications for step 4.0). □ 5.5 - Ambulatory without aid for about 100 meters; disability severe enough to preclude full daily activities; (Usual FS equivalents are one grade 5 alone, others 0 or 1; or combination of lesser grades usually exceeding those for step 4.0). □ 6.0 - Intermittent or unilateral constant assistance (cane, crutch, brace) required to walk about 100 meters with or without resting; (Usual FS equivalents are combinations with more than two FS grade 3+). □ 6.5 - Constant bilateral assistance (canes, crutches, braces) required to walk about 20 meters without resting; (Usual FS equivalents are combinations with more than two FS grade 3+). □ 7.0 - Unable to walk beyond approximately 5 meters even with aid, essentially restricted to wheelchair; wheels self in standard wheelchair and transfers alone; up and about in wheelchair some 12 hours a day; (Usual FS equivalents are combinations with more than one FS grade 4+; very rarely pyramidal grade 5 alone). □ 7.5 - Unable to take more than a few steps; restricted to wheelchair; may need aid in transfer; wheels self but cannot carry on in standard wheelchair a full day; May require motorized wheelchair; (Usual FS equivalents are combinations with more than one FS grade 4+). □ 8.0 - Essentially restricted to bed or chair or perambulated in wheelchair, but may be out of bed much of the day; retains many self-care functions; generally has effective use of arms; (Usual FS equivalents are combinations with more than one FS grade 4+). □ 8.5 - Essentially restricted to bed much of day; has some effective use of arm(s); retains some self-care functions; (Usual FS equivalents are combinations with more than one FS grade 4+). □ 9.0 - Helpless bed patient; can communicate and eat; (Usual FS equivalents are combinations with more than one FS grade 4+). □ 9.5 - Totally helpless bed patient; unable to communicate effectively or eat/swallow; (Usual FS equivalents are combinations with more than one FS grade 4+). □ 10.0 - Death due to MS

From an SDTM perspective, EDSS is typically mapped to the Questionnaires (QS) domain. There is SDTMIG QRS supplement for EDSS, released in 2013, provides controlled terminology and a standard data structure with examples.

The example below shows CDISC controlled terminology for QSTESTCD, QSTEST, and QSCAT. Subject MS01-001 had 3 EDSS data collected every 3 months after enrollment. Original results are represented using preferred terminology in QSORRES. These results are then converted to a standard numeric score in QSSTRESN, with a character representation in QSSTRESC. In some cases, original results may need minor formatting adjustments to comply with variable length requirements (e.g., representing a score of 4.5 in QSSTRESC).

qs.xpt

STUDYID	DOMAIN	USUBJID	QSTESTCD	QSTEST	QSCAT
ABC123	QS	MS01-001	EDSS0101	EDSS01-Expanded Disability Score	EDSS
ABC123	QS	MS01-001	EDSS0101	EDSS01-Expanded Disability Score	EDSS
ABC123	QS	MS01-001	EDSS0101	EDSS01-Expanded Disability Score	EDSS
QSORRES	QSSTRESC	QSSTRESN	VISITNUM	VISIT	QSDTC
Minimal disability in two FS.	2.5	2	1	Screening	2012-09-16
Fully ambulatory but with moderate disability in one FS and one or two FS grade 2; or two FS grade 3 or five grade 2.	3.5	3.5	3	Month 3	2012-12-18
Minimal disability in one FS.	2.0	2	6	Month 6	2013-02-13

In ADaM, EDSS assessments are stored in a BDS dataset, with one record per subject per assessment per visit. For CDP derivation, analysis variables are often added to facilitate profiling EDSS over time and identifying progression events. The example below illustrates CDP based on a sustained increase for at least 3 months (90 days). In this example, CDP begins on Month 12 with a 2-point increase from baseline and persists for nearly 6 months, maintaining an increase of 2 points or more during that period. On the Month 21 visit, the score decreases to 0.5 relative to the EDSS level at CDP onset, meaning end of CDP event. Variable CDPSTFL flags the start of CDP, and DIFFDY tracks the number of days between assessments to support confirmation rules. Assuming treatment starts on 17SEP2012, ADY is calculated relative to treatment start. Due to page size constraints, only key variables are shown.

adedss.xpt

USUBJID	AVISIT	ADT	ADY	DIFFDY	PARAMCD	AVAL	BASE	CHG	ABLFL	CDPSTFL
MS01-001	Screening	16SEP12	-1		EDSS0101	2.5			Y	
MS01-001	Month 3	18DEC12	93	94	EDSS0101	3.5	2	1.5		
MS01-001	Month 6	13MAR13	178	86	EDSS0101	2	2	0		
MS01-002	Month 9	14JUN13	271	94	EDSS0101	3	2	1		
MS01-003	**Month 12**	**15SEP13**	**364**	**94**	**EDSS0101**	**4**	**2**	**2**		**Y**
MS01-004	**Month 15**	**16DEC13**	**456**	**93**	**EDSS0101**	**4**	**2**	**2**		
MS01-005	**Month 18**	**17MAR14**	**547**	**92**	**EDSS0101**	**4.5**	**2**	**2.5**		
MS01-006	Month 21	18JUN14	640	94	EDSS0101	3.5	2	1.5		
MS01-006	Month 24	20SEP14	734	95	EDSS0101	4	2	2		

A key practical issue with CDP is the lack of a fully standardized definition. Differences across sponsors may include: the required confirmation duration (3-month vs 6-month), the baseline-dependent increase rules, allowable visit windows for confirmation, restriction on the magnitude or persistence of increase between the start and confirmation visits.

CDP depends on a valid baseline assessment and a confirmatory assessment within the required time window. Missing visits and window violations can lead to missed confirmations or ambiguous events. These issues should be addressed proactively in the protocol and SAP by clearly defining: the expected assessment schedule and visit windows, what constitutes a valid confirmation visit, the minimum time gap required for confirmation (and how early/late visits are handled), missing-data handling rules (including how to handle subjects without a confirmatory assessment). For example, if the confirmation of CDP is required at 6 months, the protocol should define the corresponding visits frequency, and the SAP should define the minimum acceptable interval for confirmation (e.g., 6 months x 30 days - 14 days), so that the confirmation would not be missed by few days.

CDP is frequently analyzed as time to first confirmed progression using standard survival analysis methods. This aligns well with the event-based nature of CDP and supports censoring rules that reflect follow-up duration and missing assessments. In SAS, Kaplan-Meier summaries are typically produced using PROC LIFETEST; a basic syntax example is provided below.

```
ods output Quartiles=quart;
ods output ProductLimitEstimates = ple;
proc lifetest data=inds method = km outsurv=kaplanmeier;
time aval * cnsr(1);
run;
```

FUNCTIONAL TESTS

In addition to EDSS, MS trials often include performance-based functional tests that capture disability in a more granular way. Two of the most common are the 9-Hole Peg Test (9HPT) and the Timed 25-Foot Walk (T25FW).

- 9HPT is a standardized test of upper extremity function. It measures the time required to place nine pegs into a board one at a time and then remove them as quickly as possible. The typical administration includes two trials per hand: two consecutive trials of the dominant hand followed by two consecutive trials of the non-dominant hand.
- T25FW measures mobility and leg function. It records the time it takes a subject to walk 25 feet as quickly but safely as possible, usually performed as two trials (out and back). Assistive devices are allowed if needed and should be recorded because they directly affect interpretability.

The forms below are examples of basic collection for 9HPT and T25FW during the study.

9-HOLE PEG TEST	
Is 9HPT completed?	○ Yes ○ No
If yes, please provide date of assessment	DD-MMM-YYYY

TIMED 25-FOOT WALK	
Is T25FW completed?	○ Yes ○ No
If yes, please provide date of assessment	DD-MMM-YYYY

Dominant hand	□ Right □ Left
Dominant hand	**Non-dominant hand**
Is Trial 1 completed? ○ Yes Time to complete (sec) __ ○ No Reason not completed ___	Is Trial 1 completed? ○ Yes Time to complete (sec) __ ○ No Reason not completed __
Is Trial 2 completed? ○ Yes Time to complete (sec) __ ○ No Reason not completed ___	Is Trial 2 completed? ○ Yes Time to complete (sec) __ ○ No Reason not completed__

Was assistive device used?	○ Yes Assistive device _____ ○ No
Is Trial 1 completed?	○ Yes Time to complete (sec) __ ○ No Reason not completed __
Is Trial 2 completed?	○ Yes Time to complete (sec) ___ ○ No Reason not completed____

From an SDTM perspective, these assessments are typically represented in the Functional Tests (FT) domain. CDISC provides QRS supplements (including controlled terminology and example structures) to support consistent representation of functional tests in SDTM. These resources guide how to populate variables such as FTTESTCD, FTTEST, FTCAT, and related qualifiers. The evaluator is stored in FTEVAL (commonly "INVESTIGATOR"). Trial number is captured in FTREPNUM. For 9HPT, the participant's dominant hand is typically recorded in the Subject Characteristics (SC) domain (e.g., SCTESTCD = "DOMHAND", SCTEST = "Dominant Hand", SCORRES = "RIGHT" or "LEFT"). For T25FW, information on the assistive device used may be stored in SUPPFT. The assistive device is the same across both trials and FTGRPID variable is used to link parent FT records to the corresponding supplemental qualifiers.

ft.xpt

STUDYID	DOMAIN	USUBJID	FTGRPID	FTTESTCD	FTTEST	FTCAT
ABC123	FT	MS01-001		NHPT0101	NHPT01-Time to Complete 9-Hole Peg Test	NHPT
ABC123	FT	MS01-001		NHPT0101	NHPT01-Time to Complete 9-Hole Peg Test	NHPT
ABC123	FT	MS01-001		NHPT0101	NHPT01-Time to Complete 9-Hole Peg Test	NHPT
ABC123	FT	MS01-001		NHPT0101	NHPT01-Time to Complete 9-Hole Peg Test	NHPT
ABC123	FT	MS01-001	1	T25FW101	T25FW1-Time to Complete 25- Foot Walk	T25FW
ABC123	FT	MS01-001	1	T25FW101	T25FW1-Time to Complete 25- Foot Walk	T25FW
FTSCAT	FTORRES	FTORRESU		FTREPNUM	VISIT	FTDTC
DOMINANT HAND	210	sec		1	Screening	2012-09-16
DOMINANT HAND	255	sec		2	Screening	2012-09-16
NON-DOMINANT HAND	280	sec		1	Screening	2012-09-16
NON-DOMINANT HAND	288	sec		2	Screening	2012-09-16
	38	sec		1	Screening	2012-09-16
	45	sec		2	Screening	2012-09-16

Functional tests usually include multiple attempts per visit, while analysis usually relies on a visit-level summary, most commonly the mean time across available trials. At least one trial per test is required to calculate a mean. Some protocols define a maximum allowable time to complete the test (e.g., 3 minutes for T25FW); this threshold may be used to impute values for incomplete attempts, depending on prespecified rules. In addition, if an assistive device differs from baseline, some sponsors may consider the result as not comparable and exclude it from certain analyses. In ADaM, functional test results are stored in a BDS dataset. Depending on sponsor preference and traceability needs, the dataset may include either only the derived visit-level mean, or both trial-level records and a derived mean record.

In some trials, functional tests contribute to composite disability endpoints such as NEPAD. A common component is confirmed 20% progression on 9HPT or T25FW. Confirmed 20% progression is typically defined as an increase of at least 20% from baseline (worsening because higher time = worse performance) that is sustained over a pre-specified confirmation period. From a programming perspective, the derivation logic closely mirrors CDP: assessments are evaluated for a qualifying increase and confirmation within the required time window.

Changes over time in 9HPT and T25FW are often analyzed using repeated-measures mixed models reporting estimated least squares means (LS means). Models often adjust for baseline value, age, baseline EDSS, visit effects, and within-subject correlation. A procedural example is provided in the following section.

PRO: PATIENT REPORT OUTCOMES

A patient-reported outcome (PRO) is defined as "any report of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else." PRO measures are widely used in MS research and span a broad range of symptoms and quality-of-life concepts. CDISC provides general guidance for mapping questionnaires and, for many commonly used instruments, questionnaire-specific implementation guidance. Similar to EDSS, SDTMIG QRS supplements are available for several questionnaires frequently used in MS, including controlled terminology, standard dataset structures with examples, and annotated case report forms aligned with CDISC SDTMIG submission values.

One of the most widely used PRO instruments in MS trials is the Multiple Sclerosis Quality of Life-54 (MSQoL-54) questionnaire. MSQoL-54 is a multidimensional, health-related quality-of-life measure that combines generic and MS-specific content into a single instrument. It contains 54 items and integrates the Short Form-36 (SF-36) with an additional 18 MS-specific items addressing areas such as fatigue and cognitive function.
Common MSQoL-54 endpoints include changes from baseline over time in the physical health composite and mental health composite scores.
For analysis, MSQoL-54 is typically summarized into five scores:

- Physical health composite
- Mental health composite
- Satisfaction with sexual function
- Change in health
- Overall quality of life

These are derived from 12 subscales (Physical function; Role limitations-physical; Role limitations-emotional; Pain; Emotional well-being; Energy/fatigue; Health perceptions; Social function; Cognitive function; Health distress; Overall quality of life; Sexual function) according to the scoring rules specified in the Scoring Manual (Vickrey, 1995). All scores are scaled from 0% to 100%, where 100% indicates the best possible state of health. Because these are derived scores, the derivation algorithm and missing-item rules (minimum item completion thresholds, imputation rules if allowed) must be specified in SAP.

From an ADaM perspective, a BDS dataset is used to store the 12 derived subscale scores and the 5 derived composite scores. Depending on traceability requirements, individual item responses may also be included. In the example below, the first three rows represent individual questions sourced directly from QS, the next three rows represent derived subscales, and the last two rows represent derived composite scores.
admsqol.xpt

STUDYID	USUBJID	AVISIT	AVISITN	ADT	PARAMCD
ABC123	MS01-001	Screening	1	16SEP12	MSQL101
ABC123	MS01-001	Screening	1	16SEP12	MSQL102
ABC123	MS01-001	Screening	1	16SEP12	MSQL103
…					
ABC123	MS01-001	Screening	1	16SEP12	MSQLPH
ABC123	MS01-001	Screening	1	16SEP12	MSQLLP
ABC123	MS01-001	Screening	1	16SEP12	MSQLLE
…					
ABC123	MS01-001	Screening	1	16SEP12	MSQLPHC
ABC123	MS01-001	Screening	1	16SEP12	MSQLMHC
PARAM	PARCAT1	PARCAT2	AVAL	AVALC	ABLFL
MSQL1-Would You Say Your Health Is	MSQoL-54 Questionnaire	Health Perceptions	2	Very good	Y
MSQL1-Health Now Compared to 1 Year Ago	MSQoL-54 Questionnaire	Change in Health	4	Somewhat worse now than one year ago	Y
MSQL1-Health Limit Vigorous Activities	MSQoL-54 Questionnaire	Physical Health	1	Yes, Limited a Lot	Y
…					Y
Physical Health Scale	MSQoL-54 Scales		45		Y
Role Limitations due to Physical Problems Scale	MSQoL-54 Scales		60		Y
Role Limitations due to Emotional Problems Scale	MSQoL-54 Scales		65		Y
…					

Physical Health Composite Score	MSQoL-54 Composite Scores		48		Y
Mental Health Composite Score	MSQoL-54 Composite Scores		55		Y

Changes in MSQoL-54 composite scores are commonly analyzed using estimated least squares means (LS means) from a repeated-measures mixed effects model. Typical covariates include baseline score, age, baseline EDSS, visit, and within-subject correlation. An example implementation using PROC MIXED is shown below.

```
ods output lsmeans=est;
proc mixed data=inds covtest;
class visit(ref='ref_value') bledsscat (ref='ref_value') usubjid;
model chg = base visit age bledsscat /ddfm=KR solution cl residual;
repeated visit /type=un subject=usubjid
lsmeans visit/om=blds at (age)=(&mean_age) e; /* blds is a dataset with baseline
values for all subjects*/
run;
```

MRI: MAGNETIC RESONANACE IMAGING

Relapses and EDSS alone may not fully capture MS disease activity. Magnetic resonance imaging (MRI) is therefore a key tool for assessing inflammatory activity and treatment efficacy in MS clinical trials. Common MRI endpoints include T2 lesion count or volume, active (new or enlarging) T2 lesion count, and T1 lesions (gadolinium-enhancing or hypointense) count or volume. In statistical analyses, these endpoints are frequently treated as measures of on-study MRI activity, particularly through lesion counts that represent new inflammatory activity between scans.

In SDTM, the fact that an MRI procedure was performed is typically recorded in the Procedures (PR) domain. Historically, SDTMIG v3.2 referenced the Morphology (MO) domain for imaging test results; however, MO was later deprecated. The Nervous System Findings (NV) domain has since been introduced, and sponsors may implement MRI results using NV or, in some cases, a sponsor-defined custom findings domain aligned with the Findings observation class. Relationships across SDTM domains are documented in RELREC.

The example below shows selected MRI parameters collected at Month 3 visit. Variables ZMTPTREF and ZMRFTDTC identify the reference timepoint used for interpretation (e.g., Month 3 results compared with the prior scan performed at Screening). In addition, ZMEVAL is included to capture the role of the individual providing the MRI evaluation.

zm.xpt

STUDYID	DOMAIN	USUBJID	ZMTESTCD	ZMTEST	ZMORRES
ABC123	ZM	MS01-001	NUMT1GEL	T1-Gadolinium Enhancing Lesions	2
ABC123	ZM	MS01-001	VOLT1GEL	T1-Gd Enhancing Lesion Volume	1.05
ABC123	ZM	MS01-001	NUMNT2L	New T2 Lesions	1
ABC123	ZM	MS01-001	NUMET2L	Enlarging T2 Lesions	1
ABC123	ZM	MS01-001	BRAINVOL	Brain Volume	1600
ABC123	ZM	MS01-001	BVCP	Percentage Brain Volume Change	-0.87
ZMORRESU	ZMLOC	VISIT	ZMDTC	ZMTPTREF	ZMRFTDTC
	BRAIN	Month 3	2012-09-16	Screening	2012-09-16
cm3	BRAIN	Month 3	2012-09-16	Screening	2012-09-16
	BRAIN	Month 3	2012-09-16	Screening	2012-09-16
	BRAIN	Month 3	2012-09-16	Screening	2012-09-16
cm3	BRAIN	Month 3	2012-09-16	Screening	2012-09-16
%	BRAIN	Month 3	2012-09-16	Screening	2012-09-16

Typical analysis derivations include:

- New or enlarging T2 lesion count: total number of T2 lesions that are new or have enlarged since the previous scan (or since baseline, depending on SAP definition).
- Combined unique active (CUA) lesion count: commonly defined as a combination of T1 Gd+ lesions and new/enlarging T2 lesions, with rules to avoid double counting when a lesion contributes to both categories within the same interval.
- MRI activity (binary): often defined as presence of either T1 Gd+ lesions or new/enlarging T2 lesions within a period. This is frequently used as the MRI component of composite endpoints such as NEDA.
- Brain volume loss / atrophy: commonly assessed using annualized percent brain volume change. Some studies use a threshold (for example, 0.4% annualized loss) as part of an expanded disease-control concept, but the exact threshold and method are study-specific and must follow study definitions.

MRI assessments are usually evaluated via independent central review. Due to scan quality issues, some scans may be classified as non-evaluable, resulting in missing results across multiple parameters. For derived endpoints, prespecified rules are essential; a common approach is that if any required component is missing, the composite or derived parameter is not derived.
From an ADaM perspective, a BDS dataset is typically used to store MRI results from central reading and any derived parameters required for analysis.
MRI analyses can be statistically challenging due to repeated measurements, varying visit schedules, missingness from non-evaluable scans, and distributional properties (e.g., many zeros and over-dispersion for lesion counts). Statistical analysis approaches therefore vary by endpoint type, ranging from Poisson or negative binomial regression for lesion counts to mixed-effects linear models for continuous measures such as lesion volume or brain volume change.

COMPOSITE ENDPOINTS: NEDA

No Evidence of Disease Activity (NEDA), sometimes referred to as freedom from disease activity, is a composite endpoint that integrates clinical and imaging measures of MS activity. NEDA is commonly defined by the absence of:

- Relapse
- Confirmed disability progression (CDP)
- MRI activity

Additional components may be incorporated depending on the study objective. A common extension is brain atrophy, resulting in NEDA-4.

No Evidence of Progression or Active Disease (NEPAD) is an expanded composite measure that builds on NEDA by adding functional disability components. NEPAD is typically defined as meeting NEDA criteria and the absence of:

- 20% progression on 9HPT
- 20% progression on T25FW

NEDA and NEPAD are most often analyzed as binary outcomes (yes/no) over a defined period, often using logistic regression adjusted for clinically relevant prognostic factors (e.g., baseline EDSS, age, sex, prior relapse history, baseline MRI activity, region). NEDA may also be evaluated using time-to-event methods (e.g., time to first loss of NEDA). In SAS, logistic regression can be implemented using PROC LOGISTIC, as illustrated below.

```
ods output estimates=est oddsratios=oddsr;
proc logistic data=inds;
class bledsscat(ref='ref_value') /param=ref;
model neda(event='1')=age bledsscat /link=logit;
estimate "mean" intercept 1 age &mean_age bledsscat&mean_edss1 /e cl ilink;
run;
```

Typical problems in the implementation of NEDA are:

- Definition inconsistency: NEDA/NEPAD definitions vary between sponsors (elements, cutoffs, confirmation periods), which leads to a lack of comparability between studies.
- Sensitivity over time: NEDA may be less informative early in treatment, as medical treatments might require some time to demonstrate effect.
- Impact of missing data: infrequently measured components (especially MRI, sometimes disability confirmation) can drive missing composite status. Rules for handling missingness can have an impact and no standard exists between studies.

COMPOSITE ENDPOINTS: PIRA

Progression Independent of Relapses (PIRA) is intended to capture disability progression that occurs in the absence of acute relapse activity. PIRA is typically defined as a CDP event during a relapse-free period, either because there were no qualified post-baseline relapses or because CDP onset and confirmation are sufficiently distant from a relapse (based on a pre-specified window). Confirmed PIRA at 3 or 6 months is commonly used (aligned to CDP confirmation windows).

In addition to PIRA, trials often distinguish Relapse Worsening Activity (RAW), defined as confirmed disability worsening that does not meet the criteria for PIRA. Confirmed Disability Accumulation (CDA) is then defined as any confirmed PIRA or RAW event.
Several extended composite PIRA definitions are also used. For example:

- Composite PIRA: CDP or confirmed worsening of at least 20% from baseline in 9HPT or at least 20% from baseline in T25FW.
- Modified composite PIRA: composite PIRA or decline in Symbol Digit Modalities Test (SDMT) from baseline.

Analogous composite definitions are often applied for RAW and CDA.
PIRA is commonly analyzed as a binary endpoint (yes/no) and/or as a time-to-event endpoint (e.g., time to first PIRA) using survival analysis methods.

PIRA is tempting in theory, but complicated in practice. Typical issues include:

- Absence of a single accepted definition: studies differ in relapse-free windows, confirmation rules, and eligibility criteria.
- Indirect causality: PIRA is based on timing relative to relapses, not on direct evidence of causation.
- Event recurrence and classification: repeated worsening events and relapse timing can make “in/out of relapse” classification difficult.
- Distance from relapse: required separation can range across studies (often 1-3 months), affecting event counts.
- Baseline selection: analyses may use traditional baseline, re-baselining, or roving baselines.
- Event selection rules: if several PIRA events are present, decide whether we only consider the first one.
- Handling multiple event types: for participants with both PIRA and RAW events, analysis rules must specify whether both are counted or only the first event drives the endpoint.

CONCLUSION

Multiple sclerosis trials combine various outcome types - relapses, disability progression, functional performance, MRI activity, and patient-reported outcomes - into an evidence package that must be both clinically interpretable and technically traceable. For statistical programmers, this creates a practical challenge: each endpoint family comes with its own assessment schedule, missingness patterns, and derivation rules, yet the submission requires a coherent CDISC implementation that supports reproducible analyses and transparent review.

A reliable workflow starts with early alignment on the standards and on sponsor conventions that resolve common ambiguities. In MS specifically, TAUG can accelerate consistent modeling for disease-specific concepts, but it should be treated as a supplement - the foundational standards and the protocol/SAP definitions remain the primary source of truth. Across endpoints, the most frequent sources of downstream discrepancies are not statistical methods themselves, but operational details: CDP confirmation windows, MRI reference scans and evaluability, and how composite endpoints behave when any component is missing.

From an ADaM perspective, MS endpoints benefit from structures that make the unit of analysis explicit (event-level for relapses, repeated measures for EDSS/functional tests/PRO, and interval-based MRI summaries) while preserving traceability to SDTM through robust identifiers and clear metadata. When done well, this design supports both standard analyses and credible sensitivity analyses that address missing data and assessment timing.

Overall, strong MS programming is less about implementing a single “correct” model and more about building an end-to-end, submission-ready chain: precise endpoint definitions, consistent standardization, explicit derivation logic, and transparent traceability. That combination is what turns complex MS data into results that sponsors, regulators, and clinicians can trust.

REFERENCES

ClinicalTial.gov https://clinicaltrials.gov/

Caspar E P van Munster, Bernard M J Uitdehaag, Outcome Measures in Clinical Trials for Multiple Sclerosis, 2017 Feb 9, National Library of Medicine https://pmc.ncbi.nlm.nih.gov/articles/PMC5336539/

Multiple Sclerosis Clinical Trial https://www.sciencedirect.com/topics/neuroscience/multiple-sclerosis-clinical-trial

Bernard M J Uitdehaag, Disability Outcome Measures in Phase III Clinical Trials in Multiple Sclerosis, 2018 Jun 20, , National Library of Medicine https://pmc.ncbi.nlm.nih.gov/articles/PMC6061412/

MS International Federation, Atlas of MS https://atlasofms.org/map/global/epidemiology/number-of-people-with-ms

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Daryna Teteriuk
Company: Intego Group, LLC
Address: 35B Borychiv Tik Street, Kyiv 04070, Ukraine
Work Phone: +380680188025
Email: daryna.teteriuk@intego-group.com
Website: https://integogroup.com