## Common Data Model in Multiple Sclerosis – BMSD

The Big MS Data (BMSD) network includes six MS registries, i.e. the national MS registries of the Czech Republic, Denmark, France, Italy and Sweden as well as the international MSBase. As the participating registries in BMSD all operate as individual registries although with the intention of collaborating on specific projects, there has been a need to develop a common data model (CDM) which is suitable for MS registry data. Such a CDM has been developed within the BMSD network primarily by the data managers of the Czech MS registry. The aims of a CDM have been described (Fig 1).

| | |
|---|---|
| 1 | Provide a uniform understanding of analysed data. |
| 2 | Provide precise descriptions of the data. |
| 3 | Enable federated or merged analyses on data converted to the CDM. |
| 4 | Serve as the basis for project-specific enhancements to the data model. |

Figure 1. Aims of the BMSD CDM

### Introduction

A CDM is a set of rules how to manage data in a harmonized manner. The requirements for different diseases will be quite different and our aim has been to develop a CDM for MS although there are more general models available such as OMOP (Observational Medical Outcomes Partnership). Although OMOP could be an option it was decided that a more MS specific CDM would be more useful for better fitting the specific requirements of MS projects. If there is a need in the future the BMSD CDM could be translated into a standardized structure such as OMOP.

The CDM will provide a way to harmonise and prepare MS data for analysis and will allow for data from different data sources to be analysed in a coordinated way (Fig 2).

The strongest reasons for creating a CDM were the need for transformed data to have a unified, correct interpretation across all involved data sources. Furthermore, data transformation from raw data to CDM format should always be performed by the data managers who understand their data and the focus should be on maintaining a core CDM of key variables that is as straightforward and easy to use as possible across different projects. This should result in a uniform understanding of analyzed data, precise data description and the ability to run federated / merged analysis on data converted to CDM.

Once harmonization is successfully performed, two options of further analysis are available based on the objective of the analysis and eventual legal restrains. The first involves running a meta-analysis with aggregated data. The second approach involves running analysis using a federated way, when one universal script of CDM data format is disseminated across the contributing data sources generating individual data results which are then shared with the investigator (Fig 2).

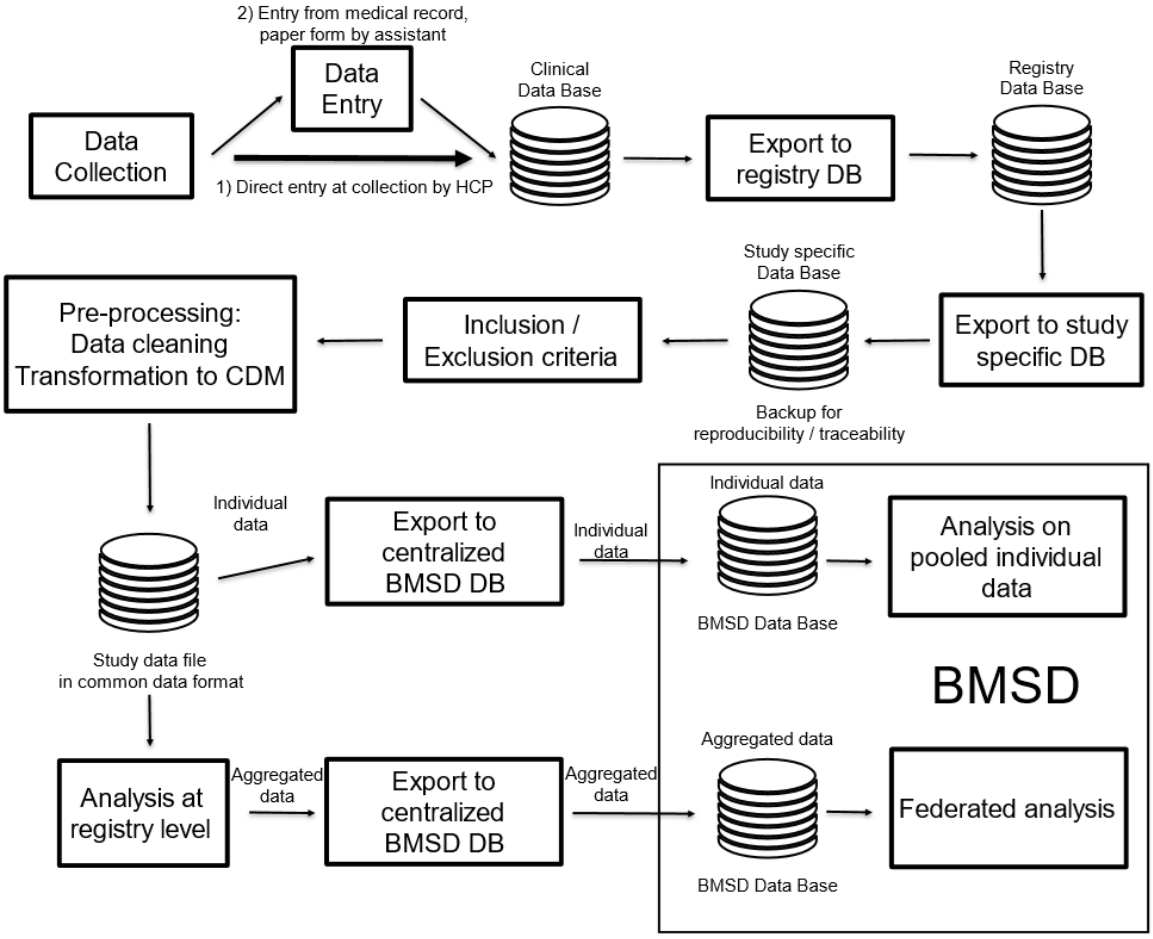



Figure 2. Pooled or federated data analysis principles.

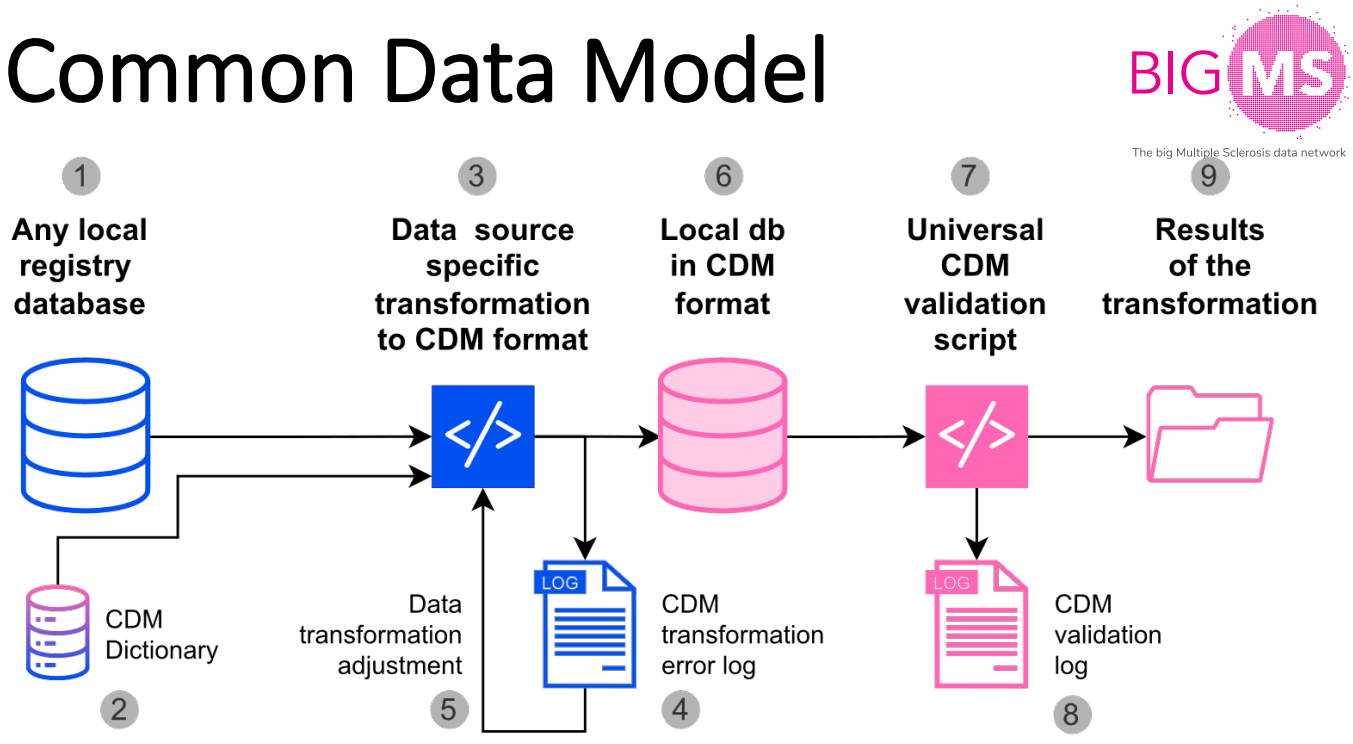
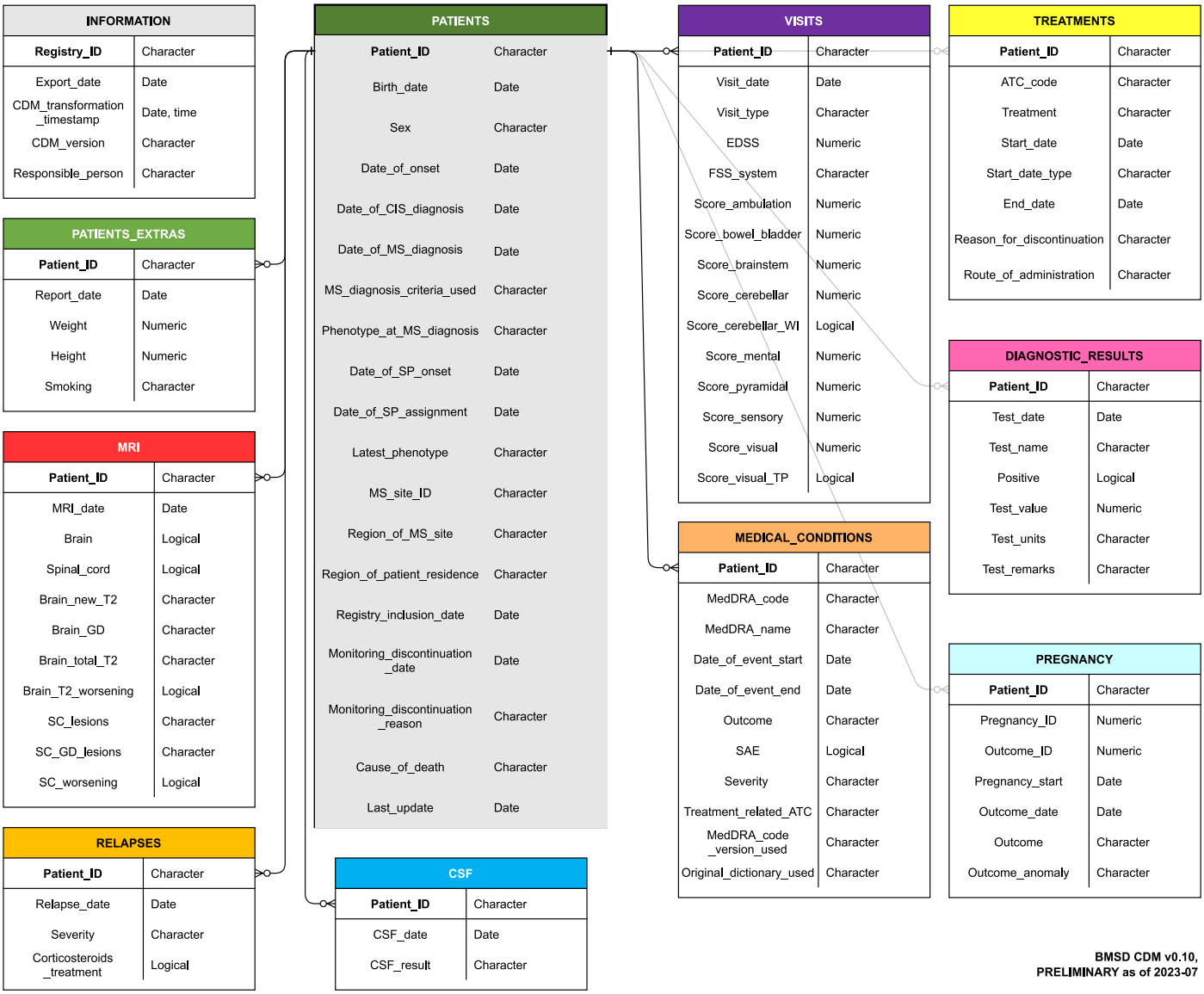



Figure 3. General Common Data Model concept

## BMSD and CDM

One important aim for creating the BMSD CDM was to offer a core CDM which can be used for various projects. The CDM structure includes various data on MS patients as described (Fig 4).




** Every variable "*date*" has additional format under "X_date_chr" in form of a text, where only the most accurate original date is captured: "YYYY" / "YYYY-MM" / "YYYY-MM-DD" / NA*

Figure 4. BMSD CDM structure

## Harmonization variable mapping instructions

The BMSD CDM will specify variables, data type, format / options and the harmonized variable principles. The formatting rules for the variables (e.g., yyyy-mm-dd) and guidelines for harmonization of the variables are shown as tables x-x below. The management of data in the CDM will ensure data being correctly harmonized and analyzed. Tables 1-9 describe the principles of how to harmonize the variables typically included in an MS study.

## Preparation of data for a BMSD study

The BMSD CDM will be used to facilitate studies including data from the BMD registries. To prepare a dataset for studies, data managers for the participating MS registries will run the CDM script which will allow each registry to create and manage their datasets locally.

## Conclusion

Including data from several MS registries into one study presents data management challenges. The use and modification of the BMSD CDM will allow for a harmonized and precis dataset which will provide a good basis for an MS study.

Although the CDM per se will of course also allow pooling of data in a shared manner, the growing complexity regarding data safety would likely favour an approach where datasets can be established, managed and analysed locally for the results to then be shared within a coordinating centre managing the common project.

# Appendices. Tables 1-11

**PATIENTS** *Key characteristics of a patient.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | Any number or alphanumeric identification of patient within respective data source used across the data tables. Initial three letters do reffer to respective data source code to prevent appearence of duplicate ID entries. For the 3 letters registry id prefix, the ISO 3166 standard alpha-3 codes to be used: Czechia (CZE), Denmark (DNK), France (FRA), Italy (ITA), Sweden (SWE) - https://www.iban.com/country-codes. For MSBase, either "MSB" to be used. Table with valid codes on the sheet "Extras". |
| *Birth_date* | Date | YYYY-MM-DD | Date of birth in the most accurate form.* |
| *Sex* | Character | "F", "M", NA | As of now, only sex of male and female are recognized in respective registries. Possible changes of sex and it's recording harmonization to be addressed. |
| *Date_of_onset* | Date | YYYY-MM-DD | Date of the first clinical manifestation of the multiple sclerosis.* |
| *Date_of_CIS_diagnosis* | Date | YYYY-MM-DD | Administrative date of the diagnosis of the clinically isolated syndrome (CIS). If the date is calculated, it shall be first visit on/after the date of fulfilling the CIS definition. *<br><br>E.g.:<br>A) for a patient with the first relaspe in 2018, and the first visit in 2019 and the second relapse + MS diagnosis in 2020, this date is in 2019.<br>B) For a for patient with the first relaspe in 2018, a calculated MS diagnosis also in 2018 and the first visit in 2019, there is no date, set to *NA* .<br>C) For a PPMS patients there is no date, set to *NA* . |
| *Date_of_MS_diagnosis* | Date | YYYY-MM-DD | Administrative date of the diagnosis of the multiple sclerosis. If the date is calculated, it shall be first visit on/after the date of fulfilling the MS diagnostic criteria. * |
| *MS_diagnosis_criteria_used* | Character | NA, "McDonald (2017)", "McDonald (2010)", "McDonald (2005)", "McDonald (2001)", "Poser (1983)" | What diagnostic criteria were used to confirm MS diagnosis. Either calculated or clinically assigned. If there is no explicit record of the criteria used in the registry, date is calculated based on diagnosis date and criteria availability e.g., McDonald (2017) for diagnosis in or after 2017. |
| *Phenotype_at_diagnosis* | Character | "RR", "PP", "SP", NA | Phenotype at the time of MS diagnosis (*Date_of_MS_diagnosis* ). If there is one relapse and dissemination in space and time, the patient is evalued as RRMS.<br><br>If phenotype at MS diagnosis is not recorded directly, it shall be deducted as follows:<br>PP patient is PP from the start.<br>RR patient is RR from the start.<br>SP patient who has information on date of conversion to SP shall be assigned as RR or SP at diagnosis based on date of diagnosis and date of conversion.<br>SP patient with no information on date of conversion to SP, shall be considered as SP at diagnosis.<br>*NA* for all other cases. |
| *Date_of_SP_onset* | Date | YYYY-MM-DD | Information on date since when the patient could had been considered as SP. If only date when clinician has recorded the assignment but not SP onset, this shall be left blank (NA).* |
| *Date_of_SP_assignment* | Date | YYYY-MM-DD | Date since when the patient was assigned/diagnosed as SP.*<br>If not collected directly, shall be calculated as the first visit on/after the *Date_of_SP_onset* . |
| *Latest_phenotype* | Character | "CIS", "RR", "PP", "SP", NA | Latest known phenotype of CIS/MS as of date of the last patient data update *(Last_update)* . |

*Table continues on the next page.*

*Table continues from the previous page.*

**PATIENTS** *Key characteristics of a patient.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *MS_site_ID* | Character | RID_yyy | Any number or alphanumeric identification of MS site within respective data source. Initial three letters (RID) do reffer to respective data source code - please see the definition in the variable *Patient_ID*.<br>yyy refers to any unique MS site alphanumeric identification valid for respective datasource. |
| *Region_of_MS_site* | Character | e.g. "FR-69M" (for Lyon), "DK-84" (for Hovedstaden), "AU-QLD" (for Queensland) | Geographic information (ISO 3166-2 code). |
| *Region_of_patient_residence* | Character | e.g. "FR-69M" (for Lyon), "DK-84" (for Hovedstaden), "AU-QLD" (for Queensland) | Geographic information (ISO 3166-2 code). |
| *Registry_inclusion_date* | Date | YYYY-MM-DD | Date from which patient is being monitored within the registry. Typically, date of provision informed consent (IC) or first monitored visit. |
| *Monitoring_discontinuation_date* | Date | YYYY-MM-DD | If patient is lost to folow-up, the appropriate date shall be also filled in. If such date is unknown explicitely, shall be calculated as a date of the last available dated record (visit, laboratory test, etc.).<br><br>If the last patient's record is dated more than 3 years prior date of registry extraction, patient shall be considered as lost to follow-up as of the date of such last record. |
| *Monitoring_discontinuation_reason* | Character | "IC withdrawal", "emigration", "death", "patients' decision", "stable condition", "lost to follow-up", "other", NA | Reason why patient is no longer followed in the registry. If the reason is unknown, but patient is no longer followed, the reason shall be "unknown". All followed patients at the time of data extraction shall have this variable set empty as not available (NA). |
| *Cause_of_death* | Character | | MedDRA code of death cause if known. If death is confirmed (set in variable *Monitoring_discontinuation_reason*), but reason is not known, value shall be set to "unknown". In all other cases shall be left blank. |
| *Last_update* | Date | YYYY-MM-DD | Date of last update of patient's information. |

**PATIENTS_EXTRAS** *Additional information on patient that shall be valid as of specific timepoint.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *Report_date* | Character | YYYY-MM-DD | Date of following records update. the appropriate date shall be also filled in. |
| *Weight* | Numeric | | Weight in kg as of *Report_date*. |
| *Height* | Numeric | | Height in centimeters as of *Report_date*. |
| *Smoking* | Character | "never", "current", "former", NA | Smoking status of the patient as of *Report_date*. If status is unknown, value shall be set to NA. |

Table 1. Patients

**VISITS** *Harmonized information on visits of a patient.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *Visit_date* | Date | YYYY-MM-DD | * |
| *Visit_type* | Character | "virtual", "physical", NA | Type of visit. Virtual does include any yet performed form of contact of the clinician and patient other than physical (telephone, video-call, etc.). |
| *EDSS* | Numeric | Values from 0 to 10, by 0.5 steps. Excluding value 0.5 ; NA (there are also visits without an EDSS evaluation) | Observed value on Expanded Disability Status Scale (EDSS) as of Visit_date. [1] |
| *FSS_system* | Character | "Kurtzke", "Neurostatus", "unknown" | Principles for **Kurtzke** shown below in the attached table. For **Neurostatus**, please see: https://neurostatus.net/media/specimen/Definitions_0410-2_s.pdf |
| *Score_ambulation* | Numeric | 0 - 12 for Kurtzke<br>0 - 15 for Neurostatus | |
| *Score_bowel_bladder* | Numeric | 0 - 5 for Kurtzke<br>0 - 6 for Neurostatus | |
| *Score_brainstem* | Numeric | 0 - 5 for Kurtzke<br>0 - 5 for Neurostatus | |
| *Score_cerebellar* | Numeric | 0 - 5 for Kurtzke<br>0 - 5 for Neurostatus | |
| *Score_cerebellar_WI* | Logical | TRUE, FALSE, NA | TRUE - Used when weakness (grade 3 or more on pyramidal) interferes with testing |
| ***Score_mental*** | **Numeric** | 0 - Normal<br>1 - Mood alteration only (does not affect DSS score)<br>2 - Mild decrease in mentation<br>3 - Moderate decrease in mentation<br>4 - Marked decrease in mentation (chronic brain syndrome-moderate)<br>5 - Dementia or chronic brain syndrome-severe or incompetent | In Neurostatus, this score is part of the *Score_cerebellar* |
| *Score_pyramidal* | Numeric | 0 - 6 for Kurtzke<br>0 - 6 for Neurostatus | |
| *Score_sensory* | Numeric | 0 - 6 for Kurtzke<br>0 - 6 for Neurostatus | |
| *Score_visual* | Numeric | 0 - 6 for Kurtzke<br>0 - 6 for Neurostatus | |
| *Score_visual_TP* | Logical | TRUE, FALSE, NA | TRUE - in presence of temporal pallor |

Reference: [1] Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale. Neurology 1983; 33: 1444-1452

| KURTZKE FSS SCALE INTERPRETATION | | |
|---|---|---|
| *Score_ambulation* | Numeric | 0 - Unrestricted<br>1 - Fully ambulatory<br>2 - ≥ 300 meters, but < 500 meters, without help or assistance (EDSS 4.5 or 5.0)<br>3 - ≥ 200 meters, but < 300 meters, without help or assistance (EDSS 5.0)<br>4 - ≥ 100 meters, but < 200 meters, without help or assistance (EDSS 5.5)<br>5 - walking range < 100 meters without assistance (EDSS 6.0)<br>6 - unilateral assistance, ≥ 50 meters (EDSS 6.0)<br>7 - bilateral assistance, ≥ 120 meters (EDSS 6.0)<br>8 - unilateral assistance, < 50 meters (EDSS 6.5)<br>9 - bilateral assistance, ≥ 5 meters, but < 120 meters (EDSS 6.5)<br>10 - Uses wheelchair without help; unable to walk 5 meters even with aid, essentially restricted to wheelchair; wheels self and transfers alone; up and about in wheelchair some 12 hours a day (EDSS 7.0)<br>11 - Uses wheelchair with help; unable to take more than a few steps; restricted to wheelchair; may need some help in transferring and in wheeling self (EDSS 7.5)<br>12 - essentially restricted to bed or chair or perambulated in wheelchair, but out of bed most of day; retains many self-care functions; generally has effective use of arms (EDSS 8.0) |
| *Score_bowel_bladder* | Numeric | 0 - Normal<br>1 - Mild urinary hesitancy, urgency or retention<br>2 - Moderate hesitancy, urgency, retention of bladder or bowel, or rare urinary incontinence<br>3 - Frequent urinary incontinence<br>4 - In need of almost complete constant catheterisation<br>5 - Loss of bladder function |
| *Score_brainstem* | Numeric | 0 - Normal<br>1 - Signs only<br>2 - Moderate nystagmus or other disability<br>3 - Severe nystagmus, marked extraocular weakness, or moderate disability of other cranial nerves<br>4 - Marked dysarthria or other marked disability<br>5 - Inability to swallow or speak |
| *Score_cerebellar* | Numeric | 0 - Normal<br>1 - Abnormal signs without disability<br>2 - Mild ataxia<br>3 - Moderate truncal or limb ataxia<br>4 - Severe ataxia, all limbs<br>5 - Unable to perform coordinated movements due to ataxia |
| *Score_cerebellar_WI* | Logical | TRUE, FALSE, NA<br>*TRUE - Used when weakness (grade 3 or more on pyramidal) interferes with testing* |

*Table continues on the next page.*

*Table continues from the previous page.*

| KURTZKE FSS SCALE INTERPRETATION | | |
|---|---|---|
| *Score_mental* | Numeric | 0 - Normal<br>1 - Mood alteration only (does not affect DSS score)<br>2 - Mild decrease in mentation<br>3 - Moderate decrease in mentation<br>4 - Marked decrease in mentation (chronic brain syndrome-moderate)<br>5 - Dementia or chronic brain syndrome-severe or incompetent |
| *Score_pyramidal* | Numeric | 0 - Normal<br>1 - Abnormal signs without disability<br>2 - Minimal disability<br>3 - Mild or moderate paraparesis or hemiparesis<br>4 - Marked paraparesis or hemiparesis; moderate quadriparesis; or monoplegia<br>5 - Paraplegia, hemiplegia, or marked tetraparesis<br>6 - Quadriplegia |
| *Score_sensory* | Numeric | 0 - Normal<br>1 - Vibration or figure-writing decrease only in one or two limbs<br>2 - Mild decrease in touch or pain or position sense, and/or moderate decrease in vibration in one or two limbs; or vibratory (c/s figure-writing) decrease alone in three or four limbs<br>3 - Moderate decrease in touch or pain or position sense, and/or essentially lost vibration in one or two limbs; or mild decrease in touch or pain and/or moderate decrease in all proprioceptive tests in three or four limbs<br>4 - Marked decrease in touch or pain or loss of proprioception, alone or combined, in one or two limbs; or moderate decrease in touch or pain and/or severe proprioceptive decrease in more than two limbs<br>5 - Loss (essentially) of sensation in one or two limbs; or moderate decrease in touch or pain and/or loss of proprioceptive for most of the body below the head<br>6 - Sensation essentially lost below the head |
| *Score_visual* | Numeric | 0 - Normal<br>1 - Scotoma with visual acuity (corrected) better than 20/30<br>2 - Worse eye with scotoma with maximal visual acuity (corrected) of 20/30 to 20/59<br>3 - Worse eye with large scotoma, or moderate decrease in fields, but with maximal visual acuity (corrected) of 20-60 to 20-99<br>4 - Worse eye with marked decrease in fields and maximal visual acuity (corrected) of 20/100-20/200; grade 3 plus maximal acuity of better eye 20/60 or less<br>5 - Worse eye with maximal visual acuity (corrected) less than 20/200; grade 4 plus maximal acuity of better eye of 20/60 or less<br>6 - Grade 5 plus maximal acuity of better eye of 20/60 or less |
| *Score_visual_TP* | Logical | TRUE, FALSE, NA<br>*TRUE - in presence of temporal pallor* |

Table 2. Visits including EDSS

**RELAPSES** *Harmonized information on relapses of a patient.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *Relapse_date* | Date | YYYY-MM-DD | Exact or imputed date of a relapse. * |
| *Severity* | Character | "mild", "moderate", "severe", "life threatening", NA | Severity of relapse. Due to heterogeneous severity assessment rules in different registries, this variable can only be used **for intraregistry analysis.** |
| *Corticosteroids_treatment* | Logical | TRUE, FALSE, NA | *Corticosteroids_treatment* shall be considered as TRUE if any corticosteroids were used to treat the relapse and FALSE if not (and is stated so in primary records).<br><br>If no direct record on corticosteroid treatment of relapse is available, but there is a record on usage of corticosteroids within 30 days from the *Relapse_date*, this variable shall be set to TRUE.<br><br>In all other cases *NA* to be set. |

Table 3. Relapses

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *MedDRA_code* | Numeric | | Lower-level term (LLT) in MedDRA clasification to be entered. If the information is originally coded in any other dictionary (e.g., ICD-9, ICD-10, SNOMED CT), it has to be converted through appropriate mapping. If conversion is not possible from the original terminiology to MedDRA, *NA* shall be set. In the core BMSD CDM mapping, the dictionaries provided by MedDRA MSSO are to be used. |
| *MedDRA_name* | Character | | Verbal English representation of the MedDRA code in *MedDRA_code* variable, as per *MedDRA_code_version_used* . |
| *Date_of_event_start* | Date | YYYY-MM-DD | Date of onset of the medical condition.* |
| *Date_of_event_end* | Date | YYYY-MM-DD | Date of resolution of the medical condition. If condition is considered as ongoing, this variable shall be left blank and *Outcome* set as "ongoing". If the outcome or current status is unknown, both *Date_of_event_end* and *Outcome* , shall be left blank (set as *NA* ). * |
| *Outcome* | Character | NA, “recovery”, “ongoing”, “death” | Information on the medical condition outcome. Possible values are "recovery", "ongoing", "death", and *NA* (if the outcome is unknown). |
| *SAE* | Logical | TRUE, FALSE, NA | Logical information if medical condition has fulfilled criteria of serious adverse event: results in death; is life-threatening; requires inpatient hospitalisation or prolongation of existing hospitalisation; results in persistent or significant disability/incapacity, or is a congenital anomaly/birth defect. *(ICH Topic E 2 A Clinical Safety Data Management: Definitions and Standards for Expedited Reporting)* |
| *Severity* | Character | NA, “mild”, “moderate”, “severe”, “life threatening” | If not collected, the variable is left blank *(NA)* . Due to heterogenic classification of severity across datasources, this variable shall be used rather for intraregistry analysis. |
| *Treatment_related_ATC* | Character | NA, specific ATC, or “not related to drug” | In case the medical condition is suspected to be in relation to any specific drug, please indicate here. If no treatment is suspected “not related to drug” shall be set. In case data is not collected or it is unknown, set to NA. |
| *MedDRA_code_version_used* | Character | e.g., "MedDRA 26.0", "MedDRA 25.1" ,... | Meta information on the classification used to create variables MedDRA_code and MedDRA_name. |
| *Original_dictionary_used* | Character | e.g., "MedDRA 26.0", "ICD-10 v2019", "ICD-10" | Meta information on the classification used to code original information in raw database. |
| *Original_medical_condition* | Character | optional | Any original information as it is available in the data source - ICD code, free text, etc. In the original language. |

Table 4. Medical conditions

**CSF** *Harmonized information on examination of cerebrospinal fluid in a patient.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *CSF_date* | Date | YYYY-MM-DD | Exact or imputed date of CSF examination.* |
| *CSF_result* | Character | "positive", "negative", NA | Specification whether CSF examination has supported diagnosis of MS. Therefore e.g., presence of OCB, results in "positive" value. |

Table 5. CSF

**TREATMENTS** *Harmonized information on treatments of a patient. Any treatment, including DMT shall appear.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *ATC_code* | Character | e.g., “L04AA34”, "J07BN01",... | ATC code as per WHO nomenclature. |
| *Treatment* | Character | e.g., “alemtuzumab”, "covid-19, RNA-based vaccine",... | ATC code full name interpretation. Unless specified otherwise. |
| *Start_date* | Date | YYYY-MM-DD | Date of the first administration of the treatment. If only date of prescription is known, shall be used as *Start_date* . * |
| *Start_date_type* | Character | NA, "administration", "prescription" | Information on how the start date was acquired and what does it represent - initial administration of the drug or date of prescription, if administration can not be obtained. |
| *End_date* | Date | YYYY-MM-DD | If treatment had been discontinued, a date of the last administration of the treatment shall be entered, otherwise left blank.<br><br>In case of pulse treatments i.e., alemtuzumab, ocrelizumab, rituximab or cladribine, the most exact information available shall be recorded (either each pulse or ongoing manner with first and last dose recorded). If last pulse is not known, shall be left as NA. In specific projects, this definition can be adjusted. |
| *Reason_for_discontinuation* | Character | “adverse event”, “disease activity”, “antibodies for drug”, ”JCV positivity ”, pregnancy”, “stable condition”, “patient choice”, “other”, “unknown”, “switch of the route”, NA | “adverse event” includes allergic reaction, side effects, local intolerance, general intolerance, biological intolerance, SAE;<br>“disease activity” includes EDSS progression, MRI activity, relapse activity, inefficacy;<br>“JCV positivity“ refers to John Cunningham virus positivity;<br>“pregnancy” includes desire for or planning of pregnancy, and ongoing pregnancies;<br>“patient choice” includes lack of compliance or any adherence issues on side of a patient, or patient's choice due to inconvenience;<br>“other” includes scheduled stop, discontinuation due to necessity of treating other concomitant disease;<br>“unknown” shall refer to such situation when treatment is discontinued, but there is no reason in database;<br>in case treatment is not discontinued NA shall be set;<br>if multiple reasons are reported, they shall be separated with tilda (~). |
| *Route_of_administration* | Character | “i.v.”, “s.c.”, “oral”, “i.m.”, “other”, NA | Information on route of administration of respective treatment. If the route changes, new record shall be made as previous treatment shall be discontinued as of the last date of administration and the new one initiated as of the first date of administration. |

Table 6. Treatments

**MRI** *Harmonized information on magnetic resonance imagining.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *MRI_date* | Date | YYYY-MM-DD | Date of MRI examination. * |
| *Brain* | Logical | TRUE, FALSE, NA | Does the MRI involve scan of brain structures? |
| *Spinal_cord* | Logical | TRUE, FALSE, NA | Does the MRI involve spinal cord? |
| ***REMOVED*** *Gadolinium_administered* | *Logical* | *TRUE, FALSE, NA* | *Was gadolinium contrast administered?* |
| *Brain_new_T2* | Character | "0", ">=1", "1", "2", "3", … "9", "10-20", ">20", "Unknown", NA | Lesion count to be stratified by the count into groups of "0", ">=1", "1", "2", "3", … "9", "10-20", ">20", "Unknown", NA. In case only granularity YES/NO is available, value ">=1" for "yes" and "0" for "no" is applicable. NA to be reported if information is not collected. |
| *Brain_GD* | Character | "0", ">=1", "1", "2", "3", … "9", "10-20", ">20", "Unknown", NA | Lesion count to be stratified by the count into groups of "0", ">=1", "1", "2", "3", … "9", "10-20", ">20", "Unknown", NA. In case only granularity YES/NO is available, value ">=1" for "yes" and "0" for "no" is applicable. NA to be reported if information is not collected. |
| *Brain_total_T2* | Character | "0", ">=1", "1", "2", "3", … "9", ">9", "10-20", ">20", "Unknown", NA | Lesion count to be stratified by the count into groups of "0", ">=1", "1", "2", "3", … "9", ">=9", "10-20", ">20", "Unknown", NA. In case only granularity YES/NO is available, value ">=1" for "yes" and "0" for "no" is applicable. NA to be reported if information is not collected. |
| *Brain_T2_worsening* | Logical | TRUE, FALSE, NA | Does the MRI confirm existence of new or enlarging T2 lesions? |
| *SC_new_lesions* | Character | "0", "1", ">=1", ">=2", NA | As there is variability in data sources in primary structure, allowed values are "0" if the record contains FALSE, NO or zero value; "1" if the record contains value of 1; ">=1" if the record contains TRUE, YES or nonspecific value with number of lesions, or ">=2" if there is confirmation of higher number of SC lesions. |
| *SC_GD_lesions* | Character | "0", ">=1", "1", "2", "3", … "9", "10-20", ">20", "Unknown", NA | Lesion count to be stratified by the count into groups of 1, 2, 3, … 9, "10-20", ">20", "Unknown", NA. In case only granularity YES/NO is available, value ">=1" for "yes" and "0" for "no" is applicable. |
| *SC_worsening* | Logical | TRUE, FALSE, NA | Does the MRI confirm existence of new or enlarging SC lesions? |

Table 7. MRI

**DIAGNOSTIC_RESULTS** *Harmonized information on laboratory or clinical examination.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *Test_date* | Date | YYYY-MM-DD | Exact or imputed date of test.* |
| *Test_name* | Character | i.e., "25FWT", "JCV", "absolute lymphocyte count", "CRP","Covid-19 PCR", "Covid-19 Ag",… | Name of the test performed. Based on project or study, this list shall be extended to needs of such study. |
| *Positive* | Logical | TRUE, FALSE, NA | Indication, wheter test had positive result or not, where applicable. |
| *Test_value* | Numeric | | Numeric value used in the respective test. If any units are used, they have to be indicated in *Test_units* variable. |
| *Test_units* | Character | i.e., "s", "elements/mm3",… | If numeric value is provided, used units have to be presented here. |
| *Test_remarks* | Character | | If any special test remarks are made, this variable can accomodate them. |

Table 8. Diagnostic results

**PREGNANCY** *Harmonized information on pregnancy.*

| Variable | Data type | Format / options | Harmonized variable principles |
|---|---|---|---|
| *Patient_ID* | Character | RID_xxx | *Patient_ID* to relate record in PATIENTS table. |
| *Pregnancy_ID* | Numeric | 1, 2,… | ID of pregnancy in patient. |
| *Outcome_ID* | Numeric | 1,2,… | If more than one fetus or child is present in pregnancy, this variable allows to identify each of them. |
| *Pregnancy_start* | Date | YYYY-MM-DD | Date of pregnancy start. If only last menstruation period is known, then this date is used.* |
| *Outcome_date* | Date | YYYY-MM-DD | Date of outcome - giving birth, abortion or miscarriage. |
| *Outcome* | Character | "Live Birth", "Stillbirth/Fetal Death", "Termination of Pregnancy for Fetal Anomaly (TOPFA)", "Miscarriage/Spontaneous Abortion", "Late Fetal Death", NA | Outcome of pregnancy, resp. specific fetus of the pregnancy.<br>**Live Births**: This is when a baby is born alive, regardless of gestational age.<br>**Stillbirths/Fetal Deaths**: These are pregnancies that result in the birth of a child with no signs of life at or after 20 weeks' gestation.<br>**Terminations of Pregnancy for Fetal Anomaly (TOPFA)**: These are medical procedures performed to end a pregnancy due to detected fetal anomalies.<br>**Miscarriages/Spontaneous Abortions**: These are pregnancies that end spontaneously before 20 weeks' gestation.<br>**Late Fetal Deaths**: This refers to pregnancies where the fetus dies in the womb after 20 weeks' gestation, but before birth, which is not induced. |
| *Outcome_anomaly* | Character | i.e., "Q00.1", "Q90",… If more than one anomaly is present in the child, all shall be filled and separated with pipe character "\|". | If any anomaly is present, then EUROCAT classification is used and present here. Entered is only ICD10 code (as of now Q00-Q99) |

Table 9. Pregnancy

**Registry identification code (RID)**

| RID | Country | Registry |
|---|---|---|
| CZE | Czechia | ReMuS |
| DNK | Denmark | DMSR |
| FRA | France | OFSEP |
| ITA | Italy | RISM |
| SWE | Sweden | SMSreg |
| MSB | International | MSBase |

ISO3166 to be used
https://www.iban.com/country-codes

| Variable | Data type | Harmonized variable principles |
|---|---|---|
| *Information_source* | Character | If information sources shall be distinguished, researcher can use this variable in any table of the CDM for each of the rows. |

**Licences and sources**

ISO_3166-2_Locals

This works is licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).
It is free for personal or commercial use with attribution required by mentioning the use of this works as follows:
This site or product includes IP2Location™ ISO 3166-2 Subdivision Code which available from https://www.ip2location.com.

ISO_3166-2_Countries

https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv

CZE ZIP codes

https://www.ceskaposta.cz/ke-stazeni/zakaznicke-vystupy#1
https://www.ceskaposta.cz/documents/10180/3738087/csv_cobce_psc.zip/45623265-227b-5b1c-d5d6-5c842b3425d3

Table 10. Licences and sources

| Brand_name | ATC_code | Route | Indication | Type | Active_substance |
|---|---|---|---|---|---|
| AUBAGIO | L04AA31 | oral | MS | DMT | teriflunomid |
| AVONEX | L03AB07 | i.m. | MS | DMT | interferon beta-1a |
| BETAFERON | L03AB08 | s.c. | MS | DMT | interferon beta-1b |
| COPAXONE[20] | L03AX13 | s.c. | MS | DMT | glatiramer acetate |
| COPAXONE[40] | L03AX13 | s.c. | MS | DMT | glatiramer acetate |
| EXTAVIA | L03AB08 | s.c. | MS | DMT | interferon beta-1b |
| RIXATHON | L01XC02 | i.v. | MS | DMT | rituximab |
| TRUXIMA | L01XC02 | i.v. | MS, NMOSD | DMT | rituximab |
| GILENYA | L04AA27 | oral | MS | DMT | fingolimod |
| KESIMPTA | L01FA02 | s.c. | MS | DMT | ofatumumab |
| LEMTRADA | L04AA34 | i.v. | MS | DMT | alemtuzumab |
| MABTHERA | L01XC02 | i.v. | MS, NMOSD | DMT | rituximab |
| MAVENCLAD | L04AA40 | oral | MS | DMT | cladribin |
| MAYZENT | L04AA42 | oral | MS SP | DMT | siponimod |
| OCREVUS | L04AA36 | i.v. | MS PP, NMOSD | DMT | ocrelizumab |
| PLEGRIDY i.m. | L03AB13 | i.m. | MS | DMT | peginterferon-beta |
| PLEGRIDY s.c. | L03AB13 | s.c. | MS | DMT | peginterferon-beta |
| PONVORY | L04AA50 | oral | MS | DMT | ponesimod |
| REBIF[22] | L03AB07 | s.c. | MS | DMT | interferon beta-1a |
| REBIF[44] | L03AB07 | s.c. | MS | DMT | interferon beta-1a |
| TECFIDERA | L04AX07 | oral | MS | DMT | dimethyl fumarate |
| TYSABRI | L04AA23 | i.v. | MS | DMT | natalizumab |
| TYSABRI s.c. | L04AA23 | s.c. | MS | DMT | natalizumab |
| ZEPOSIA | L04AA38 | oral | MS | DMT | ozanimod |
| ZINBRYTA | L04AC01 | | MS | DMT | daclizumab |
| ENDOBULIN | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| FLEBOGAMMA | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| GAMMAGARD | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| KIOVIG | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| OCTAGAM | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| PRIVIGEN | J06BA02 | i.v. | MS | IVIG | normal human, for intravascular adm. |
| FAMPYRA | N07XX07 | oral | MS | Symptomatic | fampridine |
| ENSPRYNG | L04AC19 | s.c. | NMOSD | DMT | satralizumab |
| ROACTEMRA | L04AC07 | s.c. | NMOSD | DMT | tocilizumab |
| SOLIRIS | L04AA25 | i.v. | NMOSD | DMT | ekulizumab |
| UPLIZNA | L04AA47 | i.v. | NMOSD | DMT | inebilizumab |

Table 11. ATC codes – to be updated periodically