Global Health Economics & Outcomes Research

Non-Interventional Study Protocol

Study Protocol Number

Predicting disease activity for patients with Relapsing Remitting Multiple Sclerosis using Electronic Medical Records

|  |  |
| --- | --- |
| Authors: | Rigg John, Medin Jennie |
| Document type: | Non-interventional study protocol |
| Version number: | 01 |
| PASS Study | No |
| Release date: | 06-Jun-2016 |

Property of Novartis  
Confidential  
May not be used, divulged, published or otherwise disclosed  
without the consent of Novartis  
**NIS Protocol Template Secondary Use of Data Version 2.0 dated 15-Sep-2015**

Marketing authorization holder(s)

|  |  |
| --- | --- |
| **Marketing authorization holder(s)** | Not applicable |
| **MAH contact person** | Not applicable |

Table of contents

Table of contents 3

List of figures 5

List of tables 5

List of abbreviations 6

1 Responsible parties 8

2 Abstract 9

3 Rationale and background 12

4 Research question and objectives 12

5 Research methods 13

5.1 Study design 13

5.2 Setting 14

5.2.1 Index date 14

5.2.2 Follow up period 14

5.2.3 Look back period 14

5.2.4 Inclusion criteria 15

5.3 Variables 15

5.3.1 Baseline characteristics 15

5.3.2 Outcomes of interest 20

5.4 Data sources 21

5.4.1 Data collection schedule 21

5.5 Study size/ power calculation 21

5.6 Data management 21

5.7 Data analysis 21

5.7.1 Logistic regression 22

5.7.2 Variable importance 22

5.7.3 Performance measures 22

5.7.4 Cross-validation resampling 23

5.8 Secondary objectives 24

5.9 Quality control 25

5.10 Limitations of the research methods 25

5.11 Other aspects 26

6 Protection of human subjects 26

6.1 Regulatory and ethical compliance 26

7 Management and reporting of adverse events/adverse reactions 26

8 Plans of disseminating and communicating study results 26

9 References 27

10 Annexes 29

10.1 Annex 1 – List of drugs of interest 29

10.2 Annex 2 – Exploratory objectives 31

List of figures

List of tables

List of abbreviations

|  |  |
| --- | --- |
| AUC | Area Under the (Receiver Operating Characteristic) Curve |
| CDST | Clinical Decision Support Tool |
| DMT | Disease Modifying Therapy |
| EDSS | Expanded Disability Scale Score |
| EMR | Electronic Medical Record |
| FDA | Food and Drug Administration |
| GPP | Guidelines for Good Pharmacoepidemiology Practices |
| HIPAA | Health Insurance Portability and Accountability Act |
| ICD-9-CM | International Classification of Diseases |
| ICMJE | International Committee of Medical Journal Editors |
| IMS | Intercontinental Marketing Services |
| IRB | Institutional Review Board |
| ISPE | International Society for Pharmacoepidemiology |
| MRI | Magnetic Resonance Imaging |
| MS | Multiple Sclerosis |
| NTD | NeuroTransData |
| RRMS | Relapse Remitting Multiple Sclerosis |
| SAS | Statistical Analysis System |
| SVM | Support Vector Machines |
| STROBE | Strengthening the Reporting of Observational Studies in Epidemiology |
| US | United States |

# Responsible parties

Not applicable

# Abstract

**Title**

Predicting disease activity for patients with Relapsing Remitting Multiple Sclerosis using Electronic Medical Records

**Version and date**

Version 01

**Name and affiliation of main author**

John Rigg. PhD, Real-World Evidence Solutions, IMS Health Inc.; Jennie Medin, MPH, PhD, Novartis Pharma AG.

**Rationale and background**

The relapse remitting multiple sclerosis (RRMS) treatment landscape has undergone considerable changes since the emergence of interferon-β (Betaseron, Rebif, Avonex, Extavia) and glatiramer acetate (Copaxone) in the mid-1990s (these two product classes will herein be referred to as BRACE therapies). The increased availability of disease modifying therapies (DMTs) for the treatment of RRMS has increased the complexity of optimizing treatment choices for practicing neurologists. It is important to understand whether historical patient data can be used to predict future disease activity for RRMS patients and whether this can be used in clinical practice to improve patient outcomes.

**Research question and objectives**

This is a retrospective study to estimate future disease activity and progression for patients with given baseline characteristics based on real world data.

The primary objective of this study is to estimate the overall probability of relapse over a twelve-month period.

The secondary objectives of this analysis are to produce predictions of disease activity and progression over a twelve-month period as defined by a number of outcome measures. These outcome measures are: probability of EDSS progression, probability of confirmed EDSS progression, probability of relapse or EDSS progression, probability of relapse or confirmed EDSS progression, and probability of relapse and EDSS progression. Additional secondary objectives will estimate all of the above outcome measures under a number of different treatment scenarios, e.g. predictions for outcomes will be made for cohorts switching between pre-defined treatments.

Further, exploratory objectives produce predictions using advanced machine learning methods to compare these with baseline estimates using penalized multiple logistic regressions.

**Study design**

To address the objectives of this study, a retrospective cohort analysis will be conducted using electronic medical record (EMR) data extracted from the NeuroTransData (NTD) network of neurology practices in Germany. Data will be extracted for RRMS patients receiving BRACE therapy and will be divided into **five cohorts**:

1. patients continuing on BRACE therapy
2. patients switching to alternative BRACE therapy,
3. patients switching to first-line oral therapy,
4. patients switching/ escalating to second-line therapy and
5. a composite cohort combining the above cohorts (and randomly removing duplicates in the case of patients appearing in more than one of the above cohorts).

Separate analysis will be conducted on each of the cohorts to estimate overall disease activity/ progression (using the composite cohort) and disease activity/ progression under the different treatment scenarios.

Characteristics of patients will be assessed at the **index date** which, depending on the latest available data is likely to fall between 1st January 2010 and 30th March 2015 【**2010.1.1， 2015.3.30】**. The index date is defined as the date of switch/ escalation or in the BRACE continuation cohort as the date of the most recent visit that will **allow for 12 months of post-index data**. **A twelve-month post-index period is required and outcomes will be assessed over these twelve months**. The analysis will model six alternative, **binary outcome** measures:

1. relapse,
2. Expanded Disability Status Scale (EDSS) progression,
3. confirmed EDSS progression,
4. relapse or EDSS progression,
5. relapse or confirmed EDSS progression
6. and relapse and EDSS progression (see “Data analysis” section for full definition of variables).

It is not possible to know *a priori* which outcome may be best explained by the attributes available in the data.

**Population**

The population of the study will consist of RRMS patients receiving BRACE treatment from 1st January 2010 to the most recent date available in the **specific cut of data** **used**（**？2016？**）.

1. Patients without 360 days follow up post-index will be excluded.
2. Patients without an EDSS score within 360 days before index will also be excluded.

Patients included in the BRACE continuation cohort and the composite cohort are required to have received BRACE treatment for at least three months prior to the index date (i.e. the fourth month was the earlier possible index date).

**Variables**

Candidate predictors will include information on demographics and clinical characteristics. All count and continuous variables will be transformed into Booleans to capture possible non-linear associations between covariates and outcomes and to facilitate ready interpretability of model coefficients.

**Data sources**

NeuroTransData (NTD) database

**Study size**

Over 4000 patients meet the primary inclusion/ exclusion criteria

**Data analysis**

All continuous and count predictors will be dichotomised to capture possible non-linear associations between covariates and outcomes. The cut-off between adjacent categories will be informed from analysis of the distribution of the underlying variables along with acceptable thresholds used in clinical practice.

A logistic regression model with the elastic-net penalty will be used to estimate the probability of each of the six outcomes of interest, i.e., whether a patient experiences relapse, whether a patient experiences EDSS progression, whether a patient experiences confirmed EDSS progression, whether a patient experiences relapse or EDSS progression, whether a patient experiences relapse or confirmed EDSS progression and whether a patient experiences relapse and EDSS progression in the post-index period.

All outcome variables will be assessed in the one year post-index period. For relapse, the variable will be equal to 1 for patients who experience at least one relapse in the 12 months post-index period and 0 otherwise. EDSS progression will be defined with reference to a given patients baseline EDSS score. EDSS will be defined to capture the non-linear relationship between EDSS gain and baseline EDSS, with lower baseline EDSS requiring greater increases in EDSS to be considered progression. The following is an indicative definition of EDSS progression as used in previous Novartis studies: an increase in EDSS score greater than 1.5 when baseline EDSS is 0, an increase greater than 1 for a baseline EDSS between 1.5 and 5.5 (inclusive) or an increase greater than or equal to 0.5 for EDSS greater than 5.5. This definition may change in the main study upon further consultation of the data and clinical guidance. The confirmed EDSS progression outcome is a stronger requirement (than EDSS progression) and is equal to 1 when the following conditions are satisfied: there exists an assessment indicating EDSS progression in the post-index period and there exists another assessment at least 76 days after the first in which EDSS progression is also present (as defined above). Confirmed EDSS progression also requires no conflicting EDSS assessments between the two assessments indicating EDSS progression. Note that EDSS assessments occurring during relapses will not be considered as an outcome measure. The ‘either relapse or EDSS progression’ outcome variable will be equal to 1 if at least one outcome measure, i.e., relapse and EDSS progression, is 1 and 0 if neither are equal to 1. Similarly, the ‘either relapse or confirmed EDSS progression’ outcome variable will be equal to 1 if at least one outcome measure (i.e., relapse and confirmed EDSS progression) is equal to 1 and 0 if neither is equal to 1. The relapse and EDSS progression variable will be equal to 1 if both EDSS progression is equal to 1 and relapse is equal to 1, and 0 otherwise.

Multivariate logistic regression will be used to provide an estimate of the outcomes defined above for each patient, controlling for baseline characteristics. Logistic regression is selected since it is appropriate for binary classification, whilst retaining model transparency (compared to non-parametric alternatives, for example). The elastic-net penalty shrinks model parameters to increase generalisability while keeping the selective power among collinear variables. It is particularly useful where over-fitting (Type I errors) may be expected, which is a potential concern of this study given the relatively high number of candidate predictors in relation to the number of patients with a positive outcome for many of the cohorts. Variable importance will be computed to facilitate an understanding of the attributes most relevant to predicting future disease activity. Variable importance will be computed through both quantifying the loss in model accuracy when each variable is omitted and by the magnitude of the odds ratio.

To facilitate an unbiased assessment of model performance, ten-fold cross-evaluation and ten-fold cross-validation will be used through two nested loops. For the outer loops, all the available data in a given cohort will be randomly stratified into ten equal-sized evaluation folds and ten evaluation rounds will be conducted, with each round using one fold for evaluation and the remaining nine folds for training the model and selecting hyper-parameters. The latter, i.e., model-training and hyper-parameter-selection, will be conducted through the inner loop. In the inner loop, the nine evaluation folds will be further randomly stratified into ten equal-sized "validation" folds. Ten validation inner rounds will be carried out, with each round of training using nine validation folds and validating the trained model using the tenth validation fold for the purpose of hyper-parameter selection. This is standard practice in statistical learning theory, see Hastie et al 2009

Model accuracy will be assessed using the Area Under the Curve (AUC) as well as a plot of actual outcomes by quintile of the predicted probability. All performance metrics will be computed on the evaluation folds since this provides a good approximation of out-of-sample model accuracy (Hastie, Tibshirani & Friedman, 2013).

**Milestones**

Not applicable

# Rationale and background

MS is a chronic, demyelinating, immune-mediated disease of the central nervous system characterized by inflammation and destruction of the myelin sheath covering of nerve fibers in the brain and the spinal cord. In the United States (US), one study has cited the overall incidence as 3.6 cases per 100,000 person-years in women and 2.0 in men (Alonso & Hernan, 2008). According to the National Multiple Sclerosis Society the US prevalence is estimated to be 400,000 cases, with worldwide prevalence estimated to be 2.1 million. RRMS is the most common type of MS, affecting approximately 80–85% of all patients with MS (Bainbridge & Riekmann, 2008) and is characterized by unpredictable acute attacks (known as relapses) accompanied by worsening of symptoms, followed by periods of remission during which there is a full or partial recovery from the deficits acquired during the relapse.

While there is no cure for MS there are a number of DMTs aimed at preventing and treating relapses, preventing new attacks, managing symptoms, slowing disease progression and preventing or postponing long-term disability (Halpern et al, 2011). Interferons (Betaferon®, Avonex®, Rebif®, Extavia®) and glatiramer acetate (Copaxone®) have been available since the mid-1990s and are widely used DMTs for the treatment of MS and are known jointly as the BRACE therapies.

Since 2005 a number of new DMTs with different mechanisms of action, efficacy and safety profiles have been approved. DMTs may be grouped into three broad categories according to EU guidelines: first-line BRACE therapies, first-line oral therapies (Aubagio and Tecfidera) and second-line therapies (Gilenya and Tysabri). Physicians must now consider a broad range of factors when choosing among and between these treatment classes. Consequently, choosing the optimal treatment option for a given patient has become increasingly complex.

The goal of this exploratory study is to evaluate disease activity and progression associated with treatment decisions for a cohort of RRMS patients taken from German EMR data. The current study will use historical patient data to predict disease activity and disease progression with the intention of eventually integrating this analysis in clinical practice to improve patient outcomes. Specifically, the intention is to develop a model capable of predicting disease activity and disease progression, under different treatment scenarios, with sufficient precision to inform treatment decisions. The ultimate goal is to build a web-based tool combining these predictions, and confidence intervals, with visuals that are intuitive to both patients and physicians.

If this exploratory study demonstrates the feasibility of using historic EMR data to predict future disease activity then the results will be used in the development of the web-based clinical decision support tool (CDST).

# Research question and objectives

The aim of this study is to predict disease activity and disease progression for patients currently on BRACE therapy based on demographic and clinical attributes under different treatment scenarios, e.g. patients switching from existing BRACE therapy to a different therapy.

Specifically, the primary objective of this study is to:

1. Estimate the overall probability of a patient experiencing a relapse in the next twelve-month period given baseline characteristics

The secondary objectives of this study are to produce estimates of disease activity and disease progression for the following outcome measures (please see section 5.1 Study design for further detail):

1. Probability of EDSS progression in next year
2. Probability of confirmed EDSS progression in next year
3. Probability of relapse or EDSS progression in next year
4. Probability of relapse of confirmed EDSS progression in next year
5. Probability of relapse and EDSS progression in next year

Additional secondary objectives will seek to estimate outcomes 1 – 6 for the following treatment scenarios: BRACE continuation, BRACE switch, first-line oral switch, second-line escalation (please see section 5.1 Study design for more detail)

# Research methods

## Study design

This study will use EMR data on RRMS patients starting on BRACE treatment extracted from the NTD network of neurologists in Germany. In order to address the study objectives 5 cohorts will be analysed:

1. BRACE continuation - patients continuing on BRACE therapy;
2. BRACE switch - patients switching from BRACE therapy to new BRACE therapy;
3. First-line oral switch - patients switching from BRACE therapy to first-line oral therapy (Aubagio and Tecfidera);
4. Second-line escalation - patients switching from BRACE therapy to second-line therapy (Gilenya and Tysabri);
5. Composite - a composite cohort of all of the above cohorts.

Note that for the BRACE switch and composite cohorts it is possible in theory for a patient to appear more than once. Patients will be allowed to have multiple records across cohorts, however within each cohort one record will be chosen at random for every patient. Sensitivity analyses will be performed by altering the seed by which records are chosen.

Each of these five cohorts will be examined over six different outcome measures (relapse, EDSS progression, confirmed EDSS progression, relapse or EDSS progression, relapse or confirmed EDSS progression and relapse and EDSS progression)

EDSS progression is defined as a change in EDSS score over time, relative to baseline EDSS. The exact definition of EDSS progression will be chosen to reflect the non-linear relationship between EDSS gain and baseline EDSS and will be decided upon consultation of the data and clinical input. As an indicative definition of EDSS progression, the following has been used in a number of Novartis RRMS studies: for baseline EDSS of 0, progression is indicated by an increase of ≥1.5. For baseline EDSS of ≥1 and <5.5, progression is indicated by an increase in EDSS ≥1. For baseline EDSS of ≥5.5, progression is indicated by an increase in EDSS ≥0.5

Confirmed EDSS progression is defined as an initial assessment recording EDSS progression followed by a subsequent assessment at least 76 days after this in which EDSS is measured and found to corroborate the initial assessment, it is also required that there are no other assessments in the intervening period that do not indicate EDSS progression. The choice of 76 days as the space between assessments is chosen to allow three months between assessments, with a buffer in case patients visit sooner. The choice of 76 days is in accordance with Novartis guidelines used in previous studies using EDSS as an outcome measure.

Though specific definitions of relapses in MS vary considerably there are a number of generally accepted criteria. In the absence of objective diagnostic criteria used in clinical practice the McDonald criteria is used as a standard criteria. The McDonald criteria define a relapse as “patient-reported or objectively observed events typical of an acute inflammatory demyelinating event in the CNS [central nervous system], current or historical, with duration of at least 24 hours, in the absence of fever or infection” (Polman, Reingold & Banwell, 2010). Relapses recorded in NTD data require a combination of patient and physician reporting; physicians are trained to enter a relapse start and end date in a standardized way on the basis of patient-reported symptoms.

## Setting

The study will use the most recent cut of NTD data available, the current cut is from 01 Jan 10to 30th June 2015, though it is envisaged that more recent data may be available at the time of study execution.

### Index date

The index date will differ depending on the cohort being examined. In the BRACE switch, oral switch and escalation cohorts the index date will be the date of initiation of new therapy.

In the case of the BRACE continuation cohort the index date is the date of the most recent visit that allows for a 360 day post-index period.

### Follow up period

Patients will be examined over a 360 day period post index date

### Look back period

Patient history will be considered until 01 Jan 10 or as far back as individual patient records will support

### Inclusion criteria

To be selected into the study population, patients must meet the following criteria:

* Confirmed diagnosis of RRMS;
* At least 12 months of follow-up data post index-date;
* A baseline EDSS score

Additionally, patients in the BRACE continuation cohort must meet the following criteria:

* Prescription of BRACE therapy between 01 Jan 10to 30 Jun 14 (or to the latest date as dictated by the cut of data used)

Patients in the BRACE switch, oral switch and escalation cohorts must also meet the following criteria:

* Prescription for BRACE therapy in the 360 days prior to the index date

When assessing patients for EDSS-related outcomes, patients must also meet the following criteria:

* At least one (confirmed) EDSS progression assessment in the 1 year post-index period

## Variables

Where binary variables listed below are based on underlying count or continuous variables, the precise thresholds will be selected through assessment of the distribution of the underlying variable in addition to considerations of clinical practice. The precise thresholds chosen below are indicative

### Baseline characteristics

All variables will be binary. Unless otherwise stated, **baseline values refer to information from the first visit prior to baseline**

**Gender**

* Male

**Age at index date**

* <30 years
* ≥30 and <40 years
* ≥40 and <50 years
* ≥50 years

**Region of birth**

* Central Europe
* Northern Europe
* Southern Europe
* Elsewhere
* Missing

**Years since RRMS diagnosis as of index date**

* ≥5 years prior to index date
* ≥2 and <5 years prior to index date
* <2 years prior to index date

**Relapse history**

* At least 1 relapse in the 90 days prior to the index date
* At least 1 relapse in the 91-180 days prior to index date
* At least 1 relapse in the 181-360 days prior to index date
* At least 1 relapse in the 361-720 days prior to index date
* At least 1 relapse in the 721-1080 days prior to index date
* At least 1 relapse in the 1081-1440 days prior to index date

**Baseline EDSS**

* EDSS = 0
* EDSS = 1
* EDSS = 1.5 or 2
* 2.5 ≤ EDSS ≤ 10

Patients with EDSS scores above 2 are grouped together since it is expected that the number of patients with higher EDSS scores is low and that it is therefore less meaningful to form multiple categories.

**Pre-baseline EDSS score (to be dichotimised into binary categories similar to baseline EDSS above)**

* Latest EDSS score in 361-720 days prior to the index date (0-10)
* Latest EDSS score in 721-1080 days prior to the index date (0-10)
* Latest EDSS score in 1081-1440 days prior to the index date (0-10)

**EDSS sub-scores for 8 channels (in different periods: baseline and 1, 2 and 3 years prior to the index date)**

* Functional system pyramidal (0-10)
* Functional system cerebellar (0-10)
* Functional system brainstem (0-10)
* Functional system sensory (0-10)
* Functional system bowel and bladder (0-10)
* Functional system visual function (0-10)
* Functional system mental (0-10)
* Maximum walking distance (0-10)

**EDSS progression**

* EDSS progression from baseline (1 year prior to index) to the index date. The exact definition of EDSS progression will be decided upon consultation of the data and input from expect clinicians and will seek to capture the non-linear relationship between baseline EDSS and changes in EDSS. An example of EDSS progression used in previous Novartis studies is as follows: for baseline EDSS of 0, progression is indicated by an increase of ≥1.5. For baseline EDSS of ≥1 and <5.5, progression is indicated by an increase in EDSS ≥1. For baseline EDSS of ≥5.5, progression is indicated by an increase in EDSS ≥0.5. This definition is intended to be indicative and may change in the final study

**Pre-baseline EDSS progression (in different periods: 1, 2 and 3 years prior to the index date)**

* There is pre-baseline EDSS progression in the corresponding year
* There is no pre-baseline EDSS progression in the corresponding year

**Pre-baseline confirmed EDSS progression (in different periods: 1, 2 and 3 years prior to the index date)**

* There is confirmed pre-baseline EDSS progression (in the corresponding year)
* There is no confirmed pre-baseline EDSS progression (in the corresponding year)

**Baseline number of cranial lesions**

* ≤8
* >8
* Missing

**Baseline number of spinal lesions**

* ≤2
* >2
* Missing

**Initial RRMS symptom**

* Ataxia
* Optic neuritis
* Paresis
* Sensibility disorder
* Other
* Missing

**Whether or not the patient has familial RRMS**

**Number of biological children**

**Objective severity reported (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Number of Relapses with Physician-Reported Severity "Mild"
* Number of Relapses with Physician-Reported Severity "Medium"
* Number of Relapses with Physician-Reported Severity "Difficult"

**Subjective severity reported (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Number of Relapses with Subject-Reported Severity "Mild"
* Number of Relapses with Subject-Reported Severity "Medium"
* Number of Relapses with Subject-Reported Severity "Difficult"

**Number of relapses treated (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

**Number of relapses not treated (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

**Visit reasons (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Whether or not There is Any Visit Reason Recorded as "Complications"
* Whether or not There is Any Visit Reason Recorded as "Follow-Up"
* Whether or not There is Any Visit Reason Recorded as "Progressive Course"
* Whether or not There is Any Visit Reason Recorded as "Acute Relapse"
* Whether or not There is Any Visit Reason Recorded as "Medication Associated Symptoms"
* Whether or not There is Any Visit Reason Recorded as "Patient's Wishes"

**Degree of disability recorded (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Maximum Qualitative Degree of Disability Recorded
* Maximum Degree of Disability Recorded

**Whether or not the drug on a list of interest has been used (in different periods: in 1, 2, 3 and 4 years prior to the index date; please refer to Appendix A for the list of 42 drugs of interest)**

* Each drug of interest will be a binary variable

**Cerebrospinal fluid examination (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Whether or not There is Any Cerebrospinal Fluid IgG Production
* Maximum Cerebrospinal Fluid Cell Count
* Whether or not There is Any Cerebrospinal Fluid Oligoclonal Bands
* Whether or not There is Any Cerebrospinal Fluid Total Protein Recorded as "Increased"

**Medication-related diagnostics (in different periods: in 1, 2, 3 and 4 years prior to the index date)**

* Whether or not There is Any Cerebrospinal Fluid Pathological
* Whether or not There is Any John Cunningham Virus Titer Positive
* Whether or not There is Any L-Secretin Positive
* Whether or not There is Any Neutralizing Antibody Positive
* Whether or not There is Any Varicella Zoster Virus Titers Positive
* Whether or not There is Any Spec. AC Against Natalizumab Positive

**Date on which the patient enters the data**

**Duration on treatment prior to index date**

* Index-specific BRACE treatment for <=365 days prior to index date (BRACE continuation cohort only)
* Index-specific BRACE treatment for >365 days prior to index date (BRACE continuation cohort only)
* Whether the patient received any BRACE treatment prior to index date for <=365 days (switch / escalation cohorts only)
* Whether the patient received any BRACE treatment prior to index date for >365 days (switch / escalation cohorts only)

**Changes in treatment landscape**

In order to capture the temporal difference in the association between outcomes and attributes that may arise following the introduction of a new therapy, we will consider the following variables:

* Gilenya available at the index date (17 Mar 11)
* Aubagio available at the index date (30 Aug 13)
* Tecfidera available at the index date (30 Jan 14)

**History of disease modifying therapies (A DMT is one specific type of BRACE, first line or second line treatment)**

* Number of different DMTs used in 1 year before the index (to be broken into categories based on the distribution of the variable and the number of patients in the corresponding cohort)
* Number of different DMTs used in 1-2 years before the index (to be broken into categories based on the distribution of the variable and the number of patients in the corresponding cohort)
* Number of different DMTs used in 2-3 years before the index (to be broken into categories based on the distribution of the variable and the number of patients in the corresponding cohort)
* Number of different DMTs used in 3-4 years before the index (to be broken into categories based on the distribution of the variable and the number of patients in the corresponding cohort)

**Composite group flag (a set of binary flags used in the composite cohort indicating which sub-cohort)**

* Whether or not this patient belongs to the BRACE to BRACE switch sub-cohort
* Whether or not this patient belongs to the BRACE to first-line switch sub-cohort
* Whether or not this patient belongs to the BRACE to second-line switch sub-cohort

**Interaction terms for the composite cohort (i.e. interaction terms for the composite cohort will be created between subcohort membership and ‘important’ covariates, where the ‘important variables are the most important variables in the separate, subcohort models)**

### Outcomes of interest

All outcomes are measured over the post-index period:

* At least one relapse;
* EDSS progression;
* Confirmed EDSS progression;
* Relapse or EDSS progression;
* Relapse or confirmed EDSS progression.
* Relapse and EDSS progression

### Missing Data Handling

* Variables with a high proportion of missing values (e.g. >50%; TBD) will not be included in the model.
* Variables with a proportion of missing values below this threshold and above a lower threshold (e.g. 2%; TBD) will be represented by a separate binary flag in the model (note all variables, including count and continuous, will be dichotomized, hence missing values will be a separate category).
* Variables with a proportion of missing values below the lower threshold will receive the median value of non missing observations

## Data sources

The data for this study is retrieved from the NTD database provided by NTD Network Germany. The NTD network operates databases for different neurological and psychiatric disorders for the purpose of quality management and health research. The data of the participating neurology practices are pooled anonymously to form the database. This allows observational cohort-studies using health data routinely collected in outpatient neurology practices of NTD throughout Germany who are members of the NTD network. The exploratory study will use the most recent data from the NTD network.

MS diagnoses and type of MS are recorded as International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes; other data recorded include MS-specific medication, EDSS, relapse rate, adverse events of MS medication, date and results of magnetic resonance imaging, different quality of life-scales, and other socioeconomic parameters.

Demographic and clinical data of MS patients are documented in a standardized way digitally in-time during clinical visits at least once within 3 month periods in all patients with MS in the participating practices. All neurologists are trained to document these data in a standardized way in the digital data source and are certified EDSS-raters. This data acquisition protocol is approved by the ethical committee of the Bavarian Medical Board (Bayerische Landesärztekammer, 14.06.2012).

### Data collection schedule

Not applicable. Primary data collection will not be incorporated as secondary data sources will be utilized.

## Study size/ power calculation

All available patients in the NTD database who fulfil the relevant inclusion/ exclusion criteria will be included.

The following information relates to power calculations. The primary objective is to predict relapse progression for the composite cohort which will be measured using AUC. The approximate size of the cohort will be 4,000 patients in total and 800 with the positive outcome. Based on this information, the confidence intervals for selected AUCs are shown below based on the formula:

Where , , and

And the following is the 95% estimated confidence intervals for possible AUC values:

|  |  |
| --- | --- |
| AUC | 95% confidence intervals |
| 0.6 | [0.577, 0.623] |
| 0.65 | [0.628, 0.672] |
| 0.7 | [0.678, 0.722] |
| 0.75 | [0.729, 0.771] |
| 0.8 | [0.781, 0.819] |

## Data management

All analyses will employ SAS version 9.1 (SAS Institute Inc., Cary, NC), R version 3.2.1, Python 3.5 and PySpark 1.5.2.

## Data analysis

Logistic regression will be performed on each of the six outcomes of disease activity (relapse, EDSS progression, confirmed EDSS progression, relapse or EDSS progression, relapse or confirmed EDSS progression and relapse and EDSS progression) to provide patient level predictions. The variables included in the final predictive models will be ranked and performance statistics of model accuracy will be presented. These analyses will be performed for each of the five cohorts (BRACE continuation, BRACE switch, oral switch, escalation and composite) where cross-validation re-sampling methods will be used to optimise the models and compute out-of-sample model performance.

### Logistic regression

Logistic regression is a well-established and highly effective binary classification model. However logistic regression without a penalty is prone to over fit (i.e. produce a relatively high number of Type I errors) in situations where the number of positive outcomes in the model estimation process is limited (this is especially true for all cohorts apart from the BRACE continuation and combined cohorts). For this reason logistic regression with a penalty will be used. The elastic-net will be selected as the form of the penalty (as opposed to LASSO or ridge penalty) as this typically produces more stable results where covariates are co-linear (i.e. multi-co linearity is present) (Zou et al, 2005). It is anticipated that co-linearity may be relatively high within these samples on account of the fairly low degrees of freedom in the models (even for analysis using the relapse outcome) and the fact that many of the candidate covariates are known to be correlated (e.g. relapses and EDSS progression).

The maximum-likelihood cost function of logistic regression with elastic-net regularisation can be described using the following formula



where *n* is the total number of training data, *Xi* and *yi* are the predictor vector and label of the ith patient record, respectively, *w* is the coefficient vector of the predictors, *c* is the intercept, and *α* and *λ* are the hyper-parameters weighting the total regularisation strength and relative strengths between the LASSO () and ridge () terms, respectively.

### Variable importance

Variable importance will be calculated using two methods. The first method will use the average magnitude of every variable’s coefficient to indicate the importance, i.e., the higher the magnitude, the more important the variable. The average magnitudes will be calculated across coefficients obtained from all models trained in cross-validation (described in Section 9.7.4). The second method will use the performance (measured in AUC) decrease introduced by excluding an independent variable to indicate that variable’s importance. The larger the decrease is, the more important the corresponding variable.

### Performance measures

The key measure of model performance will be the AUC for the graph of false positive vs. true positive rate for each model prediction of outcome (Fawcett, 2006).

To compute the confidence intervals of the AUCs, the method proposed by Delong et al based on the Mann-Whitney U test will be adopted (DeLong et al, 2008). Computing confidence intervals for coefficients obtained from regularised regressions is an active area of research.(Lockhart et al 2006) The challenge arises since the distribution of parameters for the null hypothesis is unknown (the conventional normality assumption is violated). It is proposed to derive confidence intervals from a standard logistic regression, where the variables included are only those variables with non-zero coefficients kept by the elastic-net regression (for a similar approach, see Halabi et al, 2013) (Halabi et al, 2014).

Thus, computing the confidence intervals of individual predictions involves two steps. First, the most important subset of all predictors are selected (e.g., the most important ten predictors based on the average variable ranking as described above). Second, the selected predictors are then used to fit a logistic regression without regularisation. Due to the small number of the selected predictors, this logistic regression is unlikely to over fit in the absence of regularisation. With the classical (unconstrained) model estimated, the confidence intervals of the predictor coefficients can be computed by assuming that the noise in logit belongs to a Gaussian distribution. The confidence interval for the prediction for each patient can then be computed using the confidence intervals of the predictor coefficients (the model parameters) together with the values of the predictors for each patient.

### Cross-validation resampling

At each step that splits patients into folds/samples, stratified random sampling will be applied to ensure that the same proportion of positive outcomes (i.e. actual positives for the outcome of disease activity) are in each fold.

For each of the five outcomes of disease activity, two nested loops will be applied within each of the five cohorts to compute the average AUC and average variable rank. The steps are described below

1. Randomly stratify each cohort into ten evaluation folds / samples of equal size.
2. Assign one evaluation fold as the test set and the remaining nine evaluation folds as the training + validation set.
3. Further randomly stratify the training + validation set into ten validation folds of equal size.
4. For every combination of possible values for hyper-parameters α and λ:
   1. Assign one validation fold as the hyper-parameter validation set and the remaining nine validation folds as the hyper-parameter training set;
   2. Estimate a logistic regression with an elastic-net penalty using the hyper-parameter training set (i.e., the nine out of ten validation folds);
   3. Compute the AUC for the hyper-parameter validation set (i.e., the tenth left-out validation fold);
   4. Repeat steps 4 a-c for all ten validation folds;
   5. Compute the average AUC for the ten validation folds (i.e., across all ten validation folds);
   6. Repeat steps 4 a–e for all possible hyper-parameter combinations.
5. Select the hyper-parameter combination that produces the highest average AUC.
6. Use the selected hyper-parameter combination to estimate a logistic regression with an elastic-net penalty on the training + validation set (i.e., the nine evaluation folds).
7. Compute the AUC for the tenth evaluation fold using the fitted model in step 6,
8. Compute the importance ranking of variables for the tenth evaluation fold using the fitted model in step 6,
9. Repeat steps 2-8 for all ten evaluation folds.
10. Compute the average AUC and variable importance (i.e. across all ten evaluation folds).
11. Select a pre-defined number of most important variables based on the results obtained in step 10.
12. Compute a logistic regression without regularisation using the selected most important variables.

The evaluation of model performance will not use a true test set as the sample size is not sufficient to have separate training and test samples. The proposed cross-evaluation re-sampling schema (involving an inner and outer loop) is known to provide a good approximation of out-of-sample accuracy (Hastie, Tibshirani & Friedman, 2013)

## Quality control

At the study level, all aspects of the study from protocol development to the reporting of the results are conducted within the work-frame of IMS Quality Management System (QMS) and in accordance with the following policies and procedures:

* POL\_QA\_001 “Quality Management System” policy
* POL\_QC\_001 “Quality Control Strategy” policy
* SOP\_QC\_002 “Quality Control of Project Deliverables”

According to the policies and procedures above, a Quality Control plan for the study will be developed and executed, which will include quality control on study methodology, statistical analysis plan, programming, data management and analysis, study results, conclusions and study report. Furthermore:

* The study Quality Control plan will establish ownership for the execution of the individual Quality Control steps. The principle of the independence of Quality Control applies
* The Principal in Charge of the study will ensure that individuals responsible for the execution of specific Quality Control steps will have knowledge, capability and experience which are adequate for the task.
* The result of the execution of the individual steps of the Quality Control plan will be documented, and include the required corrective actions, if any.
* The execution of any required corrective action will be documented.
* The executed Quality Control plan will be subjected to a final review and approval for sufficiency and completeness from the Principal in Charge of the study

Also, the principal in charge of the study will verify training compliance of IMS employees contributing to the study, as per IMS procedure SOP\_QA\_007 “Training of Quality and Operational Standards”.

## Limitations of the research methods

This observational retrospective database study will not be able to infer causality.

Date of first MS diagnosis cannot be established within this data as the patient’s complete medical history is not available.

Some variables, such as the MRI T2 lesion numbers, are missing in a substantial part of the data. This could limit the modelling accuracy.

The final limitation concerns the way EMR data captures treatment data. In particular the EMR data is able to capture information on prescribing and dispensing of DMTs however there is no information on whether and with what regularity DMTs are taken by patients.

## Other aspects

Not applicable

# Protection of human subjects

The proposed study is a non-interventional, retrospective database analysis of data collected via routine patient care. As such, study subjects will not be placed at risk as a result of inclusion in the study sample. Additionally, the information within the database is de-identified to comply with the Health Insurance Portability and Accountability Act (HIPAA) regulations regarding the protection of human subjects

## Regulatory and ethical compliance

The secondary data source proposed for the analyses meets all HIPAA compliance standards, insuring patient anonymity. As such, approval from an IRB (Institutional Review Board) is not applicable.

Compliance with Novartis and regulatory standards provides assurance that the rights, safety, and well-being of patients participating in non-interventional studies are protected (consistent with the principles that have their origin in the Declaration of Helsinki) and that the study data are credible and responsibly reported.

This study was designed and shall be implemented and reported in accordance with the Guidelines for Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology (ISPE, 2008), the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Vandenbroucke et al, 2007), and with the ethical principles laid down in the Declaration of Helsinki.

# Management and reporting of adverse events/adverse reactions

As this is a retrospective study based on anonymised secondary data sources, safety monitoring and safety reporting on an individual case level is not applicable.

# Plans of disseminating and communicating study results

Upon study completion and finalisation of the study report, the results of this non-interventional study may be submitted for publication. Publications will comply with internal Novartis standards and the International Committee of Medical Journal Editors (ICMJE) guidelines.

# References

1. Alonso A, Hernan MA. Temporal trends in the incidence of multiple sclerosis. Neurology 2008; 71(2):129-35
2. National Multiple Sclerosis Society – Epidemiology of MS. Accessed 02/23/2012. Available at: <http://www.nationalmssociety.org/about-multiple-sclerosis/what-we-know-about-ms/who-gets-ms/epidemiology-of-ms/index.aspx>
3. Halpern R, Agarwal S, Dembek C, Borton L, Lopez-Breshnahan M. Comparison of adherence and persistence among multiple sclerosis patients treated with disease-modifying therapies: a retrospective administrative claims analysis. Patient Prefer Adherence 2011; 5: 73-84
4. Bainbridge JL, Rieckmann P, editors (2008) Multiple sclerosis. In: DiPiro JT, Talbert R, Yee G, Matzke G, editors. Pharmacotherapy: a pathophysiologic approach. : New York: McGraw-Hill Medical. 913-926 p.
5. Ingwersen J, Aktas O, Hartung,HP. Advanced in and Algorithms for the Treatment of Relapse-Remitting Multiple Sclerosis. Neurotherapeutics 2016 Jan;13(1):47-57
6. Oleen-Burkey M, Castelli-Haley J, Lage MJ, Johnson KP. Burden of a multiple sclerosis relapse: the patient’s perspective. Patient 2012; 5(1): 57-69
7. Tsang BK and Macdonell R. Multiple sclerosis- diagnosis, management and prognosis. Australian Family Physician, 40 (12), 2011: 948–55.
8. Novartis Gains FDA Approval for Gilenya Multiple-Sclerosis Drug. Bloomberg, September 22, 2010
9. Pill to treat multiple sclerosis receives approval of FDA. New York Times, September 22, 2010
10. Menge T, Weber MS, Hemmer B, Kieseier BC, von Budingen HC, Warnke C et al. Disease-modifying agents for multiple sclerosis: recent advances and future prospects. Drugs 2008; 68(17): 2445-68
11. Polman CH, Reingold SC, Banwell B. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Ann Neurol*.* 2011;69:292–302. et al.
12. Ross A, Halper J, Harris N. Assessing Relapses and Response to Relapse Treatment in Patients with Multiple Sclerosis. International Journal of MS Care: 2012 Fall; 14(3): 148–159
13. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society: Series A. 1996; 58:1.
14. Guidelines for Good Pharmacoepidemiology Practices (GPP) of the International Society for Pharmacoepidemiology. Accessed 9/12/13. Available at:

<http://pharmacoepi.org/resources/guidelines_08027.cfm>

1. Fawcett T. An Introduction to ROC analysis. Pattern Recognit Lett. 2006; 27:8.
2. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67(2):301–320, 2005.
3. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845, 1988.
4. Lockhart R, Taylor J and Tibshirani,R. A significance test for the Lasso. Annals of Statistics. 2014;42(2):413-468
5. Halabi S, Lin C, Kelly WK, et al. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. Journal of Clinical Oncology. 2014; 32 (7): 671-677.
6. Hastie T, Tibshirani R, Friedman J. 2013. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Ed. Springer
7. Cortes C, Vapnik VK. Support-Vector Networks. Machine Learning, 1995; 20: 273 – 297
8. Pan SJ, Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering; vol. 22, no. 10, 2010, pp: 1345-1359
9. Hastie T, Tibshirani RJ. Generalized Additive Models, 1990, Chapman & Hall/ CRC. ISBN 978-0-412-34390-2
10. Hanley J, McNeil B. The Meaning and Use of the Area under a ReceiverOperating Characteristic (ROC) Curve.Radiology, 1982

# Annexes

## Annex 1 – List of drugs of interest

1. Alemtuzumab
2. Azathioprine
3. Baclofen
4. Benzodiazepine
5. Bladder medication
6. Cladribine
7. Cortisone relapse treatment
8. Fampyra
9. Immunoglobulins
10. Laquinimod
11. Mitoxantrone
12. Modafinil
13. NASSR
14. NDRI
15. NRI
16. Neuroleptics
17. Ocrelizumab
18. Physiotherapy
19. Plasmapheresis series
20. Psychotherapy
21. Rituximab
22. SNARI
23. SNRI
24. SSRI
25. Sativex
26. Siponimod
27. Speech therapy
28. Stem cell transplantation
29. Tetracyclic antidepressants
30. Tizanidine
31. Tolpiseron
32. Tricyclic antidepressant
33. anti-dementia medication
34. botulinum toxin
35. cyclophosphamide
36. dantrolene
37. dopamine
38. dopamine agonists
39. methotrexate
40. no immunmodulat. therapy
41. occupational Therapy
42. regular cortisone

## Annex 2 – Exploratory objectives

Time permitting, predictions will also be produced using advanced machine learning methods and compared with the baseline estimates produced by the logistic regressions with elastic-net penalty. These methods will include Support Vectors Machines (SVM), Transfer Learning and generalised additive models. All methods will be subjected to the same rigorous cross-validation and cross-evaluation resampling methods as the logistic regression with elastic-net penalty. SVM (Cortes & Vapnik, 1995) is a highly flexible, non-parametric machine learning method which is widely employed to produce robust predictions in high dimensional, complex settings, including MS. Transfer Learning (Pan & Yang, 2010) seeks to borrow information from one, larger, domain and make predictions for a second, related, smaller domain in conjunction with domain specific information from the latter, smaller domain. This approach may be particularly appropriate in this context given that the treatment transition cohorts are relatively small compared to the much larger BRACE continuation cohort and there may be valuable information from the BRACE continuation cohort which could add explanatory power when explaining the variation in outcomes for the treatment transition cohorts. A generalised additive model is a linear model using custom designed base functions on top of the original predictors. Depending on the particular base function of choice, different desired qualities of the modelling can be achieved, such a prediction accuracy and robustness (Hastie & Tibshirani, 1990). Since there are all desired qualities for the decision support tool, it is worth exploring the effect of different base functions when training the model.